

Saisie de données géographiques et cartographie assistée par ordinateur

par Jean-Paul DONNAY

MOTS-CLES. - *Cartographie assistée par ordinateur, structures de données, topologie.*

RESUME. - *Cet article rappelle les principales structures de données spatiales utilisables, en mode vectoriel, dans le contexte de la cartographie assistée par ordinateur. Les caractéristiques de chaque structure sont exposées, ainsi que leurs avantages et inconvénients. Des exemples de logiciels ayant recours à ces structures sont signalés et un unique fond de carte simplifié illustre la mise en place des banques de données résultantes.*

KEY-WORDS. - *Computer assisted cartography, data structures, topology.*

ABSTRACT. - *This paper gives an account of the principal spatial data structures in vectorial mode for the sake of the computer assisted cartography. The main features, the usefulness and the pitfalls of each structure are underlined. Examples are named among the cartographic softwares using these structures. An outline of a simple map is used to highlight the building of the spatial data files.*

I. - INTRODUCTION

Cet article rend compte de l'étape initiale du processus d'automatisation cartographique, à savoir la saisie de données géographiques, leur structuration et leur mémorisation.

Les considérations qui suivent valent uniquement pour le mode graphique vectoriel. Dans ce cas, la saisie de données spatiales est effectuée au moyen d'un appareil nommé digitaliseur ou table à numériser. Il permet la mesure et la saisie des coordonnées planes d'un point situé dans un espace bidimensionnel dont l'origine est fixée a priori par l'utilisateur (pour plus de détail sur la manipulation du digitaliseur, voir Johannsen T.M., 1979).

On a souvent entendu dire que la cartographie assistée par ordinateur avait libéré le cartographe de la table à dessin pour l'at-

tacher à la table à digitaliser. Il est vrai, en effet, que cette étape fondamentale requiert une grande attention et une longue préparation si l'on souhaite la rendre efficace.

A.- POSITION DU PROBLEME

En cartographie thématique, on est amené à manipuler deux types de données : d'une part, les données spatiales - qui peuvent être ponctuelles, linéaires ou aréales et qui, en première approximation, peuvent se ramener à une combinaison de couples de coordonnées planes -, et, d'autre part, les données thématiques qui, elles, correspondent à des attributs qualitatifs ou quantitatifs associés aux éléments de l'espace. Les premières, les données spatiales, vont déterminer le fond de carte et elles sont essentiellement de nature topographique. Les secondes, les données thématiques, font généralement l'objet d'un traitement avant d'être représentées graphiquement selon un type cartographique adéquat. Elles peuvent appartenir à une multitude de thèmes différents, pourvu qu'ils possèdent, explicitement ou implicitement, un caractère spatial.

Ces deux ensembles de données peuvent être manipulés simultanément dans une application ; en d'autres termes ils peuvent faire l'objet d'un traitement commun. C'est le cas, par exemple, lorsque l'on ajuste une surface de tendance à un nuage de points définis par trois dimensions : en (X) et en (Y) par leurs coordonnées planes et en (Z) par leur valeur thématique. Mais, le plus souvent, les traitements que l'on est amené à faire subir aux jeux de données sont distincts et il est plus pratique, et plus prudent, de séparer dès le départ les deux ensembles de données. Ainsi, par exemple, un même fond de carte peut supporter plusieurs représentations thématiques différentes, c'est-à-dire qu'un même ensemble de données spatiales peut être associé à plusieurs variables thématiques qui subiront, éventuellement, des traitements distincts. Inversément, un même phénomène thématique peut utiliser des fonds de carte différents. A l'exemple, l'ensemble des données spatiales peut être soumis à différentes projections cartographiques. La distinction entre les deux ensembles de données est donc primordiale et il faut prévoir des opérations de saisie et de stockage différentes pour chacun d'eux.

B.- LA SAISIE DES DONNEES THEMATIQUES

La saisie des données thématiques ne pose généralement pas de problèmes. Les valeurs de la variable thématique, éventuellement accompagnées d'une étiquette identifiant l'élément spatial auquel elles se rapportent, peuvent être stockées dans un fichier grâce aux multiples dispositifs d'entrée de données alphanumériques. Un bon éditeur interactif suffit à cette tâche, mais il existe aussi des logiciels permettant, outre le simple encodage de données, la réalisation de quelques traitements plus ou moins sophistiqués. Ces logiciels sont qualifiés de tableurs, calcs ou gestionnaires de données et les fonctions qu'ils supportent vont du simple tri ou de la simple transformation de variable jusqu'aux calculs statistiques relativement évolués.

C.- LA SAISIE DES DONNEES SPATIALES

Il n'en va pas de même en ce qui concerne la saisie des données spatiales. Le dispositif d'entrée de données, le digitaliseur, présentera toujours, en sortie, les coordonnées d'un point. Par conséquent, la saisie et, surtout, la mémorisation des différents éléments spatiaux passent obligatoirement par une structuration des informations en provenance du digitaliseur. Cette structure peut être simple ou, au contraire, fort complexe. Mais il faut savoir dès à présent que les traitements que l'on pourra effectuer sur les données et, en dernière analyse, les expressions cartographiques que l'on pourra utiliser, sont conditionnés par la structure organisant la mémorisation des données spatiales.

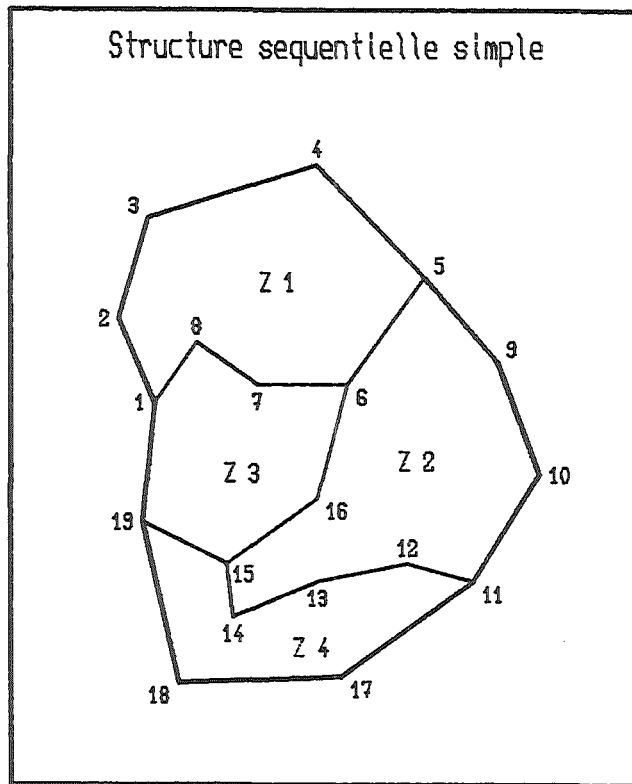
Pour mettre en évidence le rôle conditionnel des structures de données géographiques, quelques modes d'organisation de données de complexité croissante seront examinés successivement.

II. - CONDITIONS FONDAMENTALES DE STRUCTURATION DES DONNEES SPATIALES EN MODE GRAPHIQUE VECTORIEL

Pour illustrer les différentes structures abordées dans cet article, un même fond de carte élémentaire sera utilisé. Cependant, les éléments spatiaux constituant ce fond de carte sont de type aréal, soit un type incompatible, strictement, avec le mode graphique vectoriel. Cela fournit l'occasion de rappeler les principales caractéristiques du mode graphique vectoriel et les conditions fondamentales de structuration des données spatiales.

Les éléments constituant ce fond de carte sont des zones, des surfaces, dans lesquelles on peut reconnaître des communes, des provinces, des pays, etc. Or, en mode vectoriel, il n'est pas possible de saisir directement la notion de surface. On doit, pour manipuler une surface, se contenter de sa limite, de sa frontière. Cette dernière est un élément linéaire. Pour ne pas confondre une surface définie par sa frontière et un élément linéaire pur et simple, il faut ajouter à la définition de la zone un identificateur qui traduira, en quelque sorte, la liaison topologique entre la surface et sa frontière. Par conséquent, une zone sera définie, en mode vectoriel, par sa frontière et un identificateur de zone.

