
Leçons tirées d'une expérience de chorémisation automatique

Robert Laurini, Sylvie Servigne

LIRIS – UMR 5205, INSA de Lyon, Université de Lyon, 69621 – Villeurbanne
Robert.Laurini@insa-lyon.fr, Sylvie.Servigne@insa-lyon.fr

RÉSUMÉ. Cet article tire les leçons d'une série d'expérimentations faites pour obtenir automatiquement des chorèmes, à savoir des schématisations visuelles de territoires. Au-delà de la présentation rapide des résultats obtenus, il donne un catalogue de pistes de recherches pour la génération automatique de chorèmes. De plus ces derniers seront présentés comme de potentiels systèmes de représentations visuelles de connaissances spatiales, comme outils de résumés visuels de bases de données géographiques et comme systèmes d'accès à de telles bases. Les directions de recherches touchent aussi bien aux aspects cognitifs, sémantiques, sémiologiques, fouille de données spatiales, interopérabilité, etc. De plus, cet article plaide pour de nouvelles applications comme le géomarketing, l'histoire et les réseaux de capteurs.

ABSTRACT. This paper is issued from several experimentations concerning chorems, i.e. visual schematic representation of territories. Beyond a rapid presentation of results, it gives a list of research topics to be carried out. Those chorems will be presented as visual representation of geographic knowledge, visual summaries of geographic databases, and novel access methods to geographic databases. Those research topics range from cognitive, semantic, semiologic aspects to spatial data mining and interoperability. This paper advocates for novel applications such as geomarketing, history and sensor networks.

MOTS-CLÉS : Bases de données géographiques, chorèmes, résumés visuels, connaissances géographiques, sémiologie graphique.

KEYWORDS: Geographic databases, chorems, visual summaries, geographic knowledge, graphic semiology.

1. Introduction

Depuis la parution de l'article séminal de Roger Brunet (1986), ont été effectués de nombreux travaux qui se réclament de la chorématique, à savoir la façon de schématiser visuellement un territoire (Cheylan et al 1990, Brocard et al. 1995, Fontanabona 1994, etc.). Cependant, de nombreuses critiques se sont élevées contre l'approche « à la Brunet » ; il ne s'agira pas pour nous de rentrer dans ces critiques, mais de voir comment il serait possible de générer automatiquement des schématisations visuelles. Quelques premières expérimentations ont été effectuées, et l'objet de cet article sera de synthétiser les leçons tirées de ces travaux. Nous aborderons les chorèmes avec de nouvelles définitions, comme représentations visuelles de connaissances géographiques et d'outils offrant des résumés visuels des bases de données géographiques et l'accès chorématique à ces dernières.

Dans un premier temps, nous rappellerons les travaux de Brunet, puis nous préciserons notre méthode de travail en vue d'obtenir automatiquement des chorèmes. Enfin, nous dresserons un bilan, et nous donnerons quelques pistes de recherche.

2. Chorèmes, chorématique et chorémisation

Roger Brunet a inventé les chorèmes, à savoir des représentations schématisées des territoires. Ce mot vient du grec *Χώρα* qui veut dire territoire, lieu. De plus, la chorématique est une méthode de modélisation géographique basée sur les chorèmes. Pour Brunet, il s'agit d'un « outil de modélisation parmi d'autres ». Différentes critiques se sont faites jour, notamment reprochant les formes géométriques trop schématiques, et l'absence de localisation précise. Mais pour nous, le reproche essentiel repose sur la nature des renseignements utilisés pour produire les chorèmes. En effet, les auteurs donnent rarement l'exhaustivité de leurs sources, et les critiques s'appuient généralement sur d'autres sources que celles des auteurs.

Notre fréquentation des utilisateurs des technologies de l'information en matière de géographie et de cartographie nous a conduits à diverses réflexions :

– Dans de nombreuses applications, notamment pour certains niveaux de décision, des approches simplifiées mais rigoureuses s'imposent : il ne s'agit non plus de produire une cartographie détaillée, mais de montrer les points saillants ;

– De plus en plus, suite aux travaux de Ben Shneiderman (1997), l'approche dominante est celle de la « vision globale, zoom et filtrage, détails à la demande » ; en d'autres termes, il faut être capable de générer automatiquement des visions globales ;

– Les recherches actuelles sur les langages visuels montrent que les bonnes interfaces sont conviviales et intuitives (Chang 1990a et 1990b).

A partir de ces considérations, il nous a paru fondamental de produire des visualisations cartographiques simplifiées. Nous aurions pu créer un autre néologisme, mais nous avons décidé de ré-utiliser le terme de chorèmes qui convient à notre vision des choses, sans pour autant être un disciple incondtionnel de Brunet.

Un autre objectif est celui de la représentation informatique des connaissances spatiales et géographiques. En intelligence artificielle, le modèle dominant pour la représentation des connaissances se fonde sur les logiques de description (Baader et al. 2003), qui s'appuient sur des formulations mathématiques. Dans cet esprit, les connaissances spatiales sont décrites de cette façon. Mais le problème se pose depuis longtemps de découvrir un mécanisme visuel pour décrire ces connaissances ; dès lors les chorèmes pourraient être d'excellents candidats sous réserve que leur pouvoir d'expression soit égal, voire supérieur aux logiques de description.

