

EVELINE BERNIER

**UTILISATION DE LA REPRÉSENTATION MULTIPLE COMME SUPPORT À  
LA GÉNÉRATION DE VUES DE BASES DE DONNÉES GÉOSPATIALES DANS  
UN CONTEXTE SOLAP**

Mémoire  
présenté  
à la Faculté des études supérieures  
de l'Université Laval  
pour l'obtention  
du grade de maître ès sciences (M.Sc.)

Département des sciences géomatiques  
FACULTÉ DE FORESTERIE ET GÉOMATIQUE  
UNIVERSITÉ LAVAL

Avril 2002

© Eveline Bernier, 2002

## **Avant-propos**

Réaliser des études de deuxième cycle n'est jamais chose facile. Généralement, plusieurs personnes, de par leurs savants conseils et leurs encouragements, facilitent grandement ce processus. Je souhaite donc remercier certaines personnes qui ont contribué à la réalisation de cette maîtrise.

Tout d'abord, j'aimerais remercier mon directeur de recherche, M. Yvan Bédard, pour toute la confiance dont il m'a témoignée ainsi que pour m'avoir si bien dirigée. Également, je remercie mon codirecteur, M. Bernard Moulin, pour ses judicieux conseils et pour son entière disponibilité.

J'adresse également mes remerciements à toute l'équipe SIRS pour avoir répondu à mes nombreuses questions si patiemment et toujours avec le sourire. Un merci particulier à Martin et Sonia pour leur précieuse aide concernant le volet programmation.

Je désire également souligner le support financier des organismes suivants : le réseau GEOIDE, l'Association canadienne des sciences géomatiques ainsi que le Conseil de recherches en sciences naturelles et en génie.

Enfin, j'aimerais remercier plus particulièrement Sébastien pour toute l'aide et le soutien qu'il m'a apportés tout au long de ce cheminement ainsi que mes parents, qui ont su m'encourager à chaque instant.

## Table des matières

<b>CHAPITRE 1 INTRODUCTION GÉNÉRALE .....</b>	<b>1</b>
1.1. Mise en contexte .....	1
1.2. Problématique .....	3
1.3. Objectifs .....	5
1.4. Méthodologie .....	6
1.5. Synopsis .....	7
<b>CHAPITRE 2 REVUE DES CONCEPTS.....</b>	<b>9</b>
2.1. Introduction.....	9
2.2. La généralisation.....	9
2.3. L'automatisation de la généralisation cartographique.....	13
2.4. La représentation multiple .....	15
2.4.1. Échelle, résolution et granularité.....	16
2.4.2. Représentation multiple : Définition et état de l'art.....	18
2.5. Les technologies OLAP .....	24
2.6. Les origines du VUEL .....	26
2.7. Conclusion .....	27
<b>CHAPITRE 3 LA REPRÉSENTATION MULTIPLE DANS</b>	
<b>UN CONTEXTE SOLAP .....</b>	<b>28</b>
3.1. Introduction.....	29
3.2. SOLAP : les concepts de base .....	29
3.2.1. Opérations de navigation.....	32
3.3. Les vues alternatives : un nouveau besoin des outils SOLAP.....	33
3.3.1. La notion de vue.....	33
3.3.2. Les vues dans un contexte SOLAP .....	36

3.4. La nature et l'exploitation des représentations multiples dans un contexte SOLAP .....	37
3.5. Les mises à jour dans un contexte SOLAP avec représentations multiples .....	42
3.5.1. <i>Seuil de répercussion des mises à jour dans un SOLAP</i> .....	44
3.6. Développement du concept de VUEL .....	50
3.7. Conclusion .....	56
<b>CHAPITRE 4 IMPLANTATION DU CONCEPT DE VUEL.....</b>	<b>57</b>
4.1. Introduction.....	57
4.2. L'architecture système .....	57
4.3. Les données utilisées.....	61
4.4. L'application SOLAP .....	65
4.4.1. <i>Fonctionnalités de navigation</i> .....	68
4.5. Conclusion .....	73
<b>CHAPITRE 5 CONCLUSION.....</b>	<b>74</b>
<b>BIBLIOGRAPHIE.....</b>	<b>77</b>
<b>ANNEXE A COMBINAISONS POSSIBLES ENTRE LA GÉNÉRALISATION AUTOMATIQUE ET LA REPRÉSENTATION MULTIPLE.....</b>	<b>84</b>
<b>ANNEXE B MÉTAMODÈLE ORIGINAL DU CONCEPT DE VUEL .....</b>	<b>89</b>

## Liste des figures

<b>Figure 1-1</b> <i>Différentes représentations géométriques d'un même bâtiment</i> .....	3
<b>Figure 1-2</b> <i>Méthodologie</i> .....	7
<b>Figure 2-1</b> <i>Trois types de généralisation selon [Weibel &amp; Dutton, 1999]</i> .....	10
<b>Figure 2-2</b> <i>La généralisation cartographique</i> .....	12
<b>Figure 2-3</b> <i>Structure de données topologiques avec représentations multiples proposée par [Bruegger &amp; Franck, 1989]</i> .....	19
<b>Figure 2-4</b> <i>Niveaux de prise de décision</i> .....	24
<b>Figure 2-5</b> <i>Métamodèle de la représentation multiple élaboré par Devillers (2000)</i> .....	26
<b>Figure 3-1</b> <i>Exemple d'un hypercube</i> .....	32
<b>Figure 3-2</b> <i>Exemple d'une vue conceptuelle</i> .....	34
<b>Figure 3-3</b> <i>Exemple d'une vue spatiale</i> .....	35
<b>Figure 3-4</b> <i>Multiplicités géométrique, graphique et sémantique adaptée de Martel (1999)</i> .....	38
<b>Figure 3-5</b> <i>Différentes représentations géométriques pour les membres d'un même niveau hiérarchique (multiplicité géométrique multiéchelle)</i> .....	40
<b>Figure 3-6</b> <i>Deux représentations géométriques d'une même route (multiplicité géométrique uniéchelle)</i> .....	41
<b>Figure 3-7</b> <i>Deux représentations graphiques d'une même route (multiplicité graphique uniéchelle)</i> .....	41
<b>Figure 3-8</b> <i>Différentes sémantiques en fonction du niveau de granularité géométrique (multiplicités géométrique et sémantique multiéchelles)</i> .....	42
<b>Figure 3-9</b> <i>Seuil de répercussion d'une mise à jour de nature géométrique où les données détaillées sont conservées uniquement dans le système source</i> .....	44
<b>Figure 3-10</b> <i>Seuil de répercussion d'une mise à jour de nature géométrique où les données détaillées sont conservées dans l'entrepôt de données</i> .....	45

<b>Figure 3-11</b> <i>Seuil de répercussion d'une mise à jour de nature géométrique (multiplicité géométrique multiéchelle et uniéchelle) .....</i>	46
<b>Figure 3-12</b> <i>Seuil de répercussion d'une mise à jour de nature sémantique (au niveau de la classe d'objets) .....</i>	47
<b>Figure 3-13</b> <i>Seuil de répercussion et granularité d'une mise à jour de nature sémantique (au niveau de l'attribut) .....</i>	48
<b>Figure 3-14</b> <i>Seuil de répercussion d'une mise à jour à la fois sémantique et géométrique .....</i>	49
<b>Figure 3-15</b> <i>Trois vues servant à définir sémantiquement ou à représenter géométriquement et graphiquement une même réalité .....</i>	51
<b>Figure 3-16</b> <i>Métamodèle du concept de VUEL .....</i>	55
<b>Figure 4-1</b> <i>Modèle en flocon du concept de VUEL .....</i>	60
<b>Figure 4-2</b> <i>Les trois niveaux de granularité géométrique des bâtiments .....</i>	63
<b>Figure 4-3</b> <i>Exemple de hiérarchies sémantiques créées (au niveau des classes d'objets) .....</i>	64
<b>Figure 4-4</b> <i>L'architecture système .....</i>	65
<b>Figure 4-5</b> <i>L'interface de l'application SOLAP développée .....</i>	66
<b>Figure 4-6</b> <i>Différentes représentations graphiques selon la classe d'objets .....</i>	67
<b>Figure 4-7</b> <i>Forage spatial sur un seul objet .....</i>	70
<b>Figure 4-8</b> <i>Affichage de la classe générale associée à l'objet sélectionné .....</i>	71
<b>Figure 4-9</b> <i>Affichage de la classe détaillée associée à l'objet sélectionné, suite à une opération de forage descriptif .....</i>	71

# Chapitre 1

## Introduction générale

### 1.1. Mise en contexte

Récemment, de nouveaux outils d'analyse ont fait leur apparition dans le domaine des systèmes d'information. Basés sur une technologie OLAP (« On-Line Analytical Processing » ou processus d'analyse en ligne), ces outils permettent l'exploration d'une base de données traditionnelle (sans référence spatiale) de façon simple et rapide. Étant donné leur efficacité, plusieurs efforts ont été entrepris dans le but d'adapter ces outils au domaine spatial [Rivest, 2000]. Ainsi, de nouveaux outils de type SOLAP (« Spatial OLAP ») sont maintenant en émergence au sein de la communauté SIG (Systèmes d'information géographique). Ces derniers permettent de naviguer aisément, intuitivement ainsi que rapidement parmi plusieurs jeux de données géoréférencées. Un utilisateur peut donc visualiser ses données selon différents thèmes et différentes représentations (cartes géographiques, diagrammes circulaires, tableaux, etc.), tout en conservant son flux de pensée.

Parallèlement, on dénote depuis quelques années un certain engouement pour la diffusion d'informations géospaciales sur le World Wide Web. Ayant réalisé l'importance de cette

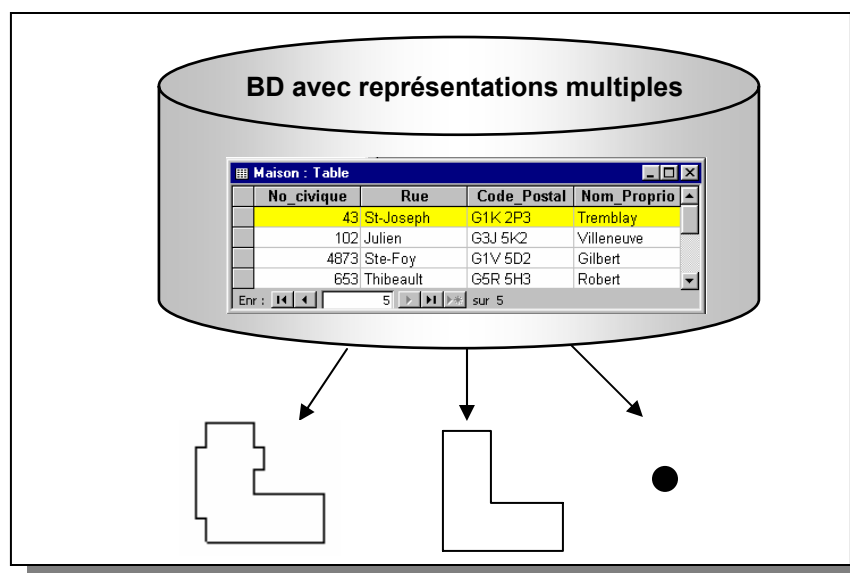
information lors du processus de décision, plusieurs exploitent maintenant la dimension spatiale de leurs données. D'un point de vue quantitatif, le Web est devenu le plus important média pour la diffusion de cartes géographiques [Kraak and Brown, 2001]. Selon Peterson (1999), environ 40 millions de cartes sur le Web sont utilisées par jour et ce, à travers le monde entier. Bien que ce média possède plusieurs avantages que ne peuvent offrir les cartes traditionnelles (ex. accessibilité, actualité et interactivité), il induit également certaines contraintes. Par exemple, les délais d'affichage de l'information géospatiale se doivent d'être à leur minimum, étant donné le caractère plutôt impatient des utilisateurs, « for users, speed is one of the biggest problems in using the Web » [Kehoe *et al.*, 1999]. Actuellement, la composante spatiale est exploitée via l'affichage de cartes généralement statiques, pré-enregistrées dans le système. Toutefois, étant donné que les technologies Web sont accessibles à une grande variété d'utilisateurs, chacun ayant des besoins spécifiques, il est primordial de disposer d'un système d'une plus grande flexibilité permettant l'affichage personnalisé d'informations géospatiales.

Dans l'optique de développer un système SOLAP sur le Web, il est nécessaire de pouvoir décrire une même réalité selon différentes représentations et/ou selon des degrés d'abstraction différents et ce, de façon quasi instantanée. L'utilisateur aura ainsi accès à des informations personnalisées en fonction de ses propres besoins. Pour ce faire, deux solutions peuvent être envisagées a priori: la généralisation automatique et la représentation multiple.

La **généralisation automatique** repose sur des algorithmes permettant de produire un jeu de données simplifié à partir d'un premier plus complexe. En fait, cette approche s'inspire de celle effectuée manuellement, processus qui consiste à « simplifier les éléments cartographiques et leur représentation, en fonction d'un besoin particulier et selon des règles précises » [OLF, 1999]. Par exemple, il est fréquent, suite à une réduction d'échelle cartographique, de devoir éliminer les détails superflus et de mettre en évidence certains éléments essentiels à la bonne compréhension du message véhiculé par la carte.



La seconde approche, **la représentation multiple**, consiste à emmagasiner, au sein d'une seule base de données, plusieurs représentations d'un même objet. Par exemple, un bâtiment pourrait être conservé sous différentes géométries, chacune présentant un niveau de détail particulier (figure 1-1.). Selon Vangenot (1998), la représentation multiple est la conséquence de la subjectivité de la perception et de la diversité des intérêts durant la phase de modélisation de la base de données.



**Figure 1-1** *Différentes représentations géométriques d'un même bâtiment*

## 1.2. Problématique

Le contexte SOLAP nécessite des temps de réponse rapides lors de la navigation dans le système. L'utilisateur peut ainsi explorer différentes avenues tout en conservant son schème de pensée. Par ailleurs, il fut démontré qu'un utilisateur moyen naviguant sur le Web tolère mal les temps de réponse supérieurs à 8 secondes [Marchand *et al.*, 2001]. Ces contraintes de performance sont à l'origine même de cette recherche. D'une part, la généralisation automatique, telle qu'on la connaît aujourd'hui, nécessite des temps de traitement élevés. Conséquemment, elle est inadéquate pour des applications qui

requièrent des réponses dans un très court délai. Également, du fait que la généralisation automatique s'inspire de celle effectuée manuellement, processus intellectuel difficile à formaliser, il est très difficile de l'automatiser entièrement. Elle requiert donc toujours une certaine intervention humaine lors des cas plus complexes. Finalement, bien que la généralisation automatique fasse l'objet de recherches depuis plus de 30 ans, son efficacité reste toujours à prouver pour certains auteurs :

*(...) satisfactory implementations of all the transformation operations (and their interactions) necessary to achieve comprehensive automated generalisation largely remain to be developed [Weibel & Dutton 1999].*

*(...) efficient methods for automatic generalization need to be developed further [Kilpeläinen, 2000]*

D'autre part, la représentation multiple semble toute indiquée lorsque des contraintes de performance doivent être respectées. Toutefois, la présence de plusieurs représentations pour un même objet au sein d'une seule base de données soulève plusieurs catégories de problèmes : l'augmentation inévitable du volume de données, les difficultés liées à la propagation des mises à jour, les risques d'incohérences, etc. [Rigaux 1994]. Également, les structures de données présentes actuellement dans la littérature sont inadéquates pour un contexte SOLAP. Majoritairement hiérarchiques, ces structures ne gèrent souvent que l'aspect géométrique des données sans égard à leur sémantique ou même à leur sémiologie.

À l'heure actuelle, la littérature abonde d'écrits concernant ces deux approches. Toutefois, elles sont souvent traitées de façon indépendante ; soit que l'on utilise le processus de généralisation automatique, soit que l'on exploite la représentation multiple. Rares sont les écrits proposant l'utilisation conjointe des deux approches au sein d'un même système. Également, les outils SOLAP étant encore pour la plupart au stade de développement, nous n'avons trouvé aucun ouvrage dédié à l'exploitation des

représentations multiples dans de tels systèmes. La problématique à surmonter est donc l'absence de méthode efficace de gestion des représentations multiples afin de rencontrer les exigences d'un SOLAP relativement à la génération des vues alternatives à la volée requises pour effectuer les opérations de navigation (forage, remontage, etc.).

### 1.3. Objectifs

L'objectif initial, alors défini au commencement de cette recherche consistait à formaliser la façon dont la représentation multiple pouvait être utilisée dans le processus même de généralisation automatique afin d'engendrer des temps de réponse acceptables pour un contexte SOLAP sur le Web. Toutefois, suite à de nombreuses réflexions et à l'émergence du concept de VUEL (Bédard *et al.*, 2000a), nous avons convenu que l'élaboration d'une telle solution risquait de dépasser le cadre d'un travail de maîtrise (quelques éléments de réflexion concernant la combinaison de la représentation multiple avec la généralisation automatique sont toutefois présentés à l'annexe A). De plus, le développement du concept de VUEL (*View Element*) offrait une nouvelle solution très prometteuse d'exploitation de la représentation multiple pour un SOLAP. Par conséquent, nous avons orienté cette recherche vers les aspects liés à l'exploitation directe des représentations multiples dans un contexte SOLAP, tant sur les plans géométrique que sémantique et sémiologique, et ceci grâce au concept de VUEL. Conséquemment, cela nous conduisait, dans le cadre de la présente maîtrise, à ne pas insérer directement dans notre solution la généralisation automatique à la volée (ce qui est réservé pour un futur étudiant). Notre **principal objectif** consiste donc à enrichir le concept initial de VUEL en relation avec le concept de la représentation multiple et à tester ce concept dans un contexte SOLAP.

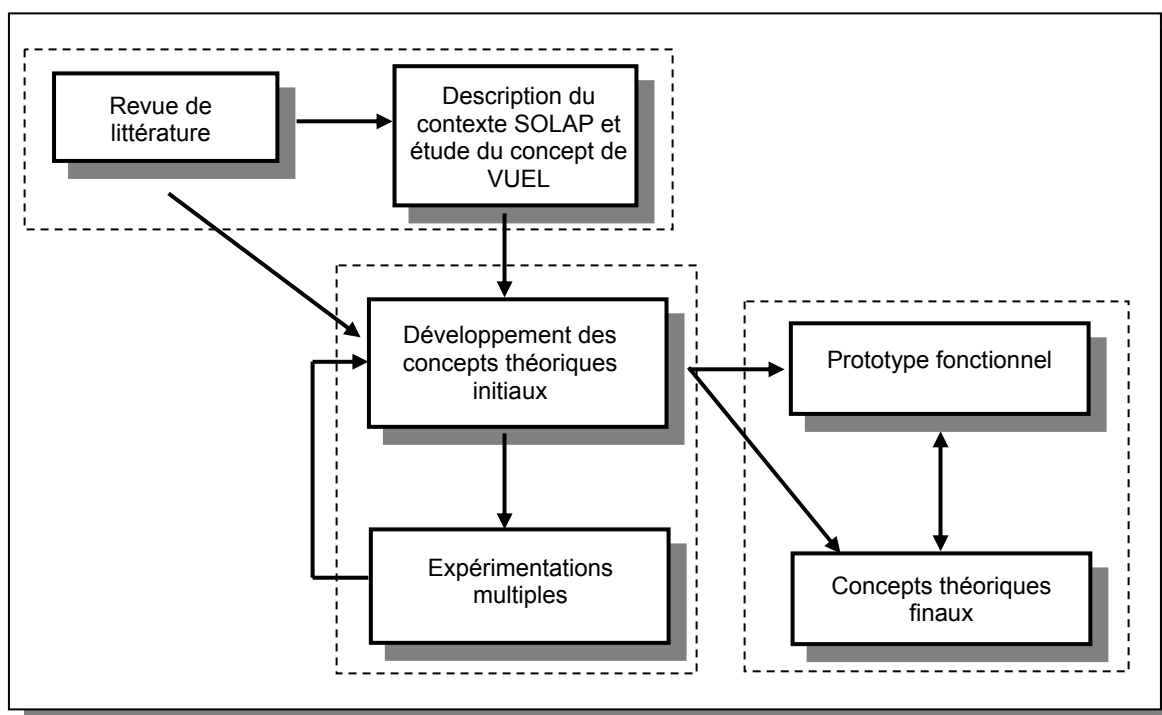
Les objectifs spécifiques de ce travail sont :

- enrichir les concepts théoriques existants en ce qui concerne la représentation multiple et la généralisation automatique;

- définir de façon précise le contexte opératoire lié à un environnement SOLAP ;
- mettre en application et adapter au VUEL les concepts théoriques définis dans Martel (1999) et testés par Devillers (2000) ;
- proposer un modèle conceptuel pour la gestion des représentations multiples dans un contexte SOLAP;
- développer un prototype SOLAP fonctionnel exploitant une base de données multidimensionnelle avec représentations multiples basée sur le VUEL.

#### **1.4. Méthodologie**

La méthode élaborée dans le cadre de cette recherche a consisté, dans un premier temps, à effectuer une étude théorique concernant les notions de représentations multiples et de généralisation automatique. Nous avons également procédé à la description du contexte opératoire lié à un environnement SOLAP ainsi qu'à l'étude de la version initiale du concept de VUEL. Par la suite, nous avons effectué le développement ou l'enrichissement des concepts théoriques initiaux, notamment ceux associés au concept de VUEL. Ces concepts ont par la suite été précisés et raffinés lors d'expérimentations effectuées à l'aide de données réelles. Celles-ci provenaient notamment de la base de données topographiques du Québec (BDTQ). Pour mener à bien ces expérimentations, nous avons utilisé les technologies Geomedia, Access, DynaGen (logiciel de généralisation de la compagnie Intergraph) ainsi que le langage de programmation Visual Basic. Finalement, ces diverses expérimentations ont permis le développement d'un prototype basé sur une structure de type VUEL. De plus, elles ont contribué à l'élaboration de concepts théoriques précis en ce qui concerne l'exploitation des représentations multiples dans un contexte SOLAP. La méthodologie suivie est illustrée à la figure suivante (figure 1-2).



**Figure 1-2 Méthodologie**

Enfin, cette recherche s'inscrit dans le cadre d'un projet actuel du réseau GEOIDE (projet DEC#9 : *Développement de techniques automatisées pour l'extraction et la généralisation de l'information géospatiale à partir de données de télédétection hyperspatiale (volet généralisation automatique)*) et est en relation avec un autre projet GEOIDE (projet DEC#2 : *Conceptualisation des fondations technologiques pour la prise de décision à l'aide du World Wide Web (volet SOLAP)*).

## 1.5. Synopsis

Ce premier chapitre a tout d'abord décrit le contexte dans lequel s'effectue cette recherche. La problématique entourant le processus de généralisation automatique ainsi que celle liée au concept de représentation multiple ont par la suite été présentées.

Finalement, les objectifs ainsi que la méthode de recherche suivie furent décrits. La suite du présent mémoire se divise en 4 chapitres.

Le chapitre suivant est destiné à une revue des concepts théoriques associés à la généralisation automatique, à la représentation multiple, au VUEL ainsi qu'aux technologies OLAP. Le chapitre 3 présente le contexte opératoire lié à un environnement SOLAP. La nature ainsi que l'exploitation des représentations multiples dans un tel contexte y sont également présentées. De plus, une nouvelle structure de données basée sur le concept de VUEL et dédiée à la gestion des représentations multiples est développée dans ce chapitre. L'implantation de cette structure est par la suite décrite au chapitre 4. Finalement, la conclusion de ces travaux ainsi que différentes perspectives de recherches sont présentées au chapitre 5.

## **Chapitre 2**

### **Revue des concepts**

#### **2.1. Introduction**

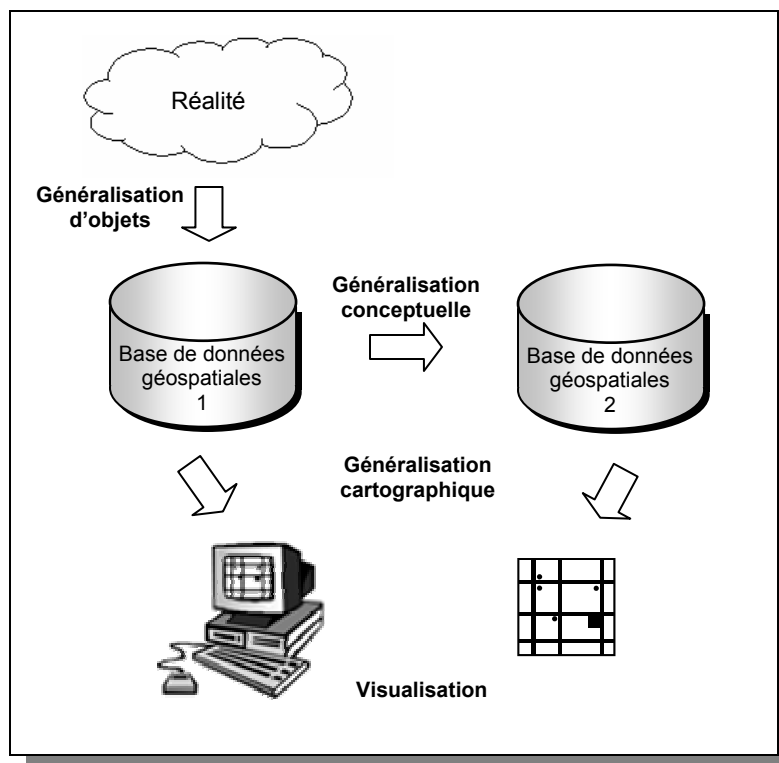
Le premier chapitre a introduit le contexte de la recherche, la problématique à surmonter ainsi que les objectifs fixés et la méthodologie suivie. Ce présent chapitre trace une synthèse théorique des concepts de généralisation automatique, de représentation multiple, de VUEL ainsi que des technologies OLAP. Par ailleurs, les concepts de généralisation et de représentation multiple sont présentés en détail dans Martel (1999) qui offre également une description des problématiques respectives, le lecteur est donc invité à consulter cet ouvrage pour plus de détails.

