



HAL
open science

L'aide à l'accès à l'information géographique: un environnement de conception coopérative d'utilisations de données géographiques

Bénédicte Bucher

► To cite this version:

Bénédicte Bucher. L'aide à l'accès à l'information géographique: un environnement de conception coopérative d'utilisations de données géographiques. Informatique et langage [cs.CL]. Université Paris VI - Pierre et Marie Curie, 2002. Français. NNT: . tel-02474452

HAL Id: tel-02474452

<https://theses.hal.science/tel-02474452v1>

Submitted on 11 Feb 2020

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.

THESE DE DOCTORAT DE L'UNIVERSITE PARIS 6

Spécialité

Informatique

présentée par

Bénédicte BUCHER

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITE PARIS 6

Sujet de la thèse :

L'aide à l'accès à l'information géographique : un environnement de conception coopérative d'utilisations de données géographiques

soutenue le : 25 juin 2002

devant le jury composé de :

Professeur Jacques ARSAC
Professeur François BOUILLE
Professeur Christophe CLARAMUNT
Professeur Anne DOUCET
Professeur Jean-Claude MÜLLER
Professeur Marinette REVENU
Monsieur Didier RICHARD
Madame Anne RUAS

Président du jury
Directeur de thèse
Examineur
Rapporteur
Rapporteur
Rapporteur
Invité
co-directeur de thèse

TABLE DES MATIERES

INTRODUCTION	9
---------------------	----------

1 PROBLEMATIQUE : AIDER L'ACCES A DES RESSOURCES EN INFORMATION (GEOGRAPHIQUE)	15
---	-----------

1.1	L'accès aux ressources en information géographique	16
1.1.1	Les ressources en information géographique	16
1.1.1.1	L'appréhension et la représentation de l'espace géographique	16
1.1.1.2	La formalisation de l'information géographique	17
1.1.1.3	Les bases de données géographiques	20
1.1.2	L'exploitation des données géographiques	22
1.1.2.1	Les contextes d'utilisation	22
1.1.2.2	Les techniques	25
1.1.2.3	Les outils : les SIG	27
1.1.3	L'accès à l'information géographique	32
1.1.3.1	L'accès aux données	32
1.1.3.2	L'accès à l'interprétation des données	38
1.1.3.3	L'accès à l'exploitation	39
1.1.3.4	L'accès à plusieurs bases de données	41
1.1.4	Bilan de l'accès à l'information géographique	42
1.2	Aider l'accès à des ressources en information	44
1.2.1	Accéder à des ressources en information	44
1.2.1.1	Les contextes	44
1.2.1.2	Le transfert de connaissances	46
1.2.1.3	L'accès	49
1.2.2	Aider l'accès	51
1.2.2.1	Faciliter l'accès d'un utilisateur à l'exploitation d'une ressource	51
1.2.2.2	Faciliter l'accès d'un utilisateur à l'exploitation de plusieurs ressources	53
1.2.2.3	Faciliter l'accès de plusieurs utilisateurs à l'exploitation de ressources	54
1.3	Bilan	55

2 ANALYSE DE L'ETAT DE L'ART	57
-------------------------------------	-----------

2.1	Etat de l'art des techniques d'aide à l'accès	58
2.1.1	Généralités	58
2.1.1.1	Solution générique	58
2.1.1.2	Un dilemme classique : expressivité et maniabilité	60
2.1.1.3	Structure de l'état de l'art des techniques	61
2.1.2	L'aide à l'accès à l'information (données, informations, connaissances)	62
2.1.2.1	L'aide à l'accès aux données	62
2.1.2.2	L'aide à l'accès aux informations	63
2.1.2.3	L'aide à l'accès aux connaissances	69
2.1.3	L'aide au raisonnement formel	71
2.1.3.1	Généralités	71
2.1.3.2	La modélisation axée sur le <i>QUOI</i> : les ontologies de domaine	73
2.1.3.3	Les modèles axés sur le <i>COMMENT</i>	76
2.1.4	Bilan des techniques d'aide à l'accès et de leur intérêt dans notre contexte	85
2.2	Etat de l'art des formalisations des <i>QUOI</i> et <i>COMMENT</i> géographiques	89
2.2.1	Généralités	89
2.2.2	Le <i>QUOI</i>	90
2.2.2.1	La représentation générale de l'information géographique	90
2.2.2.2	Des représentations dédiées à des contextes applicatifs	92
2.2.3	Le <i>Pourquoi</i>	93
2.2.3.1	Les interactions de l'homme et du monde	94
2.2.3.2	Les représentations de l'espace géographique partagées par tous	96
2.2.3.3	Les représentations de l'espace géographique dédiées à des communautés d'utilisateurs	98
2.2.3.4	La formalisation de résultats d'applications	99
2.2.4	Le <i>Comment</i>	99
2.2.4.1	La modélisation	100
2.2.4.2	La structuration de l'application	101
2.2.5	Le <i>Avec Quoi</i>	103
2.2.5.1	La création d'outils	103
2.2.5.2	Les descriptions normalisées d'outils	106
2.2.6	Bilan	110
3	NOTRE APPROCHE	111
3.1	Définition de notre contribution	112
3.1.1	Principes généraux	112
3.1.2	Définition du modèle	115
3.1.2.1	Les tâches	115

3.1.2.2	Les tâches complexes	116
3.1.2.3	Les tâches primitives	118
3.1.2.4	Rôles et domaine	119
3.1.2.5	Un langage de description de tâches	122
3.1.3	Un système d'aide à l'accès	125
3.2	Spécialisation et instanciation du modèle	131
3.2.1	Spécification du domaine géographique	131
3.2.1.1	Rendre compte du contexte de l'utilisateur	132
3.2.1.2	Intégrer les deux grandes catégories de représentations de l'espace géographique	133
3.2.1.3	Décrire les lots de données disponibles	135
3.2.2	Méthode d'analyse de tâches géographiques	136
3.2.3	La tâche d'appariement	137
3.2.3.1	Vocabulaire général	138
3.2.3.2	Détail de la méthode	141
3.2.3.3	Exploitation du modèle de l'appariement	146
3.2.4	La tâche de localisation	147
3.2.4.1	L'aspect déclaratif	148
3.2.4.2	L'aspect opérationnel	153
3.2.4.3	Exploitation du modèle de localisation	159
3.2.5	La tâche de détection de proximité et de détermination de navigation	161
3.2.5.1	Les rôles de la détection de proximité	161
3.2.5.2	Les rôles de la détermination de navigation	162
3.2.5.3	Exploitation du modèle de navigation	165
3.3	Opérationnalisation du modèle	168
3.3.1	La représentation du modèle de tâches géographiques	168
3.3.1.1	La représentation du domaine	168
3.3.1.2	La représentation des tâches	170
3.3.2	La dynamique associée au modèle géographique	172
3.3.2.1	Mécanismes de spécification	172
3.3.2.2	L'interface	177
3.3.2.3	Implémentation de la tâche de Localisation	183
3.3.3	Bilan	186
3.3.3.1	Apports à la résolution de la problématique initiale	187
3.3.3.2	Evaluation du prototype	188
CONCLUSION ET PERSPECTIVES		191
GLOSSAIRE		197

BIBLIOGRAPHIE

199

INTRODUCTION

Contexte

L'espace géographique est le monde dans lequel nous vivons. La notion de situation dans l'espace se trouve au cœur de la plupart des raisonnements, décisions, et activités de l'homme, que ce soit dans sa vie de tous les jours ou dans certaines professions. Pour cette raison, l'homme a toujours eu besoin de représentations de portions du monde géographique pour avoir une appréhension de ce monde dépassant sa propre perception. De telles représentations sont classiquement les cartes. La carte remplit trois fonctionnalités : stocker des informations géographiques, permettre des raisonnements sur ces informations, comme concevoir un itinéraire ou implanter un centre de loisirs, et être un vecteur de communication.

Depuis une quinzaine d'années, les bases de données géographiques numériques (BDG) sont produites par des instituts nationaux comme l'Institut Géographique National (IGN) pour en dériver des cartes papier ou par d'autres organismes dans le cadre d'applications plus spécifiques. Les BDG remplacent les cartes papier pour la première fonctionnalité de stockage d'informations géographiques. Elles remplissent aussi la fonctionnalité de supporter le raisonnement puisqu'elles servent de contenu à des applications de traitement de l'information géographique. Elles peuvent souvent être utiles au-delà de l'application pour laquelle elles ont été spécifiées. Cela reste vrai même si cette application est une cartographie; les informations contenues dans une base de données numériques ne sont ni entièrement transcritibles ni surtout opérationnelles sur une carte papier. Certains utilisateurs potentiels pourraient dériver de BDG des informations non explicitées dans les produits qui en ont été dérivés (application initiale ou cartes), et difficile à dériver de ces produits initiaux. Ce processus est restreint à des utilisateurs qui entretiennent des relations privilégiés avec les instituts nationaux, par exemple les utilisateurs du domaine routier et celui de la défense, ou qui produisent eux-mêmes leurs bases. Il est nécessaire de faire un effort de diffusion des BDG vers tous les utilisateurs potentiels de ces bases.

Le contexte actuel rend d'ailleurs de plus en plus pressant ce besoin de diffusion car les utilisateurs potentiels de données géographiques sont en nombre croissant. L'information géographique joue un rôle important dans la société d'information. Il est classiquement admis que 80% des décisions s'appuient sur un contexte géographique, comme le rappellent [Denègre et Salgé 96]. Il y a actuellement un accroissement du volume de données susceptibles d'être localisées dans leur emploi, c'est-à-dire dont l'exploitation nécessite de les géo-référencer. On peut citer les données du géomarketing. De plus, les logiciels de manipulation de données géographiques se vulgarisent.

Dans une optique de diffusion des données géographiques, des métadonnées géographiques, c'est-à-dire des données décrivant les BDG sont développées. Elles sont associées à la construction de bibliothèques de BDG, et plus récemment d'infrastructures nationales d'informations localisées. Etant à l'origine dédiées au transfert de données, elles comportent essentiellement des informations relatives aux protocoles de production des données et au contenu des bases. Or l'expression d'un besoin de données géographiques ne peut pas s'appuyer sur des seules informations de contenu des bases. Elle doit s'appuyer sur d'autres termes.

Ces *autres termes* doivent par exemple rendre compte de la diversité des modélisations et représentations de l'espace géographique. Le terme "modélisations" renvoie aux modèles construits pour les raisonnements d'hommes et le terme "représentations" renvoie aux modèles construits pour les raisonnements de machines. Il n'existe pas de modélisation et de représentation universelle de l'espace géographique [Raper 96] [Mennis et al. 00]. Une des causes en est que l'espace géographique –modélisé dans l'information géographique– est une réalité que chacun perçoit et représente dans son expérience. Les activités conduisant à modéliser l'espace géographique sont diverses, courantes ou spécialisées, comme la circulation routière ou l'aménagement du territoire. Les catégories utilisées pour décrire le monde varient d'une représentation à l'autre et il n'existe pas d'ontologie de l'information géographique. Les concepts des utilisateurs peuvent donc renvoyer à des contextes applicatifs divers et s'appuyer sur des termes applicatifs précis. Ces concepts peuvent ne pas être représentés dans la BDG, par exemple certaines relations entre objets, ou ils peuvent être représentés de façon très différente, par exemple le concept de ville ou de quartier. C'est le cas de concepts dont la représentation dans les BDG fait appel à des notions mathématiques sophistiquées comme la notion de situation dans l'espace. Cette notion, essentielle aux raisonnements géographiques, est explicitée dans les BDG de la façon suivante : des données géographiques représentent une partie du monde réel en décrivant des éléments de ce monde, par exemple une route ou une borne géodésique, et leurs coordonnées dans un système de référence en général mathématique. Ces coordonnées sont la représentation de la notion de situation dans l'espace. Or l'espace est continu et il est délicat de représenter par exemple les limites du Mont Blanc. Certaines entités sont ainsi représentées dans une BDG avec des choix de représentation auquel l'utilisateur n'a pas toujours accès. D'autres choix menant à diverses représentations sont aussi issus de la nécessité de discrétisation des formes géométriques ou de leur approximation par des formes connues. La nécessité de représenter des formes géométriques dans des données numériques, par exemple informatiser le contour de la côte bretonne, implique la discrétisation de ces formes, ou leur approximation par des formes connues. Ces choix éloignent la représentation de celle que l'utilisateur peut avoir.

Ces *autres termes* doivent également rendre compte des capacités de dérivation de logiciels spécifiques, les systèmes d'information géographique (SIG), qui sont dédiés à la manipulation des données géographiques. Ces systèmes proposent des fonctionnalités de systèmes de gestion de bases de données (SGBD) classique ainsi que la visualisation graphique des données, le support de requêtes spatiales, et des traitements spécifiques d'analyse spatiale et de cartographie. Ces systèmes apportent aux utilisateurs les capacités de dérivations nécessaires à la satisfaction de leurs besoins, mais ils apportent peu d'information sur la façon de les utiliser. Or la manipulation de données géographiques numériques, c'est-à-dire l'utilisation de ces logiciels, ne suit pas la même logique que la lecture d'une carte. L'œil de l'homme, lors de la lecture d'une carte papier, réalise des traitements complexes qui s'appuient sur des informations non explicitées dans la représentation cartographique comme les relations de voisinage. Typiquement, la lecture d'une carte permet de répondre rapidement à une requête comme : « trouver les grandes maisons situées à l'intérieur du virage sur la RN20, peu après le croisement avec le pont ». Or les représentations utilisées pour construire des BDG sont pour la plupart des représentations cartographiques, qui n'explicitent pas non plus ces relations. L'automatisation de la précédente requête sur des données numériques est par exemple relativement complexe puisqu'elle s'appuie sur des objets géométriques et met en œuvre des traitements mathématiques.

En définitive l'accès d'utilisateurs aux données géographiques, c'est-à-dire au potentiel d'information dérivable de ces données, pose des problèmes nouveaux par rapport à l'accès aux cartes qui ne sont pas réglés par les métadonnées actuelles. L'expression du besoin d'information géographique d'un utilisateur est très loin de la description de l'utilisation de données géographiques pouvant dériver cette information, sous une forme ou sous une autre. Les métadonnées géographiques ne sont pas suffisante pour traduire l'expression d'un besoin en une désignation des ressources utiles pour le satisfaire : elles ne tiennent pas suffisamment compte des concepts nécessaires à l'expression de besoins, et elles ne tiennent pas compte des capacités de dérivation des SIG.

Objectif de ce travail et démarche globale

Cette thèse est menée au laboratoire COGIT de l'IGN, laboratoire travaillant sur les bases de données géographiques : leur modélisation, leur entretien, leur utilisation pour dériver des informations à une résolution plus petite, leur consultation par des utilisateurs, la modélisation d'une application spécifique (la gestion d'un risque, la rédaction d'une carte).

L'objectif de ce travail est d'améliorer l'accès des utilisateurs aux bases de données géographiques. Il ne s'agit pas de l'accès aux données en tant que telles : *savoir que les données existent, où elles sont, et comment les obtenir*. Il s'agit plutôt de l'accès aux données en tant que ressources en information : *savoir comment utiliser un jeu de données, ou plusieurs jeux ensembles, pour répondre à un besoin*.

