

MODÉLISATION DE LA GÉOMÉTRIE DES OBJETS LINÉAIRES EN VUE DE LEUR GÉNÉRALISATION AUTOMATIQUE

Corinne Plazanet

IGN-France

Service de la Recherche - Laboratoire COGIT

2, avenue Pasteur - BP 68 - 94160 Saint-Mandé FRANCE

Tel : (33-1) 43 98 84 38 - Fax : (33-1) 43 98 81 71

E-Mail : plazanet@cogit.ign.fr

ABSTRACT

This paper underlines the necessity to describe the geometry of linear features prior to automated cartographic generalisation operations. We propose a hierarchical approach, which proceeds by segmenting lines and qualifying sections detected according to different criteria at several levels. Some first experimental results of segmentation and sections qualification are proposed at the end of the paper.

KEYWORDS : *Automated linear feature generalisation, segmentation, measurement, clustering, characteristic point detection.*

RÉSUMÉ

Ce papier met l'accent sur la nécessité de modéliser la géométrie des objets géographiques linéaires préalablement à toute opération de généralisation cartographique automatisée. Nous proposons un processus hiérarchique de description des lignes cartographiques basé sur des opérations de segmentation et de qualification de la géométrie à différents niveaux d'analyse. Les premiers résultats expérimentaux de segmentation et de qualification de la géométrie sont présentés à la fin de cet article.

1. INTRODUCTION À LA GÉNÉRALISATION DES OBJETS LINÉAIRES

La généralisation peut être vue comme un processus holistique de part le grand nombre de contraintes à prendre en compte [14]. Un des principaux problèmes réside dans le fait que chaque objet doit être considéré dans un contexte cartographique particulier, du fait de sa géométrie propre, de son environnement cartographique (présence d'objets voisins et inter-relations) et du niveau de généralisation visé. Nous serons donc amenés, dans le cadre d'un processus de généralisation automatique, à formaliser un grand nombre de contraintes cartographiques qui interviennent dans un tel processus.

Le sujet abordé est restreint à la généralisation cartographique des objets linéaires. Cette opération consiste schématiquement à :

- sélectionner les objets à retenir à l'échelle visée;
- simplifier les caractères géométriques non pertinents des objets retenus;
- amplifier les formes géométriques importantes;
- déplacer ces objets sans détériorer les formes locales et globales;
- harmoniser l'aspect global en gérant les conflits intervenant entre objets voisins.

Nous traitons ici des manipulations géométriques (simplification, amplification) sur des objets linéaires (notamment des routes), indépendamment des autres objets.

Un cartographe qui généralise manuellement a une perception globale et continue de la ligne. Il utilise pleinement son sens artistique et esthétique et son expérience personnelle. La plupart du temps, la connaissance qu'il utilise pour généraliser n'est pas formalisée. A quel moment décide-t-il de supprimer un virage afin d'éviter qu'il se retrouve de l'autre côté d'un talweg par exemple ? Quelles sont les catégories de connaissances qui guident le cartographe à prendre les "bonnes" décisions de généralisation ? Nous sommes à même d'ores et déjà, en regardant la manière de faire des cartographes, de formaliser les règles suivantes. Il faut :

- respecter les formes géométriques locales et globales des objets;
- préserver la topologie intrinsèque des objets;
- prendre en compte la largeur du symbole;
- prendre en compte la nature sémantique des objets.

Lors des changements d'échelle, il convient de ne pas détériorer le message cartographique que porte chaque objet. Ce message peut provenir à la fois de la nature sémantique de l'objet et de sa géométrie propre. Selon le thème de la carte, le message aura une influence plus ou moins grande sur le processus de généralisation. Par exemple, si un virage dangereux se trouve sur une route mollement sinueuse, il faudra conserver cette information (surtout dans le cas d'une carte routière), même si l'empâtement du virage est inférieur à la largeur du signe conventionnel [16]. Inversement, supprimer un virage sur une route fortement sinueuse aura une incidence moins importante sur le message porté par la route. Ces notions d'environnement de formes et de résolution sont essentielles. L'aspect contextuel de la généralisation des formes géométriques, souvent lié aux connaissances non formulées des décisions prises par le cartographe, est particulièrement difficile à modéliser.