#### **2.2. La généralisation**

D'une complexité extrême, le monde se présente à nous comme étant composé d'une infinité de détails. Toute modélisation de ce monde nécessite donc de faire abstraction de certains de ces détails afin de ne retenir que les caractéristiques importantes ou générales. Selon Martel (1997), la généralisation, dans son sens large, est « l'opération qui consiste

à augmenter le degré d'abstraction de notre description et/ou représentation des phénomènes observés, donc à en diminuer les détails, afin d'en faciliter la compréhension et/ou la communication ». Ainsi, la généralisation permet de réduire la complexité d'un phénomène en éliminant les aspects secondaires et en accentuant les éléments essentiels, en fonction d'un but visé.

Dans le contexte des bases de données géospatiales, le processus de généralisation est employé à plusieurs niveaux (figure 2-1). Une première généralisation, appelée *généralisation d'objets* (ou « *object generalization* ») [Richardson & Müller, 1995; Weibel & Dutton, 1999 ] est essentielle lors de l'acquisition des données. En effet, la base de données ne représentant qu'une portion abstraite et discontinue de la réalité, il est nécessaire de sélectionner et par conséquent d'éliminer certains éléments des phénomènes observés.



**Figure 2-1** Trois types de généralisation selon [Weibel & Dutton, 1999]

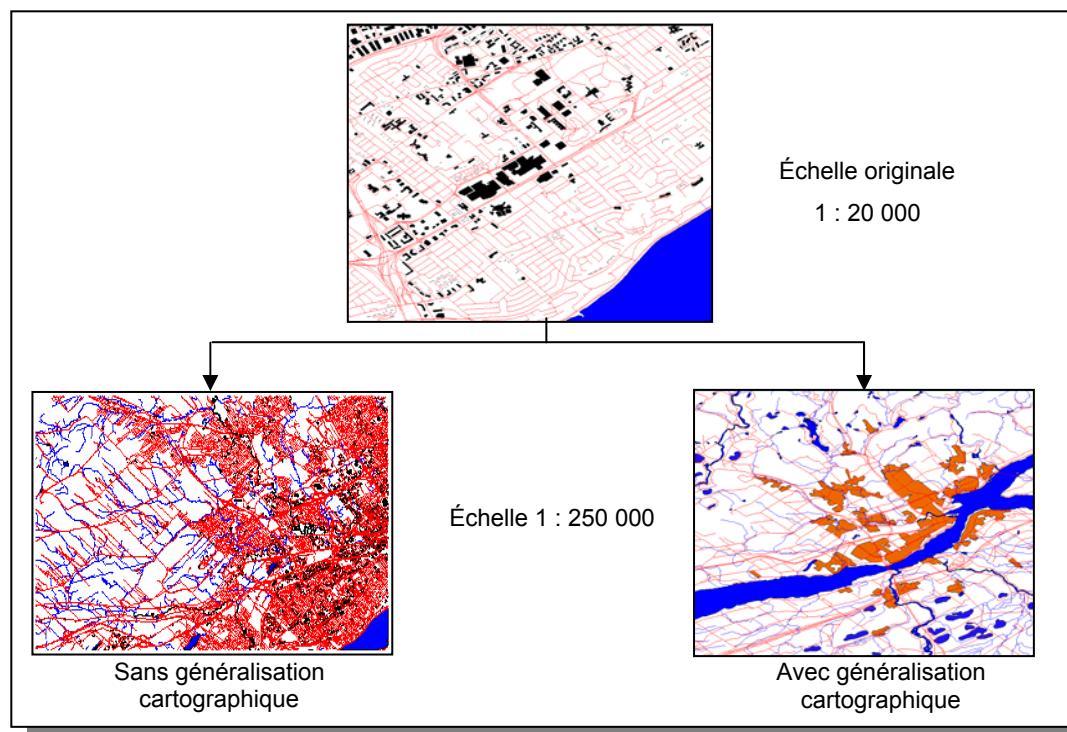


Propre au domaine numérique, le deuxième type de généralisation, la *généralisation conceptuelle* (ou « *model generalization* ») [João, 1995; Kilpeläinen, 2000; Weibel & Dutton, 1999] est employé afin de simplifier et de synthétiser le contenu d'une base de données [Rigaux, 1994]. Cet exercice permet notamment de réduire la complexité des données tant au niveau spatial que sémantique et temporel. Selon Weibel (1995), un des objectifs principaux de la généralisation conceptuelle est la dérivation de bases de données à des niveaux de résolution différents.

Enfin, contrairement au type de généralisation précédent selon lequel aucune considération visuelle, et donc aucune composante artistique ou intuitive, n'est présente lors du processus, la *généralisation cartographique* (ou « *cartographic generalization* ») est employée à des fins de visualisation. Selon Rigaux (1994), ce dernier type de généralisation a pour but principal « l'affichage des informations sur une surface de visualisation (carte ou écran). Au processus d'abstraction des données viennent donc s'ajouter des impératifs de communication visuelle qui peuvent mener à fausser volontairement, par des opérations de déplacement ou d'exagération, les valeurs sémantiques ou spatiales des objets représentés. Au cours d'une généralisation cartographique, la précision est donc secondaire par rapport à la clarté de l'affichage obtenu ».

La généralisation cartographique est habituellement employée lors d'une réduction d'échelle afin de contrôler la lisibilité des objets présents sur la surface de visualisation. Une réduction d'échelle implique inévitablement une diminution au niveau de l'espace alloué à la représentation des objets, ce qui a comme conséquence de créer de nombreux conflits. Par exemple, si l'on veut produire une carte à l'échelle 1 :250 000 à partir d'une carte à plus grande échelle (ex. 1:20 000), il est impératif d'éliminer de nombreux objets, d'en agréger d'autres, de modifier leur représentation géométrique et sans aucun doute, de déplacer certains éléments.

La figure qui suit (figure 2-2) illustre deux représentations issues d'une carte originale à plus grande échelle. La figure de gauche résulte d'un simple « zoom out » sans généralisation, ce qui se traduit par une représentation plus ou moins lisible. Au contraire, un certain degré de généralisation fut appliqué sur la représentation de droite, produisant ainsi une carte de meilleure qualité esthétique et présentant les éléments de façon claire et générale.



**Figure 2-2** *La généralisation cartographique*

En raison de ce qui précède, il semble évident que l'échelle est un facteur influençant la généralisation cartographique. Toutefois, il n'est pas le seul. Le but visé par la future carte constitue un facteur tout aussi important ; « a good map should focus on the information that is essential to its intended audience » [Weibel & Dutton, 1999]. D'ailleurs, le message que doit véhiculer la carte influence directement l'échelle à laquelle seront représentés les éléments. En effet, les phénomènes spatiaux devraient généralement être étudiés à l'échelle pour laquelle ils sont le plus pertinents [Dikau, 1990].

Le processus de généralisation s'effectue via l'enchaînement de différents opérateurs. Ces derniers ont fait l'objet de nombreuses recherches et la littérature abonde d'écrits relativement à leur description. Plusieurs auteurs proposent un ensemble plus ou moins semblable de ces opérateurs [Beard & Mackaness, 1991; Kubik & Brooke, 1992; Müller, 1990; Shea & McMaster, 1989]. Une synthèse est présentée dans Martel (1997) et regroupe les 11 opérateurs de généralisation que voici :

- |                           |                |
|---------------------------|----------------|
| 1. Raffinement            | 7. Exagération |
| 2. Reclassification       | 8. Déplacement |
| 3. Réduction de dimension | 9. Déformation |
| 4. Simplification         | 10. Agrégation |
| 5. Resymbolisation        | 11. Lissage    |
| 6. Caractérisation        |                |

### **2.3. L'automatisation de la généralisation cartographique**

Traditionnellement, la généralisation cartographique était réalisée manuellement par un cartographe expérimenté qui, partant d'une carte de base, en dérivait une autre à plus petite échelle. Aidé de son sens artistique et esthétique, celui-ci devait identifier les éléments essentiels afin de simplifier la représentation en fonction, notamment, de l'échelle et du but visé par la future carte.

Le besoin d'automatiser ce processus s'est fait ressentir il y a déjà quelques années. Les cartes ayant trouvé leur équivalent numérique et la croissante utilisation des systèmes d'information géographique (SIG) ont entraîné un besoin accru en ce qui concerne l'automatisation de la généralisation cartographique. Cette automatisation permet, d'une part, de réduire le volume de données à numériser et d'autre part, de limiter les moyens consacrés à la généralisation manuelle [Scholl *et al.*, 1996]. Par conséquent, plusieurs travaux de recherche ont, depuis plus d'une trentaine d'années, tenté de répondre à ce

besoin. Kilpeläinen (2000) identifie trois époques en ce qui concerne les recherches en généralisation automatique.

Les premiers balbutiements en généralisation automatique (1960-1975) se sont concentrés sur l'élaboration d'algorithmes, particulièrement en ce qui concerne la simplification de lignes [Ramer, 1972; Tobler, 1964]. Parmi ces algorithmes, notons celui développé par Douglas et Peucker (1973) qui est encore l'un des plus utilisés aujourd'hui. Par la suite (début des années 1980), les recherches ont porté sur l'analyse de l'efficacité des algorithmes existants. Les aspects plus conceptuels de la généralisation firent également l'objet de plusieurs travaux [McMaster & Shea, 1988]. Finalement, les récents développements dans le domaine de l'intelligence artificielle offrent de nouvelles perspectives en ce qui concerne la généralisation automatique ou assistée. Plusieurs auteurs ont travaillé à la conception de systèmes à base de règles ou de connaissances (rule-based or knowledge-based systems) [Buttenfield & McMaster, 1991; McMaster, 1991 ; Weibel, 1991]. Plus récemment, certains travaux en généralisation automatique exploitent les réseaux neuronaux [Allouche et Moulin, 2001].

Bien que la généralisation automatique ait fait l'objet de nombreuses recherches, il n'existe toujours pas de solution qui soit complète en soi [Kilpeläinen, 2000; Weibel & Dutton, 1999]. En effet, une certaine intervention humaine est toujours nécessaire afin de vérifier (et/ou corriger) certains résultats engendrés par les algorithmes ou pour guider le système dans les choix à faire lors du processus de généralisation. Également, les algorithmes ne sont, pour l'instant, efficaces que pour traiter certaines classes d'objets simples (ex. réseaux hydriques ou routiers) et uniquement lorsque ces thèmes sont pris séparément. Finalement, le processus de généralisation automatique nécessite d'importants traitements au niveau du processeur, ce qui engendre des temps de réponse majoritairement élevés [Martel, 1999; Vangenot, 1998]. Pour ces raisons, les méthodes de généralisation automatique actuelles ne peuvent répondre aux besoins des utilisateurs en temps réel.

Cette impossibilité actuelle de disposer d'une généralisation complètement automatique s'explique par de multiples raisons. Premièrement, la généralisation est perçue comme étant l'une des tâches les plus difficiles d'un cartographe [Kilpeläinen, 1992]. Située à mi-chemin entre l'art et la science, la généralisation cartographique automatique s'inspire de celle effectuée manuellement, processus intellectuel difficile à formaliser et donc à automatiser [Rigaux, 1994]. En fait, peu de règles régissent le processus de généralisation manuelle. Partant d'une vision globale de la carte à généraliser, le cartographe fait appel à ses capacités d'analyse et de synthèse afin de déterminer les bons opérateurs de généralisation à appliquer. La nature itérative de la généralisation est également considérée comme une entrave à son automatisation. Par exemple, il est fréquent que l'application d'un opérateur de généralisation visant à corriger un conflit engendre à son tour de nouveaux conflits, ce qui oblige à recommencer l'opération. Un tel processus itératif est extrêmement difficile à modéliser [Scholl et al., 1996]. Selon Plazanet (1995), « la généralisation peut-être vue comme un processus holistique de par le grand nombre de contraintes à prendre en compte (...). 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 ».

Compte tenu de la difficulté d'automatiser le processus de généralisation, il semble adéquat d'envisager une autre approche afin de fournir des représentations selon différents niveaux de détail. Cette approche consiste à les emmagasiner explicitement au sein d'une seule et même base de données spatiales; ce qui nous mène à la représentation multiple.

## **2.4. La représentation multiple**

Notre façon d'observer les phénomènes du monde réel est conditionnée par de nombreux facteurs. La culture, le passé ainsi que le domaine d'expertise en sont des exemples : ils influencent notre manière de percevoir le monde. Ainsi, chaque personne, étant pourvu

de référentiels distincts, perçoit différemment une même réalité. Par exemple, à la vue d'un bâtiment, un évaluateur foncier analysera principalement le type de revêtement, la surface habitable, alors qu'un architecte s'intéressera surtout à la structure même du bâtiment.

Dans le contexte des bases de données spatiales, un besoin indubitable est né de cette différence de perception ; celui de modéliser la réalité selon différents points de vue. Par conséquent, un même système pourra satisfaire les besoins spécifiques propres aux différents utilisateurs. Par exemple, il est fréquent que plusieurs applications exploitent une même base de données spatiales et ce, dans des contextes d'utilisation fort disparates [Vangenot, 1998]. Les besoins, notamment en termes de représentation géométrique et de résolution, peuvent donc varier d'un utilisateur à un autre. Selon Müller *et al.* (1995), chaque utilisateur devrait pouvoir analyser et voir les données au niveau de détail adapté à ses besoins. Actuellement, ceci est partiellement satisfait par le mécanisme de *vue* des bases de données qui sera discuté au chapitre suivant.

Ceci étant dit, plusieurs chercheurs proposent de conserver, au sein d'une même base de données, différentes représentations d'un même objet, ce que l'on désigne communément sous le terme de « représentation multiple ». Afin de mieux saisir cette notion, il est important de définir a priori d'autres concepts gravitant autour de cette idée.

#### **2.4.1. Échelle, résolution et granularité**

Nous associons fréquemment la notion d'**échelle** aux concepts de représentation multiple et de généralisation. On parle de généralisation de cartes à grande échelle, de représentations multiéchelles, etc. Selon l'Office de la langue française (2001), l'échelle est définie comme étant le « rapport moyen entre la distance mesurée sur une carte et la distance réelle mesurée sur la surface terrestre sans tenir compte des distorsions attribuables au facteur d'échelle et à l'altitude ». En d'autres mots, l'échelle est le rapport

entre la représentation cartographique d'un objet et la réalité terrain. De par son importance, elle est un moyen privilégié pour catégoriser les cartes. Toutefois, cette notion est beaucoup moins pertinente pour une base de données spatiales. Dans ce contexte, il est plus significatif de considérer l'échelle comme étant un niveau d'abstraction de la représentation des données.

Le concept de **résolution**, tel qu'employé dans le vocabulaire de la cartographie, désigne la taille minimale d'information représentable. Cette définition, tournée vers la visualisation de l'information, est subordonnée majoritairement par le type de support (ex. écran, papier). Or, dans un contexte de bases de données spatiales, le concept de résolution revêt une toute autre signification. Comme il fut mentionné précédemment, l'information contenue dans une base de données spatiales est représentée en fonction des choix de perception et d'abstraction de la réalité. Ainsi, les facteurs déterminant le contenu initial d'une base de données spatiales ne sont pas, comme pour la carte, la résolution du support ou l'échelle de représentation mais directement le type d'information que l'on souhaite représenter [Scholl *et al.*, 1996]. Conséquemment, la résolution d'une information contenue dans une base de données équivaut à sa résolution d'acquisition car la saisie des données est effectuée à un niveau de résolution spécifique, adapté au besoin applicatif visé [Vangenot, 1998]. Outre la dimension spatiale, la résolution peut également s'appliquer aux dimensions sémantique et temporelle d'une base de données. La première indique le niveau le plus précis selon lequel l'objet peut être décrit, alors que la seconde, la résolution temporelle, désigne le plus petit écart possible entre deux valeurs temporelles (ex. 1 heure).

Finalement, le terme **granularité** indique le « niveau de détail ou de généralisation des données emmagasinées dans une base de données » [Bédard *et al.*, 1997]. Des données détaillées correspondent à une granularité fine alors que des données agrégées équivalent à une granularité grossière. Une même information peut donc être emmagasinée dans une base de données à différents niveaux de granularité, selon les choix du modélisateur. Généralement, le niveau de granularité le plus fin correspond à la résolution même de

l'information. Toutefois, il se peut très bien qu'une base de données ne conserve que les données générales ou agrégées (ex. entrepôt de données où les données très détaillées ne sont conservées que dans la base de données transactionnelle). Dans ce cas, la granularité la plus fine ne sera donc pas équivalente à la résolution de l'information. Par exemple, la résolution d'une donnée temporelle (ex. date de naissance) peut être d'une journée (ex. 01/05/1997) mais n'être conservée qu'en mois dans la base de données (05/1997).

#### **2.4.2. Représentation multiple : Définition et état de l'art**

La représentation multiple est définie, selon certains auteurs, comme étant la coexistence simultanée de plusieurs représentations d'un même objet au sein d'une même base de données à référence spatiale [Kilpeläinen, 2000; Rigaux, 1994; Vangenot, 1998]. Cette définition nous semble très peu explicite et nous préférons employer celle exposée dans Martel (1999). Selon ce dernier, la représentation multiple désigne la « coexistence, dans une même base de données à référence spatiale ou dans un même fichier de carte numérique, de plusieurs représentations cartographiques (géométrie et sémiologie graphique) d'un même phénomène pour des fins d'utilisation différentes (à la même échelle ou à des échelles différentes) ».

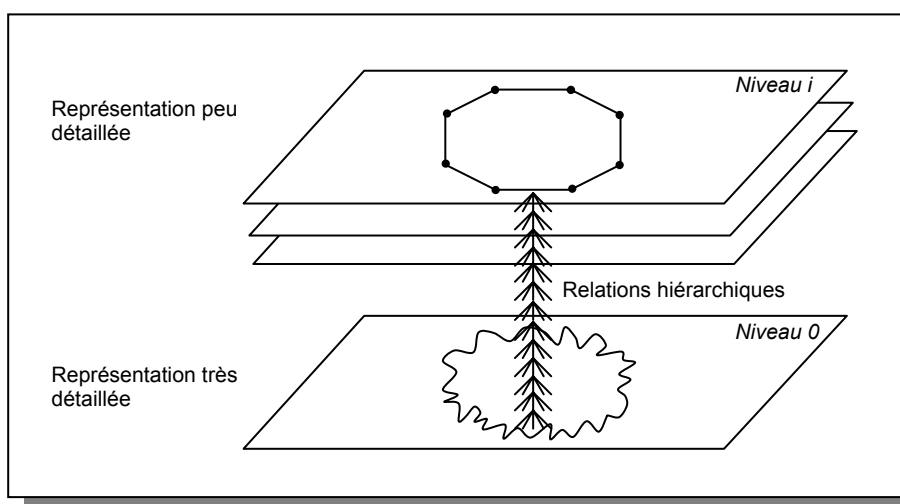
Conformément à ce qui précède, la notion de représentation multiple s'applique essentiellement aux caractéristiques visuelles des objets, conséquence intrinsèque du terme « représentation ». Si nous revenons à l'essence même du terme, l'action de *représenter* signifie « rendre à nouveau perceptible par l'un des cinq sens » [Martel, 1999]. Étant donné que seuls la géométrie et les variables visuelles (forme, taille, valeur, orientation, couleur, texture) peuvent être perçues par les sens, ces dernières sont donc les seuls éléments pouvant servir à la représentation. La sémantique est par conséquent exclue de la définition de la représentation multiple.

Les travaux en représentation multiple ont débuté sous l'initiative du NCGIA (National Center for Geographic Information and Analysis) en 1988 [NCGIA, 1989]. Depuis,



plusieurs auteurs se sont penchés sur la façon de gérer les représentations multiples. Plus précisément, la plupart des travaux concernent le « multiéchelle », une catégorie de représentation multiple où la géométrie des objets ne diffère qu'en fonction de l'échelle. Un objet possédera alors une représentation détaillée à grande échelle et une représentation simplifiée à plus petite échelle.

Certains auteurs proposent l'utilisation de **structures hiérarchiques** afin de gérer les changements de géométries dus aux changements d'échelle [Francalanci & Pernici, 1994; Jones, 1991; Kidner & Jones, 1994; Timpf & Frank (1995, 1997); Timpf, 1998]. Par exemple, Bruegger et Franck (1989) proposent une approche permettant de gérer les représentations multiples au sein de structures de données topologiques. Selon leur approche, le même objet est présent dans différentes couches, chacune ayant un niveau de résolution donné. La géométrie des objets dans les niveaux supérieurs est définie comme une agrégation de cellules des niveaux inférieurs. Les différentes couches sont donc interreliées via les relations hiérarchiques entre ces cellules (figure 2-3). L'avantage d'une telle structure réside dans le fait que les objets peuvent être présentés selon différents niveaux de détail.



**Figure 2-3** Structure de données topologiques avec représentations multiples proposée par [Bruegger & Franck, 1989]

Jones et Abraham (1986) proposent une structure hiérarchique similaire mais qui n'est pas spécialement basée sur une structure topologique.

D'autres se sont intéressés aux **structures arborescentes**, un type spécial de hiérarchie [van Oosterom, (1989, 1995); van Oosterom & Schenkalaars 1993; van Oosterom & Schenkalaars 1996]. L'idée sous-jacente à ce genre de structure est de combiner un index spatial et une structure hiérarchique distinguant les différentes représentations d'un objet selon l'échelle cartographique.

Vangenot (1998) présente **l'estampillage**, une approche permettant de définir la géométrie comme un attribut spatial à  $n$  valeurs et de gérer automatiquement l'accès à ces valeurs en fonction de la résolution. L'estampillage des associations permet d'associer à chaque instance d'une association la résolution pour laquelle elle est valable alors que l'estampillage des attributs permet d'associer plusieurs valeurs, ainsi que leur résolution respective, à un même attribut.

Devogele, Trevisan et Raynal (1996) proposent quant à eux de réutiliser les représentations monoéchelles existantes et de les relier entre elles à l'aide de **liens de correspondance**, un processus subdivisé en trois grandes étapes :

1. déclaration de toutes les correspondances et conflits entre les schémas des bases de données originales;
2. résolution des conflits et fusion des schémas;
3. appariement des données via des informations géométriques, topologiques et sémantiques, ce qui se traduit par la création des liens de correspondance.

Afin de valider leur processus, deux bases de données spatiales (BDCarto et GeoRoute) provenant de l'IGN (Institut Géographique National) ont servi à la création d'une base de données multiéchelle.

Kilpeläinen (1997, 2000) propose un modèle pour une base de données à représentations multiples (BDRM). Ce modèle est basé sur les concepts suivants :

1. la BDRM s'insère dans un environnement de généralisation conceptuelle;
2. les données emmagasinées dans la BDRM sont organisées en « niveaux »;
3. les données de chaque niveau sont définies en termes d'objets et de relations;
4. les différentes représentations d'un même objet sont reliées par des connectivités bidirectionnelles;
5. des processus de raisonnement contrôlent l'utilisation des opérateurs de généralisation conceptuelle.

La particularité de ce modèle est qu'il supporte les connectivités bidirectionnelles entre les différentes représentations d'un même objet. Ceci permet (1) la propagation automatique des mises à jour et (2) la dérivation de représentations simplifiées via des processus de raisonnement (*reasoning process*).

Par ailleurs, certains auteurs ont étudié les aspects plus problématiques liés à la représentation multiple. L'introduction d'une certaine forme de redondance (i.e. une même information est présente sous différentes formes et/ou valeurs) engendre les problèmes classiquement soulevés par la redondance dans les bases de données [Scholl *et al.*, 1996]:

1. l'augmentation inévitable du volume de données;
2. les risques d'incohérences;
3. la répercussion des mises à jour.

D'une part, l'augmentation du volume de données dépend principalement de la structure de données utilisée. Par exemple, l'approche permettant à un objet d'être présent dans plusieurs classes d'objets tend à multiplier le nombre de classes ainsi que le nombre d'objets, ce qui se traduit par une augmentation significative du volume de données.

D'autre part, la présence de plusieurs représentations pour un même objet augmente les risques d'incohérences. En fait, il devient extrêmement difficile de maintenir la cohérence au sein d'un système lorsqu'il y a présence de représentations multiples. Un des concepts clés dans la gestion des bases de données est qu'il est préférable de ne maintenir qu'une seule version d'un élément, ce qui assure automatiquement la cohérence [Buttenfield & Delotto, 1989]. Cette problématique fait donc l'objet de préoccupations majeures depuis plusieurs années; par exemple, Egenhofer, Clementini et Felice déclaraient déjà en 1994 : « we lack currently methods to assess consistency among different representations and to maintain consistently multiple representations of geographic objects ». Face à ce constat, certains chercheurs se sont penchés sur l'élaboration de méthodes permettant l'évaluation des incohérences présentes au sein des BDRM [Egenhofer, Clementini & DiFelice, 1994].

La répercussion des mises à jour représente également un aspect important des bases de données à représentations multiples. La propagation automatique des mises à jour est facilitée par l'existence de liens formels entre les différentes représentations d'un même objet. Kilpeläinen (1994, 1995) a dirigé bon nombre de ses travaux sur la façon d'automatiser cette propagation et propose le concept de « généralisation incrémentielle ». Ce concept est basé sur une approche modulaire semblable à celle retrouvée dans le domaine de l'ingénierie du logiciel. Selon son approche, chaque module représente un ensemble de données pouvant être généralisées indépendamment des autres objets de la base de données. De plus, le processus de généralisation doit également être subdivisé en différents modules. Suite à une mise à jour, seules les représentations affectées sont régénéralisées afin de correspondre à la nouvelle réalité. Ainsi, la base de données entière n'est généralisée qu'à une seule reprise.