L'accès à des ressources en information est un problème important qui dépasse le domaine de l'information géographique. La constitution croissante de « grosses » bases spécialisées et la diffusion de l'information doivent amener de plus en plus d'utilisateurs divers à manipuler des informations. Cette problématique élargie renvoie à deux activités importantes : le stockage de l'information et l'exploitation des stocks.

La constitution de stocks d'information -typiquement des bases de données- nécessite de choisir une représentation pour pouvoir entreposer des données décrivant cette information. Certaines informations se prêtent naturellement à la phase de représentation, quand elles appartiennent à un domaine déjà structuré, comme des données bancaires. Mais ce n'est pas toujours le cas, en particulier l'information géographique ne renvoie pas à un domaine naturellement structuré. Il faut alors construire un modèle, et une représentation associée à ce modèle, pour décrire des informations. Cette activité s'appuie sur les modèles associés aux méthodes d'acquisition de données, et sur les modèles les mieux adaptés aux utilisations prévues des données.

L'activité suivante, accéder à une information stockée, implique de comprendre le modèle et la représentation choisis lors de la constitution des ressources et éventuellement d'en changer si le modèle n'est pas adapté à l'utilisation que l'on veut faire des données. L'accès peut aussi impliquer de travailler avec plusieurs représentations différentes si plusieurs ressources sont exploitées conjointement.

Le cas de l'accès à l'exploitation de l'information géographique est particulièrement problématique en l'absence de représentation universelle de cette information. Ceci est aggravé par un problème moins théorique mais incontournable : il existe plusieurs formats de données géographiques, en raison de la multiplicité des formats propriétaires développés par les vendeurs de SIG, logiciels dédiés à la manipulation de ces données.

Le domaine scientifique apportant le plus d'éléments de réponses à ce problème général est l'Intelligence Artificielle (IA) et plus précisément les démarches en acquisition des connaissances. L'acquisition des connaissances est la branche de l'IA qui étudie la formalisation de connaissances : expliciter les connaissances possédées par des experts et les rendre opérationnelles dans des systèmes. Dans notre cas, ces démarches sont intéressantes car elles supportent le raisonnement formel et la répartition des connaissances de ce raisonnement entre divers acteurs. Il nous semble qu'il est justement nécessaire de mettre en place de meilleurs moyens au service d'un type de raisonnement formel particulier : la conception d'applications géographiques à partir de ressources disponibles. Il est important que ce type de raisonnement soit distribué entre plusieurs acteurs : l'utilisateur qui exprime son besoin, et un système d'aide à l'accès qui lui apporte des connaissances qu'il ne possède pas, par exemple sur les capacités de dérivation des SIG.

Ce travail est donc une démarche d'application des résultats de l'acquisition des connaissances au domaine particulier de l'information géographique.

Plan de ce rapport

Notre démarche est structurée par les questions suivantes :

- Quel est le contexte de l'aide à l'accès à l'information géographique ?
- Quelles sont les aides actuelles ?
- Quelle contribution proposons-nous?

La réponse à la première question s'appuie sur l'analyse, dans la section 1.1, de ce qu'est l'information géographique, quelles données représentent cette information et quelles en sont les fournisseurs et les utilisations. Cette première question nous amène aussi à étudier dans la section 1.2 la notion d'accès à une ressource en information et d'aide à cet accès.

La deuxième question est traitée en deux temps.

La partie 2.1 fait l'état de l'art des techniques utilisées pour faciliter l'accès à des ressources en information. Cette analyse nous permet de délimiter les types de démarches qui nous intéressent : les techniques d'aide au raisonnement formel mises en place par les travaux d'IA et qui portent sur la modélisation. Actuellement, il est préconisé de modéliser de façon distincte et reliée deux catégories de connaissances : des connaissances de description d'un domaine, le *QUOI*, et des connaissances de manipulation des objets de ce domaine, le *COMMENT*.

Le deuxième volet de l'état de l'art, 2.2, analyse les formalisations de l'information géographique et l'application du principe énoncé ci-dessus de distinction et intégration du *QUOI* et du *COMMENT*. Les démarches étudiées ne sont pas nécessairement dédiées à l'accès mais l'analyse du cas général a permis de montrer que ces démarches étaient une contribution pertinente à notre problématique. L'étude de ces démarches montre qu'elles n'appliquent pas le principe essentiel exposé ci-dessus, concernant la modélisation des connaissances. Elles n'explicitent pas la distinction entre le *QUOI* et le *COMMENT*. De plus, aucune démarche ne porte sur une formalisation globale du *COMMENT*, c'est-à-dire une formalisation décrivant des opérations élémentaires, des méthodes combinant ces opérations, et des applications mettant en œuvre ces méthodes dans des contextes d'utilisateurs.

Nous répondons enfin à la troisième question en nous appuyant sur les principes de modélisation provenant des recherches en acquisition de connaissances. Nous proposons un modèle décrivant les données géographiques, c'est-à-dire situé au niveau des métadonnées, et qui intègre des connaissances sur les *QUOI* et *COMMENT* de l'information géographique. Le *QUOI* renvoie aux descriptions de l'espace géographique construites dans plusieurs contextes : l'appréhension par l'homme de cet espace, les communautés d'utilisateurs, et la représentation de l'information géographique dans des bases de données. Le *COMMENT* comprend non seulement les outils de manipulation des données mais aussi les méthodes de manipulation de ces outils et la formulation de besoins auxquels ces méthodes permettent d'apporter des réponses.

Dans le cadre de cette thèse, il n'est pas envisageable de construire la base de connaissances complète associée à ce modèle. Nous mettons plutôt en place des éléments génériques de modélisation qui doivent se spécialiser par la suite et être instanciés pour constituer une base de connaissances de plus en plus complète.

Pour faire ce modèle, nous proposons un langage de description de tâches, méthodes et rôles. Ces notions provenant du domaine de l'acquisition des connaissances permettent d'intégrer les deux catégories *QUOI* et *COMMENT* de respecter la distinction entre ces deux catégories. Notre modèle est formalisé à l'aide d'une grammaire Backus Naur Form.

Par ailleurs, nous définissons l'architecture et les fonctions d'un système associé à ce modèle et dédié à l'aide à l'accès. Ces fonctions sont formalisées sous forme de tâches à l'aide de la grammaire. Ces dernières tâches ne sont pas des tâches géographiques. Ce sont des tâches qui réalisent une spécification coopérative d'une tâche

géographique par un utilisateur et le système. La tâche géographique spécifiée correspond à une utilisation de données géographiques qui répond au besoin de l'utilisateur.

Une étape incontournable de cette démarche est la construction d'un prototype. Nous choisissons un langage d'implémentation orienté-objet. Ce prototype permet d'évaluer la faisabilité de notre démarche c'est-à-dire sa contribution à la résolution du problème initial : l'accès d'utilisateurs à l'exploitation de données géographiques.

1 PROBLEMATIQUE : AIDER L'ACCES A DES RESSOURCES EN INFORMATION (GEOGRAPHIQUE)

Le but de ce chapitre est d'exposer le problème que ce mémoire se propose de traiter. Ce travail porte sur l'accès aux données géographiques en tant que ressources en information géographique. La formalisation de ce problème est difficile pour deux grandes raisons. La première est la complexité de l'information géographique : ces ressources sont variées, leurs manipulations possibles également. La deuxième raison est la complexité du problème général d'accès à des informations stockées. Nous consacrons une section pour détailler chacun de ces aspects indépendamment de l'autre, c'est-à-dire la nature complexe de l'information géographique d'une part et la notion d'accéder à des informations, non nécessairement géographiques, d'autre part.

Une première section, 1.1, rappelle ainsi ce qu'est l'information géographique, comment elle est exploitée, et résume la situation actuelle concernant l'accès d'utilisateurs à l'exploitation de données géographiques et les obstacles à cet accès, sans formaliser davantage ce problème.

Nous nous concentrons ensuite sur le cas général dans la section 1.2. La notion d'accès à des informations stockées est détaillée dans la section 1.2.1. Puis le problème d'aider cet accès est formalisé dans la section 1.2.2.

1.1 L'accès aux ressources en information géographique

Dans cette partie, nous présentons les éléments importants de l'accès aux ressources en information géographique, et plus exactement les éléments qui en font une problématique. L'information géographique est un domaine complexe. Nous ne cherchons pas à le présenter en détail mais plutôt à mettre en avant les facteurs de sa complexité.

Dans une première section, les processus menant à la constitution de ressources en information géographique sont introduits : l'appréhension et la formalisation de l'espace géographique, la représentation d'information géographique dans des données.

La section suivante présente les exploitations de ces ressources.

Enfin, nous détaillons les modalités de l'accès d'utilisateurs à cette information c'est-à-dire l'accès aux données et à leur exploitation.

1.1.1 Les ressources en information géographique

L'information géographique est la représentation de l'espace géographique. Le terme représentation est pris ici au sens large :

- il peut s'agir de schémas mentaux construits lors de l'appréhension d'une réalité, et guidant les appréhensions futures,
- il peut s'agir de modèles de raisonnement,
- ou encore cela peut être un schéma de données.

Nous étudions dans une première section, 1.1.1.1, l'information géographique construite dans les deux premiers cas, c'est-à-dire lors de l'appréhension de l'espace géographique par l'homme et de la représentation de cet espace dans les raisonnements de l'homme. La section suivante, 1.1.1.2, détaille les formalisations de cette information en vue de la stocker dans des données ou de la manipuler automatiquement dans des systèmes.

1.1.1.1 L'appréhension et la représentation de l'espace géographique

De façon générale, toute représentation est le résultat d'une opération d'abstraction et dépend des sujets qui la font, et donc du contexte. Un des principaux facteurs de la complexité de l'information

géographique est que les sujets construisant les représentations de l'espace géographique ne sont pas des experts partageant un même savoir dans un contexte unique et bien défini. Nous détaillons dans cette section les raisons pour lesquelles toute représentation de l'espace géographique est particulièrement subjective.

L'espace géographique est le monde qui nous entoure. Chaque homme est plongé dans l'espace géographique et est amené à le représenter en fonction de ce qu'il en perçoit dans son expérience courante. [Worboys 96] formule cela en remarquant que l'espace géographique n'est pas un "table top space", c'est-à-dire un espace observable de l'extérieur, comme des objets posés sur une table. De plus, l'espace géographique est un espace dans un monde dans lequel chacun agit, et chacun en construit une représentation dans le contexte de ses propres actions c'est-à-dire qu'il limite la représentation de l'espace géographique aux informations suffisantes pour guider ses décisions. De fait, l'information géographique est rarement au cœur d'une activité, c'est plutôt un outil supportant des activités et elle est modélisée en tant qu'outil. Par exemple, les informations retenues dans un paysage géographique ne sont pas les mêmes selon qu'on cherche à se déplacer, se nourrir, ou peindre.

A ce stade, il n'existe pas une représentation unique de l'espace géographique. L'échelle de la perception varie, ainsi que les objets perçus. Notons cependant qu'un point commun des espaces géographiques perçus est qu'ils sont rattachés à la surface de la Terre.

L'espace géographique fait également l'objet de représentations dans des communautés d'experts. Ces représentations varient en fonction de l'activité à laquelle se consacre la communauté. [Raper 96] rappelle ainsi que l'espace géographique a des significations différentes selon les disciplines. En physique, c'est un modèle unifiant la matière et l'énergie. En mathématique, l'espace est constitué d'un nombre arbitraire d'axes indépendants. En psychologie l'espace géographique est parfois vu comme un ensemble de stimuli des sens. En linguistique, l'espace géographique est un modèle de lexicalisation : les mots sont créés pour désigner des entités stables et limitées. En sciences sociales, l'espace géographique est soit une ressource, soit un cadre d'interactions. D'autres communautés se construisent non pas autour de disciplines théoriques mais autour d'activités concrètes comme le transport routier, la défense, l'aménagement ou le tourisme.

Il n'existe ainsi pas de représentation universelle de l'espace géographique. Les représentations de cet espace varient en fonction des nombreux contextes dans lesquels elles sont construites. Dans la plupart des contextes, l'information géographique est représenté en tant qu'outil supportant une activité. Ces activités peuvent prendre place dans l'expérience courante d'un individu ou dans une profession.

1.1.1.2 La formalisation de l'information géographique

Les représentations de l'espace géographique décrites ci-dessus se prêtent difficilement à une formalisation plus poussée. Détaillons les stades auxquels se posent ces difficultés.

A l'égard de l'appréhension par l'homme de son espace géographique, la principale raison du manque de représentation naturelle de l'espace géographique est que toute représentation se construit classiquement en abstrayant en objets individualisés une réalité. Or l'espace géographique est pour l'essentiel un espace continu comme le rappellent [Laurini et Gordillo 00]. Les objets constituant notre appréhension de cet espace, ou nos modèles de cet espace, ont souvent des limites conventionnelles. Par exemple il est délicat de définir précisément les limites d'une route, d'une rivière, d'une forêt ou d'un espace urbain.

Les représentations formelles utilisées pour construire des bases de données géographiques et automatiser des raisonnements sur ces données héritent nécessairement de la difficulté décrite précédemment. Les représentations de l'information géographique dans nos expériences courantes ne sont pas des représentations d'un domaine structuré et elles s'accommodent mal d'une formalisation plus poussée.

Par ailleurs, comme le souligne [Gonçalves 99], le raisonnement de l'homme dans son univers courant « s'appuie souvent, non pas sur des connaissances numériques exactes (i.e. des données quantitatives), mais plutôt sur des données de nature qualitative et relatives les unes aux autres. Les informations dont il dispose sont par exemple 'avancer de quelques mètres et prendre la rue à gauche, à l'angle de l'immeuble le plus petit', si les immeubles qui l'entourent sont tous plus grands, et non pas 'avancer de 17 mètres et prendre la rue à 190 degrés, à l'angle de l'immeuble de 53 mètres de hauteur'. » Il est clairement plus complexe de représenter, de façon exploitable, dans des données des informations qualitatives que quantitatives.

D'autres difficultés surviennent lors de la construction de bases de données géographiques. Avant de les présenter, il nous faut décrire brièvement ce qu'est une base de données géographique. Cette description est détaillée davantage dans la section suivante, 1.1.1.3.

Un élément se trouve au cœur de la plupart des activités de l'homme dans le monde et est de fait induit dans toute représentation de l'espace géographique : la notion de situation dans l'espace. Ce n'est pas un élément physique de l'espace géographique mais plutôt une notion abstraite. Une représentation servant à construire une base de données géographiques doit représenter explicitement cette notion pour que les données puissent être utilisées. Une représentation de données géographique comprend donc :

- *la représentation d'un modèle du monde réel* : Des données géographiques sont un ensemble de données représentant une partie de la Terre, comme une route ou un point géodésique, une maison, un fleuve, une limite administrative, une cote altimétrique. Ces informations sont classiquement appelées informations sémantiques. Cela dit le terme données géographiques désigne aussi parfois des données ne décrivant pas une partie de la surface de la terre mais une activité ou une propriété rattachée à une partie de la surface de la terre –c'est-à-dire une donnée géo-référencable -, comme par exemple la population d'une zone.
- *la représentation de la position par rapport à la Terre des éléments de ce modèle* : Cela implique d'une part de représenter la position. Une position est définie dans un référentiel mathématique dans lequel on connaît la position de la Terre. Cela implique d'autre part de représenter le lien entre cette position représentée et les éléments représentant l'objet du monde réel qui se trouve à cette position.
- *la représentation de la composante temporelle de l'information géographique* : En effet l'espace géographique est un espace qui évolue et dans lequel prennent place des dynamiques. Notons que les représentations actuelles de l'information géographique négligent souvent cette composante et fournissent une description statique de l'espace géographique.

