per
sge

Études rurales

La zone herbagère des Ardennes. Analyse typologique par matrice ordonnable

Jean-Daniel Gronoff

Citer ce document / Cite this document :

Gronoff Jean-Daniel. La zone herbagère des Ardennes. Analyse typologique par matrice ordonnable. In: Études rurales, n°43-44, 1971. pp. 170-214;

doi : https://doi.org/10.3406/rural.1971.1659;

https://www.persee.fr/doc/rural_0014-2182_1971_num_43_1_1659;

Fichier pdf généré le 16/05/2024

BY: $ = creative commons

JEAN-DANIEL GRONOFF

La zone herbagère des Ardennes

Analyse typologique par matrice ordonnable

Beaucoup d'utilisateurs de la représentation graphique ne conçoivent et n'utilisent l'image qu'en tant qu'illustration ou complément du texte. Le Laboratoire de Cartographie de l'E.P.H.E.-VIe Section considère que cette démarche doit être inversée. A cette fin, il a mis au point des techniques de travail où l'image permet d'appréhender de manière à la fois très complète et très souple un contenu informatif quel qu'il soit.

Loin de donner la primauté aux problèmes techniques (dessin et reproduction), une telle démarche amène tout utilisateur de « moyens graphiques » cohérents et fonctionnels à prendre du recul vis-à-vis d'eux, et à consacrer l'essentiel de son temps à la réflexion sur le seul contenu de son information, bien sûr, mais aussi sur son organisation.

Il s'agit d'abord de découvrir l'organisation interne de ce contenu, ensuite d'en interpréter les renseignements, enfin d'en communiquer les résultats.

La plupart des utilisateurs de « moyens graphiques » ignorent très souvent quel gain — quant à la compréhension et la communication du contenu de leur information — ils pourraient retirer de « moyens graphiques » employés de façon rigoureuse et méthodique, et non pas introduits uniquement sur la base soit d'une appréciation esthétique, soit du fait que « c'est plus parlant ».

Cet article s'attache dans un premier temps à donner les principes fondamentaux du traitement graphique de l'information sur un exemple très simple. Dans un deuxième temps, il en montre l'application sur un problème concret : la typologie de 300 communes de la zone herbagère du département des Ardennes*.

* Ce traitement a été effectué en collaboration avec J. BERTIN, directeur du Laboratoire ; M. BONNET, ingénieur OTAM-SARES ; J.-P. ARLIX, ingénieur OTAM-SARES, à la demande et avec le concours de la D.D.A. des Ardennes, qui en a autorisé l'utilisation pour un article essentiellement méthodologique.

Typologie de 7 communes avec 10 indicateurs

Établir une typologie communale revient à regrouper en types plus ou moins différenciés un ensemble de communes sur la base de *n* indicateurs. Le terme « indicateurs » sera préféré aux termes voisins « critères », « paramètres », « variables », à cause de sa généralité : la quantité de population, l'existence d'un bureau de poste, le remembrement sont des indicateurs.

Prenons un exemple fictif afin de saisir le rôle joué par l'image pour établir une typologie : l'utilisation de la perception visuelle pour traiter l'information constitue le domaine du traitement graphique de l'information.

Soit, pour 7 communes voisines, 10 indicateurs jugés suffisamment pertinents pour autoriser l'établissement d'une typologie. La meilleure façon de rassembler ces renseignements consiste à construire un tableau à double entrée (tableau des données). Les communes (colonnes) y sont numérotées de 1 à 7 (ordre numérique), les indicateurs (lignes) sont repérés par les lettres A à J (ordre alphabétique). Une croix (×) remplit la case correspondante chaque fois qu'il y a réponse à l'indicateur considéré, que sa formulation soit positive (A, B, C, F, H, J) ou négative (D, E, G, I). Dans ce dernier cas la croix signifie donc que pour telle commune, il n'y a pas de bureau de poste, pas d'école maternelle, pas de terrain de sport, pas d'école primaire...

1	2	3	4	5	6	7	COMMUNES
X			X				A CEG (Collège d'Enseignement Général)
	X	X		X		X	B REMEMBREMENT
X			X				C GARE SNCF
	X				X		D PAS DE BUREAU DE POSTE
					X		E PAS D'ECOLE PRIMAIRE
	X	X		X		X	F VETERINAIRE
	X				X		G PAS DE TERRAIN DE SPORT
X		X	X				H GENDARMERIE
	X				X		I PAS D'ECOLE MATERNELLE
	X	X		X		X	J COOPERATIVE AGRICOLE

Transcription graphique du tableau des données (Fig. 1).

Remplaçons la croix (×) par un carré noir. Cette simple opération améliore la perception du tableau des données, qui peut dès lors, être considéré comme une image, un graphique. Cette image sera appelée image initiale (Fig. 1).

Simplification de l'image initiale.

Elle s'opère en deux temps : tout d'abord par le reclassement des indicateurs d'après leur similitude (modification de l'ordre des lignes), puis par le reclassement des communes d'après leur similitude (modification de l'ordre des colonnes).

Reclassement des indicateurs semblables (Fig.1 → Fig. 2).

Rechercher les indicateurs semblables équivaut à trouver les lignes de distribution similaire dans l'image initiale. Une fois établies, ces similitudes permettent de redessiner une nouvelle image (Fig. 2) due à la seule modification de l'ordre des lignes, c'est-à-dire des indicateurs.

Processus de modification de l'ordre des lignes[1].

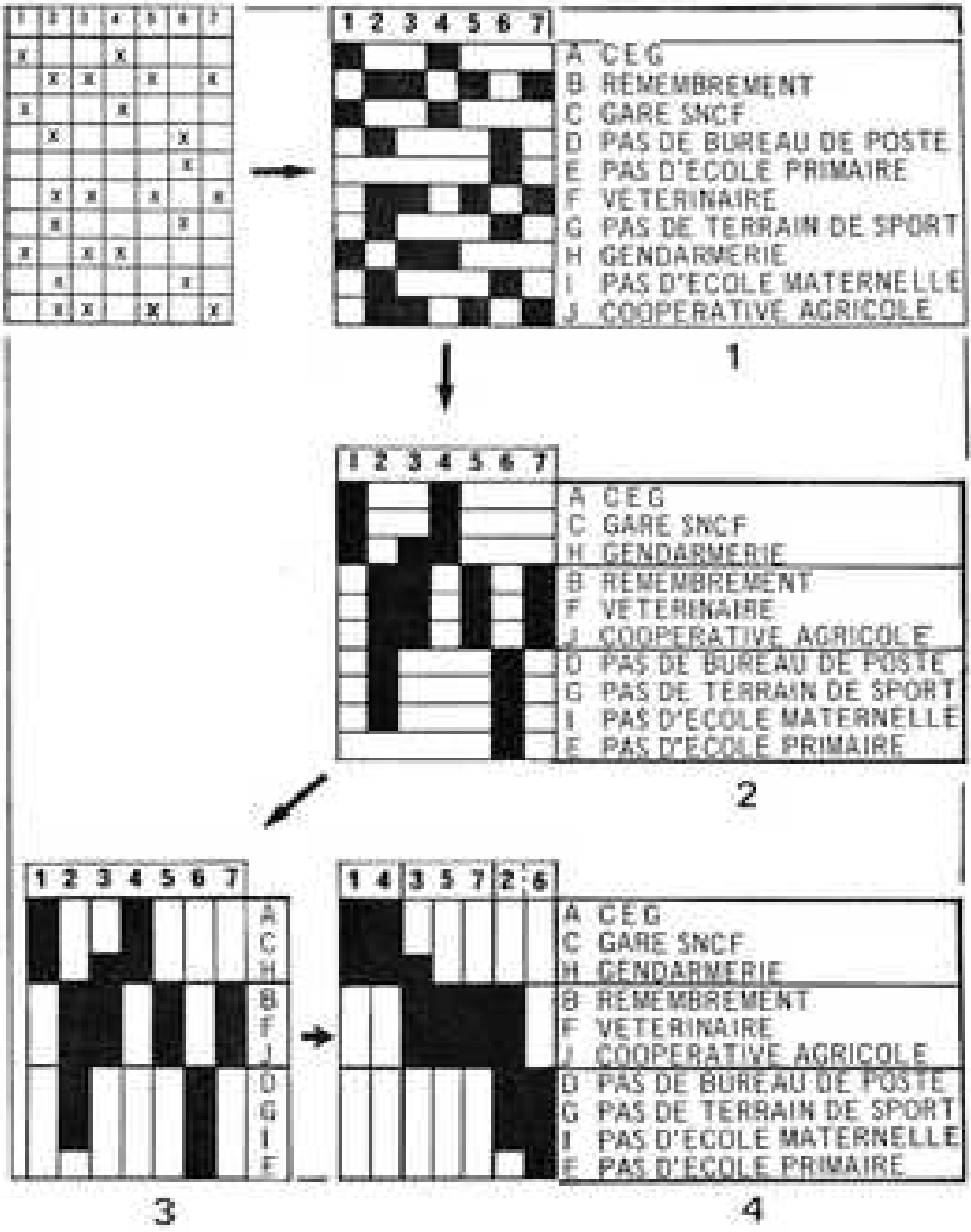
Commençons par la ligne A. Cherchons les lignes qui lui sont rigoureusement semblables : il n'y a que C. La ligne H lui étant presque semblable, ces trois lignes se classent dans l'ordre A, C, H, résultant du principe suivant : plus deux lignes sont similaires, plus elles doivent être proches l'une de l'autre. Continuons avec B. F et J sont identiques. L'ordre retenu devient alors A, C, H, B, F, J. E n'étant pas tout à fait semblable à D, G, I, il est classé en dernier. On redessine alors les lignes dans l'ordre A, C, H, B, F, J, D, G, I, E. La nouvelle image obtenue (Fig. 2) résulte de la seule modification du classement des lignes : chaque ligne est exactement similaire dans cette nouvelle image à sa correspondante de l'image initiale (Fig. 1).

Trois groupes apparaissent ainsi.

Reclassement des communes semblables (Fig. 3 → Fig. 4).

Ce reclassement doit répondre à la question suivante : quelles sont les communes formant un même type ? Pour cela, il suffit de lire l'image précédente (Fig. 2) en colonnes (Fig. 3) et de modifier l'ordre des colonnes suivant le processus déjà adopté pour les lignes. La nouvelle image obtenue (Fig. 4) s'appellera « image simplifiée ». On vérifie alors d'une part que chaque colonne de cette image simplifiée est exactement similaire à la colonne correspondante de l'image précédente (Fig. 3), d'autre part, que les renseignements portés dans chaque case de l'image simplifiée (Fig. 4), sont rigoureusement identiques à ceux de l'image initiale (Fig. 1). Au terme de ce reclassement, aucun renseignement n'a été perdu, aucun renseignement n'a été rajouté.

1. On dira également permutation.


1 2 3 4 5 6 7
A CEG
B REMEMBREMENT
C GARE SNCF
D PAS DE BUREAU DE POSTE
E PAS D'ECOLE PRIMAIRE
F VETERINAIRE
G PAS DE TERRAIN DE SPORT
H GENDARMERIE
I PAS D'ECOLE MATERNELLE
J COOPERATIVE AGRICOLE
1
1 2 3 4 5 6 7
A CEG
C GARE SNCF
H GENDARMERIE
B REMEMBREMENT
F VETERINAIRE
J COOPERATIVE AGRICOLE
D PAS DE BUREAU DE POSTE
G PAS DE TERRAIN DE SPORT
I PAS D'ECOLE MATERNELLE
E PAS D'ECOLE PRIMAIRE
2
1 2 3 4 5 6 7
A C H
B F J
D G I E
3
1 4 3 5 7 2 6
A CEG
C GARE SNCF
H GENDARMERIE
B REMEMBREMENT
F VETERINAIRE
J COOPERATIVE AGRICOLE
D PAS DE BUREAU DE POSTE
G PAS DE TERRAIN DE SPORT
I PAS D'ECOLE MATERNELLE
E PAS D'ECOLE PRIMAIRE
4

Utilisation de l'image simplifiée (interprétation).



Analyse et définition des types de communes.

a) La présence de services à caractère urbain (indicateurs A, C, H) rassemble les communes 1 et 4 dans le même type « villes ».

b) Les communes 3, 5, 7 constituent pour leur part un type caractérisé par la concentration de services agricoles (B, F, J) qui peuvent définir des « bourgs actifs ». La commune 3 a un élément commun (gendarmerie) avec le type précédent, mais ceci peut être considéré comme secondaire. Il est toujours possible, si ce détail est jugé important, d'isoler la commune 3, d'en faire un type. Seul le chercheur peut prendre cette décision[1].

c) Les communes 2 et 6 forment le troisième type. Il est parfaitement logique ici de discuter ce regroupement qui résulte d'un choix effectué par le chercheur, choix qui lui est soumis par l'image simplifiée. En effet, la commune 2 appartient à la fois au deuxième et troisième type. Devant cette situation, le chercheur a le choix entre trois solutions : ou bien la commune 2 forme un type en elle-même, ou bien elle est rattachée au deuxième type, ou bien à la commune 6. Cette dernière formule a été adoptée ici, car on a estimé que l'aspect « petits villages » était le plus important.

L'image simplifiée ne fait cependant pas disparaître les cas spéciaux. Elle les met au contraire en valeur et les présente d'une façon telle, que le chercheur effectue son choix sur des bases plus solides. Ce choix à faire n'était pas perceptible dans l'image initiale.

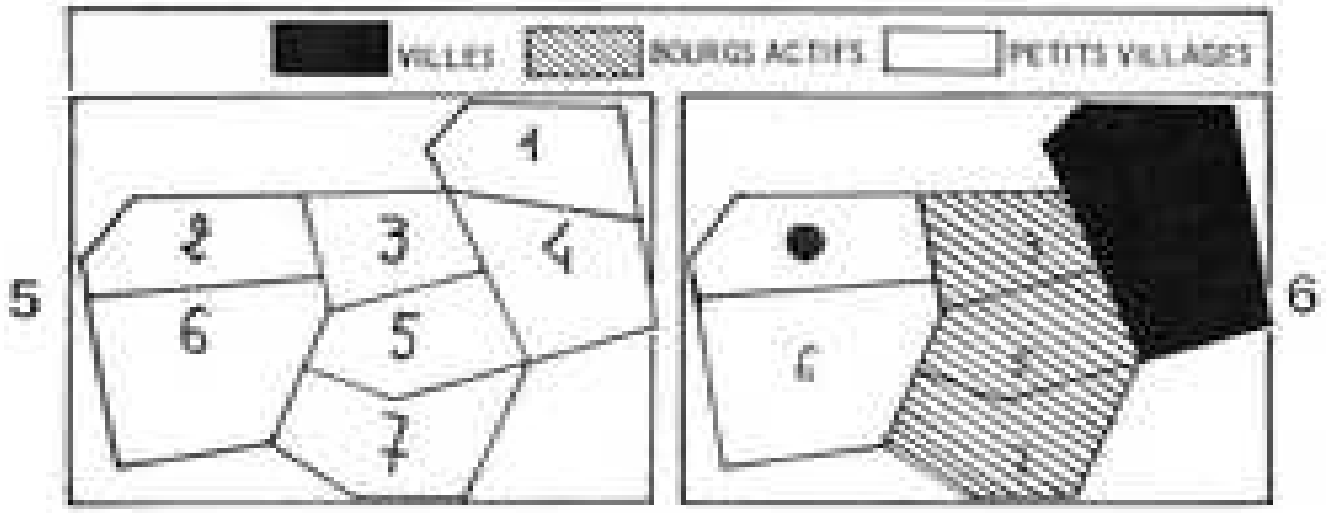
1. On entend par chercheur, celui qui a collecté les données et les soumet à un traitement de l'information : c'est le spécialiste du contenu de l'information à traiter, par opposition à l'informaticien, spécialiste des méthodes de traitement.

Acquisition du « facteur » dominant de la typologie.

La permutation des communes et des indicateurs aboutit à un ordre qui a une signification : les communes les plus urbaines (1, 4) sont à gauche dans l'image simplifiée, les communes les moins urbaines (2, 6) donc les plus rurales, à droite, les communes du deuxième type (3, 5, 7) se trouvant bien entre les deux extrêmes. Pour les indicateurs, il en va de même, mais de haut en bas.

La signification globale de cette typologie devient très claire : opposition « urbain-rural ». Chaque type est alors parfaitement différencié des autres par sa position dans cet ordre d'ensemble.

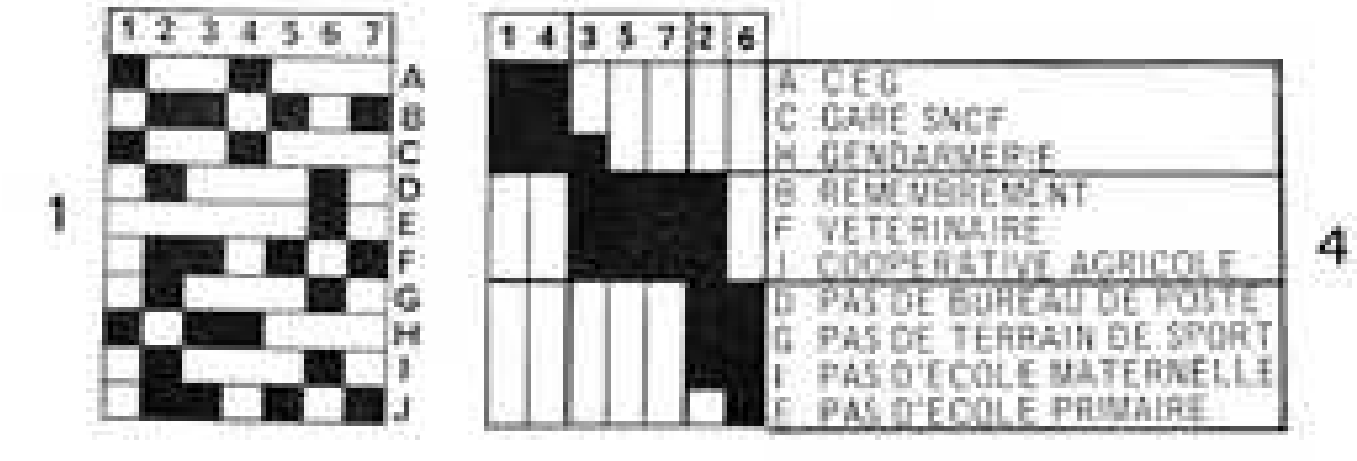
Transcription cartographique.