Mais le problème de la description des éléments du fond de carte ne s'arrête pas là. Par définition, un élément linéaire est continu. Or, le dispositif de saisie n'étant pas d'habitude analogique, l'élément linéaire ne pourra, non plus, être manipulé directement en mode vectoriel. Il faut discrétiser cet élément linéaire, le ramener à une séquence de points remarquables, définis par des critères de voisinage et d'orientation. Ce sont ces points qui, en définitive, pourront être saisis directement. Il faut cependant éviter toute confusion entre,



Fichier de zones

X	Y	X ₁	Y ₁
X ₁	Y ₁	X ₁	Y ₁
X ₂	Y ₂	X ₂	Y ₂
X ₃	Y ₃	X ₃	Y ₃
X ₄	Y ₄	X ₁₆	Y ₁₆
X ₅	Y ₅	X ₁₅	Y ₁₅
X ₆	Y ₆	X ₁₉	Y ₁₉
X ₇	Y ₇	X ₁	Y ₁
X ₈	Y ₈	-1	-1
X ₁	Y ₁	X ₁₉	Y ₁₉
-1	-1	X ₁₅	Y ₁₅
X ₅	Y ₅	X ₁₄	Y ₁₄
X ₉	Y ₉	X ₁₃	Y ₁₃
X ₁₀	Y ₁₀	X ₁₂	Y ₁₂
X ₁₁	Y ₁₁	X ₁₁	Y ₁₁
X ₁₂	Y ₁₂	X ₁₇	Y ₁₇
X ₁₃	Y ₁₃	X ₁₈	Y ₁₈
X ₁₄	Y ₁₄	X ₁₉	Y ₁₉
X ₁₅	Y ₁₅	-1	-1
X ₆	Y ₆		
X ₅	Y ₅		
-1	-1		

FIG. 1.- La structure séquentielle simple. Dans le fichier constituant la banque de données spatiales, la fermeture des zones est explicite. Le séparateur de zones correspond à un couple de coordonnées de valeurs égales à -1.

d'une part, un élément linéaire défini par une séquence de points remarquables et, d'autre part, une série d'éléments ponctuels isolés. L'introduction d'un nouvel identificateur est dès lors nécessaire. Cet identificateur traduira la relation topologique entre les points successifs de la séquence constituant la ligne. En mode vectoriel, un élément linéaire est donc défini par une séquence d'éléments ponctuels et un identificateur de ligne.

Connaissant les conditions fondamentales de structuration des données spatiales en mode vectoriel, il faut s'interroger sur les diverses manières d'organiser leur stockage.

III. - LA STRUCTURE SEQUENTIELLE SIMPLE

A.- DESCRIPTION

La manière la plus simple d'organiser la mémorisation du fond de carte consiste à saisir et à mémoriser successivement et de façon distincte les différentes zones présentes. La manipulation d'une zone s'effectuant par l'intermédiaire de sa frontière et cette dernière se ramenant, en dernière analyse, à une séquence de points, on se contentera de mémoriser, dans l'ordre, les points constituant chaque frontière de zone (fig. 1).

Le fichier résultant de cette opération est constitué d'une suite de couples de coordonnées (X, Y). Pour distinguer les différentes zones au sein du fichier, on peut utiliser un séparateur entre chaque zone : un enregistrement dont la valeur de (X) est négative, par exemple, si les coordonnées sont toujours mesurées positivement. On peut aussi utiliser une étiquette distincte pour chaque zone et associer cette étiquette à tous les points d'une zone. Enfin, pour assurer la "fermeture" des zones, il faut que le dernier point de chaque zone coïncide avec le premier. Cette correspondance peut être explicite - le point initial est répété à la fin de la zone dans le fichier - ou être réalisée par voie logicielle lors de la lecture de la banque de données spatiales.

Dans cette structure séquentielle simple, l'identificateur de zone est implicite. La structure ne supporte que des éléments spatiaux de type "zones", sauf convention contraire.

B.- AVANTAGES

L'avantage de cette structure tient à sa simplicité. D'une part, la préparation du fond de carte, précédant l'étape de saisie proprement dite, est réduite au minimum, bien qu'elle ne soit pas toujours exempte de subjectivité. Il suffit de généraliser manuellement les frontières de zones, c'est-à-dire d'identifier et de marquer sur le fond de carte les points remarquables qui doivent être digitalisés.

D'autre part, puisque chaque zone est saisie indépendamment des autres, les points constituant la frontière d'une zone sont acces-

sibles en bloc et séquentiellement. Cette disposition de l'information est compatible avec les supports séquentiels d'informations. Elle facilite également l'écriture des algorithmes de cartographie par plages. Le hachurage, par exemple, fait appel à des calculs d'intersection entre les traits de la trame et la frontière de la zone. Dans ces conditions, la structure séquentielle assure un accès rapide aux informations requises par ce genre d'algorithme.

C.- DESAVANTAGES

Les désavantages de la structure séquentielle simple sont cependant plus nombreux que ses avantages.

Les inconvénients de cette structure proviennent de la redondance d'informations contenues dans la banque de données. En effet, à l'exception des points situés sur la limite extérieure de la carte, tous les points du fond de carte apparaissent au moins deux fois dans le fichier : en fait, le nombre d'occurrences d'un point est égal au nombre de traits se croisant en ce point, diminué d'une unité si le point est situé sur la limite extérieure de la carte.

Cette redondance entraîne, bien entendu, une consommation de place d'autant plus importante sur le périphérique de mémorisation, mais cet aspect peut être considéré comme secondaire. L'inconvénient majeur vient du fait qu'à chaque occurrence d'un même point dans le fichier, il a fallu réaliser une saisie de ce point. Or, au vu de la précision du dispositif de digitalisation, il est probable qu'à des saisies distinctes d'un même point, correspondent des couples de coordonnées différents. En d'autres termes, on n'est pas assuré de la coïncidence spatiale des différents enregistrements d'un même point. Il peut en résulter, par exemple, une superposition inexacte des traits lors du tracé du fond de carte ou lors du hachurage des zones.

La redondance des enregistrements influence également le temps de saisie et le temps de tracé du fond de carte. En moyenne, le temps de saisie est doublé - si l'on considère que chaque point appartient simultanément à deux zones en moyenne - et il en va de même du temps de tracé - double tracé de la frontière commune de deux zones -.

Enfin, toute modification ultérieure du fond de carte, aussi minime soit-elle, entraînera une réorganisation généralement complexe du fichier de données puisque l'effacement et le réencodage des frontières de plusieurs zones adjacentes est presque systématiquement nécessaire.

D.- EXEMPLES

Une banque de données spatiales organisée selon la structure séquentielle simple est utilisée, en entrée, par plusieurs bibliothèques de programmes cartographiques parmi les plus célèbres.

C'est le cas, par exemple, du logiciel SYMAP dont la première version est développée dans les années 1960 par le Laboratoire de Harvard (Laboratory for Computer Graphics and Spatial Analysis, 1980).

Plus récemment, les programmes de cartographie inclus dans la bibliothèque de programmes SAS/GRAPH ont imposé une structure similaire pour leurs fichiers de données spatiales (SAS Institute Inc., 1981).

Il est clair que, grâce à sa simplicité, il est toujours possible de se ramener à une structure séquentielle et ce, quelle que soit la complexité de la structure de données effectivement utilisée.

IV. - LE DICTIONNAIRE DE POINTS

L'inconvénient majeur de la structure séquentielle simple provient de la double opération de saisie. La structure utilisant un dictionnaire de points constitue une solution intéressante à ce problème. Il ne s'agit pourtant que d'une amélioration de la première structure. Comme celle-ci, la structure à dictionnaire de points est implicitement réservée à la manipulation d'éléments aréaux.