Les bases de données géographiques comportent ainsi des informations alphanumériques décrivant les objets du monde réel, des informations géométriques décrivant les positions dans un référentiel, et des correspondances entre ces informations alphanumériques et géométriques (cf. Figure 1).

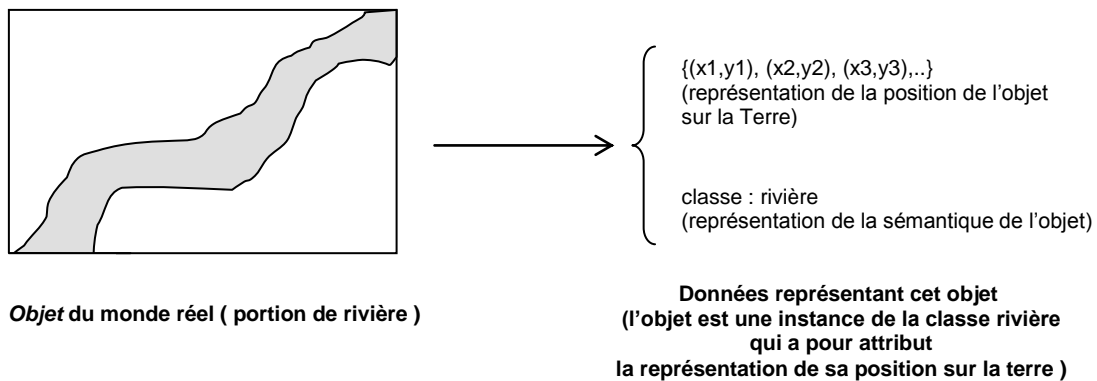


Figure 1 : Exemple de représentation d'information géographique sur une rivière : sa position par rapport à la terre est représentée par des données géométriques et l'objet du monde réel est représenté par une catégorie sémantique « rivière ». Cette représentation est partielle, elle ne comporte pas les informations relatives au contexte de la rivière. [Figure 1 : An example of geographic information representing a river : its position on the Earth surface is represented by geometric co-ordinates and the real world entity is represented by a semantic category. This is a partial representation, it does not hold contextual information.]

Lorsque les objets représentés sont des champs, c'est-à-dire des phénomènes caractérisés par une valeur en chaque point de l'espace, comme la température ou le vent, leur représentation sous forme de données est faite en mesurant des valeurs de ces champs à certains points de l'espace. En d'autres termes, la représentation de champs se fait en échantillonnant ces champs grâce à une fonction d'interpolation propre au phénomène.

La représentation de la situation dans l'espace est en partie difficile parce qu'il faut que les composantes de la représentation décrites ci-dessus, représentation du monde réel et représentation de la position, puissent être manipulées de façon cohérente : que les manipulations permises sur les informations décrivant des objets du monde réel se traduisent par des manipulations cohérentes sur les informations de positionnement de telle sorte que le résultat sémantique soit associé à une position, et vice versa.

Une autre difficulté intervient lors de la numérisation des représentations des positions. En effet ces représentations sont souvent des formes géométriques dans un référentiel mathématique et leur traduction dans une représentation numérique demande nécessairement une discrétisation de cette forme (cf. Figure 2). Cette discrétisation peut être guidée par les modes d'acquisition de l'information. C'est le cas de la télédétection où le pas de discrétisation est la taille de la cellule du capteur. Cette discrétisation peut aussi être imposée par les modes de stockage. Par exemple il n'est pas envisageable de stocker une infinité de points décrivant une géométrie. Il faut donc discrétiser la géométrie en une représentation polynomiale ou l'approcher par une courbe connue et stockable.

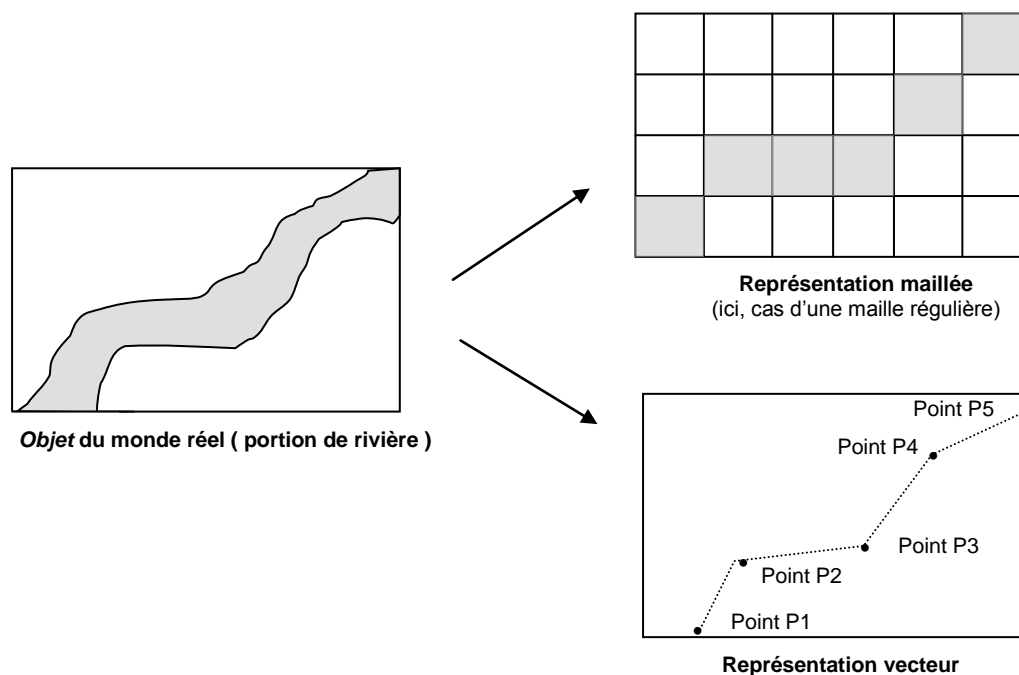


Figure 2 : Discretisations possibles de la géométrie de positionnement des objets du monde réel (ici une rivière) dans les bases de données géographiques. [Figure 2 : Possible discretisations of real world objects position geometry (here, a river) in a geographic database.]

La formalisation de représentations de l'espace géographique, i.e. la construction d'un modèle de données et le stockage effectif de données structurées selon ce modèle, pose de nombreux problèmes. La résolution de ces problèmes implique des choix plus ou moins arbitraires de la part du producteur de données.

Il existe un autre type de représentation portant sur l'espace géographique : les données localisables représentent des informations qui peuvent être attachées à une partie de la surface de la terre, comme la population d'une zone. Ces données peuvent appartenir à des domaines très divers. Une manipulation essentielle dans le traitement d'informations localisables est de représenter leur situation dans l'espace, c'est-à-dire de les *géo-référencer*. La notion de situation dans l'espace peut être représentée directement dans un système de référence si des coordonnées GPS de ces informations sont connues. Dans la plupart des cas, il est nécessaire de s'appuyer sur des données géographiques qui matérialisent l'endroit à la surface de la terre auquel ces informations sont rattachées. Des données localisables qui sont géo-référencées deviennent des données localisées, c'est-à-dire des données géographiques.

1.1.1.3 Les bases de données géographiques

Il y a quatre principaux media de stockage de l'information géographique :

- Le mot. «Nouns tend to describe stable, bounded entities, i.e. nouns are a form of spatial representation primitive »* [Raper 96].
- Le dessin : plans, cartes
- La photographie : photographies aériennes, photographies satellites.

* Les mots doivent généralement décrire des entités stables et limitées, c'est-à-dire que les mots sont une sorte de primitive de représentation spatiale. (tda)

- Les bases de données numériques : e.g. les bases topographiques décrivent les objets naturels ou construits par l'homme que l'on voit sur la Terre, les bases géodésiques décrivent des points identifiables sur Terre et dont on connaît les coordonnées dans un référentiel mathématique utilisé pour mesurer la Terre, les bases administratives décrivent des divisions de l'espace par l'homme en unités administratives, les bases numériques issues de la télédétection rendent compte des fréquences perçues par les capteurs en fonction pour chaque partie de Terre discriminée par ces capteurs.

Dans la suite de cette section nous nous concentrons sur les bases de données géographiques numériques.

Les représentations utilisées

Les représentations de données géographiques se rangent en deux classes selon l'importance relative accordée à la représentation des éléments du monde réel et la représentation des positions :

- classe 1 : ces représentations se concentrent sur la représentation d'entités sémantiques qui sont ensuite localisées dans un système de référence. C'est par exemple le cas des mots, des cartes.
- classe 2 : ces représentations sont axées sur la représentation de positions dans l'espace, auxquelles est attachée ensuite une information sémantique. C'est par exemple le cas des photographies.

Ces classes correspondent à divers mode d'acquisition de l'information et de production des données.

Les données numériques correspondant à la classe 1 sont les bases de données vecteurs. Un objet « sémantique » (représentation d'un objet du monde réel) a souvent un lien vers un objet géométrique représentant sa position, comme c'est le cas dans le SIG Arc/info. Les informations géométriques y sont représentées généralement à l'aide de primitives géométriques qui peuvent être combinées pour former des objets géométriques complexes. Notons que la conception de telles bases, décomposition en objets et choix des catégories d'objets, est largement inspirée de la carte, d'autant plus que l'essentiel des données existantes résulte de la numérisation de cartes [Brassel and Weibel 88].

Les données numériques correspondant à la classe 2 sont les bases de données maillées (en anglais raster) habituelles. Leur schéma est extrêmement réduit. L'élément de base de la représentation est le pixel qui correspond à une zone à la surface du sol. L'information sémantique est représentée comme un attribut du pixel.

La production de données géographiques

Les données géographiques pourraient être produites au cas par cas par les personnes devant manipuler des informations géographiques ou géo-référencer des informations. Cela aurait pour avantage que l'utilisateur représenterait exactement les informations géographiques qui l'intéressent, et de façon la plus appropriée pour géo-référencer ses propres informations.

Mais il y a de nombreux obstacles à cela. La constitution de bases de données géographiques et leur entretien demandent un matériel d'acquisition des données, des capacités de stockage importantes, des logiciels spécifiques, de même que du temps de saisie et une expertise. De plus, la qualité des données géographiques est un critère important pour la réussite d'un système d'information de ce type. « La gestion de la complexité implique l'information, et l'encombrement croissant du territoire appelle une information géographique fiable et détaillée » [Didier 90] p2. Or cette qualité demande une compétence et implique un surcoût.

Il est donc préférable d'utiliser, autant que possible, les BDG existantes plutôt que de construire soi-même une BDG pour une application particulière.

Ceci est d'autant plus vrai qu'un utilisateur ayant utilisé une BDG existante pour géo-référencer ses données peut potentiellement corréliser ses données avec celles d'autres utilisateurs ayant utilisé la même BDG source pour géo-référencer leurs propres données.

Ainsi il y a deux contextes de production de bases de données géographiques.

Une base de données géographiques peut être produite pour une application particulière nécessitant des données géographiques (spécifiques et non disponibles par ailleurs).

Des bases de données sont produites par des instituts (nationaux) de cartographie, comme l'IGN, soit pour produire des cartes papier, soit pour les vendre à des utilisateurs qui vont construire un système d'information autour. La mission essentielle de ces instituts est de produire des bases dites de référence.

Les données de référence permettent de localiser des phénomènes du monde perçu. De tels phénomènes peuvent être décrits par des données déjà contenues dans la base ou par des données d'un utilisateur.

Ces mêmes données de référence doivent permettre de passer d'une localisation dans un système à une localisation dans un autre. Ces bases de données peuvent être classées en :

- des bases sur l'information topographique
- des bases géodésiques
- des bases de nivellement
- des bases administratives, par exemple le plan cadastral de la Direction Générale des Impôts décrivant les éléments de références de la propriété foncière.
- des bases d'adresses postales géo-référencées
- des images numériques issues de la photographie aérienne, de la télédétection spatiale, et des orthophotos, référencées.
- ...

1.1.2 L'exploitation des données géographiques

Cette section détaille l'exploitation des données géographiques : les contextes d'utilisations de ces données (section 1.1.2.1), les techniques de manipulation de ces données dans les utilisations (section 1.1.2.2), et les outils disponibles pour manipuler les données (section 1.1.2.3).

1.1.2.1 Les contextes d'utilisation

La production de données géographiques

Un cas à part d'utilisation de données géographiques est la production de données géographiques. C'est une utilisation spécifique dans la mesure où elle nécessite des traitements d'analyse spatiale plus complexes que la plupart des autres utilisations. Nous présentons ici deux exemples d'utilisations de données géographiques à l'IGN dans le contexte de la production d'une BDG ou de la gestion de BDG.

Le premier exemple est la construction d'une base de données à partir de sources hétérogènes la BDPays®, Figure 3. Cela demande l'acquisition et l'intégration des données provenant des sources diverses ainsi que l'enchaînement de processus pour structurer et stocker une base finale.

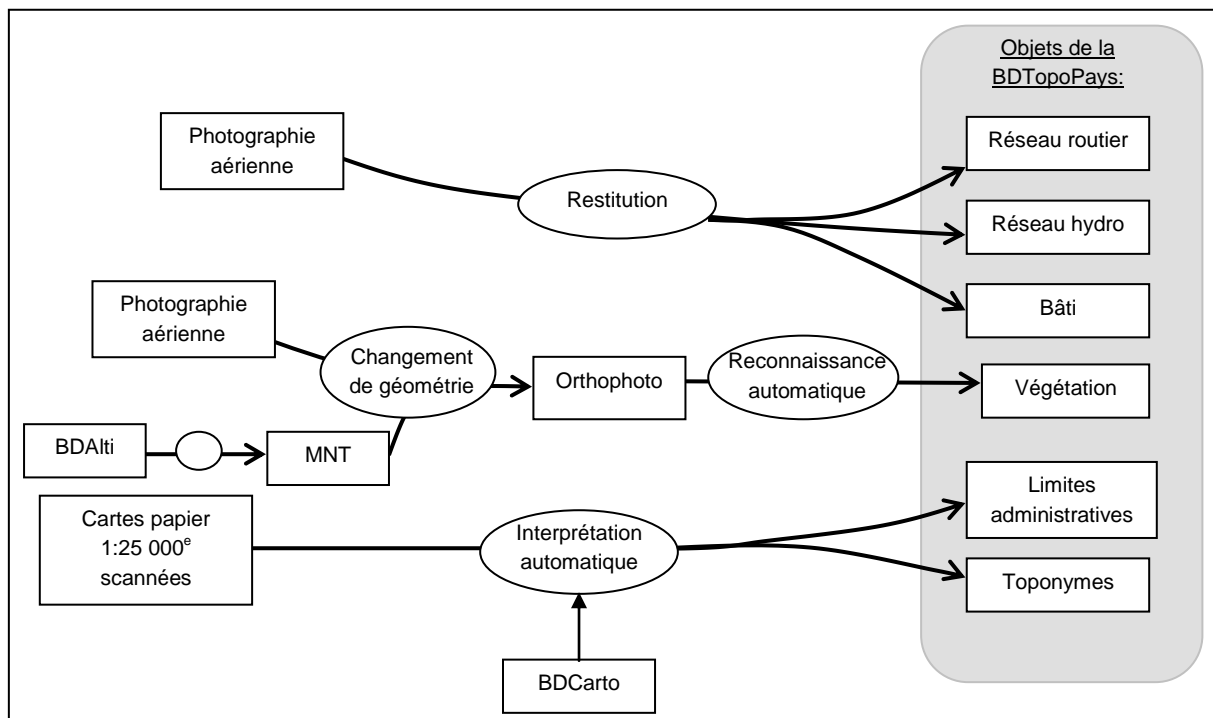


Figure 3 : La construction de BDPays®. [Figure 3 : The construction of BDPays®]

Un autre exemple d'utilisation de données géographique par un producteur de données comme l'IGN est l'intégration d'information entre plusieurs bases de données. Il y a en effet à l'IGN plusieurs BDG et non une seule de laquelle seraient dérivés tous les produits. Ceci pour plusieurs raisons :

- Les délais d'acquisition et de mise à jour sont différents selon les types d'échelle des BDG. Par exemple, la BDTopo® n'est pas encore achevée.
- La généralisation, processus de dérivation d'une BDG de petite échelle depuis une BDG de grande échelle qui implique de simplifier et modifier les objets, est très chère et son automatisation n'est pas encore maîtrisée.