Chaque commune est repérée sur la carte par son numéro (Fig. 5). La transcription cartographique des trois types définis par l'image simplifiée devient automatique. La typologie étant globalement fondée sur un ordre urbain-rural, les symboles adoptés doivent transcrire rigoureusement cet ordre. Les trois paliers (noir, gris, blanc) sont ordonnés, car le noir et le blanc représentent la plus forte opposition graphique disponible. La correspondance est parfaite (Fig. 6) : noir pour le type « villes », gris (hachures) pour le type « bourgs actifs », blanc pour le type « petits villages ».

Le cas spécial de la commune 2 est signalé par un point noir, ce qui permet toutefois de l'assimiler au type « petits villages ». La carte réalisée est donc rigoureusement conforme à l'interprétation synthétique de l'image simplifiée (Fig. 4). La carte permet de vérifier la typologie obtenue : deux communes semblables *sur cette carte*, sont proches l'une de l'autre. (Il est bien évident que ceci ne se vérifie pas pour toutes les cartes !) Cette propriété de la carte sera appelée « probabilité d'homogénéité liée à la proximité géographique » dans les exemples qui suivront.

Utilité de la simplification de l'image initiale.



— *La simplification de l'image aboutit à une mémorisation plus rapide et plus facile de l'ensemble de l'information.* L'image initiale, comme l'image simplifiée, comporte 70 cases (7 communes × 10 indicateurs = 70 cases), chaque case contenant une information élémentaire (noir ou blanc). La comparaison (et donc la mémorisation) du contenu de ces 70 cases est longue et fastidieuse dans l'image initiale. La mémorisation de l'image simplifiée est très supérieure : on perçoit 3 groupes au lieu de 70 cases ; ces groupes sont mémorisés dans un ordre significatif.

— *La simplification est obtenue sans perte d'information.* L'image simplifiée comporte rigoureusement les mêmes renseignements que l'image initiale et, dans cet exemple, le tableau des données.

— *La simplification de l'image initiale apporte un gain d'information.* Prenons la case H 3 : dans l'image *initiale,* elle a autant d'importance que les 69 autres ; sa position est neutre ; la modification de sa position (image simplifiée) permet au chercheur de la situer par rapport à toutes les autres, c'est-à-dire permet de la qualifier et de définir son originalité ; sa position est significative, elle est « chargée de sens ». Cette position significative peut être longuement commentée : elle individualise la commune 3, qui possède tous les indicateurs du type « bourgs actifs », plus un indicateur du type « villes ». Le chercheur peut, soit considérer cette case H 3 comme un détail, soit chercher à expliquer cette exception. Mais il est amené à ce choix seulement grâce à la simplification de l'image initiale.

L'utilité du traitement de l'information et son efficacité se mesurent en « gain d'information » : gain dans la mémorisation ; gain dans la connaissance du contenu de l'information : acquisition de la signification globale de la typologie, position significative de chacune des cases ; gain, pour le chercheur, dans l'utilisation de sa formation et de son expérience : l'image simplifiée lui soumet des situations nécessitant des choix, qu'il résoudra par des recherches nouvelles et par son imagination.

Les cinq opérations élémentaires d'une typologie communale.

Le processus adopté pour l'élaboration de la typologie des 7 communes de l'exemple précédent se décompose en cinq opérations élémentaires :

1. construction du tableau des données ;
2. transcription graphique du tableau des données ;
3. simplification de l'image initiale ;
4. interprétation de l'image simplifiée ;
5. communication des résultats.

Dans l'exemple de traitement qui va suivre, ces cinq opérations seront effectuées deux fois, d'abord pour une image (on dira « matrice ») de 5 indicateurs × 300 communes, dite matrice 300, ensuite pour une matrice de 101 communes × 92 indicateurs, dite matrice 100 × 100[1].

Les cinq opérations se succèdent toujours dans le même ordre, mais chacune d'elles est susceptible de prendre plus ou moins d'importance en fonction du problème posé, de la disponibilité du chercheur et de la dimension de la matrice.

Ce dernier sera pratiquement le seul responsable de l'opération située en début de traitement (construction du tableau des données) et des deux opérations de fin de traitement : interprétation de la matrice simplifiée et communication des résultats. L'exemple de traitement qui va suivre s'attache à montrer la part du chercheur dans ces trois opérations les plus longues en temps.

Une matrice de 100 lignes sur 100 colonnes va diminuer le rôle des opérations 2 et 3, qui vont durer peu de temps (8 jours au total par exemple) par rapport aux autres opérations qui demanderont au total entre 6 mois et 2 ans, et même plus. Transcription graphique du tableau des données et simplification de l'image initiale correspondent donc à des opérations secondaires, du point de vue du chercheur.

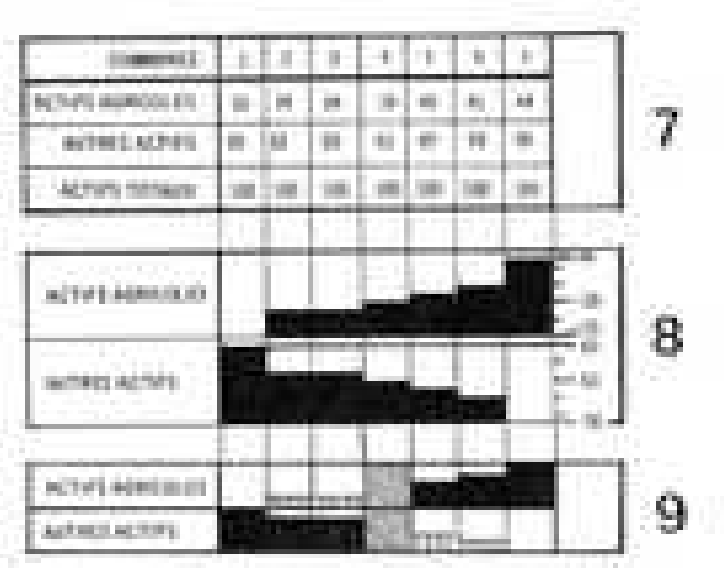
De ce fait, il ne sera donné qu'un aperçu très succinct des règles de transcription graphique et des capacités des matériels de permutation.

Opération 2 : transcription graphique du tableau des données.

Elle est totalement différente selon qu'il s'agit d'indicateurs qualitatifs ou quantitatifs.

Indicateurs qualitatifs. On appelle qualitatifs les indicateurs qui n'admettent qu'une seule réponse : oui ou non. Cette réponse est transcrite par un carré noir qui, pour des raisons de lisibilité et d'homogénéité avec la transcription des données numériques, est plus efficace que la simple croix.

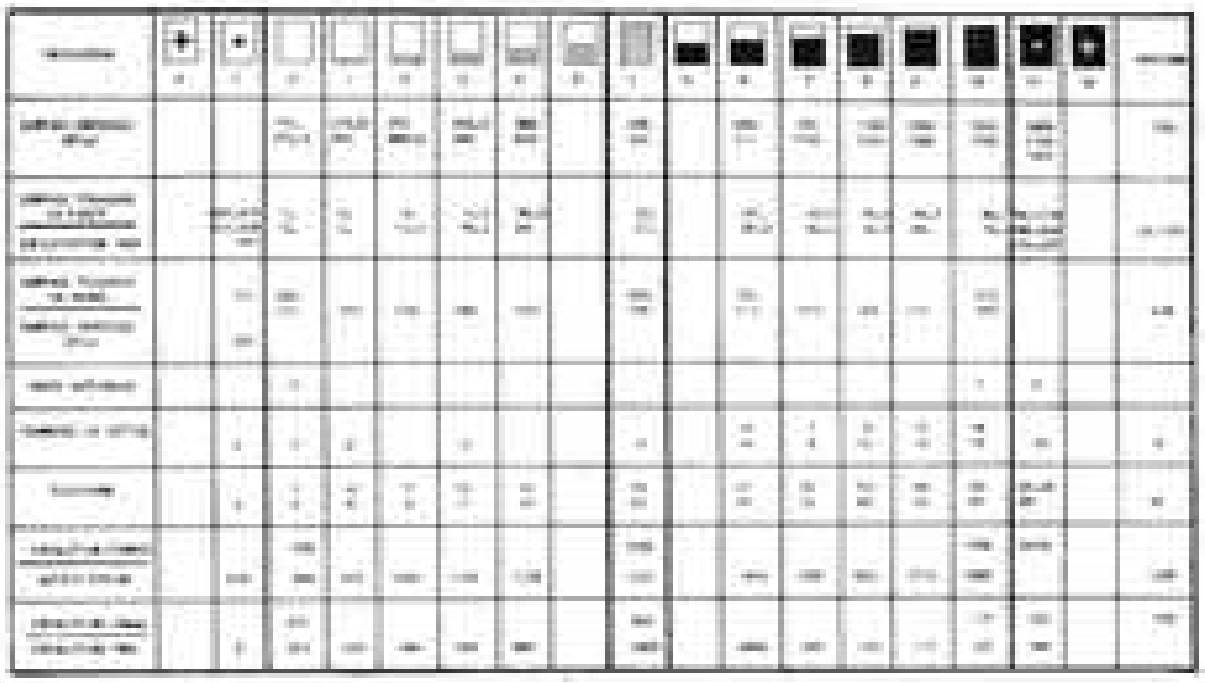
1. La dimension d'une matrice se mesure au nombre de lignes et de colonnes.



Indicateurs quantitatifs. La Figure 7 donne les valeurs chiffrées de deux séries statistiques prises comme exemple : la part des actifs agricoles, la part des autres actifs pour 100 actifs par commune. La première transcription (Fig. 8) représente les deux valeurs extrêmes de chaque série par le noir et le blanc (communes 1 et 7) et la part du noir décroît ou croît proportionnellement aux valeurs chiffrées. Cette représentation sera adoptée pour la matrice 300. La deuxième transcription (Fig. 9), adoptée pour la matrice 100 $\times$ 100, prend comme référence la valeur moyenne de chaque série transcrite en gris. Les valeurs situées au-dessus de la moyenne (et non de la médiane) sont représentées par une variation proportionnelle de noir (paliers noirs), les valeurs situées en dessous de la moyenne par une variation décroissante du gris, les deux extrêmes étant noir et blanc. Cette représentation permet de repérer les valeurs moyennes (amélioration de la lisibilité). La gamme de 11 paliers est la plus usitée : 5 noirs, 5 blancs, 1 gris moyen. Mais elle n'est pas limitative : il existe des gammes de 27 ou de 7 paliers, dont l'utilisation est liée d'une part au matériel utilisé, d'autre part à la nature du problème posé. La définition des limites de paliers dépend avant tout de l'analyse préalable des données, mais aussi des règles de perception visuelle, qu'il ne faut jamais transgresser.

Quelle que soit la représentation adoptée, toute transcription d'une série statistique numérique doit permettre : la mise en ordre des séries, ce qui est le cas de l'exemple de la Figure 7 ; la perception des similitudes (l'inversion entre les actifs agricoles et les autres actifs est perçue dans les deux représentations) ; le mélange d'indicateurs qualitatifs et quantitatifs (le oui ou non correspond aux deux extrêmes d'une série chiffrée).

Tout tableau des données transcrit graphiquement est accompagné d'un tableau de réduction en paliers similaire à celui de la Figure 10.



10

Opération 3 : simplification d'une matrice ; le matériel de permutation.

L'application rigoureuse et automatique de la loi de proximité-similitude (tout ce qui est semblable est proche dans le plan) conduit à la simplification de la matrice, que le chercheur peut contrôler puisqu'il en a l'image sous les yeux : c'est un automatisme contrôlé, « débrayable », qui aboutit à l'optimisation de la matrice.

Redessiner une matrice de 7 × 10 (70 cases) est une opération rapide. Redessiner une matrice de 5 × 300 (1 500 cases) ou de 100 × 100 (10 000 cases) est impossible. Un matériel spécial est nécessaire : il a pour fonction de permettre toutes les permutations tout en ne faisant « composer » (dessiner) la matrice qu'une seule fois. Pour une matrice 100 × 100, il suffit d'aligner dans l'ordre du tableau des données 10 000 « dominos », 10 000 « cubes » ayant sur leur face supérieure le symbole graphique convenu. Chaque domino est percé sur les côtés afin de pouvoir être tringlé dans le sens des lignes ou dans le sens des colonnes. Pour permuter les lignes ou les colonnes, il suffit alors de tringler ces lignes ou ces colonnes.

Chaque nouvelle image établie est photographiée. Les dimensions d'une matrice déterminent donc le support de permutation : papier pour les petites dimensions (10 × 10), matériel mécanique — dominos — jusqu'à 150 × 100, écran cathodique (50 × 100) dans le cas où 7 paliers seulement sont nécessaires. Un matériel en cours d'expérimentation, en 2 paliers (noir-blanc), doit permettre la permutation de 400 lignes × 200 colonnes (80 000 cases).

Typologie des 300 communes de la zone herbagère du département des Ardennes

Nécessité d'une typologie élaborée en deux phases.

La Direction départementale de l'Agriculture (D.D.A.) des Ardennes, afin de mieux adapter ses interventions à la situation locale, a jugé



11

utile de disposer d'une typologie des 300 communes de la zone herbagère de ce département, la ville de Sedan en étant exclue (Fig. 11).

La D.D.A. a souhaité élaborer cette typologie, à titre expérimental, à l'aide du traitement graphique.

Certaines contraintes matérielles ont été posées au départ : avoir un grand nombre d'indicateurs (une centaine) afin de ne pas obtenir une typologie élémentaire ; utiliser au maximum les statistiques déjà existantes (I.N.S.E.E., R.G.A.) et avoir le minimum de recherches à effectuer ; limiter le nombre de renseignements à collecter au strict nécessaire ; n'effectuer que les calculs réellement utiles.

100 indicateurs $\times$ 300 communes $=$ 30 000 données.

Traiter une telle masse d'informations pose un problème de temps, de coût et de capacité matérielle à tous les stades du traitement, considéré comme trop lourd en l'occurrence.

C'est ainsi qu'une réduction du nombre de communes en deux phases a été adoptée. La première phase a pour objectif de permettre le choix de 100 communes représentatives (matrice 300 : 5 indicateurs $\times$ 300 communes $=$ 1 500 données).

Dans la seconde phase est établie la typologie des communes représentatives, et sont opérées la vérification de la réduction et l'extrapolation des résultats aux 300 communes (matrice 100 $\times$ 100 $=$ 10 000 données).

Cette réduction préalable du nombre des communes permet d'une part un gain de temps remarquable lors des cinq opérations de traitement : fondamentalement, une matrice 100 $\times$ 100 s'avère d'une facilité d'emploi supérieure à une matrice 300 $\times$ 100. D'autre part, la capacité matérielle des matrices mécaniques ne dépassait pas, à l'époque de ce traitement, 150 lignes sur 100 colonnes, ce qui obligeait donc à cette réduction préalable.

Toutefois, le gain de temps devient inestimable surtout lors de la collecte des données sur le terrain. La présentation des séries chiffrées n'est pas toujours homogène d'une série statistique à une autre. Cependant, cette homogénéité se rencontre bien sûr plus fréquemment pour 100 communes que pour 300. Il vaut mieux également, en cas de nécessité, avoir 100 nombres plutôt que 300 à relever à la main : les risques d'erreur diminuent en proportion. Enfin, trouver des renseignements originaux pose moins de problèmes une fois effectuée la réduction du nombre de communes.

A condition de s'assurer du bien-fondé de l'extrapolation, le gain de temps est donc réellement appréciable : le résultat obtenu à partir de près de 10 000 données (1 500 dans la phase 1 et 8 500 dans la phase 2) sera voisin de celui obtenu à partir de 30 000 données (300 $\times$ 100).

Phase 1 : choix de 100 communes représentatives (matrice 300).

Cinq critères sont choisis pour la facilité de leur collecte et, *a priori*, pour leur pertinence : part de la patente dans le centime 1969 (P), équipement en eau de la commune, part des ménages agricoles pour 100 ménages ordinaires (A), solde migratoire négatif (S —) et positif (S +), densité de population 1968 (D) (Fig. 12).

Simplification de l'image (Fig. 12 → Fig. 13).

1) La matrice est divisée en deux sous-ensembles : solde migratoire positif porté à gauche ; solde négatif à droite. Ils seront traités séparément.

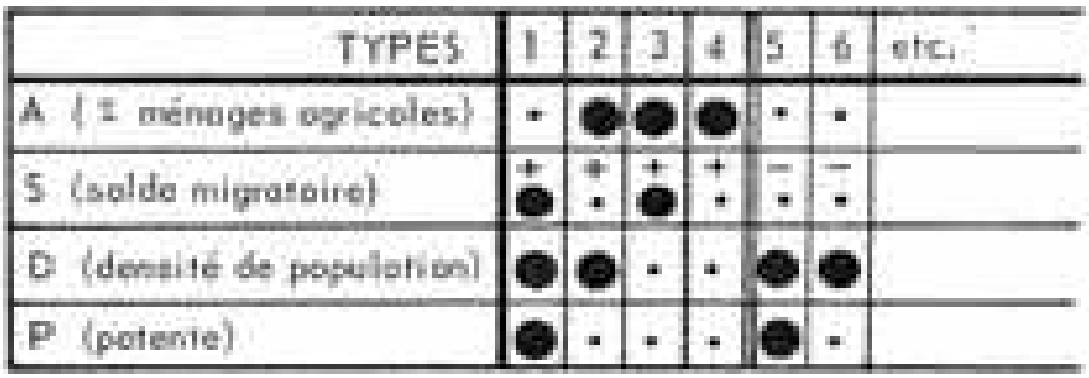
2) La permutation des colonnes fait apparaître la similitude entre D (densité de population) et P (patente) d'une part, et entre A (% de ménages agricoles) et S (solde migratoire) d'autre part.

3) La permutation des lignes permet de former les groupes DP et AS. Ils ont une distribution opposée (corrélation négative).

