



HAL
open science

Prise en compte des dépendances entre données thématiques utilisateur et données topographiques lors d'un changement de niveau de détail

Kusay Jaara

► **To cite this version:**

Kusay Jaara. Prise en compte des dépendances entre données thématiques utilisateur et données topographiques lors d'un changement de niveau de détail. Milieux et Changements globaux. Université Paris-Est, 2015. Français. NNT : 2015PESC1013 . tel-01272328

HAL Id: tel-01272328

<https://theses.hal.science/tel-01272328>

Submitted on 10 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ
— PARIS-EST

ÉCOLE DOCTORALE — PARIS-EST
Mathématiques et STIC



Prise en compte des dépendances entre données thématiques utilisateur et données topographiques lors d'un changement de niveau de détail

THÈSE

Présentée à l'université Paris-Est par

Kusay Jaara

Pour l'obtention du grade de

Docteur de l'Université Paris-Est

Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information et de la
Communication

Spécialité Sciences et Technologies de l'Information Géographique

Soutenue publiquement le 10/03/2015, devant le jury composé de:

M. Alain Bouju, Maître de conférences, HDR, Université de La Rochelle.....Rapporteur
M. Thomas Devogele, Professeur, Université de Tours.....Examineur
Mme Cécile Duchêne, chargée de recherche, IGN.....Encadrante
M. Michel Mainguenaud, Professeur, INSA de Rouen.....Examineur
Mme Anne Ruas, directrice de recherche, IFSTTAR.....Directrice
M. Eric Saux, Maître de Conférences, HDR.....Rapporteur

Résumé

Avec l'importante disponibilité de données topographiques de référence, la création des données géographiques n'est plus réservée aux professionnels de l'information géographique. De plus en plus d'utilisateurs saisissent leurs propres données, que nous appelons *données thématiques*, en s'appuyant sur ces données de référence qui jouent alors le rôle de données support. Les données thématiques ainsi saisies font sens en tant que telles, mais surtout de par leurs relations avec les données topographiques. La non prise en compte des relations entre données thématiques et topographiques lors de traitements modifiant les unes ou les autres peut engendrer des incohérences, notamment pour les traitements liés au changement de niveau de détail. L'objectif de la thèse est de définir une méthodologie pour préserver la cohérence entre les données thématiques et topographiques lors d'un changement de niveau de détail.

Nous nous concentrons sur l'adaptation des données thématiques suite à une modification des données topographiques, processus que nous appelons *migration des données thématiques*. Nous proposons d'abord un modèle pour la migration de données thématiques ponctuelles sur réseau composé de : (1) un modèle pour décrire le référencement des données thématiques sur les données topographiques par des relations spatiales (2) une méthode de relocalisation basée sur ces relations. L'approche consiste à identifier les relations finales attendues en fonction des relations initiales et des changements sur les données topographiques entre les états initial et final. La relocalisation est alors effectuée grâce à une méthode multicritère de manière à respecter au mieux les relations attendues. Une mise en œuvre est présentée sur des cas d'étude jouets et sur un cas réel fourni par un service de l'Etat gestionnaire de réseau routier. Nous discutons enfin l'extension du modèle proposé pour traiter la prise en compte des relations pour d'autres applications que la migration de données thématiques.

Mots-clés

Données géographiques, données thématiques, données topographiques, référencement spatial indirect, relations spatiales, changement de niveau de détail, intégration de données utilisateur.

Abstract

With the large availability of reference topographic data, creating geographic data is not exclusive to experts of geographic information any more. More and more users rely on reference data to create their own data, hereafter called *thematic data*. Reference data then play the role of support for thematic data. Thematic data make sense by themselves, but even more by their relations with topographic data. Not taking into account the relations between thematic and topographic data during processes that modify the former or the latter may cause inconsistencies, especially for processes that are related to changing the level of detail. The objective of this thesis is to define a methodology to preserve the consistency between thematic and topographic when the level of detail is modified. This thesis focuses on the adaptation of thematic data after a modification of topographic data: we call this process *thematic data migration*. We first propose a model for the migration of punctual thematic data hosted by a network. This model is composed of: (1) a model to describe the referencing of thematic data on topographic data using spatial relations (2) a method to re-locate thematic data based on these relations. The approach consists in identifying the expected final relations according to the initial relations and the modifications of topographic data between the initial and the final state. The thematic data are then re-located using a multi-criteria method in order to satisfy, as much as possible, the expected relations. An implementation is presented on toy problems and on a real use case provided by a French public authority in charge of road network management. The extension of the proposed model to take into account the relations for other applications than thematic data migration is also discussed.

Keywords

Geographic data, thematic data, topographic data, indirect spatial referencing, spatial relations, level of detail change, user data integration.

Table des matières

Résumé	3
Abstract	4
Table des matières	5
Remerciements	8
Introduction	9
Chapitre A. Contexte, besoin et objectif.....	12
A.1 Contexte : Utilisation combinée de données thématiques et topographiques	12
A.1.1 Définitions des données thématiques et des données topographiques de référence	12
A.1.2 Niveaux de détail et généralisation des données topographiques	13
A.1.3 Méthodes de création de données thématiques	16
A.1.4 Référencement géographique des objets	18
A.1.5 Les relations spatiales.....	20
A.1.6 La représentation des relations spatiales.....	20
A.1.7 Importance des relations entre les données thématiques et les données topographiques	21
A.1.8 Conséquence de la non prise en compte des relations entre les données thématiques et les données topographiques lors de la modification des données topographiques	22
A.2 Objectif et positionnement.....	25
A.2.1 Hypothèse de travail.....	25
A.2.2 Rôle des relations lors de traitements liés au changement du niveau de détail .	25
A.2.3 Objectif de la thèse.....	28
A.3 Etat de l'art.....	29
A.3.1 Classifications des relations.....	29
A.3.2 Taxonomie des relations	31
A.3.3 Description des relations en langage formel vs en langage naturel	32
A.3.4 Comment les relations sont décrites (d'une façon formelle).....	32
A.3.5 Modélisation des relations	34
A.3.6 Prise en compte des relations dans la généralisation	42
A.3.7 Prise en compte des relations dans l'appariement.....	46
A.3.8 Prise en compte des relations dans l'intégration de données géographiques	48
A.3.9 Prise en compte des relations dans la migration de données.....	51
A.4 Bilan du chapitre.....	51

Chapitre B. Modèle pour la migration de données thématiques ponctuelles sur réseau	54
B.1 Introduction.....	54
B.2 Vue d'ensemble du modèle proposé pour la migration de données thématiques ponctuelles sur réseau.....	55
B.2.1 Description formelle du problème	56
B.2.2 Les étapes de l'approche pour la migration de données thématiques.....	57
B.3 Extraction et modélisation des relations	59
B.3.1 Concept d'objets caractéristiques	59
B.3.2 Importance des relations initiales et importance des relations finales	61
B.3.3 La relation « porté par »	62
B.3.4 Quels sont les objets en relation et quelles sont les relations à extraire ?	64
B.3.5 Modèle de référencement des données thématiques.....	65
B.3.6 Schéma de données proposé pour le référencement des données thématiques sur les données support	70
B.4 Identification des relations finales attendues.....	81
B.4.1 Appariement entre les bases topographiques initiale et finale	81
B.4.2 Modification et transformation des relations	83
B.4.3 Détail du traitement de l'identification des relations finales attendues	85
B.4.4 Application de l'étape sur l'exemple jouet.....	86
B.5 Relocalisation des données thématiques basée sur les relations en utilisant un système multicritère.....	87
B.5.1 Phases de la relocalisation.....	87
B.5.2 Besoin d'une évaluation de la satisfaction des relations	89
B.5.3 L'évaluation du respect des relations d'une position possible en utilisant une méthode multicritère	91
B.6 Récapitulation : illustration de notre méthode de migration sur les exemples-jouets.	93
B.7 Adaptation du modèle pour d'autres cas que la migration de données thématiques ponctuelles sur réseau.....	96
B.7.1 Adaptation du modèle pour d'autres cas de migration de données thématiques	96
B.7.2 Adapter la généralisation automatique des données topographiques pour respecter les relations finales attendues	100
B.7.3 Prise en compte des relations dans un processus de carte à la demande.....	103
B.7.4 Système pour la prise en compte des relations pour tous les types traitements	107
B.8 Bilan du chapitre B.....	109
Chapitre C. Mise en œuvre de la migration de données thématiques ponctuelles sur réseau	112
C.1 Présentation de cas d'étude réel.....	112
C.1.1 Constitution de référentiels routiers par le CETE et sources d'erreurs associées	112
C.1.2 Conséquences des erreurs de référencement et besoins du CETE	117
C.2 Cas d'étude retenus pour la mise en œuvre de notre modèle	117
C.2.1 Démarche et choix des cas d'étude.....	117

C.2.2	Description de cas d'étude jouet.....	118
C.2.3	Description de notre cas d'étude réel	119
C.2.4	Différence entre les cas d'étude jouet et le cas réel	120
C.3	Mise en œuvre du modèle sur les cas d'étude	120
C.3.1	Plateforme d'implémentation	120
C.3.2	Modifications apportées à la plate-forme pour les besoins de la thèse	121
C.3.3	Préparation des données.....	121
C.3.4	L'extraction des objets caractéristiques (étape a)	122
C.3.5	Relations qui sont prises en compte et méthodes d'extraction (étape b)	126
C.3.6	Mise en œuvre de l'appariement (étape c).....	131
C.3.7	Identification des relations finales attendues (étape d)	136
C.3.8	Relocalisation basée sur les relations attendues (étape e)	137
C.4	Analyse des résultats obtenus	143
C.4.1	Résultats sur le cas d'étude jouets	143
C.4.2	Résultats sur le cas d'étude réel.....	145
C.4.3	Bilan du chapitre C.....	150
Conclusion et perspectives		152
Rappel de l'objectif.....		152
Bilan des contributions		152
Modélisation des relations entre données thématiques et données topographiques		152
Méthode pour la migration de données thématiques		153
Une méthode d'appariement partiel de réseau munis de points caractéristiques.....		153
Améliorer le paramétrage.....		153
Comparer l'appariement avec d'autres		154
Représenter les relations entre les données thématiques.....		154
Appliquer l'approche pour la migration des données linéaires et surfaciques		154
Réutiliser des parties de notre approche pour le cas de la généralisation		154
Améliorer la documentation des processus de saisie de données géographiques.....		155
Liste des figures.....		157
Bibliographie		162

Remerciements

Je remercie tout d'abord Anne Ruas et Cécile Duchêne pour m'avoir donné la chance de vivre cette expérience avec elles. Les conseils avisés d'Anne m'ont parfaitement guidé pendant mon travail, et j'ai apprécié la grande disponibilité de Cécile pour les discussions et la relecture de mes productions scientifiques.

Je remercie les membres du jury pour avoir accepté d'évaluer cette thèse et pour leurs remarques précieuses.

Je remercie Nicolas Regnauld d'avoir accepté que je puisse effectuer une partie de mon travail en Grande Bretagne à l'Ordnance Survey (OSGB). Je remercie Nick Gould et Sandrine Balley, les discussions que j'ai eues avec eux à l'OSGB étaient bien enrichissantes pour mon travail de thèse.

Je remercie Yves Bonin, le chef de projet au Centre d'Études et d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement (CEREMA) pour son aide, pour les discussions que nous avons pu avoir, ainsi que pour sa relecture. Le travail avec lui m'a permis de découvrir le coté applicatif de mon travail.

Je remercie les relecteurs de mon rapport : Lobna, Hasna, Yann et Firas pour en avoir corrigé l'écriture et avoir aidé à améliorer sa rédaction.

Je remercie l'ensemble des membres du COGIT pour leur amitié et leur accueil.

Et bien entendu je remercie ma famille : mes parents, mes frères, avec un grand merci à Waddah qui s'est déplacé en France juste pour assister à ma soutenance. Je remercie ma fiancée pour son encouragement et la force qu'elle m'a donnée pour finir cette thèse.

Introduction

Les technologies de l'information ont changé considérablement la production des cartes. Avec l'usage croissant de l'internet, les cartes web offrent plus de fonctionnalités pour le public et aussi pour les utilisateurs professionnels des données géographiques. Les producteurs mettent leurs données à disposition des utilisateurs sur le web, par exemple via des géoportails. Les données collaboratives, dont celles produites par le projet OpenStreetMap, sont de plus en plus utilisées, et certains chercheurs ont étudié les spécifications de ce type de données et proposé des méthodes pour les qualifier (Brando et al. 2011).

Suite à la grande disponibilité des données géographiques de référence pour le grand public, les utilisateurs commencent à créer leurs propres données en s'appuyant sur ces données. La saisie est souvent effectuée en utilisant des données topographiques, qui servent comme référence pendant la saisie. Un utilisateur peut être un professionnel (par exemple un gestionnaire de routes) ou un utilisateur grand public (par exemple un randonneur qui veut partager son chemin de randonnée).

Les données de l'utilisateur, que nous appelons aussi *données thématiques* dans la suite du texte, sont différentes des données topographiques. En règle générale, les spécifications des données thématiques (la précision, le niveau de détail, la position relative) sont dérivées des données topographiques sur lesquelles les données utilisateur ont été saisies.

Il est de plus en plus fréquent, pour répondre à différents besoins, de faire de la co-visualisation ou de l'analyse de données ayant une composante thématique et une composante topographique. Pour des besoins de visualisation ou d'assemblage de données, entre autres, il peut être nécessaire de modifier la composante topographique et/ou la composante thématique des données. Modifier les données topographiques peut notamment consister à les remplacer par d'autres, ou à leur appliquer un processus de généralisation de manière à réduire leur niveau de détail. La généralisation peut être nécessaire au niveau des données topographique et/ou au niveau des données thématiques.

Les données thématiques ont souvent une forte dépendance sémantique aux données topographiques de référence qui jouent pour elles le rôle d'espace support. Par exemple, des données d'accidentologie routière manipulées par les gestionnaires de routes sont fortement dépendantes du réseau routier. Les relations spatiales entre les données thématiques et les données topographiques sont porteuses de sens, et doivent donc être préservées, voire exagérées lorsque l'une ou l'autre des composants thématique ou topographique, ou les deux, sont modifiées.

Cette thèse présente une méthodologie pour le changement de niveau de détail d'un jeu de données composé de données utilisateur et de données topographiques de référence. La méthodologie proposée permet de prendre en compte les relations entre les données de l'utilisateur et les données topographiques de référence, ces relations jouent un rôle essentiel dans les traitements liés au changement de niveau de détail.

La thèse est composée de trois chapitres. Le chapitre A décrit le contexte, la problématique et l'objectif de la thèse. En fin du chapitre, nous présentons un panorama des travaux existants liés aux relations spatiales et à leur prise en compte dans des traitements automatiques tels

que l'intégration et la généralisation automatique. Notre modèle pour la migration de données thématiques ponctuelles sur réseau est exposée dans le chapitre B. La dernière partie de ce chapitre présente des propositions théoriques pour l'adaptation du modèle pour d'autres types de traitements de données comportant des composantes thématique et topographique. Le Chapitre C présente une mise en œuvre de l'approche proposée dans le chapitre B, d'une part sur des cas d'étude jouets, d'autre part sur un cas d'étude réel fourni par le Centre d'Etudes Techniques de l'Equipement d'Aix-en-Provence : le recalage de données métier ponctuelles lié à la gestion des routes sur un réseau routier différent de celui sur lequel elles sont initialement référencées.

Chapitre A

Contexte, besoin et objectif

Ce chapitre commence par définir les *données thématiques* et les *données topographiques* puis présente différents modes de création de données thématiques. Ensuite, nous nous focaliserons sur les relations spatiales, leur importance et les particularités des relations entre les données thématiques et les données topographiques. Ensuite, nous étudions des problématiques associées au traitement d'une combinaison entre des données thématiques et des données topographiques, puis nous présenterons l'objectif de la thèse. Nous finirons ce chapitre en présentant différents travaux concernant l'utilisation des relations lors du traitement de données géographiques.

Chapitre A. Contexte, besoin et objectif

A.1 Contexte : Utilisation combinée de données thématiques et topographiques

Nous étudions la prise en compte des relations spatiales lors du traitement de la combinaison de données thématiques avec des données topographiques, spécifiquement dans le cas d'un changement de niveau de détail. Le but de la partie A.1 est d'introduire les concepts liés à ces sujets (types de données, différences entre les niveaux de détails), puis l'intérêt des relations pour les données thématiques. Pour cela, nous expliquons des méthodes de saisie et des types de référencement des données thématiques sur les données topographiques en montrant le rôle des relations. Ensuite, nous montrons l'importance de la prise en compte des relations, et les conséquences en cas d'ignorance de ces relations pendant les traitements.

A.1.1 Définitions des données thématiques et des données topographiques de référence

Parmi les données géographiques, nous distinguons deux types : les *données thématiques* et les *données topographiques* (Figure 1). D'après le glossaire de la cartographie (CFC 1990) : «une carte thématique est une carte représentant sur un fond repère topographique, hydrographique, chorographique ou géographique, des phénomènes localisables de toutes natures, qualitatifs ou quantitatifs».

Nous appelons *données thématiques* les données qui sont liées à un thème particulier, par exemple, les accidents routiers, les lignes de bus, les zones d'avalanches. Les données thématiques peuvent être surfaciques (par ex. zones d'avalanches), linéaires (par ex. itinéraire de randonnées), ou ponctuelles (par ex. accidents routiers). Les données thématiques sont souvent saisies ou représentées sur des *données topographiques* ou sur une *carte topographique*. Selon le même glossaire, la carte topographique a pour but de représenter des éléments naturels ou artificiels situés sur la surface terrestre. Les *données topographiques* sont donc les données géographiques utilisées pour localiser les données thématiques.

La combinaison des données venant de plusieurs sources forment ce que nous appelons un *mashup cartographique* (McConchie 2008, p.12). Un *mashup* est défini dans l'informatique comme une application web qui utilise le contenu de plusieurs sources pour créer un nouveau service (Fichter 2013), par exemple, une page Facebook peut contenir des vidéos Youtube. Les premiers mashups cartographiques ont été nommés « map hacks » (Gibson et Erle 2006).

Les *mashups cartographiques* (en Anglais *mapping mashups*) impliquent des données cartographiques, alors le terme s'applique aux mashups qui implique une carte. Certains professionnels considèrent les mashups cartographiques comme une version moins avancée du SIG en utilisant des données moins précises (McConchie 2008, p.5).

Nous appelons un *mashup de données géographiques*, l'ensemble de deux ou plusieurs couches de données géographiques, parmi lesquelles il existe des couches thématiques et des couches topographiques. Les mashup de données géographiques sont couramment utilisés par le grand public et les professionnels.

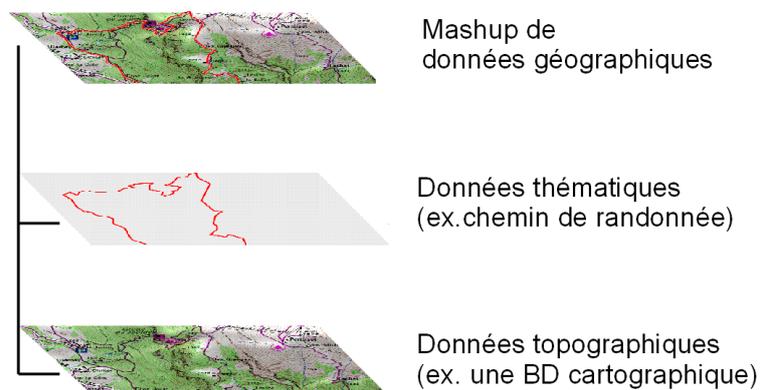


Figure 1. Un mashup entre des données thématiques et des données topographiques

Malgré ces définitions, la limite entre les données topographiques et les données thématiques n'est pas toujours nette. Par exemple, une piste cyclable peut être saisie sur un fond topographique qui ne contient pas de données sur les pistes cyclables. Mais, une piste cyclable peut aussi faire partie des données topographiques utilisées pour localiser un itinéraire de vélo.

A.1.2 Niveaux de détail et généralisation des données topographiques

Les bases de données topographiques ont un niveau de détail variable. Dans la suite, nous expliquons les différences liées au niveau de détail, puis nous introduisons la généralisation, processus qui a pour objet le changement le niveau de détail.

Niveau de détail et échelle

Le niveau de détail concerne la sélection des objets à représenter. Le niveau de détail géométrique est représenté par la résolution, la précision et l'échelle. La définition de la résolution par (David et Fasquel 1997) est «*étroitesse de l'accord entre une mesure ou une estimation et l'espérance de cette mesure ou de cette estimation*». Le Figure 2 représente la résolution pour les données maillées. La précision est l'écart maximum entre la réalité et la représentation (Ruas 2004 p.23).

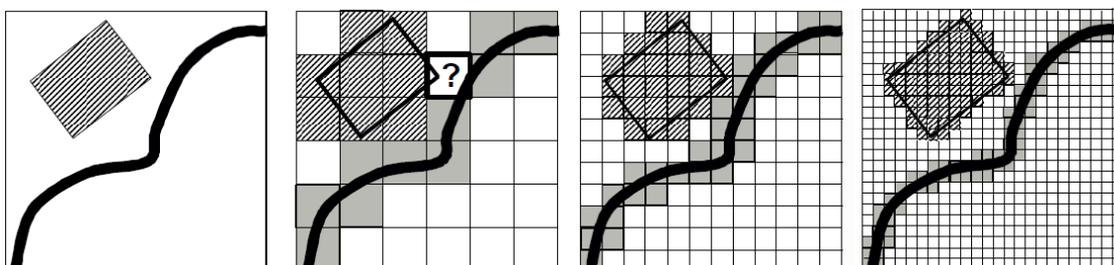


Figure 2. La résolution des données maillées: représentation d'une route et d'un bâtiment (Ruas 2004)

(Girres 2012 p.169) distingue l'échelle de représentation et l'échelle de saisie d'une base de données :

Echelle de représentation : Une base de données géographiques n'a pas d'échelle précise, mais pendant la création, elle est construite pour une *gamme d'échelles* (Ruas et Mustière 2005), par exemple, la gamme d'échelle pour la BD TOPO de l'IGN (IGN 2011) va du 1 :5 000 au 1 :50 000.

Echelle de saisie: La construction d'une base de données géographiques implique des traitements pour représenter le monde réel à l'échelle de représentation désirée, les processus de production ne sont pas toujours homogènes, en conséquence, l'échelle de saisie n'est pas toujours unique. Une base géographique peut contenir des courbes de niveau issues d'un modèle numérique de terrain, des bâtiments extraits de photogrammétrie aérienne à une certaine échelle. Par ailleurs, un opérateur qui saisit les données travaille à plusieurs échelles en même temps.

La Figure 3 montre l'échelle de saisie et l'échelle de représentation pour la BD TOPO et la BD CARTO de l'IGN.

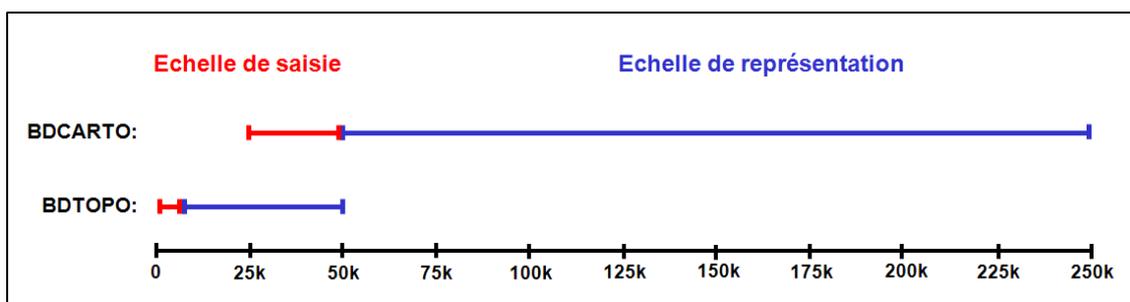


Figure 3. Gammes d'échelles de saisie et de représentation de la BD TOPO et de la BD CARTO (Girres 2012)

Selon le niveau de détail d'une base de données (la gamme d'échelle à laquelle elle correspond), on n'y trouve pas les mêmes types d'objet du monde réel, de même que sur une carte. En Figure 4, les bâtiments sont représentés au 1 :100 000 mais pas au 1 :250 000.

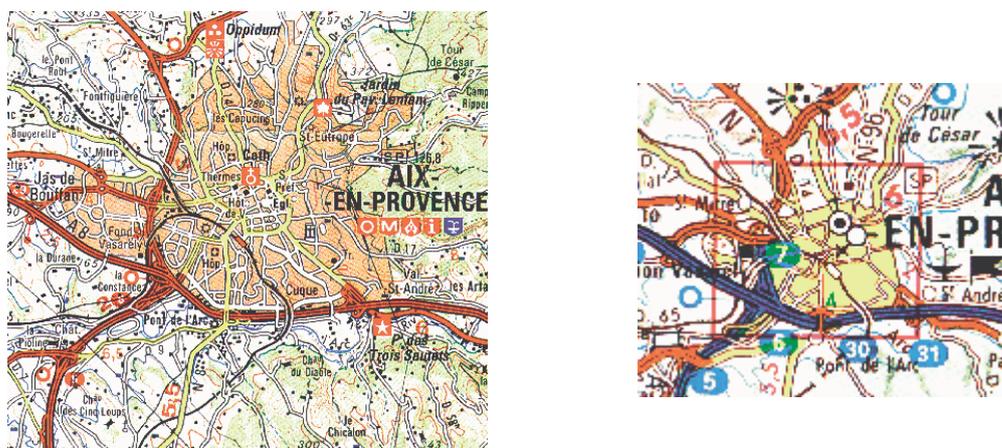


Figure 4. Différence de contenu au 1: 100 000 et 1: 250 000. Cartes IGN (Ruas 2004)

Les données topographiques diffèrent par la présence ou l'absence des entités par rapport au monde réel (David et Fasquel 1997) et le niveau de détail limite le nombre d'objets du monde réel représenté dans la base topographiques (Ruas 1999).

Le niveau de détail sémantique inclut le schéma de données et les règles de sélection (Ruas 2004). Le niveau de détail sémantique contient les réponses sur : *Quels sont l'information retenue sur l'objet ?* et *Quel est le niveau de détail de l'information retenue?* Par exemple, la largeur de la route est représentée dans une base et elle peut être décrite au mètre près ou au centimètre près.

Changer le niveau de détail par la généralisation

La généralisation est un processus d'abstraction sur les données vectorielles pour diminuer leur niveau de détail. La généralisation consiste à dériver une base avec une échelle de représentation plus petite (Ruas 2004).

Plus généralement, la généralisation peut être définie comme un processus de synthèse similaire au résumé du texte et inclut la réduction des informations en gardant l'information la plus importante (Ruas 2004). Par ailleurs, le résultat doit ressembler aux données initiales et être lisible facilement.

La généralisation ne consiste pas seulement en des manipulations sur les primitives géométriques. Le contexte doit être pris en compte, en plus des interdépendances et des formes caractéristiques des phénomènes à représenter (Mackaness 2007). Suivant le niveau de détail ou l'échelle voulue et le but de la carte, certains objets seront représentés et d'autres objets seront éliminés.

Les recherches en généralisation automatique ont pour but d'automatiser ce processus. Au-delà de l'étape de généralisation, des travaux ont été réalisés afin d'automatiser toute la chaîne de fabrication d'une carte. La carte produite contient éventuellement des données thématiques (Balley et Regnauld 2011).

Les entités géographiques sont généralisées en appliquant des opérations. Plusieurs classifications existent de ces opérateurs. La classification de (Mustière 2001) distingue trois types d'opérateurs, les opérateurs de simplification, de caricature et d'harmonisation :

1. les opérateurs de simplification : ils simplifient la forme des objets, qui deviennent plus lisibles à une échelle plus petite.
2. les opérateurs de caricature : ils ont pour but d'accentuer certains caractères des objets
3. les opérateurs d'harmonisation : ils réduisent les différences graphiques pour décomplexifier la lecture d'information.

A.1.3 Méthodes de création de données thématiques

Il existe différentes méthodes pour la création de données thématiques, nous présenterons trois méthodes couramment utilisées.

1^{er} méthode : Saisie manuelle en utilisant un fond de données topographiques

Les données thématiques peuvent être saisies manuellement par un utilisateur. La saisie est faite sur des données topographiques. Aujourd'hui, l'utilisateur obtient les données topographiques souvent via le web, plus particulièrement via les géoportails sous forme d'une carte web.

La Figure 5 montre l'exemple d'un utilisateur qui dessine son chemin de randonnée sur une carte IGN fournie par le géoportail de l'IGN.



Figure 5. Saisie d'un chemin de randonnée sur une carte IGN fournie par le géoportail de l'IGN (www.geoportail.gouv.fr)

Pendant la saisie, l'utilisateur n'a pas la géométrie exacte de ses objets, mais il connaît plus ou moins la position relative de ses objets thématiques par rapport à certains objets topographiques qui sont considérés comme des objets repères. L'utilisateur saisit donc les données en se basant sur les relations spatiales avec les objets repères qu'il a identifiés.

Le degré de cohérence avec les données topographiques est variable et lié au degré de méticulosité de l'utilisateur. La Figure 6 montre un itinéraire de vélo saisi manuellement, nous remarquons la perte de cohérence dans plusieurs parties de l'itinéraire.

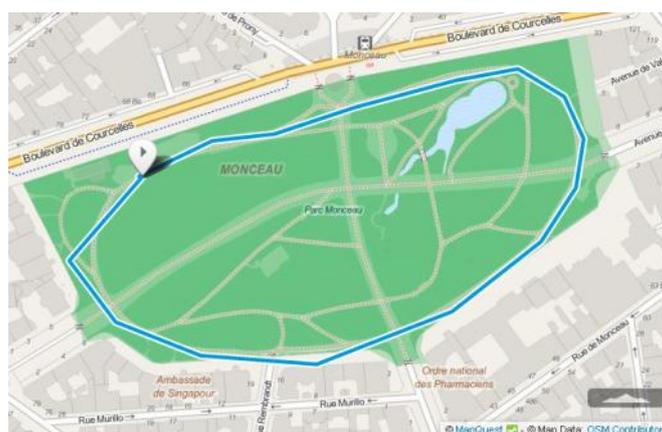


Figure 6. Exemple d'une saisie manuelle avec peu de prudence sur un fond de carte web (source www.bikemap.net)

Nous noterons pour ce type de processus que le fond topographique peut être de type raster ou de type vecteur. Si les données sont de type vecteur, il est possible de recalculer automatiquement les données de l'utilisateur pendant la saisie, cela aide à obtenir une meilleure cohérence. En effet, le recalage est effectué en se basant sur les relations topologiques qui peuvent être déclarées par l'utilisateur. Par exemple, les lignes de bus sont recalées après ou pendant la saisie pour qu'elles soient exactement sur les routes.

Cette méthode est utilisée lors de la saisie des données géographiques volontaires comme OpenStreetMap. Le rattachement des données de contenu collaboratif aux données de référence améliore la cohérence et l'utilisabilité (Brando 2013, p17).

2^{ème} méthode : Saisie professionnelle en utilisant un fond topographique et un SIG

Les spécifications facilitent l'exploitation, l'analyse et la gestion des données. Pour saisir leurs données thématiques « métier », les professionnels s'appuient souvent sur des données topographiques qui ont des spécifications bien définies, plus particulièrement des données fournies par un fournisseur institutionnel, telles que la BD TOPO ou la BD CARTO de l'IGN.

La saisie peut être effectuée avec l'aide de plusieurs outils. Par exemple, imaginons un gestionnaire de route qui veut saisir un accident qui s'est passé 50 m après le sommet d'un virage d'une route nommée. L'utilisateur cherche la route sur les données vecteur de référence. Ensuite, il cherche le sommet du virage concerné avant d'utiliser un outil pour placer l'accident à 50 mètres du sommet du virage précisément. La Figure 7 montre des données professionnelles ponctuelles (saisie des points de repère métier), qui sont saisies selon ce mode par un gestionnaire de réseau routier sur un réseau (Conseils Général du Var).

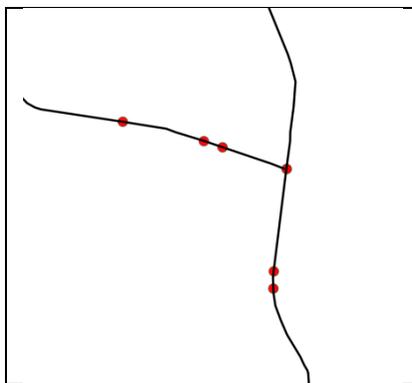


Figure 7. Réseau routier avec des données thématiques professionnelles

3^{ème} méthode : Création des données thématiques sans utiliser un fond topographique

Les données thématiques peuvent être créées sans utiliser un fond topographique. Par exemple, sur les traces GPS faites par un randonneur, la précision varie selon l'appareil et la disponibilité des signaux GPS. Le problème de ce type de données thématiques est l'absence de cohérence entre les traces et les données topographiques. En effet, les traces GPS n'ont pas souvent la même précision et la résolution des traces varie selon la couverture, ce qui pose un problème lors de l'intégration de ces données avec d'autres données géographiques (Figure 8).



Figure 8. Trace GPS sur des données IGN

Le recalage de traces GPS sur des données réseau fait d'ailleurs l'objet de recherche. Un algorithme de recalage a été proposé par (Newson et Krumme 2009) basé sur le modèle de markov caché. Cet algorithme a été étendu par (Hauert et Budig 2012) pour qu'il soit appliqué même dans le cas des données topographiques incomplètes (par exemple, les données de OpenStreetMap).

A.1.4 Référencement géographique des objets

Le but du référencement est de décrire la localisation d'un élément dans l'espace géographique. Selon (Hill 2006), le référencement (ou géoréférencement) est le fait de lier l'information à une localisation géographique. Quand nous utilisons une application de calcul d'itinéraire comme celle de Google Maps (Figure 9), l'application propose l'itinéraire (données thématiques) sur la carte avec des indications textuelles sur les routes à prendre et les changements de direction. L'application décrit l'itinéraire en donnant le lien entre l'itinéraire et les données topographiques. Autrement dit, l'application donne un référencement de l'itinéraire sur les données topographiques.



Figure 9. Calcul d'itinéraire sur Google Maps (<https://maps.google.fr/>)

Le terme concerne toutes les données peu importe leur type de géométrie (vecteur ou raster). Nous étudions uniquement le cas des données vectorielles. Les trois méthodes de référencement les plus utilisées sont décrites ci-dessous.

Référencement par coordonnées

Un endroit peut être localisé par ses coordonnées (coordonnées géographiques ou coordonnées en projection). Les coordonnées peuvent être fournies par un SIG dans le cas de la saisie sur carte ou images géoréférencées ou par des appareils dédiés comme un appareil GPS ou un smartphone équipé d'un capteur GPS. C'est le cas par exemple pour les traces de randonnées construites par les applications de géonavigation au fur et à mesure que le randonneur progresse. Le référencement par coordonnées est défini dans la norme *ISO 19111*.

Référencement par le nom

Le référencement par nom de lieu est la méthode la plus utilisée pour décrire la position d'un lieu (Hill 2006). La méthode utilise une base qui contient des noms des lieux avec leurs coordonnées absolues. Chaque objet à référencer est décrit par le nom du lieu dans lequel il se situe, et en cherchant dans la base, on obtient sa localisation absolue. Cette méthode est la plus utilisée parce que nous avons des noms pour beaucoup de lieux: les régions, les villes et communes, les montagnes, les sommets, les monuments. Un dictionnaire de noms de lieux est appelé *index géographique* (en anglais : *gazetteer*). L'index géographique est défini dans la norme *ISO 19112* (ISO 2003). Il inclut, en plus du nom et de la localisation exacte, des informations descriptives, telles que la catégorie : restaurant, monument touristique, etc. Un index géographique décrit un lien de traduction entre l'information structurée sous forme de coordonnées et l'information textuelle sous forme de phrases. La Figure 10 montre la structure d'un index géographique.

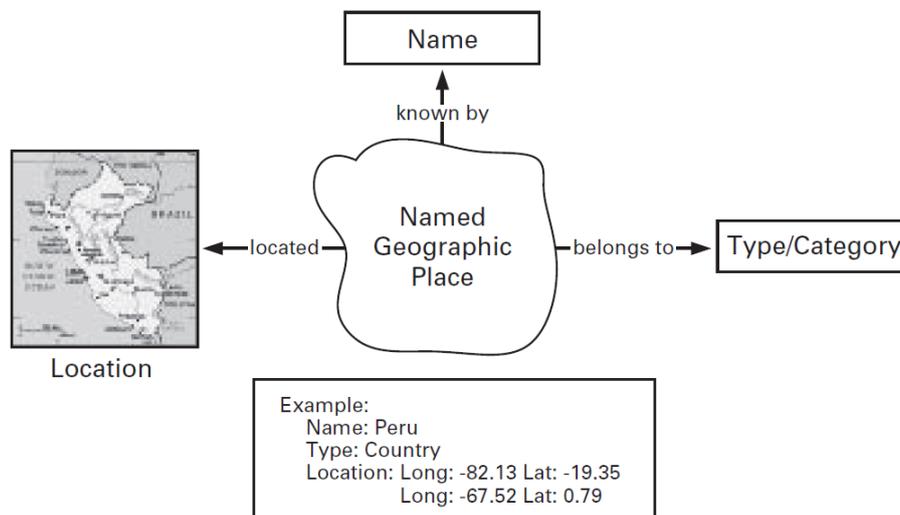


Figure 10. Un lieu nommé dans un index géographique (Hill, 2006)

Référencement par rapport aux autres objets

Ce type de référencement est fait naturellement par l'humain quand il veut retrouver sa position sur une carte. Il cherche des points de repère dans la réalité, puis il essaie de les retrouver dans la carte, et enfin il retrouve sa position sur la carte par rapport à ces points de repère. La même méthode est utilisée pour la saisie d'un accident à 50 m du virage par un

opérateur. Le référencement s'appuie sur la position relative par rapport à d'autres objets (Papadias et Theodoridis 1997; Matsakis et al. 2010 ; Clementini 2013), c'est-à-dire sur des relations spatiales entre objets. L'information de référencement qui a guidé la saisie (i.e. les relations) n'est pas souvent stockée explicitement dans les données même s'il arrive qu'elle le soit. Par exemple, un gestionnaire de routes stocke l'information sur la localisation des accidents par rapport aux points de repère qui sont des objets physiques sur la route (des bornes). L'information stockée est essentiellement la distance de l'accident par rapport à la borne la plus proche. Ce mode de référencement est utilisé lors de la saisie de données collaboratives (Brando 2013, p.68).

A.1.5 Les relations spatiales

En mathématique, le terme *relation* signifie l'association entre les éléments d'un ensemble avec les éléments d'un autre ensemble. En géoscience, les relations spatiales décrivent la position relative et l'orientation relative des objets entre eux. Par exemple, une relation de proximité entre deux bâtiments ou entre un accident et un pont, une relation d'inclusion entre un bâtiment et une zone d'avalanche, l'agencement des objets les uns par rapport aux autres, etc. Il existe une grande variété de relations (Steiniger et Weibel 2007), la Figure 11a montre un exemple de relations entre deux objets et la Figure 11b montre un exemple de relation qui concernent plusieurs objets.

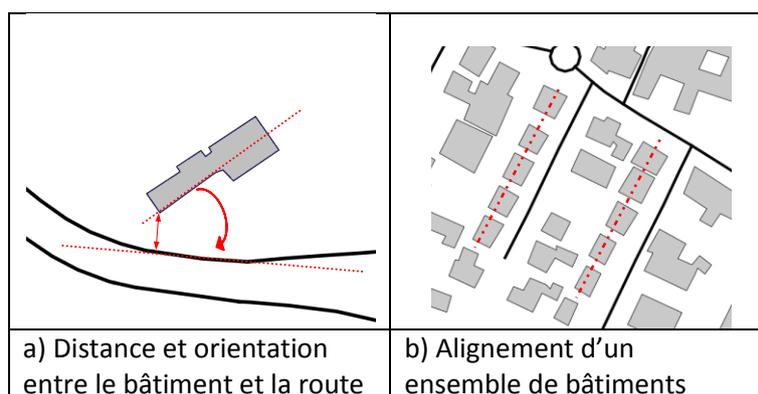


Figure 11. Exemples de relations spatiales

A.1.6 La représentation des relations spatiales

Dans une base de données géographiques, les entités sont représentées par les objets géographiques, mais les relations ne sont le plus souvent pas représentées de manière explicite. Deux entités au minimum sont nécessaires pour modéliser une relation quel que soit le type d'entités considérées (topographiques ou thématiques).

Les chercheurs ont étudié la modélisation des relations spatiales. La modélisation des relations topologiques a notamment été étudiée par Egenhofer et ses collègues (Egenhofer et Franzosa 1991 ; Egenhofer et Herring 1991 ; Egenhofer et Herring 1991 ; Egenhofer et AlTaha 1992 ; Egenhofer et al. 1993 ; Egenhofer et Mark 1995). La modélisation des relations de proximité et de direction ont été étudiées, par exemple, par (Papadias et Theodoridis 1997 ; Goyal et Egenhofer 2000).

De nombreux chercheurs reconnaissent l'importance des relations. Dans (Papadias et Theodoridis 1997) nous trouvons la phrase suivante : « *The representation and processing of spatial relations is crucial in geographic applications because very often in geographic space, relations among spatial entities are as important as the entities themselves.* ».

Mackaness et al. (2013) écrivent « *A well designed map is a silent record of many relationships that exist between the entities mapped* ».

Lors d'une modification de données, par des processus automatiques ou par un opérateur humain, les relations doivent être préservées ou prises en compte. Si un bâtiment qui est en dehors d'une forêt est simplifié, le bâtiment doit rester en dehors de la forêt : la relation topologique entre la forêt et le bâtiment est une information plus importante que la précision de la géométrie absolue du bâtiment.

Les relations représentent une partie importante de l'information géographique, mais elles ne sont pas toujours explicites et stockées comme la géométrie des objets et les attributs. Pour les rendre explicites, il faut recourir à des méthodes d'enrichissement de données (Bobzien et al. 2008)

A.1.7 Importance des relations entre les données thématiques et les données topographiques

Si les données thématiques sont saisies sur une base topographique, l'échelle de saisie et la précision de la géométrie des données thématiques sont dérivées de l'échelle et de la précision des données topographiques. Par exemple, dans le cas d'un accident qui a été saisi à 50 mètres d'un sommet de virage, si la base topographique utilisée a une précision de 5 mètres, la géométrie de l'accident sera encore moins précise (± 5 mètres + erreur de saisie). Si nous voulons décrire la position plus précisément, nous pouvons décrire la position par la relation « l'accident est à 50 mètres du sommet de virage ». Par cette information, nous avons une description plus précise de la réalité parce que nous éliminons l'erreur liée à la précision de la géométrie de la base topographique. Par ailleurs, les relations spatiales sont bien comprises par les utilisateurs. Même pour les besoins d'analyse (par exemple, l'analyse des causes d'accidents) les relations spatiales sont plus importantes que la géométrie. Alors il est préférable d'extraire et stocker la relation « l'accident est à 50 mètres du sommet de virage » à part et de l'utiliser en cas de traitement des données thématiques. Autrement dit, la prise en compte des relations lors du traitement donne un résultat plus précis ou plus représentatif de la réalité. Pour la saisie des données collaboratives, il est mieux de s'exprimer par les relations et les données qualitatives que par la géométrie (Brando 2013 p.83). Par exemple, *l'arrêt de bus est devant la poste* ou *le bâtiment est le plus haut dans le quartier*. Ces relations représentent mieux la localisation que la géométrie seule.

Puisque les données thématiques sont basées sur les données topographiques, les relations entre les données thématiques et les données topographiques ont certaines particularités par rapport aux relations spatiales entre les objets topographiques. Nous identifions deux particularités.

1^{er} particularité: l'importance des relations

Certaines relations sont sémantiquement importantes. L'importance varie selon le cas d'étude, la nature de l'objet thématique et la nature de l'objet topographique. Par exemple, la relation de proximité d'un accident avec une rivière possède a priori moins d'importance que la relation de proximité avec un sommet de virage sur la route.

L'importance est aussi liée à la méthode de saisie. Si l'opérateur qui saisit les accidents sur un réseau routier n'a que les distances à partir du début et de la fin de la route ou à partir des points remarquables sur la route, la relation de distance aura plus d'importance que les autres relations (par exemple, distances avec les bâtiments proches).

2^{ème} particularité: possibilité d'avoir une liaison de type « support » avec les objets topographiques

Pendant la saisie, certains objets topographiques jouent le rôle d'un espace support pour certains objets thématiques. Le lien entre l'objet thématique et son espace support est fort, en effet, l'existence de l'objet thématique est liée à son espace support. Le lien inclut une partie sémantique. Par exemple, les accidents routiers et les lignes de bus sont portés par les routes, un accident sur une route nommée *D27* doit rester sur cette route en cas de modification ou remplacement des données topographiques qui représentent la route.

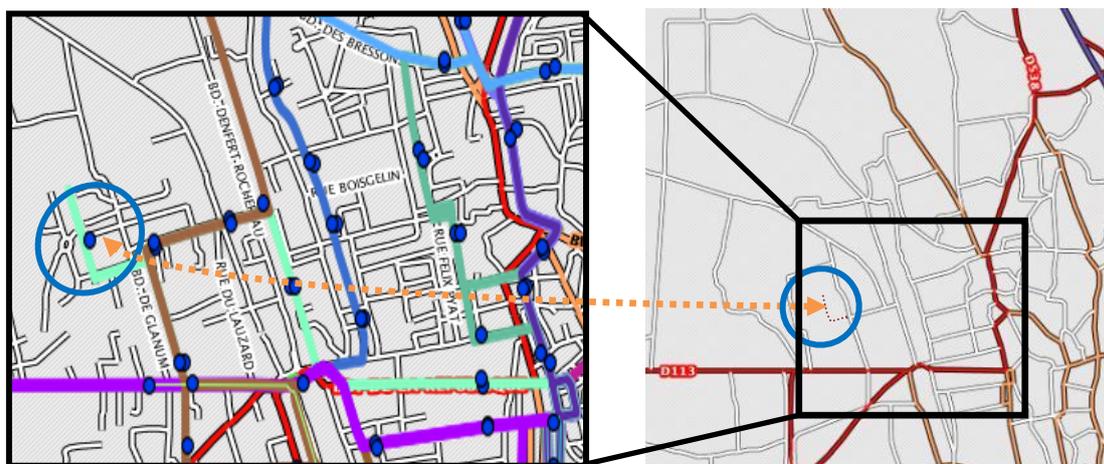
A.1.8 Conséquence de la non prise en compte des relations entre les données thématiques et les données topographiques lors de la modification des données topographiques

Après la création de données thématiques sur les données topographiques, l'utilisateur peut disposer d'une version plus récente des données topographiques, ou vouloir changer l'échelle en faisant une généralisation automatique. Si les relations entre les composantes *thématique* et *topographique* ne sont pas prises en compte, nous constatons souvent une perte de cohérence comme l'illustrent les 4 exemples ci-dessous.

Exemple de changement vers une échelle moins détaillée

En Figure 12a, nous avons un réseau de lignes de bus visualisé sur un réseau routier (nous faisons ici abstraction du décalage entre ces deux réseaux). Le réseau routier a subi une sélection entre les échelles de visualisation 18:000 et 1:32000. Lorsque un utilisateur saisit ses données sur une base topographique avec un niveau de détail plus fin, et qu'il veut visualiser ces données sur une base moins détaillée, il est possible que certains objets qui sont utilisés comme support pour les données thématiques ne soient pas représentés dans la base finale. En Figure 12, une partie d'une ligne de bus est portée par une route qui n'est pas représentée dans la base finale, cela donne un résultat incohérent lié au non respect de la relation *une ligne de bus est toujours portée par une route*. Deux solutions sont possibles, la première solution consiste à ajouter des routes dans la base finale pour une visualisation cohérente, et la deuxième solution est de juger la base comme non valable pour la visualisation des lignes de bus. Les deux solutions nécessitent de connaître les routes qui sont en relation « *porté par* » avec les lignes de bus.

Dans ce cas, la prise en compte des relations nous aide à détecter les problèmes d'incohérence lors de changement du niveau de détail, et à les résoudre.



a) Lignes de bus saisies sur une base topographiques échelle 1 :18000
 b) Base topographiques finale (échelle 1 :32000)

Figure 12. Problème de non existence de certains objets topographiques dans la base finale (source : www.géoportail.fr)

Exemple de changement vers une échelle plus détaillée

En Figure 13, le niveau de détail de la base finale est supérieur. Le chemin sur lequel l'itinéraire passe est bien représenté dans la base finale. Mais nous constatons un problème d'incohérence topologique. L'itinéraire qui a été saisi sur le chemin dans sur la base topographique initiale n'est plus cohérent avec le même chemin dans la base finale. Autrement dit, la relation « porté par » entre l'itinéraire et le chemin est perdue. De ce fait, la combinaison finale est plus difficile à lire suite à la perte de relation topologique entre l'itinéraire et le chemin.

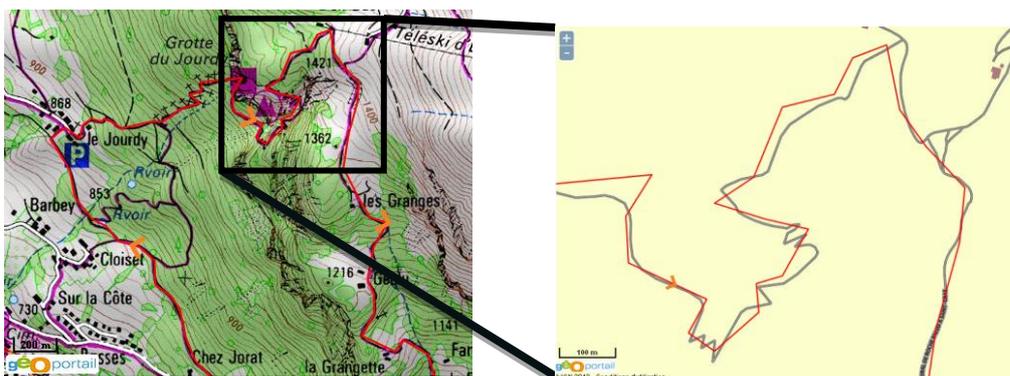


Figure 13. Remplacement d'une base par une autre plus détaillée, sans prendre en compte les relations initiales (données du géoportail de l'IGN site www.visorando.com)

Exemple de changement vers une base avec une géométrie différente

La Figure 14 montre le même problème sur les données métier routières. Les points en rose en Figure 14a sont les points de repères utilisés par un gestionnaire de routes. Afin de relocaliser les points de repères PLO (Points de localisation), le gestionnaire a repositionné les PLO en préservant la même distance curviligne par rapport au début de la route en tenant compte de la différence de longueur de la route dans la base initiale et finale, puisque la simplification affectée sur une route peut affecter sa longueur. Par exemple, considérons un PLO qui se situe dans la base initiale à 1km du début de la route, et considérons que la longueur de la route dans la base initiale est 5km. Si la route dans la nouvelle base a 5.2km de longueur, le PLO sera placé à $\frac{1}{5} \times 5.2 = 1.04$ km de distance du début de la route. Le résultat de l'application de cette méthode est illustré en Figure 14b, le problème est que les PLO ont perdu leur position relative par rapport aux sommets de virage, alors que cette information peut être critique pour les besoins d'analyse.

Le positionnement des PLO est crucial pour le gestionnaire de route parce que tous les objets métier (panneaux, feux, glissière de sécurité, etc.) sont positionnés par le gestionnaire par rapport aux PLO qui sont ses points de repère.

Le chapitre C contient une description plus détaillée de la problématique réelle d'un gestionnaire de route (CETE d' Aix), afin de localiser les accidents et les autres éléments ponctuels sur le réseau routier, et présente la mise en œuvre de notre modèle sur cette problématique.

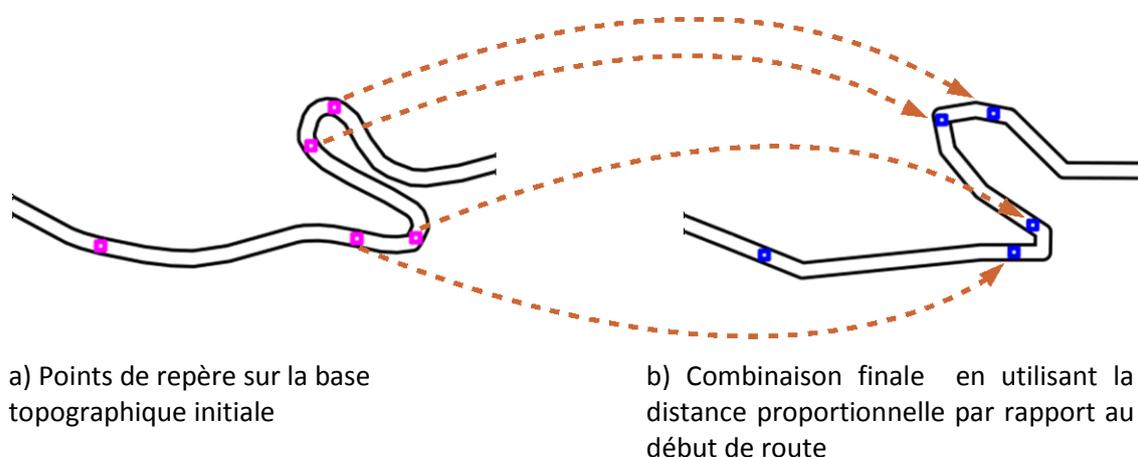


Figure 14. Repositionnement de points de repères routier (source CETE)

Exemple de changement vers une base moins complète

La Figure 15a montre l'exemple d'un itinéraire de randonnée saisi sur une base cartographique, puis visualisé sur une base cartographique différente (Figure 15b). Une partie du chemin qui porte une partie de l'itinéraire n'est pas représentée dans la base finale. La relation topologique avec le chemin a été perdue, et la base cartographique finale n'est pas valide pour le chemin de randonnée parce qu'elle ne contient pas les données topographiques nécessaires pour une visualisation cohérente.

Selon le but de la carte, certains objets topographiques qui ne sont pas en relation avec des objets thématiques ont moins d'importance que ceux qui sont en relation avec lesdits objets. En prenant l'exemple de la Figure 15b, le chemin entouré par l'éclipse en pointillés peut être effacé sans poser un problème crucial pour la localisation de l'itinéraire.

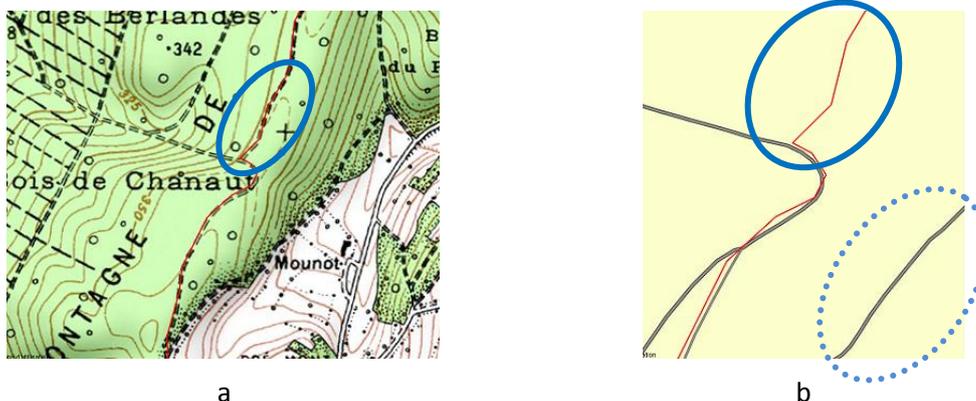


Figure 15. Un remplacement non valide d'une base topographique par une autre à cause d'une perte de relation (données du géoportail de l'IGN site www.visorando.com)

Un des chemins n'a pas été représenté dans la Figure 15. Mais il n'est pas toujours possible de tout représenter, spécialement pour les petites échelles.

Les exemples ci-dessus illustrent la problématique que nous rencontrons lors des traitements liés au changement de données de référence quand il existe des données thématiques qui s'y rattachent.

A.2 Objectif et positionnement

A.2.1 Hypothèse de travail

Dans nos travaux, nous étudions uniquement les données topographiques de type vecteur. Nous nous plaçons dans le cas où les composantes thématiques et topographiques sont cohérentes initialement, cela veut dire que les données thématiques ont été saisies en cohérence avec les données topographiques ou que les données thématiques ont été intégrées avec les données topographiques après la saisie. En cas de la non cohérence initiale, elles doivent être rendues cohérentes grâce à des techniques de conflation (Hauert 2005). La conflation ne fait pas partie des problématiques traitées dans cette thèse.

A.2.2 Rôle des relations lors de traitements liés au changement du niveau de détail

Dans le cas où nous disposons d'un ensemble de couches superposées de données géographiques, et que ces couches contiennent des composantes thématiques et des composantes topographiques, il existe plusieurs traitements possibles pour changer le niveau de détail de l'ensemble. Nous allons exposer deux cas qui nécessitent une bonne prise en compte des relations.

Premier cas : migration des données thématiques après remplacement des données topographiques

Si l'utilisateur dispose d'une nouvelle version de ses données de référence, ou s'il est obligé d'utiliser une nouvelle référence pour fusionner ses données thématiques avec d'autres, ce changement est considéré comme un remplacement des données topographiques par des données topographiques différentes. Pour rendre la combinaison cohérente, il est nécessaire de repositionner les données thématiques sur les nouvelles données topographiques. Nous appelons ce processus *migration des données thématiques* (Figure 16).

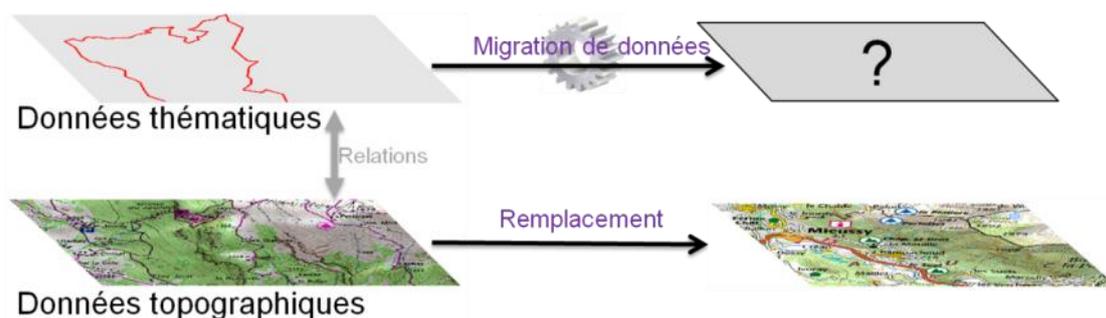


Figure 16. Migration de données thématiques lors d'un remplacement des données topographiques

La migration de données thématiques doit prendre en compte les relations initiales entre les données thématiques et les données topographiques, sinon, on risque d'avoir des incohérences, ou des pertes d'informations, comme illustré précédemment en Figure 14.