Ces BDG décrivent, à diverses résolutions, la même réalité. L'IGN doit donc maintenir la cohérence entre ces représentations ; la modification d'une donnée dans une base a des répercussions sur les données représentant la même réalité dans d'autres bases, et sur les données représentant une réalité dépendante. Par exemple, l'ajout d'un tronçon de route sur un réseau peut impliquer l'ajout de carrefours.

Il existe plusieurs cas d'utilisations conjointes des diverses représentations, dont certains contribuent au maintien de la cohérence, cf. Figure 4:

- utiliser plusieurs BDG dans une application multi-échelles, comme la navigation embarquée,
- utiliser une BDG pour guider la fabrication d'une base plus détaillée, comme la BDCarto pour créer la BDPays®,
- contrôler la qualité d'une base à l'aide d'une autre base [Laurini et al. 98],
- mettre à jour une base grâce à une autre, comme la base routière Route500 à partir de la base BDCarto®,
- intégrer des bases pour en constituer une nouvelle, comme la base routière Route120 à partir de la BDCarto® et la BDAdmin.

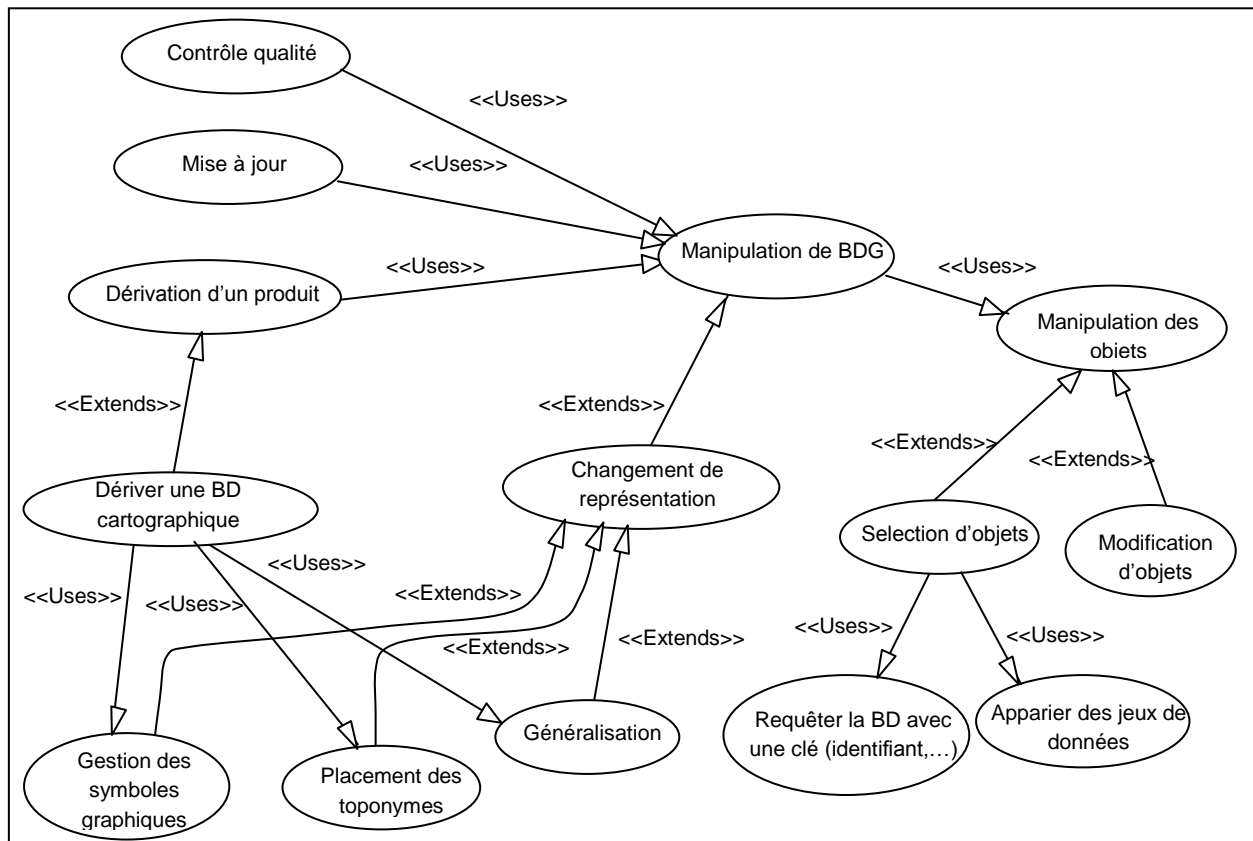


Figure 4 : Cas d'utilisations de ses BDG par l'IGN (formalisme UML). [Figure 4 : Use cases of its own GDB by IGN]

La dérivation de connaissances géographiques

Nous parlons ici des utilisations finales des données géographiques et non des utilisations qui consistent à construire un service de dérivation d'information pour des utilisateurs finaux potentiels. Par exemple nous parlerons d'une application de calcul d'itinéraire et non d'une application consistant à construire un système effectuant des calculs d'itinéraires.

Les applications des données géographiques sont classiquement résumées en une liste de questions élémentaires auxquelles ces applications apportent des réponses. Les deux principales questions sont [Frew et al. 95] :

- Où.. ? : La localisation. Exemple : où se trouve la Tour Eiffel ?
- Qu'y a-t-il là... ? : L'adressage. Exemple : quels sites remarquables y a-t-il dans ce pays ?

D'autres questions, classiquement identifiées dans la littérature [Maguire 91] et [Denègre et Salgé 96], sont :

- Et si.. ? : La simulation. Exemple : Si Paris n'existait pas, quels seraient les flux entre les principales villes françaises ?
- Quand.. ? Qu'est-ce qui a changé ? : L'analyse du changement. Exemple : quand a-t-on construit ce pont ?
- Comment.. ? : L'analyse spatiale. Exemple : Comment puis-je me rendre là ? Comment est structurée (géographiquement) cette ville ?

Des données géographiques sont nécessaires pour répondre à ces questions non pas systématiquement mais dans le cadre d'activités spécifiques : « la prévision, la planification et l'aide à la décision, [...] la gestion, l'inventaire et le suivi » [David 91].

Il existe une variété de domaines d'applications. A un domaine correspond souvent un ordre d'échelle des applications, qui respecte la résolution des objets significatifs de ce domaine, par exemple un ouragan dans le domaine de la météorologie et une maison dans le domaine du courrier postal.

Les applications à grande échelle sont de deux types [David 91]:

- applications de planification et de gestion du territoire (cadastre, urbanisme, statistique), des réseaux de transports (voiries, transports collectifs, ...), de distribution (eau, électricité, télécommunication, courrier postal,...) et d'assainissement,
- applications pour la réalisation d'équipements ou d'installations (réseaux routiers, oléoducs, gares, aéroports...).

Les applications à petite et moyenne échelles correspondent à des applications plus thématiques : agriculture, économie, tourisme, administration, démographie, climatologie, militaire, environnement. Enfin, les applications de prévision utilisent généralement des données à plus petites échelles qui résument une situation ou une organisation.

D'autres communautés d'utilisateurs sont rappelées dans [Hanigan 90] : le grand public, l'enseignement, la recherche, l'édition.

En définitive, les BDG ne sont pas exploitées par des requêtes classiques. L'information géographique dont a besoin un utilisateur est rarement explicitement représentée dans une BDG. L'exploitation de données géographiques implique des processus de dérivation. Les applications se distinguent par leur domaine, c'est-à-dire les informations thématiques associées aux données géographiques de référence, par le type de produit réalisé, et par les données mêmes c'est-à-dire la portion d'espace géographique à laquelle se rattache le problème de l'utilisateur.

1.1.2.2 Les techniques

Les utilisations décrites dans la section précédente mettent toutes en œuvre des techniques de manipulation de l'information géographique que nous rappelons ci-dessous.

Le géoréférencement : indexer, intégrer

Une manipulation répandue est le géoréférencement d'informations localisées, c'est-à-dire la représentation de la situation dans l'espace d'informations portant sur l'espace géographique. Nous appelons « empreinte » la partie de la surface de la terre à laquelle est attachée une information. Les principales raisons poussant à géo-référencer des données, sont :

- Les empreintes respectives de données géo-référencées de natures diverses peuvent servir à corréler ces données. Cela renvoie à une fonction essentielle des données géographiques : servir de référentiel.
- Les données géographiques ont un espace naturel d'indexation à deux dimensions, voire trois. Cela permet de réduire la complexité de présentation des indexes, de même qu'un tableau de N cases est plus lisible qu'une énumération de N cellules.
- L'empreinte géographique donne accès au contexte spatial de l'information géo-référencée.

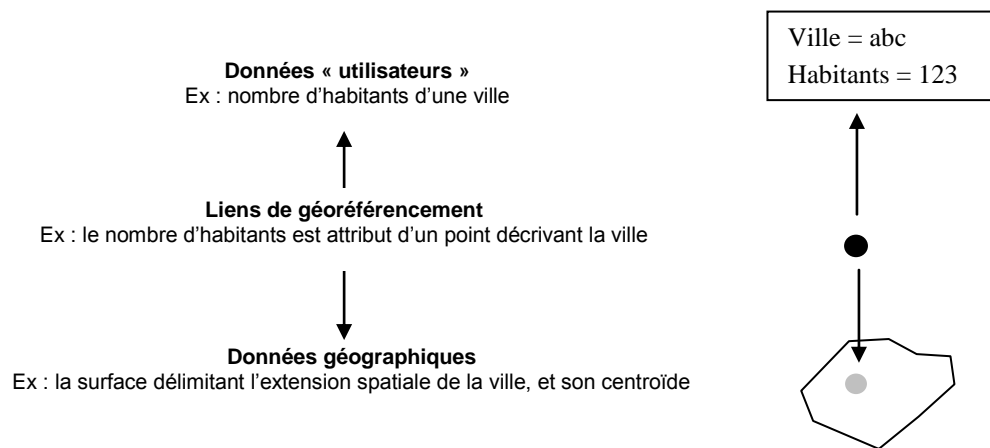


Figure 5 : Le géoréférencement de données : les données « utilisateurs » décrivent une information indirectement rattachée à une partie de la surface de la Terre, les données géographiques décrivent une partie de la surface de la Terre, les liens rattachent une information à son empreinte géographique. [Figure 5 : Data georeferencing : users data describe an information indirectly related to a portion of the earth surface, geographic data describe a portion of the earth surface.]

Nous donnons ci-dessous quelques définitions importantes concernant le référencement spatial. Un système de référence spatiale peut être de deux types : direct ou indirect. Un système de référence direct est un système mathématique rattaché à la terre généralement par la définition des coordonnées attribuées au centre de la terre. Un système de référence indirect est un système attaché à des objets du monde réel. Par exemple les systèmes linéaires rattachés au réseau routier ou une coordonnée est un point kilométrique.

La notion de référentiel spatial est plus large que celle de système de référence spatiale. Un référentiel spatial comprend :

- Le système de référence, méthode de localisation d'une information sur la surface de la terre,
- Les données de références, ensembles d'objets particuliers dont la localisation est connue, et qui peuvent être identifiés et actualisés aisément.

Des caractéristiques importantes d'un référentiel spatial sont : le degré de précision offert pour la localisation et l'universalité de son emploi.

La cartographie : représenter et communiquer

La représentation graphique de données géographiques est souvent la base de l'appréhension par l'homme des informations contenues dans ces données. La représentation cartographique permet de visualiser et d'analyser des données géographiques dans leur ensemble. Elle permet d'appréhender différents niveaux d'analyse de l'espace et des relations entre les objets non explicites dans la représentation, comme le voisinage.

Dans ce domaine, les travaux de référence sont ceux de Bertin, [Bertin 67]. L'auteur analyse les variables graphiques utilisées pour représenter des objets et le rôle de chaque variable dans la perception visuelle de relations entre les objets et catégories représentées Figure 6.

Variable \ Information perçue	Différentiation	Association	Ordre	Quantité
Forme	XX			
Taille		XX	XX	XX
Orientation	XX	X		
Couleur	XX	XXX		
Valeur			XXX	
Texture	X	XX	X	

Figure 6 : Variables graphiques définies par [Bertin 67] et leur apport dans la perception visuelle de relations entre les objets représentés. [Figure 6 : Graphical variables defined in [Bertin 67] and their input to the visual perception of represented objects.]

Selon Bertin : « la graphique est d'abord une mémoire artificielle puissante, susceptible de classements, de catégorisations, de manipulations diverses, et qui permet de découvrir soi-même ce qu'il y a à dire. »

L'analyse spatiale : enrichir

L'analyse spatiale est une activité essentielle pour la cartographie et les prises de décision.

Dans une BDG de nombreuses informations ne sont pas explicitement représentées. Typiquement, il peut s'agir d'informations reconstruites par l'œil humain lors de la lecture d'une carte, comme la relation entre objets « est proche de ».

La reconstruction de ces relations doit respecter divers niveaux d'analyse correspondant aux diverses échelles significatives des objets et phénomènes sur lesquels se font les raisonnements.

L'analyse spatiale vise à expliciter ces relations. Elle est une des principales utilisations de données géographiques : « l'information géographique parce qu'elle permet de modéliser et d'analyser les relations spatiales entre les multiples contraintes humaines, sociales, économiques, naturelles est un outil essentiel de maîtrise de cet équilibre » [AFIGEO 98].

La géographie : intervenir

Il est important de rappeler que des techniques existent pour raisonner dans l'espace géographique, autres que celles qui prévalent dans la plupart des systèmes d'information géographique actuels.

Les géographes s'attachent à trouver des relations et des lois géographiques. Ils développent des modèles permettant de guider l'intervention de l'homme dans son paysage géographique autrement que par la seule description de ce paysage et une analyse de ses équilibres [Claval 82]. Prenons l'exemple du positionnement d'une frontière dans un paysage traversé par une rivière ou une chaîne de montagnes. Une analyse physique pourrait chercher à faire passer la frontière le long de la rivière ou de la chaîne de montagne, c'est-à-dire à s'appuyer sur une « limite » naturelle. L'analyse menée par un géographe envisagerait un grand nombre de frontières non nécessairement inscrites dans le paysage et analyserait les pôles ainsi créés, c'est-à-dire les deux pays et leurs équilibres [Grasland 92].