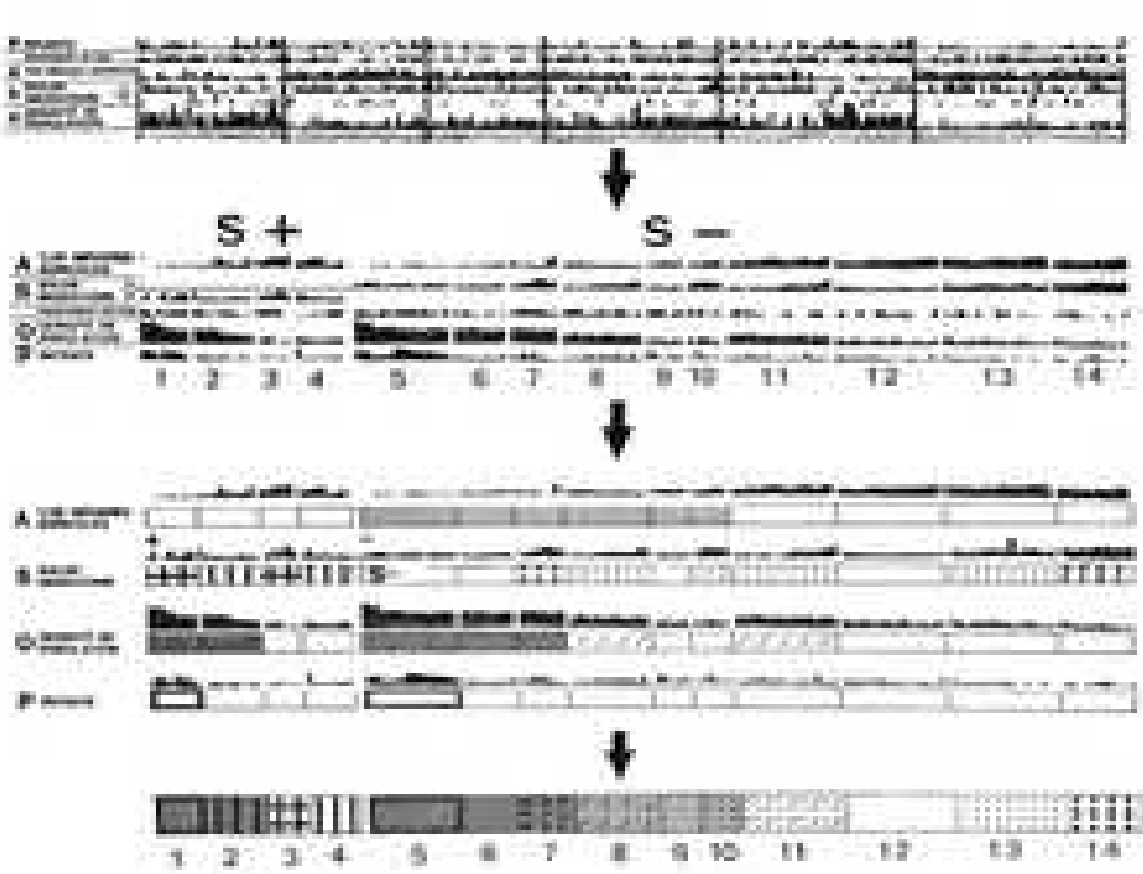
4) La ligne qui transcrit « l'équipement en eau » est conservée au centre pour juger de son parallélisme éventuel avec l'un ou l'autre groupe. La non-apparition d'un parallélisme suffisant permet, en l'occurrence, de conclure à la non-pertinence de cet indicateur et à son retrait (distribution aléatoire par rapport à l'information considérée).

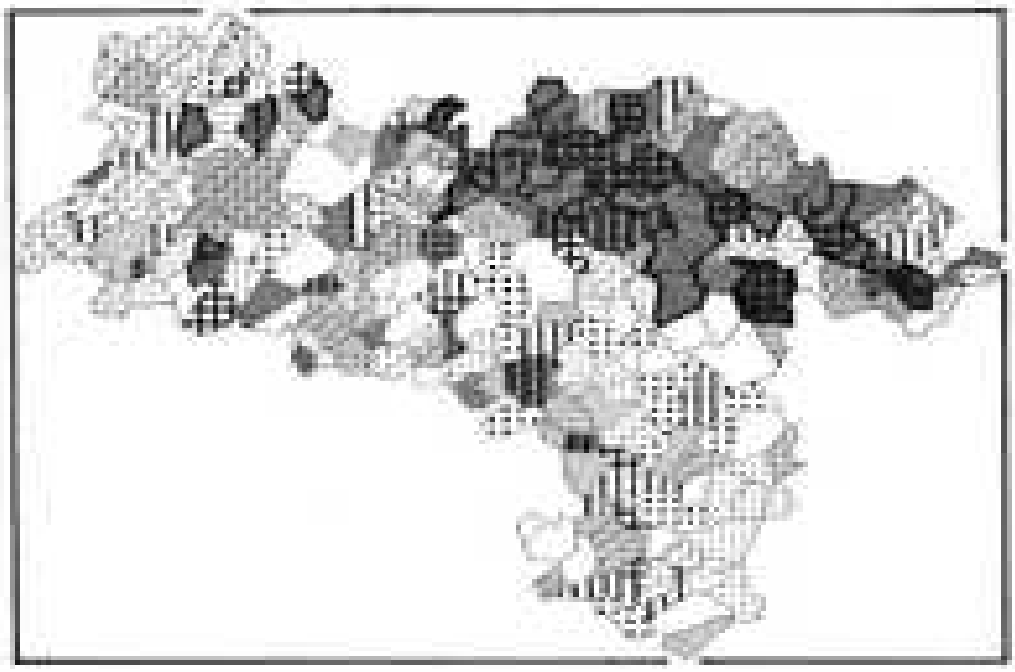
Interprétation et typologie (Fig. 14).

Chaque type est défini à partir de la combinaison des indications de A, S, D, P.



Les 14 types se retrouvent pratiquement ordonnés de l' « urbain » au « rural ».


12
S +
S −
13
1 2 3 4 5 6 7 8 9 10 11 12 13 14
14
15
1 2 3 4 5 6 7 8 9 10 11 12 13 14


16

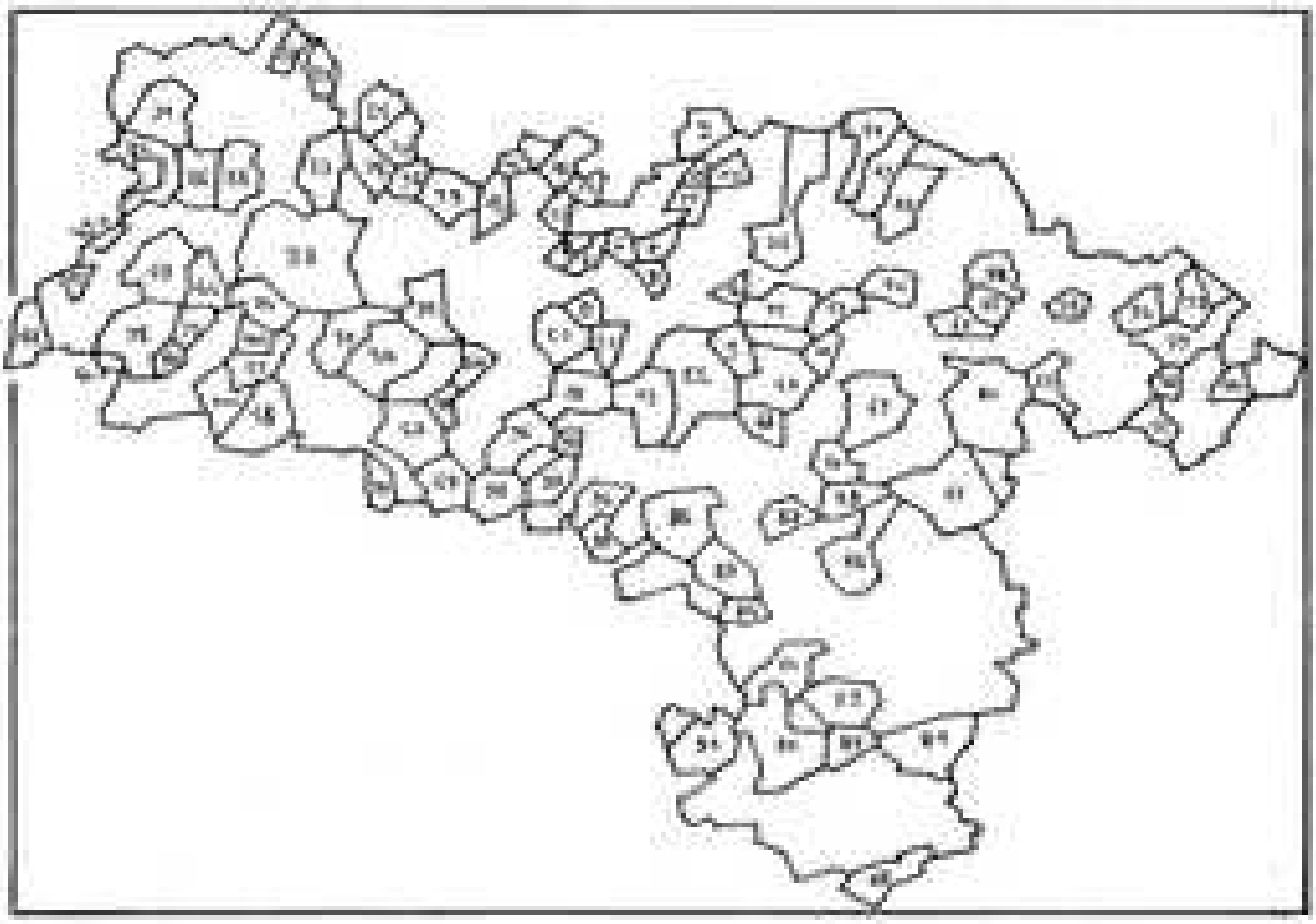
Cartographie (Fig. 14, 15 et 16).

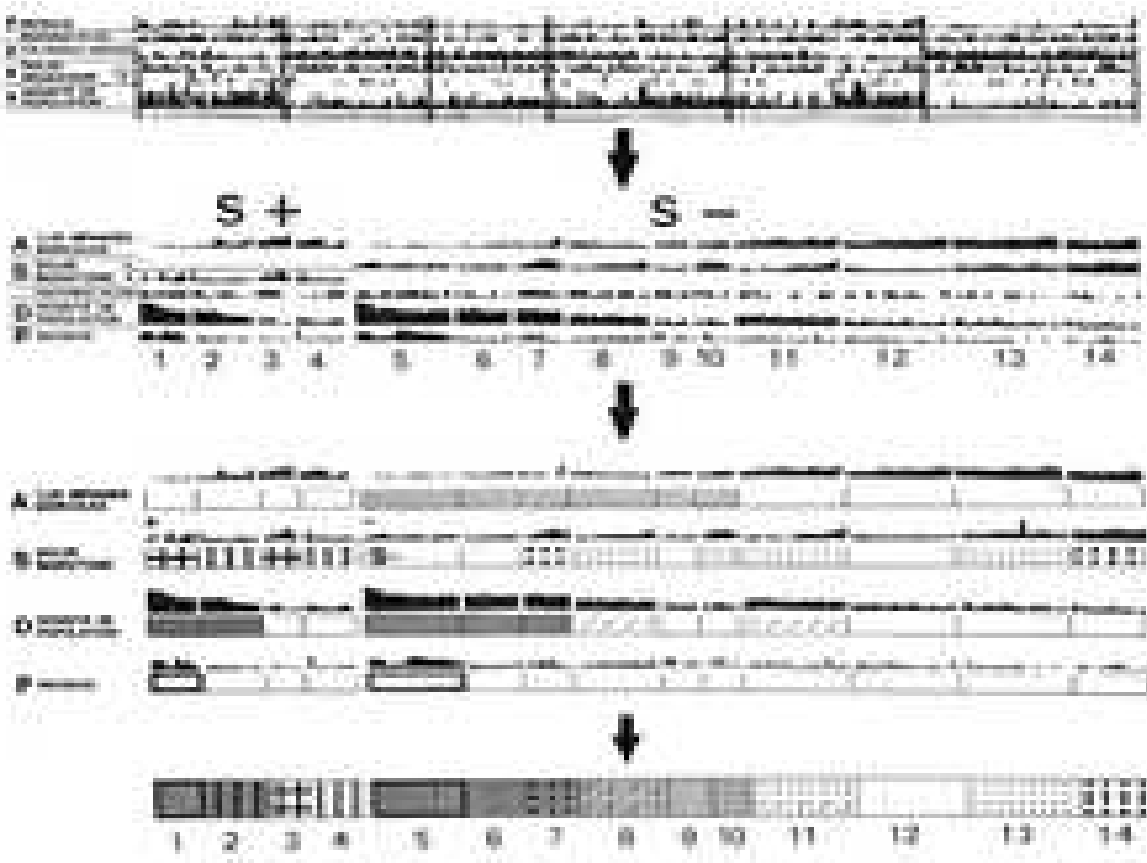
La légende détaillée (Fig. 15) résulte de la combinatoire ci-dessus. On définit, par indicateur (Fig. 14), les paliers nécessaires et suffisants pour que leur combinatoire sépare les 14 types. On choisit des signes tels que leur superposition construise un ordre visuel noir → blanc homologue à l'ordre « urbain → rural » reconnu sur la matrice et sépare efficacement les types S + et S —.

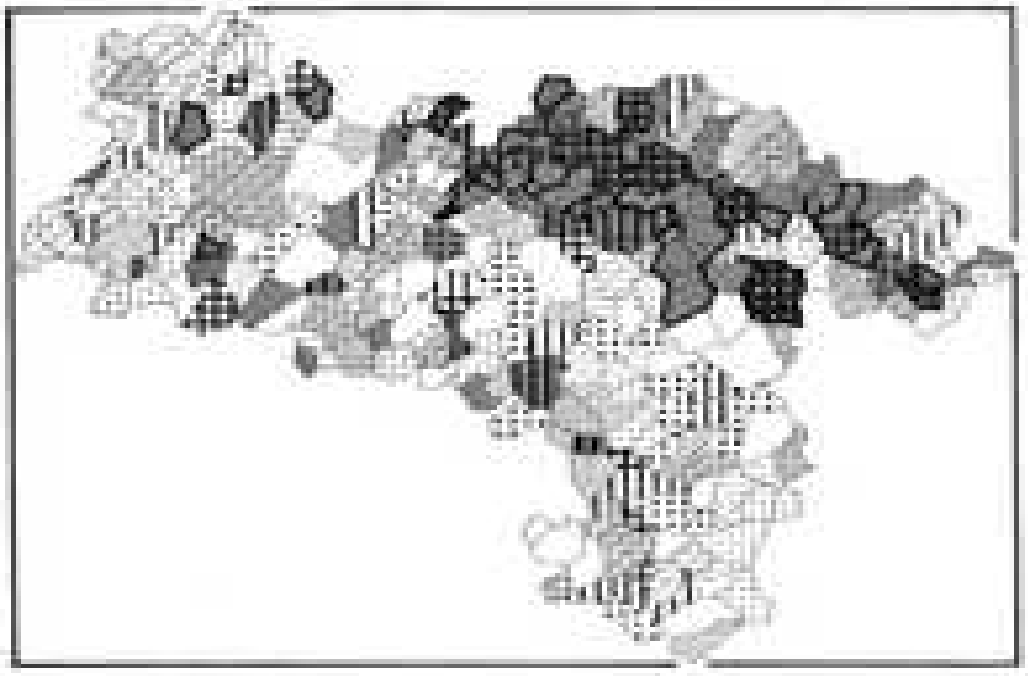
Choix de 100 communes représentatives des 300 (Fig. 17).

Ce choix est fondé en réalité sur 6 indicateurs. Aux quatre indicateurs A, S, D, P retenus pour la typologie, la cartographie introduit la notion de « proximité géographique » et en même temps le coefficient d'homogénéité lié à cette notion. Enfin le choix final, effectué par le spécialiste de la D.D.A. a fait intervenir, outre son expérience de la zone herbagère, la quantité de population, sixième indicateur. Il en résulte que si la moyenne est d'une commune sélectionnée pour trois, certaines n'en représentent aucune autre, tandis qu'une seule peut en représenter sept.

17



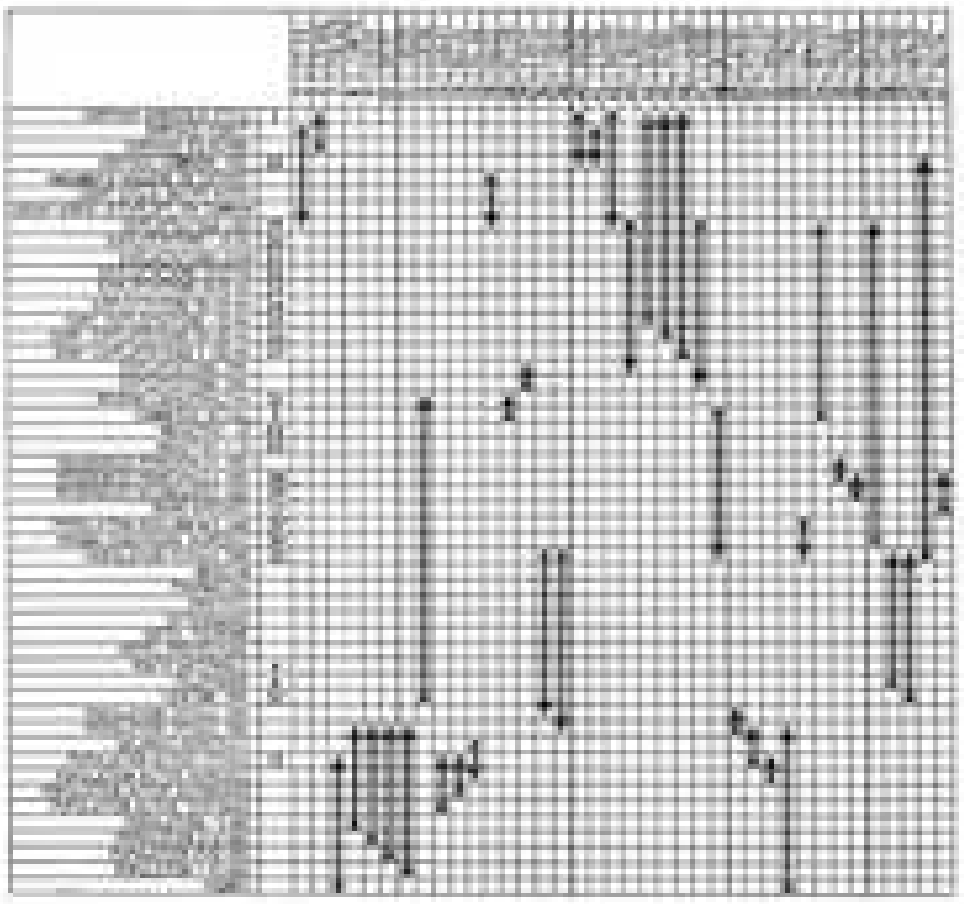

12
S +
S —
13
1 2 3 4 5 6 7 8 9 10 11 12 13 14
14
15
1 2 3 4 5 6 7 8 9 10 11 12 13 14


16

Phase 2 : typologie de 101 communes avec 96 indicateurs.

Construction de la matrice initiale.

Choisir 100 indicateurs signifie en principe trouver 100 indicateurs pertinents *a priori*, en tenant compte d'une part de la relation temps de collecte–pertinence supposée (indicateurs « coûteux » en temps), d'autre part de la relation temps de collecte–masse de calculs à effectuer.