Parmi les algorithmes de généralisation développés, nombreux sont ceux qui concernent la simplification des lignes (opérations de filtrage et de lissage) : on peut citer les algorithmes de Douglas, de Lang, de Thapa, de Jenks, de Brophy, etc. Cependant, d'après [2], [5], [7], [9], ces algorithmes sont complexes à appliquer pour tous les types de lignes (voire inutilisables dans certains cas), même par des experts en généralisation. En effet, des séquences complexes d'opérations de généralisation doivent être mises en oeuvre telles que la simplification, le lissage, la caricature avec des réglages délicats de valeurs de paramètres pour chaque étape, et ce pour chaque ligne.

La figure 1 présente respectivement de gauche à droite : un exemple de route BDCarto® (figure 1a), une généralisation manuelle au 1/250 000 (figure 1b) et le résultat d'une simplification suivi d'un léger lissage (figure 1c). Cet exemple montre que l'utilisation de l'algorithme de Douglas produit un tracé trop anguleux et inesthétique, et induit rapidement une dégradation de la topologie de l'arc. La forme Φ porte un message important pour la route, elle doit donc être conservée et amplifiée lors de la généralisation. Avec l'algorithme de Douglas, elle a été maintenue, mais devrait cependant subir une amplification à l'image de la généralisation manuelle.

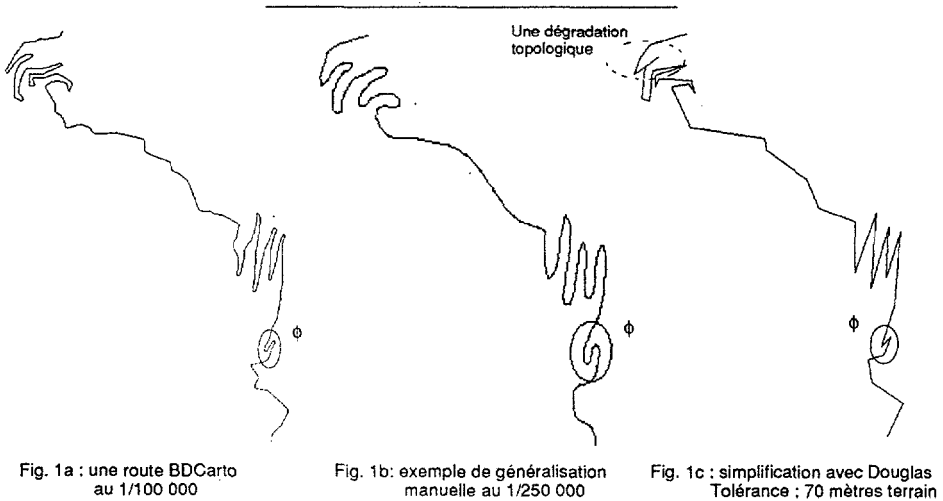


Figure 1 : Comparaison d'une généralisation manuelle et d'une généralisation automatique à l'aide de l'algorithme de simplification de Douglas

De telles considérations mettent clairement en évidence la nécessité de décrire la géométrie des objets préalablement à toute opération de généralisation. Partageant des objectifs similaires aux approches de B. Buttenfield [5] et de R. B. Mc Master [10], nous présentons dans cet article un modèle descriptif basé sur un découpage hiérarchique des lignes. Chaque section issue du découpage est considérée d'un point de vue de l'homogénéité et classifiée de manière à appliquer les algorithmes de généralisation les mieux appropriés et les valeurs de paramètres adéquates [13].

2. DESCRIPTION DE LA GÉOMÉTRIE DES OBJETS LINÉAIRES

On considère différentes formes d'une ligne à différents niveaux de perception, allant d'un niveau global (la ligne directrice) à un niveau local (un virage). Actuellement, deux critères nous permettent de décrire une ligne : la sinuosité et l'homogénéité.