A.- DESCRIPTION

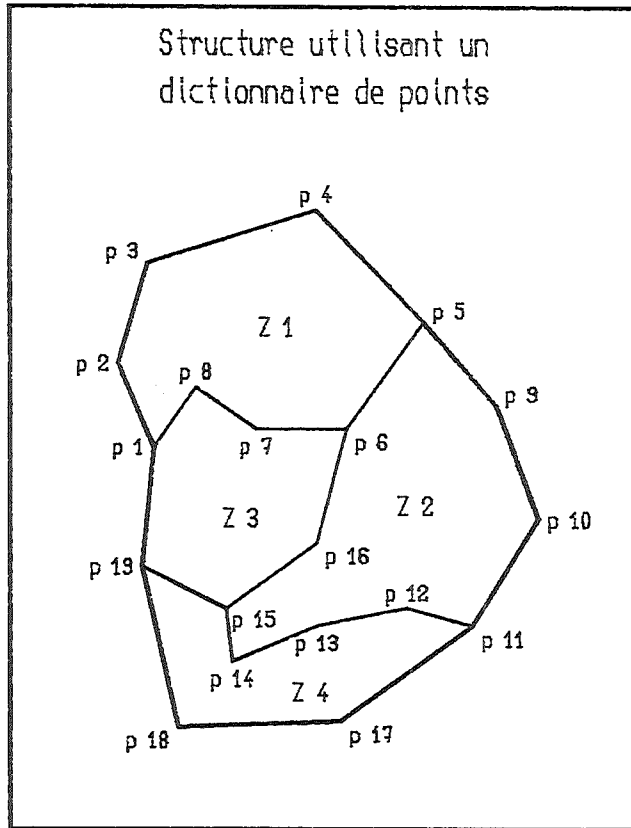
La banque de données spatiales est constituée de deux fichiers (fig. 2). Dans le premier, on mémorise les coordonnées de tous les points du fond de carte. Une étiquette identifiant le point est associée à chaque couple de coordonnées. Ce premier fichier ne contient aucune redondance.

Le second fichier est fort semblable à celui utilisé par la structure séquentielle. Chaque zone y est stockée indépendamment des autres et les zones sont distinguées par des séparateurs. Cependant, ce second fichier ne contient plus les coordonnées des points situés sur les frontières des zones, mais uniquement les étiquettes de ces points. Ces étiquettes servent de pointeurs vers le premier fichier et permettent la reconstitution des limites. Comme dans la structure précédente, ce fichier comporte une forte redondance d'informations mais, cette fois, elle porte sur des étiquettes. Le second fichier est donc indépendant de la digitalisation proprement dite.

B.- AVANTAGES

L'avantage essentiel de cette structure sur la précédente réside dans la saisie unique de tous les points du fond de carte. Il ne faut dès lors plus craindre des erreurs de superposition lors du tracé. Le temps de saisie est, lui aussi, fortement réduit.

D'autre part, les points du fond de carte peuvent être inscrits dans le dictionnaire dans n'importe quel ordre puisque, seul, l'ordre du second fichier est conséquent. Cette caractéristique peut être



Fichier de zones

Pointeur
p1
p2
p3
p4
p5
p6
p7
p8
p1
-1
p19
p15
p9
p10
p11
p12
p13
p14
p15
p16
p6
p5
-1

Dictionnaire

Etiquette	X	Y
p1	X _{p1}	Y _{p1}
p2	X _{p2}	Y _{p2}
p3	X _{p3}	Y _{p3}
p4	X _{p4}	Y _{p4}
p5	X _{p5}	Y _{p5}
p6	X _{p6}	Y _{p6}
p7	X _{p7}	Y _{p7}
p8	X _{p8}	Y _{p8}
p9	X _{p9}	Y _{p9}
p10	X _{p10}	Y _{p10}
p11	X _{p11}	Y _{p11}
p12	X _{p12}	Y _{p12}
p13	X _{p13}	Y _{p13}
p14	X _{p14}	Y _{p14}
p15	X _{p15}	Y _{p15}
p16	X _{p16}	Y _{p16}
p17	X _{p17}	Y _{p17}
p18	X _{p18}	Y _{p18}
p19	X _{p19}	Y _{p19}

FIG. 2.- La structure séquentielle utilisant un dictionnaire de points. Comme dans l'exemple précédent, la fermeture des zones est explicite et le séparateur correspond à un pointeur de valeur égale à -1.

mise à profit pour créer plusieurs banques de données spatiales sur la base d'un seul dictionnaire. Par exemple, des fonds de carte distincts représentant des zones emboîtées (provinces, arrondissements, communes) peuvent être construits sur la base d'un dictionnaire créé au niveau le plus fin du système hiérarchique (communes par exemple).

Vis-à-vis de la maintenance de la banque de données spatiales, les corrections ultérieures ne portant que sur des modifications de coordonnées peuvent être réalisées sans difficultés.

C.- DESAVANTAGES

Il reste pourtant un certain nombre de désavantages inhérents à cette structure. En pratique elle reste prévue pour la mémorisation de zones fermées. Moyennant certains artifices, il est possible de prendre en compte d'autres éléments spatiaux : un élément ponctuel, par exemple, peut être défini par un carré de quatre points très rapprochés.

La préparation du fond de carte, précédant la digitalisation, est un peu plus lourde que dans le cas de la structure séquentielle simple. Il faut, non seulement, repérer les points remarquables du fond de carte, mais aussi associer à chacun d'eux une étiquette. Celle-ci peut se ramener à un simple numéro d'ordre, codé sur un entier pour faciliter les tests et minimiser l'espace mémoire.

Si la saisie de chaque point est unique, le tracé, par contre, est toujours doublé. La redondance présente dans le second fichier de la banque de données spatiales engendre le doublement de toutes les frontières communes de deux zones voisines. Cela peut influencer la qualité du dessin et, de toute façon, cela augmente le temps de tracé.

Enfin, cette structure fait appel à une double lecture pour reconstituer les limites des zones. Si la première lecture est de type séquentiel, la seconde, utilisant les pointeurs vers le dictionnaire, est obligatoirement à accès direct. Cet aspect peut s'avérer contraignant vis-à-vis de certains périphériques de mémoire externe.

D.- EXEMPLES

La structure utilisant un dictionnaire de points est utilisée par plusieurs programmes de cartographie. Le plus célèbre est sans doute le programme CALFORM, postérieur à SYMAP mais distribué également par le Laboratoire de Harvard. CALFORM réalise une cartographie par plages traditionnelle, pour laquelle l'accès séquentiel à tous les points de la frontière d'une zone, assuré par le second fichier de cette structure, constitue un avantage (Laboratory for Computer Graphics and Spatial Analysis, 1980).

Le passage d'une structure utilisant un dictionnaire de points, à une structure séquentielle simple est immédiat. Les fichiers conçus pour CALFORM, par exemple, sont donc facilement convertis pour être utilisés par SYMAP. L'inverse n'est cependant pas évident.

V. - LES STRUCTURES TOPOLOGIQUES

Les deux premières structures qui viennent d'être décrites sont essentiellement utilisées pour le traitement de surfaces. Dans les deux cas, les structures définissent les zones en termes de points successifs, en court-circuitant, en quelque sorte, l'étape de l'élément linéaire. Les structures dont il va maintenant être question rétablissent la hiérarchie complète des éléments spatiaux. Elles offrent, grâce à cela, une plus grande efficacité et de nouvelles possibilités de traitement cartographique.

A.- DESCRIPTION

Ces structures plus performantes sont qualifiées de topologiques car elles assimilent le fond de carte à un graphe et elles retiennent, et utilisent, un certain nombre de règles topologiques qui s'appliquent aux éléments du graphe.

Les points du fond de carte sont les sommets, ou les noeuds, du graphe. Les segments reliant deux sommets sont les arêtes. Enfin, les zones sont constituées d'une séquence d'arêtes et apparaissent, par conséquent, comme des circuits dans le graphe. Chaque arête est définie par les deux sommets fixant ses extrémités et une arête ne peut appartenir, au plus, qu'à deux zones adjacentes - à une seule zone si l'arête figure sur la frontière extérieure du fond de carte -. L'arête constitue, du point de vue topologique, l'élément charnière du fond de carte.