1.1.2.3 Les outils : les SIG

Ce qui suit est largement issu du cours sur l'Echange de Données Géographiques conçu pour les élèves de l'Ecole Nationale des Sciences Géographiques par Jean-Yves Priou et Didier Richard. Des éléments sont aussi repris du cours sur les SIG conçu par Anne Ruas pour le même cycle, et des cours de Jean-Paul Donnay et Roland Billen de l'Université de Liège.

Généralités

Les données géographiques représentent des relations géométriques qui ne sont pas prises en compte par les modèles de données traditionnels et qui ne peuvent être exploitées par un SGBD classique. De plus l'exploitation de telles données ne consiste pas uniquement à faire des requêtes sur des informations signalétiques mais elle implique aussi d'effectuer d'autres traitements d'analyse et de présentation. Aussi, il existe des logiciels particuliers pour accéder aux données géographiques et les traiter : les systèmes d'information géographiques (SIG).

La catégorie des logiciels de traitement de l'information géographique numérique est plus large que la catégorie des SIG, elle comprend aussi des logiciels dédiés au traitement d'image, des logiciels métier comme ceux utilisés par le cadastre ou certaines communautés urbaines. Nous nous concentrons sur les systèmes d'information géographique, SIG, c'est-à-dire les systèmes permettant :

- d'accéder à des données géométriques et alphanumériques,
- de faire des requêtes spatiales plus ou moins évoluées, croisées avec des requêtes sémantiques.

Parmi les SIG, on distingue généralement les logiciels professionnels fournissant des fonctions de traitements avancées (ex: ArcInfo, Intergraph MGE, Geomedia, Smallworld, Lamps), les logiciels "bureautique" moins sophistiqués mais plus simples d'emploi (ex: MapInfo, ArcView, Geoconcept), les composants et les extensions dédiés par exemple à la publication de données sur le Web (ex : Geoconcept explorer, ArcIMS) et, plus récemment, les serveurs de base de données (ex : Oracle Spatial, Data Blade Informix, Spatial Data Engine d'ESRI).

L'architecture classique d'un SIG a longtemps été décomposée en un SGBD classique (par exemple la partie INFO des anciennes versions d'ARC/INFO) qui gère les informations alphanumériques, dites informations sémantiques, et un moteur spatial qui gère les informations géométriques, dites spatiales, (par exemple la partie ARC des anciennes versions d'ARC/INFO). Ce moteur spatial doit stocker les données géométriques et éventuellement topologiques, et y accéder. Cet accès se fait souvent par des arbres balancés (Quad-tree, r-tree). Le moteur spatial effectue également des traitements plus complexes de dérivation d'information exploitant la nature spatiale des données. Ce moteur, outre le fait de gérer l'information géographique, permet, dans une certaine mesure, une cohérence entre la manipulation des deux types d'informations (géométriques et alphanumériques). Certains SIG assurent un couplage fort entre données géométriques et sémantiques. Ces systèmes maintiennent automatiquement la cohérence entre les manipulations faites sur les données d'un type et les données d'un autre type et sont par exemple efficaces lors de mise à jour des données. D'autres SIG assurent un couplage faible. Ils permettent des requêtes croisées mais ne maintiennent pas automatiquement la cohérence lorsque des données sont modifiées. Les puristes réservent l'appellation SIG aux systèmes assurant un couplage fort et appellent moteurs spatiaux les SIG assurant un couplage faible. [Zeitouni et Yeh 00] proposent un parallèle entre les SGBDR classiques et les moteurs spatiaux, Figure 7.

	RELATIONNEL	SPATIAL
Données	Entier, Réel, Texte,...	Plus complexes : Point, Ligne, Région, ...
Prédicats et calculs	Tests : =, >, ... Calculs : +, /, ... Et fonctions simples	Prédicats et calculs géométriques et topologiques Tests : intersecte, adjacent à, ... Fonctions : intersection, surface,...
Manipulation	Opérateurs de l'algèbre : sélection, projection, jointure Agrégats : count, sum, ...	Manipulation mono ou inter-thèmes Sélection et jointure sur critère spatial Agrégats : fusion d'objets adjacents
Liens entre objets	Par clés de jointures	Relations spatiales (souvent) implicites
Méthodes d'accès	Index B-tree, hachage	Index R-tree, Quad-tree, Grid-file, etc.

Figure 7 : Parallèle entre les SGBD spatiaux (moteur spatiaux) et les SGBD relationnels, extrait de [Zeitouni et Yeh 00]. [Figure 7 : Parallel between spatial DBMS and relational DBMS, after [Zeitouni et Yeh 00]]

Cette situation change, comme le soulignent [Zeitouni et Yeh 00] et [Bédard et al. 00], et actuellement les données spatiales sont de plus en plus intégrées aux données non spatiales. L'architecture des systèmes est donc plutôt composée d'un serveur de données spatiales et non spatiales, au couplage plus ou moins fort, et de services qui se greffent sur ce serveur pour fournir aux clients logiciels les accès nécessaires.

Le SGBD Oracle a ainsi évolué peu à peu pour prendre en compte les données spatiales, [Riedo et Mottier 02] :

- en 1995, la version 7.1.6 intègre les points,
- en 1997, la version 7.3.3 intègre les points, les lignes et les polygones,
- en 1998, la version 8.0.4 intègre des opérateurs et index spatiaux qui améliorent les performances de 40%
- en 1999, la version 8i intègre les cercles et les arcs, et propose outre un modèle relationnel pour stocker les informations spatiales, un modèle objet.

Un autre exemple d'évolution est celle du logiciel ArcInfo qui était au début composé d'un module de gestion des données spatiales vecteur (le module ARC) et d'un module de gestion des informations sémantiques (le module INFO ou un autre SGBD). ArcInfo propose maintenant un module appelé ArcSDE qui est un serveur de données spatiales et sémantiques.

Nous détaillons ci-dessous la capacité des SIG à modéliser des informations et à traiter ces informations.

La modélisation dans les SIG

Une caractéristique essentielle d'un SIG est sa capacité de modélisation pour spécifier des objets utilisateurs, ce qui peut signifier définir des objets géographiques de référence, définir des objets utilisateurs, et définir des liens entre les objets utilisateurs et les objets géographiques.

Les principales capacités de modélisation nécessaires pour représenter des objets géographiques sont résumées dans le tableau ci-dessus. Il reprend les types définis dans la norme GML2 d'échange d'objets géographiques et détaille si ces types sont supportés ou non par les principaux SIG commercialisés. Nous nous focalisons sur les modèles vecteurs car les modèles raster (représentation par champs) sont extrêmement élémentaires.

Une caractéristique importante d'un SIG est la modélisation de la topologie. Certains SIG ne proposent pas de modèles topologiques explicites mais offrent cependant des outils d'analyse de la connectivité, comme Geomedia.

Nom et type du système →	Lamps2	ArcInfo 8 (et ArcView 8)	Geomedia	MapInfo	Geo-concept	Data Draw Cube	Small World	Mapguide
Eléments de modélisation supportés ↓								
Le point défini par un couple de coordonnées	X	X	X	X	X	X	X	X
Le point défini par un triplet de coordonnées	X	X	X		X		X	
La polyligne	X	X	X	X	X	X	X	X
La polyligne fermée	X				X	X	X	X
La fenêtre, rectangle défini par une paire de points	X			X				
La surface	X	X	X	X	X	X	X	X
La surface à trous	X	X	X	X	X	X	X	X
Collection géométrique composée de plusieurs primitives géométriques	X		X					
Collection de points	X	X	X				X	
Collection de lignes	X		X	X		X	X	X
Collection de surfaces	X		X	X	X	X	X	X
Les types d'objets sans géométrie	X	X		X			X	
Les attributs et relations dont la valeur peut être un élément géométrique	X							
La topologie	X	X				X	X	
Un objet avec plusieurs géométries	X					X		

Figure 8 : Tableau comparant les capacités de modélisation de certains SIG, vis-à-vis des types définis dans la norme GML2.

Les auteurs [Lopes et al. 97] précisent que la fonctionnalité de modélisation reste très peu développée dans la plupart des SIG. Les modèles SIG sont souvent dédiés à la cartographie, qui n'est qu'une première exploitation possible de données géographiques. Un atout certain concernant la modélisation est de supporter la modélisation objet. TIGRIS d'Intergraph a été un précurseur mais a été retiré du marché à cause de son manque de robustesse. Smallworld est également un des premiers SIG orienté-objet. C'est un SIG offrant d'excellente capacité de modélisation, mais il a longtemps été handicapé par de longs temps de traitement. Lamps2 est un SIG orienté-objet plus récent, utilisé à l'IGN. Il offre des capacités de modélisation extrêmement évoluées permettant d'attacher des comportements à un objet modélisé par l'utilisateur. Enfin, le serveur Oracle peut être associé à une sur-couche objet.

Les fonctions de modélisation de base d'un SIG devraient comprendre également les changements de système de coordonnées, les changements de structure. Les traitements de changement de structure sont essentiellement la descente de structure vers la géométrie et la remontée de structure vers la sémantique. La descente de structure consiste à passer d'un modèle comprenant des objets complexes (objets dont la géométrie est une collection de géométries) à un modèle comprenant des objets simples, ou des modèle comportant des objets simples à un modèle ne comportant que des géométries.

Lors d'une descente de structure, l'information est conservée mais n'est plus représentée de la même façon. Par exemple, dans le modèle à objets complexes, certains objets (par exemple une route) sont composés de plusieurs objets simples (par exemple un tronçon de route). Après descente de structure, le modèle ne comporte plus que des objets simples mais des attributs sont ajoutés aux objets simples pour conserver les informations précédemment décrites au niveau de l'objet complexe.

Les capacités d'importation des données sont associées aux capacités de modélisation du SIG et aux formats d'imports prévus par le concepteur. Une caractéristique importante est la modélisation de la topologie.

Les capacités d'export varient selon que le SIG autorise ou non l'export de données calculées par le SIG. Certains SIG permettent ainsi de calculer des relations topologiques mais ces relations ne peuvent être ensuite récupérées dans un fichier pour les réutiliser ensuite.

Le niveau de paramétrage des fonctions d'import/export est rarement suffisant. Un paramétrage efficace devrait permettre de filtrer les données importées par une expression sur les valeurs d'attributs. C'est important dans le contexte de données géographiques ; un jeu de données représente souvent un volume important de données et il peut être intéressant de le réduire aux données nécessaires pour le traitement. Par exemple Geomedia est un rare SIG à posséder une option de filtrage géographique qui permet de filtrer un jeu importé en ne gardant que les données portant sur une certaine zone.

Les traitements dans les SIG

Les SIG offrent plusieurs fonctionnalités. Une première fonctionnalité consiste à améliorer l'accès aux données géographiques dans la base, cela consiste à organiser des indexes spatiaux et traiter les requêtes dans ces indexes. Nous ne nous étendons pas sur ce type de traitement qui ne répond pas spécifiquement à un besoin de dérivation de données.

Les requêtes minimales proposées par un SIG sont ce qu'on peut appeler la localisation élémentaire et l'adressage élémentaire.

La localisation élémentaire consiste à accéder à la géométrie par la sémantique : "quelle est la localisation de cette commune?" (question "Où?").

L'adressage élémentaire, le traitement le plus spécifique des SIG, consiste à accéder à la sémantique par une requête sur la géométrie : "quel est l'objet situé à tel endroit?" (question "Quoi?"). Les fonctions correspondantes se distinguent d'un SIG à un autre, par exemple par des restrictions possibles sur l'élément sémantique recherché « quels sont les objets des classes c1, c2 et c3 situés à tel endroit ? ». Elles se distinguent aussi par le type de relation spatiale analysée, c'est-à-dire par la désignation de « situé à tel endroit ». Cela peut être l'inclusion dans une zone, tracée à l'écran d'une façon ou d'une autre selon le SIG, ou éventuellement désignée par un objet surfacique. Cela peut être l'intersection avec un objet géométrique, surfacique et parfois linéaire. Certains SIG proposent des requêtes plus avancées intégrant des opérateurs comme la distance maximale.

Il existe des requêtes plus poussées d'analyse spatiale ou thématique. De nombreuses approches proposent des classifications des requêtes et traitements SIG [Boursier et Mainguenaud 92][Goodchild 87][Cheylan 92][Maguire and Raper 90][Maguire and Dangermond 91]. Nous retenons les principaux traitements élémentaires relativement à l'analyse spatiale : la mesure de distance, la création de buffer (autour d'objets ponctuels, linéaires ou surfaciques), la superposition de couches, les opérateurs de manipulation de réseaux. Des traitements plus complexes sont construits à partir de ces opérateurs, comme le test d'inclusion, la mesure de direction, le calcul de graphe.

Peu de logiciels SIG possèdent des fonctions sophistiquées de généralisation. Les principaux sont CHANGE de Zeiss, et MGE/MG d'Intergraph [Ruas 99] p301. CHANGE est dédié aux grandes échelles. La généralisation sous MGE/MG est interactive et nécessite de la part de l'opérateur une expertise en généralisation, alors que CHANGE propose un processus de généralisation automatique. Les fonctions de généralisation disponibles dans les autres SIG commercialisés sont essentiellement une fonction de lissage et une fonction de filtrage. Il s'agit d'algorithmes de généralisation d'objets linéaires. Récemment, le projet européen AGENT (ESPRIT 24939) a proposé un prototype de système de cartographie automatique, construit sur le SIG Lamps2, intégrant des fonctions sophistiquées de

généralisation, fondé sur la technologie multi-agent [Lamy et al. 99][Barrault et al. 01][Duchêne et Ruas 01].

D'autres fonctionnalités importantes sont la présentation cartographique des données. Les SIG possèdent tous des fonctions élémentaires de présentation cartographique. Dans un contexte de production de cartes de qualité il est nécessaire d'acquérir des fonctionnalités avancées comme la gestion de couleurs transparentes, la gestion des écritures. De telles fonctionnalités sont dispensées par des modules spécifiques qui viennent s'ajouter à un noyau SIG, comme DATA DRAW pour Geoconcept.

La plupart des SIG proposent une interface de programmation permettant à l'utilisateur de coder ses propres fonctions, en utilisant des bibliothèques reprenant les fonctions énumérées ci-dessus. Ceci est souvent indispensable pour réaliser une application SIG adaptée au besoin de l'utilisateur. Grâce à ces interfaces, les SIG peuvent répondre à la plupart des besoins d'utilisateurs.

Il faut souligner que les divers modèles utilisées par les SIG pour effectuer des fonctions ne sont pas explicités. Une longue pratique d'un SIG permet à un utilisateur de faire des hypothèses sur ce modèle mais le modèle reste encapsulé dans le SIG. Par exemple, Geoconcept recalcule systématiquement la topologie pour effectuer des opérations sur réseaux.

1.1.3 L'accès à l'information géographique

Cette section présente les obstacles à l'accès des utilisateurs à l'information géographique. Nous distinguons plusieurs étapes dans l'accès : l'accès aux données, l'accès au contenu des données c'est-à-dire à leur interprétation et l'accès à la dérivation d'information c'est-à-dire à l'exploitation des données.