En conséquence, nous avons besoin d'un (nouveau) processus de migration de données thématiques qui prenne en entrée les relations initiales entre les données thématiques et les données topographiques (Figure 17). Pour que le résultat de la migration de données thématiques soit cohérent avec les données topographiques finales, il faut découvrir les changements entre les deux bases de données topographiques (initiale et finale). Ce changement est représenté par des liens de correspondance entre les données topographiques initiales et les données topographiques finales. Les liens sont créés en utilisant les techniques d'appariement (cf. B.2).

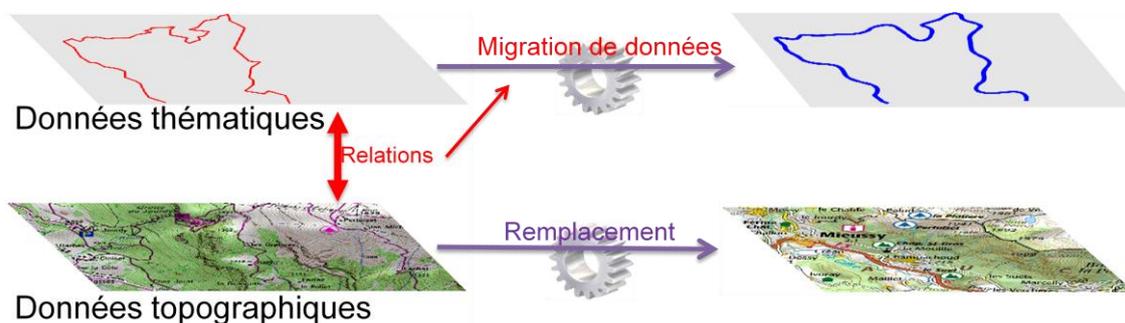


Figure 17. Migration de données thématiques lors d'un remplacement de données topographiques, avec prise en compte des relations initiales

Deuxième cas : la généralisation automatique des données topographiques

Il est possible que les données topographiques finales soient dérivées des données topographiques initiales par un processus de généralisation automatique (Figure 18). Dans ce cas, une migration de données thématiques en prenant en compte les relations sera aussi nécessaire. Une première différence avec le premier cas est qu'il est possible d'obtenir des liens de correspondance entre les données topographiques initiales et finales. Ces liens sont issus du processus de généralisation que nous avons déclenché. Les liens de correspondance issus de la généralisation (s'ils sont disponibles) sont plus sûrs que ceux qui sont issus de l'appariement. En plus, nous disposons d'information sur les traitements qui ont été effectués sur les objets lors de la généralisation. De ce fait, dans le cas de la généralisation, nous disposons d'information plus sûre sur le changement entre les deux bases de données topographiques. Cela aide à améliorer le résultat de la migration de données.

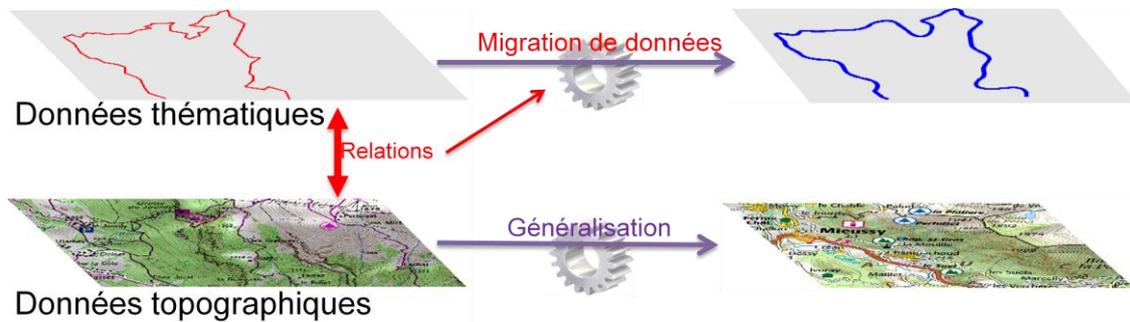


Figure 18. Migration de données thématiques suite à la généralisation de données topographiques, avec prise en compte des relations initiales

Si nous considérons le cas illustré plus haut en Figure 12, la généralisation peut supprimer des objets qui portent des données thématiques. Pour résoudre ce type de problème, il est envisageable de prendre en compte les relations initiales également lors de la généralisation (Figure 19). Par exemple, pour le cas de la Figure 12, la généralisation automatique doit prendre en compte les relations initiales *porté par* entre les lignes de bus et le réseau routier et garder les routes qui portent des lignes de bus.

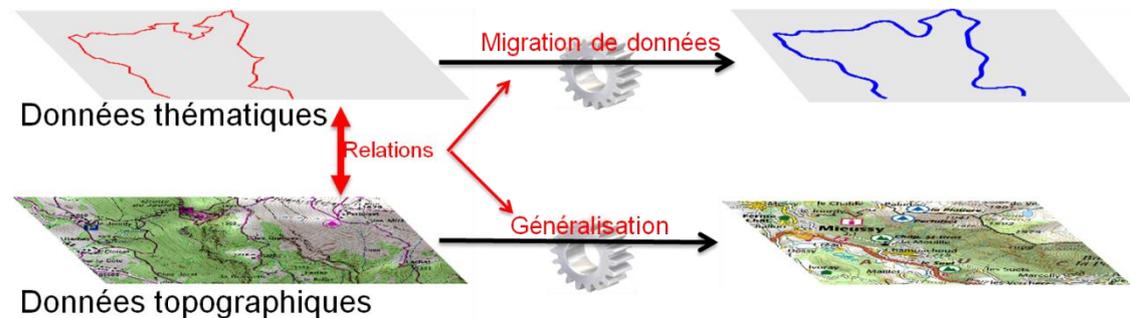


Figure 19. Prise en compte des relations initiales lors de la migration de données thématiques et la généralisation des données topographiques afin de changer le niveau de détail

Dans les deux cas, si le niveau de détail de la base topographique finale est beaucoup plus grossier que le niveau de détail initial et que le niveau de détail des données thématiques est devenu trop fin, il peut être nécessaire d'effectuer également un processus de généralisation sur les données thématiques (Figure 20). La généralisation des données thématiques donne un meilleur résultat en se basant sur les relations. Par exemple, faire un regroupement des accidents en se basant sur les relations métriques, puis représenter les accidents qui sont proches du rond point par un seul objet ponctuel au centre du rond point. Les accidents regroupés seront figurés par des cercles dont la surface correspond au nombre des accidents regroupés.

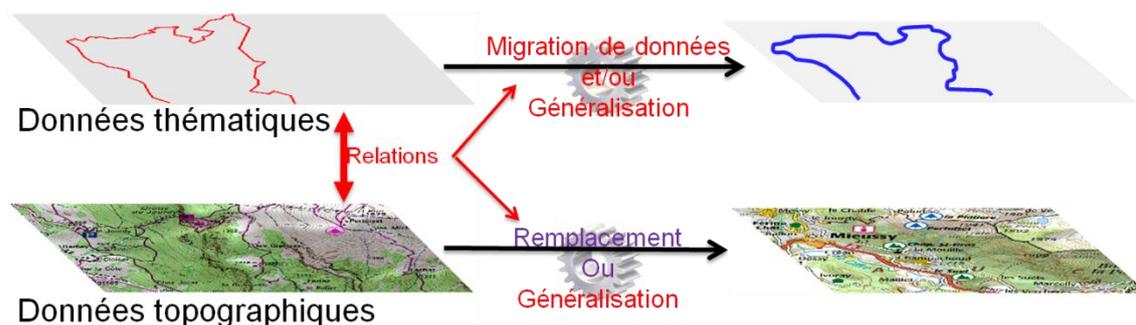


Figure 20. Utilisations possibles des relations pendant le traitement de la combinaison des données thématiques avec des données topographiques lors d'un changement de niveau de détail

A.2.3 Objectif de la thèse

Si nous avons une combinaison cohérente de données thématiques avec des données topographiques, l'objectif de la thèse est de définir une méthodologie pour préserver la cohérence entre les données thématiques et les données de topographiques lors d'un changement de niveau de détail lié à la modification des données topographiques par remplacement ou par généralisation.

Nous nous concentrons d'abord sur la migration des données thématiques en cas de remplacement de données topographiques (premier cas).

Le modèle proposé pour la migration de données thématiques implique les éléments suivants :

- Un modèle pour décrire et stocker les relations, et des méthodes pour instancier ce modèle.
- Une approche pour relocaliser les objets thématiques en prenant en compte les relations initiales modélisées.

Notre approche concerne le cas des données thématiques ponctuelles sur réseau routier. Nous discutons aussi l'extension du modèle proposé pour traiter la prise en compte des relations lors de cas de migration autres que celui de données ponctuelles sur réseau et lors de la généralisation (deuxième cas).

A.3 Etat de l'art

Maintenant que nous avons défini l'objectif de la thèse, nous allons dans cette partie présenter un état de l'art sur des questions en lien avec le sujet.

La création de données thématiques est faite en fonction des relations spatiales dont l'utilisateur dispose lors de la saisie. L'utilisateur saisit les données suivant une méthode de référencement. Pour que les relations soient prises en compte lors d'un changement de niveau de détail, nous identifions trois composantes nécessaires.

- Une modélisation des relations et de leurs propriétés. Par exemple, la relation de proximité possède une propriété qui est la distance entre les deux objets. Un autre attribut peut décrire si la distance est entre les centres des objets ou entre les bords des objets.
- Des méthodes pour extraire les relations, afin de détecter la présence ou la non-présence des relations et calculer les différentes propriétés des relations. Pour la relation de proximité, l'extraction de la relation implique une recherche dans l'entourage de chaque objet dans une distance liée à l'échelle. Quand une relation est trouvée, la méthode calcule la distance qui est la propriété principale de la relation. La modélisation et l'extraction sont liées.
- Une adaptation des traitements liés au changement de niveau de détail (migration de données thématiques et généralisation) pour que ces relations extraites et modélisées soient prises en compte. Par exemple, l'adaptation de la généralisation automatique pour interdire l'élimination d'une route qui porte une ligne de bus.

Dans notre cas, les données à référencer sont toujours des données thématiques. Pour extraire les relations de référencement de ces données sur les données topographiques, nous donnons un panorama sur les types de relations (A.3.1), puis de différents types de modélisations des relations proposées par les chercheurs (A.3.3 et A.3.4). Nous présentons une vue sur des modèles existants, pour représenter différents types de relations (A.3.5). Enfin, nous présentons quelques méthodes utilisées actuellement pour prendre en compte des relations dans des processus de généralisation (A.3.6), d'appariement (A.3.7), d'intégration (A.3.8) et de migration de données (A.3.9).

A.3.1 Classifications des relations

Dans la littérature, plusieurs classifications des relations spatiales ont été proposées, par exemple (Egenhofer et Franzosa 1991 ; Jones 1997), dont certaines dans le contexte de généralisation (Ruas 1999; Steiniger et Weibel 2007). Ces classifications sont résumées dans le tableau suivant :

Auteur	Classification
(Egenhofer et Franzosa 1991)	topologique, métrique, ordre partiel
(Jones 1997)	topologique, direction, proximité
(Ruas 1999)	topologique, angle, voisinage
(Steiniger et Weibel 2007)	topologique, géométrique, statistique, sémantique, structurelle

Tableau 1. Classifications de relations

Dans la suite, Nous allons donner les définitions des classes présentes dans le tableau.

Relations topologiques

En mathématiques, la topologie est définie comme «l'étude sur les surfaces et dans l'espace usuel des propriétés conservées par une déformation continue et liées à la notion de voisinage» (Bouvier et George 1992). Les relations topologiques décrivent l'adjacence, l'intersection et l'inclusion. Par exemple, les 8 cas modélisés dans le modèle de 4IM de (Egenhofer et Franzosa 1991) pour les polygones sont : *disjoint, touche, chevauche, égal, couvert par, à l'intérieur, couvre, contient*.

Les relations topologiques sont importantes dans notre cas parce que la relation *porté par* se traduit souvent par une relation topologique particulière (inclusion, égalité).

Relations métriques

Les relations métriques décrivent la différence en direction et en distance entre deux objets.

Relations d'ordre partiel

Ces relations décrivent la direction de l'objet par rapport aux autres objets. Par exemple, les relations *en face de, devant, derrière et au dessous*.

Relations de proximité et relations de direction

Les relations de proximité et les relations de direction sont deux sous-types de relations métriques. Les relations de proximité concernent les distances et les relations de direction ont deux sous-types : relation de direction relative (par exemple, en face de) et relation de direction absolue (par exemple, au nord de).

Relations géométriques

Les relations géométriques décrites par (Steiniger et Weibel 2007) sont liées à la différence des propriétés géométriques entre les objets, les différences ont 4 catégories : la taille, la position, la forme et l'orientation.

Relations d'angle

Les relations d'angle représentent l'écart angulaire, ou orientation relative entre les objets.

Relations de voisinage

Pour (Ruas 1999), les relations de voisinage incluent le fait qu'un objet est voisin d'un autre, et elles incluent aussi une description de la proximité entre les objets en relations (i.e. comment les objets sont voisins). Ce type de relation est aussi important dans notre cas afin de référencer un objet thématique par rapport à des objets topographiques voisins.

Relations de statistiques et de densité

Les relations statistiques décrivent la distribution spatiale d'un groupe d'objets. L'information statistique est généralement utilisée dans les cartes thématiques, plus particulièrement pour l'analyse des structures (en anglais *pattern analysis*).

Relations sémantiques

Les relations sémantiques mettent en relations les objets en fonction de leur sémantique (type, valeur attributaire, sens de la présence de l'objet à cet endroit): Les types principaux des relations sémantiques définies par (Steiniger et Weibel 2007) sont les suivants :

1. Relations de similarité : deux objets ont les mêmes caractéristiques attributaires
2. Relations de priorité
3. Relations au cas par cas (en anglais *Casual relations*) : décrivent les dépendances entre les types de données. Par exemple, les dépendances des bâtiments autour des routes avec les routes. Ce type de relation est appelé *relation fonctionnelle* par (Touya 2011, p.63).

Relations structurelles

Les relations structurelles décrivent les modèles structurels qui peuvent être perçus dans les données : par exemple, quelle classe d'objets est devant et quelle classe est mise au fond de la carte, la relation d'alignement d'un groupe de bâtiments.

A.3.2 Taxonomie des relations

Une taxonomie des relations a été proposée par (Touya et al. 2012). La taxonomie est dérivée de plusieurs classifications (Egenhofer et Franzosa 1991;Steiniger et Weibel 2007; Bucher et al, 2012). En plus des types des relations, la taxonomie (montrée en

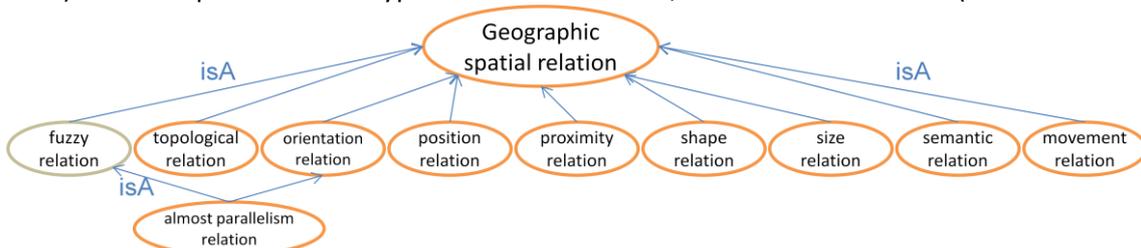


Figure 21) contient le terme *fuzzy* (en anglais) ou *floue* (en français). Une relation peut être décrite comme floue pour représenter sa clarté. Nous détaillons la description *flou* en (A.3.4, p.33).

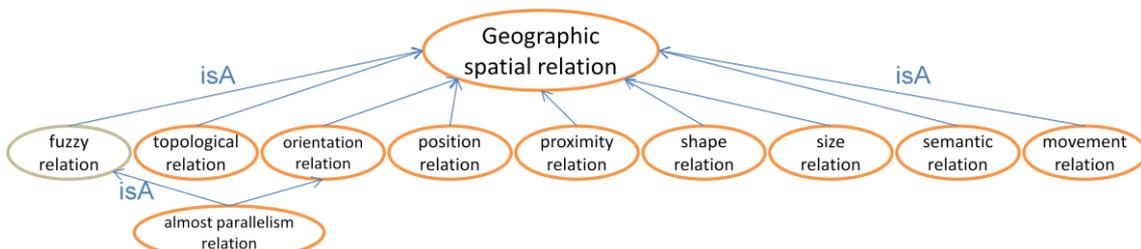


Figure 21. Taxonomie pour les relations spatiale entre deux entités géographiques (Touya et al. 2012)

Nous avons expliqué dans le paragraphe précédent les relations topologiques, les relations de proximité et les relations sémantiques. Pour comprendre les autres types de relations, la Figure 22 montre un exemple de chaque type non détaillé dans le paragraphe précédent. L'exemple Figure 22a montre une relation topologique floue, la Figure 22b montre une relation d'orientation relative, la Figure 22c montre une relation d'orientation relative floue, la Figure 22d montre une relation de position, la Figure 22e montre une relation de forme, la Figure 22f montre une relation de taille, la Figure 22g montre une relation de mouvement.

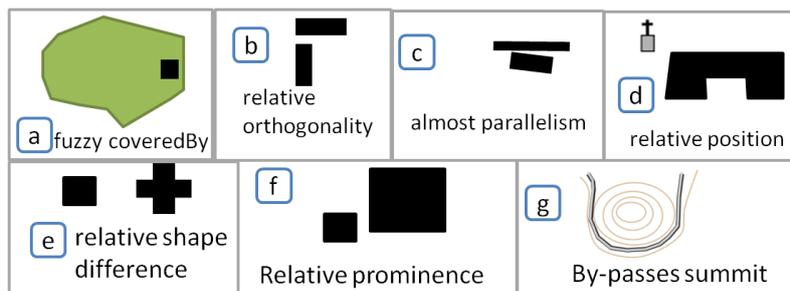


Figure 22. Exemples de relations spatiales (Touya et al. 2012)

A.3.3 Description des relations en langage formel vs en langage naturel

Nous voulons préciser que les descriptions des relations que nous utilisons dans cette thèse sont formelles. Cela veut dire qu'elles sont décrites par un ensemble limité de symboles et mots clés. Par exemple, la maison est à *l'intérieur* de la forêt. La signification des mots-clés doit être décrite, par exemple, à *l'intérieur de* peut signifier que le bâtiment est complètement à l'intérieur de la forêt. Une autre façon pour décrire les relations est d'utiliser le langage naturel. Cette description est composée de phrases utilisées par les gens selon leur perception de l'espace. Par exemple, le siège de l'entreprise *se situe à* Paris. Il est possible que le siège ne soit pas exactement dans Paris, mais la personne qui parle habite en province et selon sa perception la localisation de l'entreprise est à l'intérieur de Paris. Dans le cas des données collaboratives selon le modèle de (Brando 2013, p17), les collaborateurs peuvent saisir les relations éventuelles de leurs données en langage naturel.

Les relations en langage naturel doivent être transformées en relations formelles afin de pouvoir être utilisées dans les manipulations. Cela nécessite d'étudier la perception humaine et la signification des vocabulaires pour trouver la description formelle des relations. Certains chercheurs ont étudié la transformation des relations en langage naturel vers des relations formelles (Egenhofer et al. 1998 ; XU 2007 ; Kordjamshidi et al. 2011).

Dans notre étude, nous utilisons uniquement les relations décrites formellement. Dans la suite, le terme relation spatiale sous-entend une relation formelle.

A.3.4 Comment les relations sont décrites (d'une façon formelle)

Il existe plusieurs types de descriptions pour les relations spatiales. Les classifications diffèrent selon l'utilisation et le point de vue. Dans ce paragraphe, nous présentons certaines descriptions qui concernent les relations. Ces descriptions sont issues de la littérature.

Représentation binaire et quantitatives des relations

La représentation des relations spatiales peut être quantitative ou binaire (Touya et al. 2012). Les relations quantitatives sont représentées par des mesures qui sont les propriétés de ces relations. Par exemple, la relation de proximité possède une mesure de distance. Les relations binaires ont des conditions d'existence : elles existent ou pas selon la satisfaction de leurs conditions. Par exemple, la relation topologique *disjoint* entre deux régions existe si aucune intersection n'existe entre les deux régions. Les relations binaires peuvent être mutuellement exclusives, par exemple, deux régions ne peuvent pas être *disjointes* et *s'intersecter* en même temps. Pour prendre en compte cette réalité, les relations binaires peuvent être regroupées en familles de prédicats. Les objets ne peuvent pas avoir deux relations d'une seule famille à la fois.

Par exemple, la relation topologique entre deux régions peut être représentée par une famille de 8 prédicats (*disjoint*, *touche*, *chevauche*, *égal*, *couvert par*, *à l'intérieur*, *couvre*, *contient*). Deux régions ne peuvent pas être en relation *disjoint* et en relation *touche* en même temps.

Parmi les relations binaires, nous trouvons toutes les relations topologiques et d'autres relations, comme la relation est *parallèle à* entre la route et un bâtiment. Le modèle de (Touya et al. 2012) est présenté en détail en B.3.5.

Représentation des relations floues

Il existe des cas où il n'est pas clair si une relation binaire est vérifiée ou non (par ex. un bâtiment est *presque parallèle à* une route). Dans ce cas, la relation est considérée comme floue. Une relation binaire peut être floue quand elle n'est pas complètement présente (Duchêne et al. 2012). Par exemple, si deux objets surfaciques sont séparés par une distance trop petite en fonction de l'échelle, par exemple, il peut y avoir un doute entre deux relations *disjoints* ou *se touchant*. Aussi, la représentation d'une relation floue peut être définie en lien avec chaque relation binaire (ou en lien avec deux relations binaires dans le cas des familles de relations). Par exemple, *flou entre disjoint et se touchant*, un autre exemple est la relation *presque parallèle* (Figure 23).

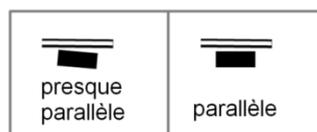


Figure 23. Relation floue et binaire

Relations définies au niveau des types et au niveau des instances

Les relations spatiales existent au niveau d'instances (le bâtiment A est en face du bâtiment B), et elles peuvent aussi être au niveau des types (les bâtiments sont disjoints avec les routes). Les relations spatiales au niveau d'instances peuvent être généralisées au niveau de type dans certaines conditions (Schwering 2006, p.43), cette généralisation a été formalisée par (Donnelly et Bittner 2005). Par exemple, si les instances d'un type donné d'objets ponctuels sont toujours sur les routes, une question peut être posée à l'utilisateur pour confirmer que la relation topologiques entre les objets ponctuels et les routes est valable pour tout le jeu de données thématiques, si oui, une nouvelle relation de type peut être créée à la place des

relations d'instances individuelles. Les relations issues de la sémantique des données sont au niveau des types (ex. un arrêt de bus est sur une ligne de bus). Les contraintes d'intégrité, par exemple celles définies par (Brando 2013, p.68) pour intégrer les données collaboratives, sont aussi au niveau des types.

Dans notre cas, la détection des relations de type permet d'optimiser certains traitements.

A.3.5 Modélisation des relations

Dans le paragraphe précédent, nous avons présenté certaines classifications pour les relations spatiales. Dans ce paragraphe, nous explorons une partie des travaux qui portent sur la modélisation de ces relations avant d'étudier l'utilisation des relations dans différents traitements dans les paragraphes suivants. Pour nos besoins, nous classons les relations en quatre catégories : topologique, proximité, direction et autre. Cette classification suit la taxonomie de (Touya et al. 2012), la Figure 24 montre les types correspondants de notre classification avec la taxonomie.

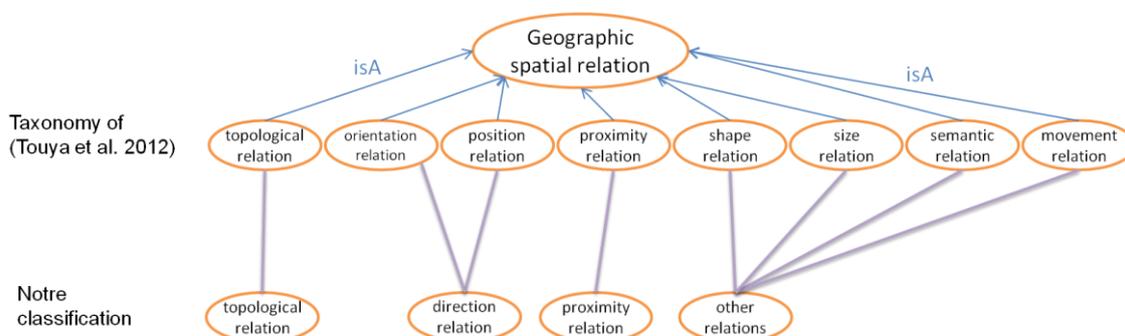


Figure 24. Correspondance de notre classification avec la taxonomie de (Touya et al. 2012)

Nous avons déjà défini les relations topologiques, les relations de proximité et les relations de direction dans le paragraphe A.3.1. Les autres relations sont les relations qui n'entre pas dans les trois premières catégories. Par exemple, la description de la position relative d'un accident le long de la route qui le porte en décrivant la forme de la route autour l'accident.

Nous étudions les modèles existants parce que nous allons les utiliser dans notre solution pour la prise en compte des relations. Ils seront utilisés dans notre modèle de référencement des données thématiques sur les données topographiques.

Nous allons présenter les modèles existants en quatre groupes : relations topologiques, relations de proximité, relation de direction et autres relations.

La modélisation des relations topologiques

Une base de données géographique décrit l'espace par la position des objets (coordonnées). Les relations topologiques contiennent des descriptions additionnelles sur l'inclusion et l'adjacence entre objets.

Les premiers modèles qui sont largement utilisés pour les relations topologiques sont les modèles 4-intersections (4IM) et 9-intersections (9IM) d'Egenhofer et ses collègues.

- Selon le modèle 4IM (Egenhofer et Franzosa 1991), les relations sont modélisées par une matrice de 4 valeurs binaires. La matrice représente la présence d'intersections non vides entre les intérieurs et les frontières des objets.
- Le modèle 9IM (Egenhofer et Herring 1991) considère l'intersection des intérieurs, des frontières et des extérieurs. Les relations topologiques sont représentées en utilisant 9 valeurs v (Tableau 2). Pour chaque valeur v , il y a deux possibilités :
 - Non vide ($\neq \emptyset$) : Il existe une intersection
 - Vide (\emptyset) : n'y a pas d'intersection.

Ces modèles sont utilisés pour les relations topologiques ligne-région, région-région et ligne-ligne. La Figure 25 montre les 8 relations topologiques entre deux régions selon le modèle 9IM. Une comparaison entre les deux modèles a été présentée dans (Egenhofer et al. 1993).

		b		
		Intérieur	Frontière	Extérieur
c	Intérieur	v_{ii}	v_{if}	v_{ie}
	Frontière	v_{fi}	v_{ff}	v_{fe}
	Extérieur	v_{ei}	v_{ef}	v_{ee}

Tableau 2. Les 9 valeurs du modèle 9IM (Egenhofer et Herring 1991)

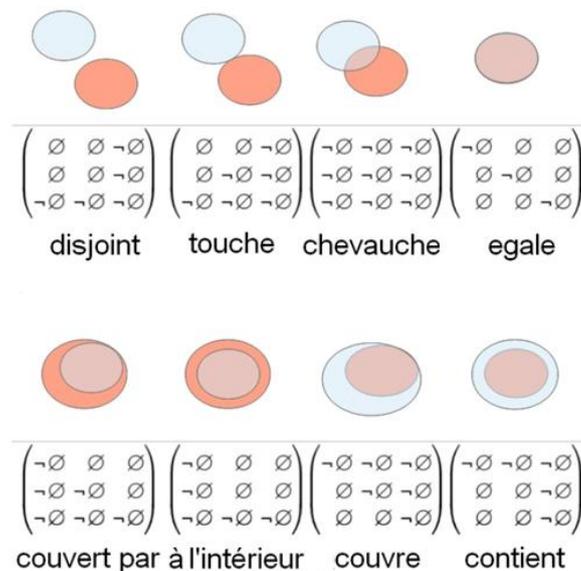


Figure 25. Les relations topologiques entre deux régions selon le modèle 9IM (Egenhofer et Herring 1991)

Il existe plusieurs extensions du modèle. Ces extensions ont été proposées par des chercheurs pour répondre à des besoins spécifiques:

- Le modèle DE-9IM par (Clementini et al. 1993) ajoute la dimension dans la matrice des intersections: dans ce modèle, les valeurs dans la matrice peuvent être
 - F :pas d'intersection
 - 0 :l'intersection est un point
 - 1 :l'intersection est une ligne
 - 2 :l'intersection est une surface

Cela permet de distinguer plus de cas. La description détaillée de ces cas est présentée dans (Clementini et Di Felice 1995). Le modèle DE-9IM est devenu un standard OGC pour représenter les relations topologiques.

- Le modèle DE-9IMBC par (Zhong et al. 2004) est une version développée du modèle DE-9IM afin de représenter les relations topologiques des objets qui ont plusieurs types de géométries. Par exemple, un objet qui est composé d'une région et une ligne. Le modèle conçu par (Kurata 2009) utilise une matrice de 3X9 valeurs pour donner une description plus fine des relations topologiques. Les relations topologiques entre les objets complexes (ex. des objets composés de plusieurs lignes ou plusieurs polygones) sont étudiées par (Schneider et Behr 2006).

Pour un modèle donné, les relations topologiques distinguées par ce modèle peuvent être organisé dans un graphe conceptuel de voisinage. La Figure 25 montre les 8 prédicats de la relation topologique entre deux régions, et leur graphe conceptuel de voisinage (Egenhofer et Al-Taha 1992). Le graphe de voisinage représente quelles situations peuvent être atteintes directement par transformation continue à partir de quelles situations initiales pour les relations topologiques pendant des traitements. La Figure 26 contient le graphe de voisinage pour les relations topologiques entre deux régions. Le graphe de voisinage des relations topologiques entre les objets linéaires a été étudié par (Reis et al. 2008) et pour les relations topologiques ligne-région par (Egenhofer et Mark 1995).

Dans notre cas, le graphe de voisinage nous aidera à identifier le degré de similarité en terme de relations topologiques entre l'état initial et un état final après migration de données thématiques que nous voulons évaluer.

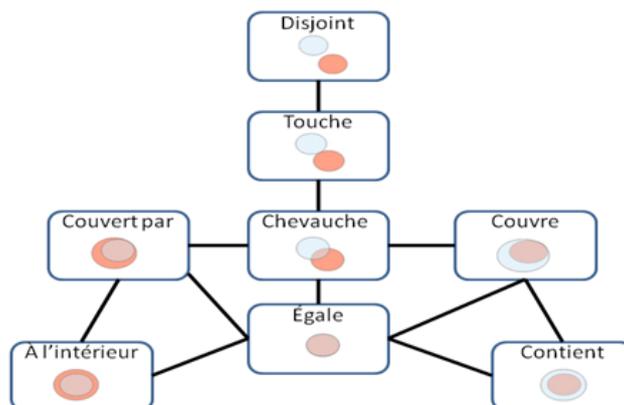


Figure 26. Graphe conceptuel de voisinage des relations topologiques entre deux régions (Egenhofer et AlTaha 1992)

Pour représenter les relations qui ne sont pas nettes, le graphe de voisinage a été développé par (Touya et al. 2012) pour y intégrer les relations floues (Figure 27).

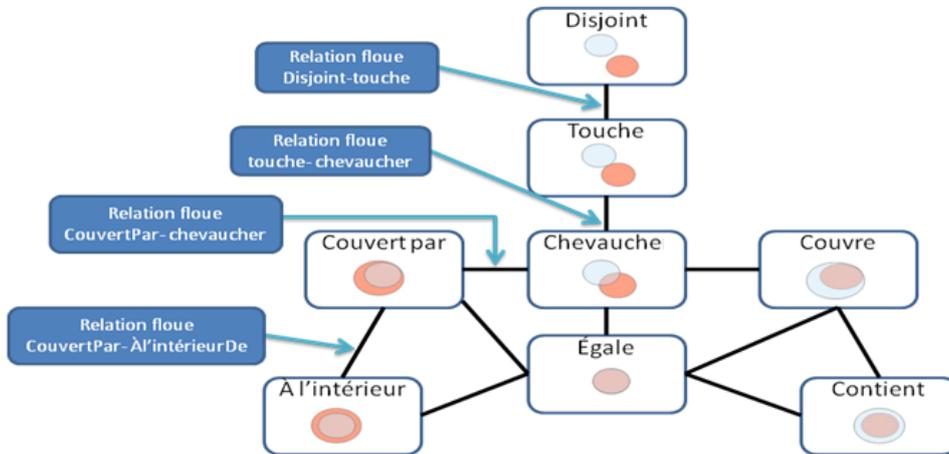


Figure 27. Extension du graphe de voisinage conceptuel pour représenter les relations floues (Touya et al. 2012)

La modélisation des relations de proximité

Les *relations de proximité* sont celles qui sont basées sur des distances (Egenhofer et Franzosa 1991). Elles sont souvent modélisées en utilisant la distance comme mesure principale. La Figure 28 montre des relations de proximité en montrant la distance à calculer pendant l'extraction de ces relations. La distance peut être la distance minimum entre les deux objets ou la distance maximum ou la distance de Hausdorff. La distance peut être avec le barycentre de l'objet, ou avec la bordure. Elle peut être mesurée de manière curviligne entre deux points sur une route donnée.

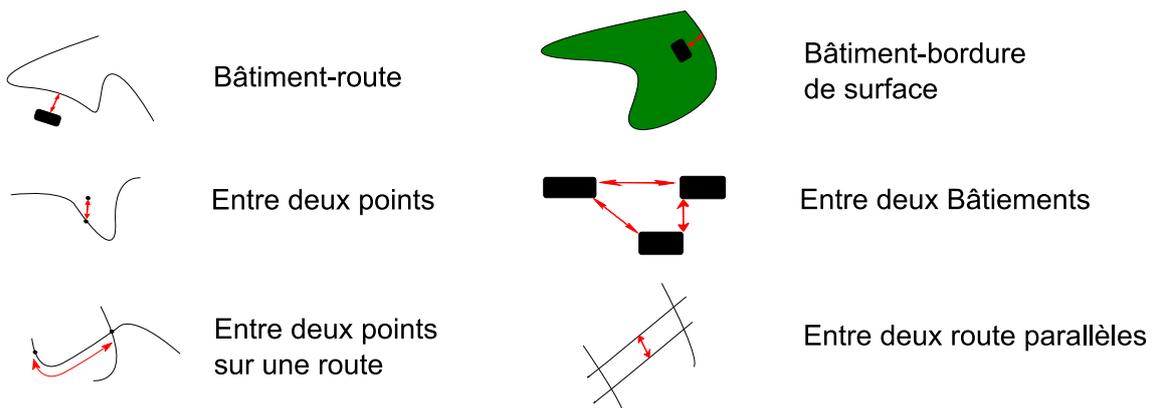


Figure 28.Exemples de relations de proximité

La modélisation des relations de direction

Les *relations de direction* décrivent la position d'un objet par rapport à un autre, elles ne sont pas symétriques, nous avons toujours un objet à référencer par rapport aux objets de

référence autour de lui (Clementini 2013). Dans ce groupe, nous distinguons deux sous groupes : les relations d'orientation relatives (ex. angle entre un bâtiment et la route) et les relations de position relative (ex. le bâtiment est au sud du carrefour).

La direction peut être décrite par une mesure de la différence d'orientation. La différence peut être modélisée par un angle et elle peut être modélisée comme une relation binaire, par exemple : parallèle ou pas parallèle, orthogonal ou pas orthogonal. Une relation d'orientation peut inclure l'angle comme une mesure principale (Figure 29). Par exemple, l'angle d'un bâtiment par rapport à une route. Lors de la modélisation, il faut choisir où nous faisons la mesure d'angle entre les deux objets en relations. Dans la Figure 29, l'angle est calculé entre la portion de la route la plus proche du bâtiment et l'orientation principale du bâtiment.

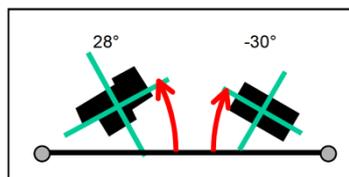


Figure 29. Orientation relative d'un bâtiment par rapport à un objet linéaire (Duchêne 2004)

Les relations de direction peuvent être représentées dans le domaine géographique par un triplet $\langle A, d, B \rangle$ où A et B sont l'objet de référence et l'objet cible (Frank 1991 ;Goyal et Egenhofer 2000). Le d est un sous ensemble non vide de ces valeurs : {N, S, E, W, NE, SE, SW, NW, 0}. Les régions correspondant aux différentes valeurs sont définies dans un repère local basé sur le rectangle minimum englobant de l'objet A.

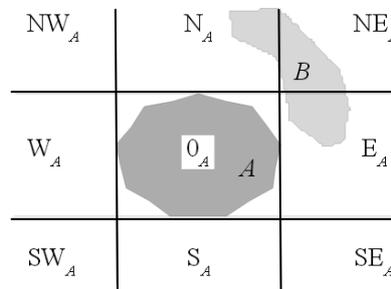


Figure 30. Relation cardinale d'orientation entre deux régions (Goyal et Egenhofer 2000)

Selon la même modélisation, la relation cardinale de direction décrite par l'orientation entre les deux régions A et B est représentée par la matrice suivante :

$$dir_{RR}(A, B) = \begin{bmatrix} NW_A \cap B & N_A \cap B & NE_A \cap B \\ W_A \cap B & 0_A \cap B & E_A \cap B \\ SW_A \cap B & S_A \cap B & SE_A \cap B \end{bmatrix}$$

$$\text{Dans l'exemple de la Figure 30 : } dir_{RR}(A, B) = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Le même principe peut être utilisé pour les relations d'orientation point-région, point-line, etc.

Une approche basée sur les *rectangles minimums englobants* a aussi été utilisée par (Papadias et Theodoridis 1997) pour modéliser les relations de position relative entre les objets. La Figure 32 montre les relations possibles.



Figure 31. Raisonnement rectangle à minimum englobant pour décrire la relation de direction (Papadias et Theodoridis 1997)

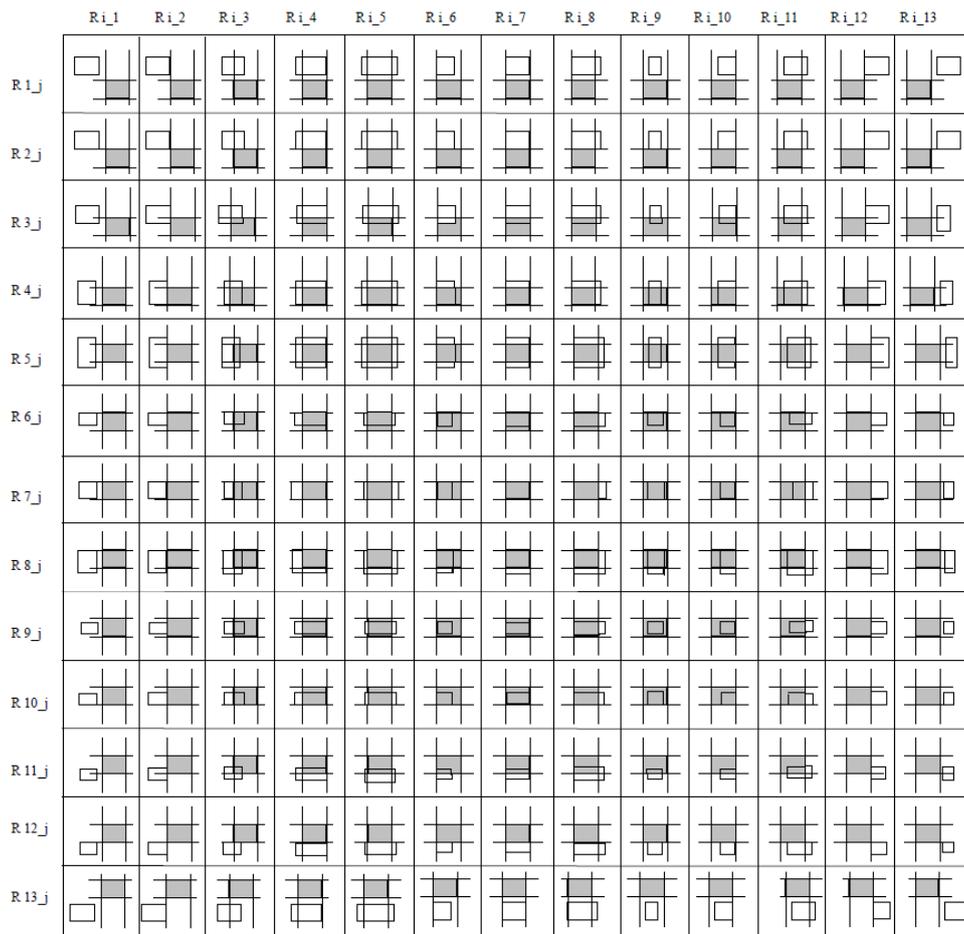


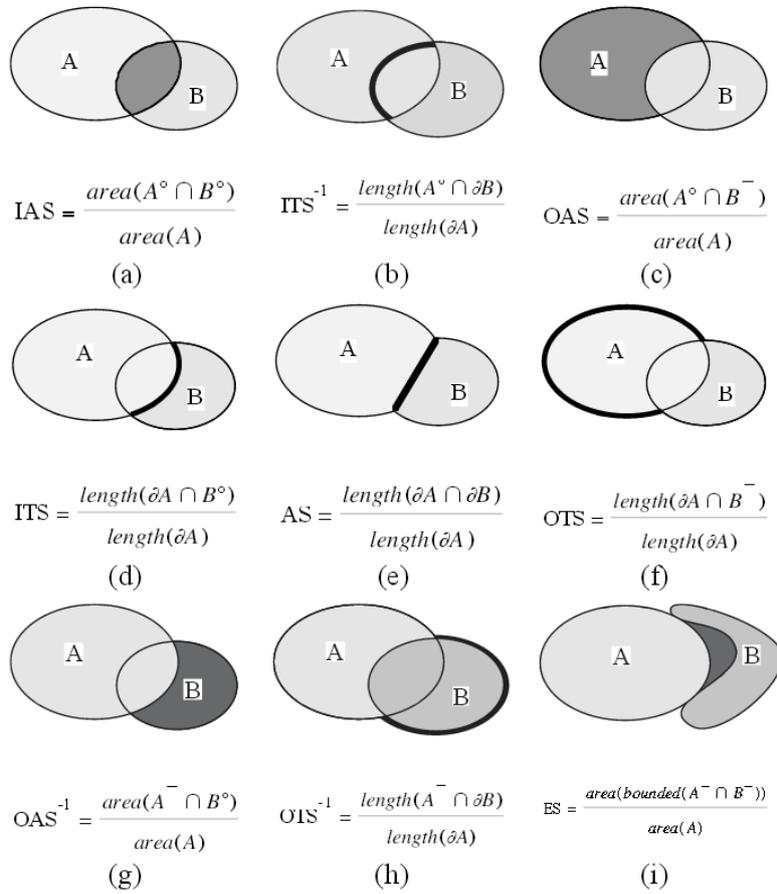
Figure 32. Relations de direction possibles en utilisant le Rectangle minimum englobant (Papadias et Theodoridis 1997)

Raffinements métriques pour les relations topologiques

Pour avoir une description précise des relations topologiques, nous pouvons faire des raffinements métriques. Ce sont des relations qui donnent une description métrique détaillée de la position d'un objet par rapport à un autre. Nous allons utiliser ce raffinement dans notre modèle de référencement des données thématiques sur les données topographiques. Selon (Egenhofer et Dube 2009), deux types de mesure peuvent être calculées pour raffiner les relations topologiques.

Le premier type de mesure est *la mesure de division* qui décrit le pourcentage des parties d'un objet intersectées ou alignées avec l'autre objet. La Figure 33 contient les mesures de divisions pour les relations topologiques entre deux régions. Les mesures ne sont pas symétriques: elles servent à référencer l'objet *B* (dans notre cas, un objet thématique) par rapport à l'objet *A* (dans notre cas, un objet topographique). Les 9 mesures de division sont organisées dans une matrice de 9 valeurs (Figure 33j), où, pour une région *X*, le symbole de l'intérieur est (X°) , le symbole de la frontière est (∂X) et le symbole de l'extérieur (X^-) .

Abréviation	Signification en anglais
IAS	Inner Area Splitting
OAS	Outer Area Splitting
OAS^{-1}	Inverse Outer Area
ES	Exterior Splitting
ITS	Inner Traversal Splitting
ITS^{-1}	Inverse Inner Traversal Splitting
OTS	Outer Traversal Splitting
OTS^{-1}	Inverse Outer Traversal Splitting
AS	Alongness Splitting



	B°	∂B	B^-
A°	IAS	ITS^{-1}	OAS
∂A	ITS	AS	OTS
A^-	OAS^{-1}	OTS^{-1}	ES

(j)

Figure 33. Mesures de divisions qui concernent les relations topologiques entre deux régions (Egenhofer et Dube 2009)

Le même principe peut être utilisé pour les relations topologiques ligne-région et ligne-ligne.

Le deuxième type de mesures métriques pour raffiner les relations topologiques est *la mesure de proximité*. Une mesure de proximité ne concerne pas seulement les objets disjoints : En Figure 34, les mesures de proximité définies par (Egenhofer et Dube 2009) sont montrées,

lesquelles une mesure de proximité entre deux régions dans le cas où l'objet à référencer est à l'intérieur de l'autre. Les mesures sont :

EC : la proximité de l'extension, l'extension de A pour que A touche B

CC : la proximité de contraction, la contraction de A pour que A touche B

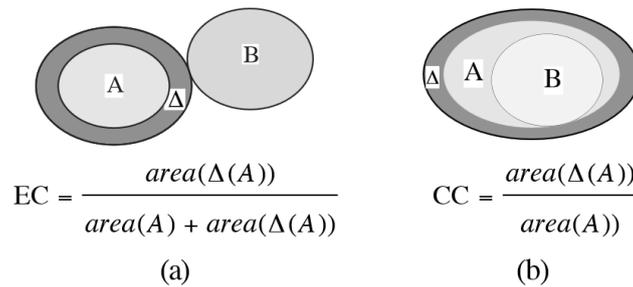


Figure 34. Mesures de proximité entre deux régions (Egenhofer et Dube 2009)

Dans ce paragraphe, nous avons donné une vue globale de la littérature sur les relations spatiales et leur modélisation en lien avec notre problématique.

Certains traitements sur les données géographiques ont été définis de manière à prendre en compte les relations spatiales entre les objets traités. Dans le paragraphe suivant, nous montrerons l'utilisation des relations dans plusieurs types de traitements en focalisant sur la partie qui concerne les relations au sein de ces traitements.

A.3.6 Prise en compte des relations dans la généralisation

Les mesures spatiales pour la généralisation

Les algorithmes de généralisation utilisent des mesures spatiales pour guider leur fonctionnement. Le but des mesures est de représenter les caractéristiques spatiales des objets. Parmi les mesures, nous pouvons citer : l'analyse des formes des routes (Barillot et Plazanet 2002), la détection de sommet de virages (Plazanet 1996), la détection d'empâtement (Mustière 2005), la forme d'un virage, l'indicateur de forme d'un bâtiment, l'orientation d'un bâtiment. Ces mesures sont utilisées aussi pour le guidage de la généralisation. Le guidage implique le choix des algorithmes et l'ordre dans lequel les algorithmes seront exécutés et sur quelles régions ou sur quels types d'objets (Ruas et Plazanet 1996). Plusieurs systèmes d'orchestration et guidage de la généralisation ont été élaborés. Dans le modèle élaboré par (Touya 2011), l'espace géographique est divisé en régions ayant des caractéristiques spatiales homogènes, ensuite, le système guide la généralisation en fonction des caractéristiques de chaque région.

Relations spatiales et contraintes sur ces relations en généralisation

Dans notre problématique, nous devons traiter le fait d'avoir un ensemble de relations initiales à prendre en compte. Le cas de la généralisation implique une gestion développée des contraintes relationnelles (e.g. un bâtiment ne se situe pas sur une route) et non relationnelles (e.g. une taille minimum pour les bâtiments).

Selon le modèle AGENT proposé initialement par (Ruas 1999) qui est un modèle multi-agent pour la généralisation topographique, les objets de la base à généraliser sont organisés en trois niveaux :

- le niveau micro : par exemple, un bâtiment.
- le niveau meso composé de groupes d'objets: par exemple, un îlot composé de bâtiment, une ville composée d'îlots.
- le niveau macro qui regroupe la totalité des objets d'un type : par exemple, les routes.

Dans le modèle AGENT, les contraintes sous-tendant la généralisation portent sur un objet (micro), un groupe (objet meso), ou une population (objet macro). Les relations ne sont pas explicitement représentées, même si elles peuvent être calculées et prises en compte implicitement à l'intérieur de l'algorithme. Les relations dans le modèle AGENT sont modélisées par des propriétés d'objets de type meso. Pour représenter une relation entre deux objets micro, il faut créer un objet meso qui contient les deux objets micro puis assigner la propriété qui correspond à la relation. Le travail de (Duchêne 2004), porte sur l'extension de modèle pour mieux généraliser les données topographiques dans les zones rurales. Elle a proposé d'exprimer les relations entre deux objets directement au niveau micro. Dans le modèle CartAcom, les contraintes relationnelles sont explicites et classées en trois types :

- Le premier type concerne la lisibilité des objets (Figure 35). Ce type inclut aussi l'interdiction de l'ambiguïté des relations pour faciliter la lecture. Par exemple, si un bâtiment est presque parallèle mais avec un faible angle avec la route, la généralisation doit ajuster le bâtiment soit pour qu'il soit parallèle ou pour qu'il ne soit pas parallèle du tout en tournant le bâtiment.

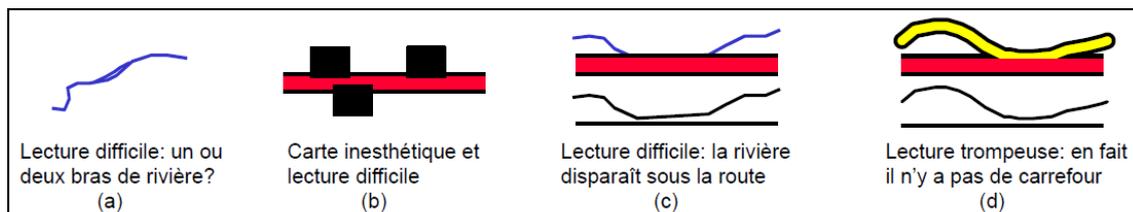


Figure 35. Contraintes relationnelles liées à la proximité des symboles (Duchêne 2004)

- Le deuxième type est la contrainte de maintien des relations afin de conserver des relations qui existent au départ (topologique, métriques...etc.).
- Le troisième type est la contrainte de cohérence géographique, pour garder les relations issues de la sémantique des objets, par exemple, la rivière intersecte la route au pont et la route mène à un point de vue (Figure 36).

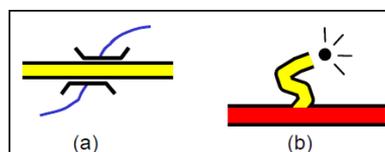


Figure 36. Contraintes de cohérence géographique (Duchêne 2004)

Dans le modèle CartAcom, comme dans le modèle AGENT, la violation d'une contrainte sur une relation déclenche un conflit, le conflit est traité en utilisant des actions. Chaque conflit a une liste d'actions possibles. La modélisation des contraintes relationnelles dans CartAcom est montrée en Figure 37.

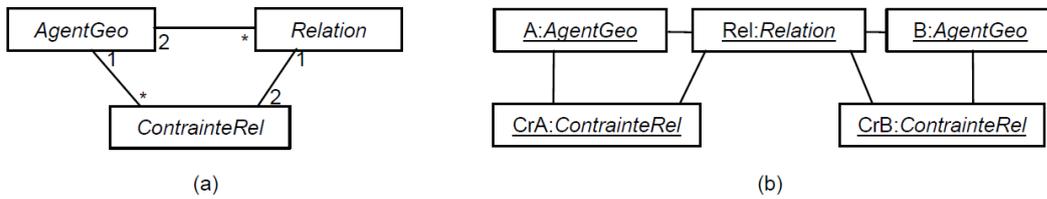


Figure 37. Modélisation des contraintes relationnelles (Duchêne 2004)
 (a) diagramme de classes (b) diagramme d'objets

Une vue plus détaillée sur les deux classes *Relation* et *ContrainteRel* est montrée en Figure 38. La relation inclut une *valeur_courante* qui représente une propriété principale de la relation (e.g.. la distance), et la *valeur_but* représente la valeur voulue de cette propriété (e.g. distance minimum). La valeur de satisfaction représente le degré de respect de la contrainte. L'importance représente le poids de cette contrainte dans l'indice global de satisfaction au niveau d'un objet. La classe *Relation* représente l'état objectif de la contrainte relationnelle, une instance de *Relation* est associée à deux instances de *ContrainteRel* (un pour chaque objet), qui guident la résolution de la contrainte par ces objets. Les contraintes sont asymétriques parce que les actions du premier objet pour la résolution des contraintes ne sont pas forcément les mêmes que les actions du deuxième objet.

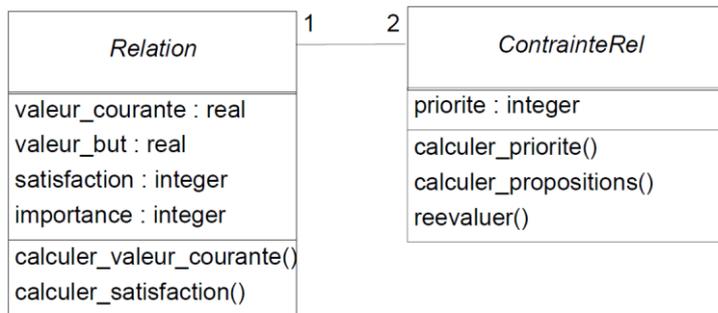


Figure 38. Modélisation des contraintes relationnelles (Duchêne 2004)

Dans le cas où plusieurs contraintes relationnelles sont à satisfaire à la fois pour un même objet, la méthode dans le modèle CartACom consiste à les agréger, si possible, afin de satisfaire un nombre maximum de contraintes. Pendant le processus de généralisation, si un conflit est lié à une contrainte relationnelle comme dans le cas du bâtiment A en Figure 39(a), la solution n'est pas toujours simple. Nous remarquons dans la Figure 39(a) qu'il n'existe pas une action possible qui respecte toutes les contraintes (Figure 39b). Pour résoudre cette problématique, la contrainte la moins importante est retirée parmi les relations. La contrainte la moins importante est la distance minimum acceptable autour du bâtiment B. Après cette relaxation de contrainte, un déplacement du bâtiment A est possible pour répondre aux différentes contraintes agrégées (la résolution de la contrainte relâchée est laissée au bâtiment B dans ce cas).

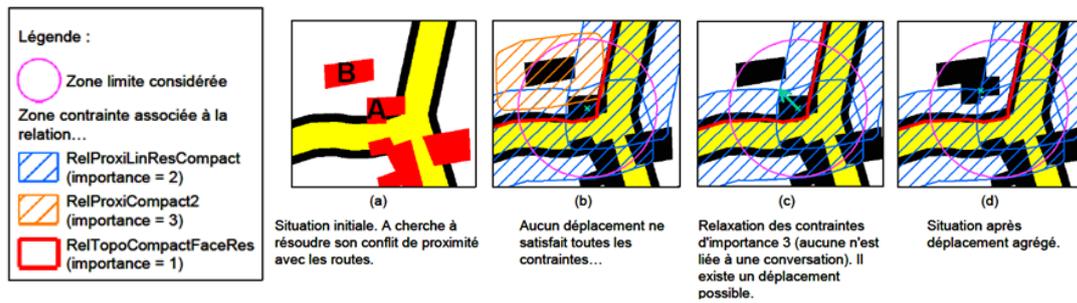


Figure 39. Résoudre la problématique des contraintes relationnelles multiples par l'agrégation (Duchêne 2004)

Le modèle CartAcom ne permet pas la modification des parties d'objets, par exemple, la déformation des courbes de niveau. Le modèle GAEL de (Gaffuri 2008) est une extension du modèle CartAcom qui inclut la déformation d'objets. La déformation est faite en ajoutant un niveau submicro. Ce niveau a des contraintes spécifiques appelées contraintes submicro. Un objet micro peut être composé de plusieurs objets submicro qui sont des agent-points. Par exemple, un agent route est composé d'un ensemble d'agents submicro de type agent-point. En termes de contraintes relationnelles, le modèle GAEL permet de respecter les relations entre un objet et une partie d'un autre objet ou entre les parties des objets (par exemple, entre les points des courbes de niveau). Nous ne traitons pas ce cas dans notre proposition, mais nous en parlerons dans les perspectives de cette thèse. Le modèle GAEL reprend le modèle de contraintes relationnelles de (Duchêne 2004) et spécifie certaines relations, valables uniquement au niveau submicro. Des actions de déformation d'objets sont proposées pour satisfaire les contraintes relationnelles des agents submicro en cas de conflits. La Figure 40 montre un cas de déformation des courbes de niveau pour satisfaire les contraintes relationnelles de proximité entre des parties de courbes de niveau.

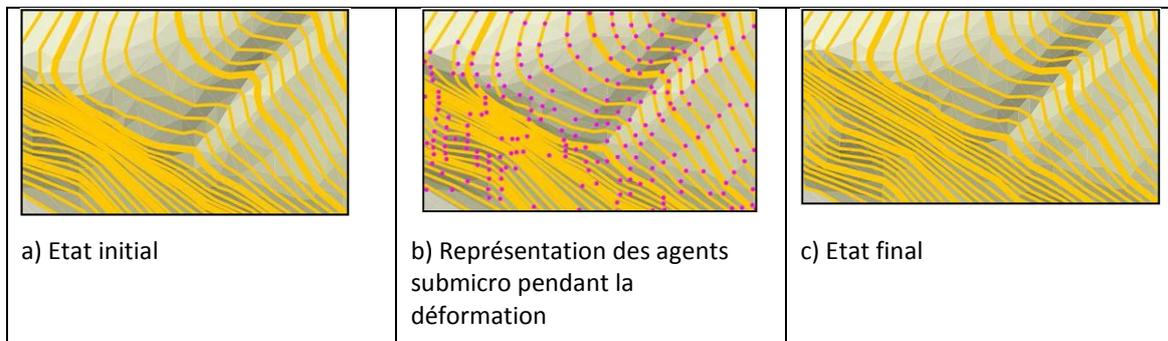


Figure 40. Déformation des courbes de niveau comme un exemple de la prise en compte des relations entre des parties d'objets dans le modèle GAEL (Gaffuri 2008)

D'autres chercheurs ont aussi étudié la déformation d'objets. Ainsi, (Arthur, et al. 2013) cherchent à conserver les relations topologiques pendant la schématisation des régions en utilisant les diagrammes de Voronoi pour assurer que les objets disjoints ne s'intersectent pas pendant la schématisation.