La distinction des différents éléments du graphe conduit, dans la pratique, à la création d'autant de fichiers séparés. La banque de données spatiales est par conséquent constituée de trois fichiers au moins. Les relations topologiques entre les éléments se traduisent, quant à elles, par une série de pointeurs entre les fichiers. En fait, seul le fichier stockant les sommets est explicitement lié à l'opération de digitalisation, les autres fichiers n'étant constitués que de pointeurs. La redondance d'informations est certes toujours présente mais elle est minimum : elle n'apparaît pas dans le fichier de sommets et ne porte que sur des pointeurs.

La définition des pointeurs, mais aussi celle des éléments spatiaux, diffèrent cependant selon les systèmes utilisant ce genre de structure. Deux exemples illustrent la structure topologique et les considérations relatives à ses avantages et désavantages seront distinguées pour chaque système.

B.- LA STRUCTURE DIME

Le concept DIME (Dual Independent Map Encoding), développé par le U.S. Bureau of the Census, est à la base d'un des premiers et des plus célèbres systèmes opérationnels faisant référence à une structure topologique de données spatiales. Il a été conçu pour permettre l'édition semi-automatique de bases de données décrivant les réseaux urbains de voirie et les unités statistiques intra-urbaines utilisés pour les recensements américains dès 1970 (U.S. Bureau

of the Census, 1970).

1. *Description.* - L'élément charnière de cette structure est l'arête qui correspond le plus souvent, en pratique, à une face d'un flot urbain. Les lignes complexes - rues courbes, lignes de chemin-de-fer, etc. - sont représentées par une série de segments successifs. La disposition de ces arêtes détermine des blocs, ou flots, pour lesquels les données du recensement sont agrégées.

La banque de données correspondante se résume à un fichier de coordonnées, mémorisant les sommets du graphe, et à un fichier d'arêtes (fig. 3). Chaque enregistrement de ce dernier fichier correspond à une arête et comporte plusieurs variables. Variables urbanistiques, tout d'abord, telles que le nom de la rue et les intervalles de numéros le long des deux côtés de la rue, pour le tronçon correspondant à l'arête. Des variables de type pointeur ensuite : deux pointeurs vers les deux sommets fixant les extrémités de l'arête et deux codes identifiant les deux blocs de part et d'autre de l'arête.

2. *Avantages.* - Cette structure permet une édition très aisée de la banque de données et elle est bien adaptée au traitement et à la mise à jour des données de recensement. Il est aisé, par exemple, d'affecter à un bloc particulier, les données relatives à un ménage en se servant uniquement de son adresse postale.

3. *Désavantages.* - Pour des traitements cartographiques diversifiés, la structure DIME est trop rigide. La reconstitution de la frontière des blocs, la reconnaissance des éléments linéaires simples, etc. demandent de nombreuses manipulations et un temps assez considérable. En outre, certains types de traitements, tels que la généralisation des lignes par exemple, sont pratiquement à exclure.

C. - LA STRUCTURE POLYVRT

Le système intitulé POLYVRT a été développé par le Laboratoire de Harvard mais, en fait, il synthétise plusieurs travaux originaux préalables. Il utilise une distinction poussée des éléments spatiaux constituant le fond de carte et a servi de référence à plusieurs autres systèmes développés par la suite (Laboratory for Computer Graphics and Spatial Analysis, 1974).

1. *Description.* - L'élément de base de cette structure reste linéaire. Il ne s'agit plus d'un simple segment mais d'une "chaîne". Les chaînes peuvent être constituées de plusieurs segments successifs. Une chaîne peut donc contenir plusieurs points intermédiaires qui doivent être distingués des noeuds fixant ses extrémités. Un point intermédiaire n'appartient qu'à une seule chaîne, tandis qu'un noeud est généralement partagé par plusieurs chaînes. Enfin, une chaîne ne peut appartenir, au plus, qu'à deux zones adjacentes.

Dans ce système, les noeuds sont les points remarquables qui supportent le graphe analogue du fond de carte et qui déterminent

Fichier de noeuds

Etiquette	X	Y
N1	XN1	YN1
N2	XN2	YN2
N3	XN3	YN3
N4	XN4	YN4
N5	XN5	YN5
N6	XN6	YN6
N7	XN7	YN7
N8	XN8	YN8
N9	XN9	YN9
N10	XN10	YN10
N11	XN11	YN11
N12	XN12	YN12
N13	XN13	YN13
N14	XN14	YN14
N15	XN15	YN15
N16	XN16	YN16
N17	XN17	YN17
N18	XN18	YN18
N19	XN19	YN19

Fichier de chaines

Etiquette	Variables urbanistiques	Noeud de depart	Noeud de destination	Zone gauche	Zone droite
C1	...	N1	N2	0	Z1
C2	...	N2	N3	0	Z1
C3	...	N3	N4	0	Z1
C4	...	N4	N5	0	Z1
C5	...	N5	N6	Z2	Z1
C6	...	N6	N7	Z3	Z1
C7	...	N7	N8	Z3	Z1
C8	...	N8	N1	Z3	Z1
C9	...	N5	N9	0	Z2
C10	...	N9	N10	0	Z2
C11	...	N10	N11	0	Z2
C12	...	N11	N12	Z4	Z2
C13	...	N12	N13	Z4	Z2
C14	...	N13	N14	Z4	Z2
C15	...	N14	N15	Z4	Z2
C16	...	N15	N16	Z3	Z2
C17	...	N16	N6	Z3	Z2
C18	...	N11	N17	0	Z4
C19	...	N17	N18	0	Z4
C20	...	N18	N19	0	Z4
C21	...	N19	N15	Z3	Z4
C22	...	N19	N1	0	Z3

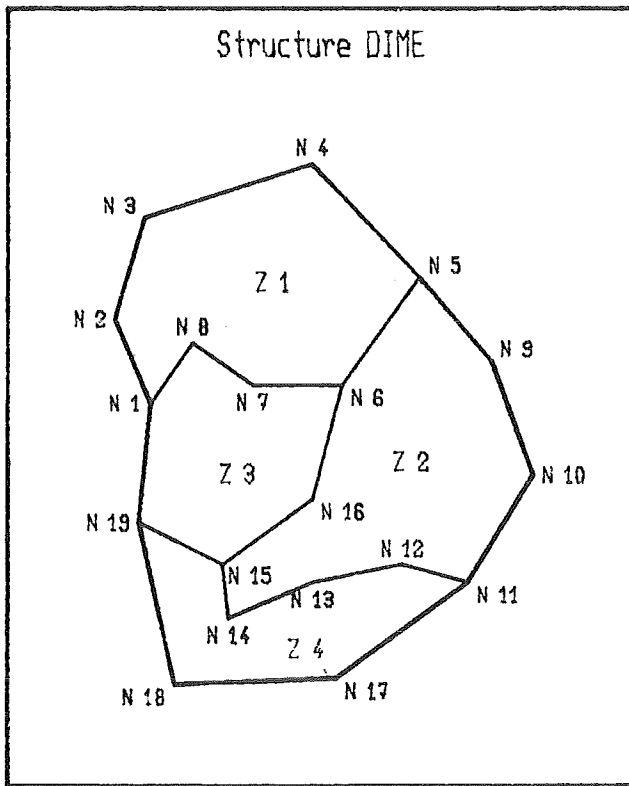


FIG. 3.- La structure du système DIME (U.S. Bureau of the Census, 1970). Les arêtes correspondent aux segments reliant deux noeuds successifs. Se rapporter au texte pour des exemples de variables urbanistiques.