- On définit de façon simple la sinuosité d'une ligne par ces nombreux changements de direction à différents niveaux de perception. Cependant il faut "affiner" cette définition en détaillant différentes facettes de la sinuosité comme par exemple les formes locales des virages (sinusoïdes, spirales...), la symétrie des virages, ou encore la sinuosité de la ligne directrice à un niveau plus global, etc ...
- On définit l'homogénéité d'une ligne par le fait qu'elle contienne ou non des portions semblables par rapport à la sinuosité (Voir figure 2). Une ligne sera donc découpée selon le critère d'homogénéité pour un niveau de perception donné. La définition de

l'homogénéité d'une ligne peut varier selon le niveau d'analyse : ainsi une section issue d'une segmentation à un niveau donné (donc homogène à ce niveau d'analyse) pourra subir une nouvelle segmentation à un niveau plus local.

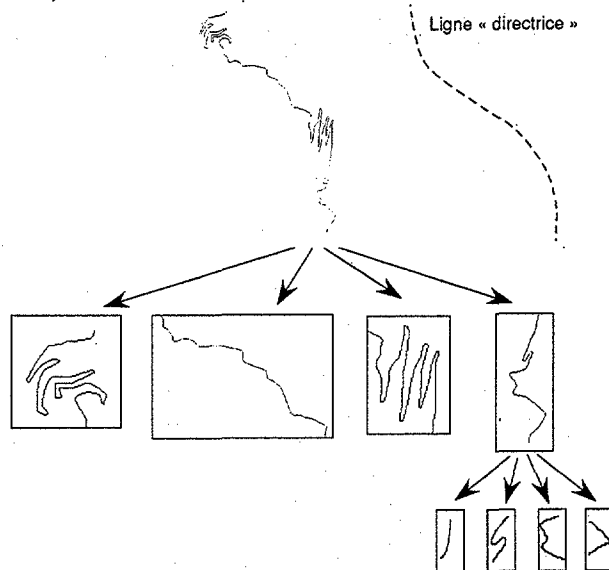


Figure 2 : Un exemple de segmentation (sur une route BDCarto®)

Bien que ces critères puissent s'avérer insuffisants pour décrire les lignes de manière objective et exhaustive, nous les considérons cependant comme primordiaux. Ils sont à la base des concepts sur lesquels est construit le processus de description proposé dans ce papier.

3. PROCESSUS DE DESCRIPTION

L'homogénéité constitue le premier critère de description. Le processus consiste ainsi à examiner de manière réursive la ligne (ou une section de ligne) et, à chaque niveau de perception : si la portion s'avère inhomogène, elle est segmentée de nouveau, alors que chaque section homogène est analysée au moyen d'une classification détaillée.

Le modèle ainsi construit par le biais d'une analyse hiérarchique a pour résultat la création d'un arbre de description (voir l'exemple de la figure 3). Chaque noeud de l'arbre correspond à une section de ligne et comporte des attributs descriptifs. La transition entre un noeud et un noeud fils peut être une segmentation ou une analyse. La racine de l'arbre correspond à la ligne originale. Plus un noeud est bas dans l'arbre, plus l'analyse correspondante est détaillée. L'homogénéité au premier niveau est qualifiée de manière grossière, alors que les feuilles correspondent à des sections plus finement homogènes qui comportent un attribut de classe de forme correspondant à un type de virage particulier (un

virage est ici défini comme une portion de courbe dont la courbure ne change pas de signe, et est délimitée par deux points d'inflexion, pour le niveau d'analyse considéré).

L'arbre descriptif ainsi construit peut alors être analysé de manière à déduire des informations complémentaires nécessaires aux décisions de généralisation, telles que, par exemple, les niveaux relatifs des formes dans les formes. Une analyse ascendante de l'arbre pourra également permettre d'améliorer la description globale de la ligne considérée.