A.3.7 Prise en compte des relations dans l'appariement

Le but de l'appariement est d'identifier les objets homologues entre deux jeux de données. Les objets appariés représentent la même entité dans le monde réel (Walter et Fritsch 1999). Pour notre problématique, l'appariement aide à connaître la différence entre les données topographiques initiales et finales pour traiter les données thématiques en prenant en compte ces changements.

L'appariement est effectué en se basant sur la ressemblance **Invalid source specified**. : deux objets A et B sont appariés s'ils se ressemblent. La ressemblance est calculée en considérant le positionnement, la nature, les relations spatiales et la forme des objets. L'appariement a des critères pour mesurer l'écart entre les objets à appairer.

(Olteanu 2008) distingue 3 catégories de critères: géométrique, relationnel et descriptif.

- Les critères géométriques : la géométrie implique la localisation, la forme, la longueur et l'orientation (Safra et al. 2006 ; Blasby et al. 2004). Il est aussi possible d'appairer une représentation surfacique et une représentation linéaire de la même entité. Une des méthodes consiste à calculer le squelette de la surface pour pouvoir l'appairer (Haunert et Sester 2004).
- Les critères relationnels : la similarité entre les objets géographiques selon les critères relationnels est calculée en comparant leurs relations spatiales. Par exemple, dans la Figure 41a, A et B sont les deux bases topographiques $A = \{a_1, a_2, a_3 \dots\}$ et $B = \{b_1, b_2, b_3, b_4 \dots\}$, et nous voulons appairer A avec B . Si a_1 est lié à b_1 , il est probable que b_2 est apparié à a_2 . Alors a_1, a_2, a_3 sont appariés avec b_1, b_2, b_3 avec l'aide de la structure spatiale **Invalid source specified**. Une autre méthode est élaborée par (Schwering 2006) pour prendre en compte les relations spatiales au niveau des types. Les relations topologiques peuvent être utilisées pour filtrer l'appariement géométrique des réseaux routiers en utilisant la théorie des graphes (Mustière et Devogele 2008). Dans la Figure 41b, la base A est représentée en rose et la base B est représentée en gris. b_1 et b_2 sont appariés avec a_1 en se basant sur les graphes calculés à partir des deux réseaux.

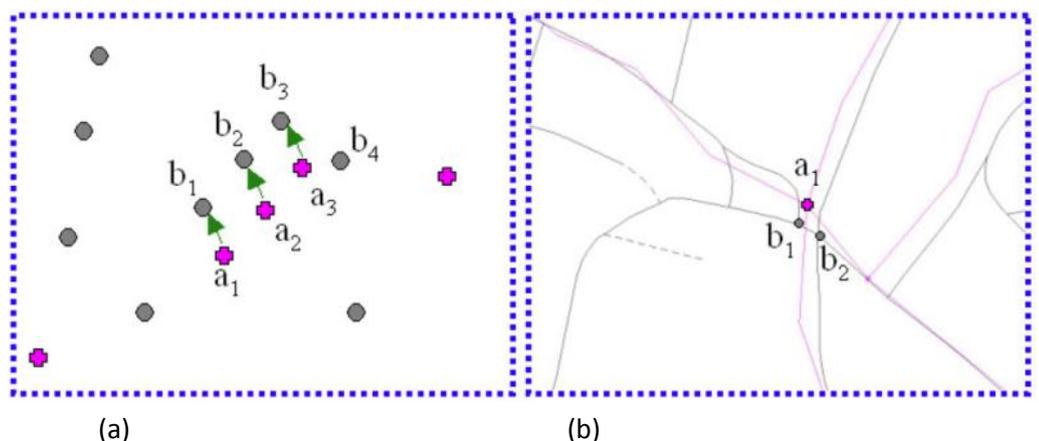


Figure 41. Exemples de l'utilisation des critères relationnels dans un processus d'appariement (Olteanu 2008)

- Critères attributaires : ces critères portent sur les différents attributs de l'objet, tels que le nom, le classement administratif, la largeur, le type d'occupation du sol ou la nature. Les attributs peuvent être quantitatifs (e.g. le nombre de voies, la largeur) ou qualitatifs (e.g. le nom, la nature). Ces attributs portent la sémantique de l'objet géographique. L'appariement sémantique lié à la nature de l'objet entre les attributs textuels a été étudié par (Mechouche et al. 2010). L'appariement proposé par (Schwering 2006) implique aussi une prise en compte de la sémantique des objets. Les spécifications des données peuvent aider aussi à l'appariement (Abadie 2012).

Les appariements proposés par (Olteanu 2008) et (Li et al. 2011) se basent sur ces trois types de critères. La Figure 42 montre les différents critères d'appariement des objets géographiques, selon (Olteanu 2008).

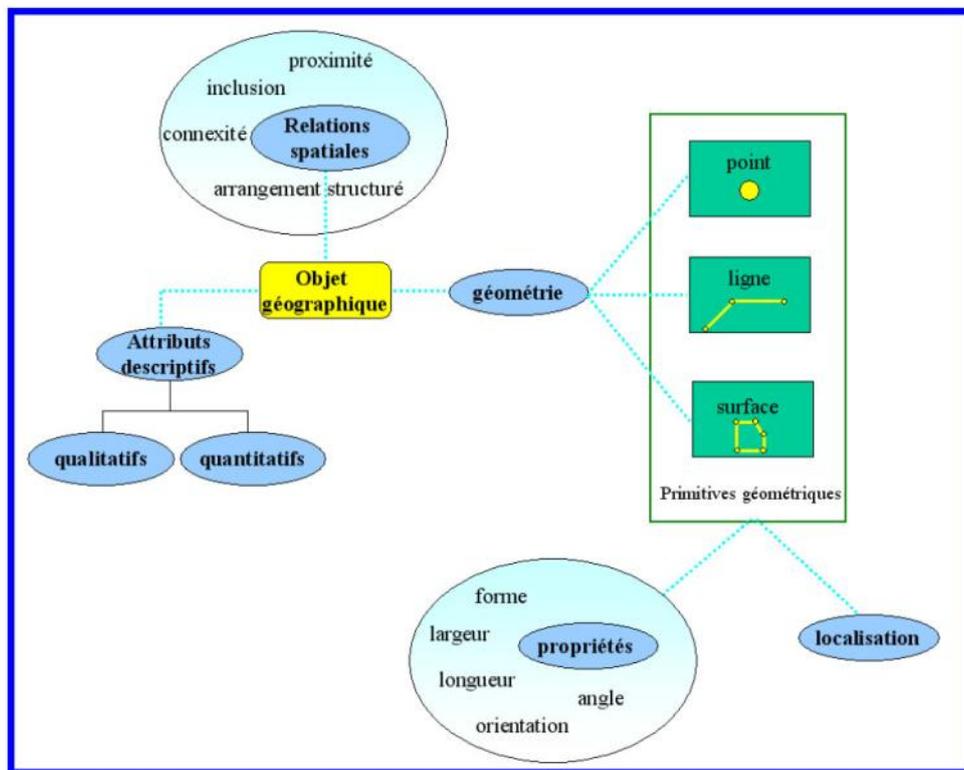


Figure 42. Critères d'appariement d'un objet géographique (Olteanu 2008, p53)

Les données géographiques à appairer ont souvent des schémas. Ces schémas sont détaillés dans les spécifications des données. La prise en compte des spécifications des données lors de l'intégration peut donner un meilleur résultat. Pour cela, il faut formaliser et appairer les spécifications. Dans l'approche proposée par (Abadie 2012), les spécifications sont formalisées en utilisant les ontologies pour pouvoir appairer les spécifications formelles (Figure 43). L'appariement des spécifications aide à appairer les schémas, et ensuite à appairer les données.

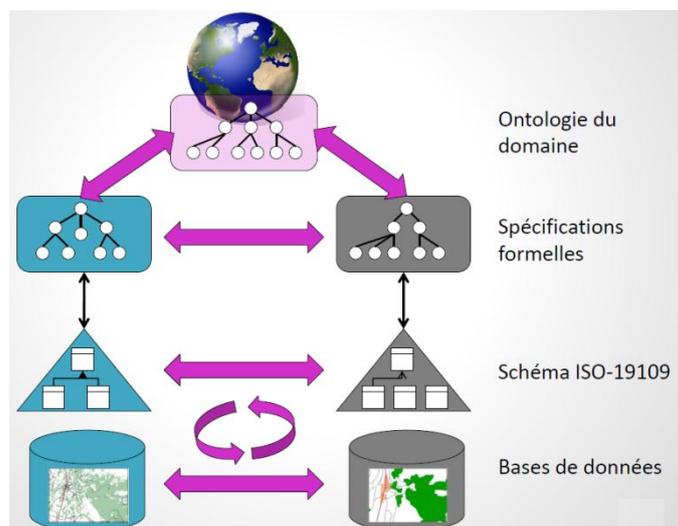


Figure 43. Modèle proposé par (Abadie 2012) pour l'intégration de données

La formalisation des spécifications proposée inclut une formalisation pour les relations spatiales trouvées dans les spécifications. Par exemple: la règle définie dans les spécifications d'une base de données : *Les hôtels de montagne situés dans un parc national sont saisis comme refuges* est formalisée dans l'ontologie qui représente les spécifications pour pouvoir être utilisée lors de l'appariement.

A.3.8 Prise en compte des relations dans l'intégration de données géographiques

L'intégration est un processus qui rend plusieurs jeux de données compatible entre eux afin qu'ils puissent être visualisés sur une seule carte (Rhind et al. 1986). Le but de l'intégration est de trouver une seule sémantique pour l'ensemble des données à intégrer. C'est aussi d'éliminer des incohérences liées à la redondance de données (Devogele et al. 1998). Autrement dit, l'intégration de données implique l'intégration des schémas et des données.

Nous nous intéressons à la prise en compte des relations dans l'intégration parce que nous cherchons à bien intégrer les données thématiques dans le jeu de données final. Notre intégration sera basée sur les relations. Ainsi dans ce paragraphe, nous étudions les méthodes d'intégration qui prennent en compte les relations.

Il faut noter que l'intégration peut impliquer une phase d'appariement afin d'identifier les objets redondants entre les jeux de données à intégrer. Une phase d'adaptation de schéma peut être aussi nécessaire. La gestion des données collaboratives représente un cas compliqué d'intégration parce que données sont par défaut hétérogènes, et elles ont différents niveaux de détail, cela rend la cohérence plus difficile à obtenir.

La prise en compte des relations spatiales est une partie essentielle du processus d'intégration. La Figure 44 montre un exemple de superposition des données venant de plusieurs sources. En Figure 44a, la rivière doit passer par la ligne de la vallée, mais avec une simple superposition, cette relation n'est pas vérifiée, ce qui génère une incohérence logique qui se voit visuellement. Sur la Figure 44b, les courbes de niveau ne doivent pas intersecter le lac. Selon (Butenuth et al. 2007), le rôle de l'intégration n'est pas juste superposer les données, mais aussi il faut qu'elle rende les relations explicites.

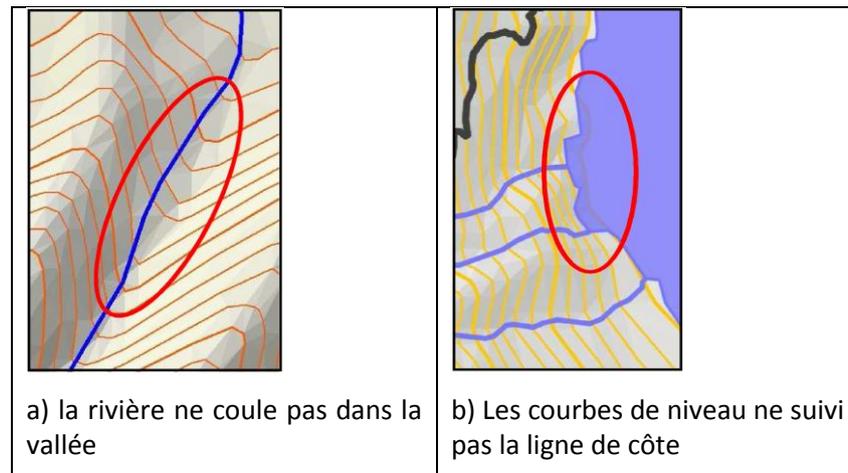


Figure 44. Exemple des incohérences liées à l'intégration et à la généralisation (Gaffuri 2008)

Certains travaux visent à étudier la détection des conflits ou l'incohérence liée au non-respect d'un type de relations. Par exemple, le travail de (Wallgrün 2013) concerne les relations topologiques des objets surfaciques avec les autres entités géographiques, par exemple, un pont représenté par une surface doit inclure une route. Une méthode pour satisfaire ces relations est proposée afin d'obtenir une meilleure cohérence, mais le déplacement est la seule action dans la méthode. Cette méthode peut être intégrée dans l'ensemble des déplacements possibles d'un objet pour satisfaire des relations topologiques.

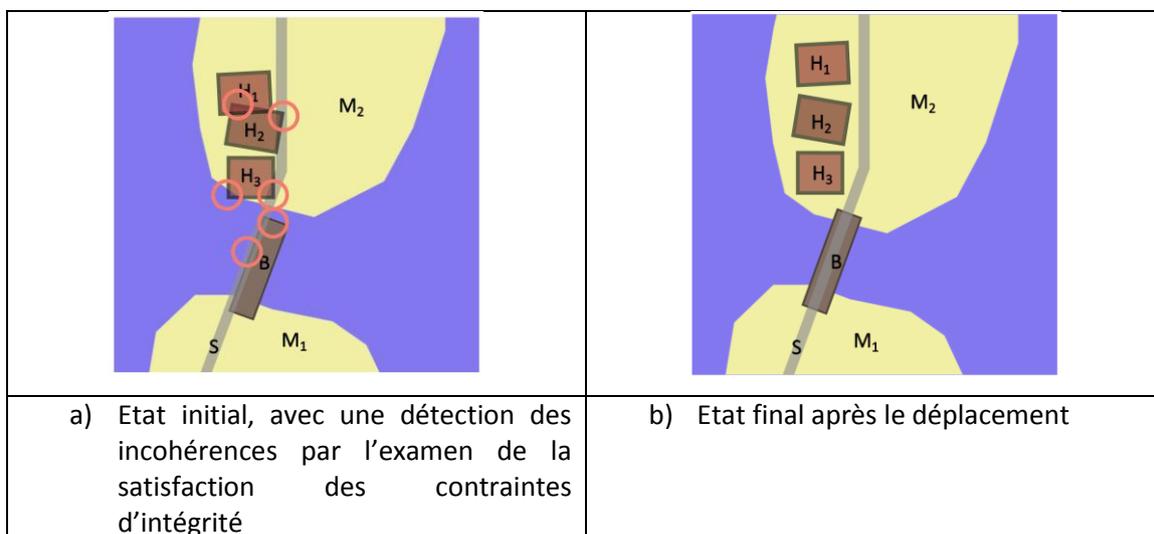


Figure 45. Améliorer le résultat de l'intégration de données par déplacement basé sur les contraintes d'intégrité (Wallgrün 2013)

L'incohérence liée à la différence de niveau de détail est abordée par (Touya et Brando 2013). La méthode proposée détecte le niveau de détail et les incohérences liées à des violations de contraintes sur des relations. L'évaluation du niveau de détail est faite en utilisant des mesures géométriques et aussi sémantiques (ex. la source de données). La Figure 46 montre quelques

cas remarquables dans OpenStreetMap. La figure montre des violations de certaines relations topologiques entre les bâtiments et les données d'occupation du sol.



Figure 46. Cas d'incohérences liées à la différence de niveaux de détail (Touya et Brando 2013) (la couleur gris sombre représente les zones urbaines)

Un modèle pour la gestion et l'intégration des données collaboratives est proposé par (Brando 2013). Le modèle a pour objectif d'évaluer et réconcilier les contributions venant des contributeurs pour obtenir un contenu collaboratif cohérent. Le modèle est composé de deux parties : l'évaluation pour la détection des incohérences et ensuite la résolution pour corriger l'incohérence.

La détection des incohérences est basée sur les contraintes d'intégrité. Les contraintes d'intégrité décrivent les relations qu'il faut maintenir pour préserver la cohérence logique des données. Le modèle propose un vocabulaire formel pour les entités, les propriétés et les relations. Ce vocabulaire aide à formaliser l'information saisie par le contributeur. Parmi cette information, nous avons les relations spatiales qui concernent les données saisies par lui. Par exemple, le bâtiment est en face de la poste. Le vocabulaire est rempli d'une façon collaborative par les utilisateurs et il est enrichi par des éléments extraits automatiquement par différentes ressources comme DBpedia, WordNet, Wikipédia. Les incohérences sont détectées en testant la satisfaction des contraintes d'intégrité par des opérations spatiales.

Pour résoudre les incohérences trouvées, le modèle inclut des méthodes correctives. Par exemple, pour la contrainte d'intégrité, *les abribus ne doivent pas intersecter les tronçons de routes*, la méthode de correction utilise une méthode appelée *déplacer*, qui calcule l'orientation de l'abribus par rapport au tronçon de route, puis déplace l'abribus pour corriger l'erreur. Un autre exemple est illustré en Figure 47. L'incohérence dans ce cas est liée à la position relative du bâtiment par rapport à la route et la méthode de correction *déplacer* a été exécutée.

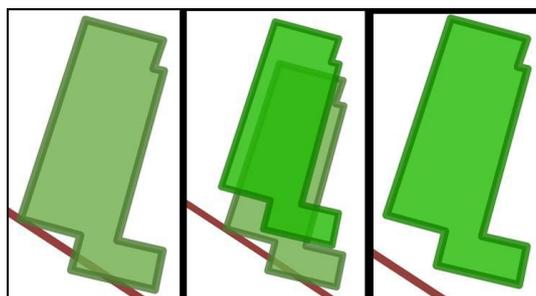


Figure 47. Exemple d'une incohérence sur OSM concernant un bâtiment chevauchant une route corrigée par la méthode de (Brando 2013)

Notons que la méthode de correction d'incohérences proposée par (Brando 2013) ne permet pas d'agréger plusieurs relations.

A.3.9 Prise en compte des relations dans la migration de données

Nous avons introduit la migration de données thématiques qui fait l'objet de notre thèse en A.2.2 (p.25). Il s'agit de repositionner des données thématiques sur des données topographiques après remplacement de celles-ci. Quelques méthodes de migration de données existent dans la littérature.

(Lu et Arikawa 2013) proposent une méthode de migration de données de l'utilisateur, que l'utilisateur saisit sur un sketch (par exemple, sur le plan de métro). La méthode est basée sur un appariement manuel par l'utilisateur entre le sketch initial et des données topographiques (qu'on appelle données topographiques finales). La migration de données est effectuée en préservant les relations de proximité avec les points appariés. La méthode est basée uniquement sur les relations métriques fournis par le lien d'appariement et ignore des autres relations. Cette ignorance engendre des incohérences.

Pour relocaliser des données ponctuelles sur réseau, une méthode utilisée par le CETE d'Aix consiste à conserver la distance proportionnelle. Avec cette méthode de relocalisation, certaines relations peuvent être perdues, par exemple la relation avec le sommet de virage (voir Figure 14). Nous détaillons le cas de la migration de données ponctuelles de CETE dans le chapitre C de cette thèse.

A.4 Bilan du chapitre

Notre problématique focalise sur la prise en compte du référencement des données saisies par l'utilisateur lors de traitements liés au changement de niveau de détail, en particulier, la migration de données thématiques.

Dans ce chapitre, nous avons étudié la combinaison des données thématiques avec les données topographiques. Nous avons présenté des méthodes de création des données thématiques en montrant le rôle des relations pendant la création. Nous avons discuté le rôle des relations dans les différents traitements quel que soit le type de données (thématiques ou topographiques). Le cas des données thématiques sur des données topographiques représente un cas spécifique où les relations ont encore plus d'importance. Cette importance est justifiée par des exemples qui montrent des incohérences liées à l'ignorance de ces relations pendant des traitements.

Nous avons regroupé les besoin en deux cas : le premier cas est celui du remplacement des données topographiques et le deuxième cas est celui de la généralisation automatique des données topographiques.

Notre objectif est de proposer un modèle de référencement par les relations entre les données thématiques et les données topographiques, puis d'étudier l'adaptation des processus de changement de niveau de détail pour prendre en compte les relations spatiales représentées par le modèle de référencement. Nous proposerons un modèle pour migration de données en prenant en compte les relations initiales. Par la suite, nous étudierons l'extension de ce modèle pour résoudre d'autres problématiques liées au non respect des relations spatiales lors des traitements des données.

L'état de l'art donne une vue générale sur les travaux de modélisation des relations. Un type de relation peut avoir plusieurs modélisations suivant le but de la modélisation et la finesse de description demandée. La prise en compte des relations n'est pas nouvelle, alors nous avons expliqué différentes méthodes utilisées par les chercheurs pour la prise en compte des relations dans différents traitements. Généralement, Il existe des solutions efficaces pour les cas simples où nous avons un seul type de relation. Par contre, si les relations sont nombreuses et de types différents, la solution devient plus complexe et peut impliquer une suppression de relations à respecter.

Nous allons présenter dans le chapitre B notre contribution pour la prise en compte des relations. Nous commençons par présenter notre modèle pour la migration de données thématiques, ensuite, nous étudierons l'extension du modèle pour les autres traitements, notamment pour la généralisation automatique. Une mise en œuvre de notre modèle sur des cas jouets et sur des cas d'application réels fournis par un gestionnaire de réseau routier est présentée dans le chapitre C.

Chapitre B

Modèle pour la migration de données thématiques ponctuelles sur réseau

Dans ce chapitre, nous présentons un modèle pour la prise en compte de la dépendance entre les données thématiques et des données topographiques lors de la migration de données thématiques. Ce processus est souvent utilisé dans le cas d'un changement de niveau de détail. Nous commençons par donner une vue générale de notre approche. Notre modèle de référencement des données thématiques sur les données topographiques est ensuite présenté et détaillé. Pour gérer la problématique d'avoir un certain nombre de relations à respecter, le processus utilise une méthode multicritères.

Dans la dernière partie de ce chapitre, nous présentons d'autres applications possibles de notre modèle. Ainsi, nous expliquons l'utilisation de notre modèle dans différents traitements qui nécessite une combinaison de données thématiques et de données topographiques.

Chapitre B. Modèle pour la migration de données thématiques ponctuelles sur réseau

B.1 Introduction

Comme nous l'avons annoncé dans l'objectif (A.2.3), nous focaliserons notre approche sur la migration de données thématiques ponctuelles sur réseau. La plupart des éléments du modèle proposé seront réutilisables pour les autres cas présentés en A.2.2 (migration des données autres que les données ponctuelles sur réseau et généralisation).

Ce chapitre est organisé comme suit. Dans un premier temps, nous donnerons une vue d'ensemble de notre approche pour la migration de données thématiques de type ponctuel sur réseau (B.2). Dans cette partie, nous présenterons une brève description de différentes étapes de l'approche. Ensuite, les différentes composantes du modèle proposé seront présentées de B.3 à B.5. Les étapes seront illustrées sur des cas d'étude jouets en B.6. Ensuite, nous discuterons l'extension du modèle pour plusieurs types de traitements sur la combinaison thématique-topographique (généralisation, migration de données, carte à la demande, etc.).

Présentation des exemples-jouets

Les deux exemples-jouets suivants qui concernent la migration de données seront utilisés comme des exemples-jouets pour la suite. Tous les deux concernent des accidents routiers (données thématiques) localisés sur les routes (données topographiques). Dans les deux exemples, le but est d'effectuer une migration de ces données thématiques vers une base topographique moins détaillée. Les images de gauche montrent la combinaison données thématiques/données topographiques initiale. Celles du milieu montrent les données topographiques finales et celles de droite montrent le résultat d'une migration à l'aide d'un recalage par abscisse curviligne proportionnelle (voir Figure 14, page 24 et les exemples associés).

Dans le premier exemple (Figure 48a, Ex.1), nous avons un accident en face d'un bâtiment. Dans la base topographique finale (Figure 48b, Ex.1), la route est moins détaillée et plus lisse. Si l'accident est relocalisé en se basant sur la distance proportionnelle par rapport à la route dans la base topographique finale, il ne sera plus en face du bâtiment (Figure 48c, Ex.1), alors que cette relation est peut-être importante pour l'utilisateur.

Le deuxième exemple contient deux accidents (Figure 48a, Ex.2). Le premier est proche du rond point et le deuxième est un peu plus loin. Le rond-point est représenté par un carrefour dans la base topographique finale (Figure 48b, Ex.2). Si ces accidents sont relocalisés par abscisse curviligne dans la base topographique finale, la relation de proximité de l'accident avec le rond point sera perdue (Figure 48c, Ex.2), malgré l'importance éventuelle de cette relation pour l'analyse des accidents.

L'objectif de notre travail de thèse est donc de proposer un modèle pour que ces relations soient mieux maintenues.

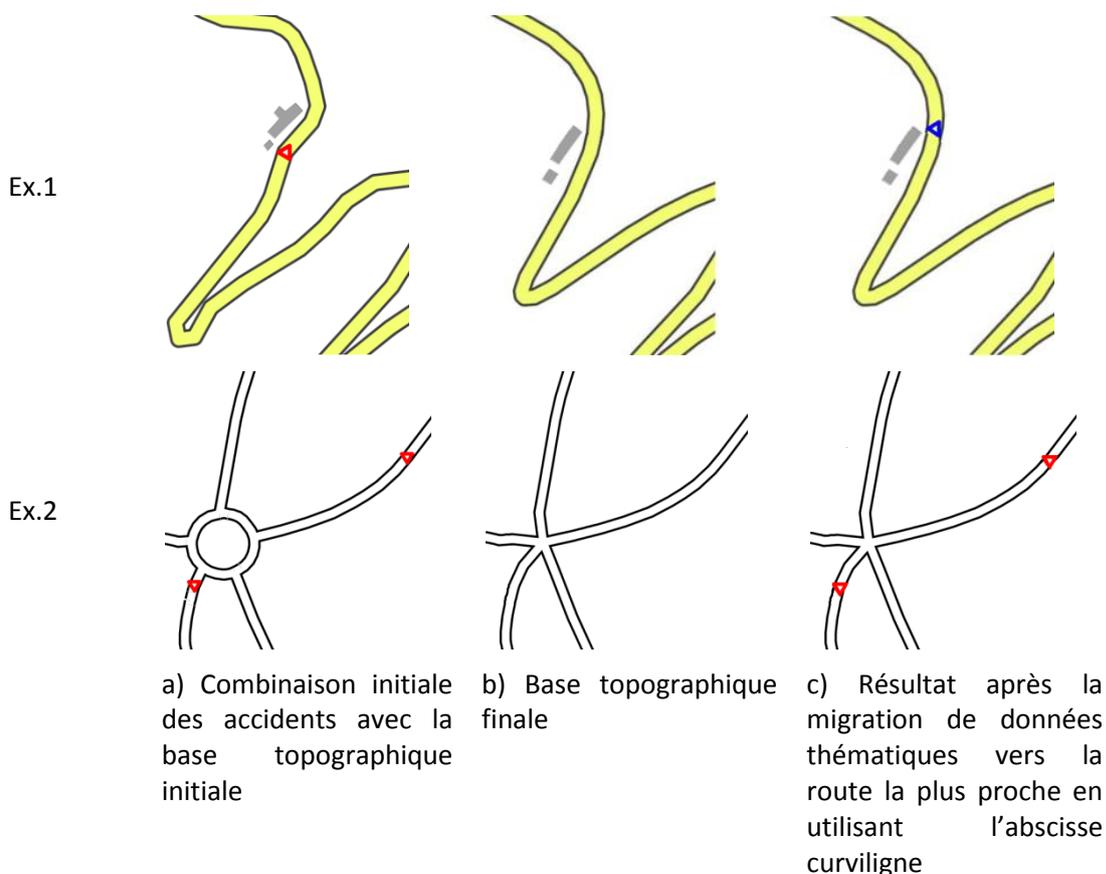


Figure 48. Notre exemple-jouet : migration d'accidents sur un nouveau réseau routier en utilisant uniquement l'abscisse curviligne (c)

B.2 Vue d'ensemble du modèle proposé pour la migration de données thématiques ponctuelles sur réseau

Nous avons vu dans le premier chapitre le rôle important des relations lors de traitements d'une combinaison de données thématiques-données topographiques. Parmi les traitements étudiés, nous avons discuté le cas de la migration de données thématiques. Ces relations ont encore plus d'importance si les données sont saisies par l'utilisateur. En effet, pendant la saisie des données thématiques sur les données topographiques, les relations spatiales sont prises en considération par l'opérateur d'une manière ou d'une autre suivant le mode de saisie (discuté en A.1.3).

Nous proposons une approche pour la prise en compte des relations lors de la migration de données thématiques. Le but de la migration de données thématiques (présentée dans paragraphe A.2.2) est de relocaliser les données thématiques sur les données topographiques finales (Figure 49 reproduction de la Figure 18).

Notre approche ne se focalise pas sur un seul type de relation (comme la plupart des travaux antérieurs), mais permet d'en prendre en compte plusieurs types à la fois. L'approche permet d'extraire et représenter différents types de relations grâce à un modèle de représentation

que nous proposons. Le modèle rend les relations explicites et manipulables pendant les différentes étapes de l'approche.

Les données topographiques finales peuvent être une base de données différente de la base de données topographique initiale ou avoir été obtenues par généralisation des données initiales sans prise en compte des relations avec les données thématiques. La relocalisation doit prendre en compte les relations initiales et les données topographiques finales pour que la combinaison finale soit cohérente.

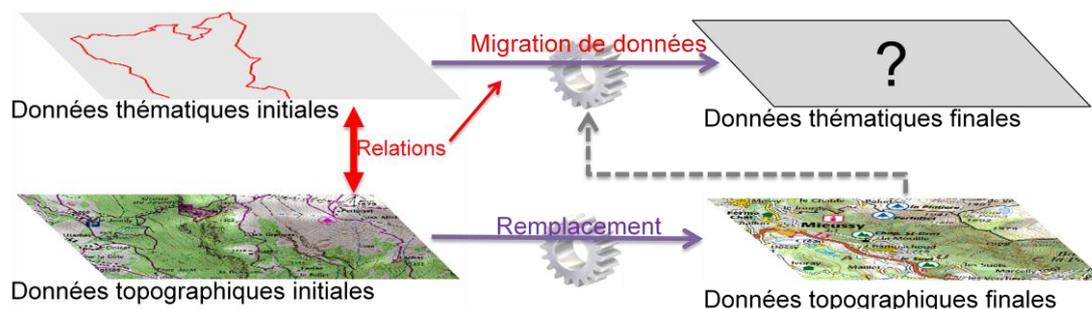


Figure 49. Migration de données thématiques suite au remplacement des données topographiques, avec prise en compte des relations initiales

B.2.1 Description formelle du problème

Nous formalisons le cas de la migration de données thématiques comme suit. Les objets topographiques (et leurs enrichissements éventuels) servent de référence pour les objets thématiques. Le référencement d'un objet thématique par rapport aux objets topographiques se traduit par le fait qu'il existe des relations spatiales porteuses de sens entre l'objet thématique et les objets topographiques. Un objet thématique peut avoir plusieurs relations avec un ou plusieurs objets topographiques. Inversement, un objet topographique peut avoir plusieurs relations avec un ou plusieurs objets thématiques. Nous considérons que chaque relation implique seulement deux objets : un objet thématique et un objet topographique.

Nous adoptons les notations suivantes :

- A est l'ensemble des objets thématiques dans les données initiales,
- B est l'ensemble des objets topographiques dans les données initiales,
- \mathcal{R} est l'ensemble des relations existant entre les objets et qui traduisent le référencement des objets thématiques par rapport aux objets topographiques, dans les données initiales,
- A' est l'ensemble des objets thématiques dans les données finales,
- B' est l'ensemble des objets topographiques dans les données finales,
- \mathcal{R}' est l'ensemble des relations existant entre les objets et qui traduisent le référencement des objets thématiques par rapport aux objets topographiques, dans les données finales.

Connaissant les données initiales thématiques (A) et les données topographiques (B), nous tenterons d'extraire et de caractériser (par des attributs) l'ensemble des relations \mathcal{R} . Cela est fait par le modèle de référencement. Ainsi donc la situation initiale est la suivante :

$$A\mathcal{R}B$$

Les données topographiques finales (B') sont déjà connues, nous cherchons à déduire (1) les relations \mathcal{R}' attendues dans l'état final, et (2) la localisation des objets thématiques dans l'état final, c'est-à-dire l'ensemble A' respectant les relations \mathcal{R}' avec B' :

$$A\mathcal{R}B \wedge B \rightarrow B' \Rightarrow ? \mathcal{R}' ? A' / A' \mathcal{R}' B' \text{ avec } \mathcal{R}' \text{ cohérent avec } \mathcal{R}$$

Après l'extraction des relations initiales \mathcal{R} , les relations finales \mathcal{R}' seront déduites, ensuite, les données thématiques A sont repositionnées pour obtenir les données thématiques finales A' . Autrement dit, notre but est de générer des données thématiques qui sont cohérentes avec les données topographiques finales en respectant les relations initiales.

B.2.2 Les étapes de l'approche pour la migration de données thématiques

Pour prendre en compte les relations dans la migration de données, nous proposons un processus en cinq étapes:

- a) L'extraction d'objets caractéristiques : un objet thématique peut être en relation avec un objet implicite de la base topographique. Si, par exemple, l'objet implicite est un sommet de virage, il faut enrichir les données topographiques par l'extraction des sommets.
- b) L'extraction des relations entre données topographiques et données thématiques dans les données initiales: les types significatifs de relations sont identifiés suivant le cas d'utilisation, ensuite, les instances de ces relations sont extraites des données initiales. La représentation de ces relations est faite par un modèle de représentation que nous proposons.
- c) L'appariement : les bases topographiques initiale et finale sont appariées, ainsi que les objets caractéristiques des deux bases.
- d) L'identification des relations finales attendues : les relations attendues au niveau de la base finale sont inférées en prenant en compte les relations initiales, les données topographiques finales et le résultat de l'appariement.
- e) La relocalisation des données thématiques : les relations inférées sont utilisées pour contrôler la relocalisation des données thématiques sur les données topographiques finales.

La Figure 61 montre l'enchaînement des 5 étapes, ainsi que leurs entrées et sorties.

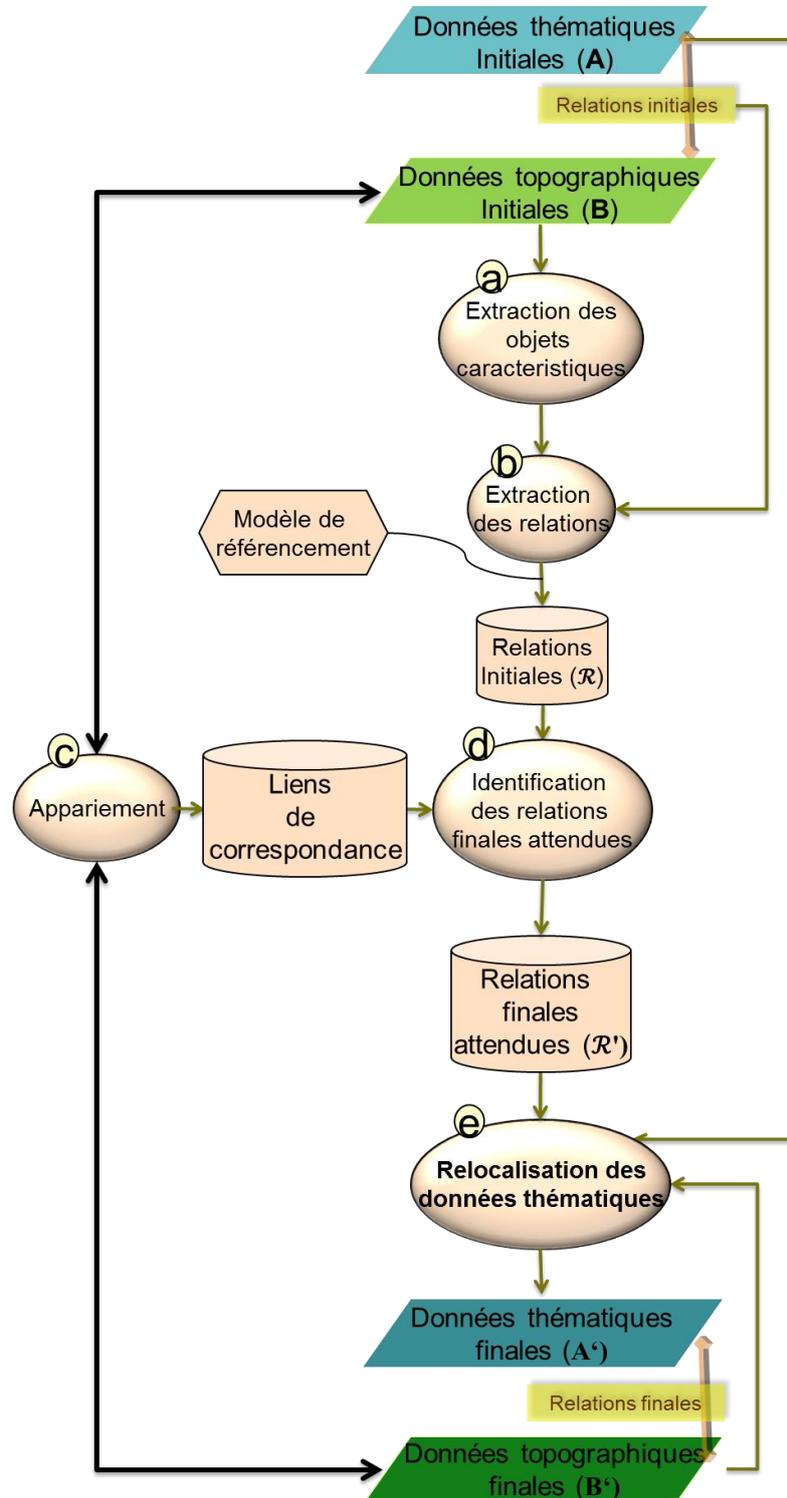


Figure 50. Processus de migration de données thématiques proposé

Dans les parties B.3 à B.5, nous allons détailler les 5 étapes. La partie B.3 contiendra des détails sur les deux premières étapes. Nous y détaillerons le concept d'objets caractéristiques et présenterons les méthodes pour les extraire (B.3.1). Nous allons également présenter les types de relations considérées pour référencer les objets thématiques sur les objets topographiques et proposer une méthode pour leur assigner une importance (B.3.2 à B.3.4). Puis nous présentons notre modèle pour la représentation de ces relations de référencement (B.3.5 et B.3.6).

La partie B.4 concerne les étapes (c) et (d). Nous commençons par introduire l'idée que les relations finales attendues peuvent être différentes des relations initiales. La différence est due aux différences entre les bases topographiques initiale et finale. Nous présentons ensuite l'appariement de données comme un processus permettant d'identifier ces différences (B.4.1), puis nous décrivons comment nous proposons d'identifier les relations finales souhaitables (B.4.2).

En B.5, nous détaillons l'étape finale de relocalisation (étape (e)), qui permet de positionner les données thématiques sur les données topographiques finales. Le résultat attendu de la relocalisation est un jeu de données thématiques finales dont les relations avec les données topographiques finales correspondent au mieux aux relations finales attendues extraites à l'étape (d).

Pour illustrer les différentes étapes, nous présenterons leur application sur nos exemples-jouets en B.6.

B.3 Extraction et modélisation des relations

B.3.1 Concept d'objets caractéristiques

Les relations permettant de référencer un objet thématique par rapport aux données topographiques n'impliquent pas toujours un objet topographique explicite qui existe sous forme d'instance dans la base de données. Parfois elles impliquent des objets implicites. Par exemple, un accident peut être avant ou après l'intersection entre une route et une voie ferrée. Nous avons besoin d'enrichir les données pour pouvoir représenter ces relations. L'enrichissement se traduit par la création d'objets, d'attributs ou de relations. Le concept d'enrichissement est introduit par (Ruas et Lagrange 1994), dans le contexte de la généralisation. Par exemple, en ville, les îlots ou "pâtés de maisons" peuvent être créés comme les faces du graphe topologique des routes; les routes peuvent être caractérisées par leur sinuosité. Les relations sont aussi considérées comme de l'information caractéristique qui enrichit les données (Neun et al. 2004).

Dans le contexte de la généralisation, des travaux sur l'enrichissement des données portent sur la reconnaissance de certaines structures comme les groupes d'îles (Steiniger et al. 2006) ou les alignements des bâtiments (Christophe et Ruas 2002; Steiniger et al. 2008). D'autres travaux concernent l'enrichissement de réseaux routier (Heinzle et Anders 2007; Touya 2010). Certains travaux décrivent les concepts géographiques à l'aide d'une ontologie. La méthode de détection utilise cette description pour trouver les instances correspondantes dans les données (Lüscher et al. 2008). La Figure 51 illustre une partie d'une ontologie utilisée par

Lüscher et al. (2008) pour l'enrichissement des données. Avec cette modélisation, les bâtiments alignés peuvent être extraits.

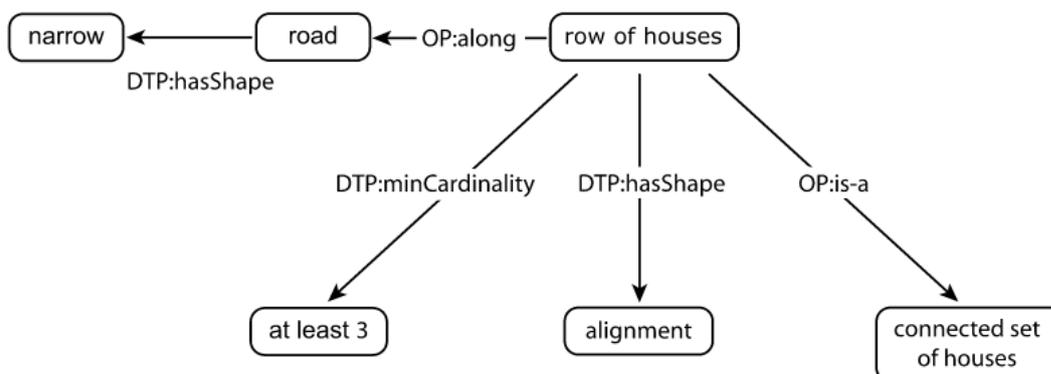


Figure 51. Une partie d'une ontologie qui décrit les maisons de villes mitoyennes (terraced houses en anglais), l'ontologie est adaptée pour l'enrichissement des données. DTP : propriété du type de données. OP : propriété d'objet. (Lüscher et al. 2008)

L'enrichissement de données est aussi utilisé pour aider à appairer deux bases topographiques qui ont des niveaux de détails différents. Des travaux sur l'appariement de réseaux routier (Mustière et Devogele 2008; Bang et al. 2010) utilisent la théorie de graphes après explicitation du graphe sous jacent aux objets à appairer. Pour appairer des objets ayant une géométrie de dimension différente à des niveaux de détail différents (par exemple, une route ou une rivière est représentée par un objet surfacique dans une base de données et par un objet linéaire dans une autre), (Haurert et Sester 2004) explicitent d'abord le squelette de l'objet surfacique.

Pour nos besoins de référencement par les relations, nous proposons d'utiliser le concept d'*objets caractéristiques*. Nous définissons les *objets caractéristiques* comme des objets additionnels extraits de la base topographique initiale. Ils sont choisis pour être a priori reconnaissables dans les différentes représentations topographiques de la région concernée. L'extraction de ces objets enrichit les données topographiques et permet de se relocaliser entre différents jeux de données topographiques d'une même région, même s'ils ont des niveaux de détails différents. Par exemple, nous pouvons extraire les points d'intersection entre les rivières et les routes, puis nous pouvons reconnaître les mêmes intersections dans les autres jeux de données, cela aide à relocaliser et connaître les positions correspondantes entre les jeux de données. Le concept des objets caractéristiques est déjà utilisé par (Plazanet 1996, p.139). Son travail concerne spécifiquement les objets linéaires, il s'agit d'enrichir le réseau routier en vue de sa généralisation. Les objets caractéristiques utilisés sont les points d'inflexions et les sommets de virages des routes. Afin de les conserver pendant le lissage, l'importance des virages est liée à leur résistance au lissage. Dans (Grosso 2010), les monuments religieux sont utilisés comme des objets caractéristiques pour appairer les cartes anciennes. Après extraction de ces points caractéristiques, nous pouvons expliciter plus de relations, telles que la relation qui indique qu'un accident a eu lieu sur la route avant ou après ou sur l'intersection route-rivière.

Pour le cas des accidents de la route, après une réflexion sur les objets qui peuvent servir comme des points de repère et en discutant avec un gestionnaire de routes, nous avons identifié comme objets caractéristiques : les sommets de virages sur les routes, les ronds-points, les intersections des routes avec le réseau ferré et les rivières, les carrefours, les débuts et fins de tunnels. En effet, ces objets peuvent être retrouvés par le calcul et ont une probabilité raisonnable d'être présents dans les différents jeux de données topographiques (sauf les débuts et fins de tunnels dont la présence dépend de la richesse sémantique des données). De plus, ces objets peuvent avoir une dimension explicative des accidents.

La Figure 52 montre le résultat de l'extraction automatique de ronds-points sur l'exemple-type introduit au début du chapitre, grâce à la méthode proposée par (Sheeren 2005, p. 189-191). Ce nouvel objet aide à construire le lien d'appariement entre le rond-point et le carrefour correspondant dans la base finale, et à l'explicitation de relations de proximité entre accidents et rond-point.

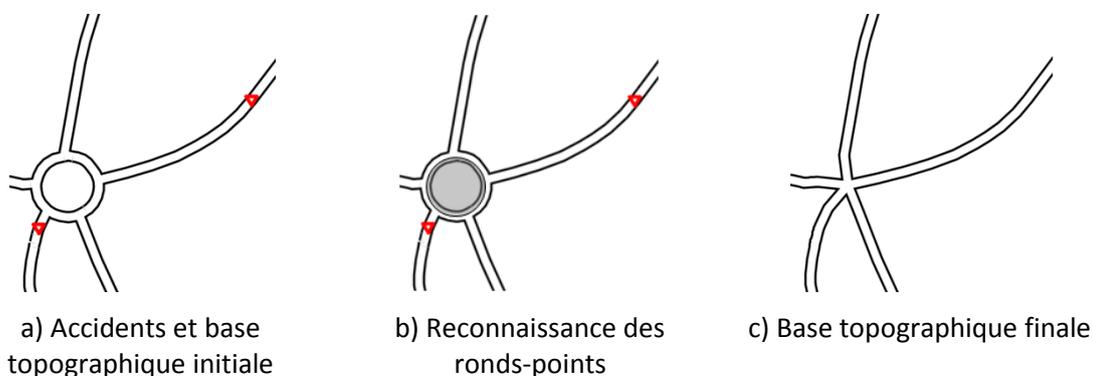


Figure 52. Extraction automatique d'un rond-point

Nous proposons de modéliser les relations spatiales. Dans notre cas, une relation spatiale a deux composants. Ces composants sont soit un objet thématique et un objet topographique, soit un objet thématique et un objet caractéristique.

Pour le besoin de modélisation, nous définissons l'*objet support* : l'objet topographique ou l'objet caractéristique qui est en relation avec un objet thématique.

Pour les relations qui concernent plusieurs objets (ex. alignement de bâtiments), un objet caractéristique doit être créé et les objets thématiques seront en relation avec ce nouvel objet créé. Par exemple, pour les bâtiments alignés, un objet caractéristique peut être créé à l'aide d'une méthode telle que celles de (Regnauld 1998; Christophe et Ruas 2002; Zhang et al. 2013)

B.3.2 Importance des relations initiales et importance des relations finales

Il existe une grande variété de relations, mais les relations à extraire dépendent des données initiales, du mode de saisie et du but de la carte. Cependant, comment savoir si une relation est importante et s'il faut l'extraire et la modéliser ou non ? Nous pensons que l'importance des relations dans l'état initial (relations initiales) ne s'évalue pas de la même façon que l'importance des relations dans la combinaison finale (relations finales).

Importance des relations initiales

L'importance d'une relation initiale est dérivée du degré de prise en compte de la relation au moment de la saisie des données thématiques. Selon le mode de saisie (discuté en A.1.3), certaines relations peuvent être prises en compte plus que d'autres. Prenons le cas d'un gestionnaire de réseau routier qui saisit ses données en se basant sur la distance curviligne par rapport au début de la route. Dans ce cas, la relation de proximité de l'objet accident avec le début de la route est plus importante que la relation de proximité de l'objet accident avec un bâtiment proche. Par contre, si le bâtiment est un bâtiment remarquable, l'importance de la relation de proximité avec le bâtiment sera la plus importante que la relation de proximité avec le début de la route. L'information sur les relations qui ont été prises en compte lors de la saisie des données thématiques peut être prédéfinie ou elle peut être récupérée de façon interactive par une interface de dialogue avec l'utilisateur. Cette importance concerne uniquement les données initiales sans considérer l'état final voulu.

Il existe deux façons pour assigner l'importance initiale des relations. La première façon est de fixer l'importance au niveau du type de relations, cela veut dire que l'importance est la même pour toutes les instances. La deuxième façon est de calculer l'importance pour chaque instance. Par exemple, pour assigner les importances aux relations de position relative d'un ensemble d'objets thématiques de type bâtiment, la position relative des bâtiments par rapport aux routes voisines peut être assignée différemment pour chaque instance suivant la distance.

Importance des relations finales attendues

Les relations finales attendues sont les relations que nous calculons en prenant en considération les préférences de l'utilisateur et le but du produit final (carte pour se localiser, carte d'analyse, etc.). En reprenant l'exemple du gestionnaire de routes, il peut préférer conserver la relation de proximité avec les bâtiments pendant la migration des données parce qu'il considère cette information importante et suffisamment précise bien que cette relation n'ait pas été vérifiée pendant la saisie.

Nous pensons que l'importance d'une relation finale doit être calculée sur la base des éléments suivants :

- L'importance de la relation initiale
- Les préférences de l'utilisateur et le but de la carte.
- Les spécifications de l'état final (l'échelle, le niveau de détail, etc.)

Nous n'avons pas étudié en détail le calcul de l'importance dans le cadre de cette thèse. Dans notre implementation, nous attribuons les importances manuellement.

B.3.3 La relation « porté par »

La relation *porté par* est une relation spéciale qui représente la dépendance forte entre un objet thématique et son espace de support topographique. Un objet thématique peut être porté par un objet support seulement s'il est proche ou inclus dans son objet support. Nous affectons à cette relation une importance maximale au sein de l'état initial et aussi au sein de l'état final. Si l'objet thématique est en relation *porté par* avec un objet topographique ou

caractéristique, l'existence de l'objet thématique sera directement liée à l'existence de son espace de support. Par exemple, un accident qui se situe sur une route nommée *D7*, ne peut pas être localisé sur une base topographique où cette route n'est pas représentée. La relation *porté par* implique un lien sémantique entre l'objet thématique et son espace de support. Avec l'absence de lien sémantique, la relation *porté par* n'existe pas. Par exemple, si un objet d'accident routier se situe très proche d'une rivière, il n'aura pas de relation *porté par* avec la rivière, parce que l'accident routier ne peut être porté que par une route ou un objet caractéristique de réseau routier (e.g. un carrefour). En effet, la relation *porté par* est une relation fonctionnelle au sens de (Touya 2011, p.63). Cela veut dire que l'importance de la relation est dérivée de la sémantique des données. Par exemple, les relations entre les points qui représentent les équipements routiers (par exemple, stations services) et les routes sont fonctionnellement importantes du fait de la sémantique des données thématiques. L'objet thématique peut être sur son espace de support et il peut aussi se positionner à côté. Par exemple, un accident routier peut être localisé dans le voisinage d'une route à cause de la précision de la saisie ou parce que c'est la position où le véhicule a été retrouvé et qu'il était sorti de la route, mais dans les deux cas, il est sémantiquement lié à la route.

La relation *porté par* est généralement définie au niveau du type. Elle peut être inférée depuis les relations constatées au niveau des instances. Par exemple, si des données ponctuelles fournies par l'utilisateur sont sur les routes dans 98% des cas, une relation *porté par* au niveau du type pour les données thématiques peut être inférée. Cette relation doit être confirmée par l'utilisateur avant qu'elle ne soit prise en compte.

L'extraction de la relation *porté par* au niveau des instances est aussi possible. Prenons par exemple un groupe de stations de météo, parmi lesquelles certaines stations se situent sur les routes, mais pas toutes. Alors seules les stations qui sont proches de routes sont portées par les routes.

L'importance de la relation *porté par* est toujours l'importance maximale. En effet, l'existence de la relation *porté par* engendre un effet sur l'importance des autres relations. Pendant l'extraction des relations initiales, le fait qu'il existe une relation *porté par* a un effet sur le calcul des importances des autres relations parce que les relations entre l'objet thématique et son espace de support sont toujours favorisées. Pour les relations qui sont entre l'objet thématique et son espace de support, l'importance de relation initiale doit être augmentée par certain pourcentage (30% par exemple).

Pendant l'identification des relations finales, l'importance initiale des relations est prise en compte. Pour cela, les relations finales avec l'objet support qui porte l'objet thématique auront une importance supérieure pendant la relocalisation de l'étape (e).

Pour la migration de données thématiques et pendant la phase de relocalisation, la relation *porté par* aide à restreindre l'espace de recherche. Par exemple, pour l'accident qui est en relation *porté par* avec la route *D7*, le nouveau positionnement de l'accident sur la base topographique finale doit être cherché uniquement sur la route correspondant à la *D7* dans la base topographique finale.

B.3.4 Quels sont les objets en relation et quelles sont les relations à extraire ?

Les relations à extraire sont celles qui existent entre l'objet thématique et des objets supports proches. Les objets supports qui sont en relation avec l'objet thématique ne sont pas uniquement les plus proches. Par exemple, en Figure 53, il faut extraire les relations de proximité entre l'accident et chacun des deux bâtiments, au cas où le plus proche serait supprimé dans la base finale.

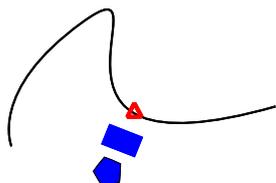


Figure 53. Objets topographiques proches d'un accident routier

Pour identifier les objets proches, nous proposons d'utiliser un critère de distance en considérant un tampon (buffer) autour de chaque objet thématique. La taille du buffer est définie en fonction de l'échelle de référence des bases de données supports initiales.

Les relations à extraire sont les relations qui peuvent être considérées lors du référencement initial des objets thématiques sur les objets supports (topographiques et caractéristiques), elles incluent les types de relations suivants:

- Relations topologiques d'inclusion
- Le positionnement fin au sein de cet objet lorsque l'objet thématique est inclus dans un objet support. Dans le cas des données thématiques ponctuelles sur réseau, le positionnement fin est décrit par distance curviligne par rapport aux extrémités du tronçon, et par rapport à des points caractéristiques situés sur le tronçon
- Relations de proximité et de direction par rapport à des objets supports proches

La liste des relations à extraire varie selon la thématique des données, les besoins de l'utilisateur, les spécifications de la base topographique finale et du produit final. Tous ces éléments devraient être modélisés pour pouvoir faire un calcul automatique des relations à extraire. Nous n'avons pas travaillé sur la modélisation de ces éléments, mais nous proposons d'associer une liste de relations pertinentes pour chaque binôme (type de données thématiques-application). Par exemple, nous avons défini une liste pour le binôme (accidents routier-analyse d'accidentologie), les détails de cette liste sont cités en chapitre C. Il existe des travaux qui visent à modéliser les préférences de l'utilisateur. Dans le travail de (Balley et al. 2014) sur la carte à la demande, les préférences de l'utilisateur sont modélisées dans des variables appelées : "variables utilisateur" (Figure 54). Les préférences incluent l'activité (par exemple la randonnée), des informations personnelles sur l'utilisateur, le contexte de l'utilisation et des préférences additionnelles. Par exemple, les bâtiments qui sont proches des accidents doivent être présents sans suppression.

User variables
<i>User need</i> Activity/Task to achieve
<i>User profile</i> Age Gender Familiarity with maps Visual disabilities Nationality
<i>Context of use</i> Position Time Season Device Typical use scale Situation
<i>User preferences</i> Likes/dislikes

Figure 54. Les variables qui modélisent les préférences de l'utilisateur dans le modèle de (Balley et al. 2014)

Dans le cas de la carte à la demande, les besoins de l'utilisateur sont modélisés et utilisés pour calculer les spécifications du produit final (Balley et al. 2012). En plus de calculer les spécifications, cette modélisation pourrait être utilisée pour identifier les types de relations à extraire et le calcul de l'importance des relations, surtout pour les relations finales. Pour chaque type de relations, il faut définir la modélisation choisie et la méthode d'extraction. Des méthodes manuelles ou semi-automatiques peuvent être exploitées pour modifier la liste des relations à extraire et les importances initiales.

Pour instancier les relations, nous proposons un modèle de référencement que nous allons détailler dans le paragraphe prochain. Le modèle est spécifique aux relations entre des données thématiques et des données topographiques. La méthode d'extraction associée à un type de relation crée une instance de ce type de relation suivant le modèle de référencement. La méthode calcule aussi les attributs des relations initiales extraites.

B.3.5 Modèle de référencement des données thématiques

Le référencement constitue un lien entre les données thématiques et le fond topographique. Des modes de référencement ont été détaillées au paragraphe A.1.4 (page 18). Nous voulons expliciter l'information de référencement pour utiliser cette information lors de la migration de données thématiques. Selon notre approche, le référencement est décrit par les relations spatiales entre les données thématiques et les données topographiques (avec leurs objets caractéristiques). La modélisation des relations que nous proposons est basée sur une ontologie qui modélise les contraintes relationnelles de généralisation. Dans la suite, nous allons présenter cette ontologie, ensuite, nous présentons l'extension de cette ontologie pour

répondre aux besoins de référencement des données thématiques. Pour comprendre le rôle de la modélisation des relations au sein de notre processus de migration de données thématiques, nous détaillons le fonctionnement de l'extraction et du stockage des relations. Enfin nous clarifions les différents traitements sur les relations et comment nous appelons les relations résultat de chaque étape de migration de données thématiques.

Ontologie des relations spatiales pour modéliser les contraintes relationnelles de généralisation

Durant l'élaboration de la présente thèse, une collaboration entre l'IGN France et l'Ordnance Survey (équivalent de l'IGN au Royaume-Uni) a été réalisée pour définir un modèle commun d'ontologie de relations et de contraintes relationnelles dans le contexte de la généralisation automatique. Le modèle proposé (Touya et al. 2012) est présenté en Figure 55.

