

# Un langage et un générateur pour représenter les résumés visuels de bases de données géographiques

Ibtissem Cherni<sup>1,2</sup>, Karla Lopez<sup>1</sup>, Robert Laurini<sup>1</sup>, Sami Faiz<sup>2</sup>

<sup>1</sup> LIRIS – INSA de Lyon  
69621 – Villeurbanne Cedex – France  
{Ibtissem.Cherni, Karla.Lopez, Robert.Laurini}@insa-lyon.fr

<sup>2</sup> Faculté des Sciences Juridique, Economique et de Gestion de Jendouba  
Avenue de l'UMA - 8189 Jendouba – Tunisie  
Sami.faiz@insat.rnu.tn

## Résumé

*Les chorèmes sont des représentations schématisées du territoire. Ils permettent de représenter visuellement des connaissances géographiques alors que jusqu'à présent les chorèmes avaient été conçus manuellement par les géographes en utilisant leurs propres connaissances mentales du territoire. C'est dans ce contexte qu'est né un projet international afin de découvrir automatiquement les connaissances géographiques à partir d'une fouille d'une base de données géographiques, et de les visualiser en s'appuyant sur la théorie des chorèmes. Cet article présente ce projet en mettant l'accent sur ChorML, langage de description des résumés visuels intermédiaire entre les résultats de la fouille de données et la visualisation, un générateur permettant de traduire les requêtes SQL de fouille de données en ChorML.*

## Abstract

*Chorems are schematic representations of territory and they can represent visually geographic knowledge. Until now chorems were made manually by geographers using their own mental knowledge of the territory. So, a new international project has been launched in order to automatically discover spatial patterns from a geographic database and layout the results, based on chorems theory. This article presents this project and stresses ChorML as a language for representing geographic knowledge to be displayed in a subsequent subsystem, together with a generator translation data mining SQL queries into ChorML.*

**Mots clés :** chorème, base des données géographiques, résumés, connaissances géographiques.

**Keywords:** chorem, geographic databases, summaries, geographic knowledge.

## 1. Introduction

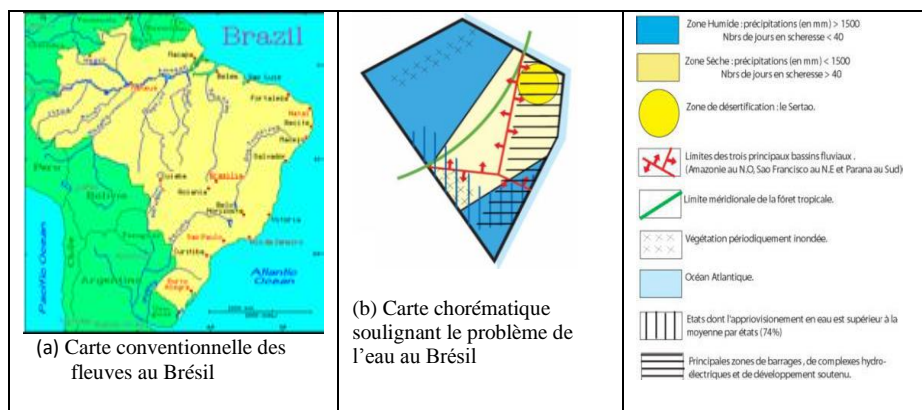
Alors que les données géographiques sont des informations portant sur des objets et événements situés sur le globe terrestre, les chorèmes [BRU 86] sont des représentations schématisées des territoires et représentent la structure et l'organisation de ces territoires. Pour ces raisons, ils semblent constituer une solution intéressante pour générer des résumés de bases de données géographiques afin de faciliter l'aide à la décision spatiale. Dans ce cadre, un projet international, intitulé ChEVIS (Chorem and Visualisation System) entre la France, le Mexique, l'Italie et la Tunisie vise à définir des solutions cartographiques capables de représenter les informations issues de fouille des données géographiques en s'appuyant sur les chorèmes.

Cet article a pour objectif de présenter ce projet en mettant l'accent sur le langage de description des chorèmes nommé ChorML (Chorem Markup Language), ainsi que son système de génération. En premier lieu, nous définissons les chorèmes grâce à un exemple, puis nous présentons le projet ChEVIS. Ensuite, nous décrivons le langage ChorML et en dernier lieu le système de génération du code ChorML.