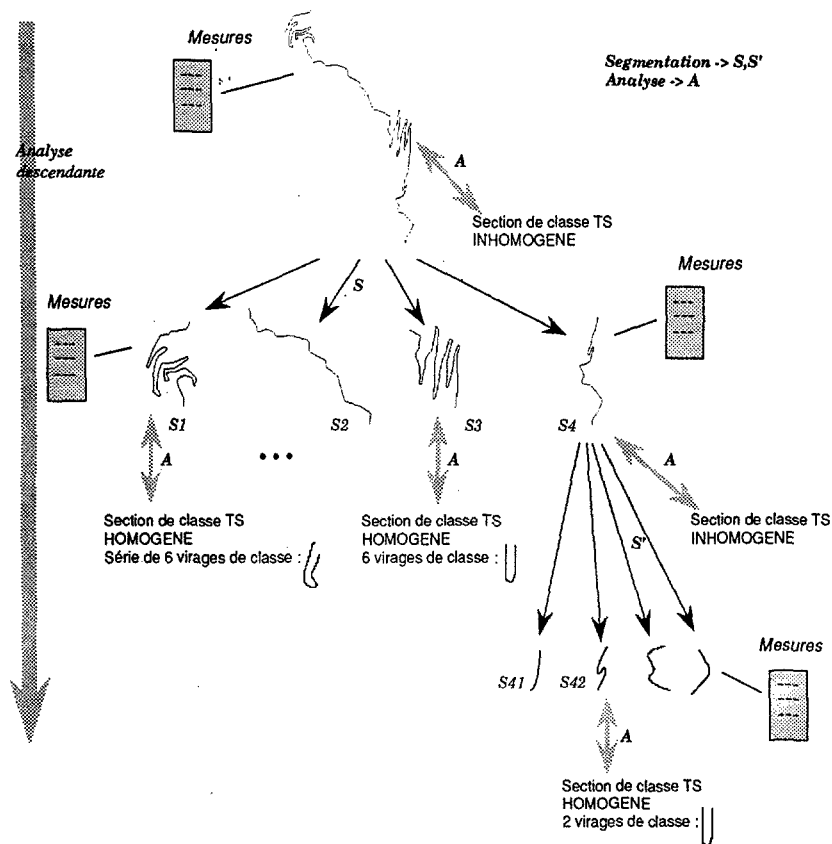


Figure 3 : Une description hiérarchique de ligne

Le processus pratique utilisé

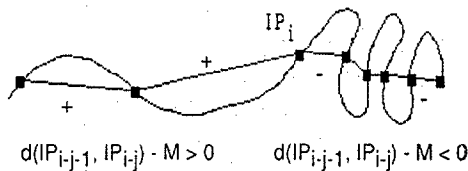
Bien que restant imprécis et pouvant être considérés selon différents points de vue, les critères de description cités ci-dessus sont complexes à modéliser. Cependant, nous avons tenté quelques expérimentations à partir d'une première définition exposée ci-dessous.

De manière générale dans la littérature aussi bien en psychologie qu'en intelligence artificielle, et plus récemment en cartographie, la détection des formes passe par la détection de points caractéristiques ([1], [6], [11], [15]). D. D. Hoffman & W. A. Richards ont suggéré que les courbes peuvent être découpées aux minimas de courbure. F. Attneave a démontré que l'information est concentrée aux points d'un contour où la direction change le plus rapidement : les angles ou les pics de courbure. Ces deux types de points caractéristiques (points d'inflexion et maxima de courbure) nous servent de points d'appui à la segmentation et aux mesures calculées sur chaque section segmentée.

Ainsi, à partir des points d'inflexion, on aboutit notamment à une première définition du critère d'homogénéité :

Soit $IP(\lambda)$ l'ensemble des points d'inflexion détectés pour un niveau d'analyse donné λ , et soit $M(IP(\lambda))$ la moyenne des distances $d(IP_i, IP_{i+1})$ entre les points d'inflexion consécutifs.

Soit $D(IP(\lambda))$ les séquences de déviation $D_i = d(IP_i, IP_{i+1}) - M(IP(\lambda))$, et S_i les signes de ces D_i . Les points d'inflexion IP_i tels que $S_{i-2} = S_{i-1} \neq S_i$ définissent des lieux de segmentation potentielle des lignes. Ces points sont retenus comme étant des points critiques. Une ligne dont les inter-distances entre points d'inflexion varient de façon significative est alors qualifiée d'inhomogène. (figure 4).