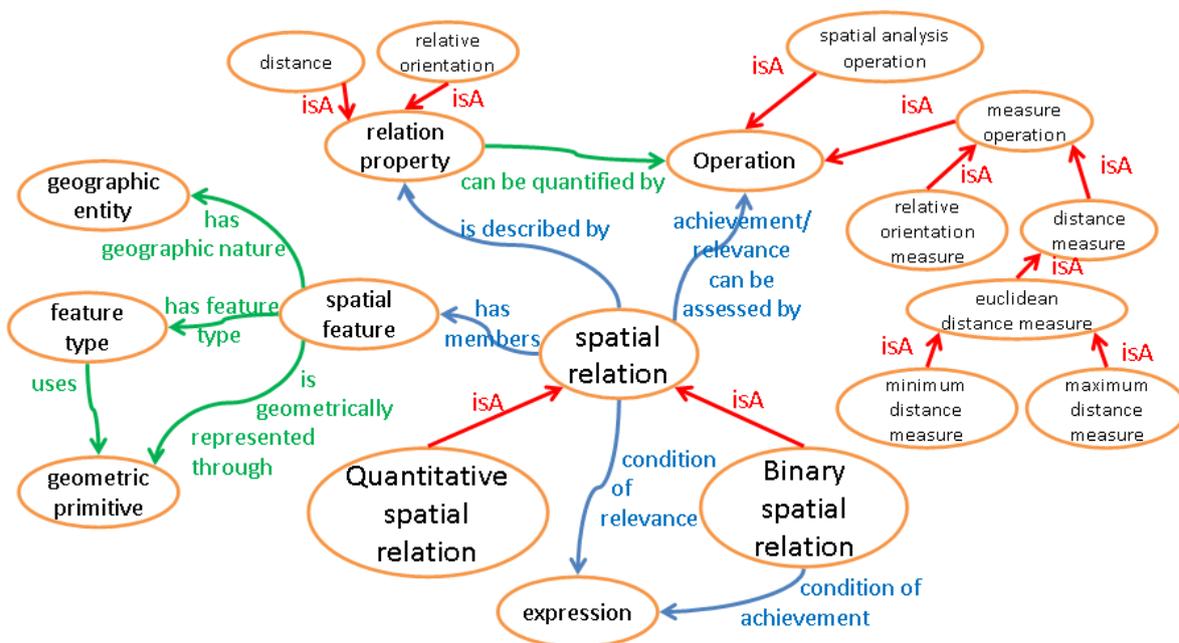


Figure 55. Modèle pour les relations spatiales, proposé par (Touya et al. 2012)

En prenant l'exemple d'une relation de proximité entre un pont (objet géographique ponctuel) et un bâtiment (objet géographique surfacique), l'ontologie sera instanciée de la façon suivante (Figure 56):

- Les deux membres de la relation spatiale sont le pont et le bâtiment, chacun des deux membres possède une nature géographique, un type d'entité géographique et une primitive géométrique,
- La relation a une seule propriété : la distance entre le pont et le centroïde du bâtiment,
- La distance est calculée par une fonction de mesure, qui est la distance euclidienne.

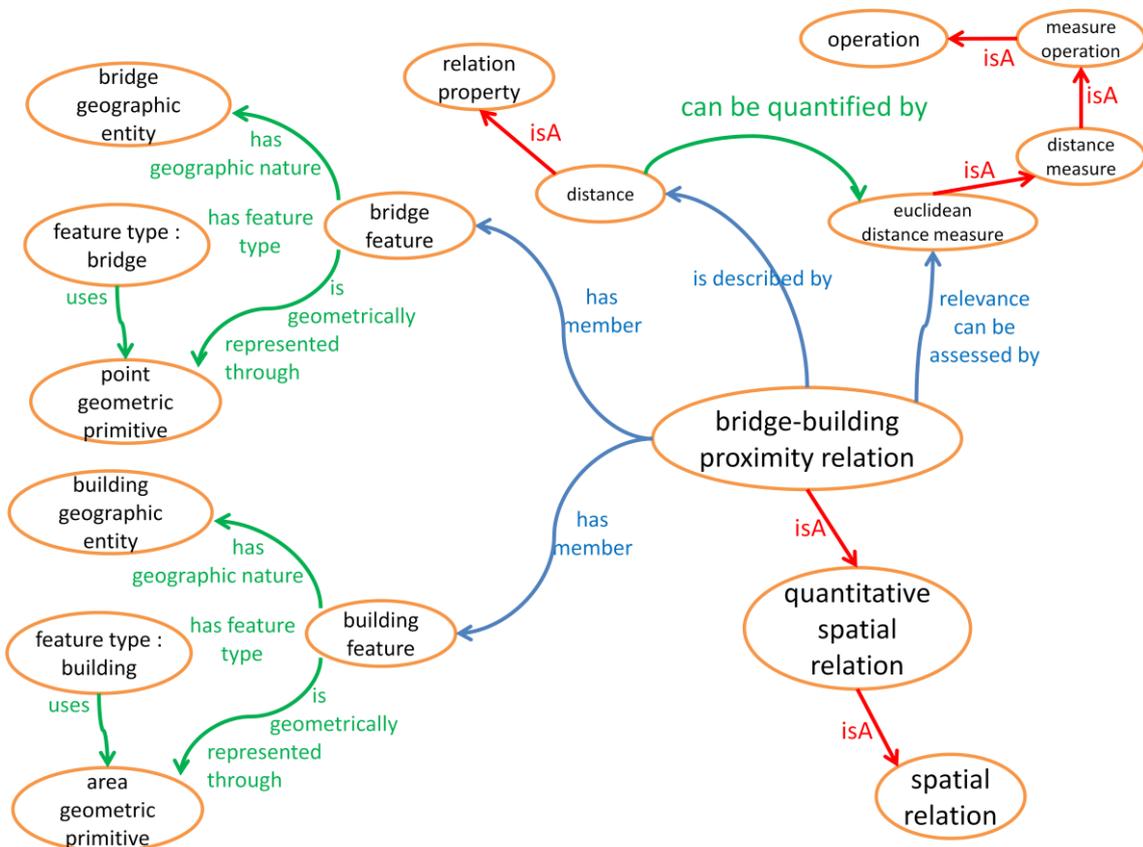


Figure 56. Instanciation du modèle de (Touya et al. 2012) pour la relation de proximité pont-bâtiment

De ce modèle, nous utilisons les éléments suivants:

- Les relations spatiales sont soit quantitatives c'est-à-dire représentées par une mesure quantitative (ex. : distance, angle...etc.), soit binaires c'est-à-dire représentées par un prédicat qui peut être vérifié ou non (cf. A.3.3).
- Les relations peuvent être pertinentes à extraire ou pas, par exemple, si deux bâtiments sont séparés par une rivière, il y a lieu de se demander si c'est pertinent d'extraire leur relation de proximité.
- Une relation quantitative peut toujours exister entre les objets (pas de condition d'existence), mais il n'est pas toujours pertinent de l'extraire.
- Chaque relation possède des propriétés qui la décrivent et permettent de décider de conserver ou non des détails sur la relation lors d'un changement de niveau de détail. Par exemple, la relation de proximité entre un objet thématique ponctuel et un bâtiment peut être caractérisée par une propriété de distance entre l'objet et la bordure du bâtiment.
- Une relation a des membres qui sont les objets en relation.
- Les relations sont détectées et caractérisées par des méthodes dédiées.
- La relation peut être presque présente, mais pas complètement présente, par exemple, la relation *presque parallèle*. Elle peut être entre deux prédicats, par exemple entre *disjoint* et *touche* (Figure 27, p.37).

Dans notre approche, une relation binaire est associée à un prédicat comme *est parallèle* ou *est contourné par*. Pour une paire d'objets donnés, un prédicat peut être vérifié ou pas. Il existe des familles de prédicats telles qu'un couple d'objets ne peut pas vérifier en même temps deux relations de la même famille. Par exemple, en se basant sur le modèle 9-intersections (9IM) (Egenhofer et Herring 1991), les relations topologiques entre deux régions peuvent être représentées par une famille de 8 prédicats (disjoint, touche, intersecte, égale, couvert par, à l'intérieur, couvre, contient). Le modèle de (Touya et al. 2012) considère chaque prédicat comme une relation séparée, donc les relations sont regroupées en famille de relations. Dans notre modèle de référencement, nous appelons "relation" ce que (Touya et al. 2012) appellent "famille de relation", et le prédicat qui est vérifié par les objets constitue pour nous une propriété de la relation. Par exemple, (Touya et al. 2012) considère une relation d'intersection, alors que nous considérons une relation topologique décrite par une propriété "prédicat topologique" qui peut valoir "intersecte".

Extension de l'ontologie des relations de (Touya et al.2012)

Nous voulons spécifier et prolonger l'ontologie parce qu'elle n'est pas assez précise pour permettre de modéliser les relations de référencement entre données thématiques et topographiques. Dans notre modélisation, nous considérons qu'une relation lie deux objets : le premier est un objet thématique et le deuxième est un objet topographique.

Nous proposons une spécialisation du modèle de (Touya et al. 2012) conformément à ce qui précède en spécifiant les objets spatiaux (spatial feature). Notre proposition est illustrée en

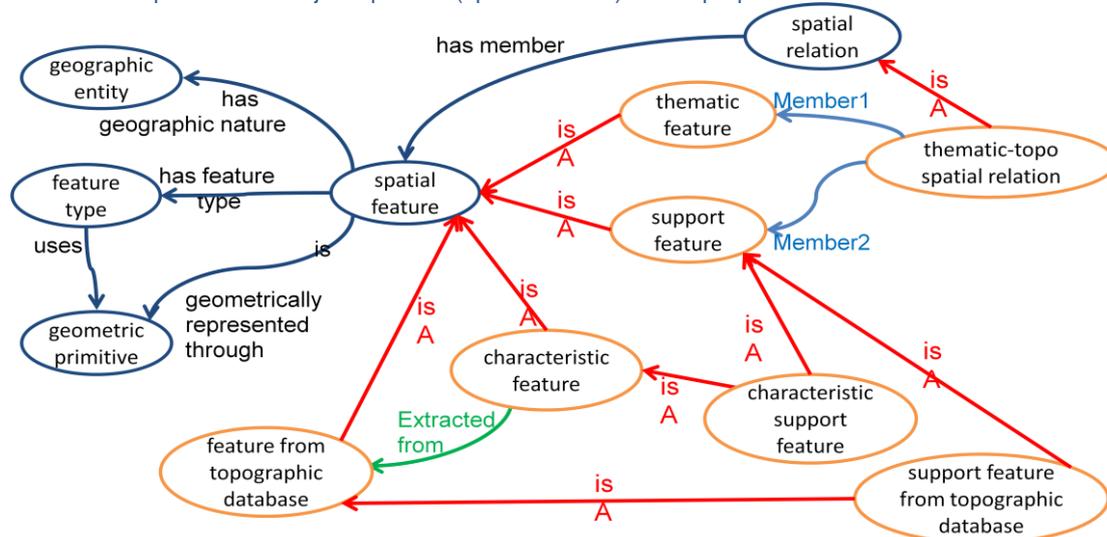


Figure 57 (les parties déjà existantes du modèle sont colorées en bleu foncé). Dans notre spécialisation du modèle, une relation spatiale a deux membres: le premier membre est thématique, et le deuxième membre est un objet support. Un objet support peut être un objet de la base de données topographique ou un objet caractéristique qui est construit à partir d'objets de la base de données topographique. Un objet topographique peut devenir un objet support ou pas, et un objet caractéristique peut devenir un objet support ou pas.

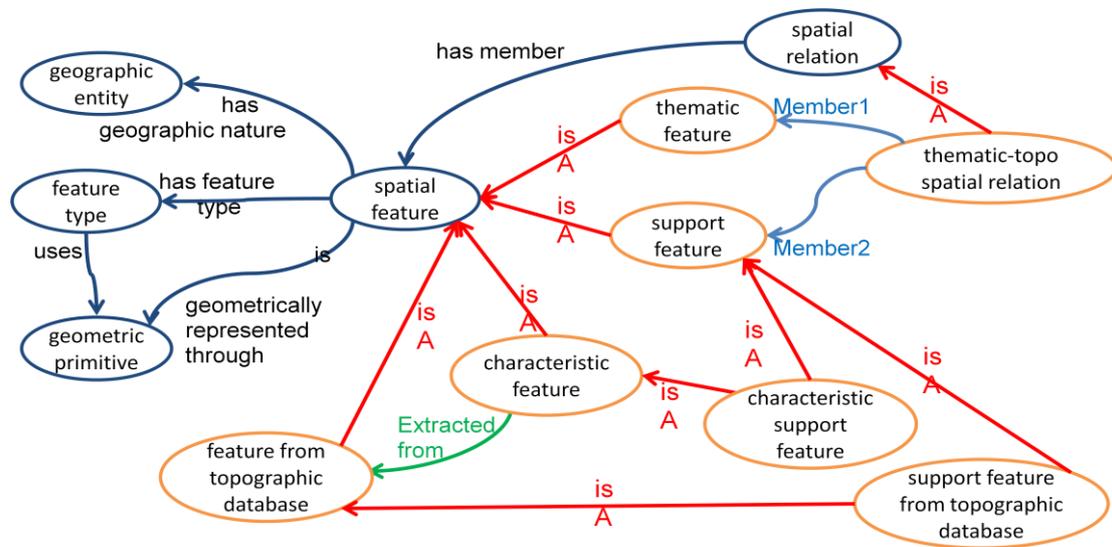


Figure 57. Extension de l'ontologie pour représenter les relations entre les données thématiques et les données topographiques

Fonctionnement de l'ensemble des éléments pour extraire et stocker les relations

Avant de détailler notre modèle de référencement des objets thématiques sur les données topographiques, nous montrons en Figure 58 les différents éléments qui participent à l'extraction et au stockage des relations. Ces éléments illustrent les étapes (a) et (b) de notre processus de migration de données thématiques tel que décrit en page 58 (Figure 50). Les objets caractéristiques doivent être extraits avant l'extraction des relations. Le modèle de référencement sert pour représenter et stocker les relations extraites. Les objets supports sont tous les objets caractéristiques et topographiques qui sont en relation avec les données thématiques. Ces objets doivent être clairement identifiés à part, car pendant la phase d'appariement des données topographique initiales et finales (étape (c) dans le processus de migration de données), les liens d'appariement pour ces objets seront particulièrement importants parce qu'ils jouent un rôle essentiel à la relocalisation des données thématiques.

En Figure 58, nous résumons les étapes (a) (l'extraction des objets caractéristiques) et (b) (l'extraction des relations). La figure montre aussi l'information stockée à la fin de ces deux étapes.

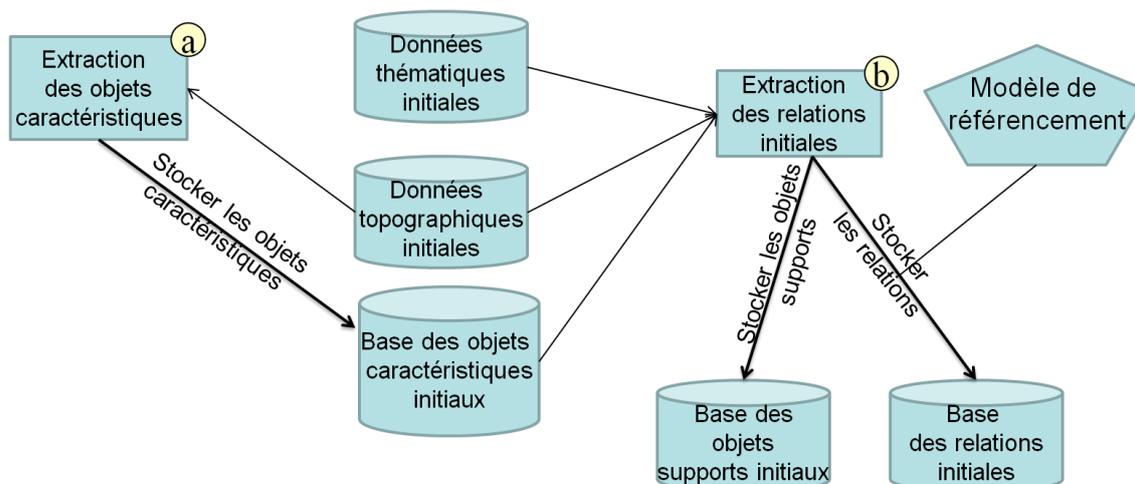


Figure 58. Fonctionnement de l'ensemble des éléments pour extraire et stocker les relations

Catégories des relations en vue du processus de migration de données

Pour expliquer l'effet des différentes étapes sur les relations au sein de processus de migration de données thématiques, nous classifions les relations en nous basant sur les types d'objets en relation. Nous identifions trois catégories de relations (Figure 59):

- Relation initiale: entre un objet thématique initial et un objet support initial, Cette catégorie de relation est générée lors de l'étape (b)
- Relation finale attendues: relation espérée entre un objet thématique et un objet support final. Cette catégorie de relation est générée lors de l'étape (d)
- Relation finale: entre un objet thématique et un objet support final. Cette catégorie de relations est constatée dans le résultat final de traitement (après l'étape (e)).

En plus de ces trois catégories, nous avons des relations courantes: ce sont des relations constatées dans les données en cours de recalage. Nous constatons cette catégorie de relation pendant l'étape de relocalisation (e). En effet, nous pouvons avoir des phases intermédiaires avant d'obtenir le résultat final. Pendant ces phases, nous appelons les relations "relations courantes".

Ce classement est similaire aux notions de valeur initiale, valeur courante, valeur but du modèle AGENT (Ruas 1999).

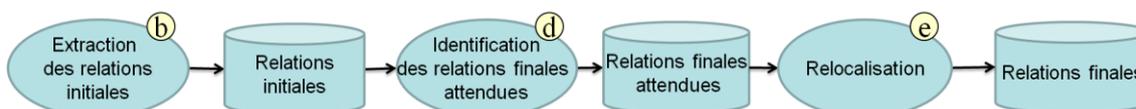


Figure 59. Les différentes catégories de relations dans le processus de migration de données

B.3.6 Schéma de données proposé pour le référencement des données thématiques sur les données support

Le but de notre modèle de référencement est d'explicitier les relations et de faciliter leur manipulation. Il permet de créer des instances qui représentent les relations. Ces instances

peuvent être manipulées (comme nous allons voir dans l'étape *d*) et elles font parties des entrées pour la phase de relocalisation (étape *e*). Notre modèle ne remplace pas les modélisations de relations que nous avons étudiées dans l'état de l'art, mais il les encapsule pour répondre à nos besoins.

Le modèle de référencement proposé est valable pour représenter les relations entre les données thématiques et les données support. Ce modèle est utilisé pour représenter les trois groupes de relations

La Figure 60 montre une version simplifiée du schéma de données proposé. Ensuite, nous détaillerons chaque partie dans les paragraphes qui suivent.

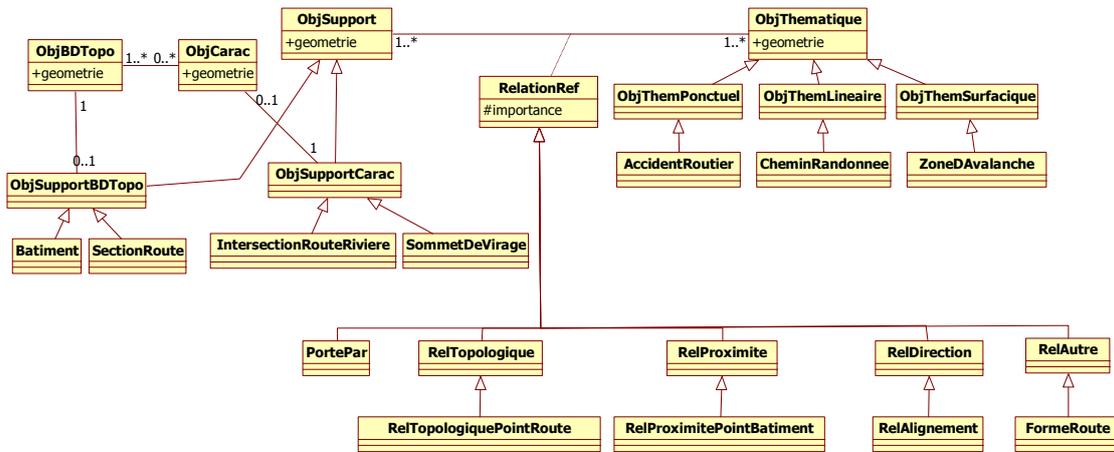


Figure 60. Version simplifiée de notre modèle de référencement des données thématiques sur les données topographiques

Comme indiqué dans le paragraphe précédent, une relation de référencement (classe *RelationRef*) lie deux objets : un objet support (*ObjSupport*) et un objet thématique (*ObjThematique*). La classe *RelationRef* est une classe d'association entre *ObjSupport* et *ObjThematique*. Un objet support peut être en relation avec un ou plusieurs objets thématiques. Un objet thématique peut être en relation avec un ou plusieurs objets supports. Cette modélisation est utilisée pour modéliser les relations initiales entre données topographiques et thématiques initiales et les relations finales attendues (entre données topographiques finales et données thématiques en cours de recalage). L'attribut *importance* de la *RelationRef* représente l'importance initiale ou finale de la relation (discuté en B.3.2). Un objet thématique peut être ponctuel (par exemple un accident routier), linéaire (par exemple un itinéraire de randonnée) ou surfacique (par exemple une zone d'avalanche). Nous distinguons ces trois types parce que les relations qui sont valables pour un type d'implantation géométrique, ne sont pas toujours valables pour les autres. De plus, les traitements sont différents dans la plupart des cas en fonction du type d'implantation géométrique. Par exemple, la migration de données ponctuelles est différente de la migration de données linéaires.

L'objet support est issu soit d'un objet caractéristique (la classe *ObjSupportCarac* hérite de la classe *ObjCarac*), soit d'un objet topographique (*ObjSupportBDTopo* hérite de *ObjBDTopo*). Un objet caractéristique est issu d'un objet topographique (e.g. un sommet de virage) ou plusieurs objets topographiques (e.g. une intersection route-rivière).

En A.3.5, nous avons classifié les relations de référencement en cinq catégories qui correspondent aux cinq classes de haut niveau sous la classe *RelationRef* de la Figure 60 :

- La relation spéciale *porté par*
- Relation topologique
- Relation de proximité
- Relation de direction
- Autre : pour les autres types de relations qui n'entrent pas sous les catégories précédentes. Par exemple, la description de la position relative d'un accident le long d'une route en modélisant la forme de la route autour l'accident. Pour les données linéaire, la relation de mouvement *contourne* entre sous cette catégorie.

Notons que chaque classe de relation lie une seule classe thématique et une seule classe topographique. Par exemple, une classe de relation de proximité Point-Bâtiment est uniquement une relation entre un objet thématique ponctuel et un objet topographique de classe bâtiment. La classe de l'objet thématique en relation et la classe de l'objet support n'est pas forcément au niveau le plus bas. Nous pouvons définir des relations qui ne sont valables que pour les accidents, définir d'autres relations qui sont valables pour tous les objets de type ponctuel.

Un type de relation peut être modélisé en utilisant plusieurs modélisations, par exemple, les relations topologiques peuvent être modélisées en utilisant les modèles 4IM ou 9IM ou DE-9IM ou DE-9IMBC (ces modèles sont détaillés dans A.3.5). Chaque modélisation doit être représentée par une classe séparée.

La modélisation des objets supports

L'objet support est soit un objet de la base de données topographique, soit un objet caractéristique issu de la base topographique. Un objet topographique peut être ponctuel, linéaire ou surfacique. La dimension des objets peut changer entre la base de données initiale et la base de données finale.

Certains objets sont traités différemment selon leurs propriétés, notamment leur implantation géométrique. Par exemple, un bâtiment peut être considéré comme un objet surfacique et il peut, aussi, être considéré comme un point lors de traitements (Duchêne 2004). En suivant la logique proposée par (Duchêne 2004), les objets peuvent être regroupés en ajoutant un niveau de classe abstraite intermédiaire entre la classe *objet support* et les classes concrètes (ex. *bâtiment*). Le regroupement est effectué en fonction de leurs propriétés vis-à-vis de la migration de données thématiques, en particulier des types de relations de référencement qu'ils vont avoir avec les objets thématiques et de la manière de les calculer, qui dépendent entre autre du type d'implémentation géométrique.

Pour chaque cas d'étude, il faut effectuer trois tâches:

- identifier les relations à extraire (cf. B.3.4)
- déterminer des types d'objets caractéristiques qui doivent être extraits
- concevoir les méthodes d'extraction de ces objets

Les objets caractéristiques à extraire se déduisent des relations à extraire et des objets déjà présents dans la base de données topographique.

Les deux tâches sont effectuées manuellement, les critères de choix sont liés au but de la carte, aux préférences de l'utilisateur et aux spécifications des bases initiales et finales. Par exemple, si les données thématiques sont des accidents et si le but de la combinaison finale est d'analyser les causes des accidents, les objets caractéristiques à extraire incluent au moins les ronds points, les intersections route-rivière et les virages (représentés sous forme de sommets).

Lors de l'extraction, un objet caractéristique garde un lien avec les objets d'origine à partir desquels il est construit (Figure 61). Le lien sera utilisé plus tard pendant la phase de relocalisation pour retrouver le correspondant de l'objet caractéristique dans la base topographique finale. Un objet caractéristique peut avoir un lien avec un objet topographique (e.g. sommets de virage d'un tronçon routier) ou deux (e.g. intersection route rivière) ou plus que deux (e.g. rond-point).

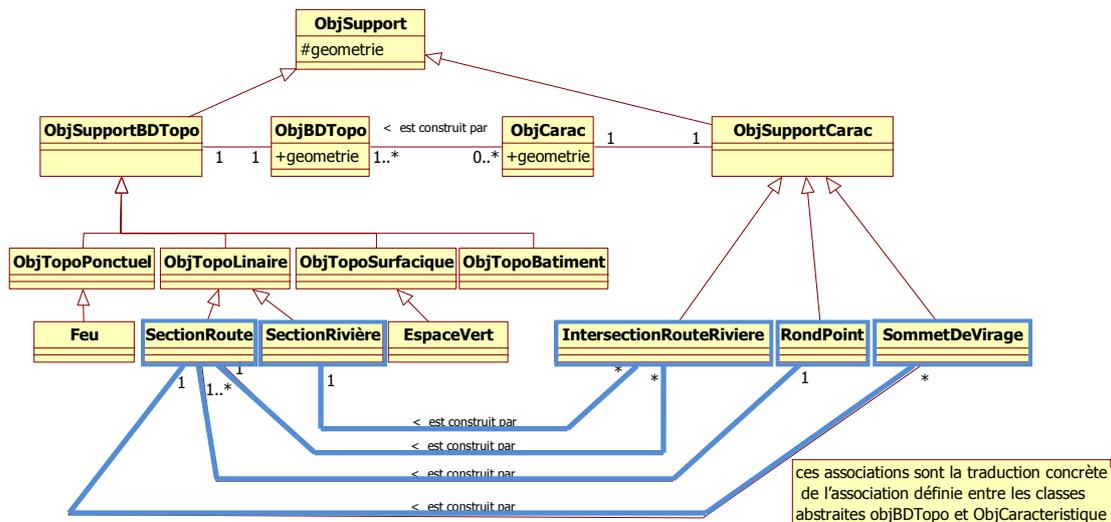


Figure 61. Modélisation des objets supports

Modélisation de la relation « porté par »

Cette relation décrit le lien sémantique fort entre l'objet thématique et son espace support. L'objet thématique ne peut exister sans son espace de support. Cette relation n'a pas d'attributs: elle existe ou elle n'existe pas entre un objet thématique et un objet support (cf. le diagramme de classe en Figure 60). L'existence d'une relation *porté par* entre un objet thématique et un objet support aura un effet pendant l'identification des relations finales

attendues. Cela, parce que les relations spatiales entre l'objet thématique et son espace de support sont prioritaires sur les relations entre l'objet thématique et les autres objets.

Détails sur la modélisation des relations topologiques

Une relation topologique est décrite par un prédicat, qui prend ses valeurs dans une énumération dédiée à la modélisation de la topologie choisie. La Figure 62 montre la modélisation de deux relations topologiques. La première est la relation topologique région-région suivant le modèle 9IM et la deuxième est la relation topologique point-ligne selon une modélisation proposée par nous, dans laquelle l'objet ponctuel peut être "sur" la ligne (inclus à l'intérieur de la ligne ou confondu avec l'un de ses extrémités), clairement disjoint de la ligne ou dans une situation entre ces deux relations, compte tenu de la résolution de données et des seuils de lisibilité pour des données finales affichées sous forme d'une carte (cf. A.3.5). Chacune de ces deux classes de relation topologique est associée à une énumération. L'énumération contient les prédicats décrivant les différentes configurations spatiales modélisées.

Le but de la méthode *calcule()* au sein de la classe *RelTopologique* est de trouver le prédicat qui correspond à la relation topologique actuelle entre l'objet thématique et l'objet topographique. Par exemple, pour la classe de relations topologiques région-région de la Figure 62, la méthode remplit les neuf valeurs booléennes de la matrice 9IM (voir Figure 25 p.35) en faisant des calculs géométriques sur les objets, puis en déduit le prédicat qui correspond à la relation topologique actuelle et met ce prédicat dans l'attribut *predicat*.

Dans la taxonomie proposée par (Touya et al. 2012) en

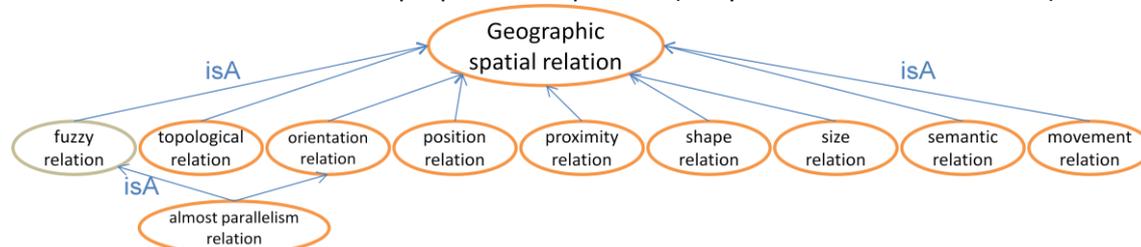


Figure 21 (page 31), une relation peut être floue (fuzzy en anglais). L'utilisation des relations ou des prédicats flous nous permet de clarifier les relations qui peuvent être jugées douteuses dans l'état final. Pour représenter des relations topologiques floues, nous proposons deux choix :

- soit ajouter un prédicat qui indique qu'une relation est presque présente. Le nom du prédicat doit commencer par le mot *presque*, par exemple, un bâtiment est *presque parallèle* avec la route.
- soit ajouter un prédicat qui indique que la relation est entre deux autres prédicats. Le nom de prédicat doit commencer par le mot *entre*, par exemple, le prédicat *entreDisjoint-Touche*.

Par contre, une relation finale attendue ne sera justement jamais floue (même si la relation existant de fait entre les deux objets concernés peut être évaluée comme floue initialement ou en cours de recalage).

En regardant les prédicats en Figure 62, nous avons utilisé le premier choix pour la relation topologique point-linéaire et nous avons utilisé le deuxième choix pour la relation topologique région-région.

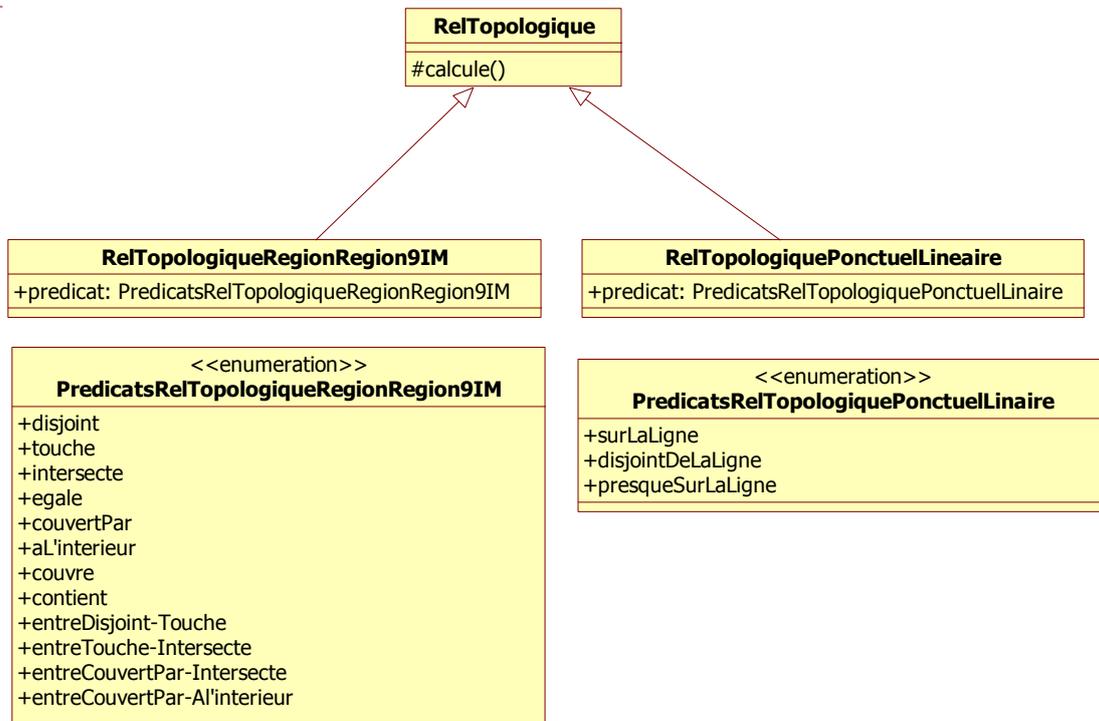


Figure 62. Représentation des relations topologiques

Détails sur la modélisation des relations de proximité

Une relation de proximité peut être modélisée de manière quantitative (elles sont alors décrites par une propriété de distance), ou par des mots descriptifs qui représentent la distance (loin, proches, très loin, très proche).

Une relation de proximité entre deux objets est valable au sein d'un espace support des deux objets (de la relation): carte, ou autre objet qui doit alors porter les 2 objets.

Pour représenter la proximité de manière quantitative, cet espace support doit être muni d'une distance (espace métrique). Par exemple, distance euclidienne sur la carte, distance curviligne le long d'un objet linéaire.

Pour deux objets qui appartiennent à un espace support muni d'une distance, la relation de proximité peut toujours être définie mais elle n'est pas toujours pertinente. La pertinence des relations varie. Par exemple, la distance sur la route pour les accidents routiers est pertinente.

Dans le cas d'une distance curviligne le long d'un objet support linéaire, il est nécessaire d'orienter le support et d'utiliser une distance orientée (qui n'est alors plus une distance au sens mathématique du terme puisque elle n'est pas symétrique). Nous orientons la route en

fonction du sens de la polyligne qui constitue sa géométrie dans la base de données (Figure 65). Si l'orientation n'est pas la même entre les bases initiale et finale, la distance finale attendue sera inversée.

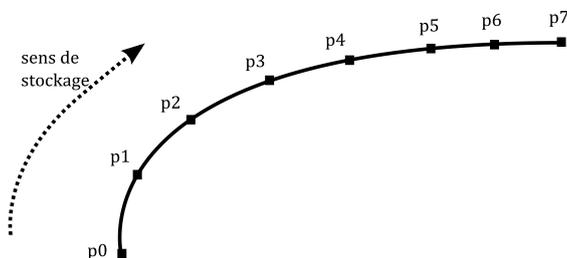


Figure 63. Orientation de la route basée sur sa géométrie

Par exemple, si nous avons un accident $a1$ qui est à 50 mètres d'un pont $pont1$ sur une route, deux positions sont possibles pour l'accident (avant et après le pont). Orienter la route et utiliser une distance curviligne orientée permet de définir la position de l'accident sans ambiguïté. En utilisant cette méthode, si l'accident est avant le pont la distance $dist(pont1, a1)$ vaudra $-50m$, s'il est après le pont la distance $dist(pont1, a1)$ vaudra $+50m$. Dans la Figure 67, la distance $dist(pont1, a1)$ vaut $-50m$.

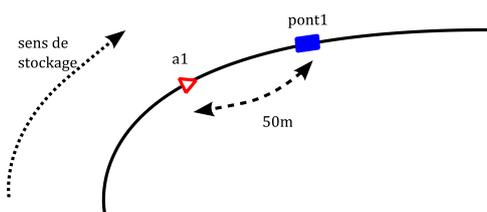


Figure 64. Représentation d'un accident qui s'est passé à proximité d'un pont

La Figure 65 présente deux exemples de relations de proximité. La première relation est entre un point sur le réseau et un rond-point. La deuxième est entre deux points sur le réseau.

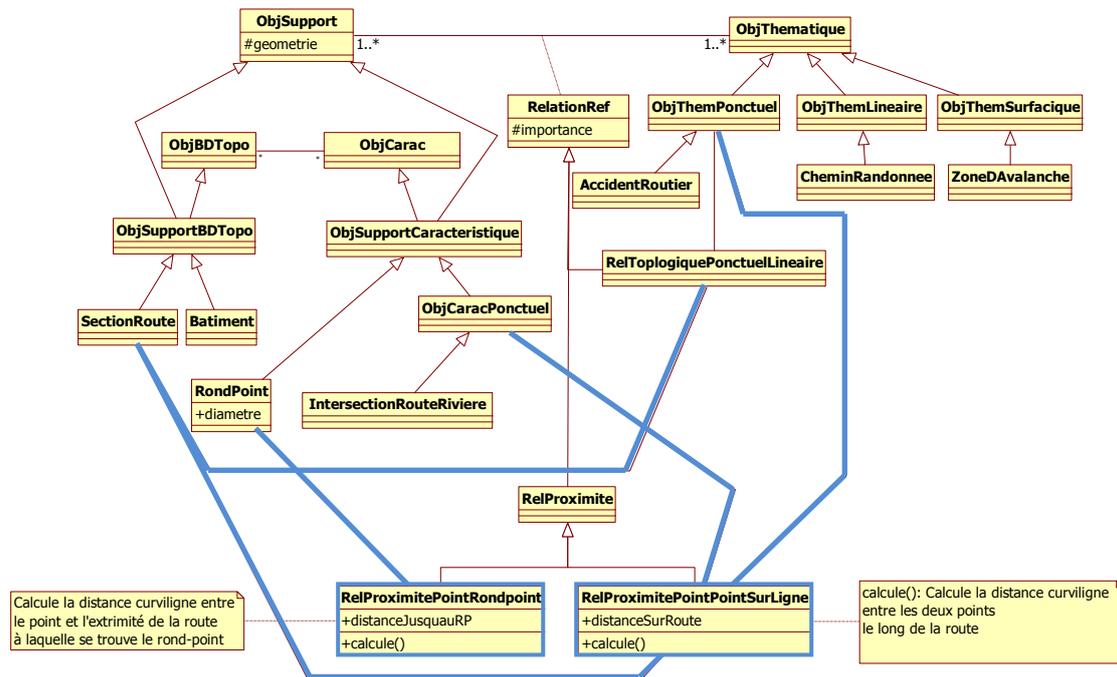


Figure 65. Représentation des relations de proximité dans le modèle de référencement

Détails sur la modélisation des relations de direction et des autres relations

La direction est représentée soit par un angle (par exemple, l'orientation absolue du segment liant un accident et un bâtiment), soit par une direction (Papadias et Theodoridis 1997). Les attributs varient selon la modélisation choisie. En

Figure 66, nous présentons un exemple d'une modélisation des relations de direction en utilisant la modélisation de (Goyal et Egenhofer 2000) et la modélisation de (Papadias et Theodoridis 1997).

Pour la catégorie "autre relation", nous trouvons les relations qui n'entrent pas sous les autres catégories comme les relations dite "de mouvement" (par exemple, la route contourne l'église), la relation de positionnement fin d'un point sur la route (par exemple, la forme de la route autour d'un accident).

Exemple d'une modélisation multiple pour une même relation

Une même relation peut être modélisée de plusieurs façons. Suivant cette logique, dans notre modèle, il peut y avoir plusieurs classes de modélisation pour une même relation. En

Figure 66, nous montrons un exemple de la modélisation multiple d'une même relation. La relation à modéliser est la relation de direction entre un objet thématique surfacique et un objet topographique surfacique. Elle est représentée de deux façons différentes. La première façon utilise la modélisation de (Goyal et Egenhofer 2000), basée sur une matrice de 9 valeurs entières (cf. A.3.5 et Figure 27 page 37). La deuxième façon utilise la modélisation de (Papadias et Theodoridis 1997) basée sur des *Rectangles Minimum Englobants* et les intervalles d'Allen.

Cette modélisation décrit la direction en 169 prédicats organisés dans une matrice 13x13 (voir Figure 32, p39). Dans notre modèle, cette modélisation est traduite par la définition d'une classe qui possède deux attributs, le premier est l'ordre horizontal R_i et le deuxième est l'ordre vertical R_j . La méthode *calcule* fait le calcul nécessaire pour détecter le prédicat parmi les 169 prédicats possibles, puis assigne les bonnes valeurs aux R_i et R_j .

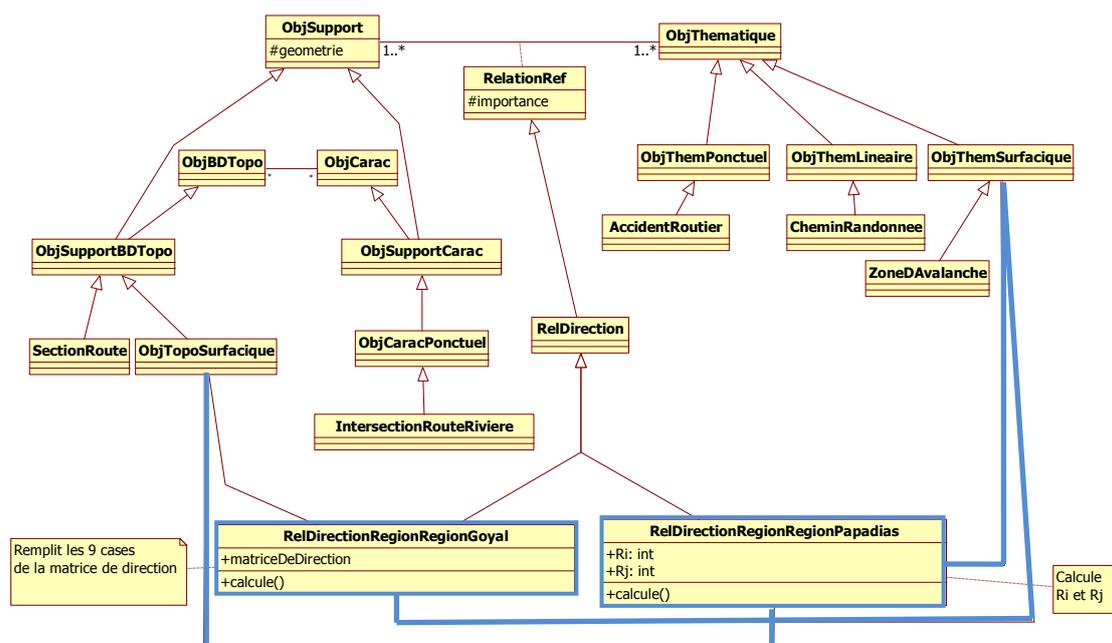


Figure 66. Représentation des relations de direction dans notre modèle

Instanciation du modèle pour l'exemple jouet

La Figure 67 montre le diagramme d'objet une fois les relations initiales instanciées pour notre deuxième exemple-jouet (voir B.1, p.54). Le diagramme inclut les éléments suivants:

L'objet thématique :

a1 : l'objet thématique accident

Les objets topographiques en relation :

tronçon1: le tronçon de route qui porte l'accident *a1*

tronçon2, tronçon3... tronçon6: les tronçons du rond point

Les objets caractéristiques :

rondPoint1: l'objet support qui représente le rond point

carrefourSimple1: l'objet support qui représente le début du tronçon1

Les objets supports directement issus des objets topographiques:

tronçonSupport1, tronçonSupport2.... tronçonSupport6, les tronçons supports issus de tronçon1, tronçon2...tronçon6 respectivement

Les relations :

porteParR1 : relation *porté par* entre l'accident *a1* et le tronçon *tronçonSupport1*

proximiteR1: relation de proximité entre l'accident *a1* et le début de la route *pointDébutTronçon1*.

proximiteR2 : relation de proximité entre l'accident *a1* et le rond point *rondPoint1*

topologiqueR1: relation topologique d'inclusion de *a1* sur le tronçon 1

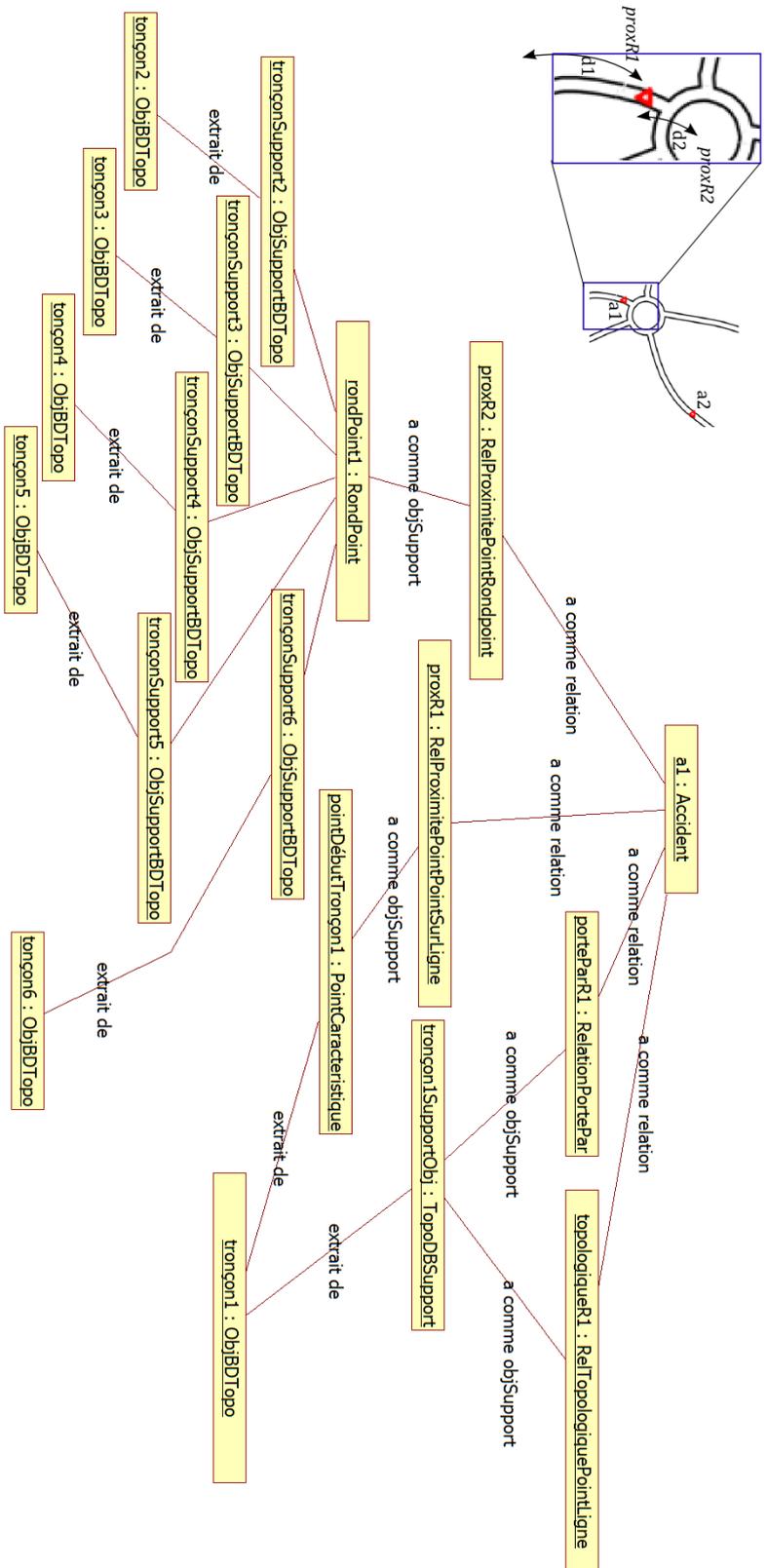


Figure 67. Diagramme d'objets, illustre la représentation de relations en utilisant notre modèle

B.4 Identification des relations finales attendues

Suivant l'approche de la migration de données thématiques que nous proposons (voir B.2), et après l'extraction des relations initiales (étape (b)), nous disposons des relations qui concernent la combinaison initiale. Pour bien localiser les données thématiques sur la base topographique finale en prenant en compte ces relations, il faut générer les relations correspondantes attendues dans la combinaison finale. Pour le faire, il faut tenir compte de la variation de la représentation du monde réel entre la base topographique initiale et la base topographique finale. En B.4.1, nous décrivons comment l'appariement de données permet de mettre en évidence cette variation. En B.4.2, nous expliquons pourquoi les relations finales attendues ne sont pas forcément identiques aux relations initiales. En B.4.3, nous expliquons notre méthode pour identifier les relations finales attendues.

B.4.1 Appariement entre les bases topographiques initiale et finale

La différence entre les bases topographiques initiale et finale peut concerner le type de la géométrie, la taille, la localisation ou la forme des objets, le type de l'objet support. Les objets peuvent être agrégés avec d'autres ou supprimés. La dimension de leur géométrie peut changer. Par exemple une rivière est représentée par une surface puis par une ligne. Ces changements induisent une modification des relations spatiales entre les données thématiques et les données topographiques dans la combinaison finale. Pour cela il faut un lien d'appariement pour modéliser ces changements pour pouvoir les prendre en compte pendant l'identification des relations finales attendues.

Le rôle de l'appariement dans notre modèle est de déterminer les liens de correspondance entre les données topographiques initiales et les données topographiques finales. Les liens aident à découvrir certains changements, comme la non représentation d'un objet dans la base finale. L'appariement peut être complété par des méthodes additionnelles pour extraire des informations sur les différences de représentation entre les bases topographiques initiale et finale. Par exemple, un rond point de la base initiale qui est représenté par un carrefour dans la base finale, un changement de taille, une agrégation (par exemple un bâtiment agrégé avec d'autres). En Figure 68, nous illustrons des liens de correspondance pour un cas jouet.

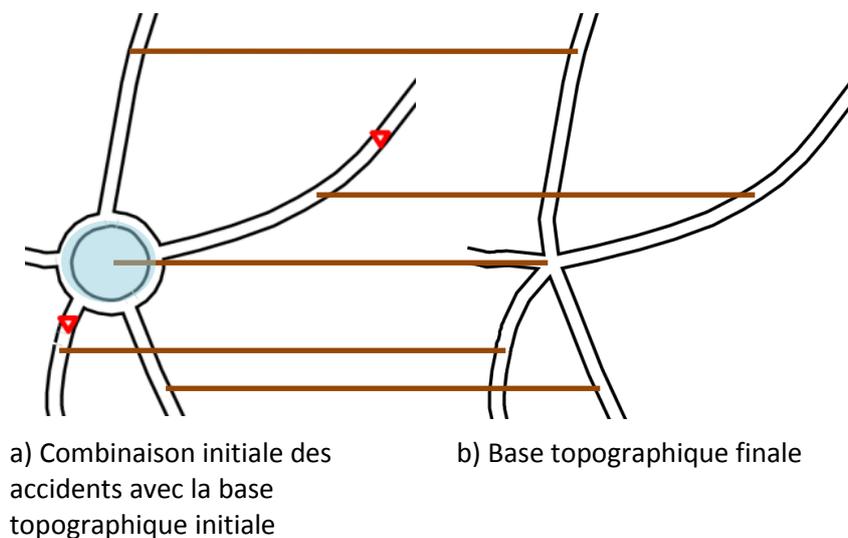


Figure 68. Appariement des données topographiques

Les liens d'appariement peuvent être de cardinalité 1-1 ou 1-N ou N-M ou N-1 ou 1-0.

Cardinalité	Signification
1-1	Un objet du monde réel est représenté par un objet dans chacune des deux bases
N-1	N objets dans la base initiale se retrouvent agrégés en 1 objet dans la base finale qui peut être de même niveau d'abstraction. Par exemple, N tronçons de route appariés avec un tronçon de route. L'objet final peut avoir un niveau d'abstraction différent. Par exemple, N tronçons de route qui forme un rond-point. Ils sont appariés avec un carrefour.
1-N	Découpage d'un objet (si le niveau de détail augmente entre les bases initiale et finale)
N-M	Découpage différent d'un objet du monde réel en objets géographiques dans les deux bases
1-0	L'objet de la base initiale n'est pas représenté du tout dans la base finale

Tableau 3. Les différentes cardinalités des liens d'appariement et leur signification

Les routes n'ont pas le même découpage entre une base et une autre, alors pour améliorer le résultat d'appariement des réseaux routiers, nous avons besoin d'assurer une continuité. La continuité peut être assurée par la construction des *strokes*. En effet, les routes sont susceptibles d'être vues comme des grandes lignes continues même avec les intersections avec d'autres routes, ces grandes lignes sont appelées des *strokes* (Thomson et Richardson 1999). La construction de *strokes* à partir d'un réseau routier utilise des critères attributaires

(deux tronçons de routes portant le même nom de route appartiendront au même stroke), et des critères d'angles (à une intersection, deux tronçons de route dont l'angle est proche de l'angle plat appartiendront au même stroke). Des travaux ont été menés pour assurer une continuation des *strokes* même à travers les ronds-points comme (Luan et Yang 2010).

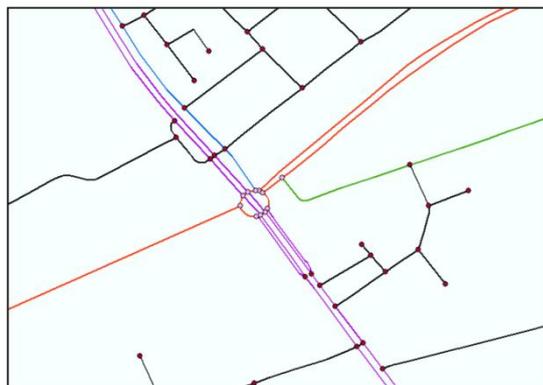


Figure 69. Exemple de création des strokes avec une continuation à travers des ronds-points (Luan et Yang 2010)

Pour recalculer les données thématiques, le cas le plus favorable est le cas 1-1. Avec les strokes, nous pouvons diminuer les liens N-1, 1-N et N-M. L'appariement des strokes peut être complet (les deux strokes se correspondent complètement y compris les deux extrémités de chaque stroke) ou partiel (un stroke est apparié avec une partie de l'autre stroke ou l'inverse).

B.4.2 Modification et transformation des relations

Pourquoi nous avons besoin de modifier ou transformer les relations entre les objets thématiques et supports

En fonction de l'information fournie par le lien d'appariement sur les changements entre les bases de données topographiques initiale et finale, certaines relations initiales doivent être légèrement modifiées par ajustement de leurs valeurs d'attributs. D'autres relations doivent être transformées en une relation d'un autre type. Le changement a pour but de faire correspondre au niveau d'abstraction et la résolution de la base topographique finale et assurer la visibilité. Prenant l'exemple d'un accident proche d'un carrefour initialement, lors d'un changement de niveau de détail supposons que le carrefour occupe plus de place : soit on change la relation de proximité pour que l'accident soit clairement disjoint du carrefour, soit on transforme la relation en relation d'inclusion avec le carrefour. Dans l'exemple jouet de B.1, la relation initiale de proximité avec le rond point doit être modifiée, parce que le rond point est représenté par un carrefour dans la base topographique finale. Pour l'accident qui est proche du rond point, la relation de proximité est convertie en relation *porté par* avec le carrefour (objet caractéristique) de la base finale. Dans certains cas, les relations doivent être modifiées même si l'objet de référence est identique dans les deux bases topographiques. Par exemple, si on imagine un bâtiment disjoint mais proche d'une zone d'avalanche dans les données initiales, il peut être pertinent de modifier sa relation en une relation d'adjacence dans les données finales, si la base topographique finale a un niveau de détail moindre et si la distance initiale était faible par rapport au nouveau niveau de détail.

But de l'étape et entrées/sorties

Le but de cette étape est d'identifier, pour chaque objet thématique, les relations finales attendues avec les données topographiques et leur importance.

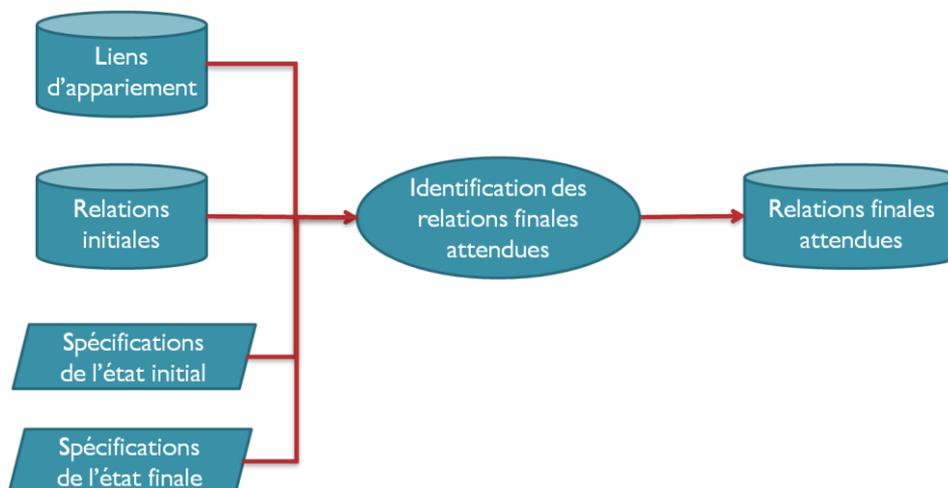


Figure 70. Entrées/sorties de l'étape d'identification des relations finales attendues

Les éléments à prendre en compte pendant l'identification des relations finales attendues pour un objet thématique donné sont les suivants (Figure 70) :

- les relations de l'objet thématique avec les objets supports initiaux,
- les liens d'appariement entre les objets supports initiaux et les objets topographiques et caractéristiques de la base topographique finale,
- les préférences de l'utilisateur et le but de la carte,
- les spécifications de l'état initial (niveau de détail, etc.),
- les spécifications de l'état final.

Ces cinq éléments sont les entrées du module. Les relations finales attendues sont la seule sortie.

Il est possible que les préférences de l'utilisateur soient déjà représentées dans les spécifications, c'est le cas pour le traitement de la carte à la demande proposé par (Balley et al. 2012) présenté en B.7.3.

Modélisation des spécifications de l'état final

Les spécifications de l'état final contiennent des informations sur le produit voulu. Pour prendre les spécifications en compte pendant l'identification des relations finales attendues, il faut que ces spécifications soient modélisées. Les spécifications modélisées seront ensuite prises en compte pendant les modifications et les transformations des relations initiales.

Dans le cadre de la thèse, nous n'avons pas testé l'utilisation d'une modélisation des spécifications pendant la modification et transformation des relations. Les tests ont été réalisés en prenant en compte une description non formalisée de l'état final. Pour le cas

d'étude réel que nous allons détailler en chapitre C, la description de l'état final et les préférences de l'utilisateur sont communiquées par l'utilisateur concerné. La modification et la transformation des relations initiales sont effectuées manuellement en prenant en compte cette description non formalisée.