## 2. Les chorèmes

Un chorème [BRU 86] est défini comme une représentation schématisée d'un espace géographique : ils servent à visualiser les caractéristiques importantes d'un territoire, et ainsi facilitent sa compréhension dans une optique de prise de décision.

Un motif géographique [DEL 09] ou pattern est une régularité intéressante d'un certain phénomène découvert dans une base des données géographiques. Les motifs peuvent être utilisés comme point de départ pour la localisation des phénomènes spatio-temporels et des relations entre eux (cf. figure 1).



**Figure 1.** Fleuves au Brésil ; (a) carte conventionnelle ; (b) carte chorématique soulignant le problème de l'eau dans ce pays ; (c) légende. [LAF 05].

Si la figure 1a donne une carte classique des fleuves au Brésil, elle ne dit rien sur les grands problèmes que doit affronter ce pays dans le domaine de l'eau, alors que

la Figure 1b (ou carte chorématique) met en évidence, par un formalisme associé la nature et la localisation des problèmes. Cette carte chorématique a été conçue par un géographe à partir de ses connaissances mentales sur ce pays. L'objectif de notre projet de recherche est, en partant d'une base de données géographiques, d'extraire les connaissances importantes et de les visualiser de manière à donner un résumé visuel du contenu de la base de données.

### 3. Présentation générale du projet ChEVIS

Le projet de recherche ChEVIS développé en commun par plusieurs laboratoires de recherches s'est donné pour objectif d'extraire les chorèmes à partir de la fouille des données géographiques. En d'autres termes, les connaissances géographiques extraites ne seront pas rédigées avec la logique descriptive, mais par des représentations visuelles. La figure 2 montre l'architecture du système ChEVIS (Chorem Extraction and Visualisation System) qui est composé de deux sous-systèmes, le système d'extraction des chorèmes (Chorem Extraction System) et le système de visualisation des chorèmes (Chorem Visualization System) reliés par les divers niveaux du langage ChorML.

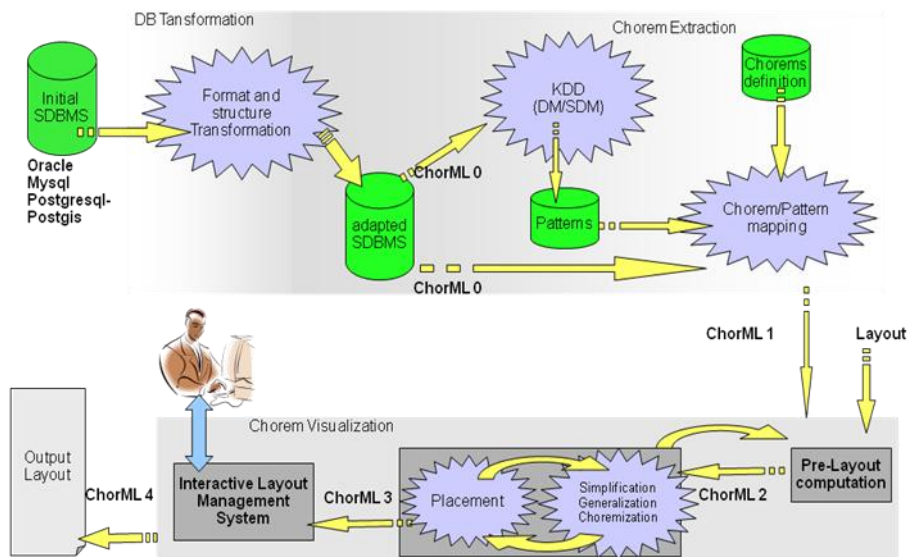


Figure 2. Architecture du système [LAU 08]

#### 3.1 Le système d'extraction des chorèmes

Comme dit précédemment, le point de départ est une base de données géographique dont il s'agit d'extraire les motifs les plus importants. Dès lors, le système d'extraction des chorèmes [DEL 09] permet de transformer une liste de motifs en chorèmes. Dans une première phase, un dictionnaire des chorèmes est

importé dans la base de données pour être employé dans la phase suivante du processus d'extraction des chorèmes. Par la suite, le système exécute un processus d'association entre les motifs et les chorèmes, ainsi qu'un processus de réduction, afin d'obtenir un ensemble restreint de chorèmes. En effet, du processus d'extraction, une grande quantité de motifs peuvent être obtenus, en risquant ainsi de surcharger d'un nombre excessif de chorèmes de la carte chorématique en cours de construction. En termes quantitatifs, l'objectif est d'obtenir entre cinq et dix chorèmes. Les deux phases permettent la participation des utilisateurs afin de sélectionner les motifs et les chorèmes les plus significatifs. Finalement, le système exécute le calcul des relations spatiales existantes parmi les éléments du chorème. Ainsi, des fonctions d'Oracle Spatial sont utilisées par le système et les résultats sont codés en ChorML, qui sera ensuite envoyé au système de visualisation [LAU 07].