18

La Figure 18 dite « tableau de création d'indicateurs » permet de mieux contrôler ces relations. C'est un instrument de travail qui a trois fonctions :

— Limiter la collecte des données à l'essentiel : 38 indicateurs nouveaux (en haut de gauche à droite), dits « indicateurs créés », sont obtenus par calcul de 52 indicateurs réellement collectés, dits « indicateurs de base » (à gauche, de haut en bas). De ces 52 indicateurs nécessaires au calcul, seuls 27 seront conservés pour le traitement (indicateurs de base numérotés).

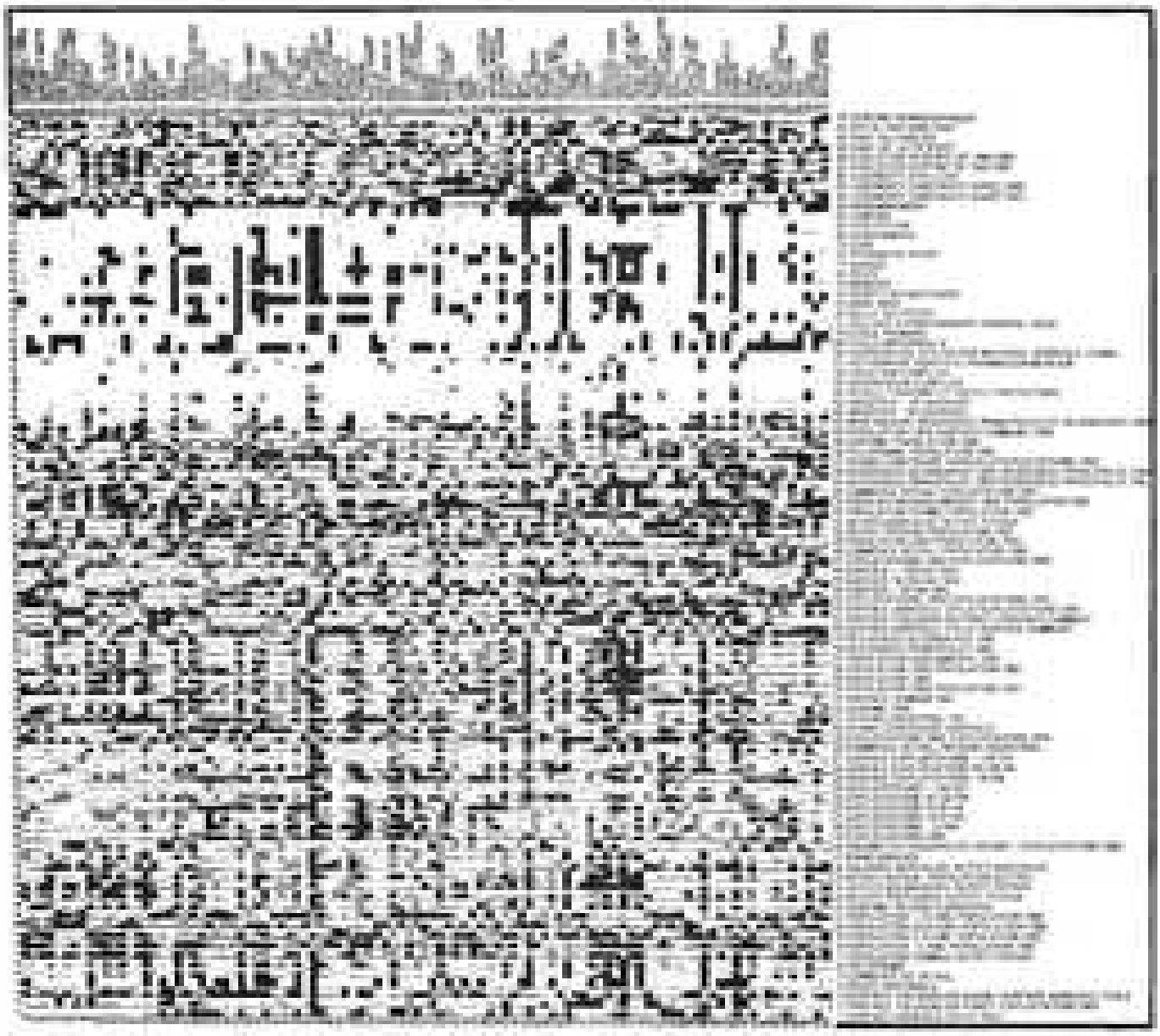
— Assister l'imagination : la vision d'ensemble de tous les rapports possibles aide à penser à des rapports nouveaux, à éliminer les répétitions, à contrôler la signification des indicateurs (relation temps de collecte–pertinence supposée).

— Servir de référence : le départ de la flèche indique le numérateur ; l'arrivée, le dénominateur du rapport effectué.

L'ordinateur effectue les calculs, construit le tableau des données qui, transcrit graphiquement, donne la matrice initiale (Fig. 19). Les communes sont rangées par canton, les indicateurs dans l'ordre de collecte.

La liste des 92 indicateurs est constituée par des indicateurs de base conservés, des indicateurs créés et des indicateurs qualitatifs (lignes 64 à 83) ; 4 autres indicateurs seront rajoutés en cours de traitement.

19



Optimisation de la matrice (Fig. 20).

L'optimisation est obtenue en deux temps :

— un temps où s'applique la loi de proximité-similitude (automatisme visuel = simplification) ;

— un temps où le chercheur optimise la matrice en effectuant les choix que lui a soumis la simplification. Ce second temps aboutit au découpage des lignes et des colonnes.

La Figure 20 montre directement la matrice optimisée, l'opération de simplification, nécessaire avant toute optimisation, ne présentant qu'un intérêt mineur, du fait de sa technicité.

L'optimisation de la matrice conduit à retenir 8 types de communes et 20 groupes d'indicateurs. Un groupe d'indicateurs de distribution voisine est dénommé « facteur » et repéré par une lettre alphabétique. En A, les quatre indicateurs correspondent à ceux de la matrice 300. Il manque l'équipement en eau et les ménages agricoles ; B, C, D, E donnent une image cohérente, diagonalisée, où se perçoivent des ensembles, des « paquets » de noir et de blanc assez compacts ; F, G, H ont en revanche une distribution très incohérente, dite « aléatoire » ; I et J ont été mis de côté à la demande de la D.D.A.

L'optimisation de la matrice aboutit donc à un ensemble de 62 indicateurs (A à E) réellement simplifié, donnant une image cohérente (il sera appelé « système 1 ») et à un ensemble de 21 indicateurs de distribution « aléatoire » dénommé « système 2 ».

La Figure 20 correspond en réalité à la seule optimisation du système 1. Les colonnes blanches séparant les 8 types de communes se trouvent toujours justifiées par une alternance « noir-blanc » d'au moins l'un des facteurs B, C, D ou E (suivre par exemple la colonne de séparation entre les types de communes 4 et 5).

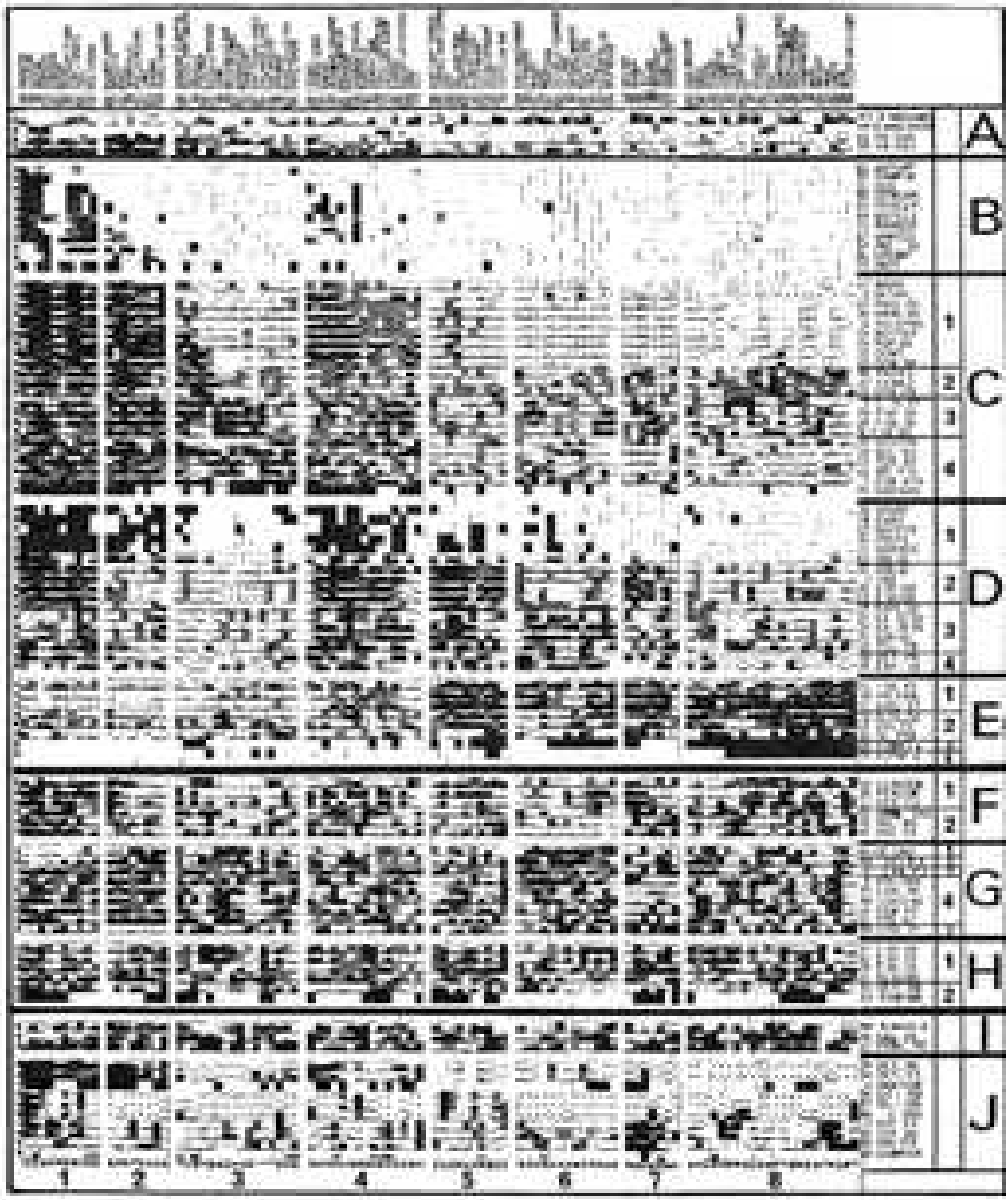
Les mêmes colonnes ne séparent rien dans le système 2, surtout pour les facteurs G et H, qui apparaissent comme des bandes noires, avec des blancs parsemés de façon aléatoire.

Le rangement des communes et des 8 types définis par les séparations est chargé d'une signification caractéristique du seul système 1 : l'opposition « urbain-rural », ce qui n'a rien d'inattendu, bien sûr ! Dans le type 1 (en haut à gauche) se trouvent les communes importantes dont 5 chefs-lieux de canton. A l'opposé, type 8 (à droite), se regroupent les villages très « agricoles ».

Le rangement des indicateurs de ce même système 1 va des indicateurs « urbains » (B et C : CEG, Banque, médecins, quantité de population, etc.) aux indicateurs ruraux (D et surtout E : absence d'école primaire, part du foncier, actifs agricoles, etc.).

L'optimisation du seul système 2, effectuée indépendamment, amènera à détruire l'ordre du système 1, pour en obtenir un autre dont il faudra découvrir la signification.

20



Matrice d'interprétation du système 1.

La Figure 21 est établie par le chercheur directement à partir de la matrice optimisée et ne concerne donc ici que le système 1. La matrice d'interprétation généralise les résultats de l'optimisation, en réduisant le nombre d'informations à mémoriser (perte d'information) au prix d'un enrichissement de la connaissance du contenu (gain d'information).

Il y a *perte d'information*, par rapport à la matrice optimisée. Les 100 communes sont « ramenées » d'abord à 15 types (les colonnes de la Figure 21), puis à 8 types. De même, les 62 indicateurs sont réduits d'abord à 17 « groupes » (les lignes de la Figure 21), puis à 5 facteurs. Enfin, les 11 paliers de la matrice optimisée sont « remplacés » par 4 « appréciations » (fort, moyen, faible, aléatoire), plus grossières, bien sûr, mais d'une mémorisation aisée.

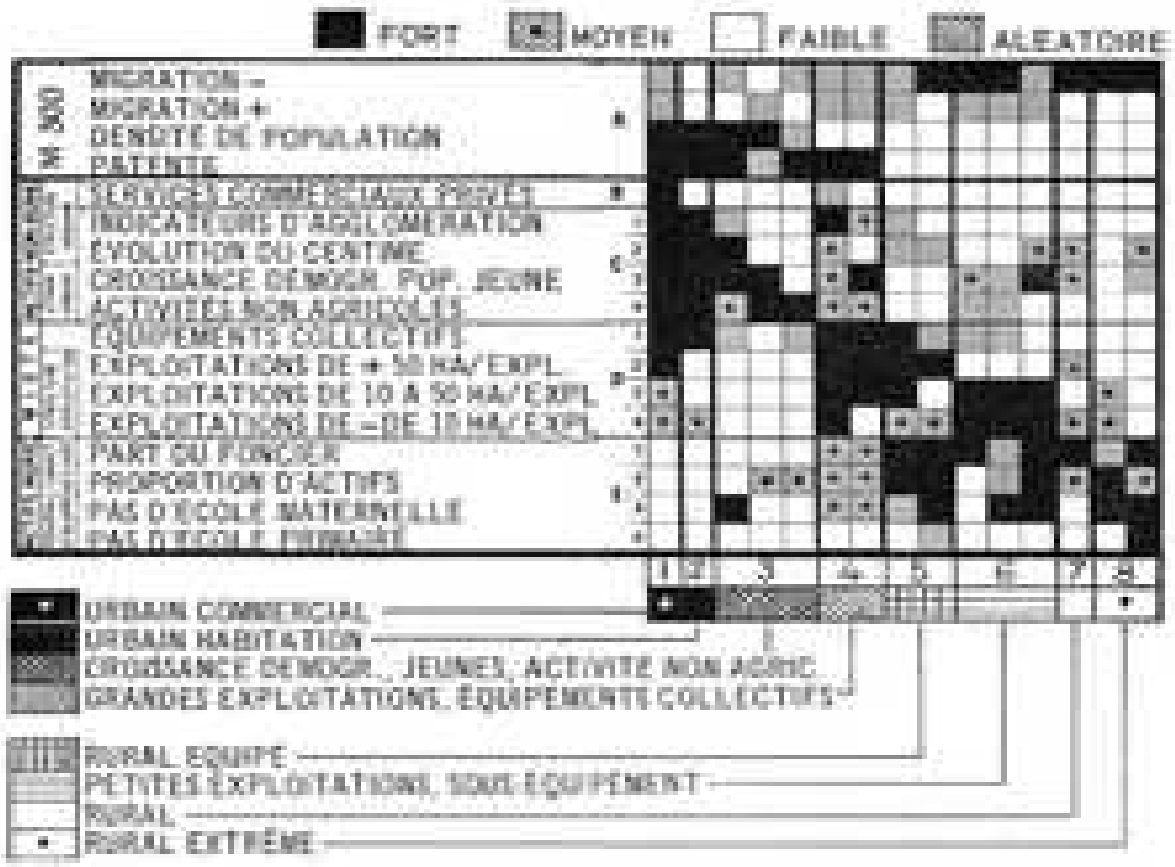
Il y a *gain d'information*. Types de communes et groupes d'indicateurs sont classés d'après l'opposition urbain-rural, ce qui permet la mémorisation de certains détails. Le gain le plus important reste toutefois, pour ce système 1, la dénomination des groupes d'indicateurs et des types de communes. Anodin en apparence, ce travail demande du temps et l'expérience du terrain. En effet, le choix des termes fait parfois appel à des éléments d'information non statistiques, ponctuels, qui « enrichissent » éventuellement les définitions trop élémentaires. Par exemple, faute d'avoir pu la demander au spécialiste de la D.D.A., la dénomination des types de communes 7 et 8 est restée très vague : « rural » et « rural extrême » n'indiquent pas la nature de la différence entre ces deux types.

Le gain d'information correspond à un accroissement de la qualité dans la connaissance d'une information. Il est difficilement mesurable et relève spécifiquement du travail d'interprétation du chercheur.

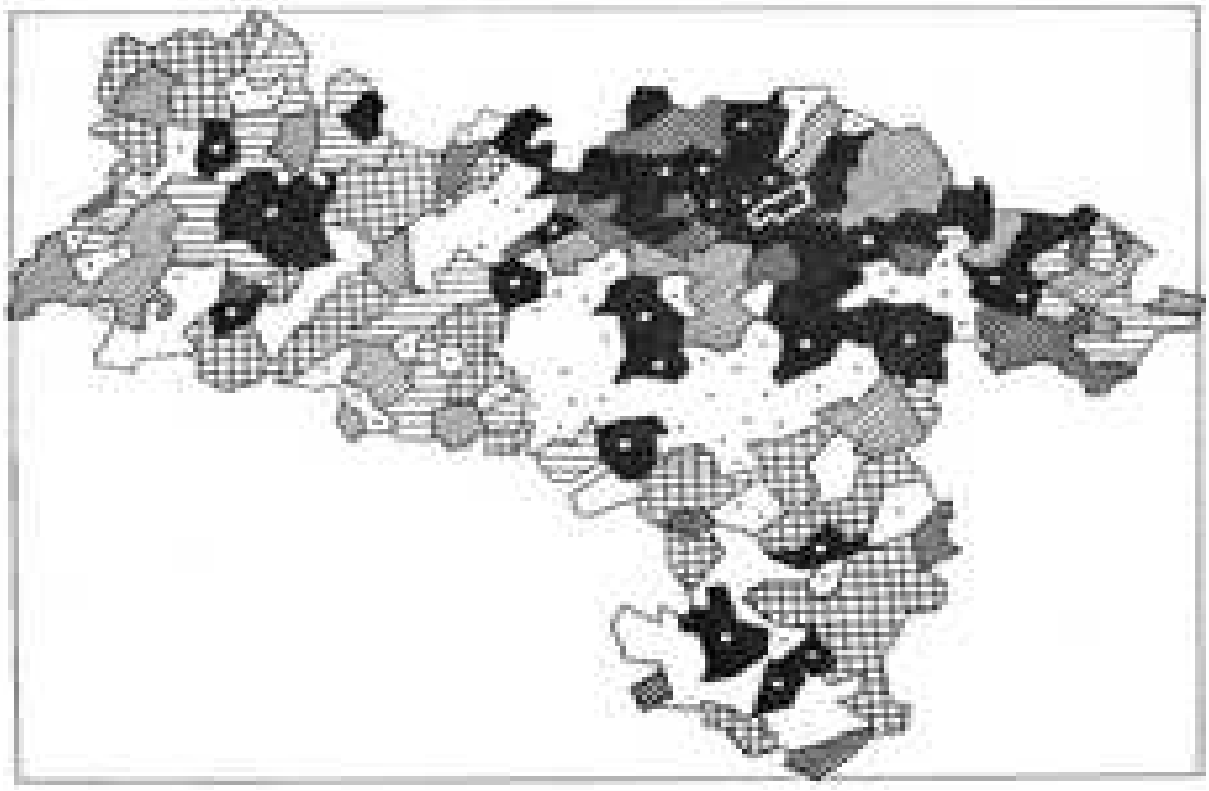
Le chercheur pondère les résultats de la matrice optimisée, d'après la *pertinence réelle* des indicateurs. L'acquisition de l'opposition urbain-rural ne présente guère d'intérêt, du fait de son évidence. On aurait pu se contenter d'indicateurs peu coûteux en temps de collecte, et efficaces du point de vue d'une typologie urbain-rural. Par exemple, les indicateurs E3 et E4 (absence d'école maternelle et d'école primaire) sont d'une très haute sélectivité, compte tenu du fait que ces renseignements s'obtiennent facilement. D'où la place qui leur est accordée dans la matrice d'interprétation, alors que le facteur B — services commerciaux privés — « résume » à lui seul 11 indicateurs de signification parfaitement redondante.