Dans le cadre de la carte à la demande, des travaux ont été menés pour modéliser les spécifications du produit final attendu avec l'intégration des préférences de l'utilisateur dans les spécifications. Le but de cette modélisation est de pouvoir prendre les spécifications en compte pendant la fabrication automatique du produit final (une carte ou des données traitées). Ces travaux sont présentés en B.7.3.

B.4.3 Détail du traitement de l'identification des relations finales attendues

Nous proposons de définir des règles de transformation pour les relations. Ces règles reflètent les préférences de l'utilisateur et les spécifications de la carte. Chaque règle définit la relation finale attendue et calcule ses attributs.

Les règles doivent être identifiées a priori pour chaque type de relation, en fonction des transformations que les objets topographiques sont susceptibles de subir. Il n'est pas nécessaire de connaître la base topographique finale pour définir les règles. Certaines sont liées au message cartographique, en particulier pour les relations liées aux seuils de perception (par exemple, une relation de proximité avec un rond point devient une relation « porté par » avec le carrefour).

Un des calculs effectués pendant l'application des règles est le calcul des importances des relations finales attendues. Nous avons présenté en B.3.2 la différence entre l'importance des relations initiales et l'importance des relations finales. L'importance d'une relation initiale est dérivée du degré de la prise en compte de la relation au moment de la saisie. Pour la relation finale, l'importance est calculée en fonction de l'importance de la relation initiale, des préférences de l'utilisateur, du but de la carte et des spécifications de l'état final (l'échelle, le niveau de détail, etc.).

Les modifications possibles sont les suivantes :

- Une relation initiale génère une relation finale attendue de même type
- Une relation initiale génère une relation finale attendue de type différent
- Une relation initiale ne génère pas de relation finale attendue (suppression)

Pour des applications différentes que la migration de données (par exemple, la généralisation automatique), une règle peut définir plusieurs transformations possibles, et des transformations interdites.

Pour des cas différents de celui des données ponctuelles sur réseau, les règles de modification pour les relations topologiques peuvent être définies avec l'aide d'un graphe conceptuel de voisinage (Egenhofer et Mark 1995). La Figure 71 montre un extrait du graphe de voisinage défini par (Egenhofer et Mark 1995) pour une ligne et un polygone, qui peut servir à modifier la relation topologique entre un arc et un objet surfacique. Une règle basée sur ce graphe pourrait être : « si la relation initiale est (r3) et si la distance entre l'objet surfacique et l'extrémité de l'arc dans la base topographique finale est inférieure à 100m, alors la relation finale attendue sera (r5) ».

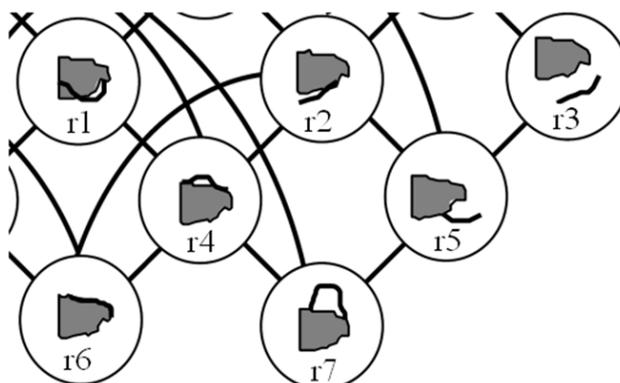


Figure 71. Une partie du graphe conceptuel de voisinage des relations ligne-polygone (Egenhofer et Mark 1995)

B.4.4 Application de l'étape sur l'exemple jouet

Pour notre exemple-jouet (Figure 72), l'accident en bas à gauche a une relation initiale de proximité avec le rond-point – il est à d mètres du rond point. Le rond-point est représenté par un carrefour dans la base finale, alors la relation doit être modifiée. L'appariement entre les objets supports initiaux et finaux fournit le lien d'appariement entre le rond-point et le carrefour. A partir de cette relation initiale, l'étape d'identification des relations finales attendues peut générer une des trois relations finales suivantes :

- relation finale attendue *porté par* avec le carrefour => l'accident sera sur la position (1) à l'état final
- relation finale attendue *de proximité* avec le carrefour avec une distance égal à (d) => l'accident sera sur la position (2) à l'état final
- relation finale attendue *de proximité* avec le carrefour avec une distance égal à ($d + \text{rayon du rond point}$) => l'accident sera sur la position (3) à l'état final

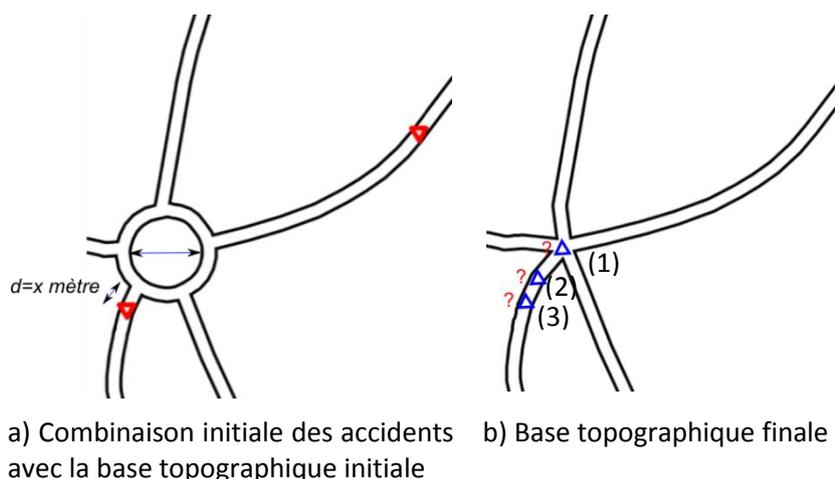


Figure 72. Le besoin pour la modification des relations

Pour choisir entre ces trois possibilités, il faut considérer les spécifications et les besoins modélisés de l'utilisateur. En plus, il faut considérer les seuils de perceptibilité des relations.

Ici, en nous plaçant dans le cas où la base de données finale est destinée à être visualisée pour des analyses d'accidentologie, nous avons défini dans la règle de transformation un seuil arbitraire de 50 mètres terrain (correspondant à 2 mm sur la carte) en-dessous duquel la première option ci-dessus est retenue. L'accident en bas à gauche, pour lequel la distance d est inférieure à ce seuil, sera donc porté par le carrefour (position 1) dans la base finale.

Dans le paragraphe suivant, nous allons détailler la relocalisation qui est la dernière étape (e). La mise en œuvre de toutes les étapes sur des données réelles est détaillée dans le chapitre C.

B.5 Relocalisation des données thématiques basée sur les relations en utilisant un système multicritère

Après l'identification des relations finales attendues entre les données thématiques et les données supports, il faut relocaliser les données thématiques pour faire en sorte que les relations attendues soient satisfaites au maximum. Il s'agit d'un problème d'optimisation sous contraintes.

Pour chaque objet thématique ponctuel à relocaliser, une recherche locale autour de sa position initiale est faite en discrétisant l'espace possible.

La relocalisation comporte trois étapes : (1) détecter des positions possibles (2) Choisir la meilleure position du point de vue du respect des relations finales attendues (3) Raffiner le positionnement au voisinage de la localisation trouvée.

Ces différentes étapes sont présentées en B.5.1. La notion de satisfaction d'une relation est expliquée plus en détails en B.5.2, et la méthode multicritère utilisée est présentée en B.5.3.

B.5.1 Phases de la relocalisation

Phase 1 : détecter des positions possibles

Le but de cette phase est de détecter un ensemble de positions possibles pour l'objet thématique ponctuel dans la base finale. Si l'objet thématique est porté par un objet linéaire final de référence, l'objet linéaire est discrétisé afin d'obtenir un ensemble de points. L'ensemble des points venant de la discrétisation est élagué en utilisant un buffer autour la position initiale de l'objet thématique : cet élagage a pour but d'éliminer les positions trop éloignées donc impossibles. La Figure 73 illustre la détection des positions possibles pour deux accidents.

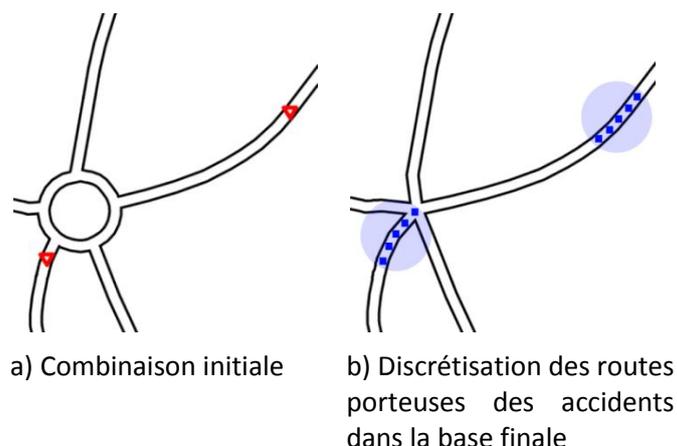


Figure 73. Discretisation de la route porteuse de l'accident pour extraire les positions possibles

Si l'objet thématique n'est porté par aucun objet de référence, l'espace délimité par le buffer autour la localisation initiale de l'objet thématique sera discretisé pour obtenir des positions possibles.

Phase 2 : Choisir la meilleure position du point de vue du respect des relations finales attendues

Le but de cette étape est de trouver la position qui respecte au maximum les relations finales attendues. Nous considérons la recherche d'une localisation optimale du point de vue des relations attendues comme un problème de décision multicritère, où les satisfactions des relations sont les critères. La satisfaction d'une relation indique à quel point cette relation est présente. Cette notion est détaillée en B.5.2. Nous utilisons la décision multicritère parce qu'il n'est pas toujours possible de satisfaire toutes les relations à la fois. Nous constatons que le fait de s'autoriser à ignorer certaines relations afin de satisfaire les autres au mieux donne un meilleur résultat qu'essayer de trouver absolument un compromis entre toutes relations. Pour aider à justifier notre choix de l'utilisation d'une méthode multicritère, nous prenons l'exemple de la Figure 74. Dans cet exemple, l'accident est situé proche de l'intersection route-rivière et en face du bâtiment (Figure 74a). Si nous cherchons à respecter les deux relations au mieux, nous allons avoir un résultat qui respecte moyennement les deux relations (Figure 74b). Par contre, si nous utilisons une méthode multicritère, nous pouvons obtenir des résultats qui respectent certaines relations et en ignorent d'autres. La Figure 74c montre le résultat d'une méthode multicritère avec pondération plus forte de la relation de position relative par rapport au bâtiment. Cela nous a permis d'ignorer la relation de proximité avec l'intersection route-rivière, relation que nous aurions prise en compte s'il n'y avait pas eu la relation accident-bâtiment. Nous pouvons remarquer que le résultat en Figure 74c est meilleur du point de vue du positionnement relatif des objets que le résultat en Figure 74b.

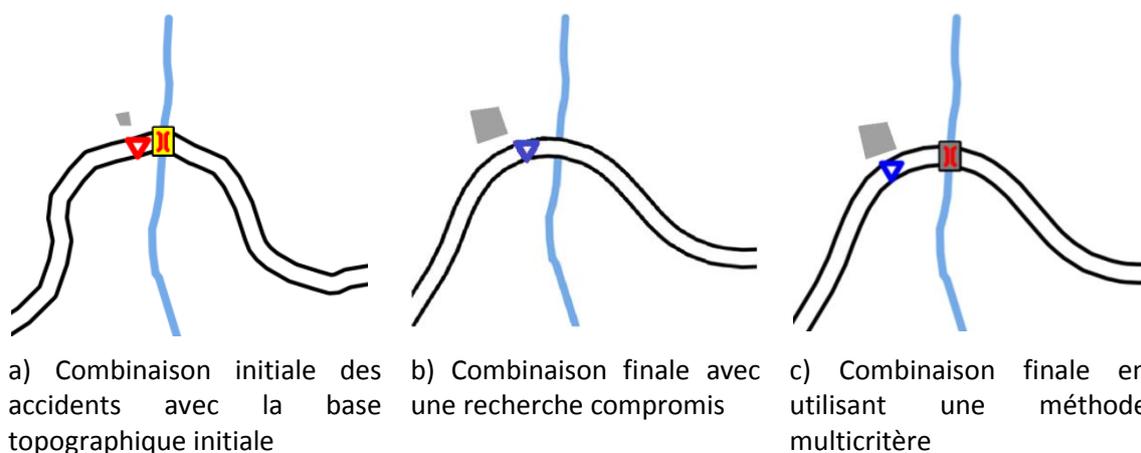


Figure 74. La différence en utilisant une méthode multicritère pour la relocalisation avec pondération plus forte de la relation de position relative par rapport au bâtiment

Dans notre processus, le niveau de satisfaction de chaque relation est d'abord calculé pour chaque localisation possible résultant de la discrétisation issue de la première étape. Puis un système multicritère est appliqué pour obtenir une évaluation des localisations possibles.

Après une étude sur les méthodes multicritère existantes, nous avons décidé d'utiliser la méthode PROMETHEE II (Preference Ranking Organization Method for Enrichment Evaluations) proposée par (Brans et Mareschal 2005), parce qu'elle permet de favoriser des solutions selon des préférences externes. Les préférences qui concernent les relations sont les importances des relations finales attendues définies à l'étape (d).

PROMETHEE II est classée parmi les approches par agrégation partielle, ce qui signifie qu'elle compare chaque paire de solutions possibles selon l'ensemble des critères. Après agrégation, le résultat de cette méthode est un classement des positions possibles, ordonnées selon la satisfaction des relations attendues qu'elles permettent d'obtenir.

Phase 3 : Raffiner le positionnement au voisinage de la localisation trouvée lors de la 2ème étape

A cette étape, un voisinage de la position trouvée à l'étape 2 est discrétisé avec une équidistance plus petite, afin de trouver une solution meilleure que celle qui a été trouvée à la 2ème étape en appliquant à nouveau le système multicritère.

B.5.2 Besoin d'une évaluation de la satisfaction des relations

Pour pouvoir utiliser la méthode multicritère pendant les phases 2 et 3 de la relocalisation, il faut pouvoir évaluer la satisfaction de chaque relation. L'évaluation doit montrer le degré de respect des relations finales attendues. Nous proposons d'introduire *une mesure de satisfaction* d'une relation. Cette mesure quantifie à quel point une relation est vérifiée par une paire d'objets. La méthode de mesure dépend de la nature de la relation et des propriétés qui la décrivent.

Par exemple, pour la relation de proximité entre un objet thématique ponctuel et un bâtiment, nous voulons mesurer la satisfaction s_r d'une position possible p_{candidat} pour l'objet thématique final. Nous proposons une mesure de satisfaction qui considère que la relation de proximité n'est pas satisfaite du tout $s_r(p_{\text{candidat}}) = 0$, si la différence entre la distance attendue d_{attendue} et la distance du candidat d_{candidat} est supérieure ou égale à un seuil seuilMax_r . Par contre, la relation de proximité est considérée comme complètement satisfaite $s_r(p_{\text{candidat}}) = 1$, si la différence entre la distance attendue d_{attendue} et la distance du candidat d_{candidat} est inférieure ou égale à un seuil seuilMin_r . Si la différence est entre seuilMin_r et seuilMax_r mètres, la satisfaction varie selon la différence (Figure 75).

Formellement, la mesure de la satisfaction S de la relation de proximité <objet thématique ponctuel, bâtiment> peut être définie de la façon suivante :

$$\begin{aligned} &\text{Si } |d_{\text{candidat}} - d_{\text{attendue}}| < \text{seuilMin}_r \text{ alors} \\ &\qquad\qquad\qquad s_r(\text{candidat}) = 1 \\ &\text{Si } |d_{\text{candidat}} - d_{\text{attendue}}| > \text{seuilMax}_r \text{ alors} \\ &\qquad\qquad\qquad s_r(\text{candidat}) = 0 \\ &\text{Sinon} \\ &\qquad\qquad\qquad s_r(\text{candidat}) = \left(\frac{\text{seuilMax}_r - |d_{\text{candidat}} - d_{\text{attendue}}|}{\text{seuilMax}_r - \text{seuilMin}_r} \right) \end{aligned}$$

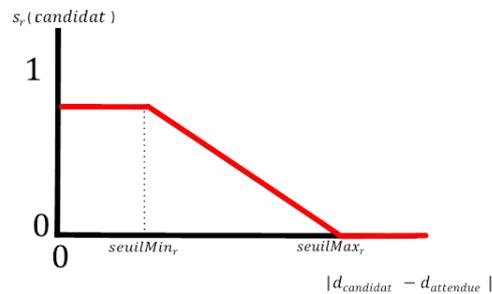


Figure 75. Mesure de satisfaction pour la relation <objet thématique ponctuel, Bâtiment>

Pour l'échelle de 1 :25000, nous avons fixé les deux seuils ($\text{seuilMin}_r = 30$ et $\text{seuilMax}_r = 100$) empiriquement, mais ces seuils doivent être basés sur l'échelle et les préférences de l'utilisateur (d'une façon similaire de la construction des règles de transformation des relations).

Pour les relations binaires, un graphe de voisinage peut aider à détecter le niveau de satisfaction. Si nous reprenons le cas des relations topologiques ligne-région (Figure 71), on peut par exemple définir la mesure de satisfaction suivante :

Si le prédicat du candidat est le prédicat attendu	$S=1$
Sinon s'il existe un arc entre le prédicat attendu et le prédicat du candidat	$S=0.5$
Sinon	$S=0$

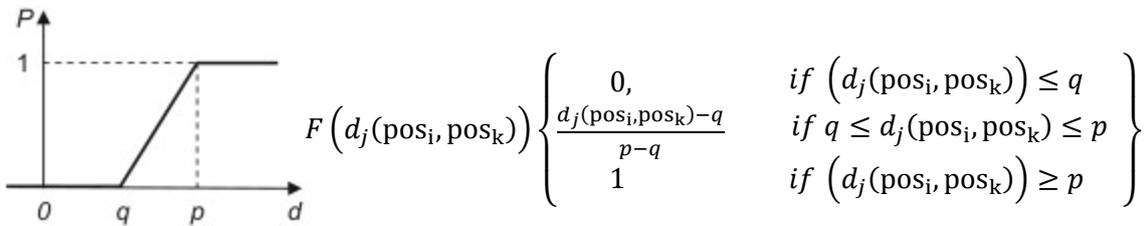
B.5.3 L'évaluation du respect des relations d'une position possible en utilisant une méthode multicritère

Détails sur la méthode multicritère utilisée

PROMETHEE II est une méthode multicritère de décision. Elle est classifiée comme une approche d'ordre partiel. Cela signifie qu'elle est basée sur la comparaison deux à deux des décisions possibles.

Notons $A = \{pos_1, pos_2, \dots, pos_n\}$ l'ensemble de positions possibles et $G = \{g_1(pos), g_2(pos), \dots, g_k(pos)\} | pos \in A$, les mesures de satisfaction de relations (i.e. les critères). Chaque critère a un poids qui représente son importance. Le poids du $j^{\text{ème}}$ critère est noté w_j . Dans notre cas, le critère est l'importance de la relation numéro j . La méthode calcule la préférence d'une alternative pos_i par rapport à une autre alternative pos_k concernant un critère en utilisant la fonction suivante :

$$P_j(pos_i, pos_k) = F(g_j(pos_i) - g_j(pos_k)) = F(d_j(pos_i, pos_k)) \in [0,1].$$



Autrement dit, la préférence de la position pos_i sur la position pos_k pour la relation j varie de 0 à 1. Elle vaut 0 lorsque la satisfaction de la relation est proche pour les positions pos_i et pos_k (différence inférieure au seuil q défini). Elle vaut 1 lorsque la satisfaction de la relation j pour la position pos_i est suffisamment supérieure à celle de la position pos_k (différence supérieure au seuil p défini) et elle varie linéairement entre les deux seuils .

Les valeurs de p et q doivent être identifiées pour chaque critère.

La préférence de pos_i par rapport à pos_k en se basant sur tous les critères est calculée comme suit (n est le nombre de critères) :

$$\pi(pos_i, pos_k) = \sum_{j=1}^n P_j(pos_i, pos_k) w_j$$

La préférence est calculée entre chaque paire de propositions. En fin de ce calcul, un graphe de surclassement complet (en anglais : complete outranking graph) est calculé.

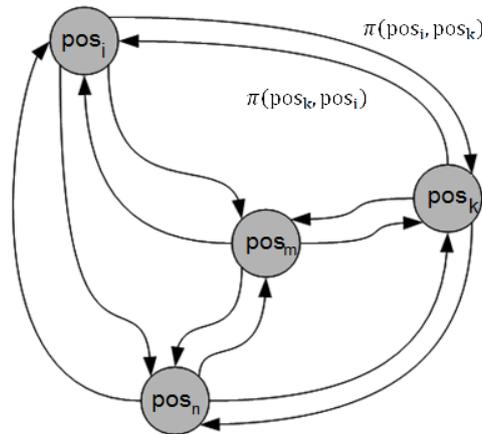


Figure 76. Graphe de surclassement complet en utilisant la méthode Prométhée II

Pour chaque position, on évalue alors à quel point elle est meilleure, et à quel point elle est pire que les autres.

Le flux de surclassement positif $\phi^+(pos)$ indique à quel point une alternative pos est mieux que les autres.

$$\phi^+(pos) = \frac{1}{n-1} \sum_{x \in A} \pi(pos, x)$$

Le flux de surclassement négatif $\phi^-(pos)$ indique à quel point une alternative a est moins bien que les autres.

$$\phi^-(pos) = \frac{1}{n-1} \sum_{x \in A} \pi(x, pos)$$

Dans Prométhée II, les alternatives sont ordonnées selon le flux net de surclassement $\phi(pos)$

$$\phi(pos) = \phi^+(pos) - \phi^-(pos)$$

La meilleure solution est l'alternative qui a le meilleur flux net de surclassement.

Comment la méthode est utilisée dans notre cas

Dans notre cas, les alternatives A sont les localisations possibles de l'objet thématique ponctuel.

Les critères pour nous sont les relations attendues, et les évaluations G pour nous sont les mesures de satisfaction des relations finales attendues. Le poids pour nous est l'importance de relation. Les valeurs p et q de la fonction de préférence doivent être identifiées pour chaque relation d'une façon empirique.

B.6 Récapitulation : illustration de notre méthode de migration sur les exemples-jouets

Notre méthode a été implémentée dans CartAGen (Renard et al. 2010), module de généralisation de la plate-forme de recherche GeOxygene (Grosso et al. 2012). Cette partie illustre la méthode sur deux exemples-types (Figure 77), et montre les résultats obtenus. La migration de données thématiques est appliquée comme décrit ci-dessous.

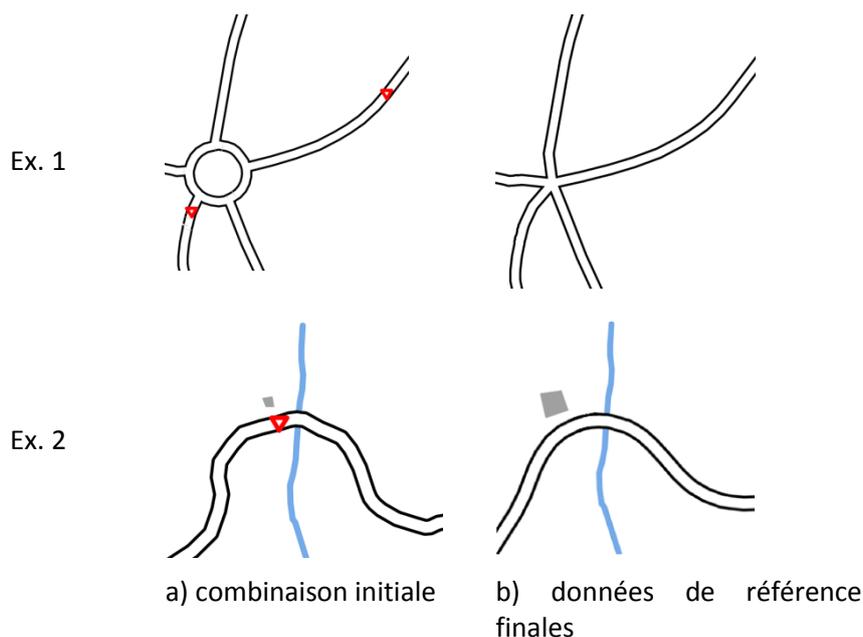


Figure 77. Données de référence initiales et finales avec la localisation initiale des accidents

a) Extraction des objets caractéristiques. Comme il est montré en Figure 78, les ronds-points sont identifiés comme des objets caractéristiques. Ils sont extraits du réseau routier par recherche des faces rondes dans le graphe dérivé du réseau routier en utilisant l'indice de Miller, selon la méthode proposée par (Sheeren 2005) (qui sera détaillée en chapitre C). Dans le deuxième exemple, l'objet caractéristique « intersection route-rivière » est extrait en utilisant une simple intersection entre objets linéaires.

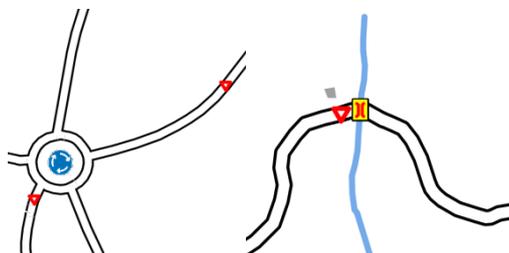


Figure 78. Extraction des objets caractéristiques initiaux

b) Extraction des relations. D'abord, nous devons identifier la route qui porte l'accident, ensuite, explorer le voisinage de chaque accident afin de détecter tous les objets supports

possibles (i.e. tous les objets topographiques ou caractéristiques pouvant avoir avec l'accident une relation traduisant son référencement par rapport aux données initiales). Dans l'exemple 1, le rond point est trouvé comme un objet caractéristique support parce qu'il est proche de l'accident et adjacent à la route qui porte l'accident. Dans l'exemple 2, l'intersection route-rivière et le bâtiment sont trouvés comme des objets supports possibles. L'étape suivante est d'extraire toutes les relations entre les accidents et les objets supports. Toutes les conditions de pertinence des relations sont testées, les relations extraites pour chaque accident sont les suivantes :

- Relations *porté par* entre l'accident et une route, i.e. extraction de la route hôte (porteuse de l'accident) pour tous les accidents
- Relation de proximité entre l'accident et le bâtiment
- Relations de proximité point-point entre les accidents et les objets caractéristiques ponctuels portés par la route
- Relations de proximité point-rond-point entre les accidents et les ronds-points qui sont adjacents à la route hôte

Dans l'exemple 1, les relations suivantes sont extraites :

Deux relations *porté par*, une pour chaque accident avec les routes concernées
Deux relations de proximité point-rond-point, une pour chaque accident

Dans l'exemple 2, les relations suivantes sont extraites :

Une relation *porté par* pour l'accident avec la route concernée
Une relation de proximité point-point avec l'intersection route-rivière
Une relation de direction entre l'accident et le bâtiment
Une relation de proximité avec le même bâtiment

c) Appariement. Les objets caractéristiques finaux sont extraits, le carrefour pour le premier exemple, et l'intersection route-rivière pour le deuxième exemple (Figure 79).

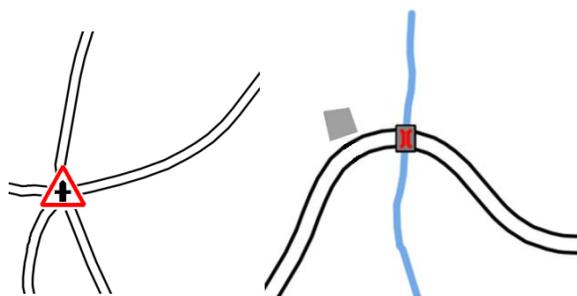


Figure 79. Extraction des objets caractéristiques finaux

Un appariement des données topographiques initiales et finales est d'abord réalisé. Dans notre implémentation, nous utilisons la méthode d'appariement automatique développée par Mustière et Devogele (2008) pour appairer le réseau routier, et un appariement simple basé sur un identifiant unique pour les bâtiments. Ensuite, les objets caractéristiques initiaux et finaux sont appariés. Dans nos exemples, un lien de correspondance est établi entre le rond point et le carrefour dans le premier exemple, et un lien entre les deux intersections route-rivière dans le deuxième exemple.

d) Identification des relations finales attendues. Dans le premier exemple, le rond-point devient un carrefour. Selon les règles qui sont définies a priori sur les relations de proximité point-rond-point, si le rond-point devient un carrefour, la relation devient une relation de proximité point-point. Par ailleurs, parce que nous nous plaçons dans l'hypothèse où la carte résultante est dédiée à l'étude des accidents routiers, si un accident est trop proche d'un carrefour (ici 50 m), le carrefour doit être considéré comme porteur de l'accident. C'est le cas de l'accident a1 ici, qui est à une distance inférieure à 50m du rond-point. Donc la relation finale attendue avec le carrefour est "porté par" qui se traduit par une relation topologique d'égalité. L'accident a2 est à une distance supérieure à 50m du rond point, la relation attendue dans l'état final sera donc une relation de proximité avec le carrefour, caractérisée par une distance égale à la distance initiale jusqu'au rond point plus le rayon du rond point. Dans le deuxième exemple, les relations finales attendues seront identiques aux relations initiales.

Il est à noter que les modifications de relations effectuées sont dictées par l'usage de la carte finale (ici par hypothèse une étude d'accidentologie). Pour un autre cas d'utilisation tel qu'une carte de navigation en temps réel, il aurait été préférable dans le premier exemple d'augmenter la distance entre l'accident a1 et le carrefour plutôt que de placer l'accident sur le carrefour, pour bien montrer que l'accident ne bloquait pas la totalité du carrefour.

e) Relocalisation de données thématiques. Les routes finales qui portent des accidents sont discrétisées avec un pas de 5 mètres sur leur partie située à moins de 300 mètres de la localisation initiale de l'accident. Les relations attendues sont évaluées pour chaque localisation possible ainsi définie. Les valeurs de l'évaluation sont assignées dans le système multicritère, qui identifie la meilleure solution. Une deuxième discrétisation avec un pas de 1 mètre est faite pour obtenir une localisation plus précise. La Figure 80 présente le résultat obtenu, qui correspond au résultat attendu.

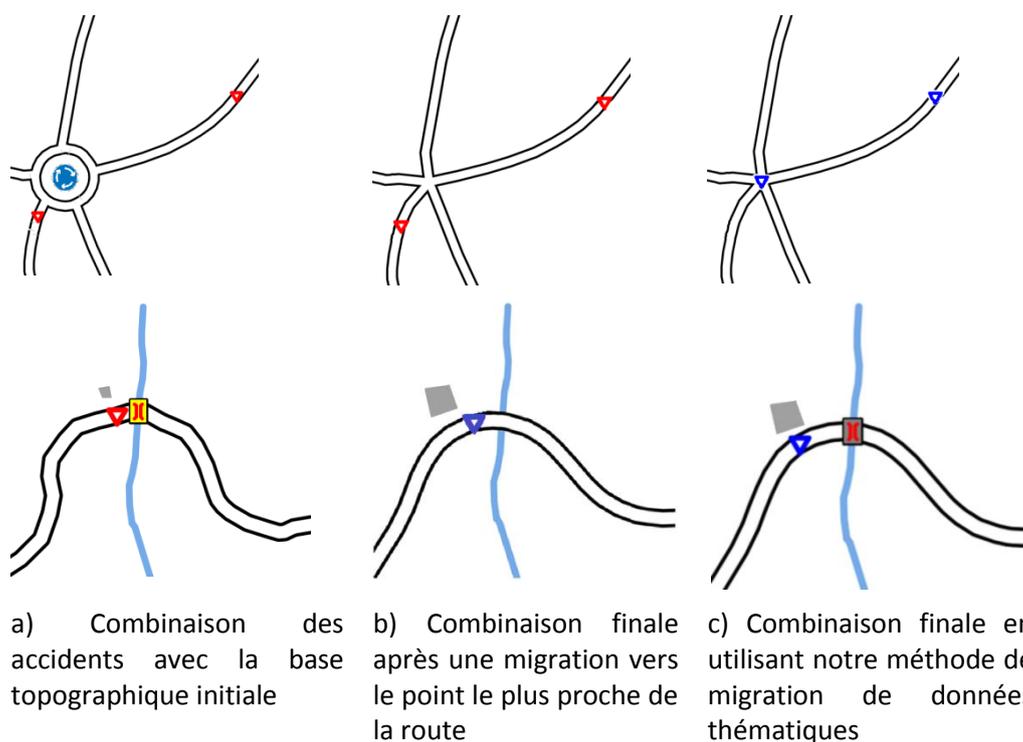


Figure 80. Résultat de migration de données thématiques

B.7 Adaptation du modèle pour d'autres cas que la migration de données thématiques ponctuelles sur réseau

Dans cette partie, nous discutons des extensions possibles du modèle pour répondre à d'autres besoins que la migration de données ponctuelles sur réseau. Premièrement, nous allons discuter la migration de données pour toutes les dimensions des données thématiques (linéaire, surfacique) et pour toutes les dimensions des données topographiques. Ensuite, nous allons discuter des approches pour la prise en compte des relations pendant deux processus autres que la migration de données thématiques: la généralisation automatique et la génération de carte à la demande.

B.7.1 Adaptation du modèle pour d'autres cas de migration de données thématiques

Pour la migration de données linéaires et surfaciques, les étapes pour la migration de données sont presque les mêmes que les étapes de la migration de données ponctuelles sur réseau, sauf la dernière étape (étape e). L'étape de relocalisation est remplacée par une étape de génération de données thématiques finales en respectant les relations finales attendues extraites lors de l'étape (d).

Les étapes sont illustrées en Figure 81. Le principe des quatre premières étapes est inchangé, bien que le détail de ces étapes soit différent. Nous allons présenter les détails de chaque étape dans la suite.

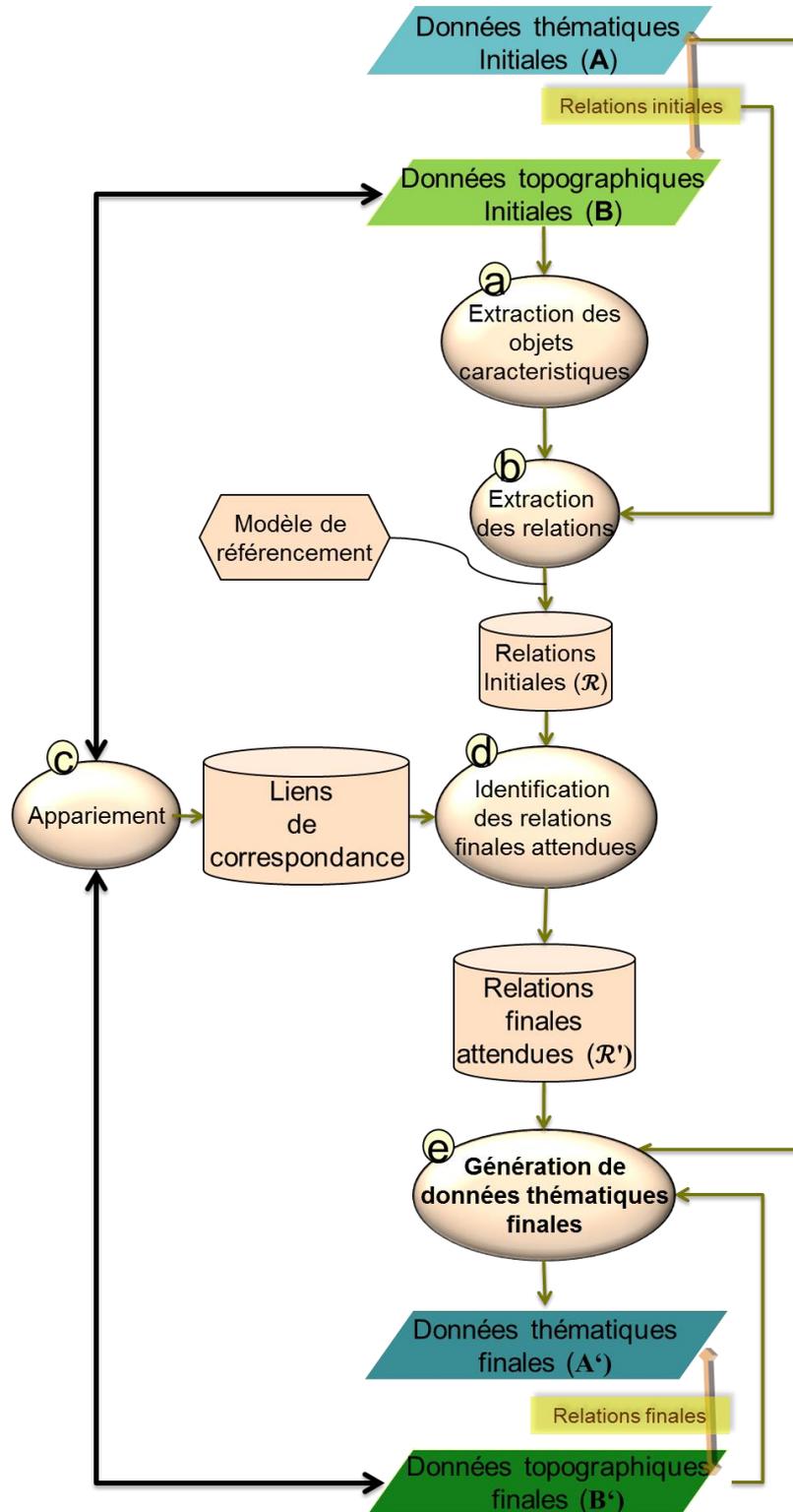


Figure 81. Processus de migration de données thématiques (version générique)

Migration de données thématiques linéaires sur réseau après un remplacement de données topographiques

Dans la suite, nous allons expliquer les modifications à apporter au modèle pour pouvoir traiter le cas de la migration de données linéaires après un remplacement de données topographiques.

Extraction et représentation des relations des données thématiques linéaires

Dans le cas de données thématiques linéaires, l'extraction des objets caractéristiques suit la même logique que celle utilisée pour la migration de données ponctuelles. Par contre, l'extraction de relations est plus compliquée parce que les relations peuvent être partielles. Une relation partielle est une relation entre une partie de l'objet thématique et un objet support. Dans (Gaffuri 2008), ce type de relation a été utilisé pour rendre les courbes de niveau cohérentes avec le réseau hydrologique (Figure 40, p.45). Les objets thématiques linéaires doivent être divisés pour pouvoir assigner les relations. Chaque objet thématique doit être modélisé par un ensemble de nœuds et arcs. Chaque nœud et chaque arc peuvent avoir des relations. La division est faite en calculant les relations avec les données du support. Une modélisation similaire a été utilisée dans (Touya et al. 2012), illustrée en Figure 82, où une relation d'"alignement" est modélisée entre une portion de l'itinéraire de vélo et la route. La relation d'alignement est modélisée par deux points appelés *points de convergence*. Pour nous, un point de convergence est un nœud.

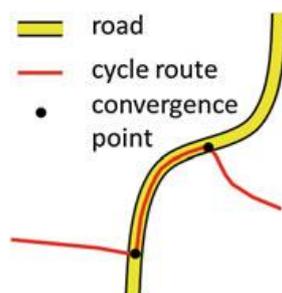


Figure 82. Une partie d'un itinéraire de vélo qui suit la route (Touya et al. 2012)

Pour générer les arcs et les nœuds du graphe, nous proposons d'abord de discrétiser les objets thématiques linéaires en points consécutifs thématiques, puis de calculer les relations de chaque point thématique. Ensuite, il faut regrouper les points consécutifs qui ont la même relation dans un arc et identifier les deux extrémités de l'arc comme des nœuds.

Pour représenter les relations, le modèle de référencement que nous avons utilisé pour les données ponctuelles sur réseau est toujours valable. Il faut juste ajouter les relations qui concernent les données thématiques de type linéaire. Un ajout sera le lien entre l'arc thématique et ses deux nœuds thématiques, ça fait un lien thématique-thématique. Le modèle de référencement avec les changements est illustré en Figure 83. Par rapport à l'exemple de la Figure 82, les points de convergence sont des nœuds thématiques et la portion entre les deux points de convergence est un arc thématique.

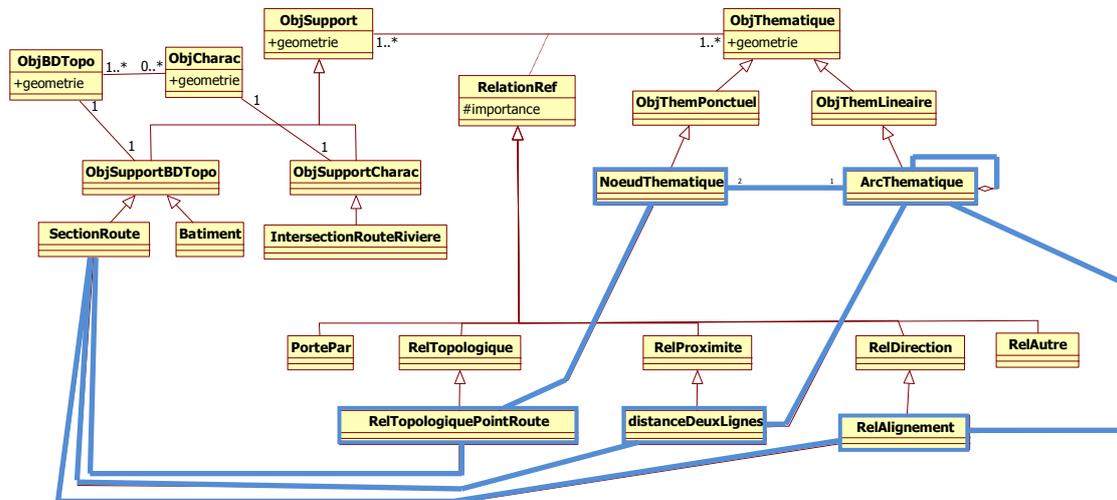


Figure 83. Modèle de référencement en cas de la migration de données thématiques linéaires

Génération de données linéaires en se basant sur les relations attendues

Après l'extraction et la modélisation des relations initiales, l'appariement des objets supports et l'extraction des relations attendues sont faits suivant la même logique que celle utilisée pour la migration de données ponctuelles sur réseau. L'étape finale sera la génération des données thématiques finales en se basant sur les relations finales attendues. Le processus comporte donc les deux points suivants :

- relocalisation des nœuds thématiques d'après la méthode de migration de données ponctuelles.
- génération des arcs entre les nœuds en respectant les relations des arcs.

Le cas simple pour la deuxième étape est d'avoir une seule relation à considérer. Par exemple, un itinéraire de vélo qui est en relation d'alignement avec une route.

La même approche peut être utilisée si les données topographiques sont des limites parcellaires, parce que les limites parcellaires forment un graphe de la même manière que le réseau routier.

La migration de données thématiques surfaciques après un remplacement de données topographiques

Si les objets thématiques de type surfacique sont assez grands, il faut modéliser l'objet thématique en bordure et intérieur. La bordure a ses propres relations et l'intérieur aussi a ses relations à calculer. Les relations de bordures sont à extraire de la même façon que celle utilisée pour l'extraction des relations des données thématiques de type linéaire (paragraphe précédent). Parmi les relations de l'intérieur, nous avons l'intersection avec d'autres objets.

L'étape (e) de la migration de données surfaciques sera d'effectuer la migration de bordures en se basant sur la méthode de la migration de données linéaires et en respectant les relations de l'intérieur.

B.7.2 Adapter la généralisation automatique des données topographiques pour respecter les relations finales attendues

Dans cette partie, nous étudions la prise en compte des relations thématique-topographique lors de la généralisation de données topographiques (Figure 84). Il s'agit d'effectuer la généralisation des données topographiques et la migration de données thématiques en même temps.

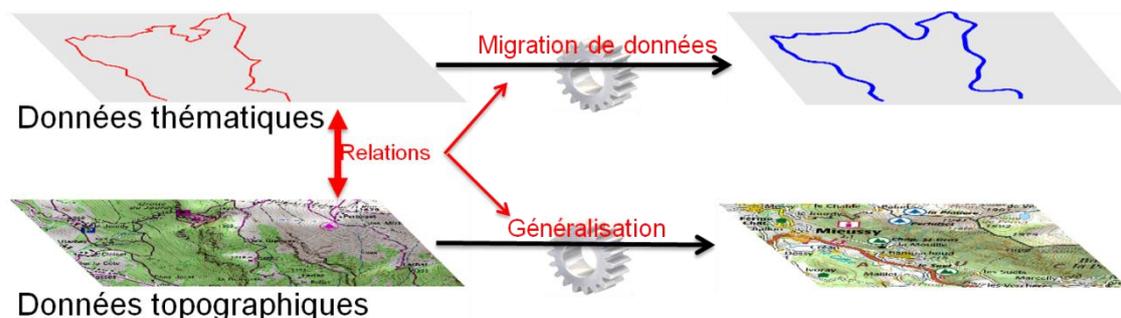


Figure 84. Prise en compte des relations initiales lors de la migration de données thématiques et la généralisation des données topographiques afin de changer le niveau de détail

Description formelle de la problématique lié à la prise en compte des relations pendant la généralisation de données topographiques

En prenant les mêmes notations qu'en B.2.1 :

- A est l'ensemble des objets thématiques dans les données initiales,
- B est l'ensemble des objets topographiques dans les données initiales,
- \mathcal{R} est l'ensemble des relations existant entre les objets et qui traduisent le référencement des objets thématiques par rapport aux objets topographiques, dans les données initiales,
- A' est l'ensemble des objets thématiques dans les données finales,
- B' est l'ensemble des objets topographiques dans les données finales,
- \mathcal{R}' est l'ensemble des relations existant entre les objets et qui traduisent le référencement des objets thématiques par rapport aux objets topographiques, dans les données finales.

Connaissant les données initiales thématiques (A) et les données topographiques (B), il s'agit d'abord d'extraire et de caractériser (par des attributs) l'ensemble des relations \mathcal{R} . Cela est fait par le modèle de référencement. Puis il s'agit d'obtenir des données topographiques finales (B') généralisées en respectant les relations \mathcal{R} initiales.

La problématique est la suivante :

Pendant la généralisation automatique, les données topographiques finales (B') seront construites par la généralisation. Nous cherchons à déduire (1) les données topographiques généralisées en respectant les relations initiales (2) les relations \mathcal{R}' attendues dans l'état final, et (3) la localisation des objets thématiques dans l'état final, c'est-à-dire l'ensemble A' respectant les relations \mathcal{R}' avec B' :

$$A\mathcal{R}B \Rightarrow ?\mathcal{R}'?B'?A' / A'\mathcal{R}'B'$$

Propositions pour effectuer la généralisation en respectant les relations initiales

Plusieurs stratégies sont possibles:

- Soit généraliser les données topographiques, puis effectuer une migration de données (ce cas est très similaire à la migration de données après un remplacement). Ce qui change dans ce cas est que les liens de correspondance sont dérivés des informations issues de la généralisation et non par appariement. Alors la seule étape qui change est l'étape (c). En plus, nous pouvons accéder au déroulement du processus de généralisation, cela nous permet de proposer une meilleure localisation. Par exemple, si on sait qu'un bâtiment a été agrégé pendant la généralisation, cette information est plus précise que le résultat d'appariement.
- Soit effectuer les deux processus ensemble, ce choix nous oblige à créer des modules qui peuvent aider à intégrer la prise en compte des relations dans le processus de généralisation. Par exemple, adapter les relations extraites pour pouvoir les convertir en contraintes relationnelles.

Pour le deuxième cas, ce qui change par rapport à la migration de données, est qu'on peut contraindre la généralisation pour que la généralisation des données topographiques prenne en compte leurs relations avec les données thématiques. En prenant l'exemple de la Figure 12, nous cherchons à effectuer une généralisation automatique qui ne supprime pas les routes alignées avec des lignes de bus.

Pour pouvoir respecter les relations avec des données thématiques pendant la généralisation, nous proposons deux méthodes. La première méthode est de définir des contraintes liées aux relations. Dans le paragraphe suivant nous discutons l'ajout des contraintes liées aux relations avec des données thématiques en utilisant le modèle CartACom de (Duchêne 2004). La deuxième méthode est spécifique à la généralisation en utilisant un système multi-agents (Ruas 1999; Duchêne 2004; Gaffuri 2008; Touya 2011). La méthode consiste à créer des agents pour les objets thématiques, l'agent doit pouvoir mesurer la satisfaction de ses relations, proposer des changements pour respecter ses relations.

Les contraintes relationnelles extraites des relations thématique-topographique

Pour pouvoir prendre en compte les relations thématique-topographique, nous voyons qu'il faut modifier le processus de généralisation et ajouter de nouvelles contraintes issues des relations thématique-topographique attendues. Nous supposons que le modèle de généralisation automatique utilisée est un modèle multi-agent. Nous allons parler uniquement de l'adaptation du modèle CartACom de (Duchêne 2004).

Les contraintes des relations thématique-topographique doivent être modélisées suivant la même modélisation utilisée au sein de la généralisation. En Figure 85 (reprise de la Figure 37), la modélisation de représentation de contraintes de (Duchêne 2004) est illustrée. Cette modélisation concerne les relations entre objets de manière générale, elle a été testée mise en œuvre pour les relations topographique-topographique.

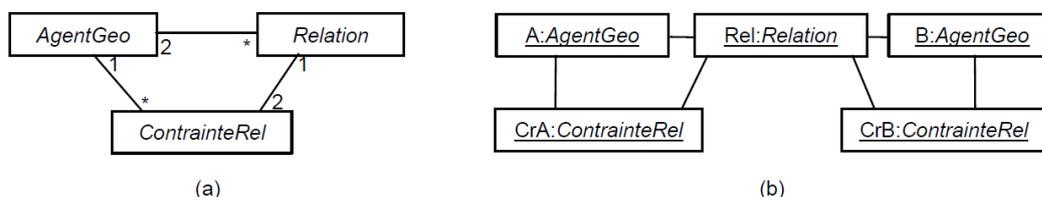


Figure 85. Modélisation des contraintes relationnelles (Duchêne 2004)
 (a) diagramme de classes (b) diagramme d'objets

Nous cherchons à adapter cette modélisation à nos besoins. Pour modéliser les contraintes relationnelles dans notre cas, nous proposons la modélisation suivante (Figure 86).

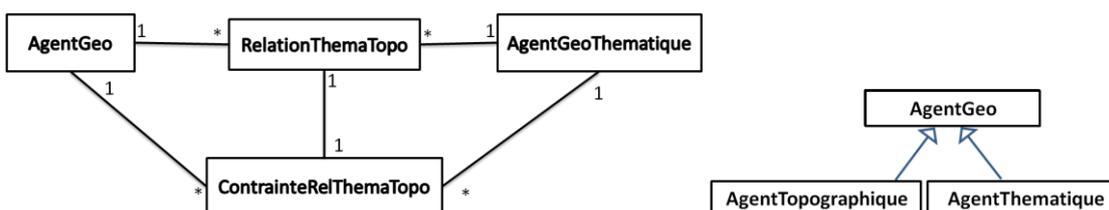


Figure 86. Modélisation proposée des contraintes relationnelles thématique-topographique au sein du modèle CartACom de (Duchêne 2004) - diagramme de classes

D'abord, il faut créer les agents qui représentent les données thématiques, nous les appelons *AgentGeoThématique*. Ensuite, il faut définir la méthode de calcul de la satisfaction de la contrainte en se basant sur les mesures de satisfaction des relations présentées en B.5.2. L'importance de la relation thématique-topographique peut être utilisée pour évaluer les conflits. En plus, une liste d'actions possibles doit être définie pour chaque conflit.

Modélisation des liens de correspondance et des traitements qui ont été effectués pendant la généralisation

Lors de la généralisation automatique, les liens de correspondances peuvent être plus riches. En plus des objets homologues, les liens et les changements des données topographiques doivent être modélisés. Ensuite, la modélisation de ces changements peut être utilisée lors de l'identification des relations finales attendues. Cette modélisation doit être intégrée au sein du modèle de référencement. Les changements possibles peuvent être organisés dans une liste. Les liens de correspondance et la liste de changements nous aident à mieux anticiper les relations finales attendues. Nous proposons d'associer une énumération pour chaque type d'objet (point, ligne, surface). Par exemple, en Figure 87, La classe *LienCorrespondance* lie deux objets supports et contient une liste de toutes les différences entre les deux objets supports concernés. Les différences sont organisées sous forme d'une liste de valeurs prises dans l'énumération.

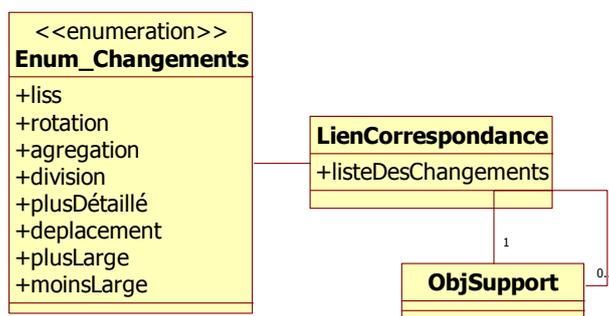


Figure 87. Modélisation de liens de correspondance entre les objets supports initiaux et les objets supports finaux

En prenant le premier exemple de nos exemples-jouets (Figure 48, p.55), l'objet support qui est la route peut avoir un lien de correspondance avec la route finale lissée. La liste de changement de ce lien de correspondance doit contenir un seul membre : *liss*. Ces changements peuvent être extraits par l'historique des opérations qui ont été effectuées pendant la généralisation automatique des données topographiques initiales.

Résumé sur la prise en compte des relations thématique-support lors de généralisation automatique des données topographiques

Pour que les relations thématique-support soient prise en compte lors de la généralisation automatique, nous avons proposé des premières idées pour que les relations soient transformées en contraintes relationnelles pour la généralisation. En plus, nous avons proposée une modélisation simple pour stocker l'historique des opérations qui ont été effectuées sur les objets topographiques initiaux lors de la généralisation.

Après l'étude de la prise en compte des relations thématique-support lors de la migration des données thématiques (B.7.1) et la généralisation des données topographiques (B.7.2), nous allons présenter nos idées pour la prise en compte de ces relations dans un processus de carte à la demande.

B.7.3 Prise en compte des relations dans un processus de carte à la demande

Le but de cette partie est de montrer le rôle des relations thématique-topographique pendant un processus automatique de dérivation de carte à la demande à partir de données topographiques et thématiques, et comment notre modèle peut être utilisé pour mieux prendre en compte ces relations. Cette partie est le résultat d'une collaboration avec l'Ordnance Survey en Grande Bretagne (OSGB) qui est l'homologue de l'IGN en Grande Bretagne. Dans la suite, nous présentons ce qu'est un processus de carte à la demande, puis nous étudions un modèle de spécifications proposé pour un tel processus. Le processus de carte à la demande que nous décrivons est proposé par (Balley et Regnaud 2011). Le modèle est ensuite analysé pour extraire les besoins liés à la prise en compte des relations. Pour répondre aux différents besoins, nous présentons, dans la partie suivante, une adaptation de notre modèle pour représenter, stocker et prendre en compte des relations.

Introduction sur la carte à la demande

Avec la vaste utilisation des données géographiques, et la diversité des besoins en terme de niveau de détail, préférences de visualisation, schémas de données, etc., il devient de plus en plus nécessaire de trouver des moyens pour répondre aux différents besoins plus rapidement.

Le but d'un processus de dérivation de carte à la demande est de produire une carte ou des données automatiquement suivant des besoins exprimés par un utilisateur. Le processus peut utiliser plusieurs sources de données, les données peuvent être thématiques fournies par un utilisateur ou topographiques. Il existe différents travaux sur la carte à la demande. Notre but n'est pas d'étudier les modèles existants, mais plutôt d'identifier les parties qui concernent les relations, et de proposer des solutions pour améliorer la prise en compte des relations. Comme exemple de système de carte à la demande, nous allons utiliser celui proposé par (Balley et al. 2012). Les étapes principales d'un processus de carte à la demande sont illustrées en Figure 88.

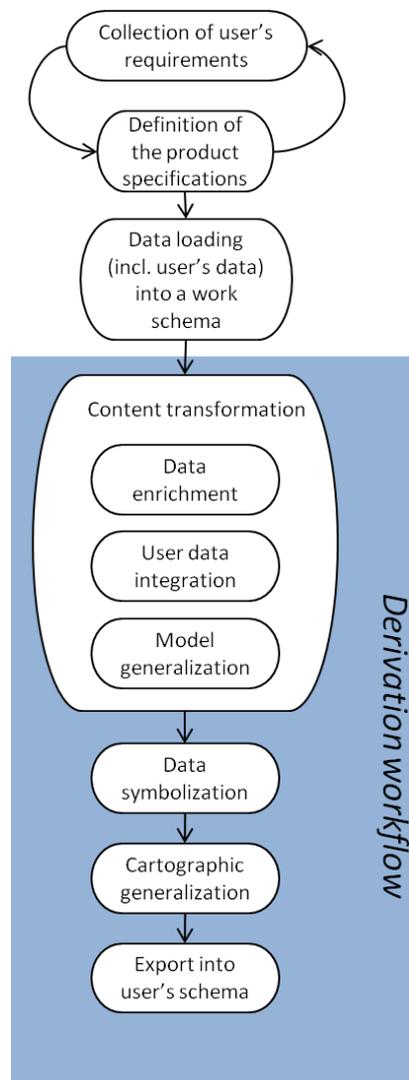


Figure 88. Les étapes principales d'un processus de carte à la demande (Balley et Regnaud 2011)

Modèle de spécifications pour la carte à la demande de (Balley et Regnaud 2011)

Le modèle de spécifications présenté par (Balley et Regnaud 2011) est inspiré du modèle de (Gesbert 2005). Le modèle de spécifications permet de décrire formellement la carte voulue par l'utilisateur. La série de processus à effectuer sur les données est ensuite décidée en se basant sur cette description formelle. L'acquisition des préférences de l'utilisateur et leur transformation en des spécifications est faite à l'aide d'un système de dialogue selon l'approche de (Christophe 2009). Le modèle de spécifications est indépendant de la source de données géographiques. De plus, il ne contient pas l'information sur les processus à effectuer (l'utilisateur décrit ce qu'il veut, pas comment il souhaite l'obtenir). Une version simple du modèle est illustrée en Figure 89. La classe *mapped concept* est la classe principale. Elle représente tous les types d'objets de monde réel qui sont utilisés lors de la fabrication de la carte. Les concepts sont organisés en une ontologie. Le modèle permet de spécifier les thèmes (les types d'objets géographiques à représenter) et de spécifier les objets à utiliser, le niveau de détail, la symbolisation, le schéma de données et le besoin d'intégration. En plus, il contient la description des données de l'utilisateur et comment les référencer. Les besoins sont présents dans les spécifications sous forme de contraintes.