Soit IP_i l'ensemble des points d'inflexion,

Soit M , la moyenne des distances $d(IP_i, IP_{i+1})$

On retient IP_i comme *Point critique*

Figure 4 : Détermination des points critiques

A partir des points critiques, la géométrie de chaque section homogène est qualifiée à l'aide d'un nouvel ensemble de mesures. On aboutit ainsi à la classification de chaque section homogène pour un niveau d'analyse donné λ . Une telle classification peut prendre

en compte non seulement la sinuosité, mais également d'autres critères tels que la forme des virages par exemple. Les mesures sont choisies de manière à représenter une première facette de la sinuosité. Pour chaque virage défini comme une portion de courbe entre deux points d'inflexion consécutifs (figure 5), les mesures suivantes sont calculées :

- la distance euclidienne *base* entre les points d'inflexion,
- la hauteur *h* du triangle (S_i, IP_i, IP_{i+1}),
- la longueur curviligne *l* entre les points d'inflexion IP_i et IP_{i+1} ,
- le ratio *l/base* entre la longueur curviligne et la distance euclidienne de IP_i à IP_{i+1} ,
- le ratio *h/base* entre la hauteur et la distance euclidienne de IP_i à IP_{i+1} .

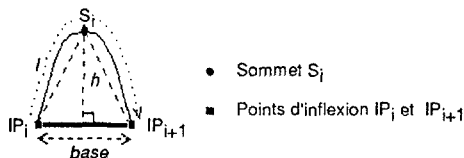


Figure 5 : Mesures effectuées sur un virage

Pour une section donnée, les valeurs moyennes ou médianes sont calculées pour toutes ces mesures, et ainsi utilisées dans un processus classique de classification.

Un arbre de description est ainsi construit du niveau global vers le détail, basé sur une segmentation récursive et une classification. Les points d'inflexion sont détectés sur une ligne préalablement lissée au moyen d'une convolution avec un filtre gaussien (de paramètre σ). La progression dans l'analyse est assurée par le réglage de la détermination des points d'inflexion : plus le niveau d'analyse est bas dans l'arbre, plus la valeur de σ est faible. Une série de plus en plus détaillée de points d'inflexion est ainsi à la base des opérations de segmentation et de classification de la sinuosité. Le choix de la valeur de σ est calculé automatiquement pour chaque ligne. Cette valeur est choisie empiriquement proportionnelle à $L \times \log(L)$, où L est la longueur curviligne des sections.

Notons le fait que les séries de mesures considérées pour évaluer les critères ne fournissent pas en l'état une vue exhaustive de ceux-ci. Il semble clair que l'homogénéité ne se limite pas aux inter-distances entre points d'inflexion. Une des premières améliorations envisagées consisterait à ajouter de nouvelles mesures (tel que la hauteur *h* des virages, figure. 5), et à décomposer le critère d'homogénéité selon plusieurs aspects. Ces critères peuvent alors être pris en compte successivement à plusieurs niveaux d'analyse. Une étape supplémentaire souhaitable consisterait à concevoir une formalisation globale intégrant plusieurs facettes de l'homogénéité. Cependant, à la lumière des expérimentations déjà réalisées, nous pensons qu'en prenant en compte successivement différentes mesures d'homogénéité, nous devrions raisonnablement obtenir des résultats probants.

De même, la classification selon la sinuosité est déterminée à partir d'un ensemble restreint de mesures. Le principal problème réside dans le fait que différents types de formes (d'un point de vue perceptuel) produisent des valeurs très proches (voir la figure 3, la section S1 de la ligne en haut à gauche comparée à la section S3). Il est possible de calculer un grand nombre de mesures à partir de la représentation proposée, ou également à partir des

mesures déjà proposées dans la littérature ([8], [5], [13]). Le problème consiste à choisir un sous-ensemble approprié pour chacune des tâches du processus à différents niveaux d'analyse, qui soient normalisées et non corrélées.