Une autre direction de recherches est celle des résumés de base de données. S'il est classique de fournir un résumé d'un document sous divers noms (abstract pour une publication, résumé exécutif d'un document pour un décideur, etc.), il devient urgent de générer des résumés de base de données. Dès lors, de nombreux travaux de recherches se sont fait jour afin de produire un exemple-type de contenu d'une base de données, ou bien de produire un texte résumant la structure et le contenu de la base de données.

Si résumer textuellement la structure est relativement facile, résumer textuellement le contenu est plus ardu, car il est nécessaire de trouver les éléments les plus importants. Hélas, on bute encore sur la définition de l'importance car parmi les préalables, il faut s'accorder sur la légitimité de l'importance : l'est-elle pour l'ensemble des humains, ou bien liée à un profil d'utilisateur ou de décideur ?

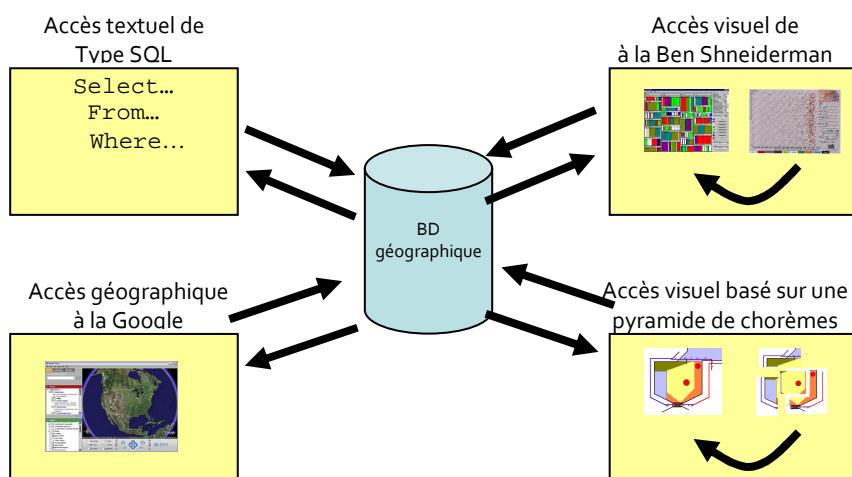


Figure 1. Différentes façons d'accéder à une base de données géographique, accès textuel de type SQL, visuel à la Ben Shneiderman (vision globale, filtrage, zoom, détail à la demande), accès géographique « à la Google », et accès visuel basé sur une pyramide de chorèmes

A cet égard, la fouille de données se borne à extraire les motifs les plus fréquents, dits aussi patterns (Shekhar et al., 2004) ; et il est courant d'extraire plusieurs milliers de motifs ; l'étape suivante dite « extraction des connaissances » se donne pour objectifs de filtrer ces motifs pour présenter ceux qui peuvent être utiles pour résoudre des problèmes. Pour aller plus loin, un des objectifs est donc non pas de résumer textuellement une base de données géographiques, mais d'offrir un résumé visuel pour donner une idée des problématiques à certaines échelles (Laurini et al. 2006, Del Fatto et al. 2007, De Chiara et al. 2009). Dès lors, on pourrait

imaginer, non pas un unique résumé au niveau le plus haut (par exemple pays), mais également des résumés intermédiaires (régions, départements) afin d'arriver graduellement aux informations de base, pour former une hiérarchie, ou mieux une pyramide de résumés. Dès lors, les chorèmes pourront de plus être utilisés comme système d'accès aux bases de données géographiques (Figure 1).

Un présupposé essentiel est celui de l'organisation de cette pyramide de chorèmes, laquelle nécessite une adaptation des formes simplifiées : si on suppose une simplification de la France et des simplifications géographiques des diverses régions de France, l'union de ces dernières donne-t-elle la simplification de la France ou quelque chose de différent ? Dès lors, l'articulation hiérarchique des chorèmes devra être abordée.

En définitive, pour nous, toute représentation cartographique visuelle simplifiée doit posséder les caractéristiques suivantes (i) être basée sur un ensemble de symboles (chorèmes) bien délimités, (ii) être une base pour la représentation des connaissances spatiales et géographiques de manière visuelle, (iii) être un outil pour les résumés visuels de bases de données géographiques, et (iv) permettre un accès graduel aux bases de données, du plus général au plus spécifique.

3. Méthode de travail

Devant cet état de fait, un programme de travail a été mis au point impliquant plusieurs laboratoires d'informatique ayant effectué des travaux en systèmes d'information géographique, à savoir : le LIRIS de l'INSA de Lyon, le laboratoire de Géomatique du Tec de Monterey, Campus de Puebla (Mexique), le DMI de l'Université de Salerno (Italie) et l'Université de Jendouba (Tunisie).