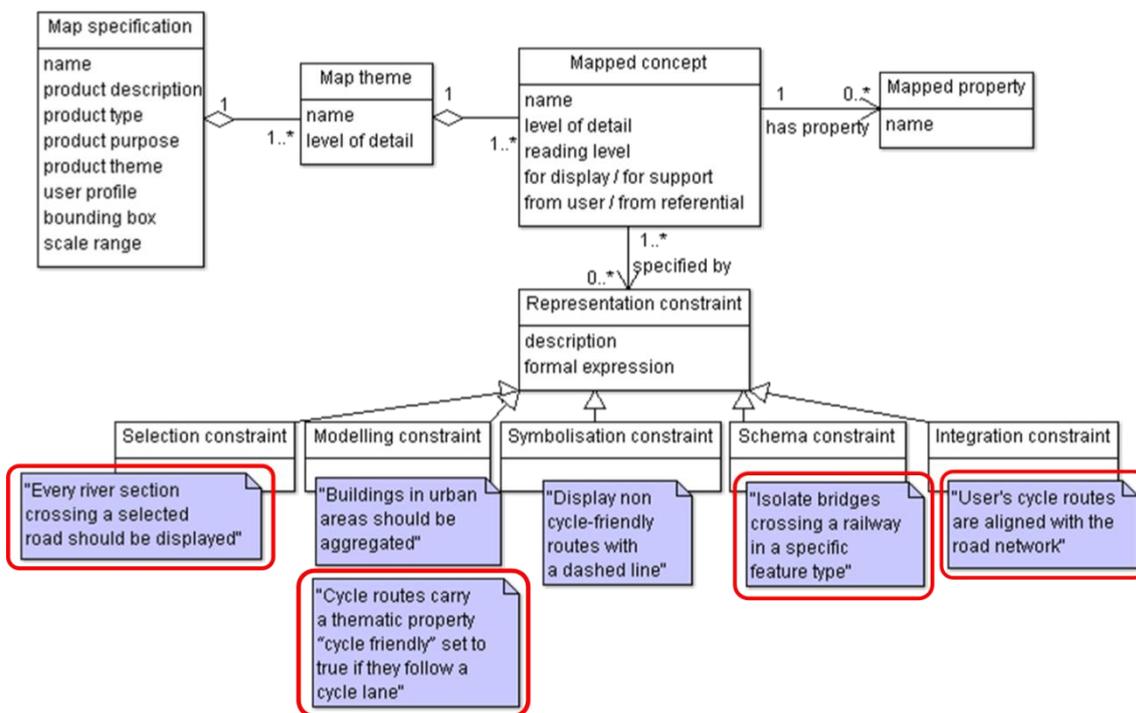


Figure 89. Modèle de spécifications de carte à la demande proposé par (Balley et al. 2012).

Les contraintes liées à des relations sont distinguées par des carrés rouges

Analyse des spécifications pour identifier les relations sur lesquelles elles s'appuient et les processus impactés

Notre but est de trouver les éléments de spécifications qui sont liés aux relations. Nous faisons cela pour proposer un modèle qui facilite la prise en compte de ces relations pendant les différents traitements de la carte à la demande.

En regardant les exemples de contraintes en Figure 89, nous remarquons que certaines contraintes de représentation sont liées aux relations (nous les avons distinguées par un carré rouge). En analysant ces contraintes, nous trouvons les points suivants :

- Les relations peuvent être un critère de sélection, par exemple, la contrainte *chaque section de rivière qui intersecte une route sélectionnée doit être représentée dans la carte*.
- Les relations peuvent être utilisées pour enrichir les données sémantiquement, par exemple *si l'itinéraire de vélo empreinte une piste cyclable, l'attribut "itinéraire sécurisé" est mis à vrai* (traitement concerné : transformation de schéma).
- Certaines contraintes concernent l'ajout de certains objets caractéristiques, par exemple, *les intersections de routes avec les voies ferrées doivent être représentées par des entités*.
- Les spécifications peuvent exprimer une relation à respecter pendant les traitements, par exemple, *les itinéraires de vélo doivent être recalés sur les routes les plus proches* (traitement concerné : sélection).
- Certaines contraintes concernent les relations topographique-topographique, et d'autres contraintes concernent les relations thématique-topographique.

Les choix pour adapter un traitement pour respecter les relations thématique-topographique

Nous identifions deux possibilités en qui concerne la façon dont un traitement est mis en œuvre dans le cadre de la carte à la demande pour prendre en compte des relations thématique-topographique :

- soit paramétrer le traitement pour que les relations soient prises en compte. Par exemple, fournir les relations pour une méthode de sélection.
- soit refaire le traitement pour prendre les relations en compte. Ce choix est pris dans certains cas où le simple paramétrage n'est pas suffisant, n'est pas possible ou si l'utilisateur veut donner une importance maximale aux relations. Par exemple, pour la migration des accidents dans les deux exemples-jouets que nous avons présentés au début de ce chapitre. Dans ces exemples, les méthodes existantes de migration de données (recalage au plus proche, distance proportionnelle, etc.) ne donnent pas le résultat désiré, et il n'existe pas de façon de paramétrer les processus de migration existants en se basant sur les relations. Pour cela, nous avons choisi d'élaborer un processus entier de migration de données thématiques en se basant sur les relations.

Pour pouvoir mettre en œuvre ces deux choix, nous proposons de définir un système dérivé de notre processus de migration de données. Le but du système est d'exécuter des tâches ou des processus avec le respect des relations initiales thématique-topographique.

B.7.4 Système pour la prise en compte des relations pour tous les types de traitements

Dans cette partie, nous présentons un système qui aide à la prise en compte des relations au sein des processus qui concerne des données qui incluent des couches thématiques. Les processus ne concernent pas forcément la carte à la demande. L'avantage de ce système est la réutilisabilité des différentes parties. Par exemple, l'extraction des relations initiales est une tâche importante quel que soit le processus concerné (généralisation, migration de donnée ou intégration, etc.).

Les propositions de cette partie n'ont pas été implémentées et elles peuvent être vues comme des perspectives de cette thèse.

L'idée est de procéder, dans un premier lieu, à la construction du système. Ensuite, si nous avons un processus et que nous voulons le modifier pour prendre en compte les relations spatiales, il faut l'adapter pour l'intégrer au sein du modèle. L'avantage du modèle est de réutiliser les éléments communs entre les différents processus pour la prise en compte des relations.

Le système est composé des modules suivants (Figure 90) :

- module pour l'extraction des objets caractéristiques
- module pour l'extraction des relations initiales
- modules pour l'identification des relations finales attendues

Un processus qui est basé sur les relations utilise les relations attendues stockées. Pour les processus externes (e.g. un processus externe de sélection de réseau), une interface est mise en place pour recevoir des demandes des processus et des systèmes externes.

L'interface du système envoie des relations par demande, un module ou un processus externe reçoit les relations et calcul le paramétrage qui correspond. Par exemple, pour un processus de sélection de réseau, notre système envoie les relations d'alignement d'une route avec les lignes de bus, ensuite, un processus externe change le paramétrage pour que les routes qui sont alignées avec une ligne de bus ne soient pas supprimées par la sélection.

L'interface du système sert aussi à définir les règles de modification et transformations des relations en fonction de processus externes.

En plus de la définition des règles de transformation des relations, le système doit être paramétré par rapport au processus (interne ou externe). Parmi ces paramètres :

- la liste de types d'objets caractéristiques à extraire
- la liste de types de relations initiales à extraire et l'importance de chaque relation par rapport aux besoins de l'utilisateur et/ou par rapport au processus à exécuter

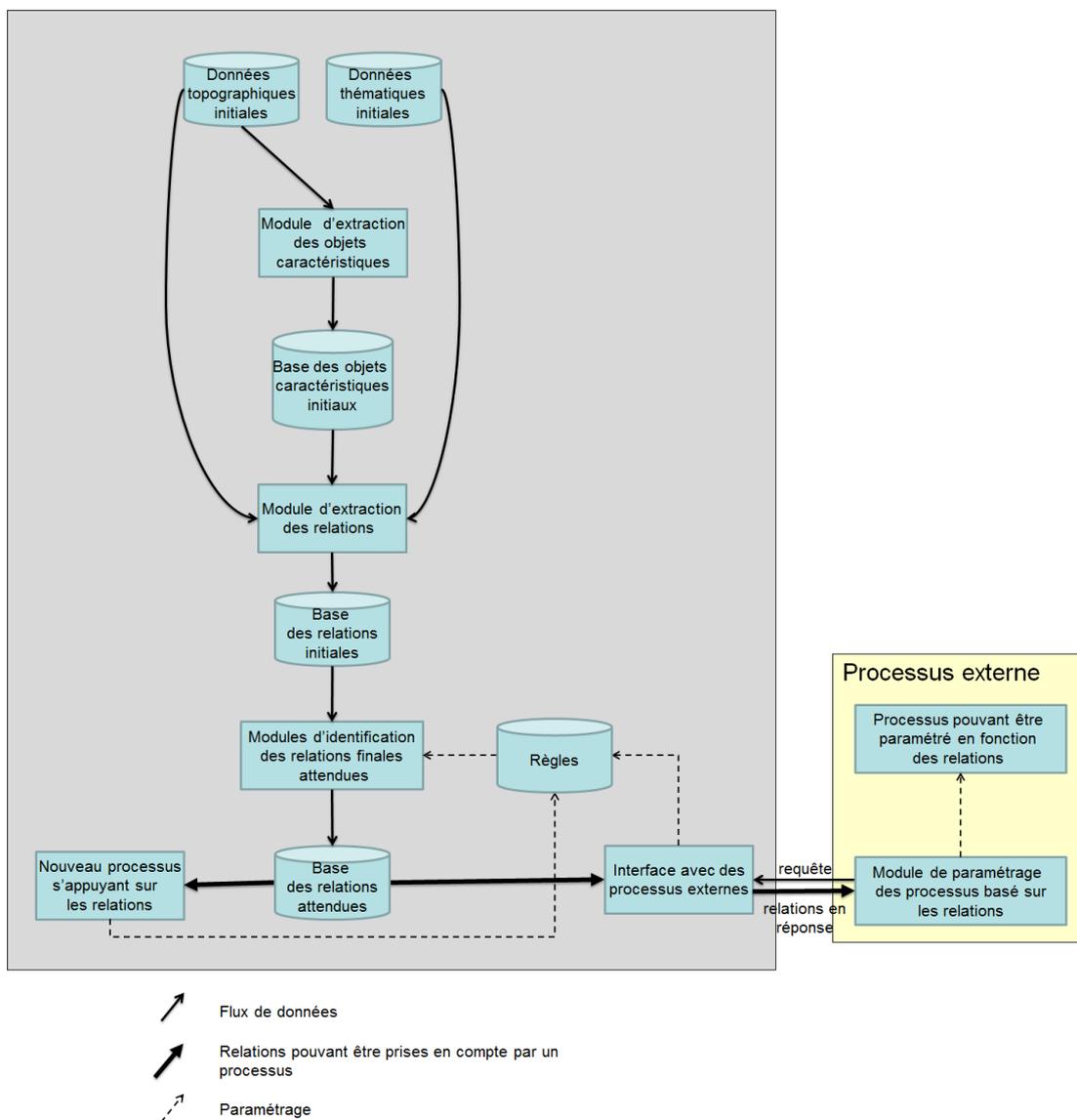


Figure 90. Vue d'ensemble du système pour la prise en compte des relations

Suivant notre approche, les processus qui sont à l'intérieur ou à l'extérieur du système demandent des relations :

- la demande peut concerner un objet topographique, et le processus cherche à savoir si cet objet est en relation avec des objets thématiques. Par exemple, une méthode de sélection qui cherche les routes qui sont en relation d'alignement avec les pistes cyclables.
- un processus peut demander les relations qui concernent un objet thématique, comme dans le cas de la migration de données thématiques. Par exemple, chercher les relations finales attendues qui concernent un accident donné, pour pouvoir le repositionner sur les données topographiques finales.

Les processus reçoivent les relations modélisées suivant notre modèle de description des relations. Nous ajoutons des méthodes dans le modèle de description des relations. Les processus au sein du système doivent être adaptés pour intégrer la satisfaction des relations au sein de leurs étapes. Par exemple, pour le placement d'un accident sur un nouveau jeu de données topographiques, une relation topologique a une fonction *isSatisfied* qui permet de savoir si un placement possible pour l'accident satisfait la relation topologique concernée.

En regardant le modèle de description des relations en Figure 60 (p.71), chaque relation inclut un lien vers l'objet thématique et l'objet du support.

Si nous avons un processus existant et que nous voulons prendre en compte les relations initiales entre les données thématiques et les données topographiques, pour savoir s'il faut seulement un paramétrage ou s'il faut redéfinir le processus au sein de notre système, il faut effectuer les tâches suivantes :

- Etudier les besoins qui sont en lien avec la prise en compte des relations spatiales,
- Etudier les différents paramètres du processus concerné, et distinguer les paramètres qui sont en lien avec les relations,
- Etudier la possibilité de prendre en compte les relations par un paramétrage du processus basé sur les relations,
- Refaire ou modifier le processus, si la prise en compte des relations est impossible en utilisant un paramétrage basé sur les relations,
- Choisir la liste des relations initiales qui sont importantes pour le processus concerné et qui sont indiquées dans les spécifications. S'il existe plusieurs modélisations de la même relation, il faut choisir une modélisation,
- Dans le module d'identification des relations finales attendues, il faut définir les règles de modification et de transformations.

B.8 Bilan du chapitre B

Après la présentation la problématique de la thèse dans le chapitre A, nous avons détaillé notre approche pour mieux prendre en compte des relations entre les données thématiques et les données de référence. Nous avons choisi de commencer par détailler notre méthode pour la migration de données thématiques sur réseau, cas le plus étudié durant la thèse. Le processus de migration est composé de 5 étapes. Ces étapes sont réutilisables pour la migration d'autres types de données (linaire, surfacique). Nous avons aussi détaillé notre vision pour pouvoir réaliser une extension de l'approche pour l'adapter à d'autres traitements que la migration.

L'implémentation de l'approche pour le cas d'étude réel fourni par le CETE a révélé un nombre de problèmes que nous n'avions pas vu pendant l'étude théorique ou pendant l'implémentation sur des exemples-jouets, notamment pour l'appariement. La description de l'implémentation et les problèmes rencontrés lors de la mise en œuvre seront détaillés au chapitre C.

Chapitre C

Mise en œuvre de la migration de données thématiques ponctuelles sur réseau

Dans ce chapitre, nous présentons une mise en œuvre de notre approche sur des cas d'étude-jouets, et sur un cas réel. Le cas réel concerne un problème parmi plusieurs exposés par le centre d'étude technique de l'équipement (CETE) d'Aix.

Nous nous focalisons les problèmes qui concernent la gestion des routes, puis nous focalisons sur la problématique de la migration de leurs données métiers entre les différents référentiels et les méthodes employées actuellement. Ensuite, nous détaillons l'implémentation de notre approche et les problèmes rencontrés lors de cette implémentation ainsi que les solutions apportées. Enfin, nous présentons et analysons les résultats obtenus.

Chapitre C. Mise en œuvre de la migration de données thématiques ponctuelles sur réseau

Le but du chapitre est de présenter en détail une application réelle de notre approche, et de donner les détails techniques de la mise en œuvre de l'approche. Nous allons d'abord présenter la demande du CETE d'Aix pour illustrer des besoins réels en lien avec le changement de niveau de détail (C.1). Ensuite, nous présenterons les cas d'études sur lesquels nous avons testé notre approche (C.2). Il s'agit de cas d'étude jouets et d'un cas d'étude réel. Les détails techniques de la mise en œuvre sont présentés en C.3. Enfin, les résultats sont discutés en C.4.

C.1 Présentation de cas d'étude réel

Les centres d'études techniques de l'équipement (CETE) font partie du réseau scientifique et technique du Ministère de l'Ecologie, du Développement Durable et de l'Energie qui est l'une des tutelles de l'IGN. Ces centres contribuent à la conception et à la mise en œuvre des politiques publiques du Ministère. Dans le cadre de cette thèse, nous avons coopéré avec le CETE Méditerranée¹ qui se trouve à Aix-en-Provence. Les activités du CETE Méditerranée couvrent de nombreux domaines comme la sécurité routière, la gestion des risques, l'exploitation de la route, des analyses sur le transport, l'aménagement des villes, etc. Parmi ses tâches, le CETE Méditerranée est chargé de la constitution d'un référentiel d'information sur le réseau routier. Dans la suite, nous allons utiliser le nom CETE pour référer au CETE Méditerranée.

C.1.1 Constitution de référentiels routiers par le CETE et sources d'erreurs associées

Référentiel « historique » tabulaire

Jusque dans les années 90, les gestionnaires du réseau routier national du Ministère ne géraient pas leurs données en deux dimensions mais par un référencement linéaire à une dimension à partir de nœuds commun en 2 dimensions. La localisation était précisée par la distance suivant l'abscisse curviligne le long de la route et non pas par la position 2D ou 3D.

Dans ce mode de localisation, le système s'appuie sur des bornes physiquement implantées sur le terrain (Figure 91) voire aux carrefours. Le référencement spatial est effectué par des mesures de distances sur les routes. Les données thématiques ponctuelles sont localisées par leur distance curviligne par rapport aux carrefours et aux bornes.

¹ Le CETE Méditerranée est devenu, au 1er janvier 2014, la direction territoriale Méditerranée du Cerema, établissement public issu de la fusion des 8 CETE, du SETRA du CERTU et du CETMEF.



Figure 91. Une borne sur la route

Plus précisément, les routes sont identifiées par un identifiant ou nomenclature, puis les carrefours et les bornes sur la route (appelées PR) sont identifiés et localisés par leur distance selon l'abscisse curviligne dont l'origine est prise au début de la route. Un objet thématique ponctuel est localisé par sa distance curviligne par rapport à la borne la plus proche. La Figure 92 montre un extrait de données thématiques représentées dans ce système, ici des objets linéaires localisés chacun le long d'une même route (parapets et glissières) décrits par leurs points initial et final.

_ID_DR	ROUTE	PRD	ABD	PRF	ABF	LATERALISAT	POSITION	NATURE	DATE
DRD0002-0-10	83 D0002	0	923	0	949	Droite	Extérieur	Parapet	01/05/2008
DRD0002-3-10	83 D0002	3	59	3	147	Gauche	Extérieur	Parapet	01/05/2008
DRD0002-3-20	83 D0002	3	61	3	114	Droite	Extérieur	glissière métal	01/05/2008
DRD0002-3-30	83 D0002	3	317	3	381	Gauche	Extérieur	Parapet	01/05/2008
DRD0002-4-10	83 D0002	4	609	4	733	Gauche	Extérieur	Parapet	01/05/2008
DRD0002-5-10	83 D0002	5	408	5	472	Gauche	Extérieur	Parapet	01/05/2008
DRD0002-5-20	83 D0002	5	413	5	469	Droite	Extérieur	Parapet	01/05/2008
DRD0003-27-20	83 D0003	27	864	27	866	Gauche	Extérieur	Parapet	01/05/2008
DRD0003-28-10	83 D0003	28	403	28	409	Gauche	Extérieur	Parapet	01/05/2008

Figure 92. Extrait d'un recensement d'objets métier : glissières de sécurité du département du Var (Bonin 2010). (PRD, ABD) et (PRF, ABF) correspondent à la description de la position du point initial (respectivement final) de l'objet sous forme (numéro du PR, distance curviligne par rapport au PR).

C'est pourquoi, les principaux logiciels de gestion d'un réseau routier ressemblaient plus à des tableurs (ou à des SGBD) qu'à des SIG. La gamme logicielle *Visage* axée sur cette modélisation est encore très largement utilisée mais n'inclut pas de représentation cartographique. Pour cela, une extension à Mapinfo a été développée, mais cela ne répond pas complètement aux besoins.

Cette architecture *tabulaire* de données répond aux principaux besoins des gestionnaires (analyse, études statistiques, métrés, calculs de budget, schémas d'itinéraires, etc.).

Référentiel basé sur la BD CARTO de l'IGN (appelé RIU V1)

Le premier référentiel routier en 2D constitué par le ministère (appelé RIU pour Référentiel Inter Urbain) date des années 90. Le référentiel est conçu pour répondre au besoin d'échanges entre applications et ajouter une dimension cartographique aux exploitations. Cela a été rendu possible grâce à la disponibilité de données géographiques de référence homogènes sur l'ensemble du territoire métropolitain et en particulier la BD CARTO de l'IGN qui a été le choix naturel de référentiel support.

Le ministère de l'Équipement a ainsi défini le MERIU V1, modèle d'échange qui répond aux besoins de gestion des routes à l'échelle d'un département. Il est utilisé dans un grand nombre d'applications.

Un RIU est le résultat d'une mise en œuvre du modèle MERIU sur des données géographiques. Autrement dit, le RIU comprend les données conformes au MERIU.

Plusieurs applications ont été construites en prenant en compte ce modèle, telles que Concerto (gestion de l'accidentologie), Orchestral (exploitation d'itinéraires). Il est encore utilisé pour actualiser les référentiels départementaux (réseau de l'État et réseau des Conseils Généraux) ainsi que dans l'application Concerto toujours employée dans les Directions Départementales des Territoires (DDT) et certains Conseils Généraux.

Le schéma du RIU V1 est fortement adhérent au schéma de la BD CARTO. En effet, les classes de la BD CARTO contiennent la géométrie pour les classes qui sont liées à la gestion de la route. La Figure 93 présente une petite partie du schéma du RIU V1 pour montrer le lien entre RIU V1 et BD CARTO. Selon le RIU V1, la voie de communication est « *un axe routier sur son ensemble depuis son origine jusqu'à sa fin indépendamment des départements ou communes traversés* ». La section de voie a pour but d'associer des éléments routiers à un tronçon géométrique de la BD CARTO. Enfin, la classe carrefour est associée à la classe *Nœud routier* de la BD CARTO.

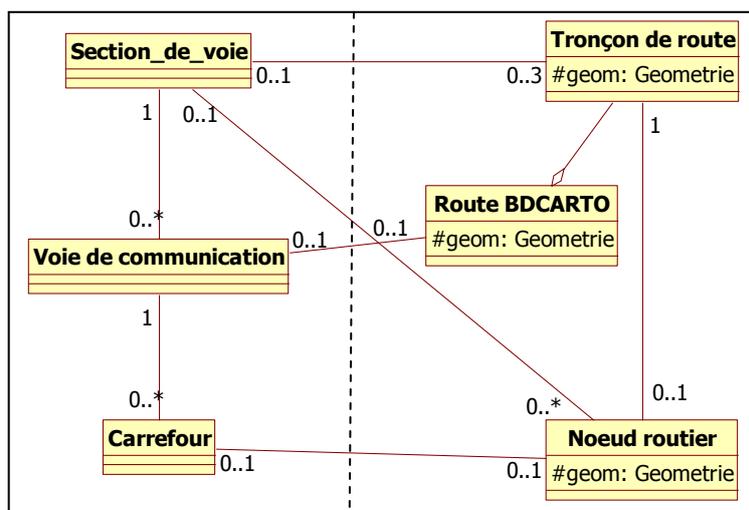


Figure 93. Un schéma simplifié du RIU V1 pour montrer le lien avec le schéma de la BD CARTO

La localisation géométrique est assignée par une liaison avec une base géographique (la Figure 93 illustre cette liaison). La liaison est axée sur les carrefours. Ces carrefours sont connus dans la base métier par leur couple (PR, distance curviligne depuis le PR) et leur distance curviligne depuis le début de la route. Pour relocaliser les données métier (tabulaires) vers une base conforme au RIU V1, la première étape est de trouver les objets homologues entre la base métier et la base topographique. Après un appariement manuel ou par des informations sémantiques entre les carrefours de la base géographique et les carrefours de la base métier, tous les éléments entre couples de carrefours appariés sont interpolés depuis la base métier.

Référentiel basé sur la BD TOPO de l'IGN (appelé RIU V2)

Avec l'arrivée des données plus détaillées de la BD TOPO, le ministère a construit un nouveau référentiel routier (RIU V2), et un nouveau modèle d'échange associé (MERIU V2)

Le MERIU V2 n'impose pas de modèle physique aux gestionnaires, au contraire du MERIU V1 qui répond au schéma de la BD CARTO. La géométrie des objets est indépendante de la base métier. Chaque objet peut avoir une localisation qui peut être issue de n'importe quelle ressource (BD CARTO, BD TOPO ou Route 500, etc.). Cela donne la possibilité de supporter en parallèle plusieurs systèmes de localisation. Pour migrer les informations qui sont saisies dans le RIU V1, le gestionnaire se base sur des données de la base « historique » tabulaire.

Le principe de localisation des données thématiques métier retenu pour le MERIU V2 est similaire à celui du MERIU V1 à savoir une interpolation entre points de repère par distance proportionnelle.

Une des nouveautés du MERIU V2 est la localisation des données thématiques ponctuelles par rapport à des PLO (points de localisation) qui est une extension du principe des points de repère.

La notion de PR a été élargie en utilisant des PLO, qui incluent les PR avec d'autres points de repère qui peuvent être : un point de localisation sur une route ou un dispositif d'échange, un PR, une adresse, un point remarquable sur le terrain, un début et/ou une fin de route, un changement de profil de chaussée...

La construction de ces PLO et leur implantation le long de la route sont effectuées à partir de la base tabulaire par l'interpolation de leurs distances curvilignes entre deux carrefours identifiés simultanément dans la base métier et sur le référentiel topographique.

Sources d'incohérence de localisation entre la base métier et les RIU

Il existe plusieurs sources d'incohérence de localisation entre la base métier et les RIU :

- Erreur dans le choix du carrefour lors de l'appariement.
- Flou sur l'appréciation pour définir les carrefours (incertitude sur leur abscisse dans la BD métier et difficulté à résumer un carrefour complexe à un seul point).
- La longueur d'un tronçon de route (de carrefour à carrefour) peut ne pas être la même dans les référentiels initial et final, soit parce que la géométrie de la route est très différente (notamment quand le niveau de détail est différent), soit parce que la position des carrefours n'est pas la même (cf. cas décrit ci-dessous en Figure 94 et Figure 95).
- La variation de la géométrie peut alors être non homogène le long de la route : lorsqu'un tronçon de route est allongé par déplacement du carrefour final, toute la portion initiale de ce tronçon reste quasiment identique, seule la fin du tronçon est différente. De même si une série de virages appartenant au tronçon de route est fortement simplifiée, seule la géométrie de cette portion est modifiée.

En appliquant la méthode reposant sur la seule distance curviligne, une mauvaise localisation des PLO sur le profil en long de la route engendre mécaniquement des erreurs de localisation

des données thématiques. En effet, la localisation relative par rapport à d'autres objets n'est pas prise en compte. Même le profil de la route autour des points de repères n'est pas pris en compte (courbure par exemple). En conséquence, les données thématiques ne sont pas nécessairement bien relocalisées sur les nouvelles données de référence. En Figure 94, nous illustrons le problème sur un exemple réel. En Figure 94a, les données sont construites à partir de la BD CARTO. Elles sont conformes au MERIU V1. La transformation inclut la modification du schéma, simplification, division et fusion, des suppressions de ronds-points, etc. En Figure 94b, les données de la BD TOPO sont transformées pour qu'elles soient conformes au MERIU V2. Nous remarquons le grand décalage du carrefour entre les deux bases topographiques de référence (le carrefour noté a' de la BD TOPO a été apparié avec le carrefour noté a de la BD CARTO). Cela engendre des différences de localisation entre les deux bases parce que les objets sont localisés uniquement par la distance proportionnelle à partir du carrefour.

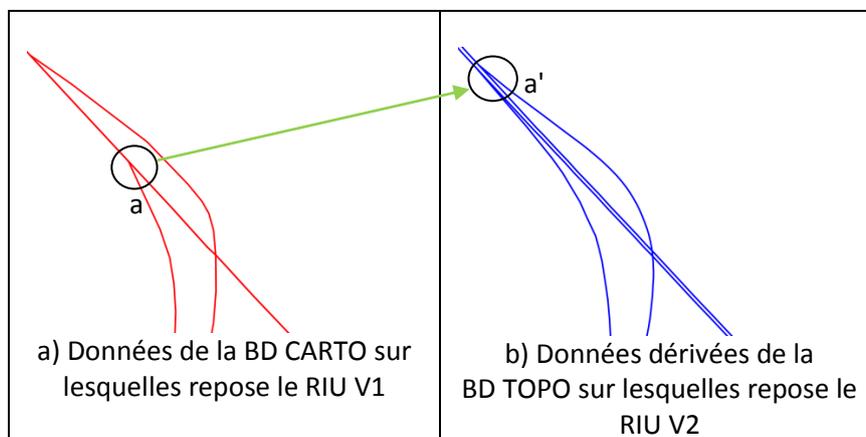


Figure 94. Exemple réel pour expliquer les erreurs lors de la migration de données thématiques à base de distance proportionnelle

La Figure 95 illustre la différence de localisation liée à la différence de la géométrie du carrefour illustrée en Figure 94. Les étoiles rouges sont les PLO qui représentent les bornes qui existent sur le terrain (PR). Ces PLO sont un résultat de migration à partir de la première base (la base historique tabulaire) vers des données conformes au MERIU V1 dérivées de la BD CARTO. Les étoiles bleues sont les PLO qui représentent les bornes sur des données conformes au MERIU V2 dérivées de la BD TOPO. Les décalages entre PLO homologues sur le tronçon de route au nord du carrefour sont illustrés par les flèches vertes. Ces décalages sont dus au fait que la longueur de ce tronçon de route a été augmentée artificiellement à cause de la différence de position du carrefour.

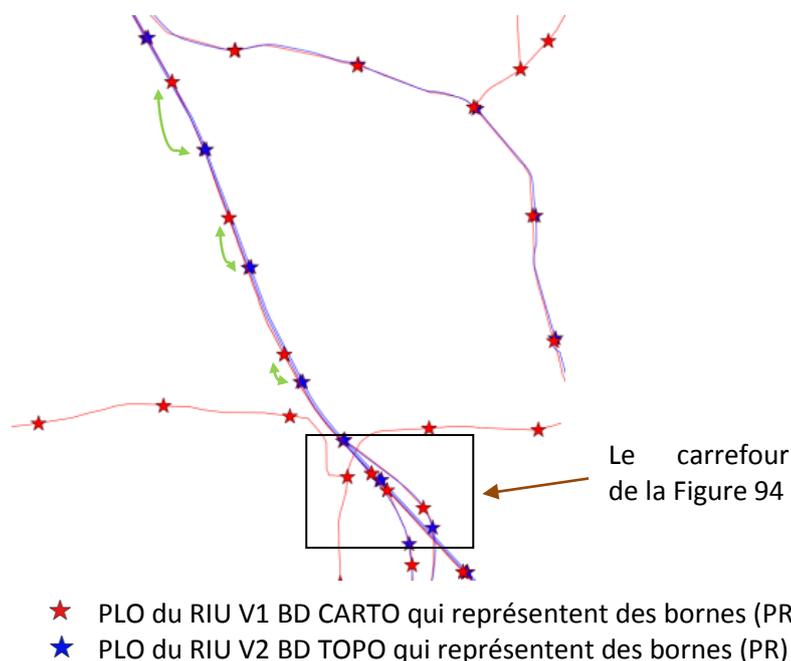


Figure 95. Erreurs constatées lors de la migration de données en utilisant la distance proportionnelle

C.1.2 Conséquences des erreurs de référencement et besoins du CETE

Le changement de référentiel est un besoin récurrent, soit pour visionner des données venant de plusieurs gestionnaires travaillant sur des référentiels différents afin de faire une analyse globale, soit pour remplacer des données topographiques de références par des données plus précises (par exemple, basculer de la BD CARTO vers la BD TOPO), soit enfin pour simplifier l'information de manière à la représenter à une plus petite échelle et sur un plus grand territoire.

Lors d'un changement de référentiel, on retrouve les sources d'erreurs similaires à celles mentionnées ci-dessus à propos du référencement des données métier tabulaires pour constituer les RIU.

Le besoin du CETE est donc de pouvoir changer des données métier (PLO et événements) de référentiel sans perdre les positions relatives des objets métier par rapport aux objets caractéristiques du réseau.

C.2 Cas d'étude retenus pour la mise en œuvre de notre modèle

C.2.1 Démarche et choix des cas d'étude

Au début de la thèse, notre objectif était la prise en compte des relations spatiales lors des changements de niveau de détail sans spécifier le type de traitement concerné.

L'exposé par le CETE des problèmes rencontrés, les données fournies, les discussions, les retours de leur part, nous ont permis de formaliser le processus de migration de données thématiques, et de mettre au point notre modèle.

Avant d'appliquer notre processus à un cas de CETE, nous l'avons appliqué aux cas jouets. Les cas jouets ont servi à choisir des valeurs pertinentes pour les paramètres, et de constater l'avantage de certaines étapes comme la relocalisation en se basant sur une méthode multicritère.

Dans la suite, nous allons d'abord décrire les cas d'étude jouets puis décrire spécifiquement le cas d'étude que nous avons pris parmi les cas d'étude présentés par le CETE.

C.2.2 Description de cas d'étude jouet

Dans nos exemples-jouets (Figure 96), les données topographiques initiales sont des données BD TOPO. Ces exemples sont déjà présentés au début de chapitre B. Chaque exemple contient des objets caractéristiques différents (rond-point, intersection route-rivière, sommets de virages). Nous avons obtenu les données finales en appliquant des fonctions simples qui simplifient les données. Dans l'exemple 1 de la Figure 96, nous avons transformé le rond-point en un carrefour. Dans l'exemple 2, nous avons déplacé le bâtiment et la route a été lissée. Dans l'exemple 3, la route a été lissée.

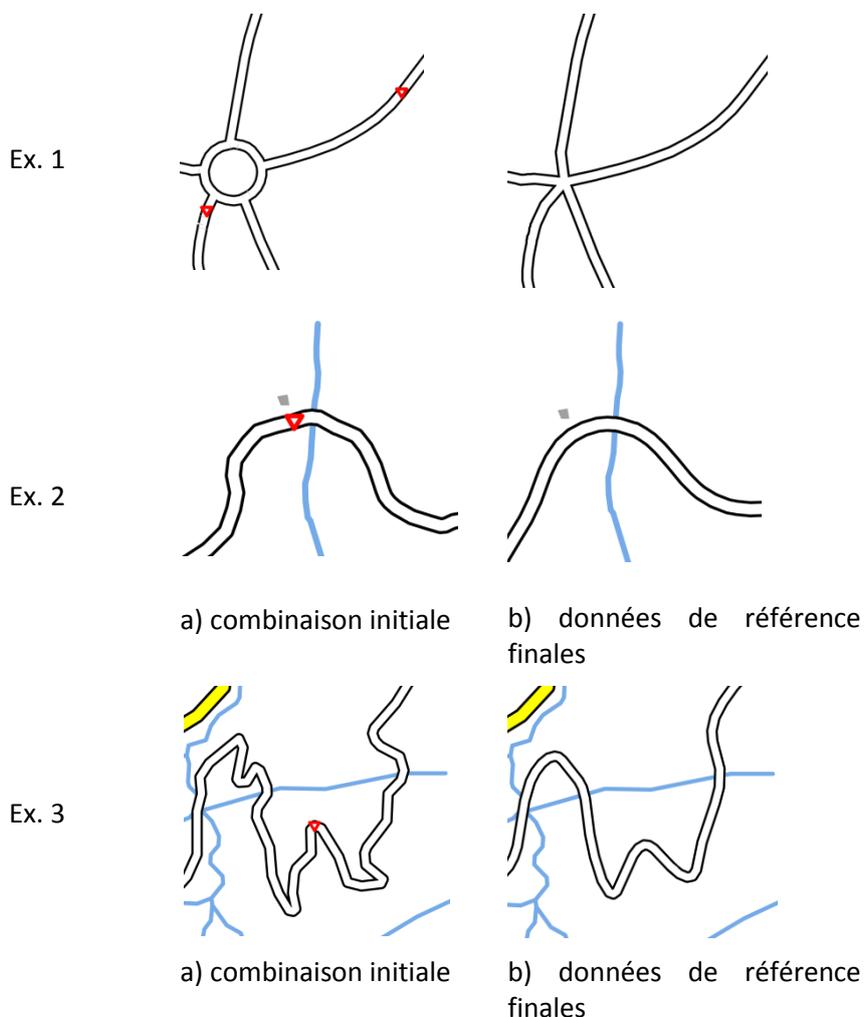


Figure 96. Des exemples-jouets sur lesquels nous avons testé notre approche pour la migration de données thématiques

Dans les exemples-jouets, étant donné que les données topographiques finales sont issues de traitements que nous maîtrisons sur les données initiales, les liens de correspondances entre données topographiques initiales et finales sont connus et nous pouvons nous passer de l'étape d'appariement (3^{ème} étape de notre approche).

C.2.3 Description de notre cas d'étude réel

Après l'analyse du problème, nous avons choisi un cas spécifique de migration de données ponctuelles sur réseau. Il s'agit d'une migration des points de localisation (PLO) du Conseil Général du département du Var (83) vers la BD CARTO. Ce cas correspond à un besoin réaliste d'intégration de données issues de référentiels différents (ici, départemental et national) pour une analyse visuelle conjointe des données.

Les données sur lesquelles nous avons appliqué notre approche

Parmi les données fournies, nous avons utilisé les données sur le département du Var.

Données initiales

1. Des données topographiques : réseau routier (uniquement les routes) constitué et utilisé par le Conseil Général du département du Var (services en charge de la maintenance et de la gestion du patrimoine routier du département). Ce réseau est issu d'un relevé GPS de précision d'ordre métrique.
2. Données ponctuelles des PLO localisés sur les données topographiques du Conseil Général du département du Var (relevé en abscisse curviligne par exploitation d'une campagne de prises de vue horizontales des routes via une caméra embarquée dans un véhicule).

Ces PLO incluent des points qui nous voyons caractéristiques car ils traduisent une réalité géographique : par exemple certains PLO sont des carrefours. D'autres PLO ne sont pas considérés comme caractéristiques car n'offrant aucune correspondance avec la géométrie du réseau ; nous les considérons alors comme des données thématiques. Cette distinction dans la base de départ est rendue possible par les informations sémantiques associées.

Données finales

- Des données topographiques de réseau routier dérivées de la BD CARTO du Var

Objectif visé et évaluation

Nous appelons la méthode utilisée par le CETE, la *méthode proportionnelle*.

Nous souhaitons évaluer si notre approche permet d'améliorer le résultat de la relocalisation par la prise en compte des relations entre objets thématiques et routes : prise en compte de la forme et les relations de proximité avec les carrefours et certains objets caractéristiques comme les sommets de virages et les ronds-points.

Nous évaluons notre méthode en terme de respect des relations initiales. L'évaluation est effectuée par la comparaison entre le résultat de la méthode du CETE (en utilisant la méthode proportionnelle) et le résultat de notre méthode.

La Figure 97 montre un exemple extrait de nos données. Avec la méthode proportionnelle (actuelle), la localisation relative par rapport aux sommets de virages est perdue lors de la relocalisation en particulier, parce que la forme de la route n'est pas prise en compte.

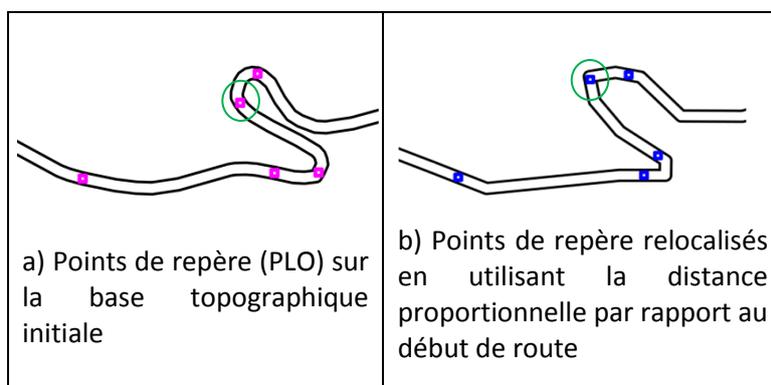


Figure 97. Données initiales et données relocalisées avec la méthode proportionnelle

C.2.4 Différence entre les cas d'étude jouet et le cas réel

Les cas d'étude jouets ont des types d'objets qui ne sont pas présents dans le cas d'étude réel choisi : bâtiments, rivières. Les objets caractéristiques avec lesquels des relations peuvent être calculées sont donc plus variés.

La grande différence se trouve dans l'étape d'appariement. Dans le cas d'étude jouet, l'appariement est trivial car nous disposons d'un lien de correspondance sans effectuer de processus d'appariement. Par contre, dans le cas d'étude réel, nous avons besoin d'effectuer un processus d'appariement.

C.3 Mise en œuvre du modèle sur les cas d'étude

C.3.1 Plateforme d'implémentation

Geoxygene est la plateforme de développement du laboratoire COGIT de l'IGN. Elle capitalise la grande majorité des codes informatiques produits dans le cadre des recherches menées au laboratoire depuis 2002. Les développements sont nombreux : appariement de données, 3D, analyse urbaine, généralisation, modèle de couleur pour la cartographie, aide à la spécification de cartes, transformation de schémas, etc. Le module de généralisation automatique au sein de Geoxygene s'appelle CartAGen (Renard et al. 2011). CartAGen est l'environnement de développement utilisé pour l'implémentation de notre approche.

C.3.2 Modifications apportées à la plate-forme pour les besoins de la thèse

Nous avons dû apporter quelques modifications au module CartAGen pour les besoins de notre thèse. Le module de données de CartAGen a initialement été conçu pour gérer un seul jeu de données. Pendant le processus de généralisation, la géométrie des objets change et certains objets peuvent être éliminés tandis que de nouveaux objets peuvent être créés (par exemple, élimination de bâtiments d'une ville pour les remplacer par un objet ponctuel « ville »). Les objets ont un attribut (*eliminated*) qui indique si l'objet a été éliminé pendant la généralisation. Chaque objet possède une géométrie initiale qui est *null* pour les objets créés pendant la généralisation. Le stockage de la géométrie initiale permet notamment de comparer le résultat de la généralisation avec l'état initial. Ce modèle permet d'optimiser l'usage de la mémoire puisque les nouveaux objets ne sont créés qu'en cas de besoin.

Pour nos besoins, il faut pouvoir stocker et raisonner sur deux jeux de données différents (un jeu initial et un jeu final), donc nous avons modifié le modèle et l'interface de CartAGen² pour pouvoir charger deux jeux de données différents ayant éventuellement des schémas de données différents.

Pour pouvoir travailler sur la migration de données entre un jeu de données et une version généralisée du même jeu de données, nous avons également ajouté la possibilité de dupliquer le jeu de données initiales en stockant des liens de correspondances entre le jeu initial et le jeu dupliqué. Après la duplication, des fonctions de généralisation peuvent être appliquées sur le jeu de données final pour obtenir un jeu de données final généralisé.

C.3.3 Préparation des données

Pendant le chargement des données du Conseil Général du Var dans notre modèle, un prétraitement a été nécessaire sur les ronds-points. Dans les données du Conseil Général, chaque rond-point est représenté par une seule section fermée comme prévu par le modèle de données qui n'a pas de vocation topologique.

Mais cela pose des problèmes pour la suite puisque certains traitements utilisés se basent sur la théorie des graphes et que nous avons besoin d'un graphe planaire pour représenter le réseau routier. Ainsi, un traitement a été fait sur les données initiales du Conseil Général pour diviser les ronds-points en sections.

Dans la suite, nous détaillons la mise en œuvre de notre processus de migration étape par étape (reproduit en Figure 98).

² L'amélioration de l'architecture de CartAGen a été réalisée avec l'aide de Guillaume Touya et Jérémy Renard que nous remercions.

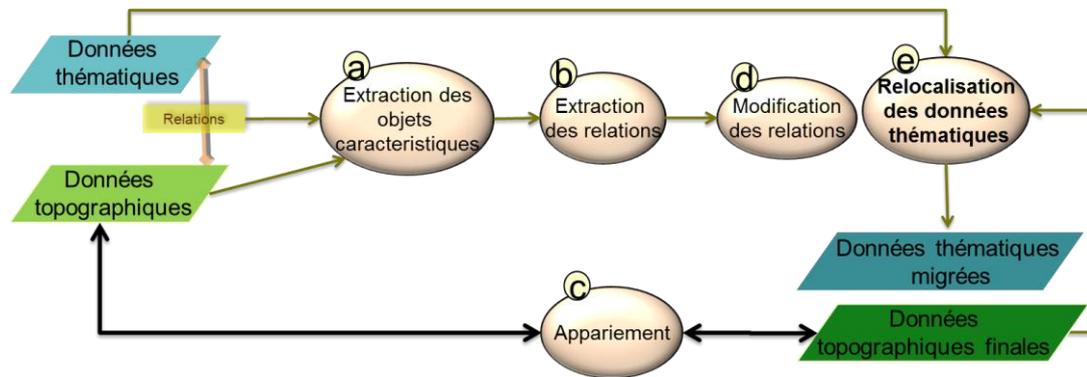


Figure 98. Le processus de migration de données thématiques

C.3.4 L'extraction des objets caractéristiques (étape a)

Identification des sommets de virages

Nous avons d'abord essayé d'appliquer la méthode de (Plazanet 1996), ensuite, nous avons réalisé une autre méthode pour répondre au mieux à nos besoins.

Méthode de Plazanet : un virage simple peut être défini comme une portion de ligne courbe entre deux inflexions consécutives significatives pour le niveau d'analyse considéré. Un point d'inflexion est un changement de tendance de direction. Un virage complexe peut être composé de virages simples. Les virages ont des critères permettant de les détecter, comme les critères de forme, de complexité (Plazanet 1996, p.71). Une méthode d'extraction des points d'inflexions a été proposée par (Plazanet 1996, p.96). La méthode consiste à appliquer un lissage pour filtrer les micro-inflexions. Le principe de détection des points d'inflexion repose sur le changement des produits vectoriels le long de la ligne. Un point d'inflexion est détecté pour chaque changement du signe de ces produits.

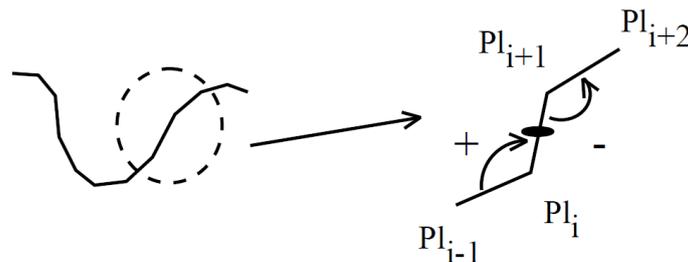


Figure 99. Principe de détection des points d'inflexion (Plazanet 1996, p.97)

Malgré l'application d'un lissage avant la détection des points d'inflexion par cette méthode, des micro-inflexions ont été malgré tout détectées. Le sommet du virage est le point le plus éloigné de la ligne des points d'inflexion en projection orthogonale.

Notre méthode : nous proposons une méthode axée sur les changements d'angles. Cette méthode n'a pas été étudiée d'une manière robuste comme la méthode de Plazanet, mais elle a donné le résultat attendu sur l'ensemble des données de test. En plus, notre méthode détecte certains points caractéristiques sur la route qui aident à bien localiser les données

thématiques. L'idée est de calculer la somme des angles entre les sections dans une distance donnée sur la route (par exemple 70 mètres pour l'échelle 1:25000). Si la somme des écarts d'angles entre segments consécutifs est supérieure à un seuil donné (Par exemple, 100 degrés), nous considérons que nous avons un virage. Après la détection d'un virage, nous utilisons la même méthode que (Plazanet 1996) pour extraire le sommet de virage. La Figure 100 montre les angles entre les sections et la Figure 101 montre un exemple fictif d'un sommet de virage détecté avec un angle cumulé de 330 degrés.

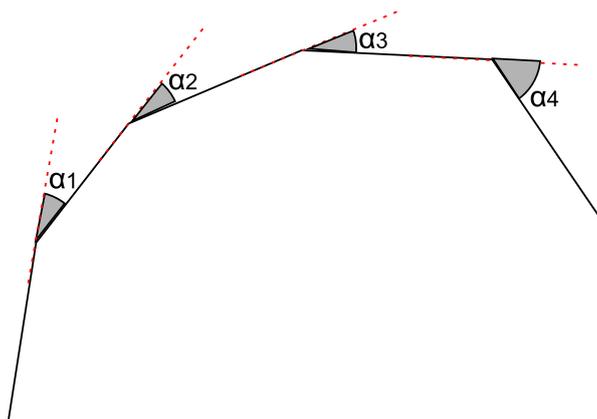


Figure 100. Détection des virages par la somme des angles

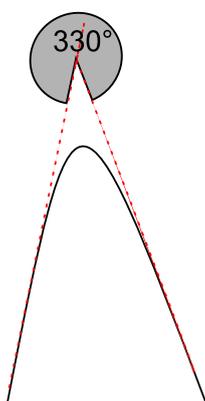


Figure 101. La somme des angles a pour but de détecter les virages

Le point délicat est le choix des valeurs paramétriques en fonction de la résolution des données. Dans notre implémentation, nous avons testé des choix de paramètres avant de choisir un jeu de paramètres. En effet, dans les deux méthodes, une étape préliminaire est d'effectuer un lissage. Le degré de lissage est variable en fonction des spécifications des données.

En Figure 102, nous montrons le résultat de l'application de notre méthode sur des données fournies par le CETE. Les données de la Figure 102a sont issues de la BD TOPO, les données de la Figure 102b sont des données du Conseil Général du Var et les données de la Figure 102c sont des données RIU V1 issues de la BD CARTO. Les Figure 102b et Figure 102c représentent la même région géographique.

Pour les points p6 et p7 de la Figure 102a, nous remarquons que les deux points sont clairement des points caractéristiques sur la route, mais un changement de tendance de direction n'est pas présent entre les deux points. Alors seul l'un des deux est détecté par la méthode de (Plazanet 1996).

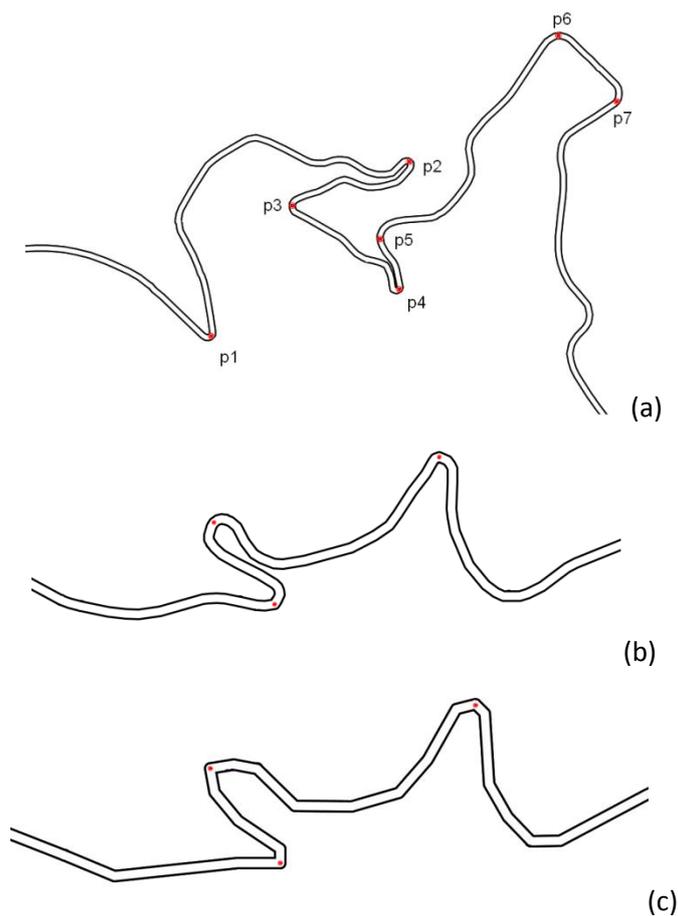


Figure 102. Des sommets de virage détectés en utilisant notre méthode de détection

Ronds-points

L'extraction des ronds-points est cruciale pour le CETE, parce que la représentation des ronds-points peut varier entre les différentes bases de données topographiques. Selon l'échelle, un rond-point peut être représenté par un carrefour simple sous forme d'objet ponctuel ou par l'ensemble des sections le constituant. Ce changement doit être pris en compte pendant la migration de données thématiques. Dans certaines bases de données topographiques, les ronds-points sont facilement identifiés par des attributs, par exemple, l'attribut « type » dans la BD CARTO. Dans les autres cas, notamment si les ronds-points sont complexes (avec des pattes d'oies), il faut les extraire en utilisant des méthodes dédiées. Dans notre implémentation, la méthode de l'extraction utilisée est la méthode de (Sheeren 2005, p.190). En plus de l'extraction, l'auteur propose une modélisation qui permet de classifier les sections

attachées au rond-point, car il faut distinguer les pattes d'oies et les routes principales parmi les sections attachées.

La méthode consiste à effectuer des analyses spatiales, elle est composée de trois étapes :

1. Construction d'une carte topologique à partir des tronçons de route et des nœuds routiers
2. Analyse de la forme des faces, chaque face qui a la forme d'un rond-point sera reconnue comme un rond-point (si la face est circulaire) en utilisant l'indice de circularité de Miller (Miller 1953).
3. Les ronds-points extraits sont ensuite combinés aux nœuds routiers comme des « grand ronds-points ».

Carrefours

Les carrefours sont extraits par la construction d'une carte topologique à partir des routes. Ensuite, il faut rendre le graphe planaire pour extraire tous les nœuds. Un nœud est reconnu comme un carrefour, si son degré est supérieur à deux. Un objet caractéristique de type carrefour a un lien avec toutes les routes auxquelles il est connecté.

Intersections route-rivière

Les intersections route-rivière ont été extraites et utilisées sur l'exemple jouet (par exemple, le cas en Figure 77 p.93)³. La méthode d'extraction est la suivante :

1. Réunir les routes et les rivières ensemble pour obtenir un jeu de données qui contient juste leur géométrie.
2. Garder un lien vers l'objet d'origine des objets fusionnés.
3. Construire une carte topologique à partir de données fusionnées.
4. Transformer la carte topologique pour la rendre planaire.
5. Tester les nœuds résultats après l'étape précédente. Un nœud est reconnu comme une intersection route-rivière s'il a au minimum un arc venant d'une route et un arc venant d'une rivière.

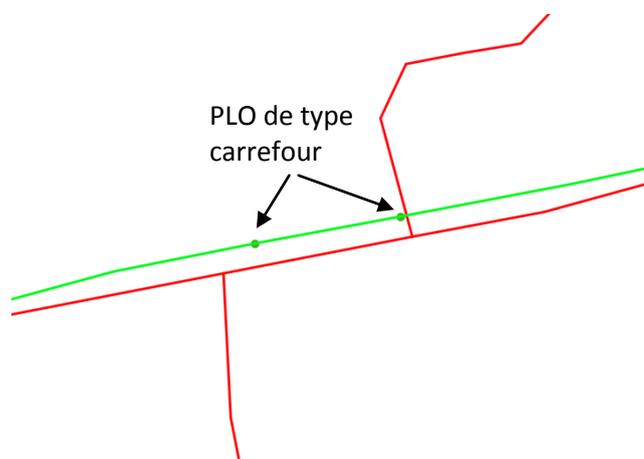
Objets caractéristiques supplémentaires issus de la sémantique des objets thématiques

Certains objets caractéristiques peuvent être fournis par l'utilisateur ou par le fournisseur des données topographiques. C'est le cas des débuts de tunnel. Dans notre cas d'étude du CETE, les PLO sont stockés dans une seule classe et le type de PLO est donné par les attributs. A titre d'exemple, sur un autre jeu de données sur le département du Vaucluse, les PLO qui sont saisis ont également un attribut, TYPEPLO, qui indique leur typologie (carrefour, début d'une section de double chaussées d'une route, début et fin de tunnel, etc.). Ces PLO sont considérés comme des points caractéristiques puisqu'ils peuvent correspondre à des objets caractéristiques qu'on

³ Nous remercions Julien Perret, chargé de recherche à l'IGN, qui nous a aidé à manipuler la carte topologique implémenté dans GeOxygene de manière pertinente pendant le développement de cette méthode.

peut extraire des données topographiques finales (la BD CARTO de l'IGN, dans lesquels les tunnels sont représentés). En plus, ces objets caractéristiques initiaux ont des informations sémantiques sur leur localisation ou sur leurs relations.

La Figure 103 montre la superposition des bases initiale et finale sur un extrait de nos données. Les PLO ont été saisis sur le réseau du Conseil Général en utilisant la méthode de distance proportionnelle à partir de données tabulaires. Les deux PLO représentent les deux carrefours qui sont facilement identifiables dans la deuxième base topographique (en rouge). Il faut donc que, lors de leur migration de la base initiale (en vert) vers la base finale (en rouge), les PLO soient positionnés sur les carrefours. L'information qui indique que le PLO est un carrefour est stockée dans l'attribut TYPEPLO.



Dans l'exemple de la Figure 103, les PLO portent un attribut traduisant leur localisation sur un objet caractéristique des données topographiques (ici, un carrefour), mais d'autres PLO n'ont pas ces informations. Par exemple, certains PLO du Conseil Général correspondant à des PR sont matérialisés sur le terrain par des marques au sol qui ne peuvent avoir de correspondance dans la base topographique finale. Cette information n'aide pas à la localisation précise du PLO. Les PLO qui n'ont pas une information claire et certaine sur leur localisation seront considérés comme des données thématiques et relocalisés pendant la phase de migration des données.

C.3.5 Relations qui sont prises en compte et méthodes d'extraction (étape b)

Relation « Porté par » <objet thématique ponctuel, route>

La relation « porté par » (discuté en B.3.3 p.62) représente le lien fort entre l'objet thématique ponctuel et son espace de support, nous avons proposé de découvrir cette relation d'une façon semi-automatique en examinant les relations d'inclusion ou de proximité entre les données thématique et une classe de données topographiques.

La méthode du calcul du cas général a été présentée B.3.3, mais dans notre implémentation, nous considérons que les objets thématiques ponctuels sont tous portés par les routes. Ceci est due au fait que c'est la seule relation utilisée par le CETE.

Relation topologique <objet thématique ponctuel, route >

Le but de cette relation est de décrire la relation topologie de l'objet thématique ponctuel avec les routes qui l'entourent. La relation topologique point-ligne est calculée en définissant un buffer autour du point thématique, si le buffer intersecte une route, une instance de relation point-ligne est créé. Selon notre modélisation (illustrée en Figure 62 p.75), il faut distinguer le prédicat parmi les trois prédicats (sur la ligne, proche de la ligne, ou entre les deux). Le prédicat est choisi par une comparaison de la distance entre l'objet thématique ponctuel et le point le plus proche de la route. Nous avons défini trois plages de distance (illustrées par une barre colorée en Figure 104)

- Distance > S2 => le point thématique est proche (illustrée en noir en Figure 104)
- Distance < S1 => le point est sur le route (illustrée en bleu en Figure 104)
- Distance < S2 et Distance > S1 => le point est entre les deux cas précédent (flou) (illustrée en vert en Figure 104)

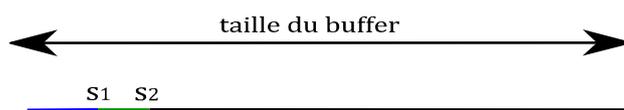


Figure 104. Comment choisir le prédicat suivant la distance entre le point et la route

Relation de distance curviligne <objet thématique ponctuel, point caractéristique>

La distance est utilisée pour définir une relation de proximité entre deux objets ponctuels sur une route. Par exemple, la distance entre l'objet thématique ponctuel et l'intersection route-rivière (Figure 96).