Fichier de zones

Etiquette	Liste de pointeurs de chaine
Z1	C1 C2 C3
Z2	C4 C5 C6 C2
Z3	C3 C6 C7 C8
Z4	C9 C7 C5

Fichier de chaines

Etiquette	Nbre de points	Noeud de depart	Noeud de destination	Zone gauche	Zone droite
C1	3	N1	N2	0	Z1
C2	0	N2	N3	Z2	Z1
C3	2	N3	N1	Z3	Z1
C4	2	N2	N4	0	Z2
C5	3	N4	N5	Z4	Z2
C6	1	N5	N3	Z3	Z2
C7	0	N5	N6	Z4	Z3
C8	0	N6	N1	0	Z3
C9	2	N4	N6	0	Z4

Fichier de noeuds

Etiquette	X	Y
N1	XN1	YN1
N2	XN2	YN2
N3	XN3	YN3
N4	XN4	YN4
N5	XN5	YN5
N6	XN6	YN6

Fichier des points intermediaires

Etiquette de chaine	Liste de coordonnees des points
C1	[Xp1,Yp1] [Xp2,Yp2] [Xp3,Yp3]
C3	[Xp4,Yp4] [Xp5,Yp5]
C4	[Xp6,Yp6] [Xp7,Yp7]
C5	[Xp8,Yp8] [Xp9,Yp9] [Xp10,Yp10]
C6	[Xp11,Yp11]
C9	[Xp12,Yp12] [Xp13,Yp13]

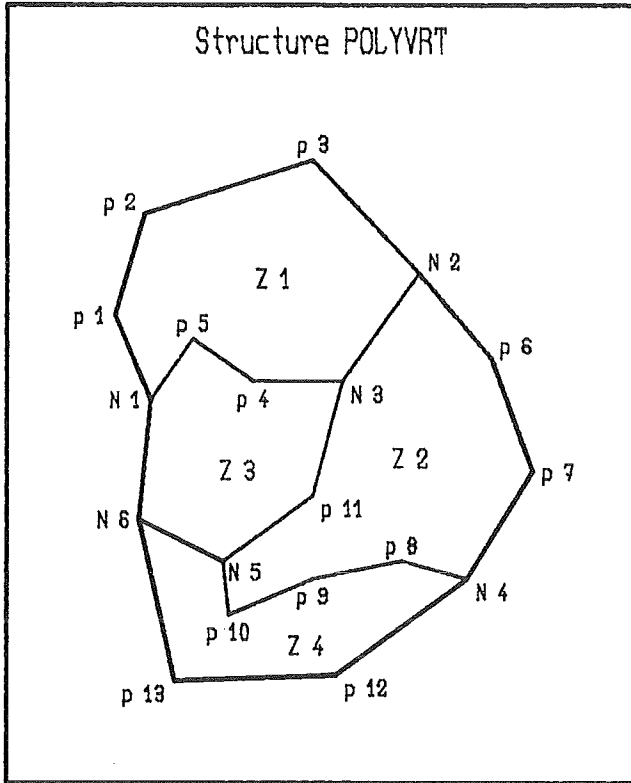


FIG. 4.- La structure POLYVRT
(Laboratory of Computer Graphics and Spatial Analysis, 1974).

toute la structure de voisinage des zones. Les points intermédiaires, par contre, peuvent être considérés comme porteurs d'une information de détail et cette distinction entre noeuds et points intermédiaires pourra être mise à profit pour réaliser une généralisation du fond de carte.

Les zones sont définies par une séquence de chaînes. A la limite, une seule chaîne peut constituer la frontière, dans le cas d'une île par exemple - ici, en outre, les deux noeuds fixant les extrémités de la chaîne coïncident -.

La banque de données comporte un fichier de noeuds qui est distinct d'un fichier de points intermédiaires. Le fichier de chaînes contient des enregistrements constitués uniquement de pointeurs : vers les noeuds situés aux extrémités, vers les éventuels points intermédiaires et vers les deux zones situées de part et d'autre de la chaîne. Le fichier de zones reprend, pour chaque zone, les pointeurs vers les chaînes nécessaires à la reconstitution des frontières (fig. 4).

2. *Avantages* - Ce système est beaucoup plus riche que tous les précédents et il permet d'envisager tous les traitements cartographiques classiques. Toutes les corrections et mises à jour de la banque de données spatiales sont réalisées très aisément. On peut éviter le doublement du tracé, inévitable avec les deux premières structures, en utilisant le seul fichier de chaînes comme input de la procédure de tracé.

La structure supporte simultanément tous les types d'éléments spatiaux puisqu'à côté des zones, utilisant toute la structure, il est possible de manipuler des éléments ponctuels et linéaires simples, n'utilisant qu'une partie de la structure POLYVRT. Les éléments ponctuels n'utilisent que la sous-structure des noeuds tandis que les éléments linéaires ignorent la sous-structure des zones.

Il est évidemment très simple de reconstituer, au départ de POLYVRT, les structures élémentaires non topologiques utilisées par les programmes de cartographie cités précédemment. Une séquentialisation de la banque de données peut aussi être requise pour stocker et accéder aux informations spatiales sur un support mémoire de type séquentiel.

3. *Désavantages* - Ce système possède plusieurs inconvénients mais la plupart sont inhérents à toute structure topologique.

C'est ainsi que la préparation du fond de carte, précédant la digitalisation, est particulièrement lourde. Il faut distinguer et étiquetter les noeuds, les chaînes et les zones et il faut marquer les points intermédiaires. D'autre part, la structure implique un accès aléatoire à tous les fichiers constituant la banque de données. L'accès aux enregistrements étant effectué par l'intermédiaire de plusieurs pointeurs, le temps nécessaire à la lecture des informations, pour la reconstitution d'une frontière par exemple, peut devenir non négligeable.

D'autres inconvénients sont liés à la structure POLYVRT elle-même. Elle n'est pas très bien adaptée au traitement des zones emboîtées. La relation biunivoque entre les chaînes et les zones, par l'intermédiaire de leurs pointeurs respectifs, empêche la construction

de plusieurs types de zones, par exemple des zones hiérarchiques emboîtées, sur la base de sous-structures noeuds et chaînes identiques. En outre, POLYVRT ne permet pas le traitement des enclaves. La définition de plusieurs limites distinctes - frontières extérieures et intérieures (zone comportant une enclave) ou plusieurs frontières extérieures (cas de l'archipel) - n'est pas prévue par le système, or c'est là un problème fréquent en cartographie. Des adaptations plus ou moins complexes de la structure originale sont proposées par la littérature pour remédier à ces quelques lacunes du système (voir ainsi : Peucker T.K. et Chrisman N., 1975; Edwards R.G., Durfee R.C. et Coleman P.R., 1977 et Corona Burgueno J.F., 1980).

D. - AMENDEMENTS DE LA STRUCTURE POLYVRT

La structure examinée maintenant est fort voisine de la structure POLYVRT. Cependant, elle comble certaines lacunes propres à cette dernière et, d'autre part, elle est légèrement simplifiée de manière à rester compatible avec l'utilisation d'une configuration micro-informatique, pour laquelle les temps d'entrée-sortie et la capacité de mémoire sont toujours critiques (Donnay J.-P., 1985).

Cette nouvelle structure topologique utilise des éléments identiques à ceux de POLYVRT (fig. 5). Les noeuds, les chaînes et les zones conservent la même définition et constituent autant de fichiers distincts. On constate que deux fichiers distincts sont réservés pour les zones, mais il s'agit d'une simple question de souplesse dans l'édition de la banque de données. Cette distinction reste de toute façon transparente pour l'utilisateur.

La première modification importante consiste à supprimer les pointeurs de zones au sein du fichier de chaînes. On a signalé que ces pointeurs faisaient, en quelque sorte, double emploi avec les pointeurs de chaînes présents dans le fichier de zones et que cette relation biunivoque empêchait la définition de zones emboîtées. Cette suppression allège donc la structure et permet de lever une des limites de POLYVRT. Il est clair cependant que cette suppression diminue la capacité d'édition de la banque de données puisque certaines fonctions précises vont nécessiter un examen complet des fichiers pour retrouver les relations topologiques supprimées.