### 3.2 Le système de visualisation des chorèmes (carte chorématique)

Le système de visualisation des chorèmes [DEL 09] permet de transformer une liste de chorèmes dans une représentation visuelle; deux phases différentes sont développées par ce système, la création des chorèmes et leur modification.

La création des chorèmes a été développée à travers un système multi-agent qui exécute trois phases : (1) la préparation des dessins, (2) le calcul des coordonnées (depuis longitude/latitude en pixels) puis (3) le placement optimal, qui par des déplacements et simplifications successives détermine le meilleur agencement possible. Cette dernière est effectuée à travers trois opérations (cf. Figure 3) :

- L'opération de simplification qui détermine une version simplifiée des données géométriques, en réduisant le nombre de sommets des polygones de la forme originale. Cette technique est basée sur des fonctions spatiales, qui incorporent l'algorithme de Ramer-Douglas-Peucker (RDP) et ses variantes [RDP 73].
- L'opération de chorémisation qui conditionne la simplification en associant à la composante spatiale du chorème déjà simplifié, la forme d'un polygone régulier avec un faible nombre de côtés.
- L'opération de satisfaction des contraintes topologiques entre les chorèmes afin de respecter les règles de positionnement relatif, notamment après la phase de simplification ; par exemple pour vérifier que les ports sont bien sur la partie terrestre et non pas au milieu de la mer.

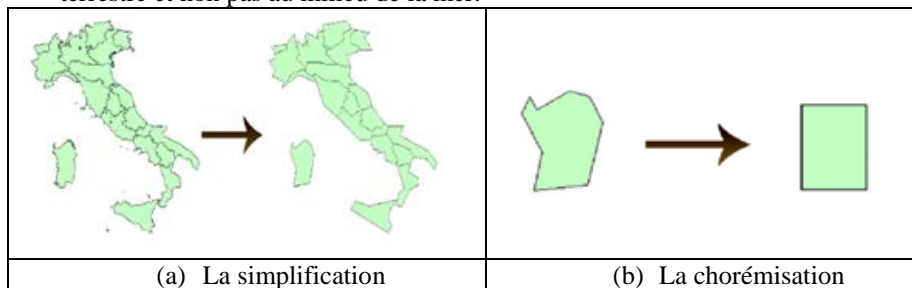
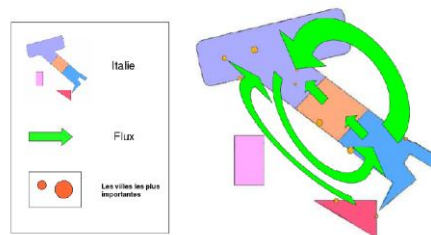


Figure 3. Les deux phases principales de la visualisation des chorèmes [DEL 09].



**Figure 4.** Exemple de carte chorématique produite par le système CheVIS représentant les flux de migration dominants en Italie.

Un exemple de carte chorématique est donné dans la Figure 4. Finalement, la phase de modification des chorèmes fournit aux utilisateurs la possibilité d'éditer celle-ci (Chorem Editor).

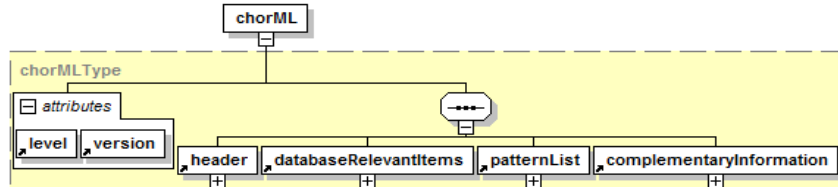
#### 4. ChorML : Chorem Markup Language

ChorML [COI 08] est un langage de type XML, multi-niveaux, dont l'objectif est de mémoriser les informations relatives aux chorèmes et de permettre la communication de telles informations entre les différents modules du système. Ces niveaux sont les suivants :