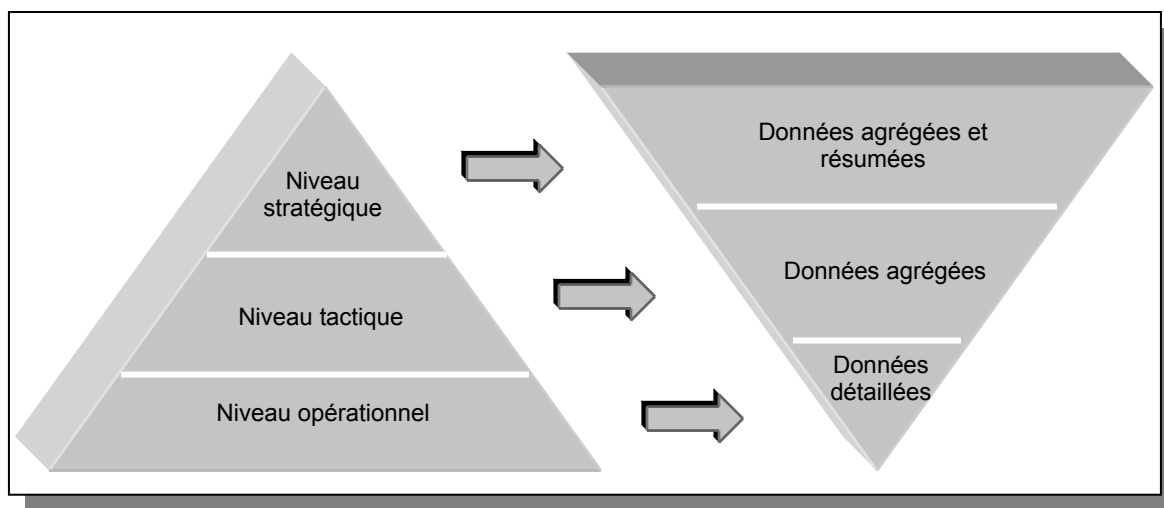
Badard (2000) présente un mécanisme permettant un haut niveau d'automatisation dans le processus de mise à jour d'un SIG avec représentations multiples. Ce mécanisme s'étend de l'extraction de l'information d'évolution dans une base de données de référence jusqu'à l'intégration et la propagation de cette information dans les systèmes

dérivés. Ce mécanisme s'appuie essentiellement sur des processus d'appariement géométrique et sémantique plutôt que sur des liens explicites entre les différentes versions d'un même objet. Le mécanisme développé fut testé dans le cadre de deux cas d'étude impliquant des bases de données de l'IGN et présentant des niveaux d'abstraction ou d'échelle différents.

La majorité des travaux en représentation multiple se sont concentrés sur l'élaboration de structures hiérarchiques afin de gérer les différentes représentations d'un même objet. Or, les relations du monde réel ne sont pas toujours hiérarchiques [Devogele 1997]. Également, la plupart des travaux existants réfèrent au multiéchelle. Par conséquent, un changement géométrique est invariablement associé à un changement au niveau de l'échelle. Toutefois, il est également possible qu'un même objet puisse posséder différentes géométries pour une même échelle (représentation multiple uniéchelle) [Martel, 1999]. Les normes de représentation qui diffèrent d'un organisme à un autre en sont un exemple concret. De plus, la majorité de ces travaux ne s'intéressent qu'à l'aspect géométrique des données sans vraiment considérer les aspects sémantique et graphique. Un environnement SOLAP exige de prendre en considération l'ensemble de ces éléments. En fait, les données y sont analysées autant d'un point de vue géométrique, que graphique et sémantique et la navigation s'effectue à travers les différents niveaux de détail (multiéchelle) aussi bien qu'à l'intérieur d'un même niveau (uniéchelle). Les systèmes doivent donc gérer les multiplicités de nature géométrique, sémantique et graphique, et ce, d'une manière multiéchelle ou uniéchelle. À l'heure actuelle, aucune des approches existantes afin de gérer la représentation multiple ne couvre l'ensemble des besoins liés à un environnement SOLAP. Il est donc nécessaire de définir de nouvelles structures afin de satisfaire les exigences d'un tel contexte.

## 2.5. Les technologies OLAP

À l'heure actuelle, les données spatiales sont gérées et exploitées par des outils de type transactionnel. Ces outils, axés sur la mise à jour, la sécurité et l'intégrité des données, restreignent de façon significative l'analyse pouvant être effectuée sur ces données. Leur structure, généralement relationnelle normalisée (i.e. la duplication des données est à son minimum), requiert un nombre élevé de joints entre les différentes tables. Ceci ralentit considérablement les temps de réponse, complexifie les requêtes et présente donc une entrave majeure à l'analyse rapide et intuitive des données spatiales. De plus, les données emmagasinées dans ce type de structures sont très détaillées et ne représentent que la situation actuelle, ce qui signifie qu'aucune donnée historique n'est conservée dans le système et qu'aucune donnée agrégée n'est rapidement accessible. Or, la prise de décision stratégique, souvent effectuée par des dirigeants peu familiers avec les technologies informatiques, requiert une grande facilité d'interrogation des données ainsi que des temps de réponse réduits à leur minimum. Aussi, ces types d'analyse nécessitent généralement des données agrégées et résumées incluant aussi bien des données historiques qu'actuelles afin de dresser le portrait général de l'organisation et de son évolution (figure 2-4). Face à cette situation, de nouveaux outils, notamment les outils OLAP (« On-Line Analytical Processing »), spécifiquement dédiés à l'analyse de l'information ont fait leur apparition sur le marché [Codd *et al*, 1993; Thomsen, 1997].



**Figure 2-4** Niveaux de prise de décision

Les outils OLAP se définissent comme « une nouvelle catégorie de logiciels axés sur l'exploration et l'analyse rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation » [Caron, 1998]. Leur structure multidimensionnelle permet de conserver les données selon les principales thématiques de l'entreprise et présente une performance d'interrogation accrue comparativement à l'approche transactionnelle [Bourgon *et al.*, 2000]. Les outils OLAP permettent à un utilisateur d'effectuer des analyses de façon intuitive et rapide afin qu'il puisse conserver son cheminement de pensée. Ces outils, souvent utilisés comme client d'un entrepôt de données ou d'un marché de données, procurent toute la performance et la puissance requises pour la prise de décision stratégique, la construction d'hypothèses d'analyse et la découverte de nouvelles connaissances.

Bien qu'il fut démontré que l'OLAP possédait un réel potentiel pour supporter l'analyse spatio-temporelle [Caron, 1998], les outils OLAP actuellement sur le marché n'exploitent pleinement que la dimension descriptive des données. En effet, même si les données peuvent comprendre une certaine composante spatiale (adresse, code postal, etc.), elles sont analysées uniquement de façon descriptive (tableaux ou graphiques statistiques) et lorsqu'elles le font de façon cartographique, le résultat est très limité. Une analyse purement nominale des données spatiales est donc une limitation des outils OLAP actuels à supporter le processus d'analyse spatio-temporelle. Afin de tirer profit de la dimension spatiale des données, un volet cartographique doit inévitablement être ajouté aux fonctions actuellement disponibles dans un outil OLAP. Des adaptations ont donc été envisagées afin de créer une toute nouvelle catégorie d'outils OLAP, les outils OLAP spatiaux ou SOLAP [Han *et al.*, 1998 ; Rivest *et al.*, 2001 ; Stefanovic, 1997 ; Stefanovic *et al.*, 2000]. De récents travaux effectués au Centre de recherche en géomatique de l'université Laval ont porté sur l'analyse de différents types d'architectures OLAP afin de déterminer la meilleure façon d'établir un lien dynamique avec les bases de données spatiales d'un SIG [Rivest, 2000]. Dans le cadre de cette recherche, différentes solutions technologiques potentielles ont été présentées et analysées et certaines d'entre elles (couplage OLAP-SIG) furent implantées et testées. Finalement, ces travaux ont permis, entre autres, d'établir les caractéristiques essentielles que devrait posséder un outil

SOLAP en plus d'identifier les caractéristiques souhaitées de l'outil SOLAP idéal. Parmi celles-ci figure la possibilité d'offrir des vues alternatives sur les données cartographiques via des mécanismes de généralisation automatique ou de représentation multiple.

## 2.6. Les origines du VUEL

Le VUEL, acronyme de *View Element*, est un concept relativement récent [Bédard *et al.*, 2000] qui est issu de nombreuses réflexions suite aux travaux de recherche effectués par Bédard avec Martel (1999) et Devillers (2000). Ces travaux avaient comme principal objectif d'explorer les possibilités d'implantation de la représentation multiple au sein des systèmes d'information géographique basés sur une structure relationnelle. Afin de formaliser le contexte d'exploitation de la représentation multiple, un métamodèle (modèle décrivant les composantes d'un modèle et non les données elles-mêmes) fut alors élaboré (figure 2-5).

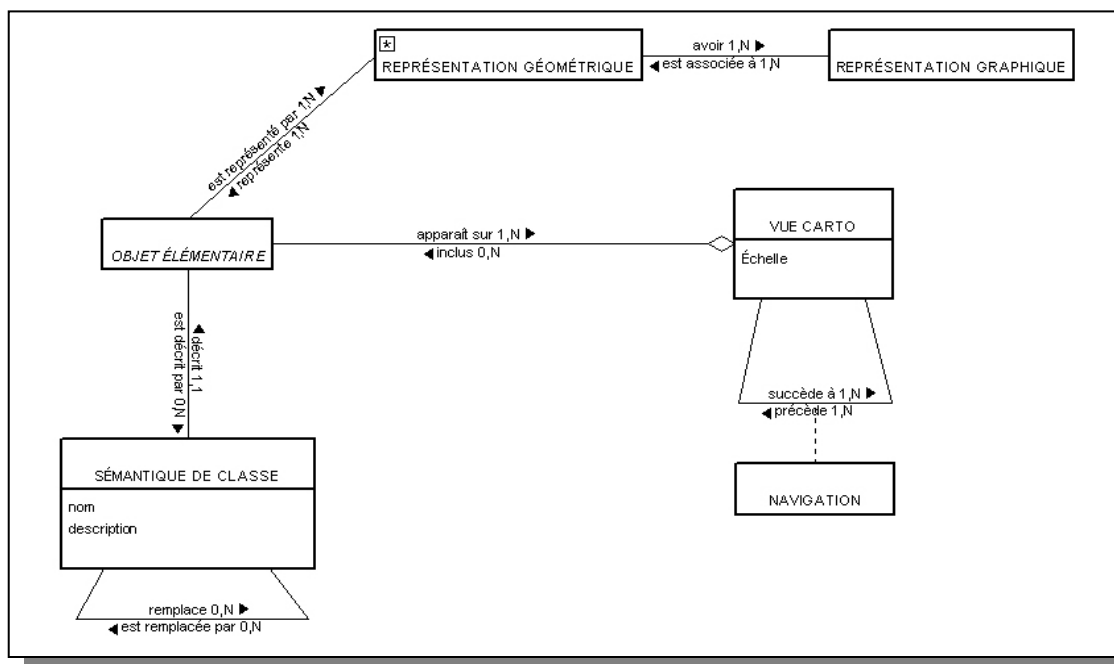


Figure 2-5 Métamodèle de la représentation multiple élaboré par Devillers (2000)



Ce métamodèle intègre les concepts de représentation multiple au contexte de navigation parmi des vues spatiales. Sans toutefois en faire une description exhaustive, les principes de base de ce métamodèle considèrent les vues cartographiques comme des agrégats d'objets élémentaires. Ces derniers peuvent être associés à plusieurs sémantiques de classe, plusieurs géométries et, implicitement, à plusieurs représentations graphiques.

De nombreuses réflexions ont par la suite donné naissance au concept du VUEL [Bédard *et al.*, 2000a]. Le métamodèle original de ce concept est présenté à l'annexe B. L'une des principales caractéristiques réside en l'adéquation du VUEL au contexte OLAP. En effet, un vuel, qui représente tout élément pouvant apparaître dans une vue quelconque, peut correspondre à différents concepts dont ceux propres à un contexte OLAP (graphique statistique, matrice, etc.). De plus, des opérations de type OLAP y sont également incluses (*drill down* et *drill up*) afin de naviguer parmi les différents vuels.

Le développement de ce métamodèle qui fut continué dans le cadre du présent mémoire avec mon directeur de recherche est présenté au chapitre suivant et son implantation fait l'objet du chapitre 4. Parmi les développements ayant été réalisés, notons l'enrichissement du volet sémantique notamment par l'ajout d'une classe servant d'intermédiaire entre le vuel et ses composantes de nature sémantique (classe d'objets, attributs, domaines de valeurs), la définition de nouvelles opérations OLAP spatiales, sémantiques et graphiques (ex. forage, remontage) et l'ajout de relations récursives de généralisation pour chacun des volets (i.e. géométrique, sémantique et graphique).

## **2.7. Conclusion**

Ce chapitre a présenté en premier lieu les notions relatives au processus de généralisation suivies des difficultés rencontrées lors de son automatisation. Celles-ci sont principalement dûes à la nature intuitive et itérative de la généralisation, qui font de cette dernière, un processus difficile à formaliser et à réaliser rapidement de manière

automatique. Par la suite, le concept de représentation multiple fut discuté principalement en ce qui a trait à sa gestion au sein d'une base de données géospatiale ainsi qu'aux problématiques associées. Les principales approches existantes furent présentées et sont majoritairement multiéchelles et basées essentiellement sur l'aspect géométrique des données. Le fondement des outils OLAP ainsi que leur applicabilité au domaine spatial suivirent. Finalement les préceptes du concept de VUEL furent discutés. Le chapitre qui suit introduit les différentes notions sous-jacentes à l'exploitation et à la gestion des représentations multiples dans un contexte SOLAP en plus de présenter les développements dont a fait l'objet le concept de VUEL.

## **Chapitre 3**

### **La représentation multiple dans un contexte SOLAP**

#### **3.1. Introduction**

Alors que le chapitre précédent a introduit les principaux concepts à la base de cette recherche, le présent chapitre se concentre sur les aspects liés au contexte SOLAP. Notamment, on y discute des concepts fondamentaux des outils SOLAP en plus de définir la notion de vue. Également, la nature et l'exploitation des représentations multiples au sein d'un outil SOLAP sont présentées. Finalement, le concept de VUEL fait l'objet d'une description plus détaillée.

#### **3.2. SOLAP : les concepts de base**

Tel que mentionné au chapitre précédent, la particularité d'un outil SOLAP est qu'il permet l'analyse spatio-temporelle rapide et intuitive des données. Plus précisément, Bédard (1997) définit un outil SOLAP comme une « plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle simples et rapides des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation via un affichage

cartographique, tabulaire ou en diagramme statistique ». En fait, un outil SOLAP permet, en plus des fonctionnalités actuellement offertes par un outil OLAP traditionnel, la visualisation cartographique des données ayant une composante spatiale géométrique. Un tel outil se compose de deux volets, l'un descriptif et l'autre cartographique. L'analyse des données peut être effectuée via l'un ou l'autre des volets ou encore, via les deux volets concurremment. À l'instar des outils OLAP traditionnels, les outils SOLAP reposent sur une approche multidimensionnelle. Cette approche est basée sur les concepts de *dimensions* et de *mesures*, et sera brièvement décrite dans les paragraphes qui suivent. Pour une description plus complète, le lecteur est invité à consulter [Caron, 1998; Rivest et al., 2001; Thomsen, 1997; Thomsen et al., 1999].

D'une part, les dimensions désignent les différentes thématiques selon lesquelles les données d'une organisation peuvent être analysées. Chacune de ces dimensions se compose de membres qui sont organisés hiérarchiquement selon leur niveau de granularité. Par exemple, pour une dimension *Région géographique*, les termes *Ville*, *Province* et *Pays* pourraient être utilisés pour désigner les niveaux hiérarchiques alors que *Toronto*, *Ontario* et *Canada* seraient des membres de chacun de ces niveaux respectivement.

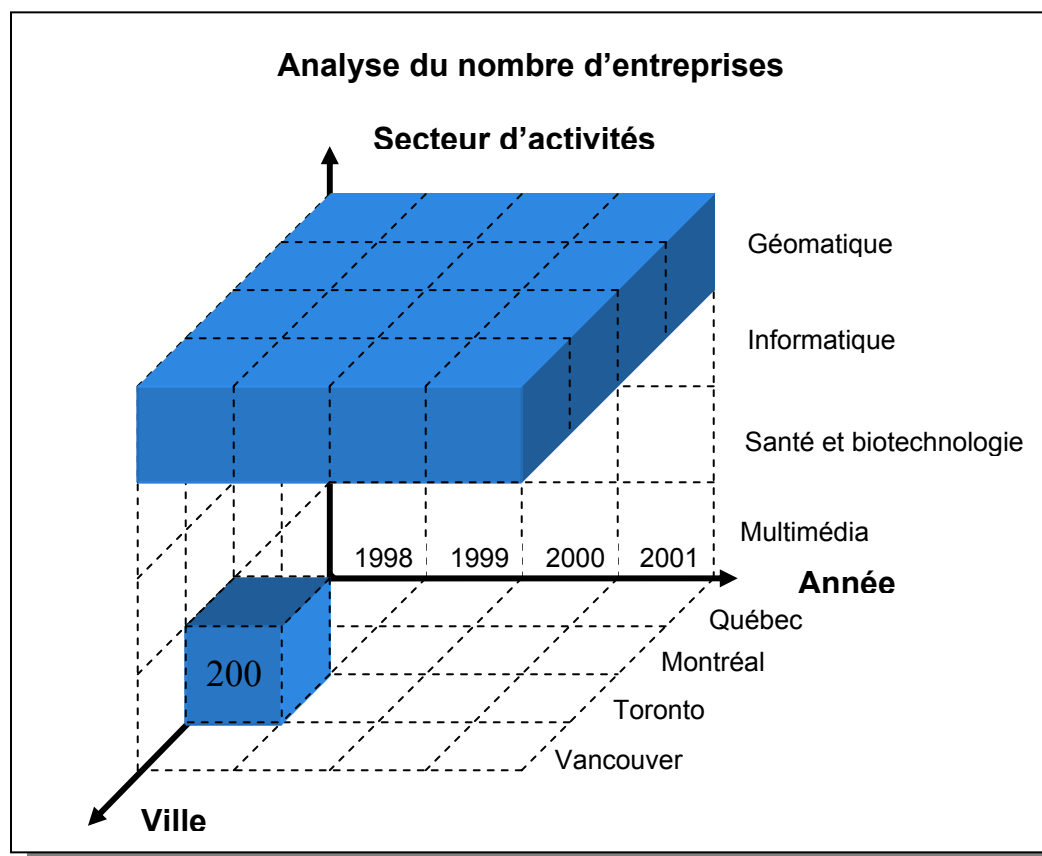
Au sein d'un outil SOLAP, cinq catégories de dimensions peuvent être identifiées : les dimensions temporelles, les dimensions descriptives, les dimensions spatiales descriptives, les dimensions spatiales géométriques, les dimensions spatiales hybrides [Bédard et al., 2000b; Han et al., 1998 ; Rivest et al., 2001]. Une **dimension temporelle**, tel que son nom l'indique, ne contient que des membres ayant une nature temporelle (ex. une dimension *Temps* composée des niveaux *Année*, *Mois* et *Jour*). Une **dimension descriptive** contient des membres dépourvus de composantes spatiales (ex. une dimension *Produit*). Une **dimension spatiale descriptive** contient des membres ayant une composante spatiale strictement nominale (ex. nom de lieu, code postal). Cette dernière dimension peut être implantée de manière similaire à une dimension descriptive puisqu'aucune visualisation cartographique n'est possible avec ces seules données

[Bédard *et al.*, 2000b]. Une **dimension spatiale géométrique** contient des membres possédant une composante spatiale géométrique pour tous les niveaux de la hiérarchie. Une **dimension spatiale hybride** contient des membres ayant une composante spatiale géométrique ou descriptive, selon leur niveau hiérarchique. Ainsi, il est possible que les membres situés aux niveaux inférieurs de la hiérarchie (i.e. ayant une granularité fine) possèdent une composante spatiale géométrique alors que les membres situés aux niveaux supérieurs (i.e. ayant une granularité grossière) détiennent une composante spatiale descriptive seulement (**dimension spatiale géométrique vers descriptive**). Par exemple, pour une dimension composée des niveaux *Ville*, *MRC*, *Province* et *Pays*, seulement les deux premiers niveaux (*Ville* et *MRC*) pourraient être représentés cartographiquement alors que les niveaux *Province* et *Pays* seraient présentés à l'utilisateur d'une manière nominale seulement. Mentionnons que le contraire est également possible, i.e. **dimension spatiale descriptive vers géométrique**.

D'autre part, les mesures sont des données généralement numériques sur lesquelles portent les analyses et sont structurées selon les différentes dimensions. Pour un outil SOLAP, deux types de mesures sont possibles : les **mesures descriptives numériques** et les **mesures spatiales**. Le premier type concerne les mesures qui sont actuellement représentées dans les outils OLAP et qui ne possèdent aucune composante spatiale. Le deuxième type, les mesures dites spatiales, représentent les mesures ayant une composante spatiale géométrique. Un exemple de ce type de mesure serait l'ensemble des polygones résultant de la combinaison des membres polygonaux provenant de deux dimensions spatiales (ex. limite municipale et bassin versant).

Finalement, la construction multidimensionnelle résultant de la conjonction de plusieurs dimensions forme ce qu'il convient d'appeler un « hypercube » dans lequel chaque cellule est définie par un seul membre de chaque dimension et représente la valeur d'une mesure à l'intersection de ces dimensions. La figure suivante illustre un hypercube formé de trois dimensions (*Année*, *Ville* et *Secteur d'activités*) dont l'analyse porte sur le *Nombre d'entreprises* (la mesure). Le cube du bas indique que pour l'année 1998, le

nombre d'entreprises oeuvrant dans le domaine du multimédia à Toronto se chiffrait à 200. De manière similaire, la strate du haut représente l'agrégation du nombre d'entreprises géomatiques situées dans chacune des villes et ce, pour les quatre dernières années (figure 3-1).



**Figure 3-1** Exemple d'un hypercube

### 3.2.1. Opérations de navigation

La navigation au sein d'un outil SOLAP peut être effectuée grâce à l'utilisation de différentes opérations. Parmi les opérations de base figurent le *pivot*, le *forage*, le *remontage* ainsi que le *forage latéral*.

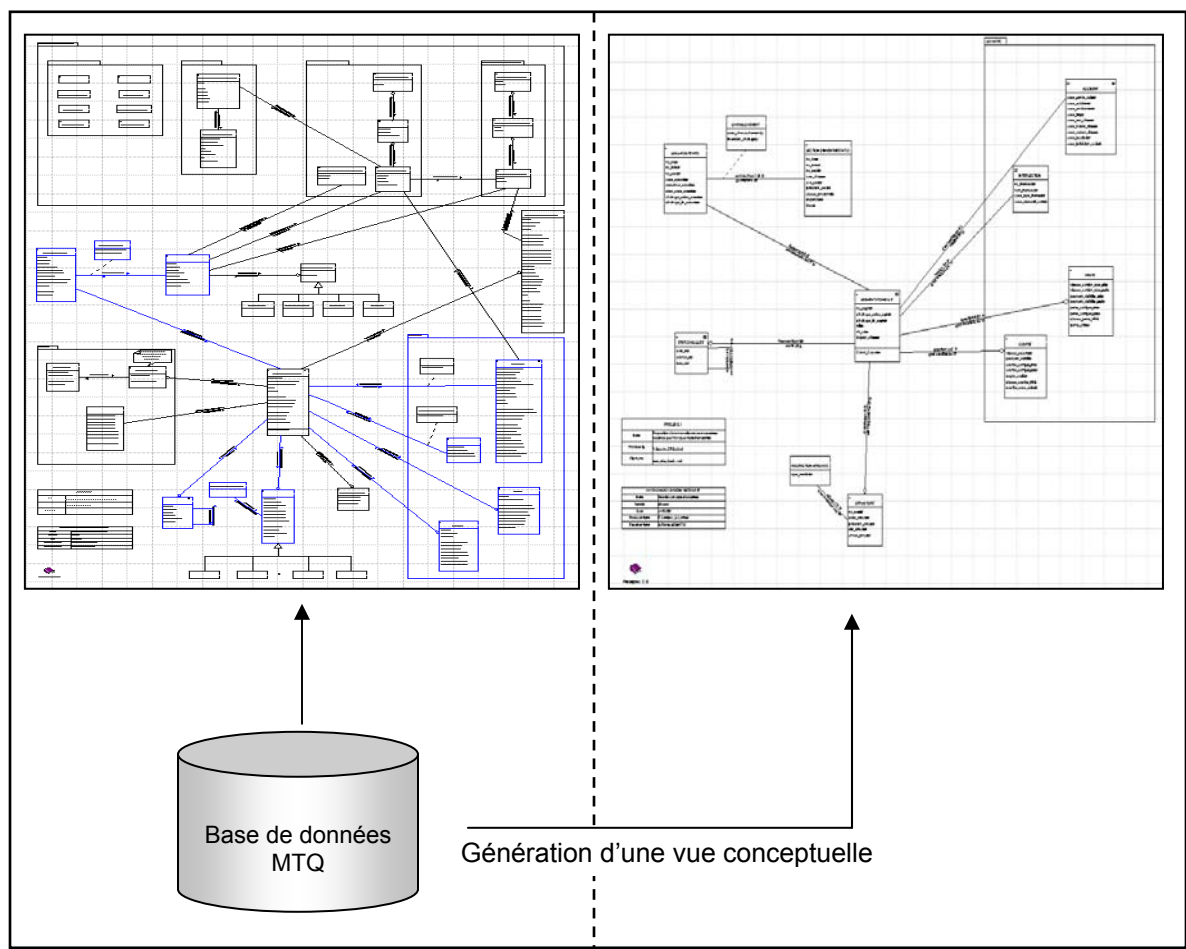
Étant donné que seul un nombre restreint de dimensions peuvent être visualisées simultanément, l'opération de **pivot** (*swap*) consiste à interchanger deux dimensions et permet ainsi d'analyser les données sous différents axes ou thèmes. L'opération de **forage** (*drill-down*) permet, pour une même dimension, de passer d'un niveau hiérarchique donné au niveau immédiatement inférieur. Il est donc possible d'analyser les données selon un niveau de granularité plus fin. Par exemple, pour la dimension *Temps*, le passage du niveau *Année* vers le niveau *Mois* consiste en une opération de forage. À l'inverse, l'opération de **remontage** (*drill-up* ou *roll-up*) consiste, pour une même dimension, à passer d'un niveau hiérarchique au niveau situé immédiatement au-dessus. L'analyse des données se fait donc à un niveau plus grossier. Finalement, le **forage latéral** (*drill-across*), permet soit (1) le passage d'une mesure à une autre pour les mêmes dimensions d'analyse et pour un même niveau de granularité, (2) le passage d'un niveau hiérarchique alternatif à un autre qui se situe au même niveau de granularité, pour une même dimension ou (3) le passage d'une strate à une autre, donc l'analyse d'un membre d'une autre dimension qui est situé dans un même niveau hiérarchique. L'ensemble de ces opérateurs peuvent être utilisés aussi bien dans le volet descriptif que dans le volet cartographique. Grâce à ceux-ci, l'utilisateur a accès rapidement aux différents niveaux hiérarchiques et peut aisément naviguer entre eux.