Nous n'abordons pas les facettes de cette problématique qui sont liées à des problèmes politiques et économiques comme les droits d'accès et les coûts. Signalons tout de même qu'un obstacle non négligeable à la distribution de données est l'évaluation de la valeur de ces données. Ce prix est souvent calculé en fonction du coût de production et non de leur valeur à l'usage.

1.1.3.1 L'accès aux données

Dans cette section nous détaillons ce qui supporte l'accès d'utilisateurs aux données géographiques.

Les métadonnées de transfert

Les travaux sur l'établissement de formats de transfert de données étudient comment enrichir le transfert avec des informations relatives à ces données et utiles pour recevoir, lire et traiter les données transférées. De telles informations sont par exemple la structure des données, Figure 9. Le terme *métadonnée* est utilisé pour désigner de façon générale les informations relatives à des données [Role 99].

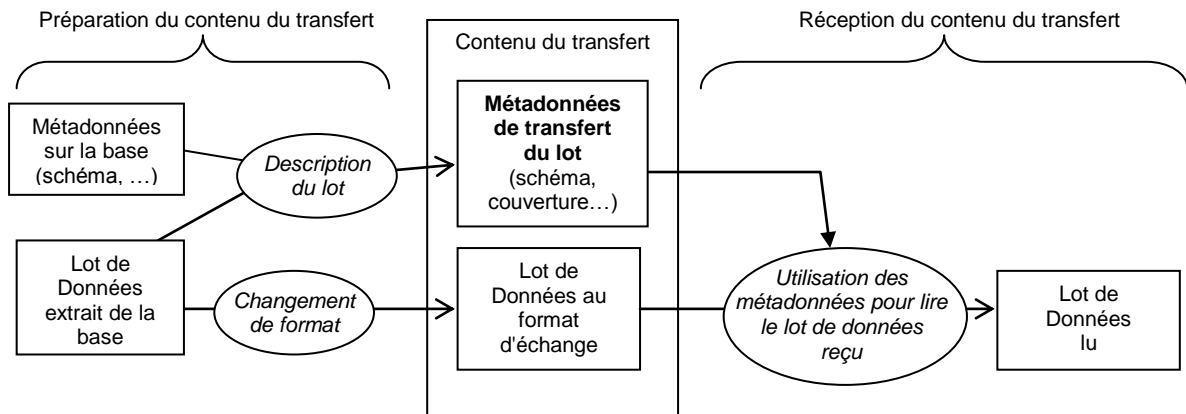


Figure 9 : Utilisation de métadonnées pour structurer des données transférées.

Par exemple, l'IGN utilise le format FEIV (Format d'Exchange Interne Vecteur) pour transférer ses données en interne. Les éléments de ce modèle sont décrits Figure 10.

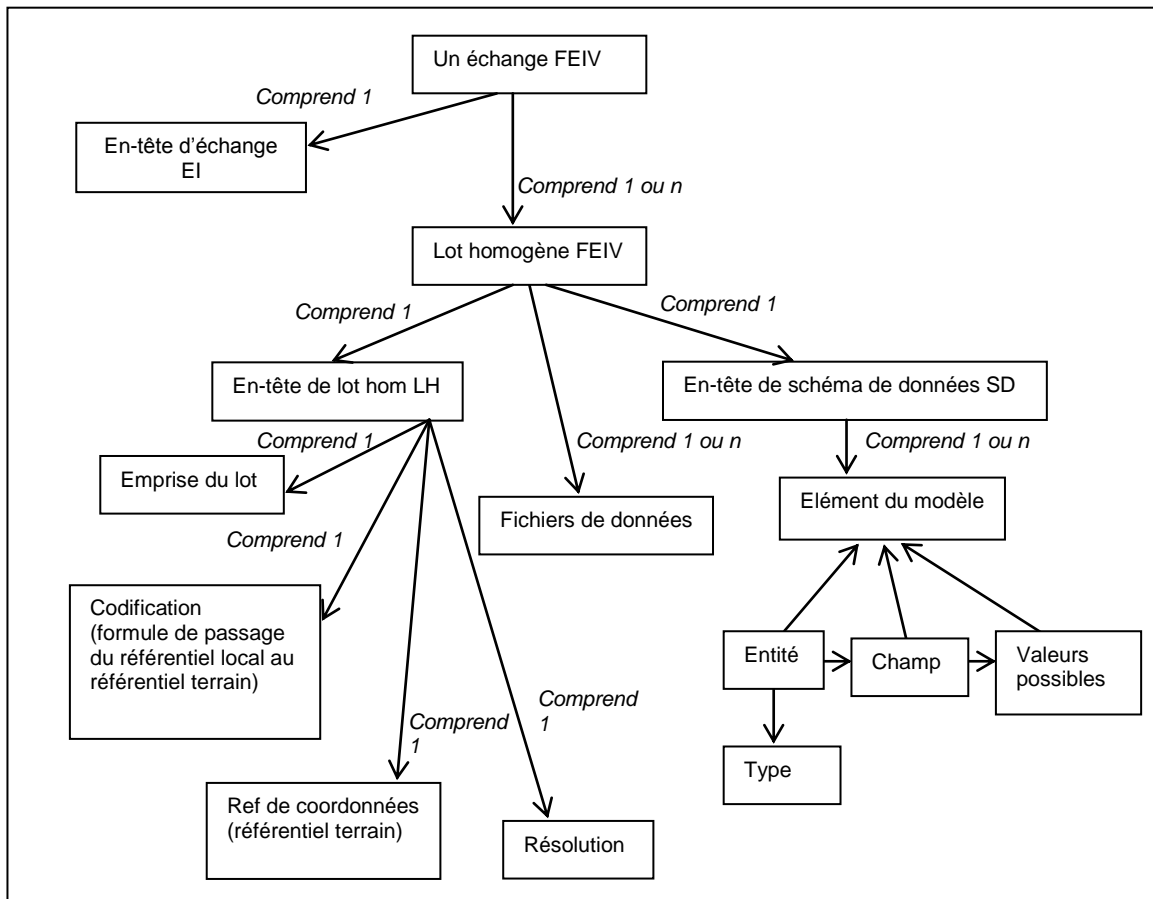


Figure 10 : Modèle FEIV utilisé au sein de l'IGN pour transférer des données. [Figure 10 : The FEIV model used inside IGN to exchange data.]

Les premiers travaux sur les métadonnées ont pris place dans le contexte du transfert de données. Depuis, des métadonnées sont également proposées dans un autre contexte, celui du catalogage de données, c'est-à-dire de l'accès aux données, Figure 11.

Les métadonnées de catalogage

La constitution de métadonnées de catalogage va de paire avec :

- la mise en place de standard de métadonnées, qui permettent de répertorier dans un même catalogue des lots de données hétérogènes,
- et la mise en place d'infrastructures d'information sur lesquels sont disponibles les données cataloguées.

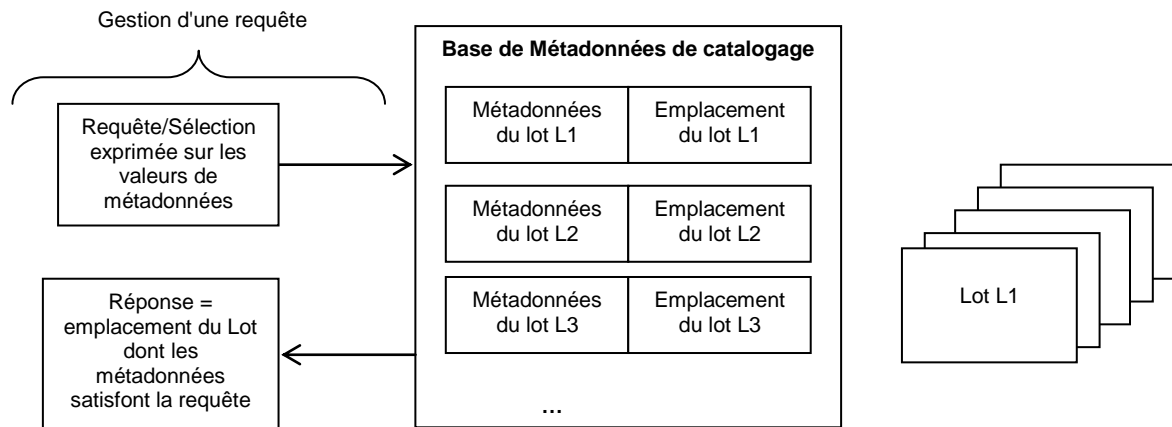


Figure 11 : Utilisation de métadonnées pour le catalogage de données.

Nous distinguons trois niveaux d'analyse des métadonnées de catalogage :

- les principes, ou une analyse des besoins en métadonnées,
- les modèles proposés qui satisfont plus ou moins bien ces principes,
- les métadonnées disponibles qui peuvent être des bases de données structurées selon les modèles standard de métadonnées ou des fichiers de données décrivant des jeux de données géographiques.

Les principes des métadonnées de catalogage sont largement étudiés par le groupe Global Spatial Data Infrastructure [GSDI/TWG 00]. Ils distinguent trois niveaux d'utilisation de métadonnées nécessitant des informations différentes :

- les *métadonnées de découverte* doivent permettre de répondre à la question : "quel jeu de données contient les données que je recherche?"
- les *métadonnées d'exploration* doivent permettre de répondre à la question : "le jeu de données identifié convient-il pour l'analyse que je veux faire?"
- les *métadonnées d'exploitation* doivent permettre de répondre à la question : "comment obtenir et utiliser les données que je veux"

Les auteurs précisent que les métadonnées de découverte sont organisées en plusieurs catégories :

- *quoi?* : le titre et la description du jeu,
- *pourquoi?* : les raisons pour lesquelles le jeu a été constitué et ses utilisations possibles,
- *quand?* : la date de création du jeu et les modalités de mise à jour à venir,
- *qui?* : le producteur des données, et les utilisateurs auxquels les données sont destinées,
- *où?* : la couverture du jeu de données,
- *comment?* : le protocole de production des données et comment on peut y accéder.

Les métadonnées d'exploration doivent détailler la description du jeu en intégrant par exemple des informations de qualité. Enfin, les métadonnées d'exploitation doivent contenir des informations encore plus spécifiques comme par exemple la projection utilisée, le schéma applicatif.

La constitution de normes de métadonnées géographiques fait l'objet de nombreux travaux [Günther and Voisard 97] [Luzet 98] [Spéry et Libourel 98]. Ces standards sont :

- la norme du FGDC, Federal Geographic Data Committee : FGDC Content Standard for Digital Geographic Metadata.
- la norme expérimentale du Comité Européen de Normalisation, CEN/TC287,
- la norme, en cours de définition, de l'organisation internationale de standardisation, ISO/TC211, repris par le consortium OpenGIS.

La première norme, reprise en partie par les suivantes, est FGDC/CSDGM. Les métadonnées se présentent sous la forme de champs associés à des valeurs. Les champs sont organisés en paquets :

- L'identification : description textuelle, couverture temporelle, couverture spatiale, mots-clés, personne à contacter, restrictions d'accès, informations techniques
- La qualité : précision, historique et complétude
- Les caractéristiques spatiales : mode de positionnement
- Le système de référence spatiale : projection et système de coordonnées
- Les entités et leurs attributs
- La distribution
- La référence des métadonnées

Ces standards ne comportent pas toute l'information recommandée dans les principes énumérés ci-avant, en particulier ils ne décrivent pas les utilisations des jeux de données. Ils ont été initialement développés dans le contexte de la gestion des données par le producteur, de leur transfert, et n'ont pas été véritablement améliorés pour traiter le contexte de la compréhension de l'utilisateur et l'accès à l'information [Lataillade 97]. En particulier, les métadonnées actuelles fournissent peu de description du sens des données, c'est-à-dire de leur lien avec la réalité [Hedorfer and Bianchin 98]. Cela est mis en évidence par exemple dans le travail de [Spéry et Libourel 98]. Les auteurs catégorisent le contenu des métadonnées selon trois niveaux d'abstraction ne dépendant pas de la compétence technique des utilisateurs :

- Les Ontologies (base de connaissances décrivant un domaine donné).
- Les métadonnées relatives au protocole (i.e. à la production des données) et au schéma (l'organisation des données dans le produit).
- Les métadonnées portant sur les données elles-mêmes.

Leur originalité par rapport aux autres approches est d'introduire un niveau ontologique qui est l'élément-clé pour l'accès au sens des données. Dans les grands standards, [Spéry et Libourel 98] étudient la part des ontologies. Elles sont utilisées pour renseigner la métadonnée *identification* des normes FGDC et ISO, et la métadonnée *classification* de la norme CEN. Les ontologies décrivant les valeurs possibles de ce champ (*identification* ou *classification*) sont des nomenclatures. Une nomenclature est un ensemble de catégories, chacune étant associée à une définition explicite des éléments qu'elle contient, et de liens entre ces catégories du type « *est une sorte de* ». Nous verrons par la suite que ces types d'ontologies ont une efficacité encore limitée dans une optique d'aide au partage de connaissances, 2.1.3.2.

Un autre défaut de ces standards est leur complexité. Ils peuvent en effet présenter jusqu'à 200 champs. Cela freine leur utilisation pour décrire les BDG existantes. Certains travaux comme ceux de [Hedorfer et Bianchin 98] cherchent à faciliter l'implémentation des standards de métadonnées en automatisant une partie de la production des métadonnées, c'est-à-dire du renseignement des divers champs sur une base donnée.

Par ailleurs cette complexité rend leur exploitation par un utilisateur difficile. D'une part il lui faut parcourir de nombreux champs, d'autre part les informations décrites dans ces champs ne sont pas toujours interprétables par l'utilisateur. Un système de catalogage qui reposerait sur ces standards de métadonnées devrait nécessairement offrir à l'utilisateur une navigation simplifiée dans ces champs et des interfaces conviviales pour exprimer des requêtes sur les valeurs de ces champs.

Soulignons enfin une limite importante des métadonnées géographiques actuelles. Le lien entre une métadonnée et des données se situe entre la métadonnée et la base de données. Or ce qui devrait compter dans la caractérisation d'une métadonnée ce sont les liens qu'elle entretient avec les données et plus précisément les liens permettant de sélectionner des données utiles. En effet de telles métadonnées permettent l'exploration d'un ensemble de bases de données mais non pas des portions sélectionnables d'une même base de données. Typiquement, les métadonnées comme la couverture (ou emprise) et le schéma pourraient être associées à des opérateurs qui permettent de les spécifier (sélection d'une couverture plus petite, sélection d'éléments du schéma) pour sélectionner un lot de données.

Pour ce qui est des métadonnées disponibles, le FGDC doit obligatoirement être utilisé par les producteurs de données américains pour décrire leurs données et il existe donc des métadonnées effectives. Par contre il n'existe pas de telle obligation en Europe. Le Geographical Data Description Directory, GDDD, a été créé en simplifiant le standard CEN pour décrire effectivement les principales bases de données des différentes agences de cartographie en Europe. Actuellement le GDDD décrit plus de 300 produits de 36 agences différentes. Il comprend les rubriques suivantes :

- Aperçu : une description textuelle de la ressource, avec les contacts utiles
- Informations commerciales : couverture géographique, conditions de vente, modalités d'usage, formats...
- Information techniques : spécifications, sources des données, paramètres de qualité
- Organisme : description de l'agence productrice

Par ailleurs, il existe de nombreux fichiers constitués par le producteur de données pour gérer ses données et qui sont de fait des métadonnées décrivant les jeux de données gérés. Ces fichiers décrivent par exemple les extensions des différentes cartes d'une même série. Ces métadonnées sont complexes à acquérir car leur existence n'est pas toujours connue et les personnes qui les produisent ne prévoient pas leur réutilisation comme métadonnées de catalogage. De plus l'utilisation de telles métadonnées se heurtent à la multiplicité et au manque de stabilité des représentations utilisées pour les stocker. Ces fichiers sont en effet de fichiers de petite taille, stockant peu de données et ils ne font pas l'objet d'une structuration définie. Lorsque le producteur modifie son fichier de gestion de données, il peut en modifier la structure s'il souhaite par exemple rajouter une information.