- *Le niveau 0* est composé de XML et GML [GML 89]. Son objectif est de stocker la BDG en y incluant des informations complémentaires comme les métadonnées [ISO 2003] et des données extérieures au territoire. Par exemple : La France ; généralement les noms des pays limitrophes ne sont pas stockés dans la BD initiale alors qu'il est courant de les voir mentionner dans les cartes chorématiques.
  - *Le niveau 1* de ChorML est également une combinaison de XML et de GML. Il spécifie les procédures et les résultats des algorithmes de fouille des données spatiales. Les éléments du langage sont en particulier :
    - Les informations de caractère général, contenant l'identificateur de la carte, le nom de la carte, le nom de l'auteur, la date de création, le système de référence, le nom de la base des données originale, la dernière date de mise à jour, etc.
    - La liste des motifs et leur origine (traçabilité) dans laquelle les données géographiques sont codées en GML.
    - Une pré-légende, qui contient une description en format texte de chaque chorème.
    - La liste des relations topologiques et non-topologiques parmi les chorèmes.
  - *Le niveau 2* de ChorML est une combinaison de XML et de SVG [SVG 99]. Les éléments complémentaires de ce niveau sont en particulier :
    - La liste des chorèmes simplifiés résultant de la phrase de Chorem Création.
    - La liste des chorèmes résultant de l'opération de Chorem Editing.
    - Une légende qui contient une description en format texte et visuelle de chaque chorème.
    - La liste des relations topologiques et non-topologiques parmi les chorèmes.
- Si le niveau 0 est essentiellement une extension de GML, et le niveau 2 de SVG,

## Un langage et un générateur pour données géographiques

le niveau 1 est le plus original et sa structure est donnée Figure 5 (COI 08).



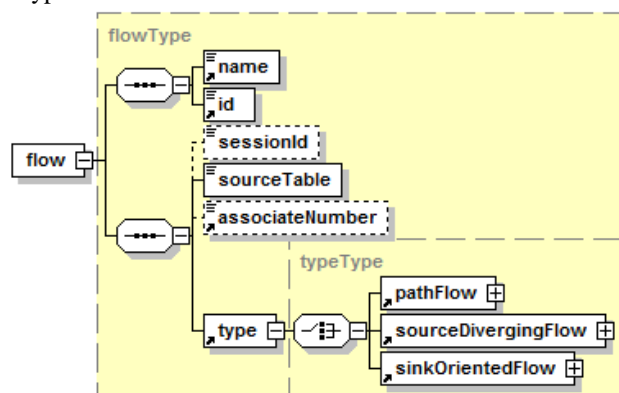
**Figure 5.** La structure de ChorML [COI 08].

Le tableau expose un exemple d'évolution d'un élément géographique ponctuel selon les niveaux de ChorML.

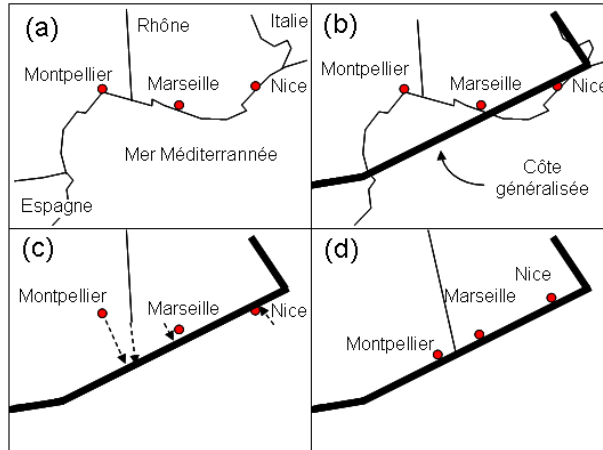
Niveau	Représentation
Niveau 0	Point de longitude / latitude, représentant une ville avec sa population.
Niveau 1	Point de longitude / latitude et son importance s'il est choisi comme motif ; autrement il n'apparaît plus.
Niveau 2	Point avec coordonnées en pixel, représenté par un cercle de rayon déterminé avec une couleur associée.

**Tableau 1.** Présentation d'un point à travers les différents niveaux de ChorML

Suite à notre analyse des chorèmes, il a été convenu de ne garder que les motifs suivants faits, flux, clusters et relations de co-localisation, et à titre d'illustration, la Figure 6 donne la structure de représentation des flux. De même, il a été décidé de coder les contraintes topologiques en se basant sur les relations topologiques au sens d'Egenhofer (EGE 89) en considération de manière à les respecter suite aux opérations de schématisation. A titre d'exemple, dans la Figure 7, les contraintes topologiques à conserver seront le fait que les ports restent sur le continent (point-surface), et que les fleuves se jettent dans la mer (ligne-surface) encodées par des contraintes de type touch-inside.