Cartographie du système 1.

La carte (Fig. 22) transcrit rigoureusement l'opposition urbain-rural en ordonnant visuellement, du noir au blanc, les 8 types de communes. La zone urbaine de Sedan forme un bloc noir.


FORT
MOYEN
FAIBLE
ALEATOIRE
MIGRATION −
MIGRATION +
DENSITE DE POPULATION
PATENTE
INDICATEURS D'AGGLOMERATION
ÉVOLUTION DU CENTRE
CROISSANCE DEMOGR. POP. JEUNE
ACTIVITÉS NON AGRICOLES
ÉQUIPEMENTS COLLECTIFS
EXPLOITATIONS DE + 50 HA/EXPL.
EXPLOITATIONS DE 10 A 50 HA/EXPL.
EXPLOITATIONS DE − DE 10 HA/EXPL.
PART DU FONCIER
PROPORTION D'ACTIFS
PAS D'ÉCOLE MATERNELLE
PAS D'ÉCOLE PRIMAIRE
URBAIN COMMERCIAL
URBAIN HABITATION
CROISSANCE DEMOGR., JEUNES, ACTIVITE NON AGRIC.
GRANDES EXPLOITATIONS, ÉQUIPEMENTS COLLECTIFS
RURAL EQUIPE
PETITES EXPLOITATIONS, SOUS ÉQUIPEMENT
RURAL
RURAL EXTRÊME

21



22

Optimisation du deuxième système.

Le système 1, ayant été reconnu et interprété, est totalement abandonné pour permettre l'optimisation des 21 indicateurs du système 2 (Fig. 23) et la réintroduction des indicateurs mis de côté lors de l'optimisation du système 1 : le système 2 passe de 21 à 28 indicateurs, les autres rejoignent le système 1. L'image du système 2 est devenue cohérente, au détriment de celle du système 1.

En outre, sont introduits trois indicateurs sur le solde migratoire (soldes positifs - soldes négatifs - soldes voisins de zéro), négligés lors de la construction de la matrice et qui s'intègrent au système 1. Un nouvel indicateur sur la taille des exploitations vient compléter le système 2. En effet, au vu des résultats du premier système, le spécialiste de la D.D.A. a pensé que le rapport des exploitations de moins de 20 ha sur le total des exploitations recensées en 1969 par commune devrait permettre de repérer les communes en retard sur le plan du développement agricole (l'exploitation moyenne de la zone herbagère fait entre 40 et 50 ha). (Cf. F1, Fig. 23 indicateur n° 95.)

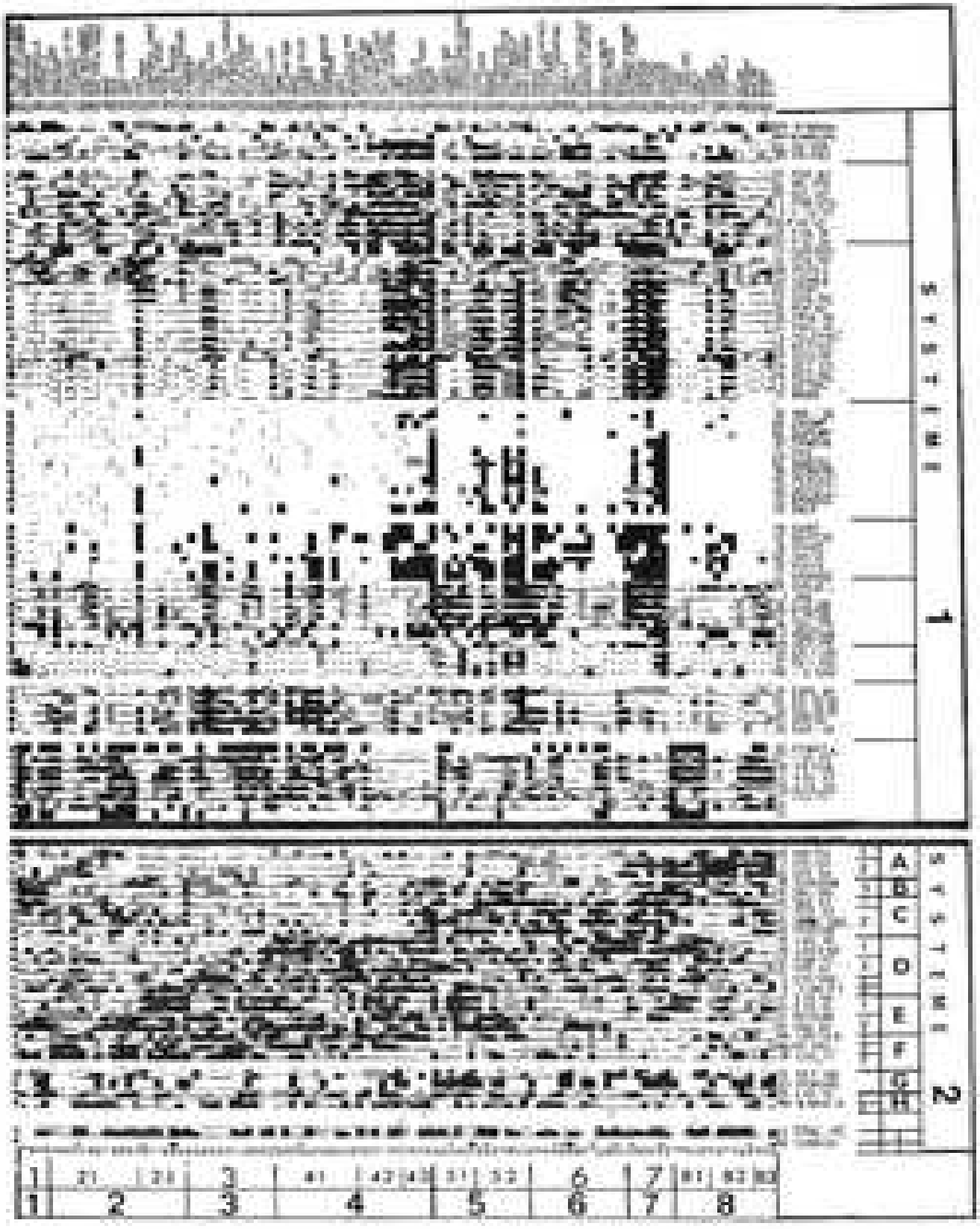
Au fur et à mesure de la progression du traitement, le chercheur est donc libre d'introduire de nouveaux éléments de comparaison, souvent suggérés par l'approfondissement de l'interprétation. Il lui est toujours facile, également, de retirer les indicateurs non pertinents, redondants, et de tester alors la pertinence d'autres qui prennent leur place.

Le traitement par système est une forme de pondération. Le système 1, avec ses 62 critères, peut n'être que d'un intérêt très réduit pour le chercheur : hiérarchiser les communes en « urbain-rural » peut ne pas être son objectif. En revanche, avec seulement 28 indicateurs, le système 2 peut y répondre. La découverte *a posteriori* de ce système constitue un gain d'information appréciable.

La délimitation des regroupements de communes et d'indicateurs effectués sur la base de la loi de proximité-similitude, c'est-à-dire le découpage des types et des groupes d'indicateurs, tient compte de l'originalité, de l'inattendu de ce deuxième système. Le niveau le plus fin divise les communes en 14 types (numérotés 1, 21, 22, 3, 41, 42, 43, 51, 52, 6, 7, 81, 82, 83) et les indicateurs en 19 groupes, dont 12 indicateurs individualisés (voir C1, D3, D4, F1, F2, F3, H1, H2, I, J). Il y a donc plus de subdivisions pour ce deuxième système composé de 28 indicateurs, que pour le système 1, qui en comporte pourtant près de 70.

Une telle finesse, délibérément voulue, ne facilitera pas l'interprétation, mais elle est nécessaire étant donné la complexité et l'hétérogénéité des indicateurs présents.

23



Interprétation du deuxième système.

La matrice d'interprétation du deuxième système (Fig. 24) est constituée de 19 groupes d'indicateurs (lignes), de 14 types de communes (colonnes) et de 4 paliers.

L'ordre global de ce système est beaucoup plus difficile à trouver. Il est constitué à la fois d'indicateurs dits de dynamisme agricole (salariés agricoles, I.V.D.[1]–retraite accélérée–remembrement) et d'indicateurs dits de niveau de vie (rénovation de l'habitat, structure démographique). Ils s'organisent autour de la dimension des exploitations : exploitations de plus de 50 ha (voir A, B, C, G) ; exploitations de moins de 50 ha (voir E, F).

A cette hiérarchie, vient se superposer un groupe d'indicateurs (Fig. 24′) où dominent les notions d'habitat moderne, de zone herbagère, de structure démographique.

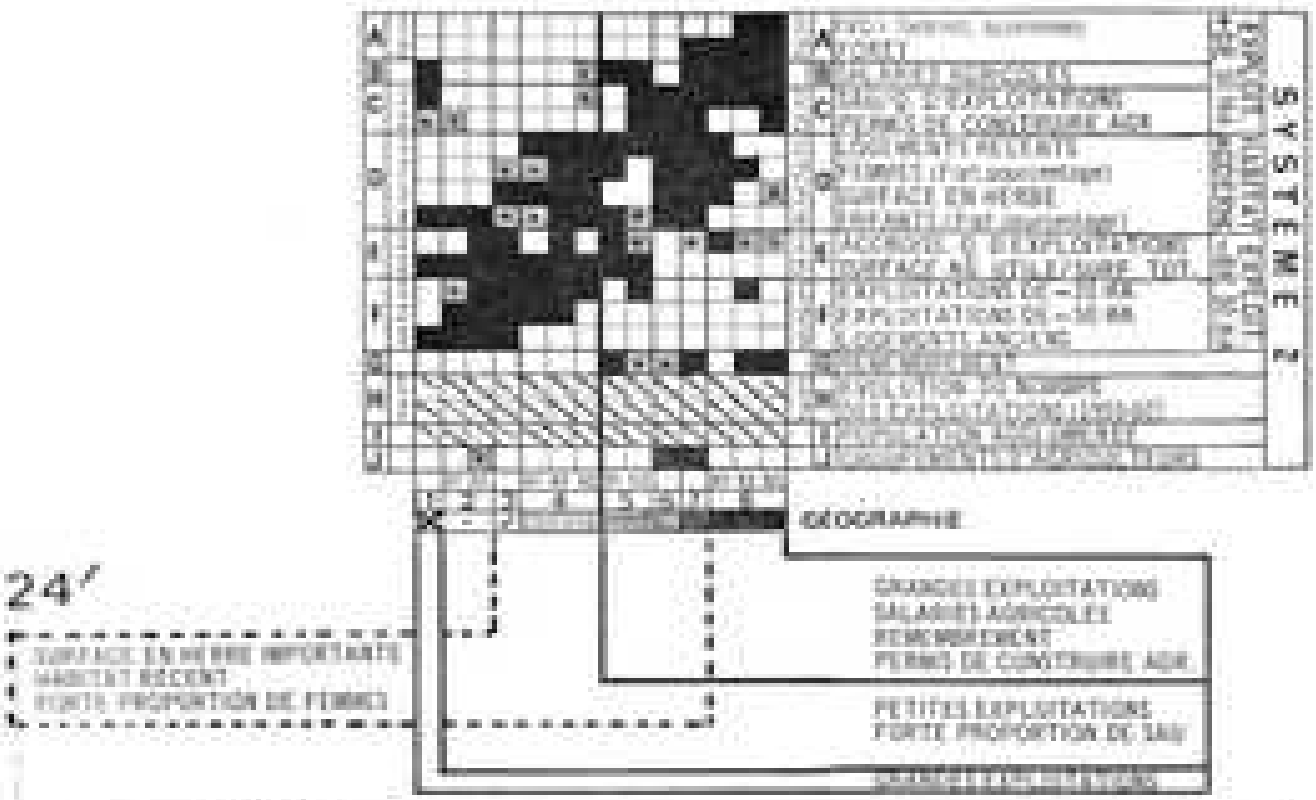
Les aspects démographiques de ce deuxième système (D1, D2, D4) ne peuvent être réellement interprétés que par un bon connaisseur de la zone herbagère. En revanche, les aspects agricoles (A, B, C, D3, F, G, J) se comprennent plus facilement, étant moins spécifiques à cette région.

Prenons le facteur F. Il est composé d'une part de deux indicateurs (F1 et F2) sur la taille des exploitations, ce qui apparaît « logique », d'autre part d'un indicateur d'habitat (F3). Ce mélange constitue un gain, et devient « logique » lui aussi quand l'interprétation en est faite. En effet, on lie fréquemment, et implicitement, la notion de tradition, de refus du progrès, aux petites exploitations. Ce lien semble ici exact et se trouve renforcé par l'indicateur d'habitat F3, puisqu'il signale les communes ayant tendance à avoir les plus fortes proportions de logements datant d'avant 1871. Sous-équipement rural et sous-équipement agricole se rejoignent donc pour les communes des types 2 et 3.

Prenons le cas des communes qui peuvent être considérées comme modernes (types 5, 6, 7, 8), parce qu'elles ont tendance à avoir de grandes exploitations (ce qu'indique C1, part de S.A.U. disponible par exploitation). Cette notion de modernisme y est renforcée par C2 — permis de construire agricole ; B — salariés agricoles ; A1 — I.V.D. allouées pour 100 exploitations ; et deux autres indicateurs qualitatifs (réponses oui-non) : G — remembrements effectués ou en cours et J, présence d'au moins une C.U.M.A.[2] ou (et) d'un G.A.E.C.[3] dans la commune.

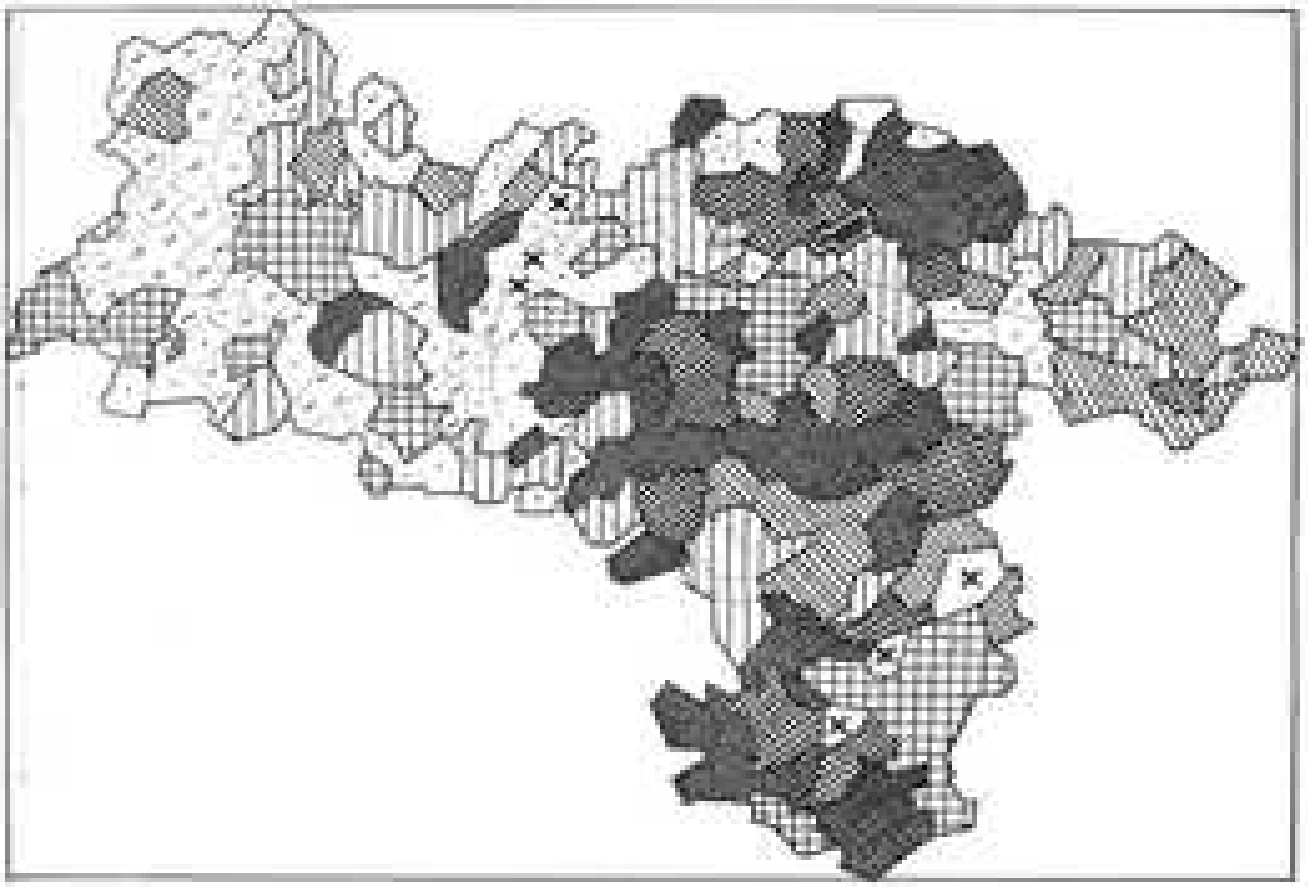
La transcription en 8 paliers ordonnés du noir au blanc suivant la dimension des exploitations, sous-entendu suivant le degré de modernisme agricole, donne une carte où la spatialisation s'avère remarquable : les communes « dynamiques » se regroupent à l'est de la zone.