### 3.3. Les vues alternatives : un nouveau besoin des outils SOLAP

#### 3.3.1. La notion de vue

Dans notre domaine de recherche, une vue est définie comme un sous-ensemble exact ou modifié d'une base de données. Par exemple, un utilisateur, selon son domaine d'application, ne pourrait être intéressé que par quelques classes d'objets et quelques attributs présents dans la base de données. Selon le mécanisme de vue, il n'aurait ainsi accès qu'aux éléments qui lui sont d'intérêt. La figure qui suit illustre une vue conceptuelle dérivée de la base de données routières du ministère des Transports du

Québec (MTQ). Elle ne présente qu'un sous-ensemble des éléments présents dans la base de données, notamment ceux en relation avec les accidents de la route. Une telle vue pourrait même être modifiée (ex. inclure des classes d'objets rebaptisées selon le vocabulaire de l'utilisateur ou des attributs nouveaux qui sont dérivés des données stockées).

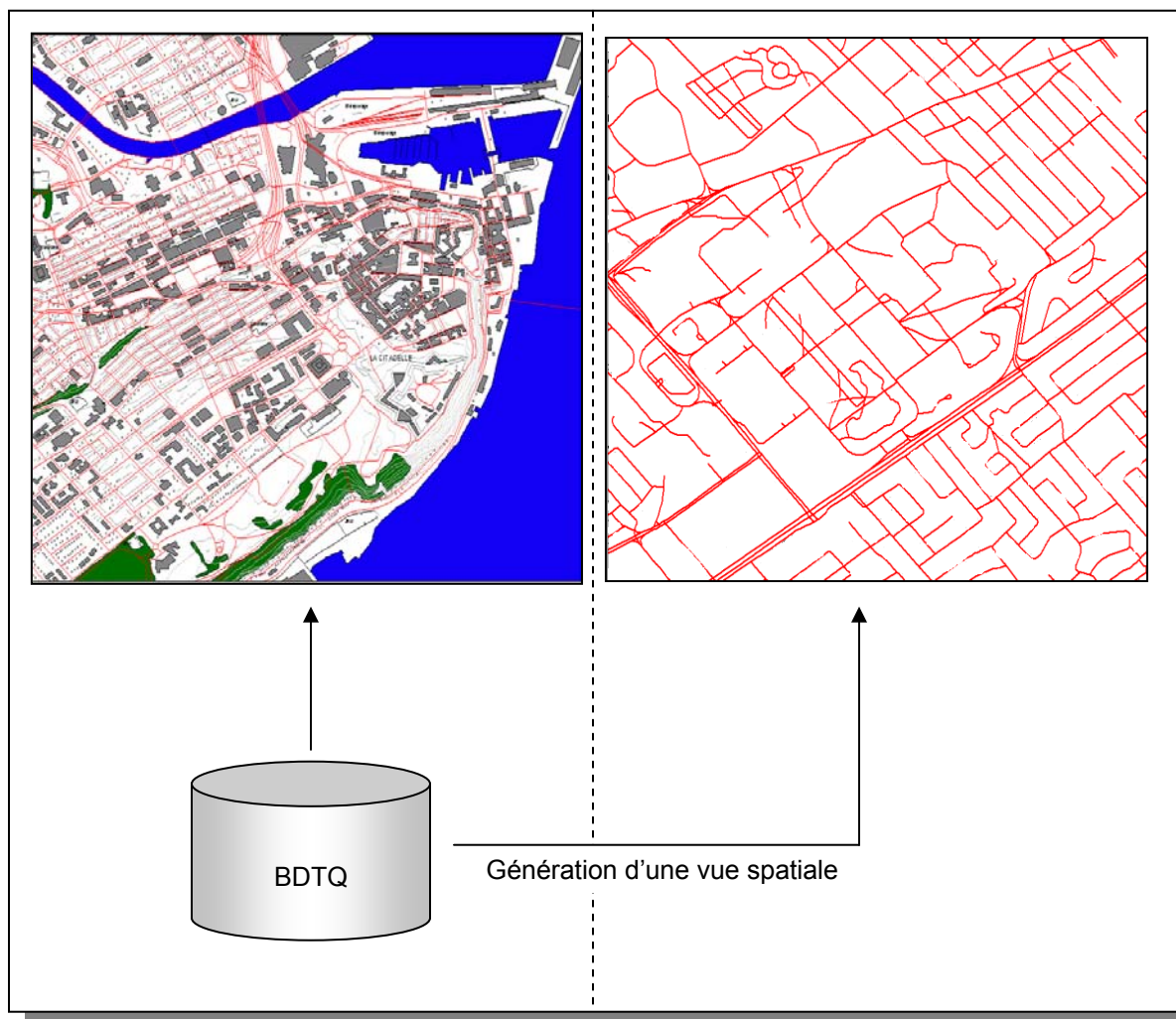


**Figure 3-2** Exemple d'une vue conceptuelle

Bien que la notion de vue soit présente dans la littérature des bases de données depuis bon nombre d'années, l'aspect spatial de cette dernière est un concept relativement récent. Conséquemment, peu d'auteurs se sont penchés sur la notion de "vue spatiale". Selon Claramunt et Mainguenaud (1995), une vue est composée d'un ensemble d'atomes, chaque atome étant défini par un ensemble de relations, un ensemble d'opérateurs spatiaux et non-spatiaux ainsi qu'un ensemble d'opérateurs de visualisation. La figure



suivante représente une vue spatiale pouvant être générée à partir de la base de données topographiques du Québec (BDTQ). Elle ne présente en fait que les éléments appartenant à la classe d'objets *Route*.



**Figure 3-3** Exemple d'une vue spatiale

Les systèmes relationnels gèrent le mécanisme de vue en permettant la dérivation de tables virtuelles à partir des tables existantes dans la base de données. Celles-ci, bien qu'elles ne soient pas physiquement emmagasinées dans la base de données, peuvent faire l'objet de requêtes et même, dans certains cas, peuvent être modifiées.

### 3.3.2. Les vues dans un contexte SOLAP

Actuellement, les données emmagasinées dans une structure multidimensionnelle et pouvant être analysées par un outil SOLAP ne possèdent qu'une seule représentation géométrique. Par conséquent, les phénomènes ne peuvent être analysés que selon un seul point de vue et préférablement qu'à une seule échelle. Toutefois, il pourrait être intéressant, par exemple, de diminuer l'échelle de visualisation afin d'analyser un même phénomène mais pour une plus grande couverture territoriale. Cependant, un simple « zoom out » aurait probablement comme conséquence d'engendrer de nombreux conflits entre les objets et résulterait en une représentation plus ou moins lisible (figure 2-2). En réalité, une diminution minime de l'échelle de représentation pourrait ne pas avoir de répercussion au niveau de la cohérence et de l'esthétique des informations représentées. Toutefois, une diminution significative pourrait résulter en une représentation illisible et ainsi nuire au processus d'analyse en cours. Il serait donc important d'offrir, pour un même objet, différentes représentations plus ou moins simplifiées, afin de continuellement présenter à l'utilisateur des vues lisibles et cohérentes. Évidemment, les éléments à analyser à chaque niveau de réflexion ne sont pas toujours les mêmes mais il est impératif de retrouver un lien hiérarchique (autant d'un point de vue cartographique que sémantique) entre les différentes classes d'objets (ex. zone urbaine → construction) (figures 3-5 à 3-8).

Également, il serait intéressant de conserver différentes représentations géométriques d'un même membre mais pour un même niveau de granularité. L'utilisateur aurait ainsi la possibilité d'analyser les données via les représentations géométriques les plus appropriées à ses besoins. Par exemple, une rue pourrait être représentée avec une géométrie polygonale dans une application cadastrale et une géométrie linéaire dans une application pour l'analyse du trafic (figure 3-6).

Bref, afin que l'utilisateur puisse exploiter pleinement les capacités d'analyse d'un outil SOLAP, ce dernier devrait pouvoir offrir des vues **alternatives**, de façon quasi-

instantanée, via des mécanismes de généralisation automatique et/ou de représentation multiple [Rivest *et al.*, 2001].

Conformément à ce qui fut présenté au chapitre deux, la généralisation automatique actuelle ne peut pas répondre aux contraintes de rapidité inhérentes à un outil SOLAP. Les temps de traitement élevés ainsi que la nécessité d'une intervention humaine lors du processus en sont les raisons principales. De plus, même si la généralisation automatique pouvait être utilisée pour produire plusieurs versions d'un même objet selon différents niveaux de granularité, certaines géométries supplémentaires devraient être emmagasinées explicitement dans la base de données. En effet, les différentes représentations géométriques d'un même objet, associées à un même niveau de granularité, ne peuvent normalement pas être déduites les unes des autres. Il s'agit généralement de nouvelles géométries saisies qui n'ont aucun lien entre elles hormis le fait qu'elles représentent le même objet (ex. polygone représentant la toiture d'une maison et polygone représentant les fondations d'une maison). Finalement, dans le cadre de cette recherche, la généralisation automatique ne peut être utilisée que pour produire, en prétraitement, les représentations géométriques simplifiées d'un même objet. Ces dernières devant être par la suite emmagasinées explicitement dans la base de données à l'aide d'une structure de données supportant les représentations multiples de façon adéquate. En somme, l'unique moyen efficace d'offrir des vues alternatives dans un contexte SOLAP se résume au stockage et à la gestion explicites des représentations multiples.

### **3.4. La nature et l'exploitation des représentations multiples dans un contexte SOLAP**

Lors de la définition de la représentation multiple, il fut énoncé que dans son sens strict, ce terme n'englobait que les caractéristiques visuelles d'un objet, soit la géométrie et les variables visuelles [Martel, 1999]. Toutefois, dans le contexte SOLAP, l'analyse des

données peut également porter sur les caractéristiques sémantiques d'un objet (via le volet descriptif). Étant donné qu'il est incohérent de parler de représentation multiple sémantique, le terme « multiplicité » sera donc utilisé désormais pour éviter toute confusion lexicale. Ce terme ayant déjà été défini dans [Martel, 1999], nous n'en ferons qu'une brève description.

Il existe trois catégories de multiplicité soit : la multiplicité géométrique, la multiplicité graphique et la multiplicité sémantique (figure 3-4).

		MULTIÉCHELLE	UNIÉCHELLE									
MULTIPLICITÉ	GÉOMÉTRIQUE	<p>Vue A      Vue B      Vue C</p>	<p>Vue U      Vue V      Vue W</p>									
	GRAPHIQUE	<p>Vue A      Vue B</p>	<p>Vue U      Vue V</p>									
	SÉMANTIQUE	<table border="1"> <tr> <td>Construction</td> <td>Vue A</td> </tr> <tr> <td>Batiment</td> <td>Vue B</td> </tr> <tr> <td>Maison</td> <td>Vue C</td> </tr> </table>	Construction	Vue A	Batiment	Vue B	Maison	Vue C	<table border="1"> <tr> <td>Lot</td> <td>Vue U</td> </tr> <tr> <td>Unité d'évaluation</td> <td>Vue V</td> </tr> </table>	Lot	Vue U	Unité d'évaluation
Construction	Vue A											
Batiment	Vue B											
Maison	Vue C											
Lot	Vue U											
Unité d'évaluation	Vue V											

**Figure 3-4** Multiplicités géométrique, graphique et sémantique adaptée de Martel (1999)

La **multiplicité géométrique** réfère à la présence d'un même concept sous différentes représentations géométriques. Lorsque ces différentes géométries sont associées à des échelles différentes, la multiplicité est dite « multiéchelle ». Par exemple, un bâtiment peut posséder une géométrie détaillée (ex. polygone détaillé) à grande échelle et une géométrie simplifiée (ex. point) à une plus petite échelle. Au contraire, lorsqu'elles sont associées à une même échelle, on parle alors de multiplicité « uniéchelle ». Par exemple,

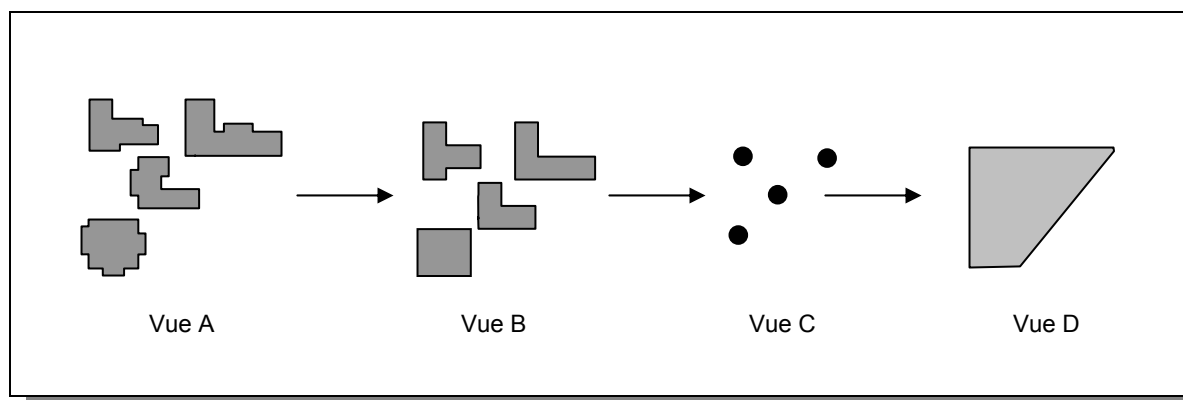
la diversité des méthodes et normes de saisies des objets peut engendrer différentes représentations d'un même objet pour une même échelle. Ainsi, un bâtiment peut être cartographié suivant sa toiture, ses fondations, etc.

La seconde catégorie, la **multiplicité graphique**, implique qu'un même objet puisse être représenté avec différentes variables visuelles (forme, dimension, valeur, orientation, couleur et texture) et selon différentes règles de sémiologie graphique. À l'instar de la multiplicité géométrique, elle peut être associée à un changement d'échelle (multiplicité graphique multiéchelle) ou non (multiplicité graphique uniéchelle). Par exemple, les symboles utilisés pour représenter certains objets (ex. écoles, églises) peuvent varier selon l'échelle d'affichage ainsi que selon le domaine d'application.

Finalement, la **multiplicité sémantique** signifie qu'un même élément graphique possède différentes sémantiques, donc peut représenter différents concepts. Par exemple, un élément graphique peut être défini comme une *résidence* à grande échelle et une *construction* à plus petite échelle, ou encore, un polygone peut représenter un lot cadastral pour une application et une unité d'évaluation foncière pour une autre application (étant donné qu'au Québec, ces deux notions législatives sont souvent identiques). On parle alors de multiplicité sémantique multiéchelle et uniéchelle respectivement (figure 3-4).

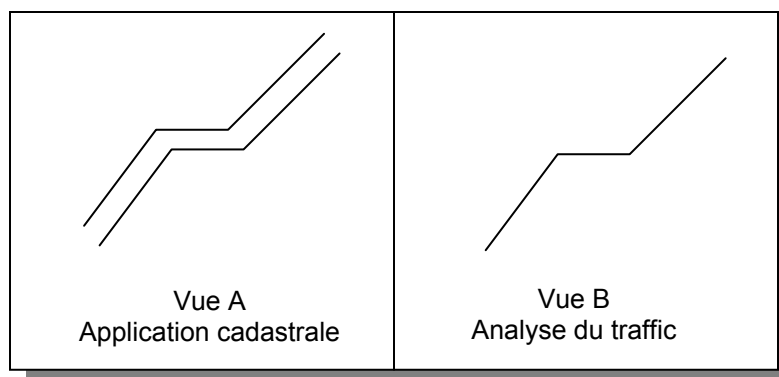
Tel que discuté à la section 3.3., la structure actuelle des outils SOLAP permet la navigation **entre** les différents niveaux hiérarchiques d'une ou plusieurs dimensions. Il est donc possible de forer ou de remonter dans une dimension mais il est actuellement impossible de naviguer à l'intérieur d'un **même niveau hiérarchique**. En réalité, les membres de chacun des niveaux hiérarchiques d'une dimension sont présentés à l'utilisateur suivant une seule représentation géométrique ou graphique et sont associés à une seule sémantique.

L'exploitation des trois types de multiplicités définis précédemment facilite la navigation au sein d'un outil SOLAP. En fait, chaque membre d'un même niveau hiérarchique d'une dimension peut être associé à une ou plusieurs multiplicités (géométrique, sémantique et graphique), ce qui permet de naviguer à l'intérieur de ce niveau hiérarchique. Par exemple, un membre peut posséder différentes géométries, chacune présentant un niveau de granularité donné. De ce fait, l'utilisateur peut *zoomer* de façon intelligente dans un niveau hiérarchique, c'est-à-dire qu'en augmentant l'échelle de visualisation, les objets seraient présentés avec plus de détails et vice versa. Prenons l'exemple d'une dimension dont le niveau hiérarchique inférieur se compose de bâtiments. À grande échelle, ceux-ci pourraient être représentés avec une géométrie polygonale détaillée. Au fur et à mesure que l'échelle de visualisation diminuerait, les mêmes bâtiments pourraient être représentés avec des géométries plus simplifiées (polygones simplifiés, points) allant même jusqu'à être finalement agrégés en un seul polygone à plus petite échelle (figure 3-5).



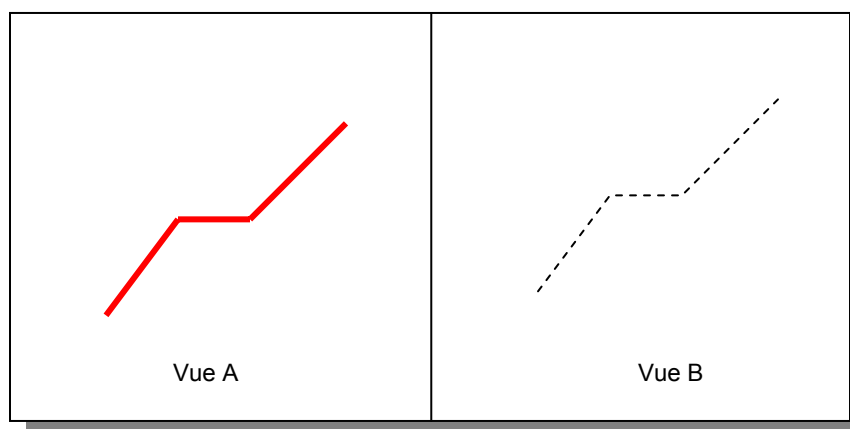
**Figure 3-5** *Différentes représentations géométriques pour les membres d'un même niveau hiérarchique (multiplicité géométrique multiéchelle)*

Également, en permettant la multiplicité géométrique uniéchelle, l'utilisateur peut analyser ses données via les représentations géométriques les plus appropriées à ses besoins. Par exemple, selon l'application, l'utilisateur pourrait visualiser les routes selon une géométrie polygonale ou linéaire (figure 3-6).



**Figure 3-6** Deux représentations géométriques d'une même route (multiplicité géométrique uniéchelle)

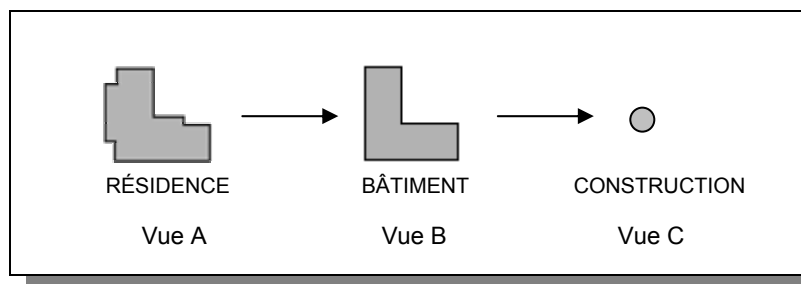
Par ailleurs, du point de vue graphique, il est possible de représenter un même élément selon différentes variables visuelles, ce qui permet une plus grande flexibilité de représentation et par conséquent, permet de répondre aux besoins des différents utilisateurs. La figure 3-7 montre une route représentée via des variables visuelles différentes répondant à deux besoins spécifiques.



**Figure 3-7** Deux représentations graphiques d'une même route (multiplicité graphique uniéchelle)

Enfin, en exploitant la multiplicité sémantique, un même élément peut être défini de plusieurs façons. Par exemple, la sémantique associée à un membre peut varier en fonction du niveau de granularité associé à sa représentation géométrique. Un objet

représenté avec beaucoup de détails sera habituellement associé à une description précise alors que le même objet, représenté avec une géométrie simplifiée, sera défini de façon plus générale (figure 3-8).



**Figure 3-8** *Différentes sémantiques en fonction du niveau de granularité géométrique (multiplicités géométrique et sémantique multiéchelles)*

Le même raisonnement s’applique également aux multiplicités sémantiques uniéchelles. Ainsi, différentes sémantiques, définies avec un même degré de granularité, peuvent être associées à un même élément. Le réseau routier en est un exemple concret. En effet, les routes, selon les normes du MTQ, appartiennent à la classe d’objets « route » alors que celles emmagasinées dans la BDTQ appartiennent à la classe d’objets « voie de communication ».

Somme toute, l’exploitation des multiplicités augmente la flexibilité de représentation et de définition des objets en plus d’accroître les possibilités de navigation au sein d’un outil SOLAP.

### **3.5. Les mises à jour dans un contexte SOLAP avec représentations multiples**

Plusieurs travaux de recherche ont porté sur le processus de mise à jour au sein des bases de données avec représentations multiples. Quoique nous soyons conscients de la difficulté associée à ce processus, le contexte SOLAP nécessite d’apporter certains éléments de réflexion.



Tout d'abord, les outils SOLAP sont généralement utilisés en tant que clients d'un entrepôt de données spatiales ou d'un marché de données. Un entrepôt de données se définit comme « une collection de données portant sur les sujets touchant une organisation, intégrée, variant dans le temps et non-volatile pour supporter le processus de décision d'une organisation » [Inmon *et al.*, 1996]. Contrairement aux systèmes transactionnels qui ne stockent que les données courantes, les données emmagasinées dans un entrepôt ne peuvent être modifiées ou détruites. Ainsi, les seules mises à jour possibles ne consistent qu'en l'ajout de nouvelles données. Par conséquent, deux types de données sont conservées, les données courantes et les données historiques. Un élément de temps est tout simplement associé à la donnée indiquant sa période de validité.

Au sein de tels systèmes, les mises à jour s'effectuent a priori au niveau du système source. Les données y sont modifiées afin de correspondre à la nouvelle réalité. Dans le but de mettre à jour l'entrepôt de données, il est donc nécessaire d'y ajouter les données ayant fait l'objet d'une mise à jour dans le système source. Toutefois, aucune donnée n'est détruite. Par conséquent, une valeur temporelle est ajoutée à l'ancienne donnée afin d'indiquer que sa période de validité se termine alors que celle associée à la nouvelle donnée débute. Mentionnons que la fréquence de mise à jour de l'entrepôt de données peut être différente de celle associée au système source. Par exemple, ce dernier peut être mis à jour quotidiennement alors que l'entrepôt de données peut l'être à chaque mois ou même à chaque année.

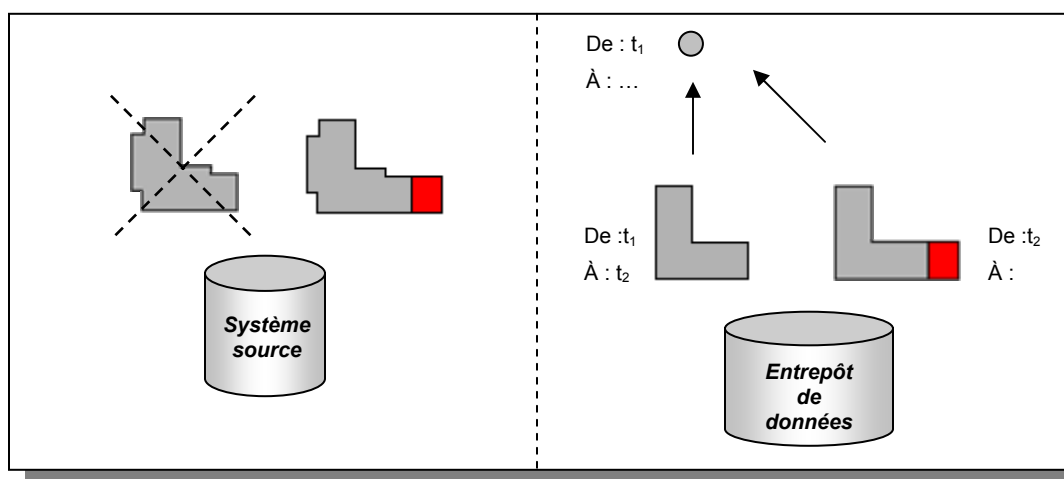
Différentes architectures peuvent être envisagées pour un entrepôt de données spatiales. Notamment, certaines conservent les données détaillées uniquement au niveau du système source alors que d'autres les emmagasinent également au niveau de l'entrepôt. Cette distinction, ainsi que plusieurs autres facteurs, influencent le processus de mise à jour dans un système SOLAP avec représentations multiples étant donné que ces outils sont généralement clients d'un entrepôt de données.

### 3.5.1. Seuil de répercussion des mises à jour dans un SOLAP

Bien qu'un membre puisse posséder différentes représentations et/ou valeurs, une mise à jour peut n'affecter qu'un sous-ensemble de celles-ci. Effectivement, il existe un seuil de répercussion pouvant varier d'une mise à jour à une autre. La présente section discute et illustre les différents niveaux de répercussion possibles pour une mise à jour.