Relation « distance curviligne » <objet ponctuel, une extrémité de la route porteuse >

Cette relation sert à référencer l'objet thématique ponctuel le long de la route qui le porte.

Cette relation n'a qu'un attribut qui est la distance curviligne sur la route par rapport à l'extrémité de la route. La distance est calculée entre le début de la route et le point le plus proche de la route par rapport à l'objet thématique ponctuel (qui est confondu avec l'objet thématique ponctuel s'il est inclus dans la route).

Relation de proximité <objets thématique ponctuel, bâtiment>

Cette relation est extraite par la création d'un buffer autour du point thématique, si un bâtiment intersecte ce buffer, une instance de la relation est créée. La relation a un seul attribut qui est la distance. La distance est calculée en mesurant la distance euclidienne entre le point thématique et le centroïde du bâtiment.

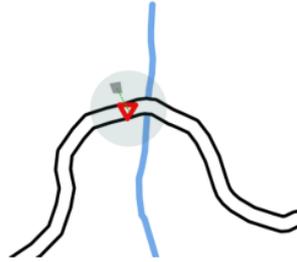


Figure 105. Calcul de la relation de proximité entre point thématique et un bâtiment

Relation de proximité <objet thématique ponctuel, rond-point>

Si une des deux extrémités de la route est un rond-point, une instance de la relation de proximité point, rond-point est créée. La relation possède un seul attribut qui représente la distance entre le point thématique et la fin de la route sur l'entourage du rond-point. Nous allons aussi avoir besoin du diamètre du rond-point, mais cette information est déjà stockée dans l'objet support.

Relations de direction <objet thématique ponctuel, bâtiment>

Cette relation est extraite par la création d'un buffer autour du point thématique, si un bâtiment intersecte ce buffer, une instance de la relation est créée. La relation a un seul attribut qui est l'orientation dans l'absolu modulo 2π du segment orienté joignant le centroïde du bâtiment au point thématique. Si nous avons le point thématique $p(x_p, y_p)$ et le bâtiment bat ayant pour centroïde $c_{bat}(x_c, y_c)$, alors

si $x_p = x_c$ alors

$$\text{anglePointBatiment} = +\frac{\pi}{2} \text{ si } y_p > y_c, \text{ et } \text{anglePointBatiment} = -\frac{\pi}{2} \text{ si } y_p < y_c$$

sinon

$$\text{si } x_p > x_c \text{ alors } \text{anglePointBatiment} = \text{Arctan}\left(\frac{y_p - y_c}{x_p - x_c}\right) \text{ (compris entre } \frac{\pi}{2} \text{ et } +\frac{\pi}{2})$$

$$\text{si } x_p < x_c \text{ alors } \text{anglePointBatiment} = \text{Arctan}\left(\frac{y_p - y_c}{x_p - x_c}\right) + \pi \text{ (compris entre } +\frac{\pi}{2} \text{ et } +\frac{3\pi}{2})$$

Relation de forme de la route autour des objets ponctuels <objet ponctuel, route>

Le but de cette relation est de décrire la forme locale de la route autour d'un point sur la route. La relation de la forme de route aide à décrire la position de l'objet thématique sur la route. Le but est de pouvoir repositionner l'objet thématique dans une position relativement similaire. Par exemple, en Figure 106, l'objet thématique doit rester à la sortie de virage dans la base finale.

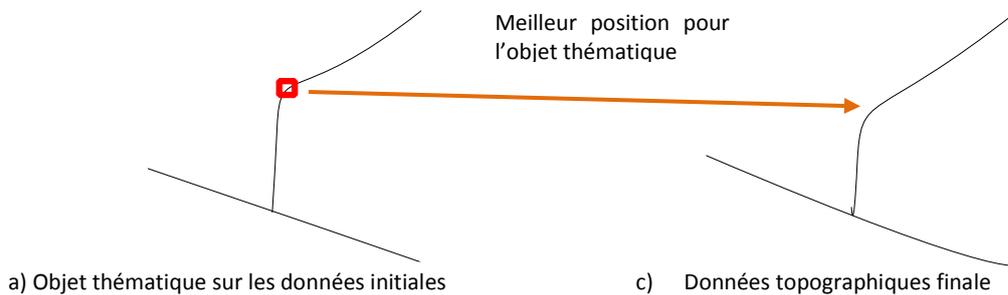


Figure 106. Exemple d'un objet thématique qui se trouve à la sortie du virage

Notre méthode est inspirée de (Buttenfield 1991) où le but était de décrire le niveau de détail d'un objet linéaire par des mesures géométriques. La méthode est axée sur la création d'une fenêtre carrée autour du point thématique, la taille du carré est déterminée de manière empirique en fonction de l'échelle. Ensuite, on calcule l'intersection de l'objet linéaire qui porte le point thématique avec la fenêtre, puis on prend les mesures illustrées dans la Figure 107 :

$Pt(x_p, y_p)$ est l'objet thématique ponctuel

$Ps1(x_{ps1}, y_{ps1}), Ps2(x_{ps2}, y_{ps2})$ sont les points résultant de l'intersection entre l'objet de référence linéaire et la fenêtre

Les mesures considérées sont :

$$dx1 : x_{ps1} - x_p$$

$$dx2 : x_{ps2} - x_p$$

$$dy1 : y_{ps1} - y_p$$

$$dy2 : y_{ps2} - y_p$$

$anglePs1Ps2$: orientation absolue modulo 2π du segment orienté joignant $Ps1$ et $Ps2$, calculé comme pour la relation de direction <objet thématique ponctuel, bâtiment> ci-dessus:

si $x_{ps1} = x_{ps2}$ alors

$$anglePs1Ps2 = +\frac{\pi}{2} \text{ si } y_{ps2} > y_{ps1}, \text{ et } anglePs1Ps2 = -\frac{\pi}{2} \text{ si } y_{ps2} < y_{ps1}$$

sinon

$$\text{si } x_{ps2} > x_{ps1} \text{ alors } anglePs1Ps2 = \text{Arctan}\left(\frac{y_{ps2}-y_{ps1}}{x_{ps2}-x_{ps1}}\right) \text{ (compris entre } -\frac{\pi}{2} \text{ et } +\frac{\pi}{2})$$

$$\text{si } x_{ps2} < x_{ps1} \text{ alors } anglePs1Ps2 = \text{Arctan}\left(\frac{y_{ps2}-y_{ps1}}{x_{ps2}-x_{ps1}}\right) + \pi \text{ (compris entre } +\frac{\pi}{2} \text{ et } +\frac{3\pi}{2})$$

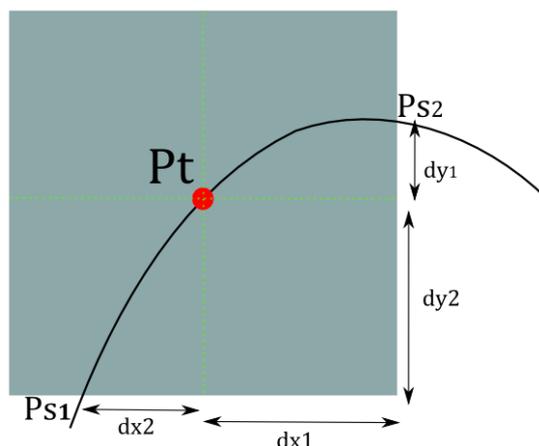


Figure 107. Mesures pour représenter la forme de la route autour d'un point thématique

La relation est donc caractérisée par les cinq valeurs ($dx1, dx2, dy1, dy2, anglePs1Ps2$).

L'inconvénient de la méthode est que les changements de direction à l'intérieur de la fenêtre ne sont pas pris en compte. Pour minimiser des erreurs éventuelles dues à cet inconvénient, il faut bien choisir la taille de la fenêtre : pas trop petite pour représenter suffisamment la forme de la route et pas trop grande.

Figure 108 montre deux formes de route qui sont représentées par les mêmes attributs selon notre méthode bien que le profil soit différent. En effet, la taille du carré correspond au niveau de détail des données de référence initiales. Dans notre implémentation nous avons choisi empiriquement la taille du carré (100X100 mètres).

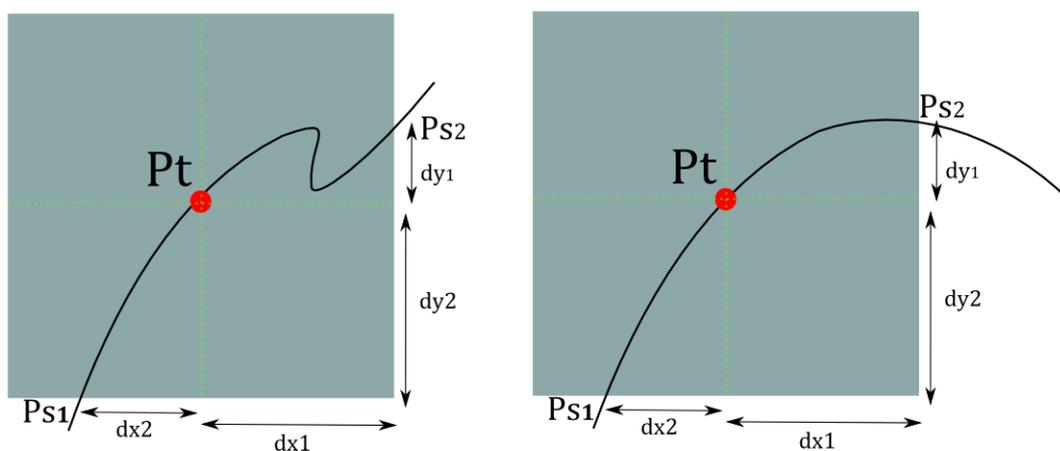


Figure 108. Possibilité d'erreur due à une taille de fenêtre très grande

A cause de contraintes de temps, nous n'avons pas pu faire de test complet de la méthode, mais nous l'avons testée sur les données réelles du CETE. Grâce à la prise en compte de cette relation, nous avons réussi à améliorer la relocalisation des points thématiques (voir le résultat à la fin de ce chapitre).

C.3.6 Mise en œuvre de l'appariement (étape c)

Nous avons discuté en (B.4.1, p.81) l'étape d'appariement. Il a pour but d'extraire les liens de correspondance entre la base initiale et la base finale. Les liens de correspondance concernent tous les objets supports : les objets topographiques et les objets caractéristiques.

Pour les cas d'étude jouets et le cas du CETE, on a besoin d'apparier les objets caractéristiques.

Pour apparier les réseaux routiers, nous avons essayé d'appliquer la méthode d'appariement proposée par (Mustière et Devogele 2008). Cette méthode concerne l'appariement de réseaux routiers qui ont deux niveaux de détails différents. La méthode trouve les liens d'appariement quand le réseau final est plus détaillé que le réseau initial. Nous avons fait ce choix aussi parce que nous disposons d'une implementation assez robuste de cette méthode dans la plateforme Geoxygene. L'implementation de cette méthode d'appariement a été testée pour apparier les données de la BD CARTO avec la BD TOPO. L'application de la méthode n'a pas donné les résultats attendus pour notre cas d'étude, car les données ne sont pas assez complètes. En regardant la Figure 109a, les données du Conseil Général de var représentent seulement la partie gérée par le Conseil Général.

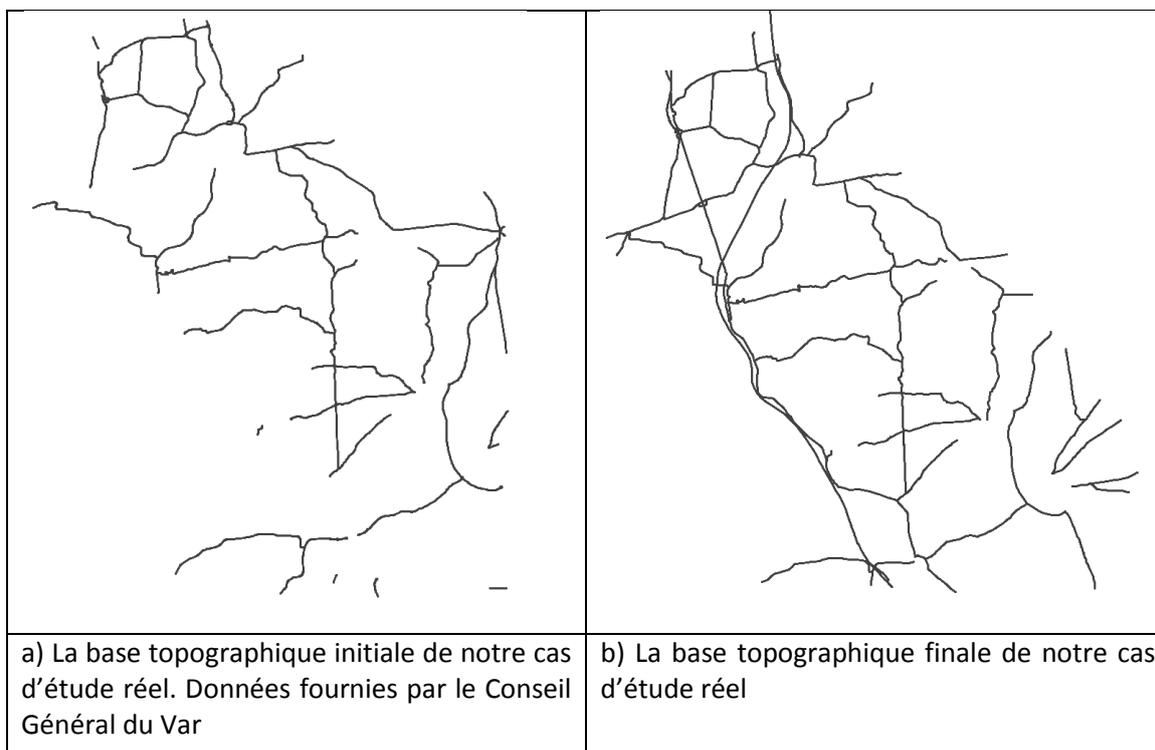


Figure 109. Un extrait des bases topographiques initiale et finale fournies par le CETE

Le découpage des routes est différent entre les bases de données. Alors pour assurer une continuité, nous avons proposé d'utiliser les strokes (voir la définition des strokes en B.4.1).

Par ailleurs, nous identifions les besoins spécifiques suivants :

- Il faut garder l'ordre entre les objets thématiques ponctuels sur chaque route

- Certains objets caractéristiques sont fournis par le CETE, mais ils ne sont pas tous appariés. Il faut qu'ils soient relocalisés par distance curviligne à partir d'un objet caractéristique apparié.

Pour répondre à nos besoins, nous proposons un appariement inspiré de l'appariement de (Mustière et Devogele 2008). L'appariement apparie les objets caractéristiques et trouve des portions des routes qui sont correspondantes. Notre processus d'appariement proposé est composé de 4 étapes :

1. Construire deux graphes : un graphe qui contient tous les objets supports initiaux qui sont sur la route et un autre graphe pour les objets supports finaux.
2. Construire les strokes sur les graphes initial et final
3. Apparier les objets caractéristiques
4. En utilisant les strokes, distinguer les portions des strokes qui sont appariées

Etape 1 : Construction des graphes

Nous proposons de construire pour chacun des états (initial et final) un graphe qui contient tous les objets supports qui sont sur la route. Par exemple, si nous avons la combinaison montrée en Figure 110a, le graphe construit à partir de cette combinaison sera celui de la Figure 110b. Pour chaque objet thématique ponctuel ou support, un nœud est créé. Le nœud contient un lien vers l'objet support initial. Les arcs sont créés à partir d'une division des routes par les objets.

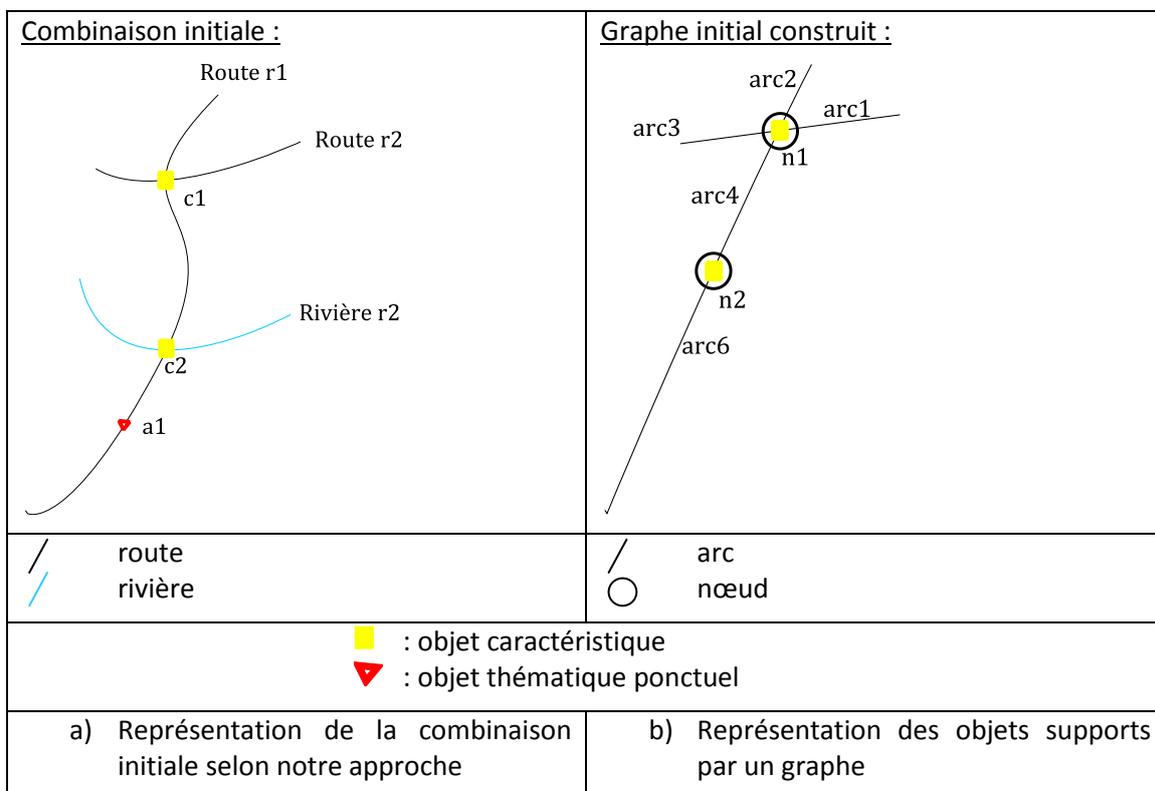


Figure 110. La construction du graphe

Etape 2 : Construction des « strokes »

Les *strokes* permettent d'améliorer le résultat d'appariement, parce qu'ils nous permettent d'assurer une continuité. Il faut que les strokes obtenus soient continus même dans le cas où nous avons des ronds-points, car un rond-point peut ne pas être représenté dans l'autre base. Par exemple, le rond point illustré en Figure 111 est supprimé dans la base topographique finale. La relocalisation des objets thématiques qui se situent sur le rond point nécessite un stroke qui continue tout le long de la route et ne doit pas être coupé par le rond-point. Pour construire les strokes, nous avons utilisé la méthode de (Touya 2007), basée sur des critères attributaires et d'angularité des routes, et qui est implémentée dans GeOxygene. Les ronds-points sont détectés avant la construction des strokes et chaque rond-point est modélisé comme un nœud dans notre graphe. La construction des strokes est faite à partir de ce graphe.

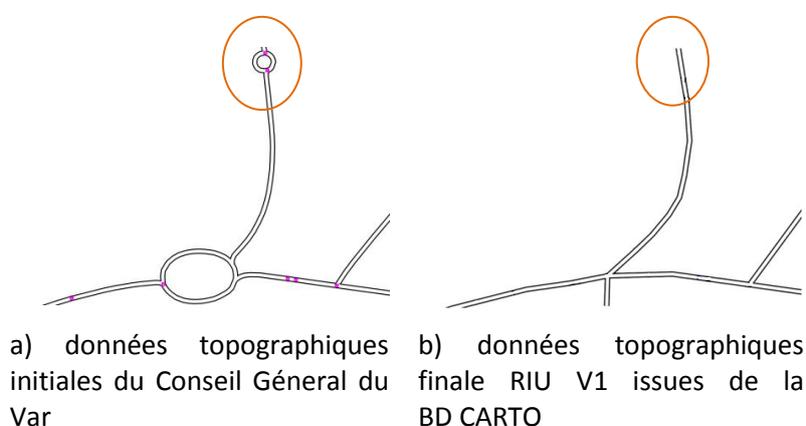


Figure 111. Le besoin de dépasser les ronds-points pendant la construction des strokes

En prenant l'exemple type de la Figure 110, un stroke peut être construit le long de la route1. Pour le graphe, le stroke passe par la liste d'arcs suivante : (arc2, arc4, arc6).

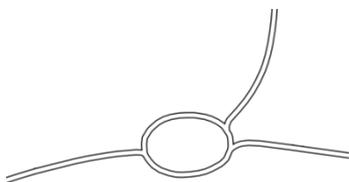
Etape 3 : Appariement des objets caractéristiques

Selon notre définition des objets caractéristiques (cf. B.3.1 p.59), « sont choisis pour être a priori reconnaissable dans les différentes représentations topographiques de la région concernée ». Il faut alors les retrouver dans la base finale. Il convient d'extraire d'abord les objets caractéristiques dans la base finale, puis de les appairer. Dans notre cas d'étude, cinq appariements ont été mis en œuvre :

1. <rond-point, rond-point>
2. <rond-point, carrefour>
3. <carrefour, carrefour>
4. <PLO (de type carrefour), carrefour>
5. <Sommet de virage, sommet de virage>

Les quatre premiers appariements sont effectués en se basant sur la toponymie des routes. Les sommets de virage sont appariés par la distance. Parmi les résultats d'appariement, nous avons constaté que certains ronds-points ne sont pas reconnus parce que leur forme n'est pas

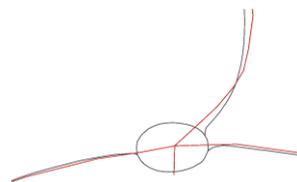
suffisamment ronde. Par exemple, dans la Figure 112a, le rond point n'est pas reconnu pendant la phase de l'extraction des objets caractéristiques.



a) données topographiques initiales du Conseil Général du Var



b) données topographiques finale RIU V1 issues de la BD CARTO



c) données topographiques initiales et finales

Figure 112. Cas spécial concernant l'appariement <rond-point, carrefour>

Etape 4 : distinguer les portions des strokes qui sont appariées

Suite à l'appariement des objets caractéristiques, certains ne seront pas appariés. Par exemple, en Figure 113, *c1* sera apparié avec *ca1* et *c3* sera apparié avec *ca3*, mais *c2* ne sera pas apparié. Alors lors de cette étape, la portion du stroke initial de *c1* à *c3* sera appariée avec la portion du stroke final de *ca1* à *ca3*.

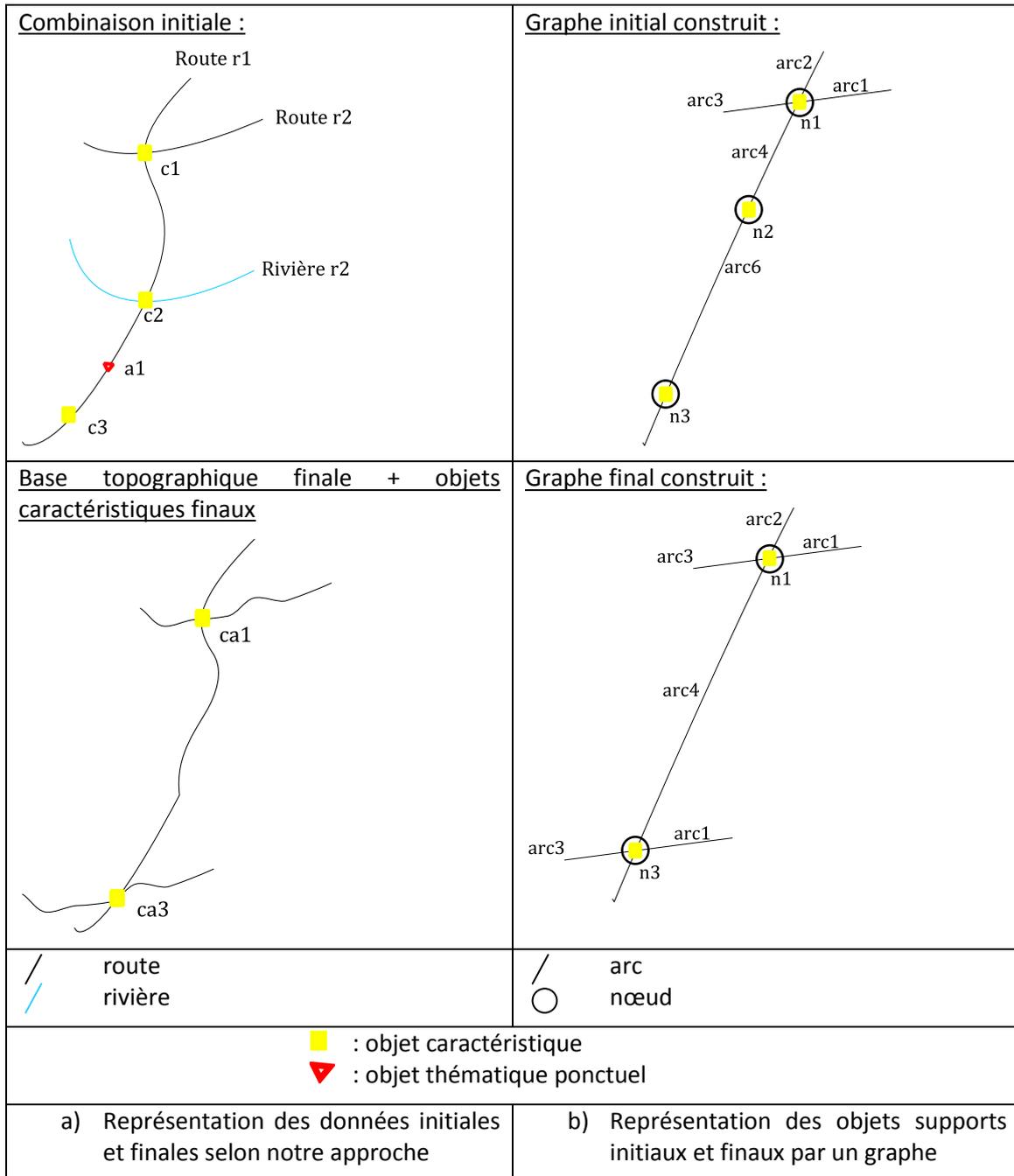


Figure 113. La construction du graphe

C.3.7 Identification des relations finales attendues (étape d)

Dans cette étape, nous supposons que nous avons les liens de correspondance entre les objets supports. Les liens de correspondance sont de type 1-1, 1-0, et N-1, les autres types ne sont pas étudiés dans le cadre de la thèse.

Le but de cette étape est d'identifier les relations attendues au sein de l'état final. Pour chaque relation, la méthode d'identification génère une relation finale attendue. La relation attendue est entre l'objet thématique final et l'objet support final qui est l'homologue de l'objet support initial dans les données finales.

Le processus d'identification d'une relation attendue à partir d'une relation initiale est guidé par des règles de modification de type condition/conclusion, les conditions dépendant du type du lien d'appariement et des spécifications modélisées de l'état final. Dans notre implémentation, nous n'avons pas mis en œuvre une modélisation des spécifications, mais nous avons défini des règles en dur pour un cas d'utilisation portant sur l'analyse de la localisation d'accidents, discuté avec le CETE.

Un objet support final est généré à partir de chaque objet topographique ou caractéristique qui joue un rôle de support pour un objet thématique. Les modifications de relations telles que définies en B.4.2 ne sont pas nombreuses dans notre cas d'application réel parce que le niveau de détail n'est pas très différent entre l'état initial et l'état final.

A noter que l'importance de la relation finale attendue fait partie des éléments à calculer (nous avons discuté ce calcul en B.3.2). Dans nos implémentations, nous avons choisi une seule valeur d'importance (0.5) pour toutes les relations initiales et finales sauf la relation *porté par*, à laquelle nous avons affecté la valeur d'importance 1. Dans le cas général, il faut que les importances traduisent les spécifications du résultat voulu.

Dans la suite, nous allons détailler quelques règles de transformation des relations initiales en relations finales attendues.

Modification de la relation porté par <objet ponctuel, route>

Si la route porteuse est appariée, une nouvelle relation *porté par* est créé avec la route correspondante. Si la route porteuse n'est pas appariée, aucune relation correspondante n'est créée (lien de type 1-0).

Modification de la relation topologique <objet ponctuel, route porteuse>

Cette modification à notre cas expérimental est similaire à la modification de la relation *porté par*. Ceci vient de la différence de niveau de détail entre l'état initial et l'état final n'est pas assez grande. Le remplacement de l'objet support avec l'objet correspondant est le seul changement. Lorsque le changement de niveau de détail est important (par exemple, de 1 :25 000 vers 100 000), la modification peut inclure un changement de prédicat pour ne plus avoir de prédicats flous. Ce changement est lié au choix de l'utilisateur. Par exemple, une relation floue entre touche et disjoint peut être transformée en relation touche en cas de diminution de niveau de détail.

Modifications des relations de distance curviligne <objet ponctuel, point caractéristique> et <objet ponctuel, début de la route porteuse >

Si l'objet thématique final dispose d'une relation attendue de type *porté par* avec une route, et si l'objet caractéristique initial est apparié, une relation finale attendue du même type est créée avec la même distance curviligne. Si l'objet thématique final n'a pas de route porteuse ou si l'objet caractéristique n'est pas apparié, cette relation initiale sera perdue.

Modification de la relation de proximité point-bâtiment

Si le bâtiment est apparié, la distance reste la même et l'objet support final est l'homologue de l'objet support initial (le bâtiment).

Modification de la relation de proximité entre un objet thématique ponctuel sur la route et un rond-point

Si le rond-point est apparié avec un autre rond-point, une relation finale attendue du même type sera créée en mettant la même distance.

Si le rond-point est apparié avec un carrefour comme dans le premier exemple de la Figure 96, la relation est transformée en relation point-point sur la route avec une distance égale à (la distance initiale+la distance de la fin de la route jusqu'au centroïde du rond-point initial). Par contre, si l'objet thématique initial est très proche du rond point, la relation finale attendue sera une relation *porté par* avec le carrefour. Nous avons fixé un seuil de 50 mètres pour tout nos cas d'étude, au-delà de ce seuil, l'objet thématique est considéré non-proche. Pour d'autres cas d'étude et d'autres spécifications de l'état final, il faut choisir un seuil différent. Si le rond-point n'est pas apparié, aucune relation finale attendue ne sera générée.

Modification de la relations de direction <objet ponctuel, bâtiment>

Si le bâtiment est apparié, les attributs restent les mêmes et l'objet support est remplacé par le bâtiment support final correspondant. Dans le cas général, une règle peut être définie pour supprimer ou modifier la relation si le bâtiment est agrégé avec un autre.

Modification de la relation de forme de la route autour de l'objet thématique ponctuel

La relation attendue possède les mêmes attributs, mais l'objet de support est la route finale issue de l'appariement.

C.3.8 Relocalisation basée sur les relations attendues (étape e)

Après l'identification des relations finales attendues, nous cherchons à relocaliser les objets thématiques ponctuels en respectant le plus possible les relations attendues. Cette étape a été expliquée en (B.5, p.87). Pour chaque objet thématique ponctuel, la relocalisation inclut trois phases :

- Discrétisation de l'espace possible pour obtenir des positions possibles dans l'état final,
- Calcul de la satisfaction des relations pour chaque position possible,
- Choix de la meilleure position en utilisant le système multicritère PROMOTHEE.

Discrétisation de l'espace possible

Dans nos cas d'étude, les données thématiques ponctuelles sont toujours portées par des routes. Pour un objet thématique ponctuel, la route finale porteuse de l'objet thématique est identifiée en retrouvant la relation *porté par* au sein des relations finales attendues.

La route porteuse de l'objet thématique est ensuite discrétisée. Pour tous nos cas d'études, nous avons pris une équidistance d'un mètre entre deux positions possibles consécutives. Parmi les points résultant de la discrétisation, les points qui sont dans un buffer de 200 mètres autour la position initiale sont retenus pour être des positions possibles pour notre objet thématique dans l'état final. Dans le cas général, l'équidistance et la taille de buffer sont choisies par rapport aux spécifications, mais nous les avons choisies empiriquement et codées en dur pour nos cas d'étude.

La définition des mesures de satisfaction

Le but de cette étape est de donner une mesure de satisfaction de chaque relation. Nous avons choisi de donner une valeur entre 0 et 1 pour chaque évaluation. Pour les relations binaires, la valeur d'évaluation est 0 ou 1, pour les autres relations (comme les relations de proximité), les valeurs varient entre 0 et 1.

Nous avons défini une méthode (*getNode*) pour chaque type de relation. Cette méthode prend en entrée la position possible et elle donne en sortie une valeur entre 0 et 1. Pour les relations binaires, la sortie de cette fonction est soit 0 soit 1. Les méthodes de mesure de satisfaction proposées ont été testées pour nos cas d'étude. Chaque méthode commence par le calcul des attributs de la relation entre l'objet candidat et l'objet support final, puis, la satisfaction est calculée en se basant sur les différences entre la valeur des attributs calculés et celle des attributs de la relation attendue. Les seuils dans ces fonctions sont fixés manuellement par nous, empiriquement. Dans le cas général, ces seuils doivent être définis en prenant en compte l'échelle initiale et finale et les préférences de l'utilisateur.

Les mesures de satisfaction des différentes relations utilisées dans notre implémentation sont les suivantes.

Satisfaction de la relation de distance curviligne :

La satisfaction est calculée par comparaison entre la distance attendue et la distance curviligne du candidat. La satisfaction augmente si la différence diminue. Si la différence entre la distance curviligne attendue et la distance curviligne du candidat est supérieure à un seuil fixé, nous donnons une satisfaction s_r égale à 0. Nous avons fixé ce seuil empiriquement ($seuil_r=50m$) pour tout nos cas.

$$|d_{candidat} - d_{attendue}| > seuil_r \Rightarrow s_r(candidat) = 0$$

$$|d_{candidat} - d_{attendue}| \leq seuil_r \Rightarrow s_r(candidat) = 1 - \left(\frac{|d_{candidat} - d_{attendue}|}{seuil_r} \right)$$

La satisfaction peut être représentée par cette courbe (Figure 114)

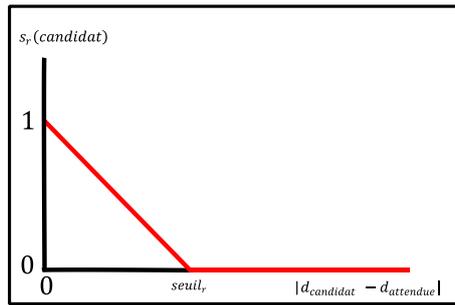


Figure 114. Satisfaction de la relation de distance curviligne

Satisfaction de la relation de proximité <objet thématique ponctuel, bâtiment> :

La satisfaction est calculée en mesurant la différence entre la distance (bâtiment, candidat), notée $d_{candidat}$, et la distance attendue. La satisfaction augmente si la différence diminue. Si la différence est au dessous d'un seuil $seuilMin_r$, nous considérons que la relation est totalement satisfaite, et si la différence est supérieur à un seuil $seuilMax_r$, la relation est jugée totalement non satisfaite pour le candidat. Entre $seuilMin_r$ et $seuilMax_r$, la satisfaction varie (Figure 115). Nous avons fixé ces seuils manuellement pour nos cas : $seuilMin_r = 10$ mètres et $seuilMax_r = 50$ mètres.

Si $|d_{candidat} - d_{attendue}| < seuilMin_r \Rightarrow$

$$s_r(candidat) = 1$$

Si $|d_{candidat} - d_{attendue}| > seuilMax_r \Rightarrow$

$$s_r(candidat) = 0$$

Sinon

$$s_r(candidat) = \left(\frac{seuilMax_r - |d_{candidat} - d_{attendue}|}{seuilMax_r - seuilMin_r} \right)$$

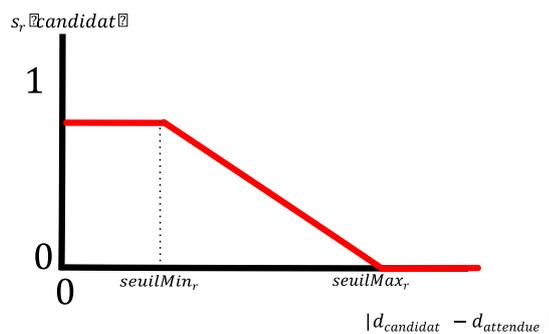


Figure 115. Satisfaction de la relation de proximité <objet thématique ponctuel, bâtiment>

Satisfaction de la relation de forme de la route <objet thématique ponctuel, route> :

La relation de forme de la route a été présentée dans ce chapitre en page 128, cette forme est modélisée par des attributs. Nous illustrons ces attributs en Figure 116 (reproduite de la Figure 107).

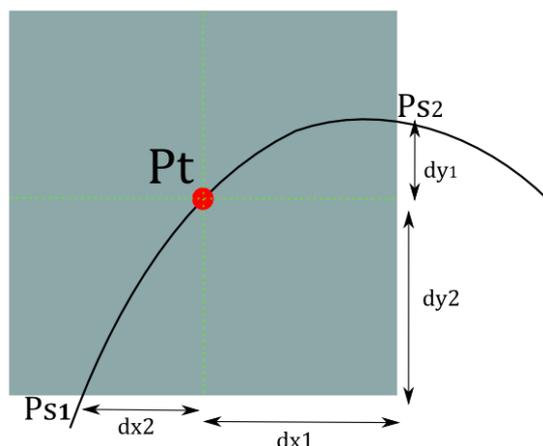


Figure 116. La forme de la route autour de l'objet thématique ponctuel (reproduite de la Figure 107)

La mesure de satisfaction cherche à comparer la forme de la route voulue avec la forme de la route d'une position possible. La Figure 117 montre toutes les positions possibles pour notre objet thématique, grâce à la mesure de satisfaction, nous pouvons choisir la position qui ressemble au plus à la forme voulue qui est la même que la forme initiale dans cet exemple.

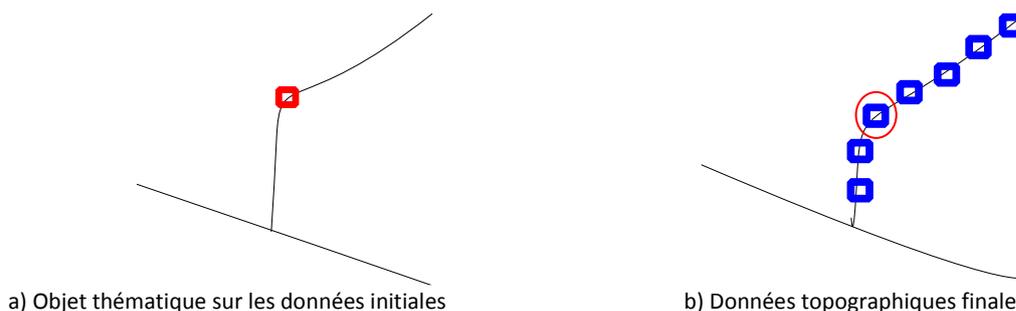


Figure 117. Exemple d'un objet thématique qui se trouve à la sortie du virage

Pour comparer la forme de la route autour du candidat avec la forme attendue, il faut d'abord calculer la relation de forme pour le candidat, on suppose que

- le positionnement initial Pt et les attributs de la relation attendue sont $(dx1, dx2, dy1, dy2, anglePs1Ps2)$
- le positionnement final Pt' et les attributs de la relation pour le candidat sont $(dx1', dx2', dy1', dy2', anglePs1'Ps2')$

Notre but est d'avoir une valeur qui représente la satisfaction de la relation en comparant ces valeurs, nous allons mesurer deux différences : différence de direction, et variation des attributs.

$$diffDeDirection = difference(anglePs1Ps2, anglePs1'Ps2') \quad \text{-en degrés-}$$

$$Variations = |dx1 - dx1'| + |dx2 - dx2'| + |dx3 - dx3'| + |dx4 - dx4'| \quad \text{-en mètres-}$$

A partir de ces valeurs, nous calculons les notes de différences (*noteDiffDeDirection*, *noteVariation*) par les règles suivantes:

$$\text{Si } (diffDeDirection < seuilMaxDD) \quad noteDiffDeDirection = 1 - \frac{diffDeDirection}{seuilMaxDD}$$

$$\text{Sinon} \quad noteDiffDeDirection = 0$$

$$\text{Si } (variations < seuilMaxVar) \quad noteVariations = 1 - \frac{variations}{seuilMaxVar}$$

$$\text{Sinon} \quad noteVariations = 0$$

$$Satisfaction = \frac{noteDiffDeDirection + noteVariations * 3}{4}$$

La Figure 116 montre comment cette relation peut aider à trouver la position correspondant dans l'état final seulement par la prise en compte de la relation de forme.

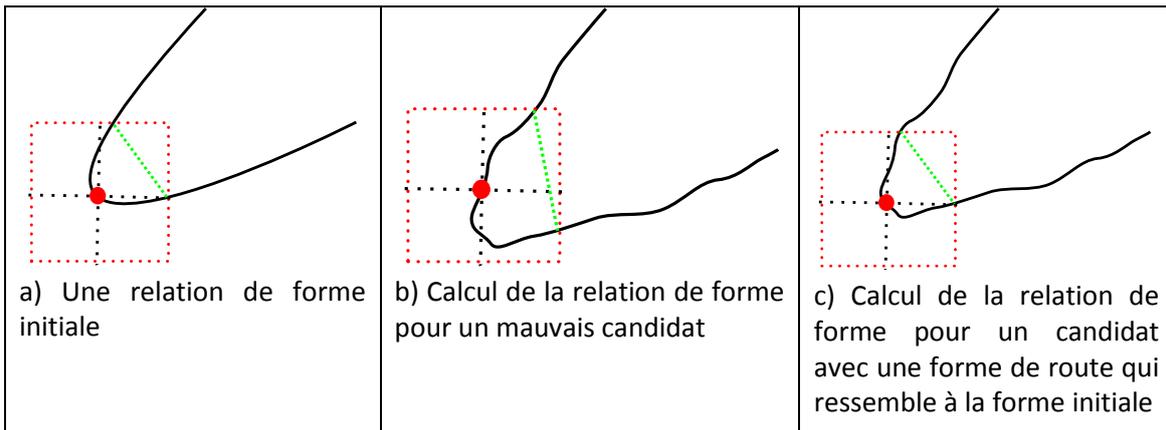


Figure 118. Satisfaction de la relation de forme de la route pour différentes positions candidates

Les valeurs *seuilMaxVar* et *seuilMaxDD* sont choisies empiriquement pour nos cas d'étude.

Satisfaction de la relation de proximité le long de la route <objet thématique ponctuel, objet support ponctuel sur la route>

Pour cette relation, nous avons choisi un seuil ($seuil_r = 30m$) pour la différence entre la distance attendue et la distance du candidat. Au-delà de ce seuil, le relation n'est pas satisfaite du tout, sinon, la satisfaction varie.

$$\text{Si } |d_{candidat} - d_{attendue}| > seuil_r \Rightarrow$$

$$s_r(candidat) = 0$$

Sinon

$$s_r(candidat) = 1 - \left(\frac{|d_{candidat} - d_{attendue}|}{seuil_r} \right)$$

Satisfaction de la relation de proximité <objet thématique ponctuel, rond-point>

Pour calculer la satisfaction de la relation de proximité avec le rond-point, nous comparons les distances jusqu'à la fin de la route, parce que la distance entre le bout de la route et le centre du rond point est la même dans la relation attendue et la relation du candidat. Si d est la distance jusqu'à l'extrémité de la route (sur le rond-point) :

Si $|d_{candidat} - d_{attendue}| > seuil_r \Rightarrow$

$$s_r(candidat) = 0$$

Sinon $s_r(candidat) = 1 - \left(\frac{|d_{candidat} - d_{attendue}|}{seuil_r} \right)$

Pour cette relation, nous avons choisi un $seuil_r = 50m$.

Satisfaction de la relation de direction <objet thématique ponctuel, bâtiment>

La relation de direction <objet thématique ponctuel, bâtiment> est décrite par l'orientation absolue modulo 2π du segment joignant le centroïde du bâtiment à l'objet thématique ponctuel (cf. page 128). La satisfaction de cette relation est mesurée en comparant l'orientation mesurée pour le point candidat ($angle_{candidat}$), et l'orientation attendue ($angle_{attendu}$). Plus précisément, nous calculons l'écart angulaire entre les deux orientations, modulo 2π , que nous ramenons entre $-\pi$ et $+\pi$. Comme le signe de cet écart angulaire n'a pas de signification particulière pour nous, nous raisonnons sur sa valeur absolue (définie entre 0 et $+\pi$), notée $écartAngulaireCandidatAttendu$. La satisfaction vaut 1 lorsque l'écart angulaire est nul, et décroît linéairement lorsqu'il augmente jusqu'à valoir 0 pour un seuil $seuil_{angle}$.

Si $écartAngulaireCandidatAttendu > seuil_{angle}$ alors

$$s_r(candidat) = 0$$

Sinon $s_r(candidat) = 1 - \frac{écartAngulaireCandidatAttendu}{seuil_{angle}}$

Nous avons choisi un seuil égal à 25 degrés, soit $seuil_{angle} = \frac{25\pi}{180}$ en radian.

Choix multicritère de la meilleure position

Dans cette phase, nous avons les positions possibles pour l'objet thématique final, et chaque position possible possède ses valeurs de satisfaction pour les relations attendues. Le but est de trouver la meilleure position en se basant sur les valeurs de satisfaction.

Nous avons une implémentation de la méthode PROMOTHEE sous licence GPL du Projet *gamma-platform*. Nous avons détaillé cette méthode en (B.5.3, p.91). Pour exécuter la méthode multicritère, il faut identifier les critères et leur poids. Pour nous, les critères sont les relations finales attendues. Le poids représente l'importance de la relation. Nous assignons les valeurs d'importance des relations aux poids. Ensuite, des candidats sont créés à partir des positions possibles issues de la discrétisation. Chaque candidat représente une position possible, un candidat possède des valeurs de satisfaction pour chacun des critères. Dans notre cas, les valeurs sont les mesures de satisfaction des relations. Après l'ajout des candidats avec les

valeurs de satisfaction, la méthode multicritère ordonne les différentes positions pour maximiser la satisfaction des relations.

	Relation1 (<i>Importance = Importance_{Relation1}</i>)	Relation2 (<i>Importance = Importance_{Relation2}</i>)
Candidat1	s_{11}	s_{12}
Candidat2	s_{21}	s_{22}
Candidat3	s_{31}	s_{32}

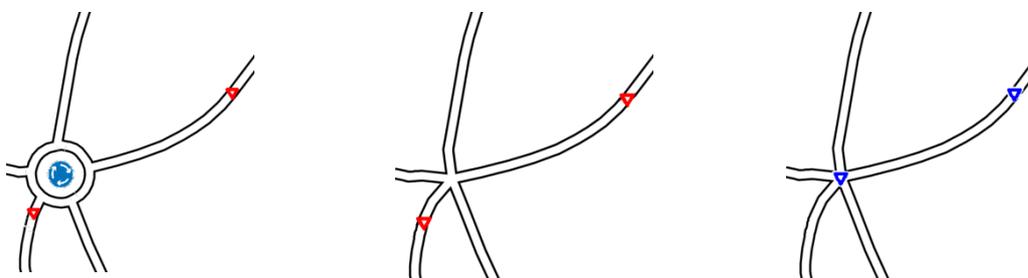
Tableau 4. Tableau des valeurs de satisfaction des candidats pour un objet thématique

C.4 Analyse des résultats obtenus

C.4.1 Résultats sur le cas d'étude jouets

Exemple 1

Avec le premier exemple, nous illustrons le fait que les relations finales attendues ne doivent pas toujours être identiques aux relations initiales. Nous avons décrit cet exemple en B.4.4, page 86). Pour l'accident qui est proche du rond point (Figure 119a), la règle de transformation de la relation de proximité accident-le rond-point fait que la relation finale attendue est la relation *porté par* avec le carrefour homologue du rond-point (Figure 119c). La Figure 119b illustre le résultat en utilisant la méthode actuelle de la distance proportionnelle.



a) Combinaison des accidents avec la base topographique initiale

b) Combinaison finale après une migration par distance proportionnelle

c) Combinaison finale en utilisant notre méthode de migration

Figure 119. Résultat de migration de données thématiques pour le cas d'un accident proche du rond-point

Exemple 2

Dans le deuxième exemple (Figure 120), nous voulons montrer l'effet de l'utilisation d'une méthode multicritère pendant la relocalisation. En Figure 120b, l'accident est relocalisé en calculant la moyenne des mesures de satisfaction des relations, puis en choisissant la position qui possède la plus grande moyenne. Le résultat obtenu avec l'utilisation de la méthode

multicritère (Figure 120c) est meilleur, parce qu'il considère les relations de l'accident avec la route, avec le bâtiment et avec le point.

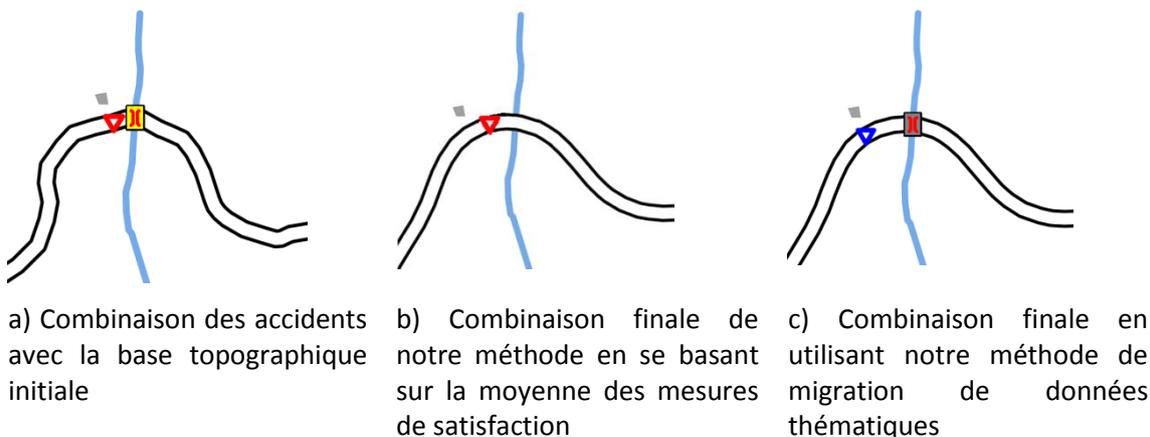


Figure 120. Résultat de migration de données thématiques et le comparer avec la migration de données thématiques sans l'utilisation une méthode multicritère

Exemple 3

Dans le troisième exemple, nous avons appliqué notre approche en nous basant sur une seule relation qui est la forme de la route autour de l'accident. Cet exemple montre la capacité à retrouver la position correspondante.

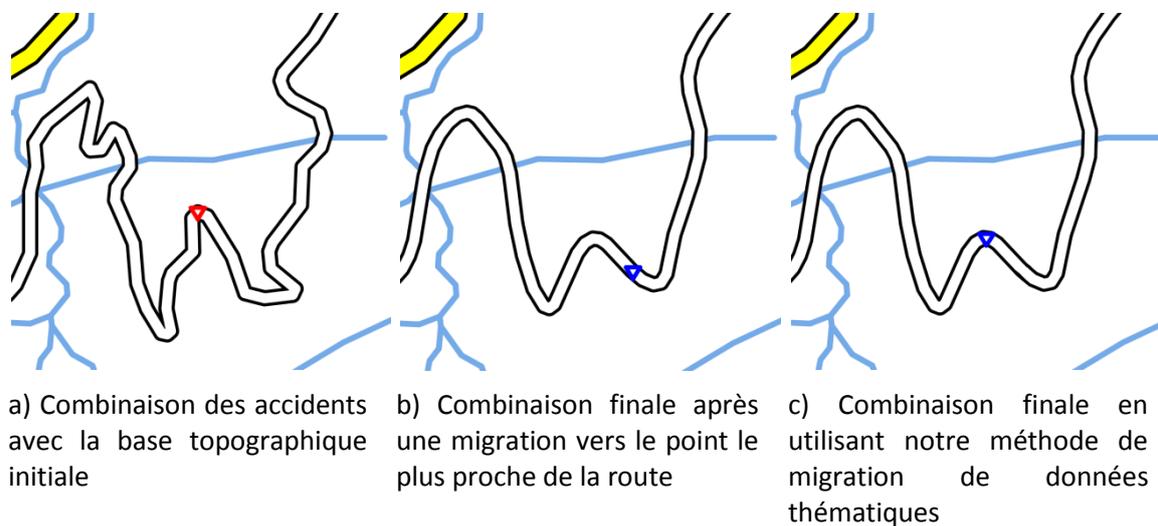


Figure 121. Résultat de migration de données thématiques en prenant en compte la relation de forme de la route

C.4.2 Résultats sur le cas d'étude réel

Nous allons montrer quelques extraits intéressants des résultats obtenus sur notre cas d'étude réel.

Exemple 4

En Figure 122, nous exposons le résultat que nous avons obtenu en appliquant notre méthode. La figure illustre comment les PLO ont la même position relative par rapport au sommet de virage. Ce résultat a été jugé par notre interlocuteur du CETE⁴ meilleur que le résultat détenu par décalage proportionnel.

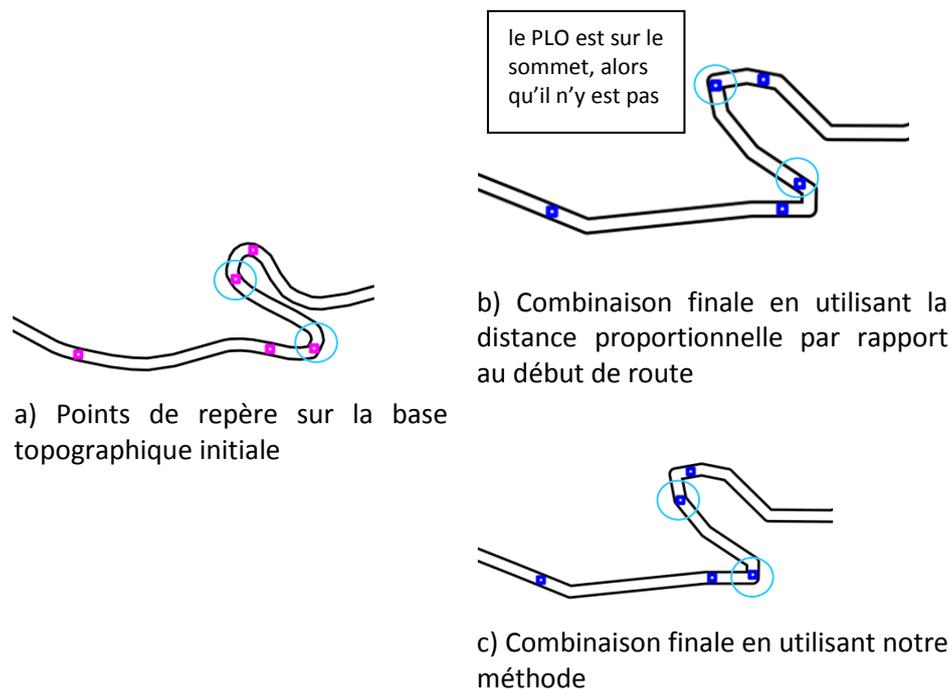
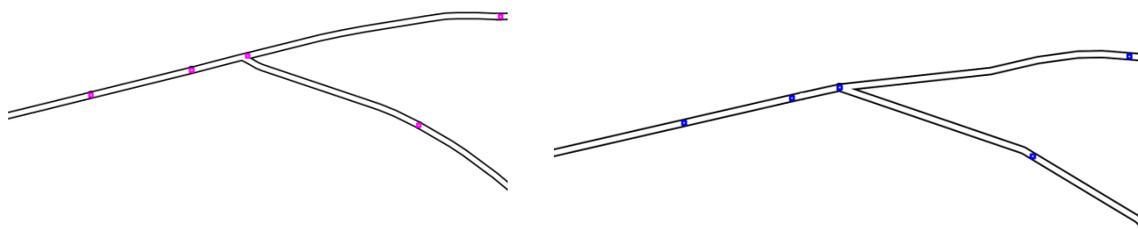


Figure 122. Résultat de l'application de notre méthode de la migration de données si la base topographique est remplacée

⁴ Nous remercions Yves Bonin, du CETE Méditerranée, pour le temps qu'il a consacré au suivi de notre travail.