Notons cependant que l'ensemble limité de mesures utilisées pour cette première expérience a conduit à des résultats encourageants.

4. RÉSULTATS EXPÉRIMENTAUX

Une partie du processus de description a déjà été expérimentée sur un ensemble de routes de la BDCarto[®] IGN, à partir des mesures décrites ci-dessus, en utilisant un outil du commerce (fonctions de "cluster analysis" du logiciel S-Plus). Le premier objectif consiste à segmenter les lignes en portions rectilignes / sinueuses / fortement sinueuses.

Segmentation

Les premières expérimentations montrent que même avec un critère d'homogénéité unique, l'étude des variations permet de déterminer un premier niveau de sections homogènes. Dans l'exemple figure 6, les virages dans la première section sinueuse (en haut à gauche) sont très serrés et fermés sur eux-mêmes, alors dans la section sinueuse du milieu, ils sont plus réguliers et en sinusoides. Une analyse plus détaillée de chaque section est nécessaire pour distinguer ces caractéristiques géométriques différentes.



Fig. 6a : Points d'inflexion détectés



Fig. 6b : Points critiques retenus

Figure 6 : Un exemple de segmentation de lignes

Classification des sections homogènes

Pour une première expérimentation, les mesures choisies à partir de l'ensemble des mesures décrites sont : la médiane du ratio $l/base$ et la médiane du ratio $h/base$. 40 lignes ont été segmentées et classifiées en utilisant le package de "cluster analysis" de logiciel S-

PLUS®. Les résultats montrent quelques problèmes, mais fournissent cependant un premier niveau de classification significatif. (Voir figure 7).

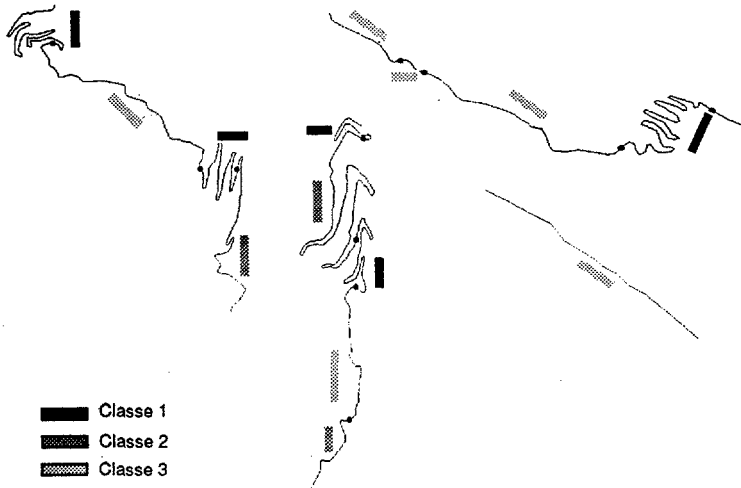


Figure 7: Un exemple de classification

5. INTÉRÊTS POUR LA GÉNÉRALISATION

A partir de l'arbre de description d'une ligne, une analyse remontante (qui n'a pas encore été intégrée dans les expérimentations déjà réalisées) devrait nous permettre de déduire des informations complémentaires de manière à guider les choix d'opérations de généralisation à appliquer localement. Cette connaissance peut être formalisée sous la forme de règles cartographiques simples telles que par exemple, pour un virage particulier :

- *règle 1* : si la distance entre les points d'inflexion est plus petite que la largeur du symbole, alors un conflit interne est détecté.
- *règle 2* : si le virage est une forme isolée dans une section très faiblement sinueuse, alors le virage considéré est un virage important qui doit être conservé.
- *règle 3* : si un virage est important et crée un conflit interne, il doit être amplifié.