1. I.V.D. : Indemnité viagère de Départ.
2. C.U.M.A. : Coopérative d'Utilisation de Matériel agricole.
3. G.A.E.C. : Groupement agricole d'Exploitation en commun.


SYSTEME 2
GÉOGRAPHIE
GRANDES EXPLOITATIONS
SALARIES AGRICOLES
REMEMBREMENT
PERMIS DE CONSTRUIRE AGR
PETITES EXPLOITATIONS
FORTE PROPORTION DE SAU
SURFACE EN HERBE IMPORTANTE
HABITAT RECENT
FORTE PROPORTION DE FEMMES

24

24′



25

Utilisation des régionalisations tirées des systèmes

Le déroulement du traitement a été ponctué par trois cartes : l'une tirée de la matrice 300 (Fig. 16), une autre déduite du premier système (Fig 22), une autre enfin du deuxième système (Fig. 35). Ci-contre et page suivante, ces mêmes cartes sont réutilisées pour être comparées mais changent de numéro de figure. La carte de la matrice 300 (Fig. 16) devient ci-contre la Figure 26. La carte du premier système (Fig. 22) est présentée deux fois : Figure 27 ci-contre et Figure 28 page suivante. Enfin, la carte du deuxième système (Fig. 25) est montrée page suivante sous le numéro de Figure 29.

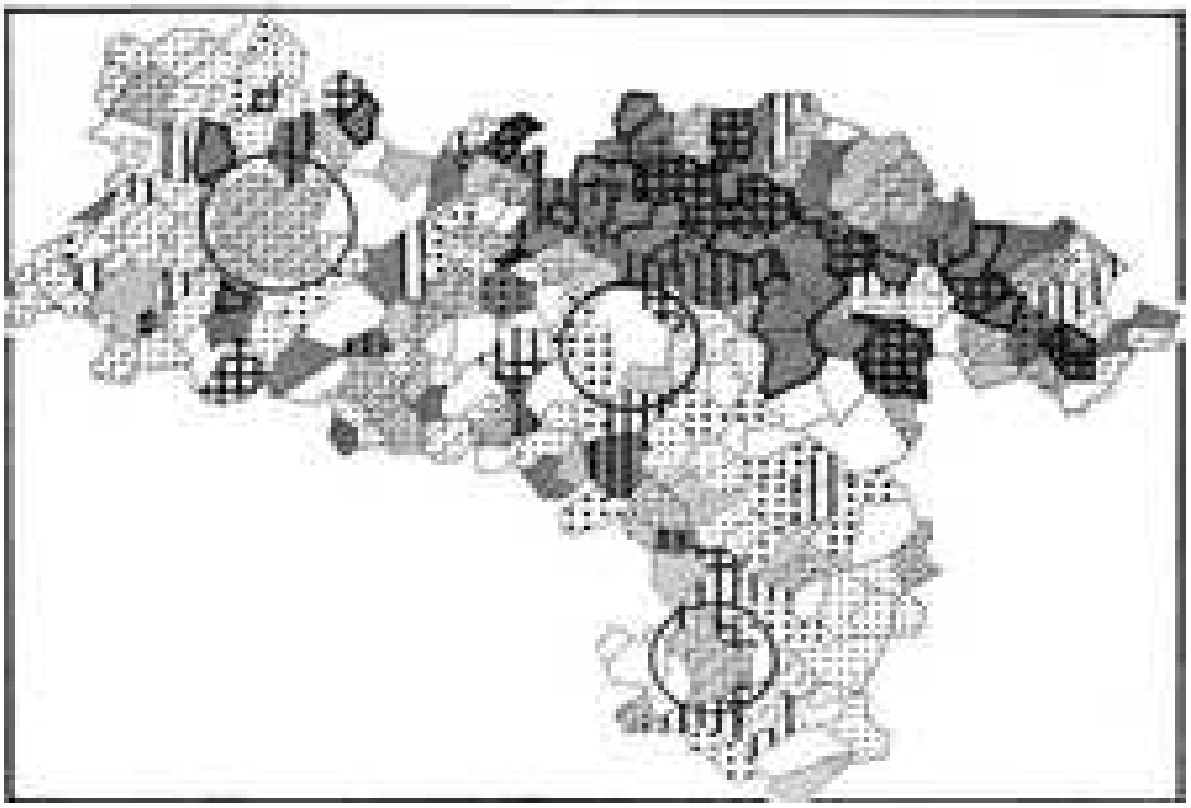
Comparaison des cartes de la matrice 300 et du système 1.

La comparaison de ces cartes suit un double objectif. D'une part, il faut vérifier le bien-fondé de la réduction, aussi bien sur le plan du choix des communes représentatives que sur celui des indicateurs utilisés. D'autre part, il est intéressant de voir comment les résultats les plus généraux — c'est-à-dire les plus spectaculaire, les plus faciles d'accès pour un lecteur non connaisseur de la zone herbagère — comment ces résultats donc peuvent être communiqués en n'utilisant que les cartes comme support graphique, comme contenant de l'information. La lecture d'une carte est en effet nettement plus aisée pour la plupart des gens qu'un tableau à double entrée, à plus forte raison si ce dernier est mis en ordre et visualisé.

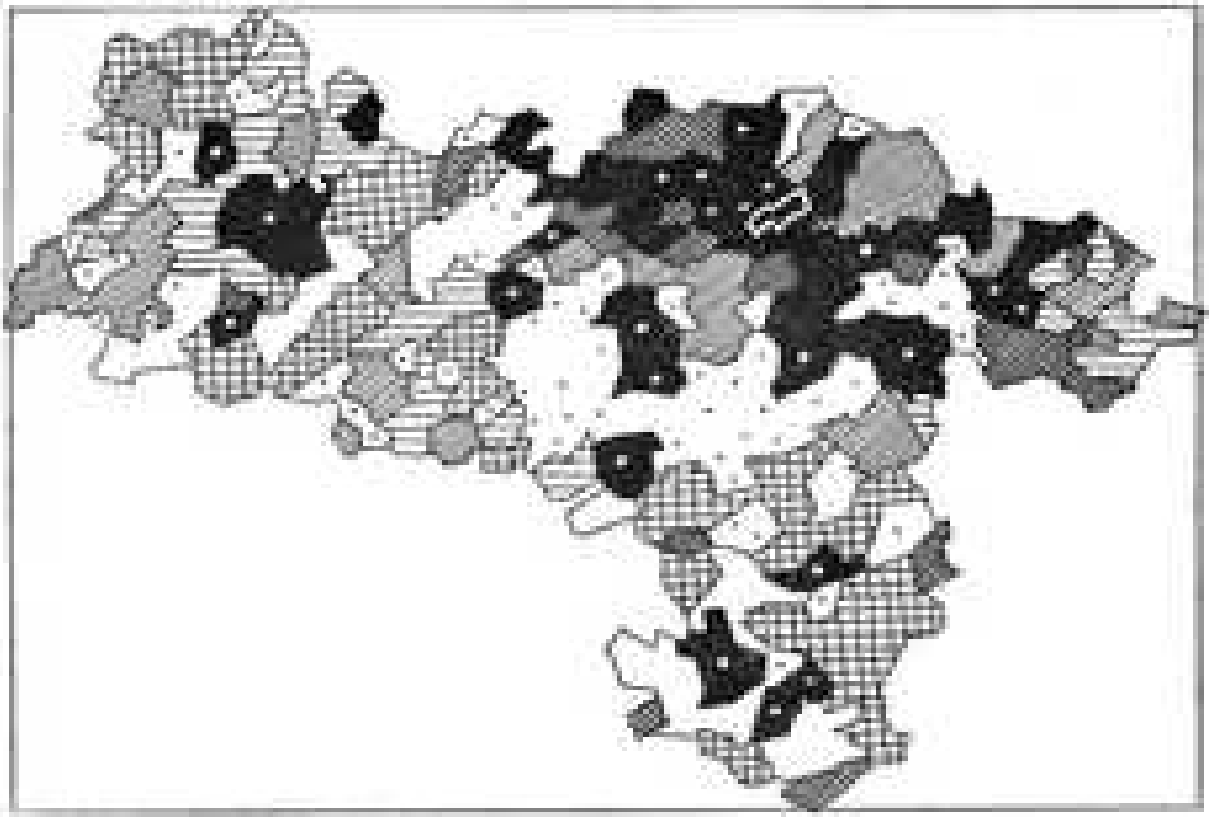
Le rangement des communes dans la matrice 300 et dans le premier système étant fondé sur la même opposition urbain-rural, la ressemblance entre les deux cartes est tout à fait normale.

Les trois différences notables (communes cerclées de la carte 26) correspondent toutes à des chefs-lieux de canton de grande superficie : la carte fait le constat de l'insuffisance de la densité de population, dans la matrice 300, comme seul indicateur de population. Cette erreur initiale n'a pas eu de conséquence sur le choix des 100 communes représentatives, la quantité de population ayant été prise en compte entretemps.

A ce stade on peut donc dire que les 100 communes sélectionnées sont bien représentatives des 300, que l'introduction de 100 indicateurs, dans le seul but d'une régionalisation (en vue d'une enquête par exemple) peut être discutée. En effet, la même régionalisation des 300 communes que celle de la Figure 27 aurait pu se faire sans réduction préalable, à l'aide soit de quelques indicateurs de population — par exemple quantité et densité de population, solde migratoire, actifs agricoles pour 100 actifs — soit d'indicateurs d'équipements — services commerciaux privés, services collectifs, absence de service, par exemple. Encore fallait-il en être sûr !



26



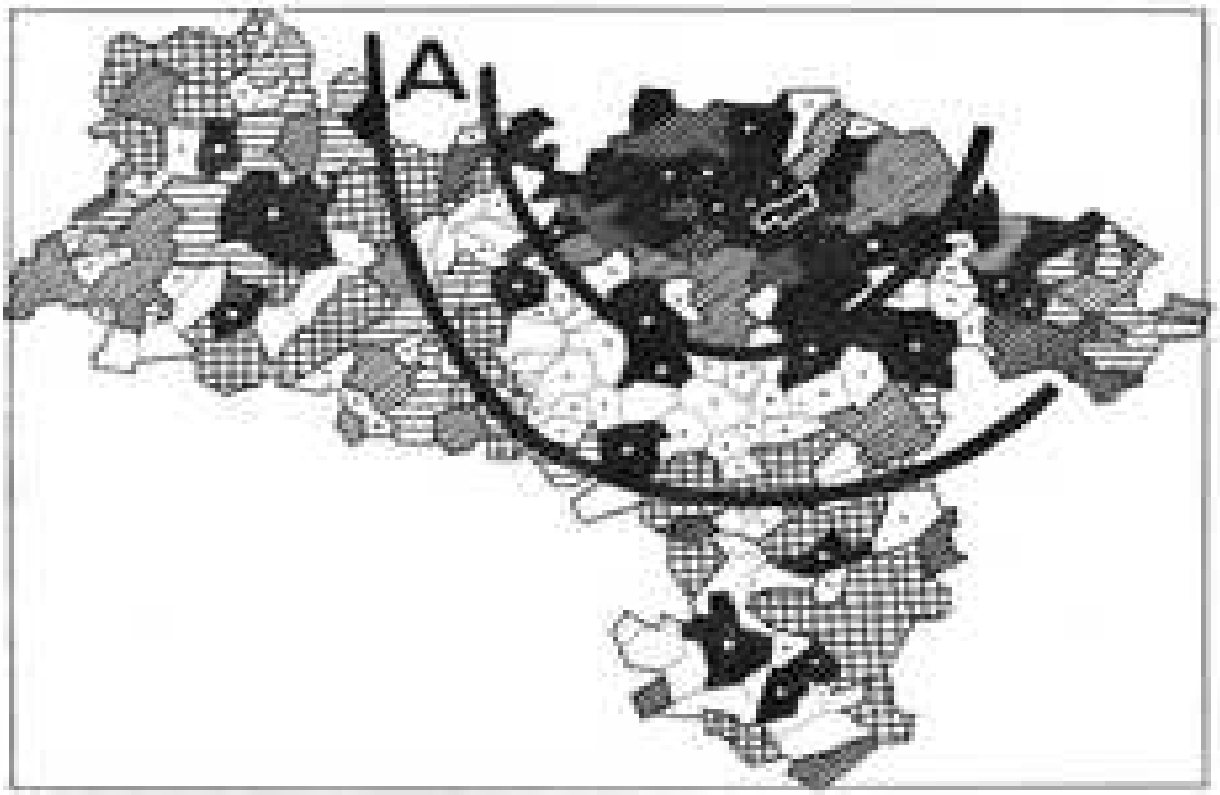
27

Comparaison des cartes des systèmes 1 (Fig. 28) et 2 (Fig. 29).

L'existence du second système, que traduit cartographiquement la dissemblance des cartes 29 et 28, justifie à lui seul la réduction préalable adoptée pour permettre une typologie fondée sur un grand nombre d'indicateurs. Pouvoir dire (et démontrer !) que la régionalisation ordonnée sur la taille des exploitations est totalement différente de la régionalisation fondée essentiellement sur la quantité de population représente un gain d'information considérable. Ce constat n'est simple qu'en apparence : d'une part il n'était pas évident *a priori*, d'autre part il ne fait que généraliser toute une série de renseignements acquis en cours de traitement, déjà connus donc, maisi mplicitement utilisés pour la généralisation. Par exemple, le système 1 a permis de juger de la pertinence des indicateurs d'équipement communal, tous de type qualitatif (réponses oui-non) et d'un temps de collecte restreint. Le système 2, quant à lui, confirme que le modernisme des exploitations est lié à leurs dimensions, et montre qu'un indicateur comme le remembrement permet de donner une idée valable du dynamisme communal, sans faire appel à des indicateurs d'évolution, qui soulèvent la plupart du temps des difficultés d'homogénéité des données.

La régionalisation tirée du premier système (Fig. 28) met en valeur la zone urbaine, suburbaine et périurbaine de Charleville-Mézières et Sedan. C'est la carte qui suggère d'englober dans la zone périurbaine les communes de type « rural extrême » (blanc pointé sur la carte). Ces dernières forment une auréole blanche parfaitement nette autour du bloc noir des communes urbaines et suburbaines (Fig. 28, A). Cette proximité brutale, sans transition, entre — disons, toutes proportions gardées — « l'extrême urbain » et « l'extrême rural » (communes en « perte de vitesse ») s'expliquerait-elle par un phénomène de « pompage » lié à l'urbanisation et accentué par des conditions agricoles et démographiques défavorables ?

La régionalisation du système 2 (Fig. 29) donne une opposition presque aussi brutale entre la partie ouest de la zone herbagère, dominée par les petites exploitations et un habitat ancien, et la partie est où les exploitations de plus de 50 ha ont tendance à être plus nombreuses. Une exception toutefois dans cette partie *est* : la zone suburbaine au sud de Sedan, où le blanc (les petites exploitations) redevient fréquent (Fig. 29, B). Cette situation visible sur la Figure 29 s'explique très bien par le fort degré d'urbanisation de ces mêmes communes, qu'indique la carte 28. Mais dans ce cas, cela devrait être vrai aussi pour l'ensemble de la zone urbaine et suburbaine, au nord de cette auréole B. L'explication n'est donc pas aussi simple et échappe à un profane de la géographie de la zone herbagère.



28



29

Superposition géographique des deux systèmes.

La superposition sur la carte du système 2 (Fig. 30) des quatre types définis comme « urbains » dans le premier système (voir la matrice d'interprétation, Fig. 21) donne la carte de synthèse des deux systèmes. L'ordre des quatre types de communes « urbaines » du système 1 est transcrit par des points de taille décroissante du type 1 au type 4 (3 cercles noirs, 1 cercle évidé).

Les cinq régions de la Figure 30 ne sont que le résultat des constats effectués par comparaison des cartes de chaque système. On a d'abord dessiné les deux limites déduites de ceux-ci : l'arc de cercle, en gros orienté est-ouest, démarquant du reste la zone sous l'influence de Charleville-Mézières et Sedan, et la coupure nord-sud du système 2 (petites et grandes exploitations). La combinaison de ces deux lignes détache plus spécialement les régions 4 et 5, car la zone sous influence urbaine est elle-même subdivisée en 3 régions, fruit de la démilitation des deux auréoles constatées précédemment.

Il est alors possible de décrire ces cinq régions de la façon la plus générale qui soit.