D'autres informations pouvant servir de métadonnées sont disponibles uniquement sous forme papier, comme par exemple les spécifications d'une BDG.

Pour conclure, l'utilisation de métadonnées dans un contexte d'aide à l'accès est encore au stade de principes. Ces principes ont l'avantage de faire l'objet d'un certain consensus mais ils sont encore peu associés à des indications sur la façon de modéliser les métadonnées en question. Il n'y a pas de lien entre ces principes et les normes de métadonnées.

Les normes de métadonnées comportent peu d'information exploitable par un utilisateur sur le sens des données ou sur les utilisations possibles des données. Elles sont plus dédiées à des contextes de transfert et de gestion de jeux de données qu'à l'accès d'utilisateurs. De plus, l'introduction dans de telles normes de champs décrivant la signification du contenu de jeux de données se heurte à un obstacle majeur : il n'existe pas d'ontologie de l'information géographique, au sens de modèle conceptuel unique sur lequel tous les acteurs manipulant de l'information géographique puissent s'accorder et qui pourrait servir à décrire le contenu en informations des BDG. Ces normes sont par ailleurs trop complexes pour être exploitées directement par des utilisateurs.

Concernant les métadonnées effectives, il existe en Europe peu de bases de métadonnées construites d'après les normes. Les métadonnées actuelles sont plutôt des fichiers de données produits dans des contextes de gestion de BDG et des informations non numériques.

Les infrastructures d'information géographique

Parallèlement aux travaux sur les métadonnées et dans une optique d'accès aux données –et non de transfert-, des catalogues de ressources en information géographique sont apparus : les *bibliothèques géographiques*, ou *geolibraries*. Elles indexent diverses bases de données par leurs composantes spatiales, appelées « empreintes spatiales » de ces ressources [Goodchild 99]. La bibliothèque géographique la plus connue est l'Alexandria Digital Library [Andersen et al. 95][Buttenfield and Kumler 96][Hill et al. 97].

Les requêtes mises en place dans un catalogue géographique doivent s'appuyer sur une information de contenu partageable sur les bases de données géographiques. Une telle information est essentiellement la notion de référence spatiale. En effet, nous avons vu précédemment qu'il est généralement possible de traduire un mode de référence spatial dans un autre. Des métadonnées typiques associées à la référence spatiale sont la couverture de la base, l'échelle ou la résolution, le système de coordonnées, les modes de positionnements possibles.

Un autre type d'information partageable entre bases de données est le nom des objets géographiques. Cette information est aussi une des plus compréhensibles pour les utilisateurs car elle ne fait pas l'objet d'une représentation différente de celle qu'en ont les utilisateurs. Les gazetiers sont des bases de données de toponymes géoréférencés. Ils peuvent servir de métadonnées d'accès à des positions géographiques : un utilisateur peut, dans une requête, utiliser un toponyme pour désigner la référence spatiale qui lui est associée dans le gazetier.

Des infrastructures sont organisées au sein de pays, les National Spatial Data Infrastructures (NSDI), pour diffuser des données à caractère localisé, généralement publiques, comme des données environnementales [Günther et Voisard 97], et également pour rendre accessibles différentes bases de données de référence comme s'il s'agissait d'une seule base [Abel et al. 99]. De fait, il y a une grande diversité, au sein même d'un pays, des fournisseurs de données géographiques et de données d'informations du secteur public, des politiques de diffusion de ces données, des disponibilités de ces données. Le besoin s'est fait sentir de donner une cohérence à l'ensemble des démarches de constitution et de distribution de telles bases dans un même pays. Une telle recherche de cohérence nécessite une politique de gestion de l'information géographique, appelée aussi NSDI. 28 pays disent posséder des NSDI et des travaux plus ambitieux visent maintenant à monter des GSDI (global au lieu de national) [Rhind 01]. Des projets travaillent ainsi à l'accès à un réseau de métadonnées portant sur l'information localisée, par exemple des BDG ou des données géologiques. C'est le cas du projet européen ESMI ou European Spatial Metadata Infrastructure prenant place dans le cadre INFO2000. Ce projet met en place un service de métadonnées spatiales pour accéder à plusieurs domaines de données localisées : les ressources géographiques et les ressources géologiques.

Les bibliothèques et infrastructures telles qu'elles existent actuellement ont montré des limites. Le retour d'expérience des utilisateurs a mis en évidence le besoin d'une interface avec des profils utilisateurs, et avec la possibilité de faire des « requêtes par l'exemple » [Goodchild 95]. Les bibliothèques géographiques deviennent actuellement des projets ambitieux [DG 99] :

“The contents of a distributed geolibrary are not limited to information normally associated with maps or images of the Earth's surface but include any information that can be associated with a geographic location. In this sense the vision thus extends far beyond the context of the National Spatial Data Infrastructure (NSDI).” *

* Le contenu d'une bibliothèque géographique distribuée ne se limite pas aux informations classiquement associées aux cartes ou aux images satellites. Il inclut aussi toute information qui peut être associée à une localisation géographique. Cette interprétation dépasse de beaucoup le contexte d'une infrastructure nationale de données spatiales. (tda)

Le GDDD a été amélioré par le projet LaClef. Ce projet permet des requêtes non seulement par mots-clés mais également par profils [LaClef 01].

Les infrastructures d'information supportent la recherche d'un lot de données que l'utilisateur sait décrire avec des caractéristiques élémentaires comme la zone géographique concernée. Par contre elles ne dispensent pas de fonctions de dérivation d'information, hormis, sur certains sites, des fonctionnalités de présentation cartographiques, de corrélations spatiales, ou de changement de systèmes de coordonnées. Ces infrastructures ne supportent pas l'expression d'un besoin d'information géographique autrement qu'en termes de contenu d'un lot de données.

1.1.3.2 L'accès à l'interprétation des données

La sémantique de données est le lien entre les données et ce qu'elles sont censées représenter. L'accès à l'interprétation de données repose sur la reconstruction de ce lien. Dans le cas des bases de données géographiques, ce lien n'est pas naturel pour plusieurs raisons exposées dans les sections 1.1.1.1 et 1.1.1.2 et que nous rappelons brièvement.

Comme il existe de nombreux contextes de formalisation de l'espace géographique, la compréhension d'une représentation de l'espace géographique nécessite la connaissance du contexte de cette représentation. Par exemple, certaines difficultés de formalisation de l'information géographique sont résolues grâce à des choix plus ou moins arbitraires, ou à des théories mathématiques sophistiquées, et il faut connaître ces choix et ces théories pour interpréter les données. Or, ce contexte est rarement explicité dans la base proposée à l'utilisateur. Les catégories utilisées pour décrire le monde réel sont toutefois spécifiées, dans des documents volumineux et trop techniques.

De plus, l'utilisateur accédant à des données a lui-même une représentation de l'espace géographique. Une difficulté majeure est l'écart entre le vocabulaire des utilisateurs et celui des bases de données géographiques.

Un même concept peut renvoyer à des représentations différentes. Par exemple un lieu-dit est représenté dans un modèle cartographique comme un objet ponctuel centré sur le groupe de bâtiments qu'il désigne, avec pour attribut l'intitulé du nom du lieu-dit. Un utilisateur s'attend souvent à ce qu'un lieu-dit soit modélisé comme un groupe de bâtiment ou une zone.

Un concept d'utilisateur peut ne pas être représenté dans les BDG. Par exemple un transporteur voulant effectuer un calcul d'itinéraire, un agriculteur voulant optimiser ses épandages, ne peuvent pas exprimer leur besoin de données géographiques à l'aide de concepts utilisés pour décrire des BDG.

Or, les contextes d'utilisation des données (ex : le vocabulaire d'un domaine d'utilisation, le modèle utilisé par des utilisateurs particuliers pour raisonner sur de l'information géographique) ne sont pas non plus explicités dans les informations rattachées aux BDG. Les bases de données géographiques n'ont pas été construites en vue d'une utilisation spécifique. Elles n'ont pu faire l'objet d'une enquête exhaustive des besoins, de sorte que leur schéma conceptuel de données n'a pu être élaboré en tenant compte des schémas propres à chaque utilisateur. Une telle enquête serait d'ailleurs impossible dans le domaine de l'information géographique. Une même base de données est pour autant exploitée dans une variété de domaines d'applications, et cela d'autant plus que cette base est chère à acquérir et complexe à maintenir. Ainsi, des données environnementales sont gérées par l'état, par exemple pour évaluer l'état des ressources naturelles, et ces mêmes données doivent aussi être accessibles au citoyen, par exemple pour connaître la qualité de l'air dans son voisinage [Günther and Voisard 97].

Une notion illustrant cette complexité est la résolution d'une base de données géographiques et son interprétation par un utilisateur.

La résolution d'une BDG dépend en partie de la résolution d'observation. Une donnée est acquise par observation et mesure d'un phénomène. La précision de cette observation renvoie à plusieurs grandeurs : un ordre de grandeur d'échelle attribué a priori par l'observateur à ce qu'il veut observer, la résolution de l'appareil d'observation.

La résolution des données correspond à la résolution d'observation, croisée éventuellement avec le pas de discrétisation de la géométrie et la précision de la numérisation.

La résolution d'une base de données est associée à une valeur d'échelle : l'échelle des cartes typiquement réalisables à partir de cette base. En effet, une base ne permet pas de produire des cartes à toute échelle.

L'interprétation d'une résolution est donc complexe pour un utilisateur qui ne sait pas discerner les notions de résolution, échelle et précision. La notion d'échelle de données, souvent utilisée, n'a pas de sens précis, c'est une notion agrégée à partir des résolutions de saisie et de stockage. Cette interprétation est particulièrement délicate quand l'utilisateur intègre ses propres données avec des données géographiques.

En conclusion, l'accès à l'interprétation des données nécessiterait l'explicitation du contexte de production des données et des choix faits lors de la représentation des données, ainsi que la prise en compte du contexte de l'utilisateur.

1.1.3.3 L'accès à l'exploitation

L'accès à l'exploitation de données géographiques renvoie au choix d'un SIG, à l'acquisition du SIG et à son utilisation sur les données.

Le choix du SIG dépend de plusieurs critères. Les principaux sont les données que l'on veut manipuler et les fonctions voulues.

D'autres critères sont la plate-forme associée au SIG, le prix, le volume de données supporté, le coût de maintenance et le coût de formation à la manipulation du système.

Certains critères sont décrits par le fournisseur de SIG. Par contre nous détaillons dans la suite en quoi la capacité d'intégration de données et de traitement de ces données proposée par un SIG n'est pas suffisamment décrite aux utilisateurs.

Evaluer la capacité d'intégration des données souhaitées dans le SIG

Il n'est pas toujours simple de savoir si des données seront correctement intégrées dans un SIG. L'utilisateur possède l'information sur les formats d'échange acceptés par le SIG, les modes de positionnement et systèmes de référence mais il ne connaît pas le modèle conceptuel de données (MCD) utilisé à l'intérieur du SIG. Les MCD des SIG ne sont en effet ni normalisés, ni explicites.

De plus, l'intégration de données dans un SIG dépend des données mêmes. Certains SIG ne gèrent par exemple pas les valeurs manquantes. C'est le cas d'AutoCad2000 qui ne sait pas importer des fichiers DXF 3D lorsque certains points de ces fichiers sont 2D mais cette information n'est pas spécifiée à l'utilisateur.

La capacité d'intégration dépend aussi de l'aptitude du SIG à effectuer des traitements d'intégration de données comme le changement de structure ou de mode de positionnement. Certaines valeurs de données peuvent correspondre à des valeurs limites du bon comportement du SIG pour certains traitements. Par exemple l'opération de géocodage est effectuée différemment selon le SIG mais ces différences ne sont pas explicites dans les modes d'emploi des SIG puisqu'ils ne décrivent pas tous les

cas d'utilisation du géocodage. C'est en testant au cas par cas que l'utilisateur constate que : tel SIG ne gère pas les abréviations dans les adresses, tel autre SIG ne gère pas les numéros de rue qui sont situés en dehors des limites des numéros connus sur la rue (ce que d'autres font en extrapolant),... Cela rend la comparaison des évaluations délicate puisqu'il faut savoir quel compromis faire entre les cas qui ne sont pas traités correctement par un SIG et traités par un autre.

De plus, une modélisation peut ne pas être disponible en standard sur un SIG mais être facilement applicable en s'appuyant sur les autres types proposés par le système. Il est significatif que lorsque nous avons fait une enquête auprès des personnes utilisant des SIG à l'IGN pour remplir le tableau Figure 8 nous avons pu avoir des réponses contradictoires sur un même SIG.

Dans de nombreux cas, des pré-traitements sont nécessaires pour adapter les données à l'application. Cela freine d'ailleurs les utilisations effectives de données géographiques en France : « le poids du coût de conversion des données dans le budget des projets et la critique sur l'adéquation des données à la demande font ressortir la nécessité d'une activité de services facilitant l'intégration et l'adaptation des données. » [AFIGEO 98]. Ces problèmes sont également importants en Europe où la constitution des bases de données géographiques est faite par diverses agences.

Lorsque des traitements sont disponibles comme les changements de structure ou de systèmes de coordonnées, il faut savoir qu'ils sont faits différemment selon les SIG et que les résultats peuvent varier grandement. De fait, une manipulation qui semble être un traitement élémentaire dans le discours de l'utilisateur se décompose souvent à l'intérieur du SIG en plusieurs opérations. Donc l'évaluation du comportement d'un SIG sur une manipulation demande de faire des tests sur un maximum de données particulières pour être sûr de tester tous les cas qui présentent de l'intérêt pour l'utilisateur. Cela nécessite du temps. Par exemple, l'IGN doit effectuer une étude sur les transformations de coordonnées de ou vers le Lambert 93 sur 2 SIG. Le temps prévu est de 2 semaines par SIG.

Evaluer les fonctions de dérivation souhaitées

Un autre critère de choix est de savoir si le SIG propose les fonctions voulues. La plupart des SIG ont les capacités de dérivation requises par les utilisateurs. Le problème est l'accès des utilisateurs à ces capacités. Même lorsque le traitement est accessible depuis l'interface standard du SIG, il ne suffit pas à l'utilisateur de connaître son nom, il lui faut aussi connaître son comportement sur ses données. Les modèles de traitement de ces systèmes ne sont jamais connus de l'utilisateur or ils ont des répercussions sur la capacité du SIG à effectuer correctement un traitement, en particulier lorsque le jeu de données présente des valeurs manquantes ou des valeurs correspondant à des limites du bon comportement du système. Cela rend l'évaluation complexe car il faut détecter ces cas limites.