L'idée de base a été de partir des bases de données géographiques (BDG) existantes en vue d'extraire les chorèmes selon deux étapes (Del Fatto 2009, Lopez-Guillen, 2010, Coimbra 2008 et Cherni-Missaoui 2009). La première consiste à effectuer des fouilles de données spatiales (spatial data mining) de manière à extraire des patterns qui préfigureront des chorèmes et dans un deuxième temps de visualiser (Figure 2). Dès lors, l'architecture informatique se décompose en deux parties articulées autour d'un langage (dialecte de XML) nommé ChorML.

3.1. Phase de fouille de données et d'extraction des chorèmes

Rappelons que l'objectif de la fouille de données ou data mining consiste, à l'instar des mineurs qui cherchent à extraire des pépites, à lancer des procédures qui extrairont des connaissances. Dans notre expérimentation, il a été décidé dans un premier temps de considérer seulement les connaissances suivantes : (i) les faits correspondant à des requêtes sur la BDG, (ii) les regroupements de zones

géographiques, (iii) les flux, (iv) les règles de co-localisation et (v) les faits aux frontières.

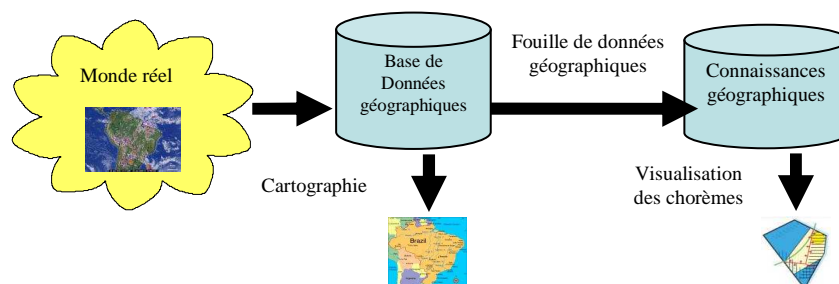


Figure 2. Les deux phases de la chorémisation automatique.

Par « fait aux frontières », nous entendons des renseignements qui ne se trouvent pas dans la base de données, mais qui permettent de mieux comprendre le contexte. En effet, prenons par exemple une base de données sur la région « Provence-Alpes-Côte d'Azur », celle-ci intégrera de nombreuses informations à l'intérieur du périmètre de juridiction. Mais dans la base de données, rien n'indiquera que la frontière sud correspond à la Méditerranée, l'est avec l'Italie, etc. Dès lors un fait aux frontières sera un nom suivi de la ligne du périmètre de la zone auquel il correspond. Plus précisément, il pourrait s'avérer important, en Italie de distinguer la frontière avec le Piémont et la Ligurie.

3.2. Phase de visualisation

De manière à assurer la lisibilité du schéma, plusieurs considérations devront être prises en compte à savoir, la taille de la carte chorématique, par exemple une page A4, l'existence ou non de couleurs (croquis en noir et blanc), le nombre maximal de chorèmes à visualiser, par exemple moins d'une dizaine.

Une fois ces paramètres connus, la phase de visualisation se décomposera en deux parties, la généralisation cartographique afin de simplifier la forme géographique du territoire, et la visualisation des chorèmes (Buttenfield, 1991).

Pour la généralisation cartographique, nous reprenons les algorithmes bien connus du domaine qui se présentent souvent comme dérivés de la méthode de Douglas-Peucker (1973). Cependant ici, le niveau de généralisation est crucial. S'il est courant d'entendre présenter la France comme un hexagone, beaucoup d'utilisateurs pensent que cette simplification est trop forte. Prenons l'exemple de l'Italie souvent comparée à une botte. Si l'on simplifie l'Italie de manière trop forte (par exemple six côtés), la botte ne sera plus reconnaissable, d'autant plus que la Sardaigne et la Sicile seront omises. Par ailleurs, il sera acceptable de généraliser la

Sicile comme un triangle, vu que les anciens Grecs l'appelaient déjà Trinacria (= triangle), et la Sardaigne comme un rectangle.

Un autre exemple connu est la simplification des Etats-Unis en rectangle : si cette simplification est acceptable à l'est, à l'ouest et au nord, la reconnaissance du sud, du côté de la Floride posera problème.