Région 1 : « urbanité » élevée, forte proportion de grosses exploitations ;

Région 2 : urbanité élevée, faible proportion de grosses exploitations ;

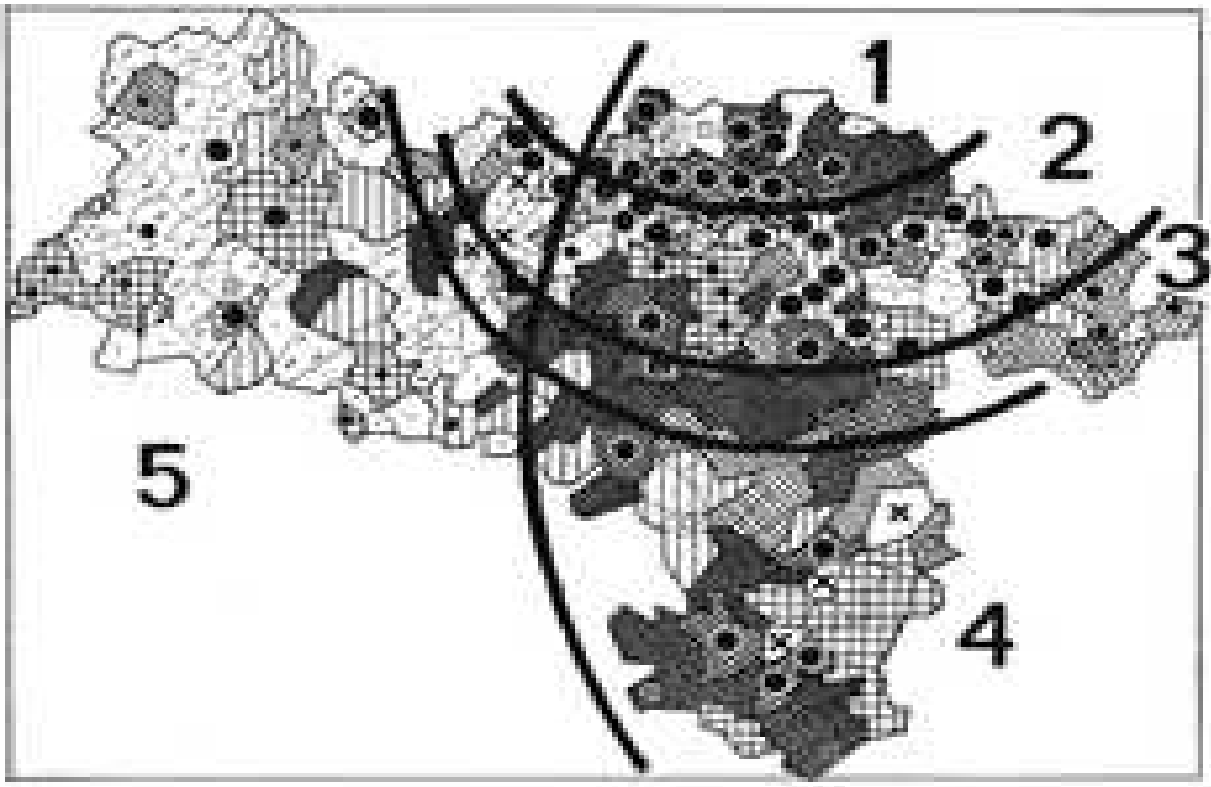
Région 3 : communes en perte de vitesse, forte proportion de grosses exploitations. C'est un peu le négatif de la région 2 ;

Région 4 : pas d'influence urbaine notable, grandes exploitations dominantes ;

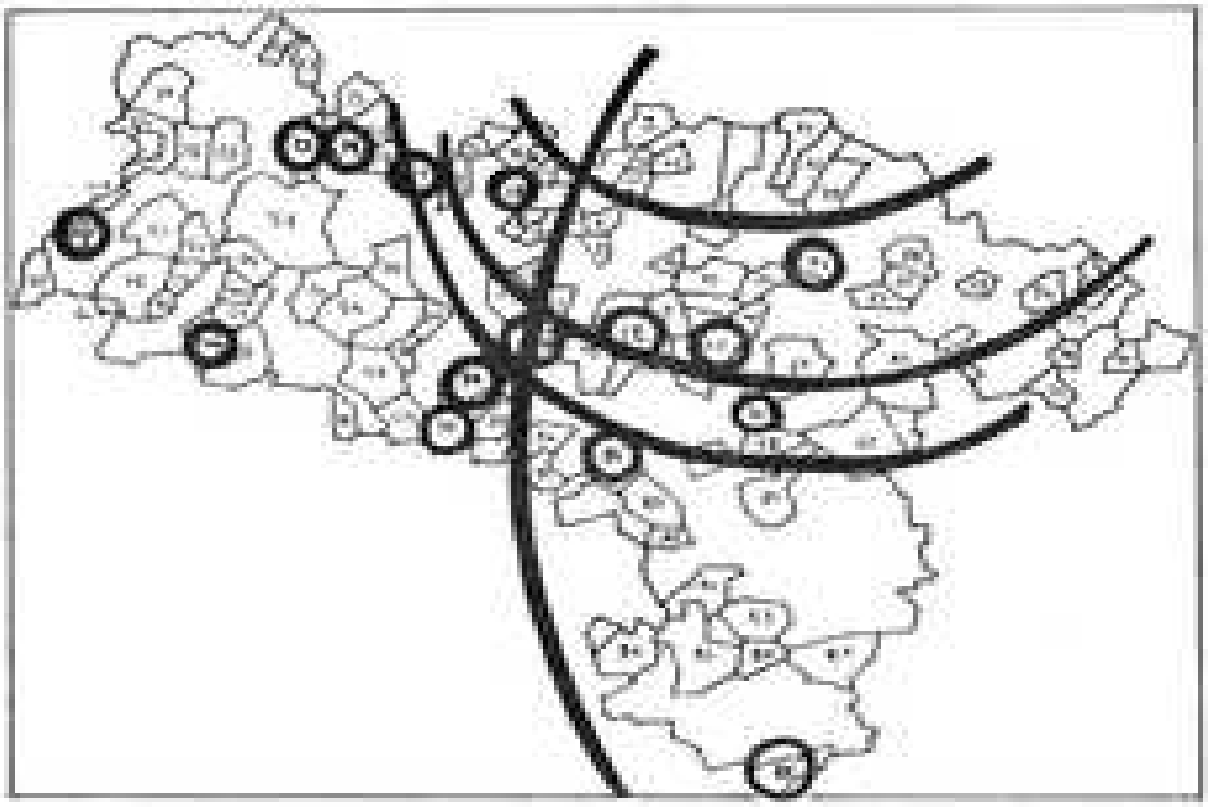
Région 5 : pas d'influence urbaine notable, dominante des petites exploitations.

La Figure 31 est composée de trois éléments très différents : les limites des régions ci-dessus, les 100 communes retenues lors de la réduction préalable (cf. Fig. 17) et, cerclées, 15 communes sélectionnées sur la seule base des 8 types du premier système (certains types étant représentés par deux ou trois communes). Le choix de ces 15 communes n'a pu être influencé par le second système, car il a été élaboré — par calcul — avant que ce système n'ait été interprété et utilisé. Cette carte montre donc la spatialisation des deux réductions, par rapport à la régionalisation en cinq types, étudiée sur la Figure 30.

Elle montre d'abord que la première réduction n'a délaissé aucune de ces régions et, surtout, a retenu un grand nombre de communes là où la régionalisation est complexe (auréoles de la zone d'influence urbaine). En revanche, elle signale que la deuxième réduction n'a sélectionné aucune commune dans la région 1 (urbanité élevée, forte proportion de grandes exploitations), ce qui, bien sûr, est gênant. Enfin, elle suggère un choix intéressant : de nombreuses communes de la deuxième réduction se trouvent à la limite de deux régions : vaut-il mieux, pour une opération d'aménagement, choisir les communes les plus typées possible, et se trouvant donc au milieu de la zone qu'elles sont censées représenter, ou vaut-il mieux, au moins en partie, prendre des communes à la frontière de deux ou trois régions et dont l'aménagement, pour cette raison, serait apparemment moins facile ?



30



31

Utilisation des valeurs numériques dans un traitement graphique

Les cartes tirées de chaque système en constituent l'exploitation la plus immédiate et la plus directe qui soit. Nous venons de voir qu'elles-mêmes amènent un gain d'information substantiel, puisqu'il a nom, régionalisation. Ce n'est pas la localisation et la lecture des phénomènes commune par commune qui sont intéressantes, mais bien la spatialisation (degré de proximité géographique) des relations entre ces phénomènes, relations découvertes à l'aide du traitement matriciel, en l'occurrence.

Il en va de même quand le chercheur utilise des nombres pour exprimer, vérifier ou améliorer ce que le traitement lui a permis de découvrir, que le « langage » de ce traitement soit graphique ou mathématique.

L'intérêt de toute exploitation numérique d'un traitement ne réside certainement pas dans la quantité de nombres que le chercheur *transmet* à un éventuel lecteur, car la mémorisation en devient très vite impossible pour l'un comme l'autre. La précision si souvent présentée comme « nécessaire » à tout traitement de l'information, ne constitue pas non plus l'avantage fondamental des nombres... ce peut être l'arbre qui cache la forêt. La précision n'est en effet pas exploitable directement en tant que telle. Le tableau de la Figure 32 va nous permettre de comprendre pourquoi.

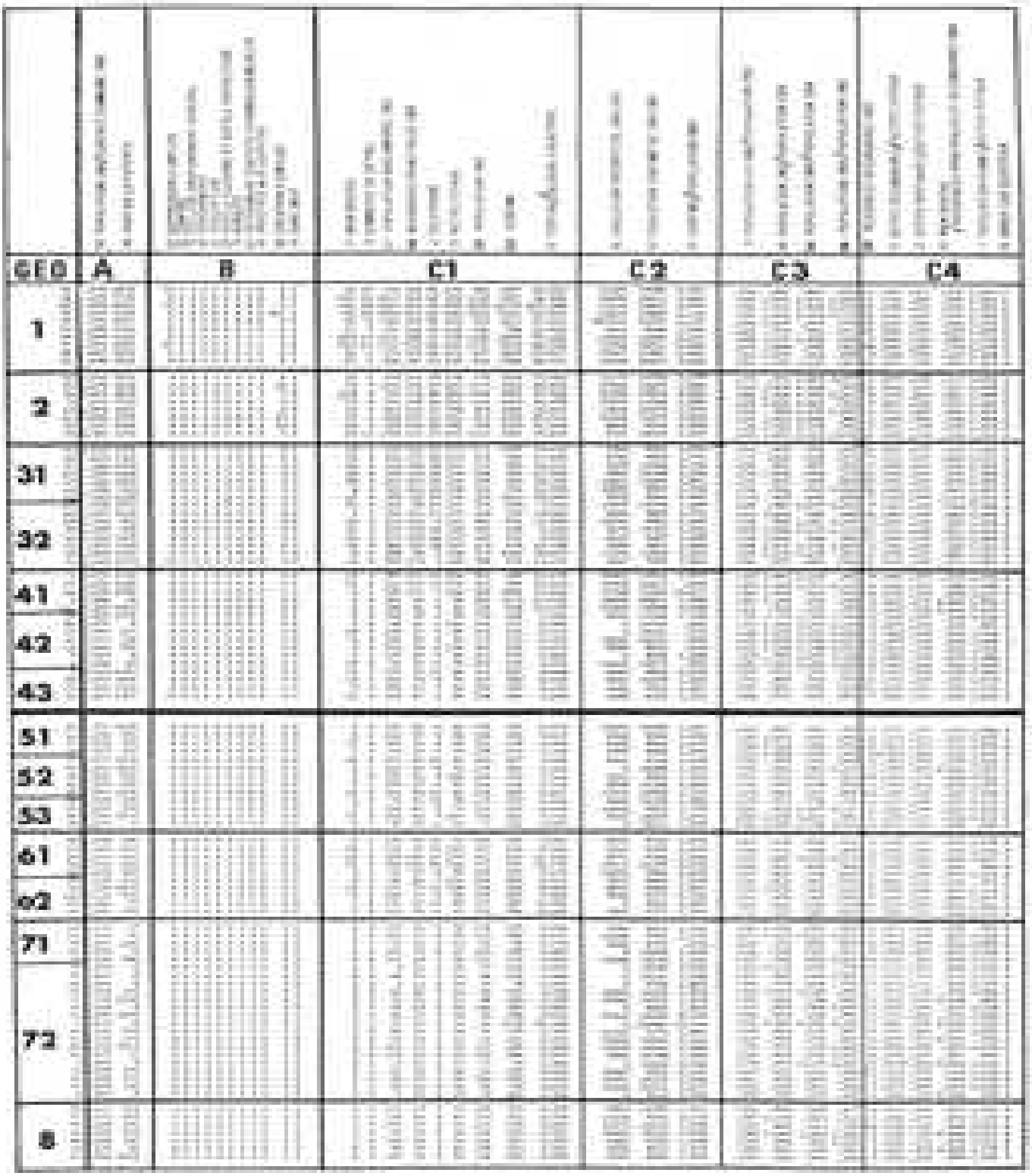
Une exploitation numérique : le tableau des données optimisé.

Ce tableau doit être compris et lu par rapport à la Figure 20 (système 1 optimisé). En effet, le rangement des communes (attention ! une commune = 1 ligne, Fig. 32) et des indicateurs est identique, à quelques exceptions près, à celui de la Figure 20 pour les groupes d'indicateurs A, B, C1, C2, C3, C4, seuls présentés ici. Le terme de transcription numérique convient parfaitement pour décrire ce tableau : les onze paliers de la matrice graphique sont remplacés par les valeurs numériques propres à chaque case.

Imaginons que nous ne possédions pas les acquis du traitement pour aborder ce tableau, qui pourtant est mis en ordre. La lecture et la mémorisation de ces nombres devient vite fastidieuse, voire impossible. Un tel tableau ne prend toute son utilité qu'après l'optimisation, car il ne s'agit plus alors de mémoriser 3 535 nombres (101 communes × 35 indicateurs), mais de tenir compte de la précision des 20 à 50 nombres que l'optimisation suggère d'aller lire de plus près. Cette précision est alors chargée de tout le gain du traitement et c'est dans ces conditions seulement qu'elle constitue elle-même un gain d'information.

Un tel tableau ne peut, en pratique, être rapidement obtenu (il y a toujours des difficultés techniques) que par ordinateur. En outre la sortie des séries chiffrées doit être calculée à l'avance, en fonction du format de l'éventuelle publication, pour éviter, autant que possible, d'avoir à lire des chiffres très petits.

32



Une exploitation numérique : les communes types.

L'utilisation combinée du tableau des données (Fig. 32) et de l'interprétation du premier système (Fig. 20, 21, 22) aboutit à une exploitation numérique où interviennent des calculs à proprement parler, mais qui est aussi à la fois plus facile sur le plan de la communication et plus intéressante pour le chercheur, qui peut y introduire directement son expérience du terrain, par exemple.

Pour effectuer ces calculs, deux ensembles ont été déterminés *a priori* : l'un appelé « ensemble urbain » (Fig. 33) ne prenant en compte que les communes classées dans les types 1 à 4 du système 1, un autre appelé « ensemble rural » (Fig. 34), formé des communes des types 5 à 8. Chaque ensemble est caractérisé par une série d'indicateurs, au nombre de 13 (Fig. 33) et de 15 (Fig. 34). Leur sélection est aisée, dans la mesure où leur pertinence est connue, étant à la base du découpage des indicateurs en groupes et facteurs, découpage que rappelle la colonne de droite de chaque tableau. Des 96 indicateurs de la matrice 100 × 100 ces deux tableaux n'en conservent donc au total que 23, dont 5 sont utilisés en commun.

Cette sélection d'indicateurs sert à trouver par calcul, la commune moyenne de chaque type, en suivant le processus ci-dessous :

— calcul de la moyenne des indicateurs de chaque ensemble : colonne moyenne « urbaine » (Fig. 33) et moyenne « rurale » (Fig. 34) ;
— calcul des moyennes de chaque type : types 1 à 4 (Fig. 33), 5 à 8 (Fig. 34) ;
— recherche de la ou des communes les plus semblables à la commune théorique de leur type ;
— introduction éventuelle d'informations *externes* au traitement pour corriger les choix effectués.

Nous avons déjà vérifié comment les 15 communes retenues se distribuent géographiquement : ce sont celles de la Figure 31. Cependant les auteurs de ce choix n'avaient à leur disposition ni cette carte, ni l'interprétation du deuxième système. Toutefois, leur choix est solide, dans la mesure où les 8 types du deuxième système sont tous représentés (lignes intitulées : « système 2 : numéro de type »).

Il est vrai qu'il avait été élaboré avec le concours des spécialistes de la zone. L'apport d'informations externes au traitement est nécessaire, car il compense l'aspect théorique de la commune moyenne type qui sert de première référence.

Réduction du nombre d'indicateurs (de 92 à 23), réduction du nombre de communes (de 101 à 15) : la quantité de valeurs numériques à mémoriser permet alors d'utiliser de façon opérationnelle la précision des nombres, d'autant plus que les acquis du traitement sont à la base des réductions opérées.

ENSEMBLE "URBAIN" : TYPES 1 à 4 DU SYSTEME 1
COMMUNES ET INDICATEURS RETENUS

Une exploitation numérique : l'étude d'un indicateur.

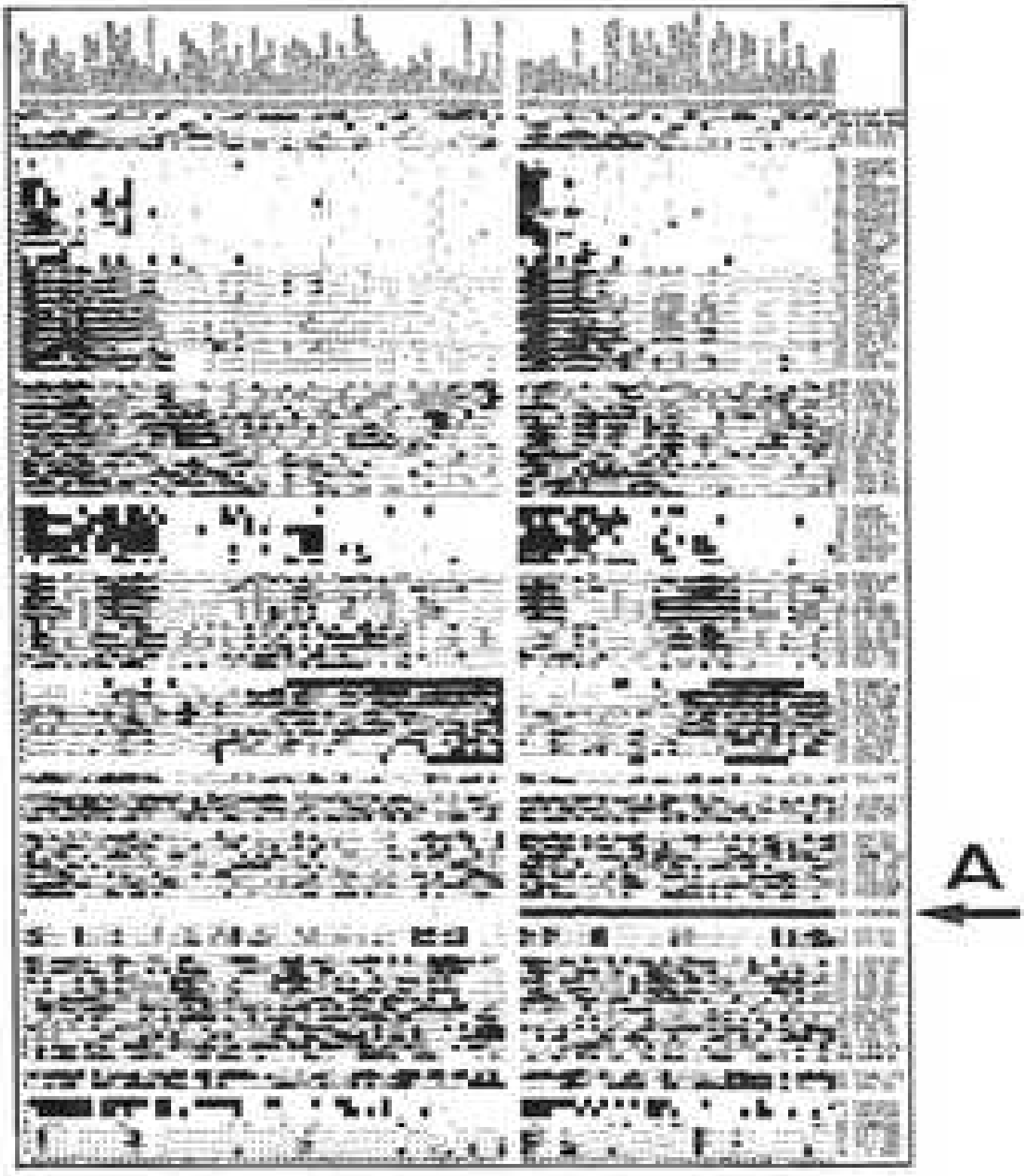
L'indicateur remembrement est privilégié dans la Figure 35 (en A), les communes remembrées étant séparées des autres et se trouvant toutes à droite de la matrice. A l'intérieur des deux blocs ainsi définis, les communes sont ordonnées d'après leur position dans le système 1 optimisé. Les communes non remembrées sont donc classées des plus urbaines aux plus rurales, ainsi que les communes remembrées. Les indicateurs conservent exactement le même rangement que celui du système 1 optimisé (cf. Fig. 20), à quelques exceptions près.