Une seconde modification, la plus importante, résulte de l'introduction d'un code supplémentaire associé à chaque chaîne. Ce code intitulé CT ou code de tracé sur la figure, correspond en fait à un attribut thématique affecté à la chaîne. En utilisant des valeurs différentes pour ce code lors de la saisie des chaînes, il est possible de distinguer, dès le départ, les types de chaînes que l'on enregistre. Cette distinction est bien d'ordre thématique puisque l'on peut associer chaque fois des valeurs différentes pour des frontières administratives communales ou nationales, pour des cours d'eau ou des canaux, pour des routes ou des autoroutes, pour les chemins de fer, pour les lignes à haute tension, etc. Le nombre de valeurs distinctes n'est pas limité et le choix des valeurs est absolument libre. Cette distinction est particulièrement pratique pour les opéra-

Fichier des points intermediaires

Etiquette	X	Y
p1	Xp1	Yp1
p2	Xp2	Yp2
p3	Xp3	Yp3
p4	Xp4	Yp4
p5	Xp5	Yp5
p6	Xp6	Yp6
p7	Xp7	Yp7
p8	Xp8	Yp8
p9	Xp9	Yp9
p10	Xp10	Yp10
p11	Xp11	Yp11
p12	Xp12	Yp12
p13	Xp13	Yp13

Liste de chaines

Etiquette	Pointeur de chaine
L1	C1
L2	C2
L3	C3
L4	C4
L5	C5
L6	C6
L7	C2
L8	C3
L9	C6
L10	C7
L11	C8
L12	C9
L13	C7
L14	C5

Fichier de chaines

Etiquette	Noeud de depart	Noeud de destination	1er point intermed.	Nbre de points	Code de trace
C1	N1	N2	p1	3	2
C2	N2	N3		0	1
C3	N3	N1	p4	2	1
C4	N2	N4	p6	2	2
C5	N4	N5	p8	3	1
C6	N5	N3	p11	1	1
C7	N5	N6		0	1
C8	N6	N1		0	2
C9	N4	N6	p12	2	2

Fichier de noeuds

Etiquette	X	Y
N1	XN1	YN1
N2	XN2	YN2
N3	XN3	YN3
N4	XN4	YN4
N5	XN5	YN5
N6	XN6	YN6

Fichier de zones

Etiquette	Pointeur vers liste de chaines	Nbre de chaines
Z1	L1	3
Z2	L4	4
Z3	L8	4
Z4	L12	3

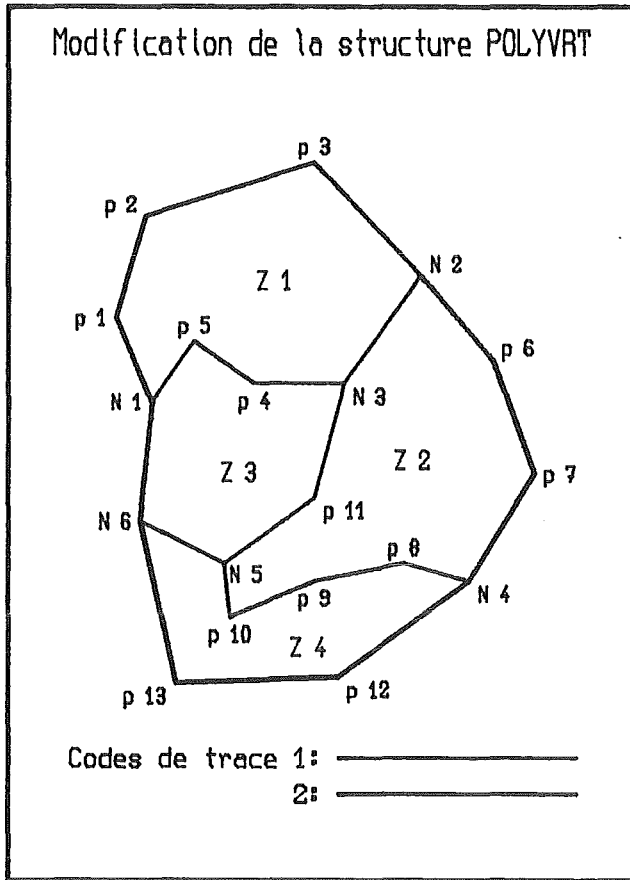


FIG. 5.- Un exemple de modification de la structure POLYVRT (Donnay J.-P., 1985). Les valeurs du code de tracé sont associées, lors du dessin, à des types de trait définis par l'utilisateur.

tions de dessin ultérieures puisque la possibilité sera donnée à l'utilisateur de faire correspondre, à chaque valeur du code, un type de tracé différent choisi, lui aussi, par l'utilisateur (fig. 6).

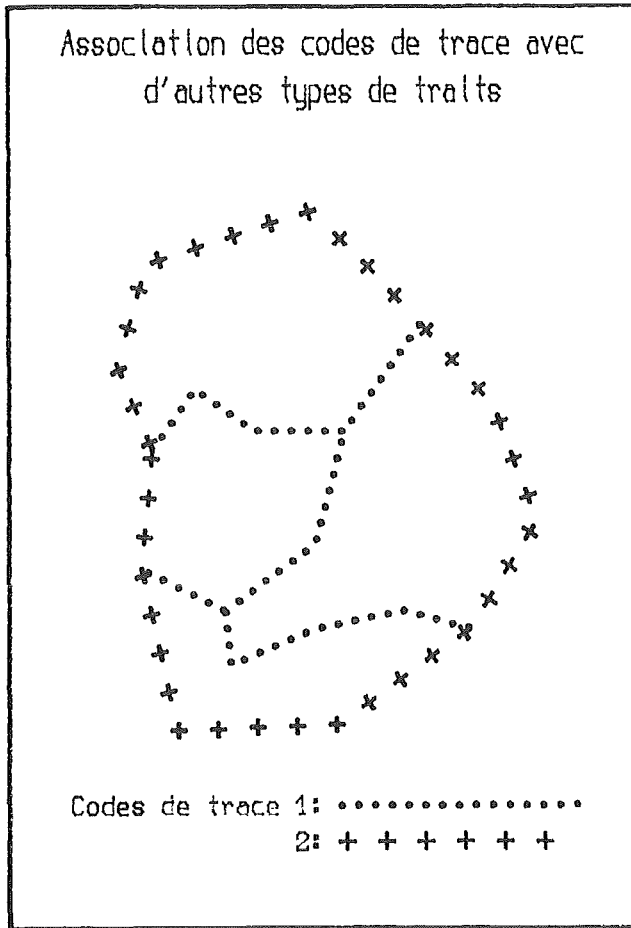


FIG. 6.- Même banque de données spatiales qu'à la figure 5 mais association, lors du dessin, d'autres types de trait aux valeurs du code de tracé.

La présence de ce code va permettre de lever d'autres limites de la structure topologique, et cela en réservant une signification particulière à deux valeurs du code.

Une première valeur sera associée à un élément linéaire concret mais dont on ne souhaite pas le tracé. Ainsi par exemple, la limite

Fichier des points intermediaires

Etiquette	X	Y
p1	Xp1	Yp1
p2	Xp2	Yp2
p3	Xp3	Yp3
p4	Xp4	Yp4
p5	Xp5	Yp5
p6	Xp6	Yp6
p7	Xp7	Yp7
p8	Xp8	Yp8
p9	Xp9	Yp9
p10	Xp10	Yp10
p11	Xp11	Yp11
p12	Xp12	Yp12
p13	Xp13	Yp13
p14	Xp14	Yp14
p15	Xp15	Yp15

Fichier de noeuds

Etiquette	X	Y
N1	XN1	YN1
N2	XN2	YN2
N3	XN3	YN3
N4	XN4	YN4
N5	XN5	YN5
N6	XN6	YN6
N7	XN7	YN7

Fichier de zones

Etiquette	Pointeur vers liste de chaines	Nbre de chaines
Z1	L1	3
Z2	L4	7
Z3	L11	4
Z4	L15	3

Fichier de chaines

Etiquette	Noeud de depart	Noeud de destination	1er point intermed.	Nbre de points	Code de trace
C1	N1	N2	p1	3	2
C2	N2	N3		0	1
C3	N3	N1	p4	2	1
C4	N2	N4	p6	2	2
C5	N4	N5	p8	3	1
C6	N5	N3	p11	1	1
C7	N5	N6		0	1
C8	N6	N1		0	2
C9	N4	N6	p12	2	2
C10	N4	N7		0	-1
C11	N7	N7	p14	2	2

Liste de chaines

Etiquette	Pointeur de chaine
L1	C1
L2	C2
L3	C3
L4	C4
L5	C10
L6	C11
L7	C10
L8	C5
L9	C6
L10	C2
L11	C3
L12	C6
L13	C7
L14	C8
L15	C9
L16	C7
L17	C5