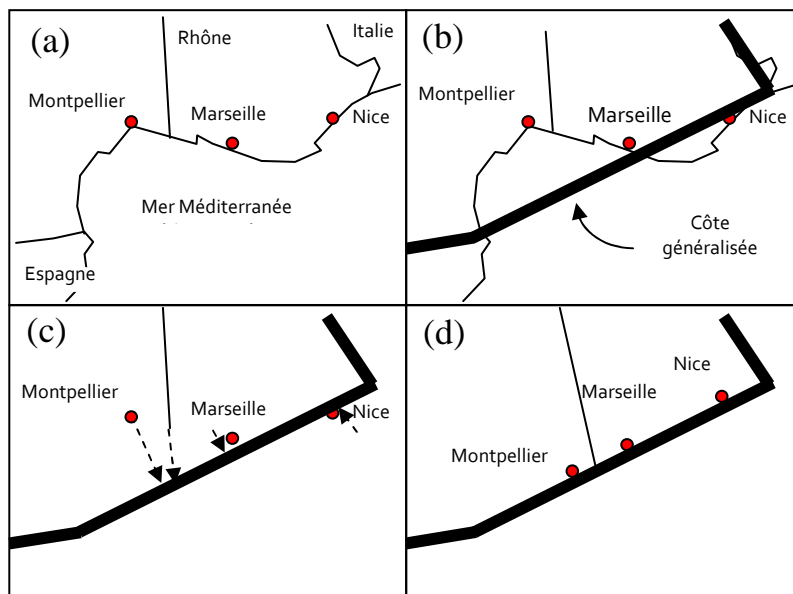


Figure 3. Déplacement des ports suite à la généralisation cartographique. (a) tracé original ; (b) côte généralisée ; (c) nécessité de déplacement des ports et de prolongement du Rhône ; (d) tracé final

De plus amples études cognitives devront être menées pour définir un niveau acceptable de généralisation permettant une reconnaissance des lieux non-ambiguë.

La généralisation cartographique pose un problème redoutable pour les villes en bordure de territoire comme les ports. La Figure 3 donne un exemple de la côte méditerranéenne où l'on voit qu'après généralisation, les villes de Marseille et de Montpellier se retrouvent à l'intérieur des terres, et Nice au milieu de la mer. Pour résoudre ce problème, il faudra imposer des contraintes topologiques à de telles villes de manière à ce qu'elles restent positionnées correctement après généralisation. Un problème voisin est celui des villes frontalières extérieures comme Genève, qui devront rester extérieures après généralisation. Ainsi, de connaissances d'un nouveau type devront être introduites, à savoir les contraintes topologiques des villes portuaires ou frontalières et les estuaires des fleuves. Signalons toutefois, que ces contraintes pourraient ne pas être données de manière déclarative, mais pourraient

être détectées par des algorithmes de type « zone-tampon » lancés sur les périmètres intérieur et extérieur du territoire en cours d'étude.

La solution adoptée (Delfatto, 2009) est un système multi-agent qui se charge de la disposition optimale des chorèmes, du respect des contraintes topologiques à l'intérieur, puis et du bon placement des informations extérieures. Voir Zoroaster 1991, Ruas et al. 1999 et Jabeur-Moulin 2005.

3.3. Langage ChorML

Comme interface entre les deux sous-systèmes (détection des chorèmes, et visualisation), un langage ChorML a été conçu afin de stocker les chorèmes découverts et leur origine. Par origine, on entend la procédure de data mining qui a découvert le chorème en cours de description ; de plus ces informations pourront nous aider à la traçabilité des chorèmes.

Trois niveaux de langage ont été définis :

- le niveau 1 correspond au contenu initial de la base de données écrit selon la norme GML (<http://www.opengis.net/gml/>) auxquelles ont été ajoutées les informations extérieures ;
- le niveau 2 correspond aux connaissances extraites par fouille de données avec les indications de traçabilité ;
- le niveau 3 correspond à la visualisation des chorèmes selon la norme SVG (<http://www.w3.org/Graphics/SVG/>) car ici il s'agit d'un dessin sur ordinateur.

Niveau	Représentation d'une ville	Représentation d'un flux
Niveau 1 (norme GML)	Surface caractérisée par les coordonnées de son polygone	Valeur d'un attribut (par exemple) flux de population ou de marchandise) entre deux zones polygonales
Niveau 2	Centroïde défini par longitude/latitude avec son importance, par exemple population, s'il est reconnu dans un pattern	Vecteur orienté selon le sens du flux entre les coordonnées des centroïdes des zones, avec son type et son importance
Niveau 3 (norme SVG)	Point avec des coordonnées en pixels, représenté par un cercle avec son rayon et sa couleur	Flèche avec des coordonnées en pixels, couleur définie par le type de flux, et épaisseur par l'importance.

Tableau 1. Représentation des points et des flux à travers les niveaux de ChorML.

A titre d'exemple, une ville et un flux seront décrits à travers ces trois niveaux de langage (Tableau 1).

Cependant des explications doivent être fournies notamment sur le niveau 1 : comme le nombre de connaissances peut être très élevé, un filtrage doit être opéré pour préciser les connaissances à garder comme chorèmes. A l'heure actuelle ce filtrage n'est pas automatisé, mais reste manuel. En d'autres termes, après filtrage seul un petit groupe de chorèmes seront sélectionnés pour la visualisation.

Loin de nous l'idée de détailler toute la grammaire ChorML ce qui exigerait de nombreuses explications (Coimbra 2008).

3.4. Expérimentations effectuées

Les expérimentations se sont faites dans plusieurs directions, généralement à partir des bases de données géographiques en ORACLE car ce logiciel incorpore de plus des outils de fouille de données et une interface cartographique.

Nous sommes partis d'une connaissance des flux de population en Italie, province par province. L'étude a abouti à structurer l'Italie en plusieurs zones et à ne garder que les flux les plus importants. La Figure 4 permet d'avoir rapidement une vision des flux dominants.