**Figure 6.** Représentation des flux [COI 08].

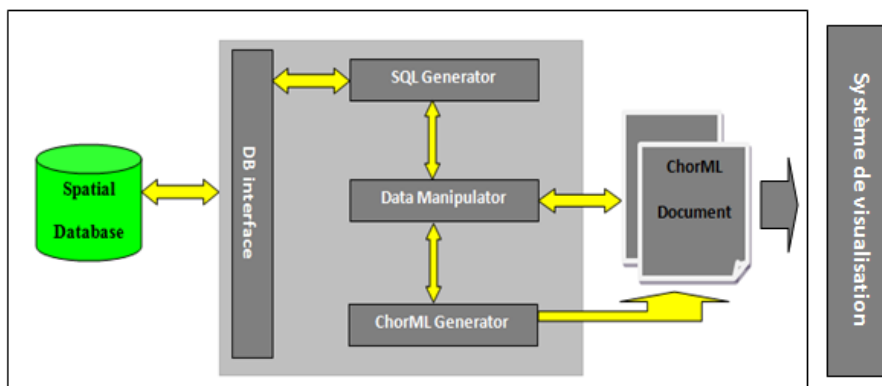


**Figure 7.** Exemples de contraintes topologiques à respecter de type point-surface et ligne-surface : (a) carte originale ; (b) opération de généralisation ; (c) éléments à déplacer ; (d) résultat final.

### 5. Génération automatique de ChorML

Comme dit précédemment, l'objectif de cet article est, une fois exposé le contexte de présenter le système de génération de ChorML en supposant partir, non pas d'une base de données géographiques en GML, mais décrite avec ORACLE Spatial. Comme ce logiciel dispose d'un arsenal important de procédures de fouille de données et de fonctions géographiques, ce générateur se présentera de fait comme un traducteur de SQL ou PL/SQL en ChorML ; pour simplifier l'exposé par la suite nous ne parlerons que de requêtes SQL. Afin de suivre la traçabilité des motifs, il faut mémoriser le texte initial de la requête et ses résultats.

Dès lors, ce système de génération des documents ChorML est composé de quatre modèles comme le montre la Figure 8.



**Figure 8.** Architecture du générateur ChorML à partir d'Oracle Spatial.

## Un langage et un générateur pour données géographiques

Dans ce qui suit nous allons présenter rapidement les quatre parties de notre système.

- a) Connexion : ce module a pour finalité d'assurer la communication entre le générateur et le SGBD utilisé (Oracle dans notre cas), et ce du moment où on lance la requête jusqu'au moment où l'on récupère le résultat. Ce module assure également le bon déroulement de l'exécution de notre connexion.
- b) Génération : c'est à travers ce module que les requêtes seront lancées et exécutées. En fait SQL Generator assure (1) la préparation et l'exécution de la requête SQL appropriée et (2) la récupération du résultat et stockage des données dans une structure bien déterminée.
- c) Manipulation : la tâche principale de ce module est de récupérer le résultat du SQL Generator et d'injecter les données dans le module ChorML Generator ; il permet aussi de faire des modifications sur le ChorML déjà généré. Ce dernier est composé de trois méthodes : (1) Une méthode qui permet de récupérer un résultat d'une requête d'un type donné (sql/plsql) afin d'appeler les méthodes nécessaires au module ChorMLGenerator, elle prend comme paramètres une requête, un nom du fichier pour stocker le résultat de génération et un type de requête. (2) Une méthode qui permet de récupérer le résultat d'une requête cluster afin d'appeler les méthodes de génération de résultats correspondantes au type de requête. (3) Une méthode de génération de contraintes qui seront intégrées dans le fichier final ChorML.