Utilisation d'une chaîne non opérationnelle
pour manipuler une enclave

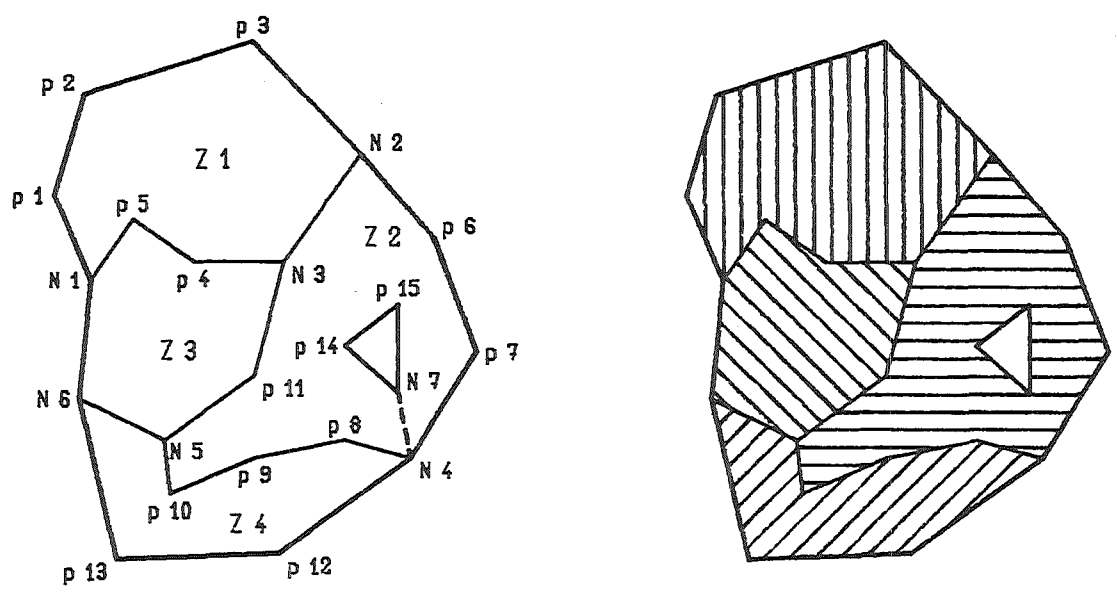


FIG. 8.- Introduction d'une enclave dans le fond de carte et prise en compte de celle-ci dans la banque de données spatiales. Le dessin de droite illustre le rôle de la chaîne non opérationnelle (C10) lors de l'application de la procédure de hachurage.

ne peuvent pas servir à la délimitation d'une plage de couleur ou d'une trame. Leur rôle consiste uniquement à relier entre elles des parties distinctes de la frontière d'une zone. On utilisera une chaîne non opérationnelle pour relier la frontière extérieure d'une zone avec la frontière intérieure délimitant une enclave. C'est aussi une chaîne non opérationnelle qui reliera la frontière continentale d'une zone avec la frontière isolée d'une île appartenant à la même zone. Cette solution peut accepter plusieurs niveaux d'enclaves ou relier un groupe de zones isolées comme dans le cas d'un archipel (fig. 8).

La chaîne non opérationnelle ne correspond à rien de concret, elle ne possédera pas de points intermédiaires. Elle pourra, éventuellement, croiser certaines chaînes réelles du fond de carte sans interférer sur les algorithmes de dessin. Elle ne devra pas intervenir sur le calcul du périmètre des frontières, mais par contre elle permettra l'évaluation exacte des surfaces des zones et le positionnement plus correct des centres de gravité des zones. Elle permet donc de résoudre simplement plusieurs problèmes importants de la cartographie par plage et facilite considérablement le dessin des éléments linéaires du fond de carte.

Pour en terminer avec les structures topologiques, il faut encore savoir que toutes celles qui viennent d'être présentées sont qualifiées de structures de premier niveau car elles servent de point de départ à d'autres structures spécialisées permettant des traitements cartographiques très particuliers. Par exemple, une sous-structure de noeuds peut servir de base à une structure de triangles utilisée par certains algorithmes d'interpolation pour la recherche d'isolignes (Debuyl D., 1985). Les sous-structures de chaînes et de zones permettent de construire des modèles de segments et des modèles de facettes, requis par des algorithmes de cartographie par prisme en 3 dimensions (Flener P., 1985). Et on pourrait multiplier les exemples de généralisation de la structure topologique.

VI. - UN EXEMPLE DE STRUCTURE RELATIONNELLE

Les structures topologiques ne constituent pas le modèle le plus complet en matière de banque de données spatiales. Si l'on se réfère à la théorie des bases de données, on constate que les structures topologiques relèvent des modèles hiérarchiques, au mieux des modèles en réseau mais qu'elles n'atteignent pas les modèles relationnels. Ce dernier type d'organisation de données est sensiblement plus évolué. Les pointeurs, les chemins d'accès et les liens entre les données sont déterminés de manière dynamique. Ils requièrent un langage de recherche utilisant notamment toute la série des opérateurs logiques. Depuis une dizaine d'années, on trouve dans la littérature, essentiellement informatique d'ailleurs, des propositions de systèmes relationnels de base de données adaptées aux données spatiales.

Un seul exemple illustrera cette approche. Il s'agit d'ailleurs d'une structure qui fait la liaison entre les structures topologiques classiques et le modèle relationnel. C'est la structure qualifiée de relationnelle et récursive élaborée par Shapiro et Haralick (Shapiro L.C. et Haralick R.M., 1980).

Le système prévoit la définition d'un nombre fini de types de structures spatiales : états, communes, rivières, etc. A chaque type, correspond un nom et un prototype de structure qui mentionne le nombre et le genre de relations qui peuvent être associées à ce type. Les relations sont définies par une table d'attributs/valeurs, où les valeurs peuvent être, soit un atome : un nombre ou une chaîne de caractères par exemple, soit un autre type de structure. La mise en place du système inclut trois dictionnaires reprenant : les noms de structures, les types de structures et les genres de relations admises pour la banque de données envisagée. Il est possible, via ce système de retrouver toutes les structures topologiques classiques mais aussi les structures propres au mode graphique raster. En outre, cette organisation, contrairement aux précédentes, admet directement l'association entre les données spatiales et les données thématiques.

Si on reprend l'exemple ayant illustré les structures précédentes (fig. 9), on peut imaginer la représentation suivante. Trois types de structures : les zones, les chaînes et les points. La table d'attributs des zones contient d'une part des atomes : surface, centre de gravité ou toute autre information de nature thématique se rapportant aux zones, et d'autre part la même table renvoie à la structure chaîne. La table d'attributs des chaînes, d'une part, renvoie à la structure des points, et d'autre part contient des attributs topologiques et thématiques propres aux chaînes : pointeurs vers les chaînes et les zones voisines et longueur de la chaîne par exemple. Enfin on peut imaginer que la table d'attributs de la structure des points ne contiennent que deux attributs du genre atome, correspondant aux coordonnées des points en question. Il est clair que l'on peut envisager des tables d'attributs/valeurs sensiblement plus complexes, même pour ce petit exemple. Cependant, l'implantation de ce genre de système n'est pas aisée et, à la limite, on peut définir un système différent pour chaque application, le plus petit dénominateur commun à toutes ces variantes restant la structure topologique simple, telle qu'elle a été présentée précédemment.

VII. - CONCLUSION

Différentes structures de données spatiales sont décrites dans cet article, depuis la structure séquentielle simple jusqu'aux structures relationnelles, en insistant plus particulièrement sur diverses versions de structures topologiques. L'énumération est loin d'être exhaustive. Elle se veut au contraire sélective, de manière à mettre en évidence la complexité croissante des concepts mis en oeuvre.

La structuration des données spatiales reste un sujet de recherche particulièrement fécond et des progrès substantiels sont encore nécessaires pour éliminer ce "goulet d'étranglement" que constitue la saisie de données. Si l'expression "cartographie automatique" sous-entend le traitement cartographique de données sans intervention humaine, la seule étape de la saisie des données enlève, dans l'état actuel des connaissances, toute signification à cette expression.