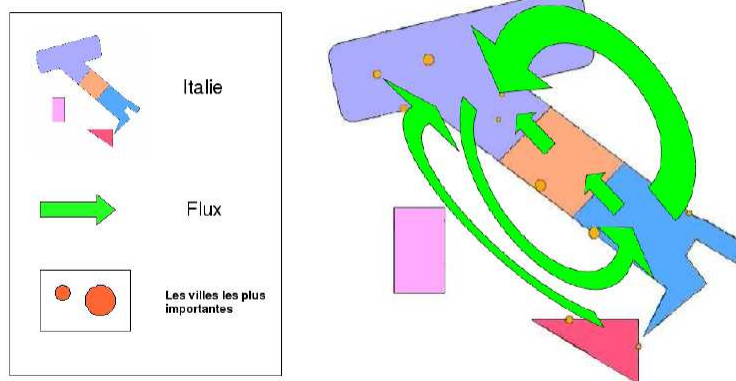


Figure 4. *Résumé visuel des migrations italiennes.*

L'autre jeu de données concerne les recensements à Puebla durant la colonisation espagnole entre les 16^{ème} et 18^{ème} siècles (Lopez-Guillen 2010, Loreto 2007). En Figure 5, à titre d'exemple, est présenté un résumé visuel qui donne la répartition spatiale entre les différents groupes de population. Il est à remarquer que la

superposition des groupes au centre correspond aux domestiques souvent recensés à la même adresse que leurs employeurs.

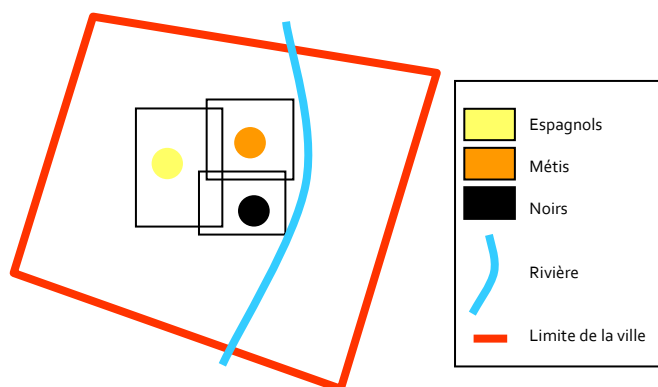


Figure 5. *Résumé visuel de la ségrégation spatiale à Puebla au 17ème siècle.*

3.5. Sur les aspects logiciels

Divers prototypes ont été montés à Lyon, Salerno et Puebla et leur intégration pose problème, notamment car les versions du langage ChorML sont différentes. Une fois ces alignements effectués, à l'heure actuelle, deux étapes-clés sont encore faites de manière manuelle, (i) le passage entre la sémantique des données (cachée dans les noms de données et de leurs métadonnées associées) et les algorithmes de fouille de données spatiales, et (ii) la réduction du nombre de motifs découverts par fouille de données de manière à choisir un nombre limité entre 5 ± 2 chorèmes pour conserver une lisibilité optimale.

4. Recommandations et directions de recherches

Afin de réaliser automatiquement les chorèmes à partir de bases de données géographiques, fort de ces expérimentations, il nous semble important d'organiser les recherches dans les directions suivantes. Cependant, comme expliqué précédemment, il faudra adjoindre aux bases de données des éléments géographiques extérieurs au territoire, lesquels devront être délimités avec précision. Certaines directions avaient déjà été mentionnées par Reimer et Dransch (2009).

4.1. Aspects cognitifs

S'il existe des polémiques sur l'acceptation des chorèmes, cela impose des recherches plus profondes sur la notion même de chorèmes et leur distinction vis-à-vis de la cartographie. Même si historiquement les chorèmes ont été inventés, utilisés et propagés par les géographes, il nous semble urgent de les réexaminer à la lumière des langages visuels.

De plus, il faudra effectuer de nombreuses recherches sur l'acceptation des chorèmes, leur compréhension par les utilisateurs ; dans ce but des tests psychocognitifs devront être mis au point, des mesures devront être élaborées de manière à établir des règles de bonnes pratiques dans ce domaine. Devant cet objectif, il nous semble incontournable de distinguer deux classes de problèmes :

- s'il s'agit d'analyser des territoires connus, l'utilisateur aura tendance à essayer de retrouver ce qu'il connaît déjà sur le territoire en question plutôt que de comprendre le message de l'auteur du chorème ;
- s'il s'agit d'un territoire inconnu ou devant lequel l'utilisateur n'a aucun a priori, la clarté du message et le niveau de compréhension devront être étudiés finement.

Un autre aspect est celui de l'importance. Peut-on définir une importance ex nihilo ? Ou bien celle-ci dérive-t-elle des profils des utilisateurs ? Dans tous les cas de figure, une fois les préliminaires cognitifs stabilisés, une formulation mathématique sera la bienvenue.