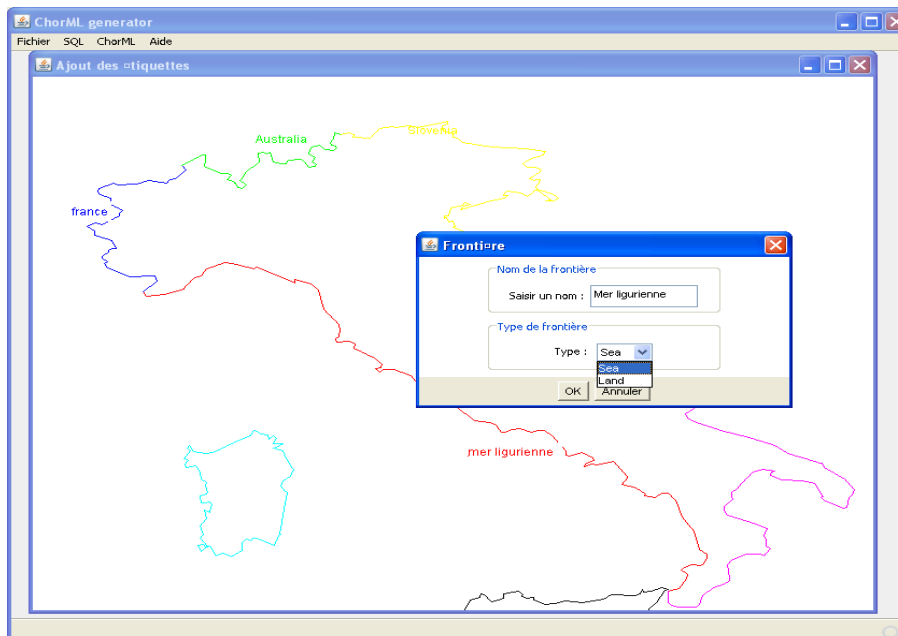


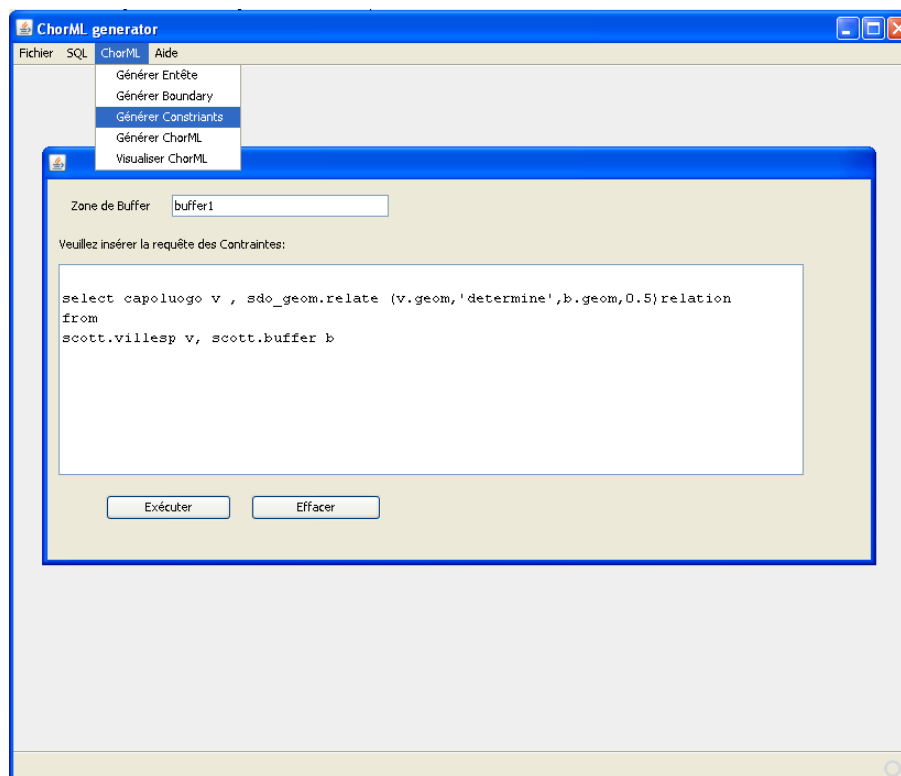
Figure 9. ChorML generator : génération des frontières.



[Tapez un texte]

I. Cherni et al.

- d) ChorML Generator : ce module a pour objectif la construction d'un document XML selon la spécification ChorML. Ce module est un composant essentiel dans l'application développée car il permet la génération de document XML selon la spécification ChorML, il contient (1) Un module qui permet de générer un document XML selon la spécification ChorML d'un élément simple ou d'un cluster ou région ; (2) Un module qui soit le plus important, c'est celui qui fait la transformation d'un élément de type SDO\_GEOMETRY en ChorML en fonction de type de ce champ récupéré de la base de données. (3) Un module qui permet de générer un document XML qui représente les contraintes ChorML. (4) Un module qui permet de stocker des requêtes SQL sous forme d'un document XML. (5) Une méthode qui permet la génération complète d'un document ChorML, elle permet de fusionner les documents générés « l'entête » et « les informations complémentaires ».