#### Mise à jour de nature géométrique

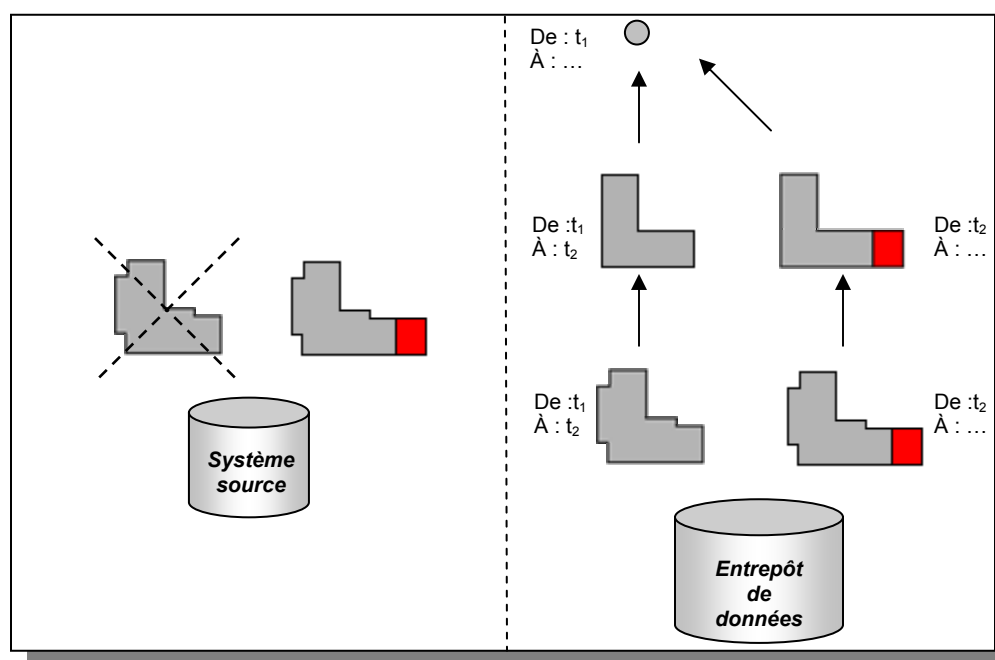
D'une part, dans le cas où les données détaillées sont exclusivement emmagasinées dans le système source, il est possible qu'une mise à jour n'ait aucune répercussion au niveau de l'entrepôt de données. Par exemple, un objet peut avoir une représentation polygonale dans le système source et une représentation ponctuelle dans l'entrepôt. Ainsi, une mise à jour affectant le polygone n'aurait de répercussion que dans le système source. D'autre part, il est possible que pour un membre possédant plusieurs géométries emmagasinées dans l'entrepôt de données, la mise à jour n'en affecte que quelques-unes. La figure suivante illustre ce deuxième cas (figure 3-9).



**Figure 3-9** *Seuil de répercussion d'une mise à jour de nature géométrique où les données détaillées sont conservées uniquement dans le système source*

Premièrement, la mise à jour s'effectue au niveau du système source. Cette nouvelle géométrie est par la suite généralisée et introduite dans l'entrepôt de données. Un intervalle de temps est associé à chaque géométrie indiquant la période temporelle selon laquelle elles sont valides. Pour cet exemple, la mise à jour n'a de répercussion qu'à un seul niveau puisque la géométrie du niveau suivant est trop simplifiée.

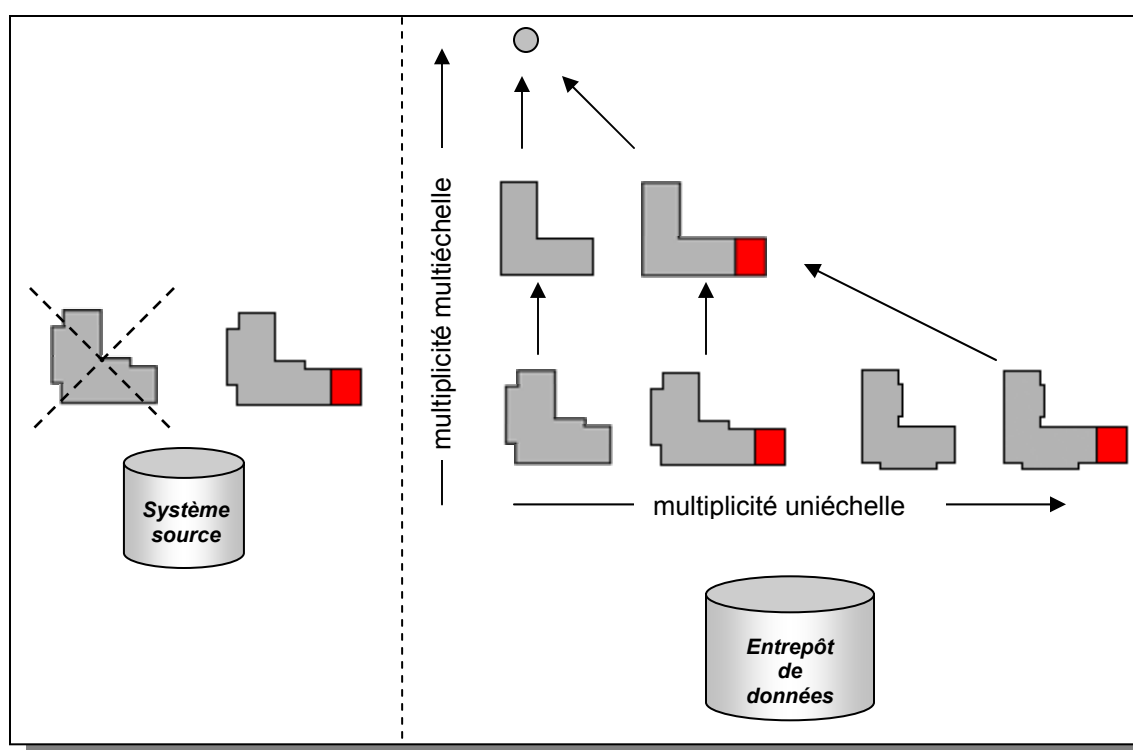
Par ailleurs, si les données détaillées sont également conservées dans l'entrepôt de données, une mise à jour aura inévitablement **au moins une répercussion** au niveau de celui-ci (figure 3-10). Par la suite, la répercussion dépendra notamment du niveau de granularité des autres représentations géométriques.



**Figure 3-10** *Seuil de répercussion d'une mise à jour de nature géométrique où les données détaillées sont conservées dans l'entrepôt de données*

Des règles peuvent donc être établies dans le but de contrôler la propagation d'une mise à jour. Par exemple, il pourrait être défini qu'une mise à jour correspondant à une surface plus petite que  $100 \text{ m}^2$  n'affecte que la représentation du niveau le plus détaillé alors qu'une mise à jour surfacique plus grande que  $100 \text{ m}^2$  affecte les trois premiers niveaux de granularité.

Jusqu'à maintenant, les exemples n'ont présenté que des cas de multiplicités géométriques multiéchelles. Comme il fut mentionné, un objet peut également posséder plusieurs géométries pour un même niveau de détail. La figure suivante illustre un cas de multiplicité géométrique à la fois multiéchelle et uniéchelle (figure 3-11). Cette dernière y est associée au niveau de granularité le plus fin de l'objet. La mise à jour engendre donc l'ajout de trois nouvelles géométries au niveau de l'entrepôt de données. Il est à noter que pour une question de lisibilité, les étiquettes temporelles n'apparaissent pas sur la figure suivante.

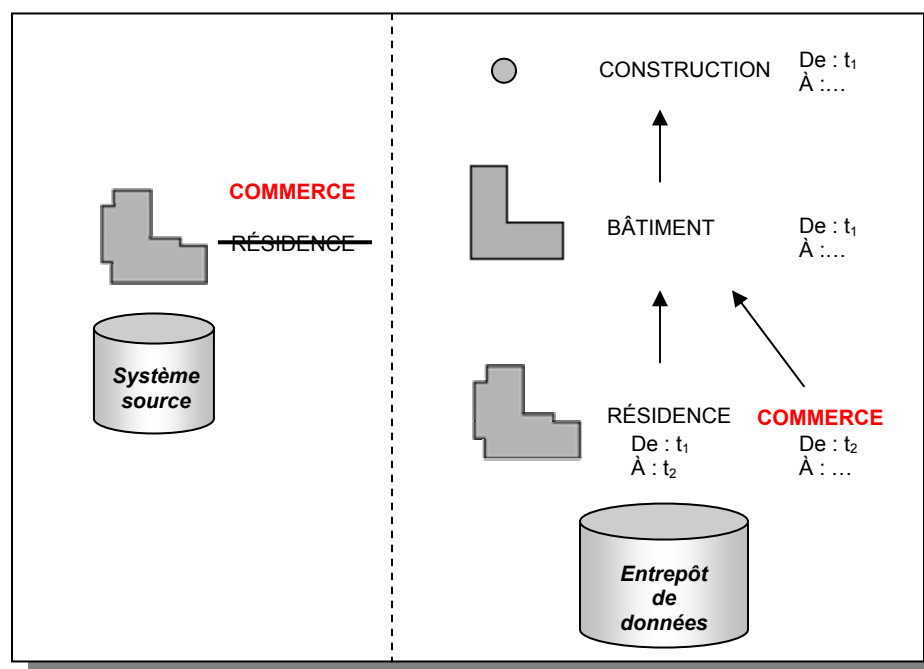


**Figure 3-11** *Seuil de répercussion d'une mise à jour de nature géométrique (multiplicité géométrique multiéchelle et uniéchelle)*

### Mise à jour de nature sémantique

D'un point de vue sémantique, une mise à jour peut concerner aussi bien les attributs d'un objet que sa nature même (classe d'objets). Par exemple, pour un objet représentant différents concepts en fonction du niveau de granularité de sa géométrie, il est possible

qu'une mise à jour n'affecte que le niveau le plus détaillé. Par exemple, si une résidence était maintenant entièrement vouée à des activités commerciales, ce changement de sémantique (changement au niveau de la classe d'objets) n'affecterait que le premier niveau de granularité sans aucune répercussion aux niveaux supérieurs. Le commerce devient, tout comme la résidence, un bâtiment au niveau suivant (figure 3-12).

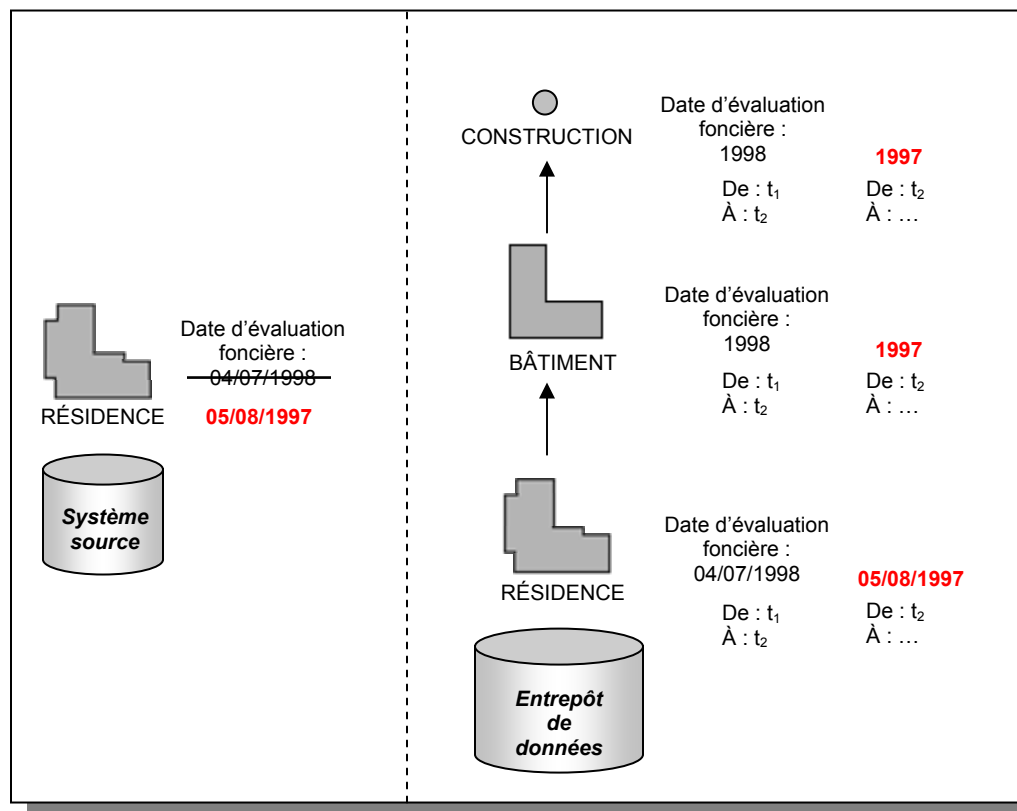


**Figure 3-12** *Seuil de répercussion d'une mise à jour de nature sémantique (au niveau de la classe d'objets)*

Également, la mise à jour d'une valeur associée à un attribut peut se restreindre à un sous-ensemble des niveaux, l'attribut pouvant ne pas être significatif ou même présent aux autres niveaux de granularité. Par exemple, la répercussion d'une mise à jour concernant le nombre de pièces d'une résidence dépend de la granularité sémantique de l'objet aux différents niveaux. En effet, cet attribut est significatif pour une résidence mais ne l'est pas pour un bâtiment ainsi que pour une construction.

D'un autre côté, il est possible que la mise à jour elle-même puisse posséder plusieurs niveaux de granularité. Notamment, selon le degré de granularité défini pour la valeur

d'un attribut, une mise à jour peut s'appliquer à tous les niveaux mais différemment. Prenons l'exemple d'un attribut « date d'évaluation foncière » associé à un objet. Au niveau le plus détaillé, la valeur de l'attribut est emmagasinée de façon détaillée (jour/mois/année). Toutefois, cette même valeur est stockée de façon plus simplifiée aux autres niveaux (seulement l'année). Ainsi, la mise à jour de cet attribut se répercute à tous les niveaux où l'attribut est présent mais selon différents degrés de détail (figure 3-13).

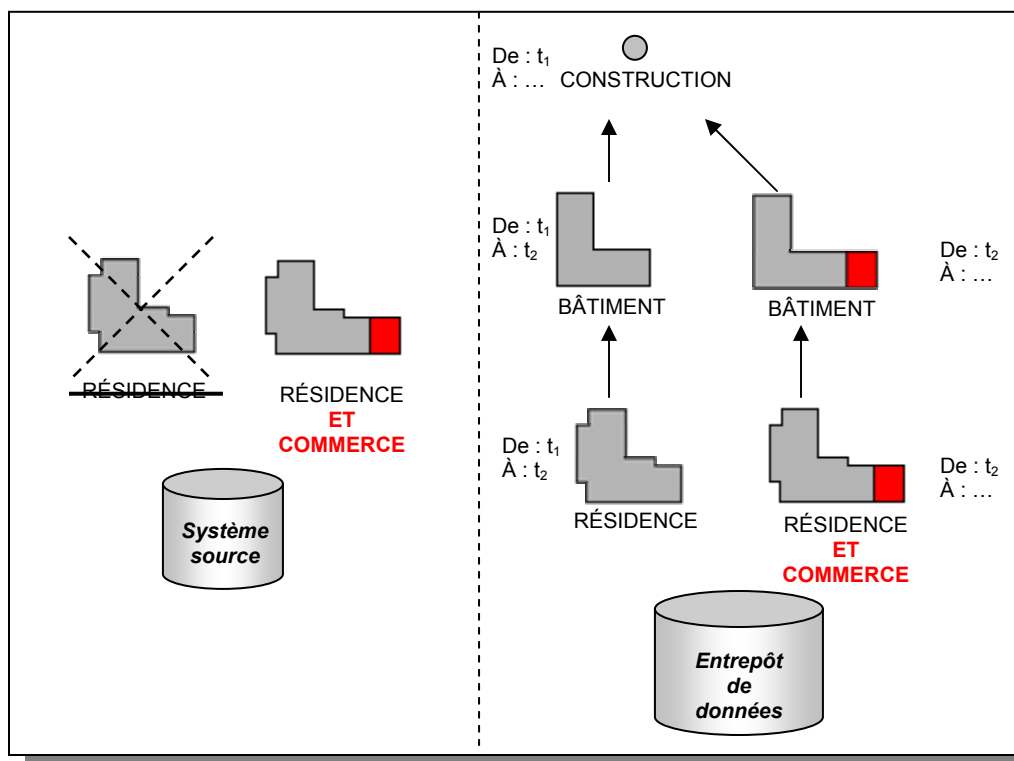


**Figure 3-13** *Seuil de répercussion et granularité d'une mise à jour de nature sémantique (au niveau de l'attribut)*

### Mise à jour sémantique et géométrique

Finalement, une mise à jour à la fois géométrique et sémantique peut avoir des seuils de répercussion différents selon chacun des points de vue. Par exemple, supposons maintenant que l'on ajoute à la résidence une nouvelle partie destinée à des fins commerciales (figure 3-14). La mise à jour au niveau sémantique, qui représente en fait la

naissance d'une nouvelle occurrence d'objets de la classe « Commerce », ne s'applique qu'au niveau le plus détaillé. De même, elle se concrétise géométriquement par l'agrandissement du polygone déjà existant pour la résidence. Toutefois, au niveau supérieur, cette mise à jour ne se traduit que par une évolution géométrique. Ainsi, du point de vue sémantique, la mise à jour n'affecte qu'un seul niveau alors que du point de vue géométrique, elle en affecte deux.



**Figure 3-14** *Seuil de répercussion d'une mise à jour à la fois sémantique et géométrique*

En d'autres termes, il existe un **seuil de répercussion** des mises à jour qui dépend entre autres de (1) la présence d'un objet (ou d'un attribut) dans un niveau hiérarchique et (2) du niveau de granularité associé à sa géométrie ou à sa sémantique.

### 3.6. Développement du concept de VUEL

Il est de notre avis que le concept de VUEL présenté au chapitre précédent offre une nouvelle solution très prometteuse afin d'exploiter la représentation multiple au sein d'un outil SOLAP.

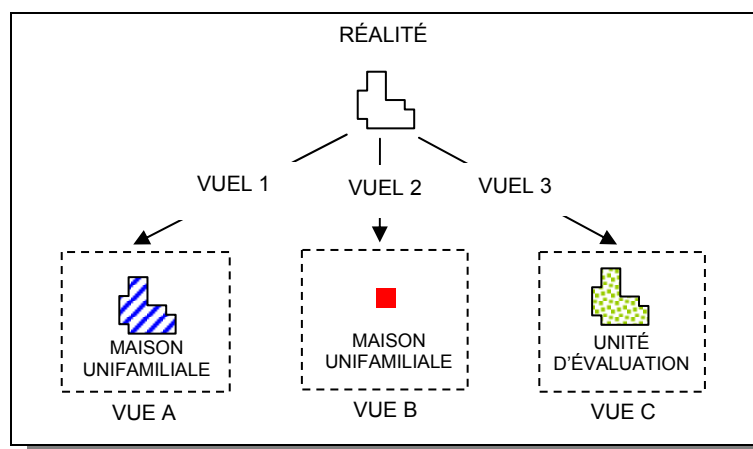
D'une part, celui-ci permet de gérer les trois types de multiplicité (géométrique, sémantique et graphique) de façon indépendante et explicite; ce qui permet une plus grande flexibilité lors de la navigation dans le SOLAP. D'autre part, contrairement à la plupart des approches présentées au chapitre deux, cette structure n'est pas basée uniquement sur l'aspect multiéchelle mais tient également compte des multiplicités possibles pour un même niveau de granularité. Enfin, il s'applique à tous les types de vues soit cartographique, tabulaire et en diagramme statistique.

Par conséquent, le concept de VUEL a été enrichi dans le cadre de cette recherche et cet enrichissement constitue une contribution importante du présent mémoire. Étant donné que ce concept a fait l'objet d'une très brève description au chapitre précédent, il est décrit plus en détail dans la présente section qui présente également les nouveautés apportées au concept de VUEL, i.e. au métamodèle de données SOLAP.

Tout d'abord, un vuel représente, tel que son nom l'indique, tout élément pouvant apparaître dans une vue quelconque [Bédard *et al.*, 2000a]. Il peut donc correspondre à différents concepts tels un polygone sur une carte, une cellule d'un tableau, une pointe de camembert ou une barre d'histogramme par exemple; il peut même correspondre à des artefacts tels que les éléments d'une légende cartographique ou d'un tableau. Tel que défini par Bédard, un vuel est la combinaison unique entre un élément visible (géométrie et sémiologie graphique) dans une vue et une sémantique particulière (ce qui est d'ailleurs représenté par les cardinalités 1 :1 de ses relations dans le métamodèle (figure 3-16)). Le vuel n'est pas l'objet sémantique per se (ex. maison) mais bien ce qu'on en voit dans une vue (ex. polygone bleu rayé ayant les attributs d'une maison). Ce même objet



sémantique peut être représenté par un point rouge dans une autre vue ou encore ce même polygone peut devenir vert piqué et représenter une unité d'évaluation avec des attributs différents pour une autre vue, et ainsi de suite (figure 3-15).



**Figure 3-15** *Trois vues servant à définir sémantiquement ou à représenter géométriquement et graphiquement une même réalité*

Le vuel a été conçu dès le début pour supporter toutes ces multiplicités sémantiques, géométriques et graphiques typiques des applications SOLAP. En ce sens, le métamodèle et le concept de vuel représentent le cœur d'un engin de gestion des vues graphiques et de leurs composantes. Naturellement, comme nous l'avons dit précédemment et en fonction de nos intérêts premiers, un vuel peut être un élément d'une vue cartographique. Cet élément pouvant être représenté géométriquement soit par des primitives (point, ligne, polygone) qui représentent un objet sémantique ou un artefact cartographique tel qu'une légende. Un vuel peut également être un élément d'une image, dans lequel cas il correspond au pixel (*picture element*) de l'image.

En raison de ce qui précède, un vuel sert donc à représenter et à définir les membres d'une dimension. Un même membre peut donc être associé à différents vuels, ce qui permettra de le représenter et de le définir différemment dans plusieurs vues.

Par ailleurs, plusieurs ajouts ont été apportés au métamodèle du concept de VUEL dans le cadre de cette présente recherche et sont discutés dans les paragraphes qui suivent.

### *1. Ajout d'une classe sémantique*

Afin de mieux gérer la sémantique pouvant être associée à un vuel, une nouvelle classe d'objets fut ajoutée au métamodèle (*sémantique*). Cette sémantique se compose d'une classe d'objets, des attributs pouvant lui être associés ainsi que des valeurs possibles (associées ou non à un domaine) pour ces attributs. Elle sert donc d'intermédiaire entre ces derniers éléments et le vuel et permet ainsi de le définir complètement.

### *2. Relation explicite entre les attributs et la sémantique*

Par ailleurs, les attributs pouvant être associés à une classe d'objets, et par conséquent à un vuel, sont maintenant reliés explicitement à la sémantique. Ainsi, contrairement à l'ancien modèle (annexe B), il n'est possible d'associer à un vuel qu'un sous-ensemble des attributs d'une classe donnée. Par exemple, pour une vue, la classe maison pourrait être associée à deux attributs (ex. adresse et date de construction) alors que pour une autre vue, cette même classe pourrait être associée à quatre attributs (ex. adresse, date de construction, valeur marchande et nombre d'étages).

### *3. Opérations SOLAP*

Des opérations de type SOLAP ont été ajoutées à chacune des composantes d'un vuel. Il est donc possible de naviguer dans les aspects géométrique, sémantique et graphique pouvant être associés à un vuel. Par exemple, un forage au niveau de la sémantique engendra une définition plus détaillée du membre représenté. De la même manière, un remontage au niveau de la géométrie représentera le même membre mais avec une géométrie plus simplifiée. Également, ce type d'opération fut inséré au niveau de la vue, qui représente en fait un agrégation de vuels. Par conséquent, il est possible d'effectuer une opération sur l'ensemble des vuels présents dans une vue. Par exemple, un forage spatial provoquera l'affichage d'une nouvelle vue composée des mêmes membres mais représentés avec une géométrie plus détaillée (lorsque possible).

#### *4. Relations récursives de généralisation*

Des relations récursives de généralisation ont été insérées dans le métamodèle. Par exemple, celle associée à la « Classe » indique que chaque classe peut être la généralisation d'une autre. Ceci permet les hiérarchies de classes en fonction du niveau d'abstraction et donc, la gestion de la multiplicité sémantique multiéchelle (ex. maison → bâtiment → construction). Des hiérarchies de valeurs (associées à un domaine) en fonction du degré d'abstraction peuvent également être gérées par le biais de la relation récursive de généralisation associée à la classe « Valeur ». Par exemple, le domaine de valeurs associé à l'attribut classe fonctionnelle des routes du MTQ peut comprendre, selon le degré d'abstraction, des valeurs détaillées (route provinciale, route régionale, route locale, autoroute régionale et autoroute nationale) ou générales (route et autoroute). Les relations récursives permettent donc de relier hiérarchiquement ces différentes valeurs. Le même principe s'applique également aux volets géométrique et graphique.

#### *5. Variables visuelles*

Une sémiologie est également définie pour chaque vuel. Cette sémiologie est déterminée par les variables visuelles que sont la couleur, la forme, la taille, la valeur, l'orientation et la texture.

Le métamodèle résultant est illustré à la figure suivante (figure 3-16). Ce dernier fut élaboré avec l'outil de modélisation Perceptory qui intègre le concept de PVL (Plug-in for Visual Language) spatial et PVL spatio-temporel aux composantes du modèle de classes du formalisme UML (Unified Modeling Language) [Bédard, 1999]. Ce qui permet la modélisation de bases de données spatio-temporelles.

La gestion temporelle dans le métamodèle du VUEL pourrait se faire par l'ajout de classes associatives au niveau des différentes relations du vuel. Une période de validité serait donc rattachée à chaque relation. Il serait ainsi facile de gérer la temporalité des

relations, et par conséquent, de fournir différentes vues selon différentes périodes de temps. Toutefois, le développement de la gestion temporelle dépasse le cadre du présent mémoire et sera traitée ultérieurement par un autre étudiant.