4.2. Aspects sémantiques géographiques

Toute base de données géographiques possède à la fois des aspects spatiaux (géométriques) et des attributs non-spatiaux (population, usage du sol, etc.). Toute schématisation d'un territoire mixe et mixera ces deux aspects.

Prenons trois bases de données géographiques de trois villes d'environ cent mille habitants, chacune utilisant une langue différente pour nommer les attributs et les valeurs d'attributs textuels. A-t-on besoin de disposer d'une ontologie géographique et urbaine multilingue pour assurer les correspondances entre les diverses bases ? Ou bien nous contentons-nous d'assurer ces correspondances « à la main » par le biais d'une interface spécialisée ?

Supposons de plus que l'on désire y étudier la ségrégation spatiale. Bien évidemment, si les attributs géographiques sont les mêmes (par exemple à l'îlot), les types de données sociologiques varieront grandement. En effet, dans certains pays les revenus sont publics, alors que pour d'autres leur publication est interdite, impliquant dès lors l'élaboration d'outils originaux pour estimer des indicateurs de richesse ou de pauvreté en partant uniquement des données publiques.

Dès lors, des recherches devront être lancées en linguistique pour établir ces correspondances, non pas de manière abstraite à partir des ontologies de domaines,

mais de manière concrète à partir des spécifications des attributs en liaison avec les métadonnées ; autrement dit à partir des ontologies locales.

Si l'on désire construire un système valable dans plusieurs langues, ces aspects sémantiques devront avoir trouvé une solution efficace.

4.3. Sur la sémiologie graphique

Si l'usage de la sémiologie graphique à la Bertin (Bertin 1967) et de la théorie des couleurs d'Itten et de leur contraste (Itten 1973), représentent à l'heure actuelle une des bases solides pour la conception des cartes, nous pensons qu'il faut aller plus loin de manière à élaborer et intégrer davantage de règles de bonnes pratiques non seulement pour la manière de valuer les « variables de Bertin », mais aussi pour la sélection des couleurs. Ces travaux devront être menés en synergie avec les recherches sur les aspects cognitifs.

Reprenons l'exemple des flux. Supposons que l'on travaille sur les flux de marchandises dans la Manche, certaines transitant par containers sur des bateaux, d'autres dans des camions montés dans les car-ferries, etc. Alors que les règles pour choisir les épaisseurs des flèches semblent simples, le choix des couleurs est plus problématique. Une solution est d'ajouter de petits schémas comme des dessins stylisés de bateaux ou de camions pour préciser la sémantique des flèches. Si l'on choisit une flèche verte, ce symbolisme pour certains Américains correspond aux flux d'argent (billet vert) alors qu'il pourrait évoquer des denrées écologiques pour d'autres personnes.

Par ailleurs se dégagent de nouvelles directions de recherche sous le vocable de « Visual Analytics » (Thomas-Cook, 2005) dans lesquelles les chercheurs essaient de trouver de nouvelles façons de « faire parler » visuellement les données. Là aussi des synergies et des fertilisations réciproques doivent se faire jour.

4.4. Aspects connaissances géographiques et connaissances visuelles

Faut-il en rester à la logique de description pour décrire les connaissances géographiques ou bien doit-on essayer de chercher d'autres représentations ? Tout nous laisse penser que les humains (voire sans doute aussi les animaux) n'ont pas attendu l'avènement de cette logique pour mémoriser et réfléchir sur les territoires, ni même celui de l'écriture. Vraisemblablement dans les cerveaux d'autres méthodes de représentations doivent exister. Et la solution visuelle nous semble une piste à étudier en priorité, et ceci pour plusieurs raisons :

- les langages de signes visuels ont existé avant l'arrivée de l'écriture, et les spécialistes nous expliquent que les lettres de l'alphabet proviennent de dessins ;
- selon l'adage bien connu « un schéma valant un long discours », une synthétisation visuelle semble une manière élégante et incontournable ;

– actuellement, les interfaces graphiques des ordinateurs et de leurs dérivés (téléphones portables, tablettes, etc.) se veulent conviviales et intuitives, c'est-à-dire sans passer par des légendes textuelles ; la diffusion et la familiarité de ces outils pousseront les utilisateurs à accepter les outils visuels ;

– les recherches actuelles d'une part en cartographie et d'autre part sur les langages visuels devraient converger vers des outils de représentation visuelle des connaissances géographiques.

Depuis longtemps, la piste des chorèmes nous a semblé un candidat sérieux pour la représentation visuelle des connaissances géographiques, vraisemblablement sous des formes et des formalismes qu'il faudrait revoir.