**Figure 10.** ChorML generator : génération des relations topologiques.

### 5.1 Etat de l'implémentation

Dans cette section, nous proposons quelques étapes de la génération d'un document ChorML par notre système ChorML Generator : La génération des

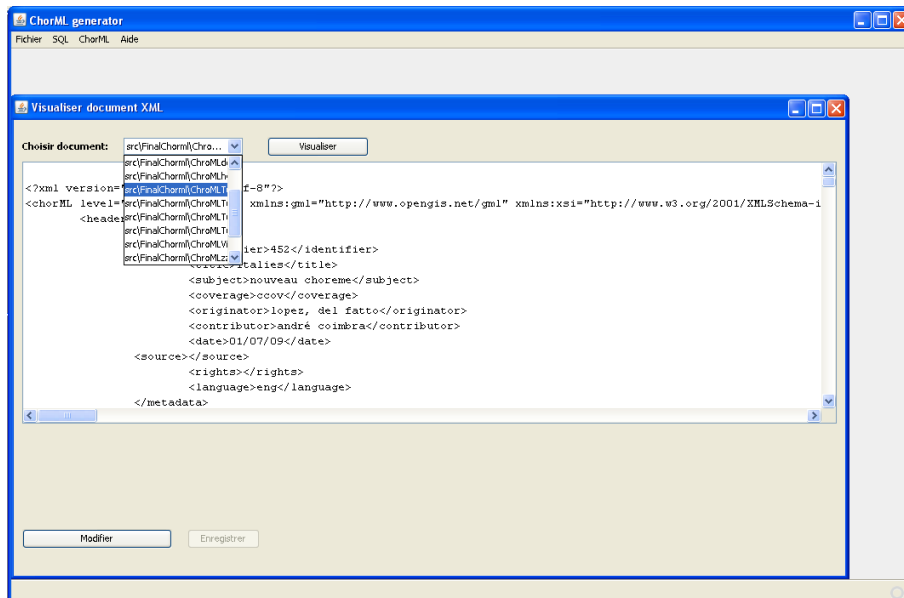
## Un langage et un générateur pour données géographiques

frontières, la génération des relations topologiques et la génération du document ChorML final.

En effet, rares sont les bases de données géographiques qui décrivent rapidement l'extérieur de leur territoire de juridiction. Afin d'assurer une plus grande intelligibilité des chorèmes, il a été décidé d'ajouter des informations extérieures comme les noms des mers, des pays voisins, etc.

La figure 10 montre l'interface de génération des frontières d'un pays à partir d'une interface qui permet de valider le nom de frontière et son type (mer, terre) aussi l'utilisateur doit avoir le choix d'ajouter des frontières régionales de la même manière que pour les frontières. Les frontières régionales sont les frontières entre un pays et une région d'un autre pays.

La génération des relations topologiques se fait par l'exécution de deux requêtes spatiales : SDO-GEOM\_RELATE et SDO-BUFFER, ces deux requêtes vont donner les relations topologiques entre les différents types géométriques. L'interface de génération des relations est montrée par la Figure 10.



**Figure 11.** ChorML generator : génération et visualisation des documents ChorML.

La Figure 11 montre la génération d'un document ChorML qui est le résultat de fusion de trois documents XML : l'entête, la requête et les informations complémentaires.

[Tapez un texte]

I. Cherni et al.

## 5.2 Validation des résultats

Après avoir généré les documents ChorML, il faudrait disposer d'un outil pour valider les résultats obtenus par notre système; c'est pourquoi nous avons décidé de transformer ces résultats en KML (Keyhole Markup Language) [KML 07] qui est un langage basé sur le formalisme XML et destiné à la gestion de l'affichage de données géographiques dans les logiciels Google Earth, Google Maps, Google Mobile et World Wind.

Dans la Figure 12, nous présentons la visualisation d'une requête de motif et de frontière ayant été transformée en KML et puis visualisée avec Google Earth.