Table Attributs/Valeurs de C1

Attributs	Valeurs
Pointeur vers zone gauche	
Pointeur vers zone droite	Z1
Chaine suivante a gauche	C4
Chaine suivante a droite	C2
Longueur de C1 (atome)	L(C1)

Liste des points | p1 p2 p3 p4 p5

Table Attributs/Valeurs de C2

Attributs	Valeurs
Pointeur vers zone gauche	Z2
Pointeur vers zone droite	Z1
Chaine suivante a gauche	C6
Chaine suivante a droite	C3
Longueur de C2 (atome)	L(C2)

Table Attributs/Valeurs de C3

Attributs	Valeurs
Pointeur vers zone gauche	Z3
Pointeur vers zone droite	Z1
Chaine suivante a gauche	C8
Chaine suivante a droite	C1
Longueur de C3 (atome)	L(C3)

Liste des points | p6 p7 p8 p1

Table Attributs/Valeurs de C4

Attributs	Valeurs
Pointeur vers zone gauche	
Pointeur vers zone droite	Z2
Chaine suivante a gauche	C9
Chaine suivante a droite	C5
Longueur de C4 (atome)	L(C4)

Liste des points | p5 p9 p10 p11

Table Attributs/Valeurs de C5

Attributs	Valeurs
Pointeur vers zone gauche	Z4
Pointeur vers zone droite	Z2
Chaine suivante a gauche	C7
Chaine suivante a droite	C6
Longueur de C5 (atome)	L(C5)

Liste des points | p11 p12 p13 p14 p15

Table Attributs/Valeurs de C6

Attributs	Valeurs
Pointeur vers zone gauche	Z3
Pointeur vers zone droite	Z2
Chaine suivante a gauche	C3
Chaine suivante a droite	C2
Longueur de C6 (atome)	L(C6)

Liste des points | p15 p16 p6

Table Attributs/Valeurs de C7

Attributs	Valeurs
Pointeur vers zone gauche	Z4
Pointeur vers zone droite	Z3
Chaine suivante a gauche	C9
Chaine suivante a droite	C8
Longueur de C7 (atome)	L(C7)

Table Attributs/Valeurs de C8

Attributs	Valeurs
Pointeur vers zone gauche	
Pointeur vers zone droite	Z3
Chaine suivante a gauche	C1
Chaine suivante a droite	C3
Longueur de C8 (atome)	L(C8)

Table Attributs/Valeurs de C9

Attributs	Valeurs
Pointeur vers zone gauche	
Pointeur vers zone droite	Z4
Chaine suivante a gauche	C8
Chaine suivante a droite	C7
Longueur de C9 (atome)	L(C9)

Liste des points | p11 p17 p18 p19

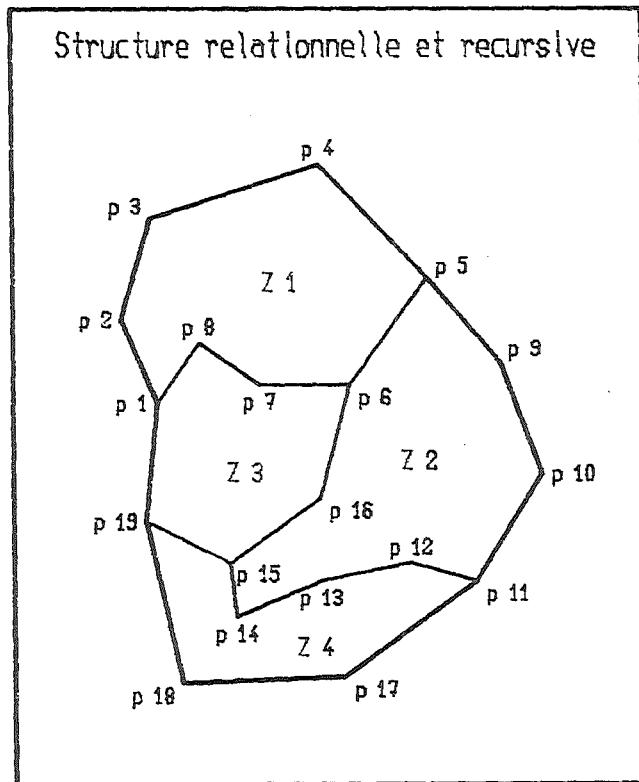


Table Attributs/Valeurs de Z1

Attributs	Valeurs
Pointeur vers 1ere chaine	C1
Position (vs 1ere chaine)	droite
Surface de Z1 (atome)	S(Z1)
Centroide de Z1 (atome)	C(Z1)

Table Attributs/Valeurs de Z3

Attributs	Valeurs
Pointeur vers 1ere chaine	C3
Position (vs 1ere chaine)	gauche
Surface de Z3 (atome)	S(Z3)
Centroide de Z3 (atome)	C(Z3)

Table Attributs/Valeurs de Z2

Attributs	Valeurs
Pointeur vers 1ere chaine	C4
Position (vs 1ere chaine)	droite
Surface de Z2 (atome)	S(Z2)
Centroide de Z2 (atome)	C(Z2)

Table Attributs/Valeurs de Z4

Attributs	Valeurs
Pointeur vers 1ere chaine	C9
Position (vs 1ere chaine)	droite
Surface de Z4 (atome)	S(Z4)
Centroide de Z4 (atome)	C(Z4)

FIG. 9.- Reconstitution des relations topologiques classiques et prise en compte d'attributs thématiques par la structure relationnelle et récursive de Shapiro et Haralick (Shapiro L.C. et Haralick R.M., 1980). Les attributs thématiques présents dans cet exemple n'ont qu'un rôle illustratif.

BIBLIOGRAPHIE

- CORONA BURGUEÑO J.F., 1980. - A Geographical Data Base, in Blaser A. (Ed.), *Data Base Techniques for Pictorial Applications*, Lecture Notes in Computer Science, vol. 81, Springer-Verlag, Berlin, pp. 347-363.
- DEBUYL D., 1985. - *Création de cartes par isolignes*, Mémoire de Graduat en informatique, Ecole Provinciale Supérieure d'Informatique de Seraing, inédit.
- DONNAY J.-P., 1985. - *Concepts, méthodes et apports de la cartographie assistée par micro-ordinateur*, Thèse de doctorat en Sciences géographiques, Université de Liège, inédit.
- EDWARDS R.G., DURFEE R.C. and COLEMAN P.R., 1977. - Definition of a hierarchical polygonal data structure and the associated conversion of a geographic base file from boundary segment format, *Harvard Papers on Geographic Information Systems*, Vol. 4, pp. 1-26.
- FLENER P., 1985. - *Cartographie par prismes*, Mémoire de Graduat en informatique, Ecole Provinciale Supérieure d'Informatique de Seraing, inédit.
- JOHANNSSEN T.M., 1979. - How to get cartographic data in computer systems, in Van Zuylen L. (Ed.), *Computer-Assisted Cartography*, A.C.I. Commission III, Enschede, pp. 37-85.
- LABORATORY FOR COMPUTER GRAPHICS AND SPATIAL ANALYSIS, 1974. - *POLYVRT Manual*, Harvard University, Cambridge (Mass.).
- LABORATORY FOR COMPUTER GRAPHICS AND SPATIAL ANALYSIS, 1980. - *LABLOG*, Harvard Graduate School for Design, Cambridge (Mass.).
- PEUCKER T.K. et CHRISMAN N., 1975. - Cartographic Data Structures, *The American Cartographer*, vol. 2, n° 1, pp. 55-69.
- SAS INSTITUTE INC., 1981. - *SAS/GRAPH User's Guide*, Cary (N.C.).
- SHAPIRO L.C. et HARALICK R.M., 1980. - A spatial Data structure, *Geo-Processing*, vol. 1, pp. 313-337.
- U.S. BUREAU OF THE CENSUS, 1970. - *Census Use Study : The DIME Geocoding System*, Report n° 4, Washington (D.C.).
-