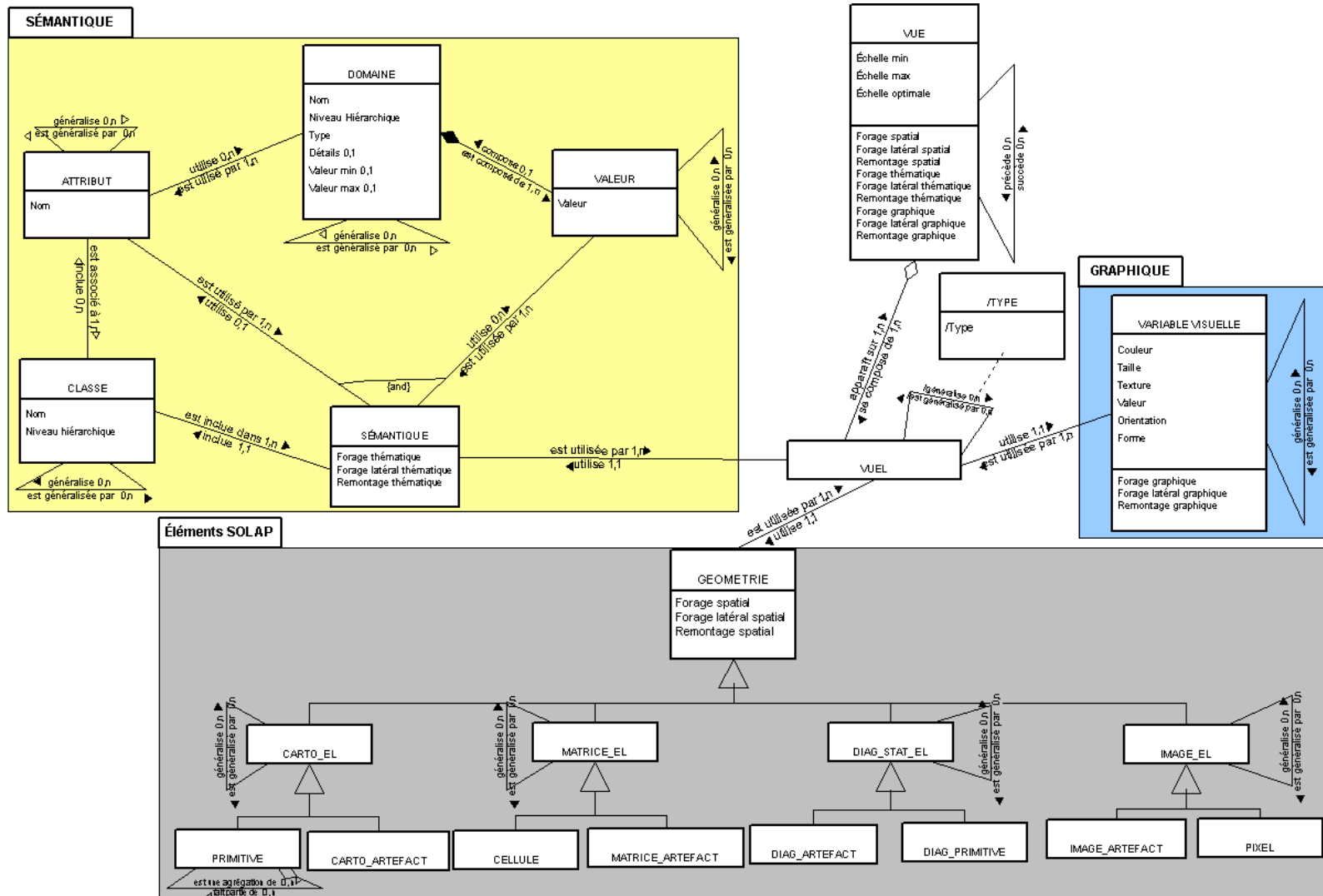


Figure 3-16 Métamodèle du concept de VUEL

### **3.7. Conclusion**

Ce chapitre a tout d'abord permis de définir les concepts de base d'un environnement SOLAP. Par la suite, un nouveau besoin pour les outils SOLAP fut identifié, celui d'offrir des vues alternatives. Considérant l'avancement actuel des algorithmes de généralisation automatique, il fut nécessaire d'envisager une approche basée sur la conservation des représentations multiples au sein d'une base de données spatio-temporelle. Par conséquent, la nature et l'exploitation des représentations multiples dans un contexte SOLAP furent précisées. De même, les facteurs influençant le processus de mise à jour dans un système SOLAP furent discutés. Finalement, une structure dédiée à la gestion des vues et des multiplicités en découlant, basée sur le concept initial de VUEL mais enrichi dans le cadre du présent mémoire, a été présentée. Le chapitre suivant est consacré à l'implantation de cette structure.

## **Chapitre 4**

### **Implantation du concept de VUEL**

#### **4.1. Introduction**

Afin de valider les notions théoriques associées à la gestion des multiplicités via le concept de VUEL, un prototype SOLAP permettant la génération de vues alternatives fut développé. Ce prototype, qui utilise des données topographiques et du réseau routier, gère les vues alternatives de type cartographique (donc ne contient pas les vues de type tabulaire ni graphique statistique), conformément aux objectifs du présent mémoire. Ce chapitre décrit l'implantation du concept de VUEL (figure 3-16) ainsi que l'interface SOLAP qui rend possible la navigation parmi les différentes multiplicités associées à un même vuel. Des copies d'écran illustrent le résultat final alors que le système développé est disponible sur l'ordinateur *Goliath* du CRG.

#### **4.2. L'architecture système**

Un outil SOLAP peut être développé suivant trois principales architectures. Celles-ci diffèrent principalement par le type de base de données utilisée pour emmagasiner les

données. Ces architectures ne seront présentées que très brièvement et donc, pour une description plus complète, le lecteur est invité à consulter [Rivest *et al.*, 2001; Thomsen *et al.*, 1999].

Premièrement, l'architecture ROLAP (*Relational OLAP*) conserve les données détaillées et agrégées au sein d'une structure relationnelle. Celle-ci doit cependant être relativement dénormalisée afin d'accélérer les temps de réponse. Bien qu'elle possède l'avantage de minimiser la redondance des données comparativement aux autres structures décisionnelles (celles-ci sont emmagasinées dans une seule structure contrairement aux architectures suivantes), elle présente toutefois l'inconvénient de générer les temps de réponse les plus longs [Thomsen *et al.*, 1999].

Deuxièmement, l'architecture MOLAP (*Multidimensional OLAP*) emmagasine les données détaillées et agrégées dans une structure multidimensionnelle propriétaire (cube OLAP). Du fait que les données sont structurées de manière similaire à l'utilisation pouvant en être faite, cette architecture présente les performances les plus élevées [Thomsen *et al.*, 1999]. Cependant, suite à une mise à jour au niveau des données de base, l'ensemble des données doivent être rechargées dans la structure multidimensionnelle en plus de devoir recalculer toutes les agrégations. Ceci diffère des autres architectures pour lesquelles seulement les agrégations doivent être recalculées [Rivest *et al.*, 2001].

Finalement, l'architecture HOLAP (*Hybrid OLAP*) résulte de la combinaison des deux architectures précédentes. En fait, les données détaillées sont conservées dans une structure relationnelle alors que les données agrégées et résumées sont emmagasinées directement au sein d'une structure multidimensionnelle. La récupération des données via l'une ou l'autre des structures s'effectue de manière transparente à l'utilisateur. Cette architecture est à privilégier lorsque les analyses requièrent principalement des données agrégées et que le volume des données détaillées est significatif [Microsoft Corporation, 2000].

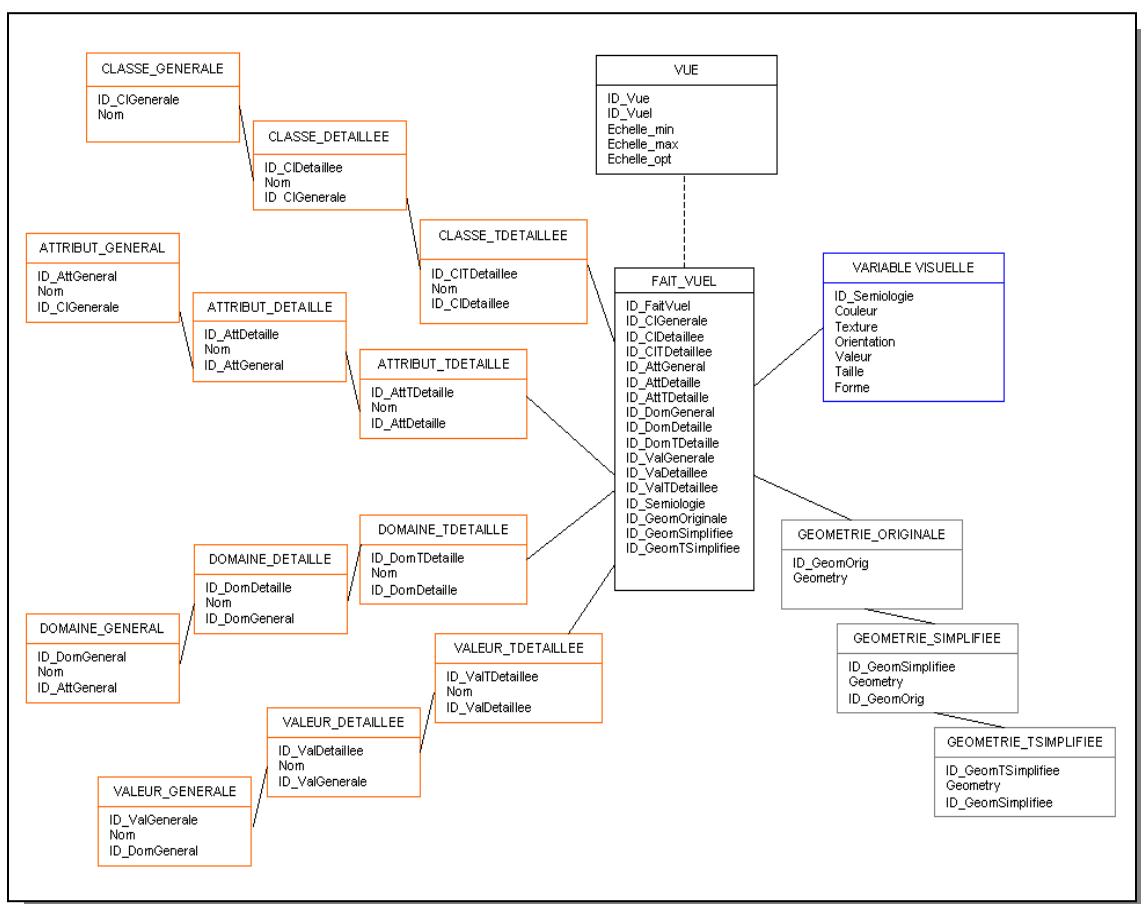


Pour ce qui est de notre implantation, l'architecture ROLAP fut privilégiée. D'une part, il était essentiel de démontrer que le concept de VUEL pouvait facilement être implanté sur une structure relationnelle commerciale (ex. Access). D'autre part, contrairement à l'architecture HOLAP, celle-ci ne requiert pas le stockage des données dans une structure parallèle (multidimensionnelle), ce qui facilite la maintenance du système. Enfin, il était souhaitable, dans un premier temps, d'expérimenter la structure VUEL dans l'environnement le plus flexible et non-contraint par une structure propriétaire (comme on rencontre dans les architectures MOLAP et HOLAP), ceci afin de concentrer nos efforts sur le concept de VUEL per se.

L'exploitation d'une base de données relationnelle à des fins d'analyse requiert une certaine dénormalisation au niveau de la structure des données. Ceci permet notamment la diminution du nombre de tables à joindre et augmente ainsi la performance d'interrogation des données. Pour ce faire, différents modèles de données simulant une structure multidimensionnelle peuvent être utilisés (ex. étoile, flocon, constellation ou mixte). Ces modèles reposent tous sur les concepts de dimensions, de mesures et de faits. Comme il fut mentionné au chapitre précédent, une dimension représente un thème d'analyse alors qu'une mesure est une donnée généralement numérique sur laquelle portent les analyses. Le fait représente la valeur d'une mesure résultant de la conjonction de plusieurs membres appartenant chacun à une dimension particulière.

La figure qui suit présente le modèle en flocon qui fut créé pour notre implantation du concept de VUEL. Ce modèle présente six dimensions (classe, attribut, domaine, valeur, variable visuelle et géométrie), chacune ayant été décomposée selon trois niveaux hiérarchiques, excepté la dimension *variable visuelle*. Nous avons décidé de limiter ce prototype à trois niveaux hiérarchiques pour des fins purement pragmatiques basées sur notre expérience, sur le temps dont nous disposons pour réaliser ce prototype ainsi que sur le fait que trois niveaux nous apparaissaient suffisant pour tester le concept de VUEL. Dans le cadre d'un projet, il faudrait évidemment définir ce nombre de niveaux en fonction des besoins des utilisateurs et ceci, pour chaque dimension. Pour les mêmes

raisons, bien qu'il soit possible de définir différentes représentations graphiques en fonction du niveau de granularité (ex. pour une granularité grossière, les objets peuvent être représentés via une seule teinte de bleu alors que pour une granularité fine, ils peuvent l'être selon plusieurs teintes de bleu), nous avons décidé que seule la multiplicité graphique uniechelle serait implantée dans le cadre de cette recherche.



**Figure 4-1** Modèle en flocon du concept de *VUEL*

Toutes les dimensions sont reliées à une table centrale, communément appelée « table des faits », qui contient les identifiants de chaque dimension. Généralement, les occurrences de cette table représentent des faits. En ce qui nous concerne, chaque occurrence ou chaque fait représente en réalité un vuel puisque celui-ci, de par sa définition, résulte de la combinaison unique d'une sémantique, d'une géométrie ainsi que d'une représentation graphique. Ainsi, la table centrale (*Fait\_Vuel*) contient tous les vuel de la base de

données qui peuvent apparaître sur une vue quelconque. Une table *vue* est également présente dans le modèle afin de montrer comment le vuel contribue à la gestion des vues et de gérer également ces vues. Cependant, elle ne constitue pas une dimension puisqu'elle ne compose pas un vuel mais représente plutôt un agrégat de vues.

Habituellement, la table des faits ne contient que les identifiants des tables les plus détaillées, celles-ci faisant l'intermédiaire entre la table des faits et les tables plus générales ou agrégées. Toutefois, pour notre application, il est possible que certains objets ne soient associés qu'à des informations générales. Par exemple, un objet peut être associé uniquement à une classe d'objets générale (ex. *Bâtiment*). Pour cette raison, la table des faits regroupe donc les identifiants de chacune des tables peu importe leur niveau hiérarchique, ce qui permet une plus grande flexibilité lors de la combinaison des dimensions (tout en conservant explicitement les liens hiérarchiques supportant la navigation entre niveaux d'agrégation d'une dimension).

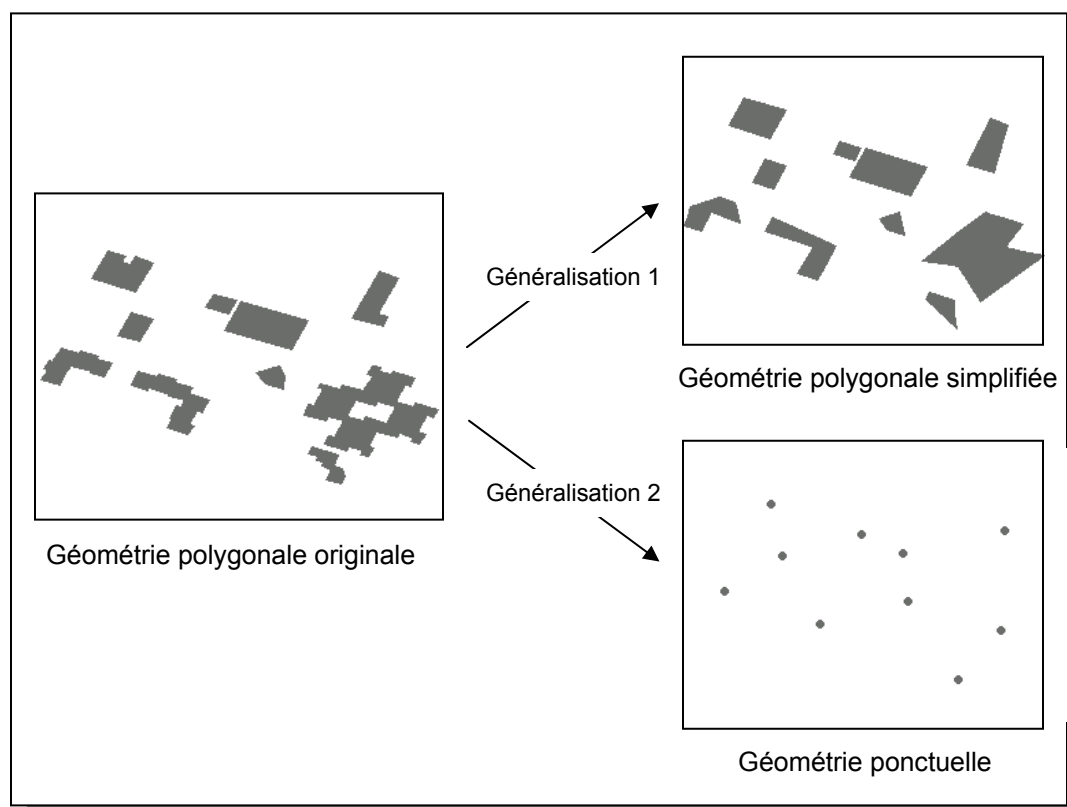
Le métamodèle de VUEL aurait également pu être implanté suivant un modèle en étoile. Celui-ci se distingue par le fait que tous les niveaux hiérarchiques d'une dimension sont conservés au sein d'une seule table. Le modèle en flocon que nous lui avons préféré est généralement utilisé lorsque le volume de données est important puisqu'il minimise le volume des tables à joindre lors de l'exécution de requêtes. Toutefois, la raison principale pour laquelle nous l'avons utilisé pour notre implantation est qu'il permet de visualiser rapidement tous les niveaux hiérarchiques associés à une dimension donnée, ce qui facilite la navigation.

### **4.3. Les données utilisées**

Les données qui furent utilisées pour notre implantation provenaient majoritairement de la base de données topographiques du Québec (BDTQ). De plus, nous nous sommes

inspirés des données du ministère des Transports du Québec (MTQ) afin de simuler certaines données routières pour mieux supporter nos différentes multiplicités.

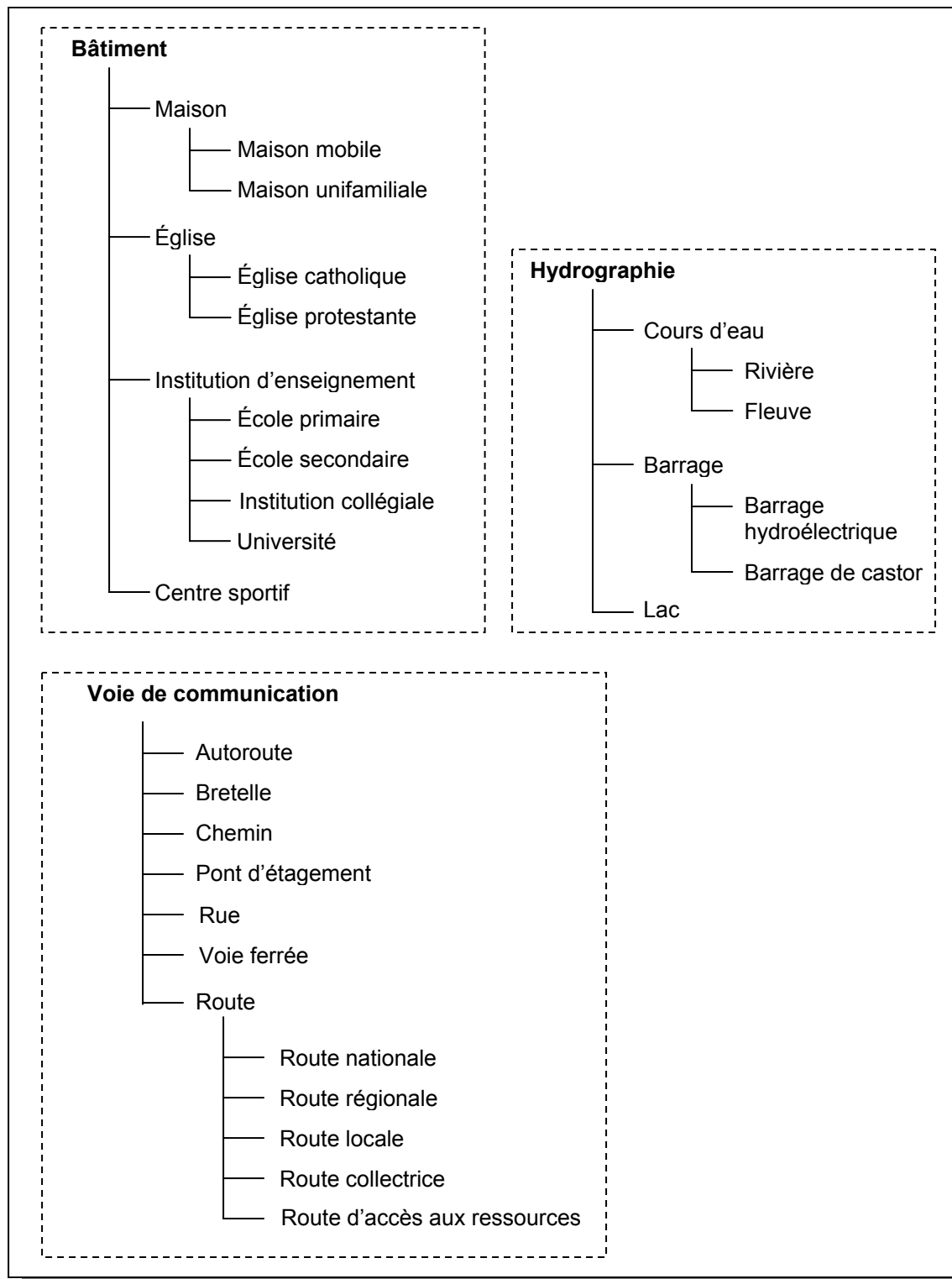
D'une part, les données de la BDTQ furent principalement utilisées afin de valider les aspects relatifs à la multiplicité géométrique. Pour les besoins de notre application, seulement les géométries concernant les voies de communication et les bâtiments ont été exploitées. Afin de disposer de données à différents niveaux de granularité, nous avons procédé à la généralisation, en prétraitement, des bâtiments. Cette généralisation s'est effectuée via le logiciel DynaGEN de la compagnie Intergraph. Ce logiciel facilite le processus de généralisation cartographique en offrant un ensemble d'outils (opérateurs de généralisation) permettant de réduire la complexité des données cartographiques. Cette généralisation peut se faire selon deux modes : automatique et interactif. En mode automatique, l'utilisateur spécifie a priori les opérateurs (ainsi que les algorithmes sous-jacents) à appliquer sur les **classes d'objets**. Ces opérateurs sont par la suite exécutés séquentiellement, selon l'ordre défini par l'utilisateur. En mode interactif, il est possible de sélectionner un opérateur et de ne l'appliquer qu'à un seul objet. Finalement, quelques opérateurs de généralisation peuvent être appliqués manuellement, ce qui permet à l'utilisateur de spécifier lui-même la géométrie ainsi que la position du nouvel objet résultant de la généralisation. À l'aide de ce logiciel, nous avons procédé à la généralisation des bâtiments possédant une géométrie polygonale. Un premier degré de généralisation a permis d'obtenir une géométrie polygonale simplifiée alors que l'application d'un plus haut degré de généralisation a conduit à l'obtention d'une géométrie ponctuelle pour ces mêmes bâtiments (figure 4-2). Ceux-ci sont donc conservés dans notre base de données selon trois niveaux géométriques.



**Figure 4-2** Les trois niveaux de granularité géométrique des bâtiments

D'autre part, afin de valider les aspects relatifs à la multiplicité sémantique (multiéchelle et uniéchelle), nous nous sommes inspirés des attributs des données routières du MTQ en plus d'en définir de nouveaux pour les bâtiments de la BDTQ. Il est important de mentionner que ces nouveaux attributs ne sont aucunement présents dans les normes de la BDTQ et n'ont été ajoutés qu'à titre d'exemple.

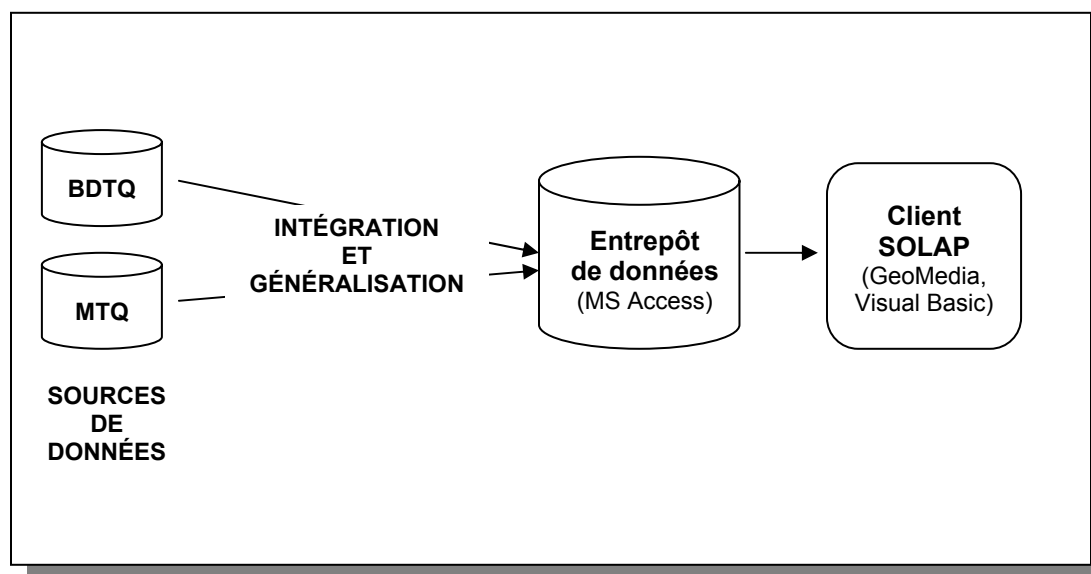
Plusieurs hiérarchies sémantiques ont donc été créées (figure 4-3), tant au niveau des classes d'objets qu'aux niveaux des attributs, des domaines de valeurs ainsi qu'au niveau des valeurs en elles-mêmes. Par exemple, le domaine de valeurs associé à l'attribut *classe fonctionnelle* des données routières se compose de valeurs générales (ex. autoroute, route) et de valeurs détaillées (ex. autoroute à deux voies, autoroutes à quatre voies, route nationale, route régionale, route collectrice).



**Figure 4-3** Exemple de hiérarchies sémantiques créées (au niveau des classes d'objets)

Également, différentes variables visuelles ont été définies pour certains objets afin de valider la multiplicité graphique uniéchelle. Ces valeurs sont gérées par le système lors de l’affichage des objets.

Finalement, l’ensemble de ces données fut stocké dans une base de données Access, structurée suivant le modèle en flocon présenté à la figure 4-1. La géométrie des objets y a été emmagasinées via le logiciel GeoMedia qui permet d’inclure des tables géométriques dans une base de données Access (figure 4-4). Celle-ci fut peuplée manuellement afin de créer les différentes multiplicités que nous voulions traiter. Également, les différentes combinaisons possibles (sémantique, géométrique et graphique) ont dû être déterminées a priori.