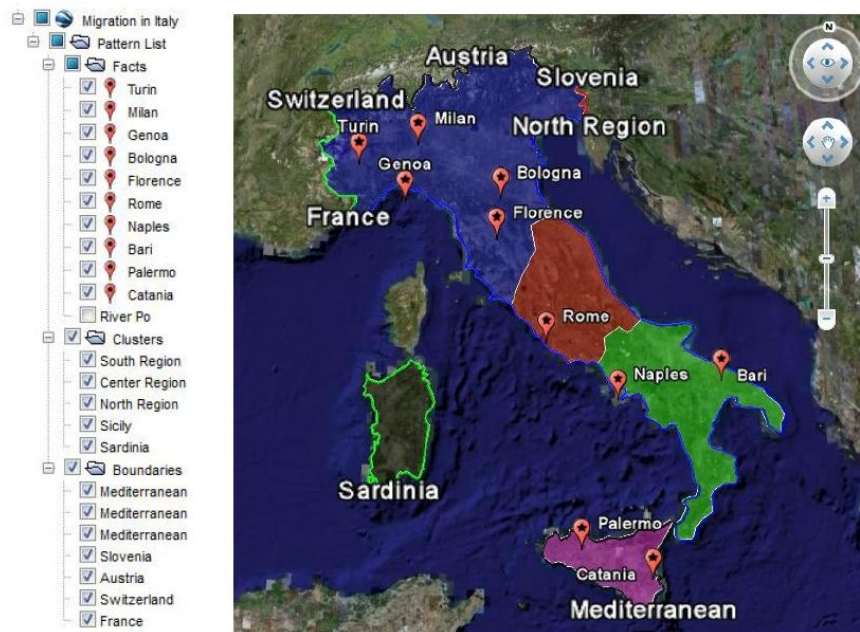


Figure 12. Visualisation avec Google Earth (cluster et frontières).

## Conclusion

Cet article présente notre projet ChEVIS qui permet de définir des solutions cartographiques capables de représenter convenablement les informations extraites d'une base des données géographique. La solution proposée se fonde sur le concept de chorèmes et sur sa capacité de synthétiser des scènes qui comprennent des objets géographiques et des phénomènes spatio-temporels en leur associant des notations visuelles schématisées.

## Un langage et un générateur pour données géographiques

Nous tenons à remercier Pierre Lauret pour la programmation des modules relatifs aux frontières et aux contraintes topologiques.

### Références

- Brunet, R (1986 ). *La carte-modèle et les chorèmes*, Mappemonde 86/4 pp. 4-6.
- André Rocha Coimbra (2008). *ChorML: XML Extension For Modeling Visual Summaries of Geographic Databases Based on Chorems*, Master project, INSA Lyon.
- VinCenzo Del Fatto (2009). *Visual Summaries of Geographic Databases by Chorems*, Ph.D. Thesis, INSA Lyon, University of Salerno.
- Egenhofer M.J (1989). *A Formal Definition of Binary Topological Relationships*. Foundations of Data Organization and Algorithms, 3rd International Conference, FODO 1989, Paris, France, June 21-23, 1989, Proceedings. Lecture Notes in Computer Science 367 Springer 1989, ISBN 3-540-51295-0, pp. 457-472.
- GML (1998). *The Open Geospatial Consortium: Geography Markup Language (GML)*, <http://schemas.opengis.net/gml>
- ISO 19115 (2003). *Geographic Information - Metadata. International Organization for Standardization (ISO)*, Geneva, 2003
- KML (2007). *Keyhole Markup Language (KML)*, (<http://code.google.com/apis/kml/documentation/>)
- Lafon B., Codemard C. et Lafon F (2005). *Essai de chorème sur la thématique de l'eau au Brésil*, (<http://webtab.ac-bordeaux.fr/Pedagogie/Histgeo/espaceleve/bresil/eau/eau.htm>).
- Laurini, R., Milleret-Raffort, F. et Lopez, K (2006). *A Primer of Geographic Databases Based on Chorems*, In proceedings of the SebGIS Conference, Montpellier, Published by Springer Verlag LNCS 4278, pp. 1693-1702.
- Laurini, R. (2008). *Visual Summaries of Geographic Databases*, 14<sup>th</sup> Int'l conf. On Distributed Multimedia System, Invited speaker.
- Douglas, D., Peucker, T (1973). *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*", The Canadian Cartographer 10(2), 112-122.
- SVG (1999) Scalable Vector Graphics (SVG), (<http://www.w3.org/Graphics/SVG/>).