Le processus de segmentation et d'analyse actuellement expérimenté n'est pas encore lié aux algorithmes de généralisation, mais plutôt à la perception visuelle. Pour progresser vers les objectifs fixés dans le cadre de la généralisation automatique, nous devons résoudre divers problèmes :

- définir quels sont les algorithmes de généralisation adaptés aux différents types de lignes. Ainsi une fois que l'on a déterminé la classe d'appartenance d'une ligne ou d'une portion de ligne, le choix de l'algorithme (ou de la séquence d'algorithmes) le mieux

adapté et des valeurs de tolérance sont implicites. Ce problème met l'accent sur un point critique lié à l'évaluation et à la compréhension des effets des algorithmes de généralisation sur les objets linéaires:

- concevoir une méthode pour segmenter et analyser les lignes, similaire à celle décrite dans ce papier, mais orientée vers une classification "opérationnelle" (i.e. liée aux algorithmes de généralisation). Néanmoins, le processus actuel de segmentation et de classification nous permet déjà d'adapter dans certains cas des choix cohérents d'algorithmes et de valeurs de tolérance. Les premiers résultats obtenus nous permettent d'ores et déjà de distinguer les classes de sections de lignes 'quasiment rectilignes', des portions 'mollement sinueuses' et des portions 'très sinueuses', de telle sorte que par exemple nous savons où appliquer correctement l'algorithme de Douglas, avec les valeurs de tolérance adaptées au type de sinuosité.

De plus, le problème de la validation des résultats de généralisation introduit en particulier le problème de l'évaluation de la qualité en termes de conservation de formes. Une utilisation potentielle de l'arbre de description pourrait consister à comparer les deux arbres correspondant à deux versions de lignes : originale et généralisée, facilitant ainsi la comparaison.

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Attneave. F. 1954 *Some informational aspects of visual perception*. Psychological Review. Vol. 61. No. 3
- [2] Beard K. 1991 *Theory of the cartographic line revisited*. Cartographica 28(4) p 32-58
- [3] Buttenfield B. 1986 *Digital definitions of scale-dependent line structure*. AutoCarto London 86 p 497-506
- [4] Buttenfield B. 1989 *Scale dependence and self similarity in cartographic lines* Cartographica Vol. 26, p 79-100.
- [5] Buttenfield B. 1991 *A rule for describing line feature geometry*. In *Map Generalization* Chapter 3 p. 150-171. Buttenfield and McMaster Ed. Longman Scientific & Technical. London 1991
- [6] Hoffman D. D. & Richards W. A. 1982 *Representing smooth plane curves for visual recognition* AAAI 82 pp. 5-8
- [7] Jenks, G. F. 1989 *Geographic logic in line generalisation* .Cartographica p. 27-42
- [8] McMaster, R. B. 1986 *A statistical analysis of mathematical measures for linear simplification* - The American Cartographer 13(2) p.103-117
- [9] McMaster, R. B. 1989 *The integration of simplification and smoothing algorithms in line generalization* - Cartographica 26 p.101-121
- [10] McMaster, R. B. 1993 *Knowledge Acquisition for Cartographic Generalization: Experimental Methods*. ESF GISDATA Work. Compiègne France Dec. 1993. To appear in "GIS and Generalization: Methodological and Practical issues" Taylor & Francis, London
- [11] Muller J. C., Wang Z. 1993 *Complex coast-line generalization*. Cartographic and Geographic Information Systems. Vol 28-2, p. 96-106.
- [12] Plazanel C., Affholder J.-G., Lagrange J.-P. & Ruas A. : Représentation et analyse de formes linéaires pour l'automatisation de la généralisation cartographique, EGIS'94, vol II. p.1112-1121
- [13] Plazanel C. *Measurements, characterisation and classification for automated linear features generalisation*. AUTOCARTO 12 Charlotte Vol. 4. pp.59-68
- [14] Ruas. A. & Lagrange J.-P. 1993 *Data knowledge modelling for generalisation*. Séminaire ESF-GISDATA. Compiègne
- [15] Thapa K. 1989 *Data compression and critical point detection using normalized symmetric scattered matrix*. AutoCarto 9 . p. 78-89
- [16] Weger G. 1993 *Cours de cartographie* IGN-France. Document interne.