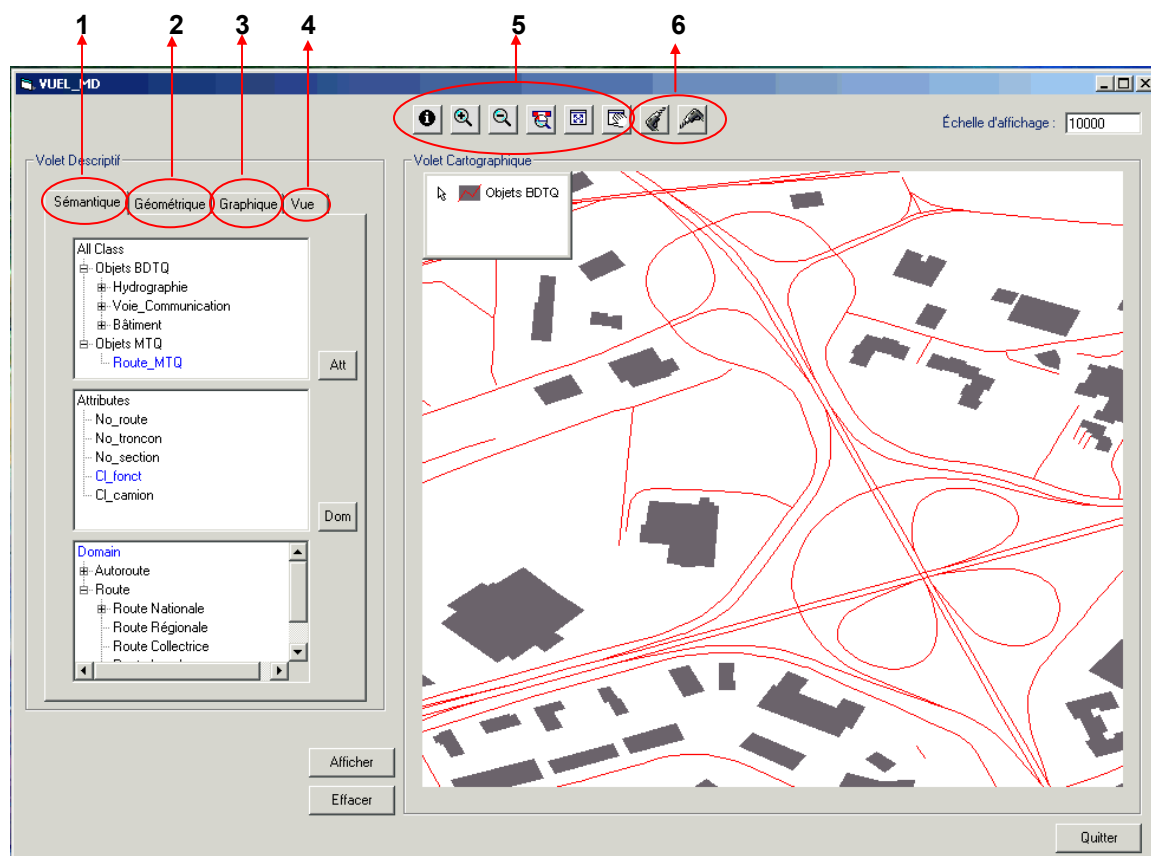


**Figure 4-4** *L'architecture système*

#### 4.4. L'application SOLAP

Dans le but de naviguer parmi les différents types de multiplicités emmagasinées au sein de la base de données relationnelle (modélisée suivant un modèle multidimensionnel), une application SOLAP a été développée. Celle-ci fut programmée en langage Visual

Basic. Les bibliothèques d'objets du logiciel GeoMedia 4 d'Intergraph ont servi au développement du volet cartographique. Les données sont accédées par l'interface SOLAP via des requêtes SQL (*Structured Query Language*) invisibles à l'utilisateur. La figure suivante présente l'interface de l'application développée.

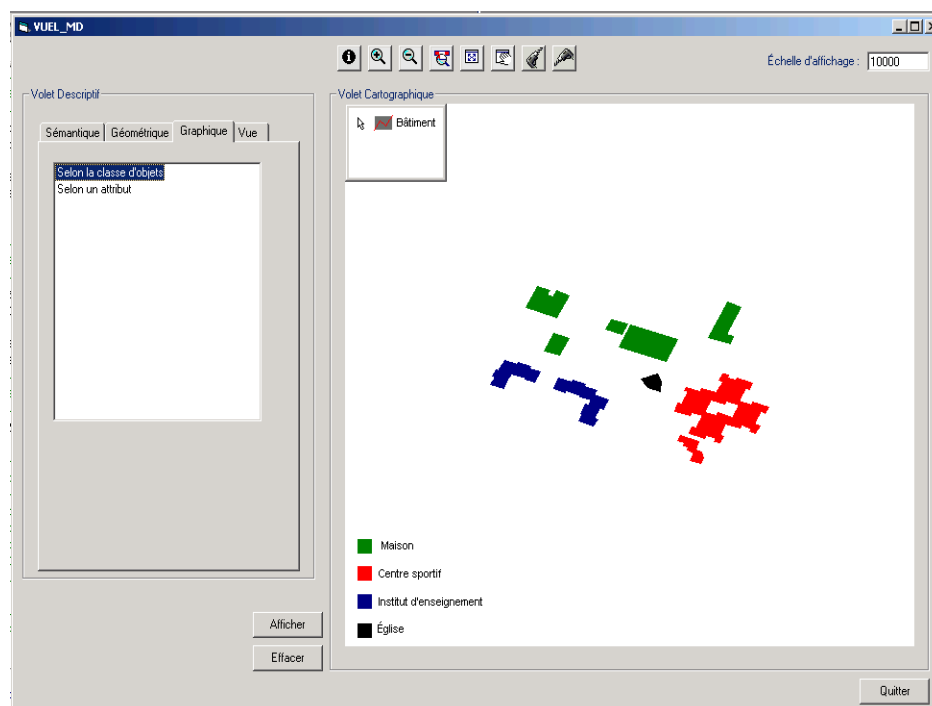


**Figure 4-5** L'interface de l'application SOLAP développée

D'une part, le volet descriptif permet la sélection des objets à visualiser. Cette sélection s'effectue premièrement au niveau des classes d'objets (partie supérieure de l'onglet sémantique (1)). Puisque celles-ci sont présentées à l'utilisateur suivant différents niveaux de granularité (visibles via un arbre expansif(+)/rétractable(-)), il est possible de sélectionner des classes d'objets générales (ex. hydrographie, voie de communication, bâtiment) ou plus spécifiques (ex. rivière, autoroute, église). Par la suite, les attributs correspondant à la classe d'objets sélectionnée peuvent être visualisés (dans la partie centrale de l'onglet sémantique) et ce, toujours selon différents niveaux de granularité (visibles via un arbre expansif(+)/rétractable(-)). Afin de raffiner sa sélection,



l'utilisateur peut également spécifier des valeurs pour certains attributs. Si ces attributs ne possèdent pas de domaine de valeurs, l'utilisateur est alors invité à entrer la valeur désirée. Dans le cas contraire, les valeurs possibles s'affichent dans la dernière fenêtre de ce volet et peuvent également avoir différents niveaux de granularité (visibles via un arbre expansif(+)/rétractable(-) ). Par exemple, l'utilisateur peut tout d'abord sélectionner la classe d'objets « Route », faire afficher les attributs correspondant à cette classe (ex. no\_route, classe\_fonctionnelle et classe\_camionnage) pour finalement visualiser le domaine de valeurs associé à l'attribut classe\_fonctionnelle. Il peut ainsi visualiser, via le volet cartographique, seulement les routes de classe fonctionnelle « régionale ». Similairement, en sélectionnant l'onglet *géométrique* (2), il est également possible de définir le niveau de granularité géométrique selon lequel les objets sélectionnés doivent être affichés. Pour notre application, trois niveaux géométriques sont disponibles et correspondent aux niveaux présentés à la figure 4-2. Via l'onglet *graphique* (3), il est possible de distinguer les éléments affichés selon leur classe d'objets ou encore selon la valeur d'un attribut grâce à l'exploitation de différentes variables visuelles (figure 4-6).



**Figure 4-6** *Différentes représentations graphiques selon la classe d'objets*

Enfin, l'onglet *vue* (4) permet à l'utilisateur de sélectionner des vues pré-enregistrées pour lesquelles l'échelle d'affichage ainsi que la géométrie et la représentation graphique des objets furent définies a priori. Il s'agit d'ailleurs là du principal onglet qui servirait à un utilisateur final alors que les trois premiers onglets serviraient surtout à créer une vue ad hoc ou permettraient à l'administrateur du système de créer des vues pour les utilisateurs du SOLAP. Il serait même possible que seul le quatrième soit disponible aux utilisateurs finaux dans une exploitation normale du SOLAP et que les trois premiers onglets ne seraient disponibles que dans une fonction avancée.

D'autre part, suite à l'affichage cartographique des objets correspondant à la sélection dans le volet descriptif, différentes fonctions de navigation sont offertes à l'utilisateur. Parmi celles-ci figurent des opérations traditionnelles de navigation spatiale (*zoom in*, *zoom out*, *pan*, etc. (5) ) mais également des opérations plus spécifiques et adaptées au SOLAP ainsi qu'aux multiplicités (6). Ces opérations ainsi que leur impact au niveau de la structure VUEL sont présentées à la section suivante.

#### **4.4.1. Fonctionnalités de navigation**

##### ***Agrandissement ou réduction de l'échelle d'affichage (zoom in, zoom out)***

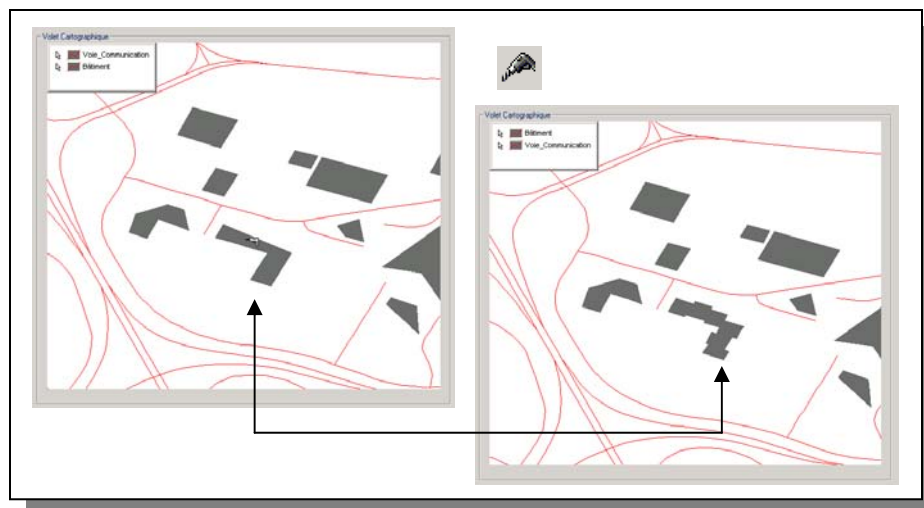
Cette fonction correspond à celle actuellement offerte par tous les SIG commerciaux. Elle n'engendre qu'un changement au niveau de l'échelle d'affichage de la vue sans égard à la lisibilité des objets. Étant donné que la représentation et la sémantique associées aux objets restent les mêmes, cette opération ne produit aucun changement de vue et par conséquent, aucun changement de vue. Il est important de se rappeler qu'un changement de vue ne survient que suite à un changement d'au moins un vue présent dans la vue (indépendamment du niveau de zoom, même si un nombre différent de vues apparaissent à différentes échelles). De plus, un changement de vue se produit uniquement lorsqu'au moins une des occurrences de relation change (géométrie, sémantique, graphique).

### ***Agrandissement ou réduction « intelligente » de l'échelle d'affichage***

Cette fonction est disponible dans certains SIG commerciaux (ex. GeoMedia) par le biais de l'opération « zoom to scale ». Cette fonction requiert la définition au préalable des échelles (minimale et maximale) d'affichage des classes d'objets. Ainsi, lorsque l'utilisateur agrandit ou réduit l'échelle d'affichage, certaines classes d'objets peuvent disparaître alors que d'autres peuvent apparaître. Par exemple, pour notre application, il est possible de spécifier certains intervalles d'échelles selon lesquelles chaque niveau géométrique doit être employé pour représenter les objets. Par exemple, il pourrait être décidé qu'au-delà d'une certaine échelle (ex. 1/100 000), les bâtiments doivent être représentés avec la géométrie ponctuelle. Ceci permet de conserver une certaine lisibilité lors d'opérations de réduction ou d'agrandissement de l'échelle d'affichage. Il est possible que cette fonction n'engendre aucun changement de vue. En effet, si le facteur d'agrandissement ou de réduction de l'échelle d'affichage est minime, il se peut qu'aucun vuel ne soit affecté et par conséquent, qu'aucun changement de vue ne soit effectué. Par contre, un changement significatif provoquera probablement un changement de vue. Cette fonction ne considère que l'aspect géométrique des objets. Par conséquent, elle ne tire profit que partiellement de toute la richesse qu'offre une structure de type VUEL.

### ***Forage et remontage spatiaux***

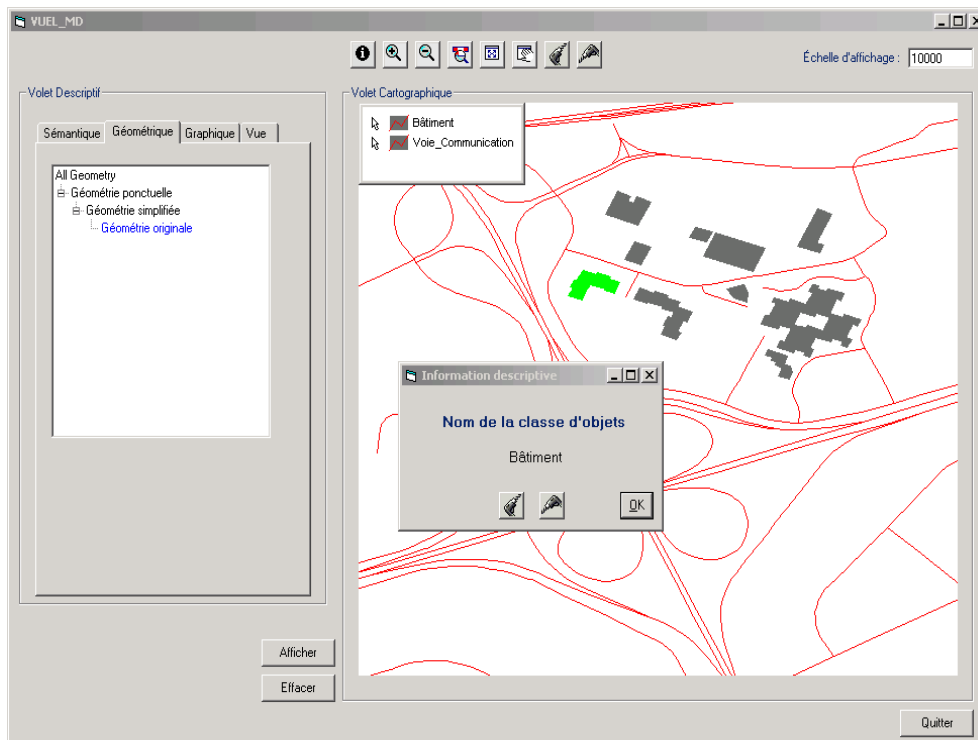
Ces opérations, typiquement SOLAP, permettent de naviguer parmi les différentes géométries associées à un même objet. Suite à la sélection d'un objet dans le volet cartographique, il est donc possible de le visualiser avec une géométrie plus détaillée (forage spatial) ou plus simplifiée (remontage spatial). De manière similaire, ces opérations peuvent être effectuées pour une vue entière. Ainsi, il est possible d'afficher tous les objets d'une même classe présents dans une vue selon un niveau géométrique supérieur ou inférieur. L'application de ces opérations engendre automatiquement un changement de vue puisque l'occurrence de relation géométrique d'un vuel est modifiée. La figure suivante illustre une opération de forage effectuée sur un seul objet d'une vue cartographique et permet de visualiser cet objet avec une géométrie plus détaillée.



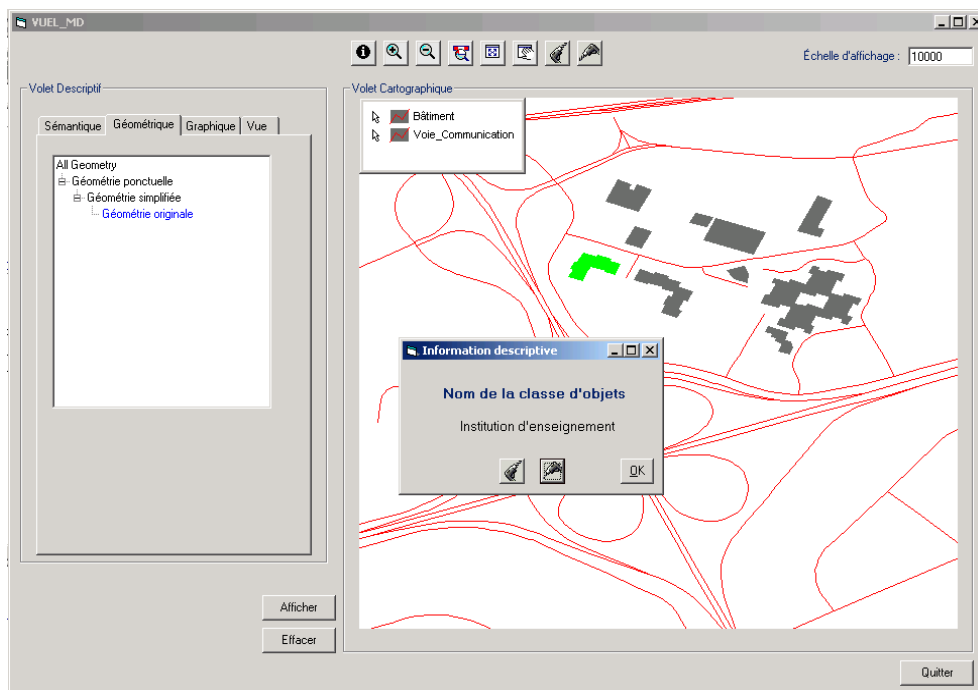
**Figure 4-7** *Forage spatial sur un seul objet*

### ***Forage et remontage descriptifs***

À l'instar des fonctions précédentes, les opérations de forage et de remontage descriptifs permettent de naviguer parmi les différents niveaux sémantiques associés à un vœl. Suite à l'application de ces opérations, un objet est donc défini de manière plus générale (remontage descriptif) ou plus précise (forage descriptif). Théoriquement, le changement de sémantique d'un objet engendre un changement de vœl, et par conséquent un changement de vue, bien qu'il ne soit pas perceptible par l'utilisateur car la représentation cartographique de l'objet n'est pas modifiée. Les deux figures qui suivent illustrent une opération de forage descriptif sur un objet sélectionné dans le volet cartographique. La première figure indique la classe générale associée à l'objet alors que la seconde, suite au forage descriptif, indique la classe détaillée associée à ce même objet.



**Figure 4-8** Affichage de la classe générale associée à l'objet sélectionné



**Figure 4-9** Affichage de la classe détaillée associée à l'objet sélectionné, suite à une opération de forage descriptif

Bien qu'elles ne furent pas implantées dans le cadre de cette recherche, il sera intéressant d'envisager deux autres opérations de navigation dans le cadre de travaux futurs, soient : la *généralisation* et la *spécialisation* qui en fait vont exploiter cette fonctionnalité de la structure VUEL ainsi que la richesse sémantique de la structure de données du domaine d'application (i.e. les relations de généralisation/spécialisation, ainsi que les relations d'agrégation du modèle conceptuel de classes d'objets de l'application développée).

### ***Généralisation et spécialisation***

Alors que les fonctions précédentes n'avaient de répercussion que dans un volet à la fois (sémantique **ou** géométrique), ces deux nouvelles opérations permettraient de naviguer au même moment dans la dimension géométrique **et** dans la dimension sémantique. En fait, elles combindraient les opérations spatiales et descriptives en une seule opération. La fonction de généralisation permettrait de remonter dans les dimensions géométrique et descriptive associées à un objet. En fait, suite à l'opération, le même objet serait représenté avec une géométrie plus simplifiée et serait associé à une sémantique plus générale. Au contraire, l'opération de spécialisation permettrait de descendre dans ces deux mêmes dimensions. Ces deux opérations pourraient également être effectuées pour une vue entière. Il va sans dire que ces opérations induiraient un changement de vue étant donné que deux occurrences de relation seraient modifiées (géométrique et sémantique). Il s'agit en réalité d'opérations synchrones (la géométrie et la sémantique changent en même temps) alors que les opérations définies précédemment étaient de nature asynchrone (où la géométrie et la sémantique ne changent pas en même temps). Ces opérations pourraient être d'une très grande utilité étant donné que pour plusieurs applications, ces deux aspects devraient varier au même moment (par choix). Par exemple, il se pourrait que les besoins d'un client relient toujours les changements de sémantique avec les changements géométriques (et peut-être même graphiques), ce qui aurait un impact sur l'optimisation de l'implantation. Celle-ci pourrait alors ne pas supporter la flexibilité liée à l'indépendance des trois aspects (sémantique, géométrique et graphique).

## 4.5. Conclusion

Ce chapitre a présenté l'implantation du concept de VUEL (i.e. son métamodèle) sur une plate-forme relationnelle mais structurée de façon multidimensionnelle, grâce à l'utilisation d'un modèle d'implantation de type flocon. Également, l'interface SOLAP, dédiée à l'exploitation des différentes multiplicités, qui a été développée dans le cadre de cette recherche fut également présentée. Celle-ci permet de naviguer facilement et rapidement parmi les différentes représentations géométriques, graphiques ainsi qu'à travers les différentes sémantiques d'un ou plusieurs objets. Finalement, la conclusion de ces travaux fait l'objet du chapitre suivant.

## **Chapitre 5**

### **Conclusion**

Étant donné l'efficacité des outils OLAP pour l'analyse des bases de données traditionnelles, plusieurs efforts ont été entrepris afin d'adapter ces outils au domaine spatial. Depuis, les outils SOLAP sont maintenant en émergence et permettent à un utilisateur non-familier avec les technologies SIG, d'explorer intuitivement et rapidement une base de données à référence spatiale. Malheureusement, à l'heure actuelle, les structures multidimensionnelles sous-jacentes à ce type d'outils ne permettent pas de représenter et de définir différemment un même objet. Ainsi, chaque objet ne possède qu'une seule représentation géométrique et graphique et n'est associé qu'à une seule sémantique. Or, afin de répondre aux différents besoins des utilisateurs, un tel système devrait permettre la description d'une même réalité sous différents points de vue et à différents niveaux d'abstraction.

Afin de combler cette lacune, la généralisation automatique et la représentation multiple constituent deux approches pouvant être envisagées. Suite à une revue de la littérature, la première approche, la généralisation automatique, s'est vite avérée inappropriée pour un contexte SOLAP. Du fait qu'elle nécessite souvent une intervention humaine lors du processus en plus de générer des temps de réponse relativement longs, elle ne peut répondre aux contraintes de rapidité liées à un contexte SOLAP.



Par ailleurs, l'étude des structures actuelles permettant la gestion des représentations multiples a permis de constater un manque de flexibilité au niveau de ces structures empêchant ainsi leur utilisation dans un contexte SOLAP. Principalement multiéchelles, elles permettent de gérer les différentes géométries d'un même objet uniquement en fonction de l'échelle. Également, les structures existantes ne gèrent généralement que l'aspect géométrique des données sans considérer les aspects sémantique et graphique qui sont essentiels pour un outil SOLAP.

Parallèlement à ceci, l'émergence du nouveau concept de VUEL au sein de l'équipe de Yvan Bédard semblait être une solution prometteuse afin de gérer les représentations multiples dans un contexte SOLAP. Le **principal objectif** de cette recherche consistait donc à enrichir le concept initial de VUEL en relation avec le concept de la représentation multiple et tester ce concept dans un contexte SOLAP. Pour atteindre cet objectif, nous avons premièrement étudié la version initiale du concept de VUEL [Bédard *et al.*, 2000] et jeté les bases des concepts théoriques liés au VUEL. Par la suite, le développement d'un métamodèle a permis d'enrichir ces concepts théoriques en plus de mieux structurer les différentes composantes du VUEL. Le métamodèle développé permet aussi bien la gestion des multiplicités géométriques que sémantiques et graphiques, ce qui constitue une innovation importante. Également, ce métamodèle est destiné à gérer autant les multiplicités de type multiéchelle que de type uniéchelle.

Ensuite, afin de valider ces concepts théoriques et vérifier la viabilité du métamodèle sur une technologie relationnelle existante, nous avons procédé à son implantation sur une plate-forme Microsoft Access. Toutefois, étant donné que cette base de données était destinée à des fins d'analyse dans un contexte SOLAP, il a été nécessaire de simuler une structure multidimensionnelle par l'exploitation d'un modèle de données particulier (de type flocon). Les données provenant de la BDTQ ont servi au peuplement de notre base de données. Finalement, le développement d'un prototype SOLAP permettant la génération de vues alternatives (cartographiques et sémantiques) et basé sur une structure de type VUEL a démontré le potentiel de ce concept à supporter les changements de vues

et les différentes multiplicités pouvant être associées à un même objet. En définitive, le concept de VUEL tel que développé dans le cadre de cette recherche est une nouvelle façon de gérer des vues cartographiques et sémantiques et surtout, constitue une toute nouvelle manière d'aborder la représentation multiple.

En terminant, il est possible d'identifier certains éléments qui pourraient être développés ultérieurement afin d'enrichir le concept de VUEL. Premièrement, il serait intéressant d'intégrer l'aspect temporel au métamodèle du VUEL. Une des solutions possibles serait d'ajouter des classes associatives à chacune des relations composant un VUEL et indiquant ainsi sa période de validité. Également, en raison des contraintes de temps, seulement les éléments cartographiques du métamodèle furent implantés dans le cadre de cette recherche. Des travaux futurs pourraient donc se concentrer sur l'implantation des éléments typiquement SOLAP tels que les tableaux et les graphiques statistiques (le modèle conceptuel le prévoyant déjà). De plus, d'autres fonctionnalités de navigation pourraient être développées pour le prototype SOLAP telles que les fonctions de généralisation/spécialisation qui furent discutées dans le cadre du chapitre quatre.