Le choix de cet indicateur est lié, d'une part, au fait qu'il se soit montré très pertinent dans le deuxième système (cf. Fig. 24), d'autre part, parce qu'il ne décrit pas une situation statique donnée mais qu'il témoigne d'une évolution, d'un certain dynamisme des communes. La Figure 35 met donc en valeur un double classement : celui du remembrement, témoin du système 2, et celui de l'ordre urbain-rural, témoin du système 1. Cela permet de poser la question suivante : quelles sont les communes faisant preuve d'un dynamisme agricole et qui n'ont pas ou peu d'équipements collectifs, indices d'un dynamisme rural? La réponse n'est claire que pour un seul type d'équipement : la majorité des communes sans école maternelle (indicateur 69) se trouvent dans les communes non remembrées. Pour tous les autres équipements, il n'y a aucune tendance affirmée. En revanche, d'autres relations apparaissent, qu'il serait trop long de commenter ici. Ces relations peuvent faire l'objet de calculs divers, pour lesquels en général les chercheurs ne manquent pas d'imagination.

Comme pour un indicateur qualitatif (réponses oui-non), la mise en ordre des valeurs numériques d'un indicateur quantitatif ne prend tout son intérêt qu'après l'optimisation. Celle-ci permet de démontrer la pertinence de l'indicateur retenu. En outre, la détermination des « valeurs-seuil », c'est-à-dire des valeurs numériques caractéristiques d'un type de communes, prend appui sur l'ensemble des indicateurs et met alors en jeu tous les acquis du traitement. Tout indicateur ainsi mis en ordre peut faire l'objet d'une carte : la proximité géographique sert à vérifier, si besoin est, le bien-fondé des valeurs-seuil. Il existe également tout un arsenal de calculs pour déterminer la relation d'un indicateur avec un ou plusieurs autres.

De tels calculs (taux de corrélation, d'entropie, coefficients de rang, moyennes diverses, fréquences, etc.) que certains chercheurs utilisent couramment, ont l'avantage d'autoriser des comparaisons rapides avec d'autres ensembles externes au traitement. Dire : « la part des ménages agricoles pour 100 ménages par commune et la densité de population par commune ont un taux de corrélation de — 0,963 dans la zone herbagère, contre — 0,922 dans la région Champagne–Ardennes » peut être utile au spécialiste, d'autant que cela ne nécessite pas un traitement de l'information lourd en temps et en matériel.

35



Perte et gain d'information, généralisation et précision.

Il est maintenant possible de reprendre tout le processus de traitement tel qu'il vient d'être décrit sous un angle plus théorique, c'est-à-dire en l'analysant selon les notions de perte et de gain d'information.

Il y a perte d'information quand, pour une raison quelconque, une partie des données originales du problème traité est abandonnée. Ainsi transcrire les 100 nombres d'un indicateur en 11 paliers constitue une perte d'information.

Il y a gain d'information quand la réorganisation interne d'un ensemble de données permet soit de découvrir de nouvelles relations, internes ou externes au traitement, soit d'avoir la certitude de l'absence de relations entre tout ou partie des données. Ainsi, la réorganisation interne de la matrice 100 $\times$ 100 a conduit à la découverte de deux systèmes.

(Ne pas confondre gain et apport d'information. Un gain d'information fait souvent intervenir des éléments explicatifs non contenus dans le traitement. Mais, ajouter une commune ou un indicateur constitue un apport d'information, dont on ignore *a priori* l'utilité pour un éventuel gain d'information.)

Un gain d'information ne s'accompagne pas nécessairement d'une perte d'information, mais une perte d'information devrait toujours en principe être justifiée par un gain équivalent.

Analysons comment les pertes d'information les plus importantes de ce traitement se trouvent justifiées par un gain d'information plus ou moins équivalent.

a) *Première perte d'information :* la réduction de 300 à 100 communes (cf. Fig. 12 à 17). L'idéal aurait été de faire la typologie en conservant et les 300 communes et les 100 indicateurs. Ce traitement idéal n'étant matériellement pas possible, une réduction préalable était nécessaire. Outre une collecte des données plus facile, cette réduction se trouve justifiée par la découverte des deux systèmes (cf. Fig. 20 et 23). Ce gain, qui est obtenu sans contrepartie équivalente au niveau de la seule matrice 100 $\times$ 100, entraîne en particulier avec lui le gain des régionalisations, dont nous avons tiré le profit maximum (cf. Fig. 26 à 30).

b) *Deuxième perte d'information :* réduction en onze paliers graphiques des données numériques de la matrice 100 $\times$ 100 (cf. Fig. 10 et 19). C'est la seule perte d'information qui soit véritablement liée au traitement graphique de l'information. Le gain correspondant ? La simplification de l'image initiale. A preuve, la lecture des 3 535 nombres du tableau des données optimisé (Fig. 32) : il n'y a aucune réduction, mais il n'y a également pas acquisition immédiate du degré d'ordre ou de désordre de ce tableau. Quand une matrice ne comporte que des indi-

cateurs qualitatifs, comme celle de l'exemple fictif initial (cf. Fig. 1 à 6), sa simplification ne nécessite aucune perte d'information.

c) *Troisième perte d'information* : les matrices d'interprétation (cf. Fig. 21 et 24). Il y a trois réductions effectuées en même temps : les communes y sont ramenées à une dizaine de types, les indicateurs à une vingtaine de groupes et quelques facteurs, les 11 paliers à quatre appréciations. Dénomination des regroupements effectués, pondération et donc choix des éléments les plus significatifs, mémorisation facilitée en constituent la contrepartie.

d) *Quatrième perte d'information* : choix de quelques communes types (cf. Fig. 33 et 34). Dans ce cas, la perte d'information ne se situe qu'au niveau des communes et des indicateurs. Ce genre de réduction ne peut toutefois avoir lieu qu'après avoir pris en compte tous les acquis du traitement. Le seul gain véritable qui vienne en échange tient non pas tellement dans la réintroduction des valeurs numériques, mais surtout dans l'acquisition de noms de communes, d'entités concrètes à partir desquelles il devient possible d'effectuer un travail à une autre échelle, celle des exploitations agricoles par exemple.

Le gain constitué par la découverte et l'exploitation des deux systèmes, et qui justifie donc la réduction de 300 à 100 communes, représente la plus grande originalité d'un tel traitement. C'est en effet à la notion de système que sont liées, d'une part, les deux cartes de régionalisation, d'autre part, la sélection des indicateurs les plus significatifs, enfin le choix des communes types. La notion de système convient en réalité aussi bien à l'élaboration des divers degrés de généralisation liés à l'interprétation des résultats, qu'à une utilisation souple et efficace de la précision, tant numérique que spatiale, nécessaire aux prises de décision.

Tous les traitements de l'information importants, à l'image de celui que nous venons de résumer, supposent pertes et gains d'information, car ces derniers correspondent en réalité à l'objectif même du traitement. Tous visent en effet à supprimer ce qui est jugé superflu en fonction de certains impératifs préalablement fixés, pour ne retenir que les éléments jugés importants et nouveaux en fonction de ces mêmes impératifs.

A l'exemple de celui dont nous venons d'étudier le déroulement, et dont nous n'avons d'ailleurs pas caché certaines insuffisances, le traitement graphique par matrice ordonnable représente une méthode de traitement où le chercheur reste constamment maître de son information, que ce soit pour contrôler les pertes d'information, ou pour tirer parti des gains d'information.

Rôle du chercheur dans le traitement matriciel

Le traitement matriciel est actuellement utilisé dans de nombreuses disciplines : urbanisme, sociologie, économie, archéologie, histoire, démographie, géographie, médecine, etc. Les champs d'application en sont donc très variés, à condition que les dimensions de l'information à traiter ne dépassent pas les capacités matérielles du support de permutation adopté. Dans le cas inverse, le chercheur dispose alors de deux solutions : soit adopter une autre technique, un autre système de traitement graphique de capacité supérieure, mais d'une utilisation moins souple que le traitement matriciel, soit ramener les dimensions de son information aux capacités du traitement matriciel, par utilisation d'une « chaîne de traitement ».

Dans l'exemple précédent, la chaîne de traitement fait appel à une première matrice en vue de réduire le nombre de communes ; l'optimisation de la seconde matrice est alors considérée comme applicable à l'ensemble des communes. Une chaîne de traitement suppose toujours une opération de réduction effectuée soit par combinaison de systèmes de traitement graphique entre eux (cartes + matrice), soit par combinaison de systèmes de traitement graphique et mathématique (analyse factorielle des correspondances + matrice).

L'automatisme ne s'applique qu'à des stades bien précis d'un système ou d'une chaîne de traitement de l'information, qu'il soit graphique ou mathématique, et pour des opérations en général de caractère élémentaire et répétitif. Plus un système ou une chaîne de traitement prendra en compte une masse importante de données, plus la collaboration entre l'informaticien et le chercheur devra être suivie.

Ce dernier sera alors en mesure de consacrer l'essentiel de son temps sur les aspects du traitement réellement importants pour lui : définition des limites et des objectifs de son étude, sélection des indicateurs, exploitation poussée de tous les résultats.

Dans l'exemple précédent, le traitement matriciel englobe non seulement l'image optimisée, la matrice elle-même en tant que support de permutation ou mémoire artificielle, mais encore toute une série de tableaux, de documents connexes, utiles à la fois au chercheur et à l'informaticien. L'informaticien contrôle chacune des cinq opérations élémentaires à l'aide de ces tableaux. Le chercheur emploie ces tableaux en leur attribuant une double fonction : avant l'optimisation, ils lui permettent de faire les choix d'indicateurs et d'en améliorer *a priori* la qualité et la pertinence ; après l'optimisation, ils servent à interpréter, communiquer et pondérer les résultats. Ce dernier point mérite d'être analysé en détail en prenant les cinq opérations fondamentales comme repère.

La relation temps de collecte–pertinence supposée : construction du tableau des données.

Appliquée intuitivement par de nombreux chercheurs, cette relation évidente entre la quantité de travail fournie en vue d'obtenir des indicateurs et la pertinence, la sélectivité, le pouvoir discriminant dont le chercheur les charge, a servi à la mise au point de deux tableaux que le chercheur doit employer avant la collecte des données.

Le « tableau de création d'indicateurs » (Fig. 18) constitue la base des commentaires du chercheur sur sa sélection d'indicateurs.

Le « tableau de pertinence des indicateurs » n'est pas présenté dans l'exemple précédent. Le chercheur regroupe ses indicateurs suivant des ensembles « homogènes », « logiques », appelés « familles » d'indicateurs. Ces familles peuvent correspondre à des divisions du tableau des données suivant la nature des données recueillies : démographie ; économie rurale, urbaine ; équipements, etc. La composition des familles d'indicateurs actuellement recommandée porte plus nettement sur la relation temps de collecte–pertinence supposée : le chercheur met noir sur blanc ses *a priori* et le niveau de sélectivité qu'il attend de telle ou telle famille d'indicateurs. Ce tableau peut être transcrit graphiquement et optimisé avec les autres indicateurs. L'utilisation combinée de ces deux tableaux permet au chercheur de construire un tableau des données où les *a priori* sont recensés, où la sélection des indicateurs est maîtrisée au maximum. Le chercheur peut d'ailleurs toujours rajouter des indicateurs en cours de traitement.

Un tableau dit d' « homogénéité » de l'information, antérieur aux deux précédents, permet au chercheur, si besoin est, d'analyser la structure et la combinatoire de son information, en vue d'adopter les systèmes de traitement les plus appropriés à ses préoccupations.

Transcription graphique du tableau des données[1].

En grande partie automatique, cette opération correspond parfois à une traduction graphique : le chercheur met en valeur certains indicateurs, soit en doublant le nombre de paliers, soit en les répétant mais transcrits inversés. Ceci est l'équivalent par exemple des dichotomies effectuées par champs d'action dans l'exemple de traitement par l'analyse factorielle des correspondances cité plus haut. Le tableau de réduction en paliers (cf. Fig. 10) est un instrument de référence.

1. Le lecteur désireux de ne pas se contenter des principes très généraux de la transcription graphique énoncés avec les figures 7, 8, 9 aura tout intérêt à se reporter à l'ouvrage de J. BERTIN, *Sémiologie graphique*, Paris, Gauthier-Villars, 1967, 431 p.

Simplification de la matrice.

L'application automatique de la loi de proximité-similitude aboutit à la simplification de la matrice. Le chercheur, en effectuant les choix suggérés par la simplification, optimise la matrice. Quand la simplification fait découvrir un ensemble d'indicateurs de distribution cohérente et un autre ensemble d'indicateurs de distribution incohérente, dite « aléatoire », chacun de ces ensembles sera appelé système, et sera optimisé et interprété pour lui-même. La simplification peut être obtenue par des cheminements différents, tous étant fondés cependant sur la loi de proximité-similitude et visant à découvrir les systèmes. Certaines informations demandent des cheminements appropriés.

L'analyse factorielle des correspondances constitue elle-même un de ces cheminements, intéressant pour son automatisme et pour sa complémentarité avec le traitement matriciel.

Dans le fascicule 35 des *Études rurales* (juil.-sept. 1969, pp. 5-135), un article décrit un exemple d'application d'un système de traitement mathématique proche du traitement matriciel, l'analyse factorielle des correspondances (cf. F. Loux et L. de Virville, « Le système social d'une région rurale : le Châtillonnais ». Précédé d'un avant-propos de J. Cuisenier). La notion de « famille » d'indicateurs y est appelée « champ d'action ». Ces champs d'action sont hiérarchisés en « sous-systèmes » ou domaines.

Le cheminement adopté pour ce traitement mathématique vise à optimiser les familles d'indicateurs, puis à optimiser globalement l'ensemble considéré.

La transcription graphique des résultats (« cartes ») utilise comme dans le traitement matriciel la notion de proximité dans le plan pour exprimer les mesures de similitude entre les individus et leurs caractéristiques. Une meilleure transcription graphique aurait d'ailleurs permis d'intégrer très facilement le troisième axe avec les deux autres sur un seul graphique (ou « carte »). Le traitement matriciel se compare par ailleurs très facilement avec tous les autres systèmes de traitement mathématique.

La relation temps de collecte–pertinence supposée–pertinence réelle : interprétation de la matrice.

Vérification de la pertinence des indicateurs.

Chaque système une fois optimisé, le chercheur doit vérifier et commenter le bien-fondé de ses hypothèses de sélection des indicateurs. Le tableau de pertinence des indicateurs est analysé en fonction de l'ordre fourni par l'optimisation. Le chercheur découvre alors des redondances (un seul indicateur aurait suffi pour exprimer la même idée) ou au contraire des familles d'indicateurs très sélectives. Ce niveau de sélectivité, rapporté au temps de collecte de ces indicateurs, donne au chercheur le niveau de pertinence.

Dans l'exemple précédent, la famille des indicateurs d'équipements s'est avérée très pertinente, car chaque type de commune était représenté par une combinaison de ces indicateurs, tous qualitatifs (« oui »-« non ») et dont la collecte n'a nécessité que peu de

temps et aucun calcul. Il va de soi qu'aucun système automatique n'est capable d'affecter un niveau de pertinence à un indicateur, ce niveau étant lié à l'interprétation des résultats.

Matrice d'interprétation.

Instrument de pondération liée à la pertinence des indicateurs, c'est également un instrument de synthèse : le chercheur dénomme les groupes d'indicateurs corrélants (on dira facteurs), ou choisit l'indicateur représentatif d'un groupe (indicateur-pilote). Il opérera la hiérarchisation des groupes d'individus et de leurs caractéristiques en fonction des niveaux d'interprétation qui lui conviennent.

Cartes de synthèse (pertinence de la proximité géographique).

La proximité géographique est utilisée comme un indicateur particulièrement pertinent de la synthèse obtenue, lorsque bien sûr l'information traitée a comme références statistiques des unités géographiques.

La probabilité d'homogénéité liée à la proximité géographique est confirmée ou infirmée. Le chercheur ne peut commenter l' « allure » de la carte et en dégager les explications fondamentales, que si la transcription cartographique de l'optimisation est correcte.

L'indicateur de proximité géographique est en outre utilisé pour l'étude de l'interdépendance des systèmes (superposition cartographique, cf. Fig. 30).

La communication des résultats.

Instrument de pondération et de synthèse, la *matrice d'interprétation* est également un instrument de communication et peut même n'avoir que cette fonction. Le chercheur généralise les résultats en ne prenant en compte que le minimum de paliers : présence, absence, moyen, « aléatoire », ou met en vedette certains indicateurs en fonction de ses interlocuteurs.

La *transcription cartographique* peut avoir une fonction équivalente : une carte en quatre régions ne communiquera pas suffisamment de renseignements pour les spécialistes locaux, qui demanderont une carte en quinze régions. La matrice permet en outre tous les classements spéciaux que le chercheur souhaite établir en vue d'applications ou de démonstrations particulières. Ces « manipulations » sont effectuées sur des bases méthodologiques très précises (y compris d'éventuelles transcriptions cartographiques) dont l'objectif est de communiquer facilement certains résultats très détaillés découverts par l'optimisation.

Les chercheurs ne doivent pas demander à un système de traitement de l'information uniquement de traiter beaucoup de données très vite :

en effet, l'interprétation des résultats ne pourra jamais en elle-même être automatisée et les chercheurs y passeront toujours beaucoup de temps.

Ne prendre en compte que les propriétés d'automatisme et de capacité d'un système de traitement — que le « langage » de traitement soit mathématique ou graphique — correspond à une sous-utilisation par les chercheurs de l'apport du traitement de l'information.

Le *gain d'information* (qualité des résultats obtenus) ; la *facilité de contrôle des opérations effectuées* (temps d'initiation nécessaire au chercheur pour comprendre la fonction de chaque opération) ; la *communicabilité des résultats* (temps nécessaire à l'interlocuteur pour saisir les résultats dans leur détail) nous semblent des propriétés tout aussi fondamentales.

L'exigence des chercheurs dans le domaine de l'interprétation devrait rendre possible la mise au point d'une méthodologie encore plus accessible, construite surtout à partir du langage graphique et en maîtrisant toutes les possibilités.