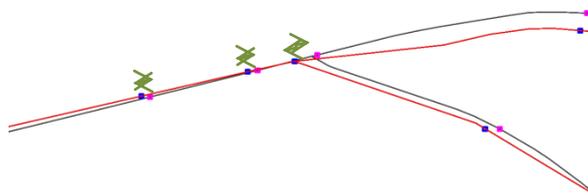
Exemple 5

En Figure 123, nous montrons un exemple de la migration de données dans le cas d'un carrefour apparié avec un carrefour. Le PLO du carrefour a été relocalisé sur le carrefour. La position du PLO initial n'était pas exactement sur le carrefour. Sur la base finale, notre algorithme relocalise le PLO sur précisément sur le carrefour. Ensuite, les autres PLO ont été relocalisés en conservant leur distance du carrefour.



a) combinaison initiale des PLO sur la base topographique initiale

b) résultat après la migration des données ponctuels vers la base finale

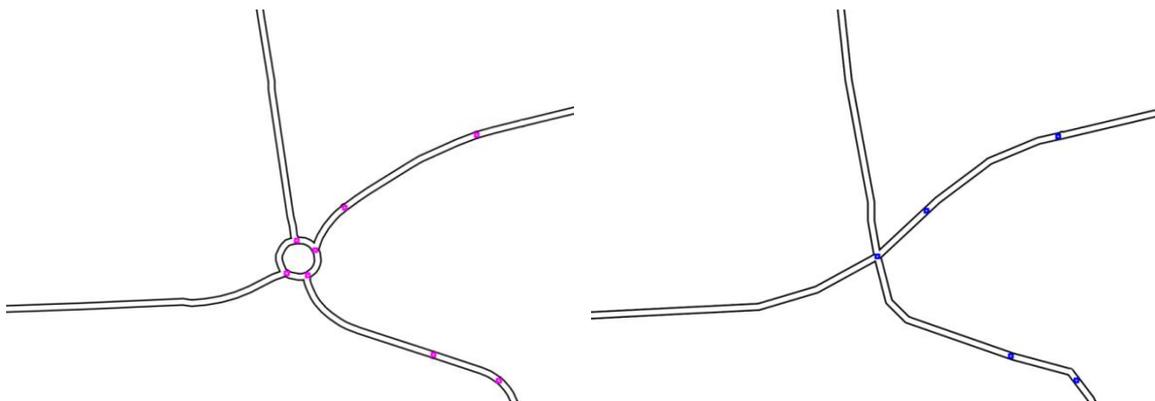


c) La superposition des données de l'état initial avec les données de l'état final

Figure 123. La migration de données thématique dans le cas d'un carrefour initial apparié avec un carrefour

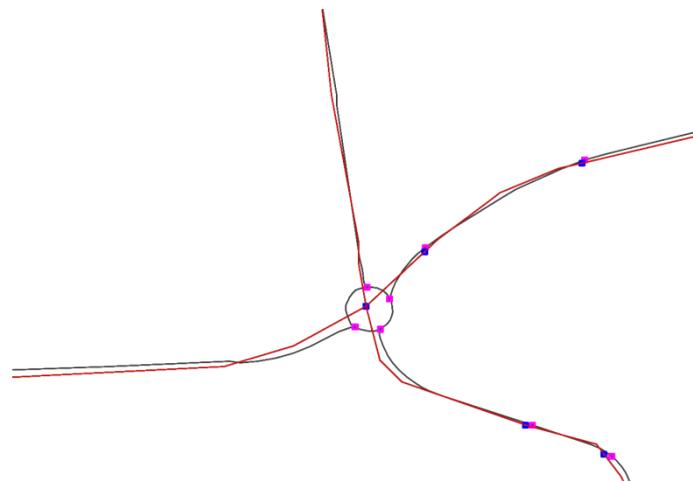
Exemple 6

En Figure 124, nous montrons un exemple de notre migration de données dans le cas où l'on a un rond point apparié avec un carrefour. Les objets thématiques qui sont sur les ronds-points ont été relocalisés sur le carrefour et les autres sont relocalisés proches de leur localisation initiale.



a) combinaison initiale des PLO sur la base topographique initiale

b) résultat après la migration des données ponctuelles vers la base finale

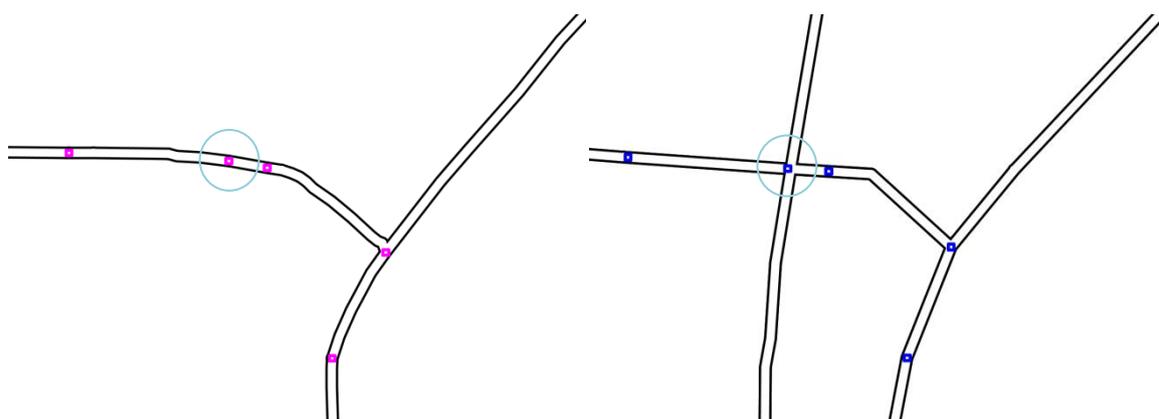


c) Superposition des données de l'état initial avec les données de l'état final

Figure 124. La migration de données thématique dans le cas d'un rond point initial apparié avec un carrefour

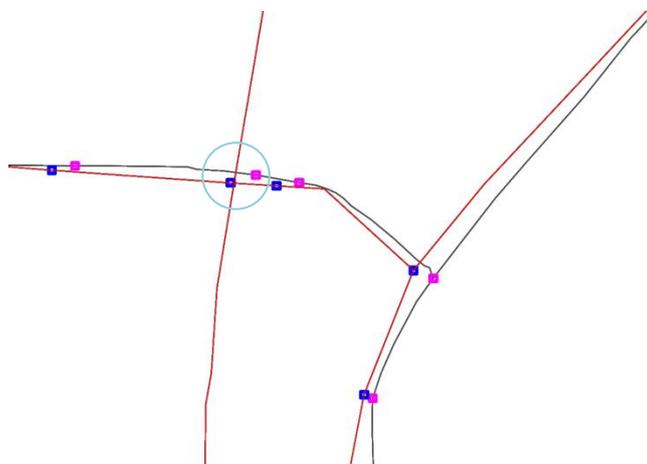
Exemple 7

En Figure 125, nous montrons un exemple d'une relocalisation d'un PLO en se basant sur l'information sémantique sur la localisation stockée dans un attribut (appelé TYPEPLO). Le PLO entouré sur la Figure 125a a un attribut indiquant qu'il s'agit d'un carrefour (ce carrefour n'apparaît pas dans la base de données initiale à cause de la sélection effectuée sur le réseau routier).



a) combinaison initiale des PLO sur la base topographique initiale

b) résultat après la migration des données ponctuelles vers la base finale



c) Superposition des données de l'état initial avec les données de l'état final

Figure 125. Exemple d'une relocalisation des PLO en utilisant l'information sémantique

Exemple 8

Nous avons constaté des cas discutables (en Figure 126). Le rond point n'a pas été reconnu, donc les connexions avec le rond point ont été appariées avec le carrefour. Il en résulte un grand décalage par rapport à l'état initial. Notre appariement a construit des liens d'appariement entre les entrées sorties du rond-point et le carrefour de la base finale.

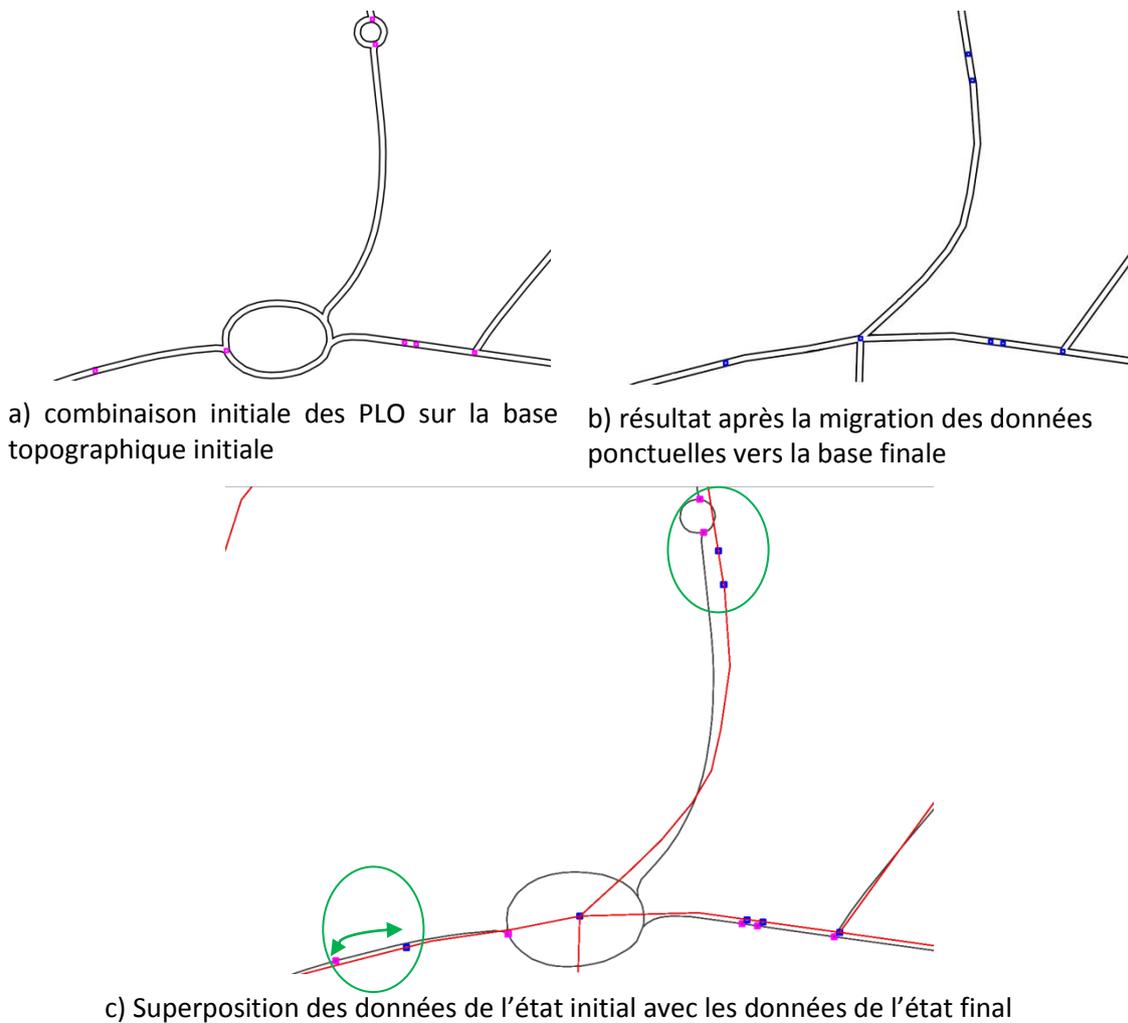


Figure 126. Exemple du cas d'un grand rond-point

Les résultats ont été discutés avec le CETE, et cette solution a été considérée comme bonne pour la problématique de la migration des données métier.

C.4.3 Bilan du chapitre C

Après la présentation de notre approche au chapitre B, nous avons détaillé dans ce chapitre la mise en œuvre pour le cas spécifique de la migration de données ponctuelles sur réseau. La plus grande difficulté rencontrée était d'obtenir un bon résultat d'appariement. Nous avons défini un processus d'appariement qui répond à nos besoins tout en utilisant les informations sémantique des routes. Les résultats obtenus nous ont permis de valider notre approche, en évaluant les résultats visuellement et par la discussion avec le CETE qui a fourni le cas d'étude réel. D'autres méthodes d'appariement peuvent être utilisées. Si la cardinalité des liens résultant de l'appariement n'est pas uniquement 1-1 ou N-1, il convient de modifier les règles d'identification des relations finales attendues pour définir les transformations qui prennent en compte tous les types de liens d'appariement.

Le paramétrage de l'implémentation a été fait manuellement et les valeurs des paramètres ont été choisies de manière empirique pour répondre au besoin de l'utilisateur tel que communiqué par le CETE. Certains résultats sont discutables et liés à notre méthode d'appariement et au paramétrage. Alors il faut tester l'effet de la méthode d'appariement et du paramétrage sur les résultats.

Nous avons évalué le résultat, soit par une comparaison avec la méthode actuelle, soit par une évaluation visuelle. D'après notre interlocuteur du CETE, les résultats sont jugés acceptables.

Conclusion et perspectives

Conclusion et perspectives

Rappel de l'objectif

Avec l'accessibilité à un public de plus en plus large des données géographiques produites par des fournisseurs nationaux ou autres, les utilisateurs ont tendance à intégrer des données de différentes sources. Certains utilisateurs saisissent leurs propres données, que nous appelons thématiques, en utilisant un fond topographique comme référence. Dans le jeu de données composé de données topographiques et thématiques ainsi créé, les données topographiques servent de support aux données thématiques. Elles permettent de localiser les données thématiques, mais aussi de les interpréter car les relations entre données thématiques et topographiques ont du sens (par exemple un accident routier sera peut-être analysé différemment selon qu'il ait lieu sur une route droite, juste avant un pont ou dans un virage). Or, les données topographiques de référence sont toujours susceptible d'être remplacées par d'autres, par exemple suite à la disponibilité d'une base topographique plus précise, pour pouvoir intégrer les données thématiques avec d'autres qui n'utilisent pas forcément la même base de référence, ou pour la visualisation à un niveau de détail moindre. Il faut que les données thématiques soient transférables automatiquement vers une nouvelle base, car le transfert manuel ou la re-saisie des données est très coûteux. Lors d'un changement de référence - pour des besoins d'intégration ou pour obtenir une version généralisée de la combinaison initiale - des incohérences peuvent être provoquées par la non-prise en compte des relations initiales entre les données thématiques de l'utilisateur et les données initiales de référence.

L'objectif de la thèse était de proposer une méthodologie pour prendre en compte les relations entre les données thématiques et les données topographiques lors d'un changement de niveau de détail, en nous concentrant sur le cas où les données de référence sont remplacées par d'autres et où il faut faire *suivre* les données thématiques, ce que nous avons appelé « migration des données thématiques ».

Bilan des contributions

Modélisation des relations entre données thématiques et données topographiques

Nous avons proposé un modèle pour la description des relations de référencement entre données thématiques et données topographiques. Le modèle inclut les éléments suivants :

- la notion d'objets caractéristiques,
- un modèle générique permettant de représenter les relations entre un objet thématique et un objet dit *support* qui peut être topographique ou caractéristique,
- la relations spéciale *porté par*,
- la spécialisation de ce modèle générique pour quelques relations valables pour des objets thématiques ponctuels sur réseau,
- la notion *d'importance* d'une relation qui traduit les préférences de l'utilisateur (lorsqu'il s'agit des relations finales attendues) ou des informations fournies par l'utilisateur sur le mode de saisie (lorsqu'il s'agit des relations initiales),
- des méthodes d'instanciation des relations.

- Une représentation des relations floues. Une relation est floue si elle n'est pas complètement vérifiée, par exemple, la relation « presque parallèle » et la relation « entre disjoint et touche ».

Méthode pour la migration de données thématiques

Nous avons proposé un processus pour la migration de données thématiques, plus spécifiquement la migration de données ponctuelles sur un réseau.

Cette méthode consiste à instancier les relations initiales entre données topographiques et thématiques, à apparier les données topographiques initiales et finales, puis à identifier les relations finales attendues en fonction des relations initiales, de l'appariement et de règles de transformation tenant compte de préférences utilisateur et du changement de niveau de détail entre les états initial et final. Nous avons notamment proposé que les relations floues soient transformées en relations nettes selon la perceptibilité à l'état final – par exemple, une relation floue « presque sur la ligne » est transformée en une relation « loin de la ligne » ou en une relation « sur la ligne » selon la différence de niveau de détail entre l'état initial et le l'état final. Une fois les relations finales attendues identifiées, les données thématiques sont relocalisées en utilisant un système multicritère, qui permet de proposer une solution même en cas d'impossibilité de satisfaire toutes les relations.

La méthode proposée contient plusieurs briques qui peuvent être adaptées pour la migration d'autres types de données thématiques (linéaires et surfaciques), ou pour prendre en compte les relations dans d'autres types de traitements (généralisation, carte à la carte). Nous avons donné en fin de chapitre B des pistes pour étendre notre modèle de migration à d'autres types de traitement.

Notre approche pour la migration de données ponctuelles sur réseau a été testée par l'application sur un cas d'étude réel fourni par le CETE Méditerranée. Les résultats ont été discutés avec le CETE et ils sont jugés satisfaisants.

Une méthode d'appariement partiel de réseau munis de points caractéristiques

Lors de la mise en œuvre, nous avons rencontré certaines difficultés techniques principalement liées à l'appariement. Nous avons proposé de les résoudre par la construction d'un graphe intermédiaire dont les nœuds sont les objets caractéristiques du réseau routier (carrefours, sommets de virages, ronds-points), puis à partir de ce graphe, d'objets de plus haut niveau que les tronçons de routes basés sur les bonnes continuités (les strokes).

Perspectives

Faciliter le paramétrage

Pendant la mise en œuvre du cas d'étude réel, le processus de migration de données ponctuelles est paramétré empiriquement par rapport aux préférences qui ont été discutées avec le CETE. Pour que le paramétrage soit mieux spécifié, il faudrait d'abord tester l'approche sur des cas plus variés avec une documentation des besoins et des paramètres utilisés, pour pouvoir ensuite fournir à l'utilisateur une assistance au paramétrage.

Comparer l'appariement avec d'autres

Nous avons utilisé une seule méthode d'appariement ad hoc fournissant des liens de type N-1 et 1-1. D'autres méthodes d'appariement peuvent être utilisées, et les règles pourraient être modifiées pour accepter tous les types de liens d'appariement. De plus, il faudrait comparer l'appariement ad hoc réalisé avec des méthodes existantes d'appariement partiel de lignes, par exemple celui de (Devogele 2000).

Représenter les relations entre les données thématiques

Notre modèle de description des relations concerne uniquement les relations entre les objets thématiques et les objets supports, mais dans certaines applications, il peut y avoir besoin de prendre en compte des relations entre les objets thématiques – par exemple les inter-distances relatives entre les accidents sur une route. Pour cela, il serait souhaitable d'enrichir notre modèle de description des relations en ajoutant plus d'objets caractéristiques et plus de relations.

Appliquer l'approche pour la migration des données linéaires et surfaciques

Nous pensons que notre approche pour la migration de données ponctuelles sur réseau peut être adaptée pour être appliquée à la migration de données thématiques de type linéaire et surfacique, par exemple la migration des pistes cyclable qui sont saisie sur la BD TOPO vers la BD CARTO. Nous avons donné des pistes pour cela en fin de chapitre B, notamment l'idée de se ramener en partie à la migration de données ponctuelles en identifiant des points de convergence et divergence entre les objets thématiques linéaires et le réseau qui les porte. Pour adapter notre processus à d'autres cas de migration, les étapes sont les suivantes :

- étendre le modèle de description des relations pour représenter les relations qui concernent chaque cas d'application,
- définir les méthodes pour extraire les relations,
- définir des règles pour l'identification des relations finales attendues,
- définir des méthodes de satisfaction des relations,
- définir un processus de génération des données thématiques finales en prenant en compte les relations attendues.

Réutiliser des parties de notre approche pour le cas de la généralisation

Nous pensons que l'approche est utilisable pour intégrer la prise en compte des relations des données thématiques avec les données de référence lors de la généralisation automatique des données topographiques ou/et des données thématiques (par exemple, la généralisation de lignes de bus et leurs données de référence).

Nous avons donné en fin de chapitre B des pistes pour utiliser les relations thématique-topographique dans un processus de généralisation. Pour guider la généralisation des données topographiques par la présence des données thématiques, nous avons identifié deux cas : celui où le processus de généralisation peut être paramétré en fonction des relations, et celui où le paramétrage ne suffit pas et où il faut réécrire le processus. Par ailleurs, la thèse en cours d'Adrien Maudet étudie notamment la modélisation des relations d'hébergement d'un objet par un autre dans un processus de généralisation basé sur une approche multi-agents (Maudet et al. 2014). L'idée est d'adapter un modèle initialement dédié à la simulation

(Picault et al. 2010) qui permet de faire varier le comportement d'un agent en fonction de l'agent qui l'héberge. Un cas d'application est le cas de l'hébergement de données thématiques par des données topographiques : la relation d'hébergement correspond alors à notre relation *porté par*. Les relations que nous avons définies, les méthodes d'extraction et de calcul de satisfaction, pourraient être réutilisées pour contraindre la généralisation.

Améliorer la documentation des processus de saisie de données géographiques

L'importance des relations dépend énormément des éléments pris en compte pendant la saisie des données, par exemple des informations issues d'une fiche descriptive d'accident rédigée par la police immédiatement après un accident, qui peuvent comporter des éléments de référencement indirect. Pour pouvoir tenir compte des éléments de référencement indirect qui ont guidé la saisie, il serait souhaitable de documenter le processus de saisie en ajoutant des métadonnées de saisie aux objets. La documentation peut être aidée par des questions à l'utilisateur (Brando 2013). Pour les données dont la saisie fait l'objet d'un protocole basé sur une fiche descriptive remplie sur le terrain puis saisie en bureau, comme les accidents routiers, il est aussi possible d'inciter la personne qui remplit la fiche à utiliser le référencement indirect en prévoyant des questions dédiées dans la fiche. Il serait par exemple intéressant de savoir comme la police situe un accident par rapport aux objets qui l'entourent, et éventuellement ceux qui semblent avoir joué un rôle dans l'accident.

Pour finir

De plus en plus de jeux de données sont en réalité des agrégats de données de différentes sources. Lors de la constitution de ces agrégats par intégration, une certaine cohérence entre données est assurée. Mais les données ont des cycles de vie. Ainsi ces jeux sont transformés, simplifiés, mis à jours. Garder la cohérence des données est un vrai défi parce que cette cohérence est porteuse de sens. Dans notre thèse nous avons proposé une méthode à base de représentation de relations et de recalage multicritères pour maintenir au mieux ces relations. Nous avons testé notre proposition sur le cas de données ponctuelles 'accrochées a' des données linéaires. Ce cas est en fait assez fréquent. Toutefois il existe aussi des données linéaires accrochées à d'autres données linéaires comme les lignes de bus ou pistes cyclables qui se réfèrent à des routes. Il s'agit de relations qui ressemblent à la relation 'porté par' que nous avons proposé. Ainsi nous pensons que le modèle que nous proposons peut s'étendre à d'autres cas. Notre méthode de migration est donc un premier pas pour améliorer la qualité et la longévité des données thématiques.

Liste des publications

Communications avec actes dans un congrès international.

Jaara K. Duchêne C. & Ruas A. (2013). Preservation and Modification of Relations Between Thematic and Topographic Data Throughout Thematic Data Migration Process, *Cartography From Pole To Pole, Selected Contributions to the 26th International Conference of the ICA*, M. Buchroithner, N. Prechtel, D. Burghardt Ed., pp. 103--118, *Lecture Notes in Geoinformation and Cartography*, Springer.

Balley S. Jaara K & Regnaud N. (2012). Towards a prototype for deriving custom maps from multisource data, *Proceedings of 15th ICA workshop on generalisation and multiple representation*.

Jaara K. Duchêne C. & Ruas A. (2012) . A model for preserving the consistency between topographic and thematic layers throughout data migration, *15th International Symposium on Spatial Data Handling (SDH'2012)*.

Touya G. Balley S. Duchêne C. Jaara K. Regnaud N. & Gould N. (2012). Towards an Ontology of Generalisation Constraints and Spatial Relations, *Proceedings of 15th ICA workshop on generalisation and multiple representation*.

Jaara K. Duchêne C. & Ruas A. (2011). Toward the generalisation of cartographic mashups: Taking into account the dependency between the thematic data and the reference data throughout the process of automatic generalisation, *Proceedings of 15th ICA workshop on generalisation and multiple representation*

Articles dans des revues internationales ou nationales sans comité de lecture.

Jaara K. Duchêne C. & Ruas A. (2013). Préservation et modification des relations entre les données thématiques et les données topographiques lors de migration de données thématiques, *Cartes & Géomatique - Revue du Comité Français de Cartographie*, n. 218, pp. 39--50

Chapitres de d'ouvrage scientifique.

Touya G. Bucher B. Falquet G. Jaara K. and Steiniger S. (2014). Modelling Geographic Relationships in Automated Environments, *Abstracting Geographic Information in a Data Rich World*, chap. 3, pp. 53--82, Burghardt, Dirk and Duchêne, Cécile and Mackaness, William Ed., *Lecture Notes in Geoinformation and Cartography*, Springer, doi:10.1007/978-3-319-00203-3_3 

Liste des figures

Figure 1. Un mashup entre des données thématiques et des données topographiques.....	13
Figure 2. La résolution des données maillées: représentation d'une route et d'un bâtiment (Ruas 2004).....	13
Figure 3. Gammes d'échelles de saisie et de représentation de la BD TOPO et de la BD CARTO (Girres 2012).....	14
Figure 4. Différence de contenu au 1: 100 000 et 1: 250 000. Cartes IGN (Ruas 2004)	14
Figure 5. Saisie d'un chemin de randonnée sur une carte IGN fournie par le géoportail de l'IGN (www.geoportail.gouv.fr).....	16
Figure 6. Exemple d'une saisie manuelle avec peu de prudence sur un fond de carte web	16
Figure 7. Réseau routier avec des données thématiques professionnelles.....	17
Figure 8. Trace GPS sur des données IGN	18
Figure 9..Calcul d'itinéraire sur Google Maps (https://maps.google.fr/)	18
Figure 10. Un lieu nommé dans un index géographique (Hill, 2006).....	19
Figure 11. Exemples de relations spatiales	20
Figure 12. Problème de non existence de certains objets topographiques dans la base finale (source : www.géoportail.fr)	23
Figure 13. Remplacement d'une base par une autre plus détaillée, sans prendre en compte les relations initiales (données du géoportail de l'IGN site www.visorando.com)	23
Figure 14. Repositionnement de points de repères routier (source CETE)	24
Figure 15. Un remplacement non valide d'une base topographique par une autre à cause d'une perte de relation (données du géoportail de l'IGN site www.visorando.com).....	25
Figure 16. Migration de données thématiques lors d'un remplacement des données topographiques	26
Figure 17. Migration de données thématiques lors d'un remplacement de données topographiques, avec prise en compte des relations initiales.....	26
Figure 18. Migration de données thématiques suite à la généralisation de données topographiques, avec prise en compte des relations initiales.....	27
Figure 19. Prise en compte des relations initiales lors de la migration de données thématiques et la généralisation des données topographiques afin de changer le niveau de détail	27
Figure 20. Utilisations possibles des relations pendant le traitement de la combinaison des données thématiques avec des données topographiques lors d'un changement de niveau de détail.....	28
Figure 21. Taxonomie pour les relations spatiale entre deux entités géographiques (Touya et al. 2012)	31
Figure 22. Exemples de relations spatiales (Touya et al. 2012)	32
Figure 23. Relation floue et binaire.....	33
Figure 24. Correspondance de notre classification avec la taxonomie de (Touya et al. 2012) ..	34
Figure 25. Les relations topologiques entre deux régions selon le modèle 9IM (Egenhofer et Herring 1991)	35
Figure 26.Graphe conceptuel de voisinage des relations topologiques entre deux régions (Egenhofer et AlTaha 1992)	36
Figure 27. Extension du graphe de voisinage conceptuel pour représenter les relations floues (Touya et al. 2012)	37

Figure 28.Exemples de relations de proximité.....	37
Figure 29. Orientation relative d'un bâtiment par rapport à un objet linéaire (Duchêne 2004)	38
Figure 30. Relation cardinale d'orientation entre deux régions (Goyal et Egenhofer 2000).....	38
Figure 31. Raisonnement rectangle à minimum englobant pour décrire la relation de direction (Papadias et Theodoridis 1997).....	39
Figure 32.Relations de direction possibles en utilisant le Rectangle minimum englobant (Papadias et Theodoridis 1997).....	39
Figure 33. Mesures de divisions qui concernent les relations topologiques entre deux régions (Egenhofer et Dube 2009).....	41
Figure 34. Mesures de proximité entre deux régions (Egenhofer et Dube 2009).....	42
Figure 35.Contraintes relationnelles liées à la proximité des symboles (Duchêne 2004).....	43
Figure 36.Contraintes de cohérence géographique (Duchêne 2004).....	43
Figure 37. Modélisation des contraintes relationnelles (Duchêne 2004) (a) diagramme de classes (b) diagramme d'objets.....	44
Figure 38. Modélisation des contraintes relationnelles (Duchêne 2004).....	44
Figure 39. Résoudre la problématique des contraintes relationnelles multiples par l'agrégation (Duchêne 2004).....	45
Figure 40. Déformation des courbes de niveau comme un exemple de la prise en compte des relations entre des parties d'objets dans le modèle GAEL (Gaffuri 2008).....	45
Figure 41.Exemples de l'utilisation des critères relationnels dans un processus d'appariement (Olteanu 2008).....	46
Figure 42. Critères d'appariement d'un objet géographique (Olteanu 2008, p53).....	47
Figure 43. Modèle proposé par (Abadie 2012) pour l'intégration de données.....	48
Figure 44.Exemple des incohérences liées à l'intégration et à la généralisation (Gaffuri 2008)	49
Figure 45. Améliorer le résultat de l'intégration de données par déplacement basé sur les contraintes d'intégrité (Wallgrün 2013).....	49
Figure 46.Cas d'incohérences liées à la différence de niveaux de détail (Touya et Brando 2013) (la couleur gris sombre représente les zones urbaines).....	50
Figure 47. Exemple d'une incohérence sur OSM concernant un bâtiment chevauchant une route corrigée par la méthode de (Brando 2013).....	51
Figure 48. Notre exemple-jouet : migration d'accidents sur un nouveau réseau routier en utilisant uniquement l'abscisse curviligne (c).....	55
Figure 49. Migration de données thématiques suite au remplacement des données topographiques, avec prise en compte des relations initiales.....	56
Figure 50. Processus de migration de données thématiques proposé.....	58
Figure 51. Une partie d'une ontologie qui décrit les maisons de villes mitoyennes (terraced houses en anglais), l'ontologie est adaptée pour l'enrichissement des données. DTP : propriété du type de données. OP : propriété d'objet. (Lüscher et al. 2008).....	60
Figure 52. Extraction automatique d'un rond-point.....	61
Figure 53. Objets topographiques proches d'un accident routier.....	64
Figure 54.Les variables qui modélisent les préférences de l'utilisateur dans le modèle de (Balley et al. 2014).....	65
Figure 55. Modèle pour les relations spatiales, proposé par (Touya et al. 2012).....	66
Figure 56.Instanciation du modèle de (Touya et al. 2012) pour la relation de proximité pont-bâtiment.....	67

Figure 57. Extension de l'ontologie pour représenter les relations entre les données thématiques et les données topographiques	69
Figure 58. Fonctionnement de l'ensemble des éléments pour extraire et stocker les relations	70
Figure 59. Les différentes catégories de relations dans le processus de migration de données	70
Figure 60. Version simplifiée de notre modèle de référencement des données thématiques sur les données topographiques.....	71
Figure 61. Modélisation des objets supports.....	73
Figure 62. Représentation des relations topologiques.....	75
Figure 63. Orientation de la route basée sur sa géométrie.....	76
Figure 64. Représentation d'un accident qui s'est passé à proximité d'un pont.....	76
Figure 65. Représentation des relations de proximité dans le modèle de référencement.....	77
Figure 66. Représentation des relations de direction dans notre modèle.....	78
Figure 67. Diagramme d'objets, illustre la représentation de relations en utilisant notre modèle.....	80
Figure 68. Appariement des données topographiques.....	82
Figure 69. Exemple de création des strokes avec une continuation à travers des ronds-points (Luan et Yang 2010).....	83
Figure 70. Entrées/sorties de l'étape d'identification des relations finales attendues.....	84
Figure 71. Une partie du graphe conceptuel de voisinage des relations ligne-polygone (Egenhofer et Mark 1995).....	86
Figure 72. Le besoin pour la modification des relations.....	86
Figure 73. Discrétisation de la route porteuse de l'accident pour extraire les positions possibles.....	88
Figure 74. La différence en utilisant une méthode multicritère pour la relocalisation avec pondération plus forte de la relation de position relative par rapport au bâtiment.....	89
Figure 75. Mesure de satisfaction pour la relation <objet thématique ponctuel, Bâtiment>	90
Figure 76. Graphe de surclassement complet en utilisant la méthode Prométhée II.....	92
Figure 77. Données de référence initiales et finales avec la localisation initiale des accidents.	93
Figure 78. Extraction des objets caractéristiques initiaux.....	93
Figure 79. Extraction des objets caractéristiques finaux.....	94
Figure 80. Résultat de migration de données thématiques.....	96
Figure 81. Processus de migration de données thématiques (version générique).....	97
Figure 82. Une partie d'un itinéraire de vélo qui suit la route (Touya et al. 2012).....	98
Figure 83. Modèle de référencement en cas de la migration de données thématiques linéaires.....	99
Figure 84. Prise en compte des relations initiales lors de la migration de données thématiques et la généralisation des données topographiques afin de changer le niveau de détail.....	100
Figure 85. Modélisation des contraintes relationnelles (Duchêne 2004) (a) diagramme de classes (b) diagramme d'objets.....	102
Figure 86. Modélisation proposée des contraintes relationnelles thématique-topographique au sein du modèle CartACom de (Duchêne 2004) - diagramme de classes.....	102
Figure 87. Modélisation de liens de correspondance entre les objets supports initiaux et les objets supports finaux.....	103
Figure 88. Les étapes principales d'un processus de carte à la demande (Balley et Regnauld 2011).....	104

Figure 89. Modèle de spécifications de carte à la demande proposé par (Balley et al. 2012).	105
Figure 90. Vue d'ensemble du système pour la prise en compte des relations	108
Figure 91. Une borne sur la route	113
Figure 92. Extrait d'un recensement d'objets métier : glissières de sécurité du département du Var (Bonin 2010). (PRD, ABD) et (PRF, ABF) correspondent à la description de la position du point initial (respectivement final) de l'objet sous forme (numéro du PR, distance curviligne par rapport au PR).	113
Figure 93. Un schéma simplifié du RIU V1 pour montrer le lien avec le schéma de la BD CARTO	114
Figure 94. Exemple réel pour expliquer les erreurs lors de la migration de données thématiques à base de distance proportionnelle	116
Figure 95. Erreurs constatées lors de la migration de données en utilisant la distance proportionnelle	117
Figure 96. Des exemples-jouets sur lesquels nous avons testé notre approche pour la migration de données thématiques	119
Figure 97. Données initiales et données relocalisées avec la méthode proportionnelle	120
Figure 98. Le processus de migration de données thématiques	122
Figure 99. Principe de détection des points d'inflexion (Plazanet 1996, p.97).....	122
Figure 100. Détection des virages par la somme des angles	123
Figure 101. La somme des angles a pour but de détecter les virages	123
Figure 102. Des sommets de virage détectés en utilisant notre méthode de détection.....	124
Figure 103. Des PLO de type <i>carrefour</i> . En vert : les PLO et le réseau routier fourni par le Conseil Général du Vaucluse En rouge : réseau routier conforme au RIU V1 issu de la BD CARTO	126
Figure 104. Comment choisir le prédicat suivant la distance entre le point et la route.....	127
Figure 105. Calcul de la relation de proximité entre point thématique et un bâtiment	128
Figure 106. Exemple d'un objet thématique qui se trouve à la sortie du virage	129
Figure 107. Mesures pour représenter la forme de la route autour d'un point thématique... ..	130
Figure 108. Possibilité d'erreur due à une taille de fenêtre très grande	130
Figure 109. Un extrait des bases topographiques initiale et finale fournies par le CETE	131
Figure 110. La construction du graphe.....	132
Figure 111. Le besoin de dépasser les ronds-points pendant la construction des strokes	133
Figure 112. Cas spécial concernant l'appariement <rond-point, carrefour>.....	134
Figure 113. La construction du graphe.....	135
Figure 114. Satisfaction de la relation de distance curviligne.....	139
Figure 115. Satisfaction de la relation de proximité <objet thématique ponctuel, bâtiment>	139
Figure 116. La forme de la route autour l'objet thématique ponctuel (reproduite de la Figure 107)	140
Figure 117. Exemple d'un objet thématique qui se trouve à la sortie du virage	140
Figure 118. Satisfaction de la relation de forme de la route pour différentes positions candidates	141
Figure 119. Résultat de migration de données thématiques pour le cas d'un accident proche du rond-point	143
Figure 120. Résultat de migration de données thématiques et le comparer avec la migration de données thématiques sans l'utilisation une méthode multicritère.....	144

Figure 121. Résultat de migration de données thématiques en prenant en compte la relation de forme de la route	144
Figure 122. Résultat de l'application de notre méthode de la migration de données si la base topographique est remplacée.....	145
Figure 123. La migration de données thématique dans le cas d'un carrefour initial apparié avec un carrefour	146
Figure 124. La migration de données thématique dans le cas d'un rond point initial apparié avec un carrefour	147
Figure 125. Exemple d'une relocalisation des PLO en utilisant l'information sémantique	148
Figure 126. Exemple du cas d'un grand rond-point	149

Bibliographie

- Abadie N. (2012).** Formalisation, acquisition et mise en œuvre de connaissances pour l'intégration virtuelle de bases de données géographiques. *Thèse en SIG*, Université Paris-Est.
- Arthur G. Meulemans W. Reimer A. Haverkort H. & Speckmann B. (2013).** Topologically Safe Curved Schematisation. *The Cartographic Journal*, Volume 50, Number 3, pp.276-285(10) August 2013.
- BALLEY S. & REGNAULD N. (2011).** Models and standards for on-demand mapping. *proceeding of the international cartographic conference (ICC 2011)*.
- Balley S. Baella B. Cristophe S. Pla M. Regnaud N. & Stoter J. (2014).** Map specifications and user requirements. *Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation*, Burghardt D. Duchêne C. Mackaness W. (Eds), Lecture Notes in Geoinformation and Cartography, Springer, pp.17-52.
- Balley S. Jaara K. & Regnaud N. (2012).** Towards a prototype for deriving custom maps from multisource data. *15th ICA Workshop on Generalisation and Multiple Representation*, 13-14 September. Istanbul (Turkey).
- Bang Y. Ga C. & Yu K. (2010).** An iterative process for matching network data sets with different levels of detail. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-4/C7.
- Barillot X. & Plazanet C. (2002).** Analyse des formes des routes. *dans Ruas A., Généralisation et représentation multiple*, Hermes Lavoisier, Paris. Chapitre 11, pp.203-224.
- Blasby D. Davis M. Kim D. & Ramsey P. (2004).** GIS Conflation using Open Source Tools. *A white paper of JUMP project*.
- Bobzien M. Burghardt D. Petzold I. Neun M. & Weibel R. (2008).** Multi-representation Databases with Explicitly Modeled Horizontal, Vertical, and Update Relations. *Cartography and Geographic Information Science*, Vol.35, Issue 1.
- Bonin Y. (2010).** Présentation de la problématique des référentiels routiers, Axes de recherche souhaités. *Document de CETE d'Aix, PCI Géomatique (DCEDI/AGIL)*.
- Bouvier A. & George M. (1992).** Dictionnaire des mathématiques, 3ième édition, *presse universitaire de France(PUF)*.
- Brando C. (2013).** Coalla : Un modèle pour l'édition collaborative d'un contenu géographique et la gestion de sa cohérence. *Thèse en informatique*, Université Paris-Est. laboratoire COGIT.
- Brando C. & Bucher B. (2013).** Stratégies pour l'édition collaborative et la gestion de la cohérence d'un contenu géographique. *Symposium ECOTECHS'2013 : Les données pour l'agriculture et l'environnement du capteur à l'indicateur, 9-10 octobre 2013, Montoldre (Allier), France*.
- Brando C. Bucher B. & Abadie N. (2011).** Specifications for User Generated Spatial Content. *14th International Conference on Geographic Information Science (AGILE'11)*, Utrecht (Netherlands), 18-21 April, pp.479-495.

Bucher B. Falquet G. Clementini E. & Sester M. (2012). Towards a typology of spatial relations and properties for urban applications. *Proceedings of the Usage, Usability and Utility of 3D City models Conference (3U3D), Nantes.*

Buttenfield B. P. (1991). A rule for describing line feature geometry. Dans *Map Generalization: Making Rules for Knowledge Representation*. Buttenfield B. (auteur), McMaster R.B. (auteur), pp.150-171.

Chen J. Li C. Li Z. & al. (2000). Improving 9-Intersection Model by Replacing the Complement with Voronoi Region. *Geo-spatial Information Science*, vol.3 , pp.1-10.

Christophe S. (2009). Aide à la conception de légendes personnalisées et originales : proposition d'une méthode coopérative pour le choix des couleurs. *Thèse en Science d'Information Géographique*, Université de Marne-La-Vallée.

Christophe S. & Ruas A. (2002). Detecting Building Structures for generalisation purposes. *10th International Symposium on Spatial Data Handling (SDH'02)*, 9-12 July, Ottawa (Canada).

Clementini E. (2013). Directional relations and frames of reference. *GeoInformatica April 2013*, Vol. 17, Issue 2, pp.235-255.

Clementini E. & Di Felice P. (1995). A Comparison of Methods for Representing Topological Relationships. Dans *Information Sciences - Applications* (pp.149-178). Elsevier ISSN 1069-0115 Vol.3, Number 3.

Clementini E. Di Felice P. & van Oosterom P. (1993). A Small Set of Formal Topological Relationships Suitable for End-User Interaction. *SSD '93 Proceedings of the Third International Symposium on Advances in Spatial Databases* , pp.277-295.

Comité Français de Cartographie. (1990). *Glossaire de cartographie.*

David B. & Fasquel P. (1997). Qualité d'une base de données géographique : concepts et terminologie. *Bulletin d'Information de l'IGN*, N°67.

Devogele T. Parent C. & Spaccapietra S. (1998). On spatial database integration. *International Journal of Geographical Information Science* 12:4, pp.335-352, OI:doi=10.1080/136588198241824.

Devogele T. (2000). Mesure d'exactitude et processus de fusion à l'aide de la distance de Fréchet discrète, *Revue internationale de Géomatique*, vol.10, numéro 3-4, 2000, Hermès, pp.359-381.

Donnelly M. & Bittner T. (2005). Spatial Relations Between Classes of Individuals. *Spatial Information Theory*, Lecture Notes in Computer Science Vol.3693 , pp.182-199.

Duchêne C. (2004). generalisation cartographique par agents communicants : le modèle CartACom. *Thèse de doctorat en informatique*, Université Pierre et Marie Curie Paris VI. laboratoire COGIT.

Duchêne C. & Gaffuri J. (2008). Combining three multi-agent based generalisation models: AGENT CartACom and GAEL. *13th International Symposium on Spatial Data Handling (SDH'08)*, 23-25 June, Montpellier (France) .

Duchêne C. Ruas A. & Cambier C. (2012). The CartACom model: transforming cartographic features into communicating agents for cartographic generalisation. *Int. Journal of Geographical Information Science* 26(9) , pp.1533-1562.

Egenhofer M. J. & Herring J. R. (1991). Categorizing binary topological relationships between regions, lines and points in geographic databases. *Tech. Report, Department of Surveying Engineering*, pp.1-28.

Egenhofer M. J. & Mark D. M. (1995). Modeling Conceptual Neighborhoods of Topological Line-Region Relations in *International Journal of Geographical Information Systems* 9 (5) , pp.555-565.

Egenhofer M. J. Rashid A. & Shariff B. M. (1998). Metric Details for Natural-Language Spatial Relations. *ACM Transactions on Information Systems*.

Egenhofer M. (2010). Qualitative Spatial-Relation Reasoning for Design in Studying Visual and Spatial Reasoning for Design Creativity. Gero J. (ed.).

Egenhofer M. & AlTaha K. (1992). Reasoning about Gradual Changes of Topological Relationships in *Theory and Methods of Spatio-Temporal Reasoning in Geographic Space, Pisa, Italy*, and in: Frank A., Campari I., and Formentini U. (eds.), *Lecture Notes in Computer Science*, Vol. 639, Springer-Verlag , pp.196-219.

Egenhofer M. & Dube M. (2009). Topological Relations from Metric Refinements *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* , Seattle, WA M. Mokbel and P. Scheuermann (eds.) .

Egenhofer M. & Franzosa R. (1991). Point-set topological spatial relations. *Int. Journal of Geographical Information Systems* 5(2) , pp.161-174. doi :10.1080/02693799108927841.

Egenhofer M. & Mark D. (1995). Modeling Conceptual Neighborhoods of Topological Line-Region Relations. *International Journal of Geographical Information Systems* 9 (5) , pp.555-565.

Egenhofer M. Sharma J. & Mark D. M. (1993). A critical comparison of the 4-intersection and 9-intersection models for spatial relations: formal analysis. *Autocarto* volume 11 issue 92 .

Elias B. (2003). Determination of landmarks and reliability criteria for landmarks. *ICA commission on Map Generalization. 5th Workshop on Progress in Automated Map Generalization*. IGN Paris, 2003.

Fichter D. (2013). What Is a Mashup. *Library of University of Saskatchewan*.

Frank A. U. (1991). Qualitative spatial reasoning with cardinal directions in *Proc. Seventh Austrian Conference on Artificial Intelligence* pp.157-167. Vienna.

Gaffuri J. (2008). Généralisation automatique pour la prise en compte de thème champs: le modèle GAEL. *Thèse en SIG*, Université Paris-Est, Marne la Vallée. Laboratoire COGIT.

Gesbert N. (2005). Étude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration. *Thèse en informatique*, Université de Marne la Vallée. Laboratoire COGIT.

Gibson R. & Erle S. Walsh J. (2005). Mapping Hacks: Tips & Tools for Electronic Cartography., *O'Reilly Media Inc.*

Girres J.-F. (2012). Modèle d'estimation de l'imprécision des mesures géométriques de données géographiques. *Thèse en SIG*, Marne La vallée: Université Paris-Est - École doctorale MSTIC. Laboratoire COGIT.

Gould N. & Cheng J. (2012). A prototype for ontology driven on-demand mapping of urban traffic accidents. *In 15th Workshop of the ICA commission on Generalisation and Multiple Representation*. Istanbul, Turkey, 13th-15th September 2012.

Goyal R. & Egenhofer M. (2000). Consistent Queries over Cardinal Directions across Different Levels of Detail in *11th International Workshop on Database and Expert Systems Applications*, Greenwich, U.K., Tjoa A.M., Wagner R., and Al-Zobaidie A. (eds.), pp.876-880, September 2000.

Grosso E. (2010). Integration of historical geographic data into current georeferenced frameworks: A user-centred approach. *e-Perimètron, International web journal on sciences and technologies affined to history of cartography and maps*, Vol. 5, No 3, pp.107-117.

Grosso E. Perret J. & Brasebin M. (2012). GEOXYGENE : « une plate-forme interopérable pour le développement d'applications géographiques », dans Bucher B. Le Ber F. *Développements logiciels en géomatique - Innovations et mutualisation* Hermès Lavoisier, Chapitre 3.

Hauert J. & Budig B. (2012). An algorithm for map matching given incomplete road data. *Proceedings of the 20th ACM SIGSPATIAL international conference on advances in geographic information systems (ACM-GIS'12)*, pp.510–513.

Hauert J.-H. (2005). Link based Conflation of Geographic Datasets. *8th ICA WORKSHOP on Generalisation and Multiple Representation*, A Coruña.

Hauert J.-H. & Sester M. (2004). Using the Straight Skeleton for Generalisation in a Multiple Representation Environment. *ICA Workshop on Generalisation and Multiple representation 20-21 August 2004 Leicester*.

Heinzle F. & Anders K. H. (2007). Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. *Generalization of geographic information: cartographic modelling and applications*, eds. W. Mackaness, A. Ruas and T. Sarjakoski, 233-253. Oxford: Elsevier.

Hill L. L. (2006). Georeferencing: The Geographic Associations of Information, *The MIT Press*. Cambridge, Massachusetts, ISBN 9780262083546.

IGN. (2011). BD TOPO version 2.1 - descriptif de contenu. *Rapport technique*.

Jones C. (1997). Geographic Information Systems and Computer Cartography. Longman.

Kordjamshidi P. Van Otterlo M. & Moens M.-F. (2011). Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language. *ACM Transactions on Speech and Language Processing (TSLP)*, Volume 8 Issue 3.

Kurata Y. (2009). Three-Valued 9-Intersection for Deriving Possible Topological Relations from Incomplete Observations. Dans *Advances in GIScience, Lecture Notes in Geoinformation and Cartography* (pp.289-308). Springer.

- Li L. Michael F. & Goodchild M. F. (2011).** An optimisation model for linear feature matching in geographical data conflation. *International Journal of Image and Data Fusion*, 2(4), pp.309–328. Available at: <http://www.tandfonline.com/doi/abs/10.1080/19479832.2011.577458>.
- Lu M. & Arikawa M. (2013).** Location-Based Illustration Mapping Applications and Editing Tools. *Cartographica* 48(2): pp.100-112.
- Luan X. & Yang B. (2010).** Generating Strokes of Road Networks Based on Pattern Recognition. *13th Workshop of the ICA Commission on Generalisation*. Zürich, Switzerland.
- Lüscher P. Weibel R. & Burghardt D. (2008).** Alternative options of using processing knowledge to populate ontologies for the recognition of urban concepts. *11th ICA Workshop on Generalisation and Multiple Representation – 20th, 21st of June 2008, Montpellier* .
- Mackaness W. (2007).** Understanding Geographic Space. *Mackaness W. Ruas A. Sarjakoski T. (eds) : The Generalisation of Geographic Information : Models and Applications*. Elsevier .
- Mackaness W. Burghardt D. & Duchêne C. (2013).** Map Generalisation: Fundamental to the modelling and understanding of geographic space. *Chapter 1 in D. Burghardt, C. Duchene and W.A. Mackaness (eds): Abstracting Geographic Information in a Data Rich World, LNG&C, Springer, Heidelberg. ISBN 978-3-319-00202-6*.
- Massiot L. Abadie N. & Bucher B. (2011).** Mining user generated Web content to characterise geographic features beyond topographical aspects. *25th International Cartographic Conference (ICC'11)*, 3-8 July, Paris (France).
- Maudet A. Touya G. Duchêne C. & Picault S. (2014).** Representation of Interactions in a Multi-Level Multi-Agent Model for Cartography Constraint Solving. *Advances in Practical Applications of Heterogeneous Multi-Agent Systems: Proceedings of the 12th International Conference, PAAMS 2014, Lecture Notes in Computer Science, Springer, pp.183--194*
- Matsakis P. Wawrzyniak L. & Ni J. (2010).** Relative positions in words: a system that builds descriptions around Allen relations. *International Journal of Geographical Information Science* 24(1): 1-23 .
- McConchie A. L. (2008).** Mapping Mashups: Participation, Collaboration and Critique on the World Wide Web, *Master thesis*. Vancouver: The University Of British Columbia.
- Mechouche A. Abadie N. & Mustière S. (2010).** Mesure de la distance sémantique entre parties potentiellement communes à deux taxonomies. *21èmes journées Ingénierie des Connaissances (IC'06)*, 8-11 juin, Nîmes (France).
- Miller, V.C. (1953).** A Quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee, Proj. NR 389-402, Tech. Rep. 3, Columbia University, Department of Geology, ONR, New York.
- Mustière S. (2001).** Apprentissage supervisé pour la généralisation cartographique, *thèse de doctorat en informatique*, université Pierre et Marie Curie Paris VI, laboratoire COGIT.
- Mustière S. (2005).** Cartographic generalization of roads in a local and adaptive approach: A knowledge acquisition problem. *International Journal of Geographical Information Science* Vol. 19, No. 8–9, September–October 2005, pp.937–955.

- Mustière S. & Devogele h. (2008).** Matching networks with different levels of detail, December 2008, *Geoinformatica*, Volume 12, Issue 4, pp.435-453.
- Neun M. Weibel R. & Burghardt D. (2004).** Data Enrichment for Adaptive Generalisation. *ICA Workshop on Generalisation and Multiple representation*, Leicester .
- Newson P. & Krumm J. (2009).** Hidden Markov Map Matching Through Noise and Sparseness. *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp.336-343.
- Olteanu A.-M. (2008).** Fusion de connaissances imparfaites pour l'appariement de données géographiques. *thèse en SIG*, Université Paris-Est .
- Papadias D. & Theodoridis Y. (1997).** Spatial Relations, Minimum Bounding Rectangles, and Spatial Data Structures. *International Journal of Geographic Information Science*. Vol.11 pp. 111-138.
- Picault S. Mathieu P. & Kubera Y. (2010).** PADAWAN, un modèle multi-échelles pour la simulation orientée interactions. In *Actes des 18e Journées Francophones sur les Systèmes Multi-Agents (JFSMA'2010)*, Mahdia (Tunisie), Cepaduès, 2010.
- Plazanet C. (1996).** Enrichissement des bases de données géographiques : analyse de la géométrie des objets linéaires pour la généralisation cartographique (application aux routes). *Thèse en SIG*, Université de Marne la Vallée, Laboratoire COGIT.
- Steiniger S. Burghardt D. Weibel R. (2006).** Recognition of island structures for map generalization. *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*. P.67-74.
- Regnauld N. (1998).** Généralisation du Bâti : structure spatiale de type graphe et représentation cartographique. *Thèse en informatique*, Université de Provence- Aix-Marseille 1. Laboratoire COGIT
- Reis R. M. Egenhofer M. J. & Matos J. L. (2008).** Conceptual Neighborhoods of Topological Relations Between Lines. *Headway in Spatial Data Handling , Lecture Notes in Geoinformation and Cartography* , pp.557-574 .
- Renard J. Gaffuri J. & Duchêne C. (2010).** Capitalisation problem in research - example of a new platform for generalisation: CartAGen. *12th ICA Workshop on Generalisation and Multiple Representation, Zürich (Switzerland)*.
- Renard J. Gaffuri J. Duchêne C. & Touya G. (2011).** Automated generalisation results using the agent-based platform CartAGen. *25th International Cartographic Conference (ICC'11) - poster session*, 7 July, Paris (France) .
- Rhind D. W. Green N. P. Mounsey H. M. & Wiggins J. C. (1986).** The integration of geographic data. *Proceeding of Austro Carto Perth. Austerian Cartographic Association*, Perth, pp.273-293.
- Ruas A. (1999).** Modèle de généralisation de données géographiques à base de contraintes et d'autonomie, *Thèse en SIG*, Université de Marne la Vallée. Laboratoire COGIT.
- Ruas A. & Bianchin A. (2002).** Echelle et niveau de détail. *Généralisation et représentation multiple*. chapitre 1, p.26-44. Paris: Hermes Lavoisier.

- Ruas A. (2002).** Généralisation et représentation multiple, chapitre Echelle et niveau de détail pp.26-44. Paris: Hermes-Lavoisier.
- Ruas A. (2004).** Le changement de niveau de détail dans la représentation de l'information géographique. *Mémoire d'Habilitation à Diriger les Recherches, spécialité sciences de l'information géographique*, Université de Marne-la-Vallée.
- Ruas A. & Lagrange J.-P. (1994).** Geographic information modelling: GIS and generalisation. *6th International Symposium on Spatial Data Handling (SDH'94)*, 5-9 September, Edinburgh (UK), pp.1099-1117.
- Ruas A. & Mustière S. (2005).** Bases de données géographiques et cartographiques à différents niveaux de détail. *Bulletin du Comité Français de Cartographie*, 185:pp.5-14.
- Ruas A. & Plazanet C. (1996).** Strategies for automated map generalisation. *7th International Symposium on Spatial Data Handling (SDH'96)*, Delft (Netherlands), 12-16 August, pp.319-336 .
- Safra E. Kanza Y. Sagiv Y. & Doytsher Y. (2006).** Efficient Integration of Road Maps. *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems, Arlington (Etats-Unis)*.
- Schneider M. & Behr T. (2006).** Topological Relationships Between Complex Spatial Objects. *ACM Transactions on Database Systems (TODS)*, 31(1), pp.39-81.
- Schwering A. (2006).** Semantic Similarity Measurement including Spatial Relations for Semantic Information Retrieval of Geo-Spatial Data. *Thesis in nature science*, Université de Münster.
- Sheeren D. (2005).** Méthodologie d'évaluation de la cohérence inter-représentations pour l'intégration de bases de données spatiales. Une approche combinant l'utilisation de métadonnées et l'apprentissage automatique. *Thèse en informatique*, Université Paris 6. Laboratoire COGIT.
- International standardization organization(ISO) (2003).** International standard : Geographic information — Spatial referencing by geographic identifiers ISO 19112 First edition.
- Steiniger S. & Weibel R. (2007).** Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science* 34(3) , pp.175–197.
- Steiniger S. Burghardt D. & Weibel R. (2006).** Recognition of island structures for map generalization. *proceedings of the 14th annual ACM International Symposium on Advances in Geographic Information Systems, ACM-GIS'06*, Arlington, Virginia, pp.67-74.
- Steiniger S. Lange T. Burghardt D. & Weibel R. (2008).** An approach for the classification of urban structures based on discriminant analysis techniques. *Transaction in GIS* ,Volume 12, Issue 1, Page 31-59, Feb 2008.
- Thomson R. C. & Richardson D. E. (1999).** The 'good continuation' principle of perceptual organization applied to the generalization of road networks. *Proceedings of the ICA 19th International Cartographic Conference*. pp.1215-1223.

Touya G. (2010). Enrichissement automatique de données par analyse spatiale pour la généralisation de réseaux. *International Journal of Geomatics and Spatial Analysis* VOL 20/2 - 2010 - pp.175-199 - doi:10.3166/rig.20.175-199.

Touya G. (2011). *Orchestration d'un processus multi-modèles de généralisation d'espaces géographiques hétérogènes. Thèse en sig.* Université Paris-Est. Laboratoire COGIT.

Touya G. & Brando C. (2013). Detecting Level-of-Detail inconsistencies in Volunteered Geographic Information data sets. *Cartographica: The International Journal for Geographic Information and Geovisualization*, Vol. 48 n°2, Selection of 26th International Cartographic Conference, 25-30 August, Dresden (Germany), pp.134-143 .

Touya G. Balley S. Duchêne C. Jaara K. & Regnaud N. (2012). Towards an ontology of generalisation constraints and spatial relations. *ICA Workshop on Generalisation and Multiple Representation*, 13-14 September, Istanbul (Turkey).

Touya G. Coupé A. Le Jollec J. Dorie O. & Fuchs F. (2013). Conflation Optimized by Least Squares to Maintain Geographic Shapes. *ISPRS Int. J. Geo-Inf*, 2, 621-644; doi:10.3390/ijgi2030621 .

Van Damme M.-D. (2013). GeOxygene : une plate-forme de recherche pour le développement d'applications SIG. *Journées Francophones Open source Géomatique (FROG'13)*, 10 juin, St Mandé (France) .

Van Oort P. (2005). Spatial Data Quality : From Description To Application. *Thèse de doctorat*, Wageningen University .

Vauglin F. (1997). Modèles statistiques des imprécisions géométriques des objets géographiques. *Thèse de doctorat*, Université Marne-La-Vallée .

Wallgrün J. O. (2013). Topological Adjustment of Polygonal Data. *Advances in Geographic Information Science*, Springer , pp.193-208.

Walter V. & Fritsch D. (1999). Matching Spatial Data Sets: Statistical Approach. *International Journal of Geographical Information Science* 13(5) , pp.445-473.

XU J. (2007). Formalizing natural-language spatial relations between linear objects with topological and metric properties. *international Journal of Geographical Information Science*.Vol. 21, No. 4 (pp.377–395).

Zhang X. Tinghua A. Stoter J. Kraak M. & Molenaar M. (2013). Building pattern recognition in topographic data: examples on collinear and curvilinear alignments. *GeoInformatica* 17, pp.1–33.

Zhong Z.-N. Jing N. Chen L. & Wu Q.-Y. (2004). Representing topological relationships among heterogeneous Geometry-Collection features. *Journal of Computer Science and Technology* Volume 19, Issue 3 , pp.280-289.