Par ailleurs, il semble évident que lorsque la généralisation automatique (aux niveaux sémantique, cartographique et graphique) engendrera des temps de réponse satisfaisants pour un contexte SOLAP, son intégration au sein d'un tel système permettra de limiter le stockage des données. En effet, les données qui pourront être rapidement généralisées à la volée pourront ne plus être emmagasinées explicitement dans la base de données, ce qui en diminuera considérablement le volume.

Enfin, des travaux sont actuellement en cours au Centre de recherche en géomatique de l'Université Laval afin d'implanter le métamodèle du VUEL selon une architecture MOLAP, donc directement dans une structure multidimensionnelle propriétaire (cube OLAP). Les résultats obtenus jusqu'alors semblent être très satisfaisants.

## Références

- Allouche, M. et B., Moulin, 2001, Reconnaissance de patterns par réseaux de neurones: application à la généralisation cartographique, soumis à *Revue internationale de géomatique* (accepté mai 2001).
- Badard, T., 2000, *Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques*, Thèse de doctorat en Informatique de l'Université de Marne-la-Vallée, spécialité : sciences de l'information géographique, soutenue le 15 décembre 2000 à Saint-Mandé.
- Beard, K. & W., Mackaness, 1991, Generalization operations and supporting structures, *Proceedings of the International Symposium on Computer-Assisted Cartography (Auto-Carto 10)*, pp. 29-45
- Bédard, Y. 1997, Spatial OLAP, Vidéoconférence. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 novembre, Montréal.
- Bédard, Y., 1999, Visual Modeling of Spatial Databases Towards Spatial Extensions and UML, *Geomatica*, Vol 53, No2, pp. 169-186
- Bédard, Y., S., Larrivée, M.-J., Proulx, P.-Y., Caron & F., Létourneau, 1997, *Geospatial Data Warehousing: Positionnement technologique et stratégique*, Rapport préparé pour le Centre de Recherche de la Défense de Valcartier, Université Laval, 79p.
- Bédard, Y., R., Devillers, E., Bernier & B., Moulin, 2000a, Automatic Generalisation or Multiple Representations: Towards a combination of both approaches, *GEOIDE DEC#9 Workshop on feature extraction and cartographic generalization*, Université Laval, Québec, 17 novembre.
- Bédard, Y., T. Merrett and J. Han, 2001, Fundamentals of spatial data warehousing for geographic knowledge discovery. In: *Geographic Data Mining and Knowledge Discovery*. Edited by H. Miller and J. Han, Taylor & Francis.
- Bourgon, J.-F., J., Brodeur, F., Lemieux, P., Marchand & S., Rivest, 2000, *Fondements théoriques et pratiques des serveurs universels pour l'exploration multidimensionnelle des très grandes bases de données*, Rapport de cours (SCG-62058), Département des sciences géomatiques, Université Laval, 67p.

- Bruegger, B.P. & A.U., Franck, 1989, Hierarchies over Topological Data Structures, in: *International Archives ASPRS/ACSM Annual Convention, AGIS/LIS*, Baltimore, Vol. 4, pp. 137-145.
- Buttenfield, B. P. & J.S., Delotto, (éditeurs), 1989, *Multiple representations : Scientific Report for the Specialist Meeting*, National Center for Geographic Information and Analysis (NCGIA), Technical Paper 89-3, 87p.
- Buttenfield, B. P. & R. B., McMaster, (éditeurs), 1991, *Map generalization : Making rules for knowledge representation*, Longman, 245p.
- Caron, P.-Y., 1998, *Étude du potentiel Olap pour supporter l'analyse spatio-temporelle*, Mémoire de M.Sc., Faculté de foresterie et de géomatique, Université Laval, 132p.
- Claramunt, C. & M., Mainguenaud, 1995, Spatial View: A dynamic and flexible vision of GIS database, *Proceedings of the DEXA International Conference and Workshop on Database and Expert System Applications*, Revell, N. and Min Tjoa, A. eds, Omnipress, London, UK, pp. 483-493.
- Codd, E.F., S.B., Codd & C.T., Salley, 1993, *Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate*, Hyperion white papers, <http://www.hyperion.com> , 20 pp.
- Devillers, R., 2000, *Patrons d'implantation de la représentation géométrique multiple dans les SIG*, Essai de M.Sc., Faculté de foresterie et de géomatique, Université Laval, 45 p.
- Devogele, T., 1997, *Processus d'intégration et d'appariement de bases de données géographiques: Application à une base de données routières multi-échelles*, Thèse de doctorat de Méthodes Informatiques, Université de Versailles, 207 p.
- Devogele, T., J., Trevisan & L., Raynal, 1996, Processus de constitution d'une base de données multi-échelle, *Revue Internationale de géomatique*, Vol. 6, No.2-3, pp. 249-263.
- Dikau, R., 1990, Geomorphic landform modeling based on hierarchy theory, *Proceedings of the 4<sup>th</sup> International Symposium on Spatial Data Handling*, Zürich, pp. 230-239
- Douglas, D. & T., Peucker, 1973, Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature, *The Canadian cartographer*, Vol. 10, pp. 112-122.
- Egenhofer, Max J., E., Clementini & P.Di, Felice, 1994, Evaluating inconsistencies among multiple representations, *Proceedings of the 6<sup>th</sup> International Symposium on Spatial Data Handling*, Taylor & Francis, Londres, pp. 901-920

- Francalanci, C. & B., Pernici, 1994, *Abstraction levels for entity-relationship schemas*, In Loucopoulos, pp. 456-473
- Han, J., N., Stefanovic & K., Koperski, 1998, Selective materialization: An Efficient Method for Spatial Data Cube Construction, *Proceedings of the 1998 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'98)*, Melbourne, Australia, pp. 144-158
- Inmon, W.H., 1996, *Building the Data Warehouse*, Second Edition, John Wiley & Sons, 410 p.
- João, E.M., 1995, *The importance of quantifying the effects of generalization*, dans: Müller, J.C., J.P., Lagrange, R., Weibel, (éditeurs), *GIS and Generalization: Methodology and Practice*, Taylor & Francis, Londres, pp. 183-193
- Jones, C.B., 1991, Database architecture for multi-scale GIS, *Proceedings Auto-Carto 10*, Baltimore, ACSM/ASPRS, pp. 1-14
- Jones, C.B. & I.M., Abraham, 1986, Design considerations for a scale-independent cartographic database, *Proceedings of the 2<sup>nd</sup> International Symposium on Spatial Data Handling*, Seattle (WA).
- Kehoe, C., J., Pitkow, K., Sutton, G., Aggarwal & J.D., Rogers, 1999, *Results of Gvu's Tenth World Wide Web User Survey*, Graphics Visualization and Usability Center [http://www.cc.gatech.edu/gvu/user\\_surveys/survey-1998-10/tenthreport.html](http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/tenthreport.html) (accedé le 01/09/2001)
- Kidner, D. B. & C.B., Jones, 1994. A deductive object-oriented GIS for handling multiple representations, *Advances in GIS Research, Proceedings of the 6<sup>th</sup> International Symposium on Spatial Data Handling*, Taylor & Francis, Londres, pp. 882-900
- Kilpeläinen, T., 1992, Multiple Representations and Knowledge-based Generalization of Topographical Data, *Int. Arch. of Photogrammetry and Remote Sensing, Commission III*, part B3, Washington, D.C., pp. 954-964.
- Kilpeläinen, T., 1994, Updating multiple representation geodata bases by incremental generalization, *Int. Arch. of Photogrammetry and Remote Sensing, Commission III /IV*, Munich, Vol. 30, part 3/1, pp. 440-447.
- Kilpeläinen, T., 1995, Requirements of a multiple representation database for a topographical data with emphasis on incremental generalization, *Proceedings of 17<sup>th</sup> International cartographic conference*, Barcelona, Vol. 2, pp. 1815-1825.

- Kilpeläinen, T., 1997, *Multiple Representation and Generalization of Geo-Databases for Topographic Maps*, Thèse de doctorat, Publications of the Finnish Geodetic Institute, No. 124, 229 p.
- Kilpeläinen, T., 2000, Maintenance of Multiple Representation Databases for Topographic Data, *The Cartographic Journal*, Vol. 37, No. 2, pp. 101-107
- Kraak, M.-J. & A., Brown, (éditeurs), 2001, *Web Cartography : developments and prospects*, Taylor & Francis, Londres, 207 p.
- Kubik, K.K. & G.M., Brooke, 1992, An intelligent interface for computer supported map generalization, *ASPRS/ACSM/RT 92 technical papers*, Vol. 3, GIS and Cartography, pp. 265-274
- Marchand, P., Y., Bédard & G., Edwards, 2001, A hypercube-based method for spatio-temporal exploration and analysis, *Geoinformatica*, soumis.
- Martel, C., 1997, *Analyse des capacités et limites de MGE/Map Generalizer pour la généralisation des cartes routières*, Rapport préparé dans le cadre du cours Projet de levés intégrés, Faculté de foresterie et de géomatique, Université Laval, 24p.
- Martel, C., 1999, *Développement d'un cadre théorique pour la gestion des représentations multiples dans les bases de données spatiales*, Mémoire de M.Sc., Faculté de foresterie et de géomatique, Université Laval, 128p.
- McMaster, R.B., 1991, Knowledge acquisition for cartographic generalization, *Proceedings of the 15<sup>th</sup> Conference of the International Cartographic Association*, Bournemouth, UK, 24 p.
- McMaster, R.B., K.S., Shea, 1988, Cartographic generalization in a digital environment: A framework for implementation in a geographic information system, *Proceedings of GIS/LIS'88*, San Antonio, Texas, pp. 240-249
- Microsoft Corporation, 2000, *Microsoft SQL Server 2000 documentation*.
- Müller, J.C., 1990, Rule-Based Generalization: Potentials and Impediments, *Proceedings of the 4<sup>th</sup> International Symposium on Spatial Data Handling*, Zürich, Vol. 1, pp. 317-334
- Müller, J.C., R., Weibel, J.P., Lagrange & F., Salgé, 1995, *Generalization : state of the art and issues*, dans Müller, J.C., J.P., Lagrange & R., Weibel, (éditeurs), *GIS and Generalization: Methodology and Practice*, Taylor & Francis, Londres, pp. 3-17
- National Center for Geographic Information and Analysis, NCGIA, 1989. The research plan for the National Center of Geographic Information and Analysis, In: *Int. Journal of Geographical Information Systems*, Vol. 3, No. 2, pp.117-136

- OLF, 2001, *Le grand dictionnaire terminologique*, version 1.1, [en ligne], Office de la langue Française, [<http://www.granddictionnaire.com>], (accédé le 1 décembre 2001)
- Peterson, M.P., 1999, Trends in Internet map use, A second look. In Touch the Past, Visualize the Future, *Proceedings 19<sup>th</sup> International Cartographic Conference*, Ottawa, Canada, edited by Keller, C.P., Section 5 : Capitalizing on new technologies (Ottawa : Organizing Committee for Ottawa ICA 1999) CD-ROM
- Plazanet, C., 1995, Modélisation de la géométrie des objets linéaires en vue de leur généralisation automatique, *Proceedings ICC'95 International Cartographic Conference*, Barcelone, Vol.1, pp. 486-495
- Ramer, U., 1972, An interactive procedure for the polygonal approximation of plane curve, *Computer Graphics and Image Processing* 1, pp. 244-256
- Richardson, D.E., J.-C., Müller, 1995, *Rule selection for small-scale map generalization*, dans Müller, J.C., J.P., Lagrange & R., Weibel, (éditeurs), *GIS and Generalization: Methodology and Practice*, Taylor & Francis, Londres, pp. 136-149
- Rigaux, P., 1994, La représentation multiple dans les systèmes d'information géographique, *Revue Internationale de géomatique*, Vol.4, No.2, pp.137-164
- Rivest, S., 2000, *Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG*, Essai de M.Sc., Faculté de foresterie et de géomatique, Université Laval, 84 p.
- Rivest, S., Y., Bédard, & P., Marchand, 2001, Towards better support for spatial decision-making: Defining the characteristics of Spatial On-Line Analytical Processing (SOLAP), *Geomatica, the journal of the Canadian Institute of Geomatics*, Vol. 55, no. 4, pp. 539-555
- Scholl, M., A., Voisard, J.-P., Peloux, L., Raynal & P., Rigaux, 1996, *SGBD Géographiques: spécificités*, International Thomson Publishing, France, 185p.
- Shea, K.S., R.B., McMaster, 1989, Cartographic generalization in a Digital Environment: When and how to generalize, Auto-Carto 9, *Proceedings of the 9<sup>th</sup> International Symposium on Computer-Assisted Cartography*, Baltimore, pp. 56-67
- Stefanovic, N., 1997, *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*, M.Sc. thesis, Computing Science, Simon Fraser University.
- Stefanovic N., J. Han, and K. Koperski, 2000, Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes, *IEEE Transactions on Knowledge and Data Engineering*, 12(6): 938-958.

- Thomsen, E., 1997, OLAP solutions : building multidimensional information systems. John Wiley & Sons, 608p.
- Thomsen, E., G., Spofford, D., Chase, 1999, Microsoft OLAP Solutions, John Wiley & Sons, 509 p.
- Timpf, S., 1998, *Hierarchical structures in map series*, Thèse de doctorat, Technical University of Vienna, Departement of Geoinformation, <http://www.geoinfo.tuwien.ac.at/publications/timpf>
- Timpf, S. & A., Frank, 1995. A multi-scale data structure for cartographic objects, *Proceedings of ICC'95* (ICA, ed.), Barcelona, Vol.1, pp.1389-1396
- Timpf, S. & A., Frank, 1997, *Using hierarchical spatial data structures for hierarchical spatial reasoning*, In Spatial Information Theory - A Theoretical Basis for GIS (Int. Conference COSIT'97), edited by Hirtle, Stephen C. and Frank, Andrew U. Lecture Notes in Computer Science Vol.1329. Berlin-Heidelberg: Springer-Verlag, pp. 69-83
- Tobler, W.R., 1964, *Numerical Map Generalization*, Michigan Inter-University Community of Mathematical Geographers Discussion, Paper No. 8, Ann Arbor, Michigan.
- van Oosterom, P., 1989, A reactive data structure for geographic information systems, *Proceedings Auto-Carto 9*, Baltimore, pp. 665-674
- van Oosterom, P., 1995, *The GAP-tree: an approach to 'on-the-fly' map generalization of an area partitioning*, dans : Muller, J.-C., J.-P., Lagrange & R., Weibel (éditeurs), GIS and generalization, Taylor & Francis, Londres, pp. 120-132
- van Oosterom, P. & V., Schenkelaars, 1993, The design and implementation of a multi-scale GIS, *Proceedings of EGIS'93*, Italie, pp. 712-722
- van Oosterom, P. & V., Schenkelaars, 1996, *Applying reactive data structure in an interactive multi-scale GIS*, dans : Molenaar, Martien (éditeur), Methods for the generalization of geo-databases, Netherlands geodetic commission, No. 43, pp. 37-56
- Vangenot, C., 1998, Représentation multi-résolution, Concepts pour la description de bases de données avec multi-représentation, *Revue Internationale de géomatique*, Vol.8, No.1-2, pp.121-147
- Weibel, R., 1991, Specification for a platform to support research in map generalization, *Proceedings of the 15<sup>th</sup> Conference of the International Cartographic Association*, Bournemouth, UK, 12 p.



Weibel, R., 1995, *Three Essential Building Blocks for Automated Generalization*, dans Müller, J.C., J.P., Lagrange & R., Weibel, (éditeurs), *GIS and Generalization: Methodology and Practice*, Taylor & Francis, Londres, pp. 56-69

Weibel, R. & G., Dutton, 1999, *Generalising spatial data and dealing with multiple representations*, dans : Longley, P., A., M., F. Goodchild, D., J. Maguire & D., W. Rhind (éditeurs), *Geographical Information Systems: Principles and Technical Issues*, Second Edition, Vol. 1, Wiley, pp. 125-155

## Annexe A

### **Combinaisons possibles entre la généralisation automatique et la représentation multiple.**

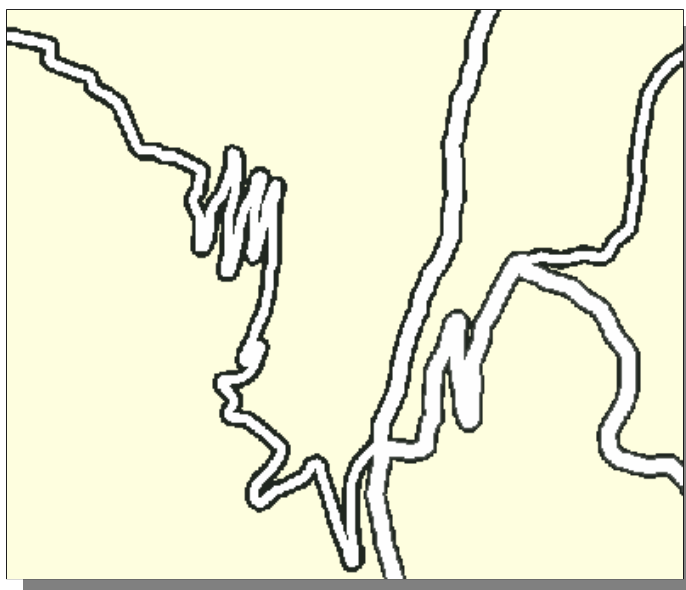
Une des caractéristiques essentielles d'un outil SOLAP est qu'il permet une analyse spatio-temporelle **rapide** des données. Ainsi, un utilisateur pourra naviguer au sein d'une ou de plusieurs dimensions tout en conservant son schéma de pensée. Cette contrainte de rapidité est similaire à celle régissant l'affichage de documents sur le web, où ceux-ci doivent être présentés à l'utilisateur dans un délai minimal (environ 8 à 10 secondes) [Marchand *et al.*, 2001]. Suite à ce qui fut présenté à la section 2.2, il semble évident que la généralisation automatique actuelle ne peut, à elle seule, permettre l'affichage de vues alternatives de façon quasi-instantanée.

Face à ce constat, il nous apparaît logique d'envisager une certaine combinaison de la généralisation automatique et de la représentation multiple afin de palier aux inconvénients de chacune et de tirer profit de leurs avantages respectifs. À l'heure actuelle, la littérature ne comporte peu ou pas d'écrits à ce propos. La plupart des auteurs expriment leurs idées concernant l'une ou l'autre des approches mais peu d'entre eux discutent d'une combinaison potentielle de la généralisation automatique et de la représentation multiple. En réalité, les seuls ayant eu des visées en ce sens ne font que souligner le fait que la généralisation, automatique ou non, peut servir à la création des différentes représentations d'un même objet [Beard & Mackaness, 1991; Vangenot, 1998; Weibel & Dutton, 1999]. Toutefois, il est de notre avis qu'une combinaison beaucoup

plus formelle pourrait être exploitée afin d'améliorer et de faciliter le processus de généralisation.

L'utilisation de patrons de représentation multiple au sein du processus même de généralisation automatique pourrait non seulement le simplifier mais également l'accélérer. Ces patrons correspondraient aux différentes représentations possibles d'un même objet et pourraient avoir été créés lors de la saisie initiale des objets en vue de leur stockage dans une base de données spatiales. Par exemple, le cartographe devant numériser un bâtiment lors de la saisie des données, pourrait le faire de plusieurs façons. Une première numérisation conforme à la représentation détaillée du bâtiment suivie d'autres numérisations plus ou moins simplifiées du même bâtiment. Ainsi, dès la saisie initiale des données, il serait possible d'associer plusieurs géométries, plus ou moins complexes, à un même objet en vue de les réutiliser lors du processus de généralisation automatique.

La généralisation automatique appliquée à des éléments linéaires tels que les réseaux routiers ou hydriques engendre généralement de bons résultats. En effet, étant donné que quatre-vingt pourcent (80%) des objets cartographiques sont linéaires, la généralisation de lignes fut amplement étudiée et par conséquent, les algorithmes concernant ce genre d'objets sont relativement efficaces [Müller *et al.*, 1995]. Toutefois, dès que ces réseaux possèdent des éléments plus complexes, la qualité du résultat se détériore rapidement. Prenons l'exemple d'une route qui se compose d'une partie très sinueuse. Lors d'une réduction d'échelle, un empâtement des virages est quasi inévitable, ce qui engendre une piètre représentation de la route (figure 1). Afin de corriger la situation, il est nécessaire d'appliquer certains opérateurs de généralisation. Toutefois, étant donné la complexité de ces opérateurs et le fait qu'il s'agisse d'un élément en composant un autre, il est excessivement difficile, voire impossible, d'effectuer cette généralisation rapidement en mode automatique.

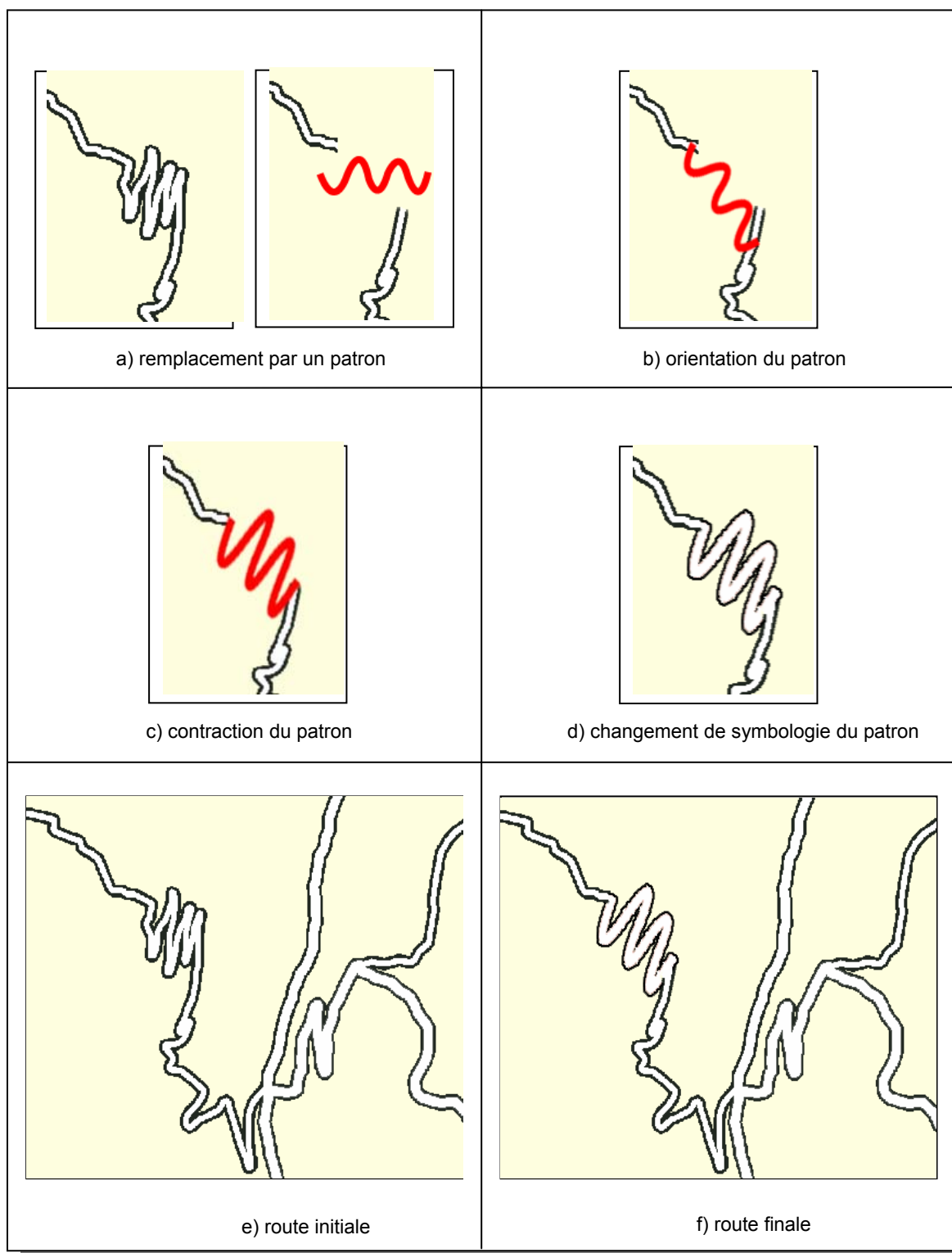


**Figure 1** *Route conflictuelle*<sup>1</sup>

En revanche, l'utilisation de patrons de représentation multiple permettrait de faciliter grandement ce processus. Ainsi, au lieu de généraliser la partie de route sinueuse, elle pourrait tout simplement être remplacée par un patron, préalablement enregistré dans la base de données, correspondant à ce type de géométrie. Évidemment, quelques manipulations (orientation, étirement, etc.) devront être effectuées sur le patron dans le but de conserver la cohérence de la route. Toutefois, il s'agit d'opérations très simples qui peuvent être effectuées rapidement (figure 2).

---

<sup>1</sup> © IGN



**Figure 2** *Utilisation de patrons de représentation multiple*

Toutefois, ce genre de processus nécessite une certaine structuration au niveau des données. De plus, l'ajout de données de nature sémantique est nécessaire afin de conserver la simplicité du procédé. Une structure de données à l'instar de celle actuellement exploitée au Ministère des transports du Québec (MTQ) est à privilégier. Les données sont emmagasinées en tant que « segments atomiques ». Un segment atomique représente la plus petite unité uniforme du réseau routier. Plusieurs critères, tels que le type de pente et le degré de courbure sont utilisés afin de segmenter le réseau routier du Québec. Pour chacun des segments, il est possible de connaître ses coordonnées d'origine et de destination. Ainsi, il est très facile de localiser un élément potentiellement problématique en vue d'une généralisation automatique et de le remplacer par un patron de représentation multiple à la position exacte.

La combinaison telle que discutée au paragraphe précédent se traduit par un couplage vertical entre la généralisation automatique et la représentation multiple ; cette dernière étant utilisée dans le cadre du processus même de la généralisation automatique. Un deuxième couplage, de type horizontal, peut également être envisagé. Selon ce couplage, les deux approches pourraient être utilisées parallèlement dans le but de produire des vues cartographiques. Ainsi, les thèmes simples pourraient faire l'objet d'une généralisation automatique alors que les thèmes plus complexes seraient emmagasinés sous la forme de représentations multiples.

## Annexe B

### Métamodèle original du concept de vuel

[Bédard *et al.*, 2000]