4.5. Aspects résumés visuels de BDG

Les bases de données géographiques ont été inventées pour gérer et comprendre les territoires dans lesquels nous vivons. Sous la notion d'échelle, pendant des décennies voire des siècles, de manière insidieuse se sont cachées des préoccupations autres que les rapports entre distance réelle et distance sur la carte, à savoir des éléments sémantiques : il ne sert à rien de cartographier tous les bâtiments sur une carte de France dans un format A4. Dès lors deux types de généralisations doivent être étudiés :

– en premier, la généralisation cartographique dont les algorithmes sont bien connus, mais dont les seuils devront être stabilisés cognitivement : est-il acceptable de représenter les Etats-Unis sous forme de rectangle ? nous pensons que des généralisations géométriques trop brutales ont fait été nocives du point de vue cognitif ;

– en second lieu, la généralisation sémantique doit être définie de manière plus formelle de manière à combiner efficacement généralisation géométrique et généralisation sémantique.

Une fois cet objectif atteint, la génération automatique des résumés visuels des bases de données géographiques pourra effectivement être résolue.

4.6. Aspects accès chorématique aux bases de données

Comme illustré Figure 1, il existe divers styles pour accéder à une base de données géographiques ; celle des chorèmes, permettant une approche graduelle, allant du plus général au particulier nous semble une piste attrayante.

Une fois une pyramide de chorèmes mise au point, il faudra de nouveau mettre en œuvre des tests psycho-cognitifs au point pour mesurer l'acceptabilité et l'efficacité de cette approche.

4.7. Aspects fouille de données géographiques

Dès lors, tout repose sur des démarches efficaces et efficientes de fouille de données spatiales. Même s'il existe des méthodes performantes pour extraire les motifs fréquents, il faudra mettre au point des méthodes originales pour extraire les motifs importants, en conformité avec une définition de l'importance.

Une démarche hiérarchique partant des tables mêmes des bases de données pour en extraire l'important, puis de combiner ces résultats, serait sans doute une approche prometteuse.

Cependant la réduction du nombre de motifs reste un problème difficile pour conserver la lisibilité de la carte chorématique ou du résumé visuel. Dès lors des recherches devront être effectuées :

- pour extraire des motifs qui aient du sens pour le décideur,
- et pour restreindre le nombre afin d'obtenir une carte chorématique lisible.

5. Vers de nouveaux domaines d'application

Jusqu'ici les chorèmes avaient été faits sur des problèmes de géographie classique (physique, démographique, sociologique, économique, etc.) et nous pensons que leur intérêt peut être étendu à d'autres domaines pour lesquels nous n'avons aucun a priori sur les structures sous-jacentes des territoires. Nous citerons :

- en histoire : l'analyse des bases de données géographiques et historiques afin de schématiser les territoires du passé ;
- en géomarketing : l'analyse spatiale, par exemple l'étude des ventes à partir des tickets de caisse lesquels seront d'un intérêt prodigieux pour les décideurs commerciaux ;
- sur les réseaux de capteurs temps réel (météorologie, pollution, etc.) : ici il s'agira non pas de déterminer des chorèmes une fois pour toute, mais surtout de les calculer en temps réel de manière à schématiser l'évolution, à mieux comprendre les structurations dans l'espace et le temps ;
- en micro-géomatique ou encore la géomatique d'affaires où les objets sont géo-localisés par cellules RFID, par exemple dans les magasins, les entrepôts, les centres commerciaux ; ici l'analyse par les chorèmes permettrait de découvrir les caractéristiques saillantes cachées de ce type d'endroit (Chamberland-Tremblay et al., 2009).

Ce texte est le fruit de nos premières expérimentations afin de générer automatiquement des chorèmes à partir des bases de données géographiques. Partis de la définition de Brunet, nos travaux nous ont menés vers d'autres directions comme la représentation visuelle des connaissances géographiques, les résumés visuels et l'accès chorématique aux bases de données et nous ont amenés à diverses questions scientifiques présentées également dans cet article.

Bibliographie

- Baader F., Calvanese D., McGuinness D. L., Nardi D., Patel-Schneider P. F. (2003) "*The Description Logic Handbook: Theory, Implementation, Applications*". Cambridge University Press, Cambridge, UK, 2003. ISBN 0-521-78176-0
- Bertin J. (1967) "*Sémiologie graphique*", Paris, Mouton/Gauthier-Villars, 1967.
- Brocard M., Lecoquierre B., Mallet P. (1995) "Le chorotype de l'estuaire européen". *Mappemonde* 3-1995. pp. 6-7.
- Brunet R. (1986) "La carte-modèle et les chorèmes", *Mappemonde* 86/4 pp. 4-6.
- Buttenfield B., McMaster R. (1991) "*Map Generalization: Making Rules for Knowledge Representation*". Longman, London. 1991.
- Chamberland-Tremblay D., Giroux S., Caron C. (2009). "Coopérer par l'espace : la microgéomatique dans l'espace intelligent", *Revue Internationale de Géomatique*, Vol.19, No 2, pp.231-244.
- Chang S. K. (1990a) "*Principles of Visual Programming Systems*" (Ed.), Prentice Hall, 1990 (ISBN 0-13-710765-X).
- Chang S. K. (1990b) "*Visual Languages and Visual Programming*" (Ed.), Plenum Pub. Co.
- Cherni Missaoui I., (2009) "*Résumés visuels de bases de données géographiques : Transformation de requêtes spatiales en ChorML*". Master en Informatique. Université de Jendouba, Tunisie. Septembre 2009.
- Cheylan J., Deffontaines J., Lardon S., They H. (1990) "Les chorèmes : un outil pour l'étude de l'activité agricole dans l'espace rural". *Mappemonde*, n° 4. 1990. pp. 2-4.
- Coimbra Rocha A. (2008) "*ChorML: XML Extension for Modeling Visual Summaries of Geographic Databases Based on Chorems*". Thèse en Master, INSA de Lyon, 2008.
- Del Fatto V. (2009) "*Visual Summaries of Geographic Databases by Chorems*". Thèse de Doctorat, 7 avril 2009. En Cotutelle INSA de Lyon et Università di Salerno (Italie).
- Del Fatto V., Laurini R., Lopez K., Loreto R., Milleret-Raffort F., Sebillio M., Sol-Martinez D., Vitiello G. (2007) "Potentialities of Chorems as Visual Summaries of Spatial Databases Contents", VISUAL 2007, 9th International Conference on Visual Information Systems, Shanghai, China, 28-29 June 2007. Springer Verlag LNCS-4781 "*Advances in Visual Information Systems*", pp. 537-548.
- De Chiara D., Del Fatto V., Laurini R., Sebillio M., Vitiello G. (2009) "Visual Analysis of Spatial Data through Maps of Chorems". In Proceedings of the 15th *International Conference on Distributed Multimedia Systems*, San Francisco, September 10-12, 2009, Edited by Chang SK, Celentano A; and Yoshitaka A. Published by Knowledge Systems Institute, ISBN 1-891706-25-X, pp. 295-300.
- Douglas D., Peuckert. (1973), "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature", *The Canadian Cartographer* 10(2), pp. 2–122 (1973)

- Fontanabona J. (1994) "Ebauche d'une grille concepts spatiaux utilisables lors de l'analyse et de la construction de cartes". *Mappemonde* 1-1994. pp.1-5.
- Itten, J. (1973). "*L'art de couleur : l'expérience subjective et le raisonnement objectif de couleur*". New York : Van Nostrand Reinhold. ISBN 0442240376.
- Jabeur N., Moulin M. (2005) "A multiagent-based approach for progressive web map generation". In Proceedings of the "On the Move to Meaningful Internet Systems "2005: OTM Workshops, LNCS-3762, (2005), pp. 99-108.
- Laurini, R., Milleret-Raffort, F., Lopez, K. (2006) "A Primer of Geographic Databases Based on Chorems". In: SebGIS Conference, Montpellier, Springer Verlag LNCS4278 (2006) pp. 1693-1702.
- Lopez-Guillen K. (2010) "*Résumés visuels des bases de données géographiques basés sur les chorèmes*", Thèse de Doctorat Informatique, 23 juillet 2010. INSA de Lyon.
- Loreto R. (2007) "*Una vista de ojos a una ciudad nueva hispana. Puebla de los Ángeles en el siglo XVIII*". Benemérita Universidad Autónoma de Puebla, CONACYT, Instituto Nacional de Astrofísica Óptica y Electrónica, 2007. ISBN: 978-968-7938-01-1.
- Reimer A., Dransch D. (2009) "Information Aggregation: Automatized Construction of Chorematic Diagrams". In Proceedings of the Workshop "GeoViz & The Digital City" (Hamburg March 3-5, 2009).
- Ruas A. Demazeau Y. Jackson M. Mackaness W. Weibel R. Lamy, S. (1999) "The application of agents in automated map generalisation". In Proceedings of 19th ICA meeting., 14-21 August, Ottawa (Canada).
- Shekhar S., Zhang P., Huang Y., Vatsavai R. (2004) "Trends in Spatial Data Mining". In: *Data Mining: Next Generation Challenges and Future Directions*. Association for the Advancement of Artificial Intelligence Press. 2004.
- Shneiderman B. (1997) "*Designing the User Interface*". Third edition. Addison-Wesley, Publishing Company. 1997. pp. 600.
- Thomas J.J, Cook K.-A. (eds) (2005) "*Illuminating the Path: The Research and Development Agenda for Visual Analytics*". Published by the National Visualization and analytics Center, US Dpt of Homeland security. ISBN: 0769523234. 186 p.
- Zoraster S. (1991) "Expert systems and the map label placement problem". *Cartographica*, 28, (1), pp. 1-9, 1991.

Robert Laurini est professeur à l'INSA de Lyon, ancien directeur du laboratoire LISI et du GDR SIGMA. Il est l'auteur de près de 200 publications sur les systèmes d'information géographique et multimédias. Ses ouvrages les plus importants sont « *Fundamentals for Spatial Information Systems* », avec D. Thompson, Academic Press, 1993 et « *Information Systems for Urban Planning* » Taylor and Francis, 2001.

Sylvie Servigne est maître de conférences au département informatique de l'INSA de Lyon et membre du LIRIS, Laboratoire d'InfoRmatique en Image et Système d'Information. Elle est l'auteur de près de 80 publications sur les systèmes de bases de données spatiales et géographiques.