De plus, certains traitements ne sont pas disponibles en standard alors qu'ils existent dans des modules additionnels, ou qu'ils peuvent être programmés à l'aide du langage de programmation offert par le SIG. En fait, plus un SIG propose de traitements, plus c'est un système complexe et moins son interface est simple.

Notons qu'il est également possible d'utiliser un site Web-SIG. Ces sites proposent des fonctionnalités SIG en ligne. Ils sont généralement développés à l'aide d'une sur-couche d'un SIG appelée extension Web-SIG. ESRI a par exemple conçu plusieurs extensions de ses logiciels commercialisés, dont la dernière est ArcIMS (Internet Map Server). Les fonctionnalités SIG ainsi proposées en ligne à des utilisateurs sont détaillées dans [Hubert 01a] [Hubert 01b]. Ceux sont essentiellement des fonctionnalités de visualisation (zoom, déplacement), ou de sélection de thèmes à cartographier. Des requêtes sont parfois possibles, soit pour sélectionner une zone d'intérêt comme dans les plans de ville,

soit pour comparer des thèmes. Certains sites proposent des fonctionnalités plus avancées comme le calcul d'itinéraire, le choix d'une projection ou d'une légende. L'auteur souligne que ces sites ne prennent pas en compte la variété des besoins et des types d'utilisateurs et sont peu flexibles. Ils proposent de plus peu de fonctionnalités d'analyse spatiale. Enfin, sur de tels sites, les fonctionnalités sont applicables aux données du site, l'accès en ligne à des services applicables sur n'importe quelles données géographiques n'est actuellement pas envisageable car les standards de métadonnées géographiques en place n'assurent pas encore l'homogénéité entre les services, comme le souligne [Abel et al. 99].

Un problème récurrent est finalement le manque de familiarité des utilisateurs avec les logiciels SIG et leur manipulation sur des BDG [Traynor and Williams 95]. La connaissance approfondie d'un SIG particulier, quel qu'il soit, demande plusieurs mois d'utilisation. A IGNConseil par exemple, pour ce qui est des 3 principaux SIG "simples" du marché, Geoconcept, MapInfo et ArcView, il existe un spécialiste de chaque. Le fait de s'être familiarisé avec un SIG donné ne signifie pas que l'accès aux autres SIG soit facilité, d'autant plus que les interfaces utilisateurs diffèrent de façon conséquente d'un SIG à l'autre. Ceci implique que peu de personnes "connaissent" des SIG comparables : une personne ayant appris par exemple à manipuler un SIG bureautique particulier préférera utiliser toujours ce même SIG pour les applications nécessitant un SIG de ce type.

Ce facteur freine les travaux de comparaison de SIG car pour les comparer il faudrait dans l'idéal en connaître plusieurs parfaitement pour effectuer correctement les tests, et pour trouver facilement les points de comparaison intéressants. Quand on possède mal un SIG (ce qui est pratiquement souvent le cas), un défaut décelé chez ce SIG peut être du en fait à une mauvaise manipulation de l'utilisateur, ou une lacune peut être due à une mauvaise connaissance des potentialités du système.

Dans le choix d'un SIG intervient donc souvent le fait que l'utilisateur se soit déjà familiarisé avec un tel outil, d'autant plus qu'il ne dispose pas dans la littérature de travaux comparant précisément les capacités de SIG.

En conclusion, le choix d'un SIG pour traiter des données est une activité complexe. Actuellement, dans les critères de choix, la facilité d'accès au SIG prévaut souvent sur l'optimisation de l'application même. De fait, il existe quelques grandes catégories de SIG (bureautique, professionnel, cartographie, traitement d'image...) au sein desquelles il est difficile de départager les SIG. A un besoin donné, correspond une catégorie dans laquelle chaque SIG apporte une réponse mais le détail de ces réponses ainsi que les différences entre ces réponses ne sont pas explicites pour l'utilisateur, ou sont difficiles à agglomérer pour faire un choix.

1.1.3.4 L'accès à plusieurs bases de données

Une situation courante est de devoir choisir une BDG parmi plusieurs alors qu'aucune de ces bases n'est dédiée à l'utilisation que l'on veut en faire. Cela implique de pouvoir évaluer le compromis à faire en choisissant certaines données plutôt que d'autres, compromis entre ce qu'apportent les ressources et ce qu'elles coûtent. En effet les ressources n'ayant pas été produites pour le besoin particulier de l'utilisateur, il se peut qu'aucune ne le satisfasse exactement. L'utilisateur doit peser le pour et le contre : tel jeu est moins précis mais plus récent, tel autre jeu nécessite une phase longue de pré-traitement mais il est riche en informations, etc. Là encore, la difficulté est de considérer ensemble des représentations différentes et de prendre en compte des critères de qualité variés.

Une autre situation courante est de devoir évaluer le contenu de plusieurs bases de données géographiques pour les exploiter de façon dépendante : utilisation multi-échelles, jointure de plusieurs

couvertures différentes. Or ces bases de données géographiques ont souvent été conçues indépendamment, par différents acteurs. Comme il existe beaucoup d'arbitraire dans la modélisation et représentation de BDG, ces bases présentent généralement beaucoup d'hétérogénéités. Il est délicat d'évaluer ensemble des BDG diverses, et éventuellement d'estimer la possibilité de les intégrer.

[Devoegele et al. 98] font ainsi une taxonomie détaillée des conflits rencontrés lors de l'intégration effective de divers jeux de données géographiques. Ils relèvent les grandes catégories de conflits suivantes : les conflits de sources de données, les conflits d'hétérogénéités (de modèle, de type de positionnement), les conflits de définition des classes, les conflits de structures, les conflits de description sémantiques et géométriques, et les conflits de données.

Ainsi, dans le cas de la construction de la BDPays décrit Figure 3 page 23, se posent les problèmes suivants :

- Problèmes de formats : il faut travailler sur une plate-forme avec des données provenant d'une autre plate-forme.
- Problèmes de référencement : il faut travailler avec des ressources utilisant des systèmes de référencement différents. L'intégration de ces ressources nécessite de les référencer dans un même système dans la base finale. Non seulement cela demande de développer des services de transformation de coordonnées mais aussi il faudrait pouvoir connaître précisément la précision des nouvelles coordonnées, qui n'est probablement plus la même que celle des coordonnées initiales.

1.1.4 Bilan de l'accès à l'information géographique

Nous venons de présenter notre vision de ce qu'est l'information géographique, la constitution de ressources en information géographique, les BDG, l'exploitation de ces ressources et les modalités de l'accès d'utilisateurs à ces ressources.

La problématique de l'accès d'utilisateurs à des données géographiques apparaît en définitive comme l'écart entre l'expression du besoin et la sélection de données utiles pour y répondre. Nous détaillons ci-dessous les éléments contribuant à cet écart.

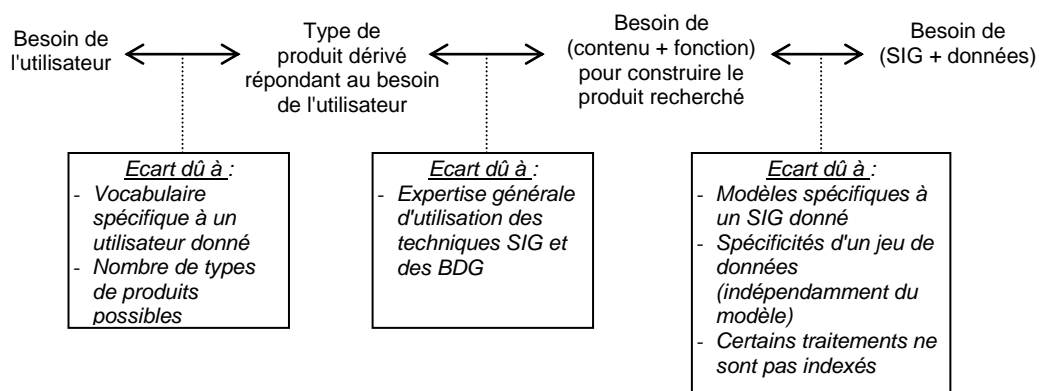


Figure 12 : Facteurs contribuant à l'écart entre l'expression d'un besoin d'information géographique et l'expression d'une sélection de données et traitements utiles pour y répondre.

L'information géographique dont a besoin un utilisateur est rarement explicitement représentée dans une BDG. C'est une information qu'il faut dériver à partir de données géographiques. C'est pourquoi notre démarche se situe en amont des SIG. Les SIG sont des systèmes d'accès à de l'information localisée dans la mesure où ils permettent de sélectionner des données à l'aide de requêtes spatiales. Ils

permettent de dériver de l'information mais ils ne développent pas de fonctionnalité d'aide à l'accès à cette dérivation dans la mesure où les produits potentiellement dérivés ne sont pas "indexés" au niveau du SIG.

L'expression du besoin renvoie souvent à des concepts utilisateurs, c'est-à-dire au contexte du besoin, et non aux ressources utiles pour y répondre. Or les utilisateurs ont leurs propres représentations de l'espace géographique qui est rarement celle utilisée par le producteur de données et qui intègre parfois des expertises spécifiques comme celle du transport ou de l'agriculture. Ils expriment leurs besoins avec des termes spécifiques.

Cette information recherchée correspond finalement à un produit dérivé. Il existe plusieurs produits dérivés des données géographiques, comme des bases utilisateurs géo-référencées, des cartes, des tableaux de mesures résultant d'analyses spatiales et thématiques. Certains utilisateurs non familiarisés avec l'exploitation de BDG peuvent ne pas savoir traduire leur besoin d'information géographique en un besoin de produit dérivé, c'est-à-dire qu'ils ne savent pas quel type d'application peut les satisfaire.

Un besoin de produit dérivé se traduit non seulement en une sélection de données mais aussi en une sélection d'outils qui fournissent des traitements sur les données. Ces deux types de sélection ne peuvent se faire de façon indépendante. Aussi les problèmes d'accès aux données sont multipliés par les problèmes d'accès aux capacités de dérivation des SIG. Les utilisateurs ne sont en effet pas toujours familiarisés avec ces capacités.

Lorsque le besoin est traduit en un besoin de contenu et fonctions, un problème extrêmement complexe demeure : choisir le format de données et le SIG les mieux adaptés pour répondre à ce besoin. Ce choix implique des compromis, des tests sur des jeux de données véritables. Le choix d'un SIG est un problème extrêmement complexe. Ce problème n'est actuellement pas encore optimisé au cas par cas et nous ne pouvons envisager de proposer un système générique d'aide au choix d'un SIG. Par contre il est possible de désigner à l'utilisateur les propriétés que le système doit présenter pour répondre à son besoin, et les conditions qu'il convient de vérifier.

Dans la section suivante nous abordons notre problématique sous un autre angle : la notion d'accès à de l'information stockée, non nécessairement géographique. L'analyse de cet angle permettra de décomposer notre problème et de structurer la diversité des obstacles à traiter.

1.2 Aider l'accès à des ressources en information

Dans cette section, nous faisons abstraction des spécificités de l'information géographique –comme par exemple la notion de situation dans l'espace- pour définir la notion d'accès à des informations stockées. Il est important de formuler cette problématique générale, pour décomposer un problème complexe en sous-problèmes.

Cela permet également de lever les ambiguïtés du langage courant et nous donner un vocabulaire de base pour notre discours, indépendant du domaine de l'information géographique.

De plus, élargir ainsi notre problématique nous amène à élargir le champ de notre état de l'art. Cet état de l'art, en ne se limitant pas au domaine de l'information géographique, nous assure de tenir compte des démarches les plus efficaces dans cette problématique d'accès et de proposer par la suite, partie 3, leur application au domaine de l'information géographique.

1.2.1 Accéder à des ressources en information

Accéder à des ressources de quelque nature qu'elles soient signifie « être en mesure de les utiliser ». Lorsqu'il s'agit de ressources en information, il est équivalent de dire qu'une personne peut utiliser une information stockée ou que cette information est devenue de la connaissance pour cette personne. Pour étudier la problématique générale de l'aide à l'accès, il donc est important de définir la notion de connaissance.

1.2.1.1 Les contextes

Nous détaillons ici les principaux contextes dans lesquels surgit la problématique de l'accès à des informations stockées. Nous considérons aussi le cas où l'information n'est pas stockée matériellement mais plutôt existe à travers le savoir d'un homme.

Accéder à de l'information stockée

Le premier contexte est l'exploitation de stocks d'informations. Nous détaillons ci-dessous les éléments importants de ce contexte.

La constitution de stocks d'information nécessite une phase de modélisation et de représentation de cette information pour pouvoir l'entreposer dans des données. L'information peut se prêter naturellement à cette phase de représentation quand elle appartient à un domaine déjà structuré. Mais ce n'est pas toujours le cas. Il faut alors construire un modèle de données décrivant l'information.

Dans ce cas, on est guidé par d'autres caractéristiques des informations à stocker qui sont :

- les modèles associés aux méthodes d'acquisition de données stockables décrivant cette information
- les représentations optimisées pour les manipulations prévues sur ces données.

Le premier cas est illustré par les exemples suivants : le stockage d'une information météorologique peut s'appuyer sur les résultats de mesures de barométrie, pluviométrie, anémométrie, et le stockage d'une information sur le comportement d'un fleuve peut s'appuyer sur les résultats de mesures effectuées par des capteurs à divers endroits du fleuve.

Le deuxième cas peut être illustré par l'organisation de tâches à effectuer dans un agenda, la représentation est ici {date(mois, jour, heure) à laquelle il faut effectuer la tâche, intitulé de la tâche}. En effet, on considère que la manipulation type de ces informations est de savoir ce qu'il faut faire à une date donnée. L'autre manipulation, consistant à savoir quand a lieu un événement, est secondaire.

L'utilisateur des informations stockées doit comprendre le modèle et la représentation utilisés pour construire les données. Il doit aussi adapter ces données au modèle qu'il utilise pour travailler sur les informations en question, modèle qui peut être différent du modèle de stockage.

Notons qu'un problème difficile à détecter survient quand l'utilisateur croit à tort avoir compris le modèle et la représentation, ou croit avoir traduit de façon correcte les données dans son modèle propre. Ceci peut entraîner des incohérences entre l'information qu'il récupère et l'information qui a été stockée.

Partager et réutiliser des connaissances

En IA, de nombreuses démarches en ingénierie des connaissances prennent place dans le contexte du partage et de la réutilisation de connaissances. Nous les détaillons dans la section 2.1.3.

La différence essentielle avec le contexte présenté avant est qu'ici les connaissances ne sont pas nécessairement stockées dans des données, elles peuvent aussi être présentes dans l'expertise d'un homme, dans le savoir global d'une communauté.

Construire des systèmes interopérables : communiquer des connaissances

Le dernier contexte est celui de la communication de connaissances entre des systèmes implémentés : l'interopérabilité.

[Vcovski 97] propose une définition précise de l'interopérabilité :

“Interoperability refers in general to the ability of various autonomous systems to bring together parts and to operate in collaboration. In most cases this means the exchange of meaningful information.

Within computing environments there are many different aspects of interoperability such as :

- Independent applications A and B running on the same machine and operating systems, i.e. interoperability through a common hardware interface.
- An application A reading data written by another application B, i.e. interoperability through a common data format.
- Application A communicating with application B by means of interprocess communication or network infrastructure, i.e. interoperability through a common communication protocol.

Besides technical issues, there are also interoperability topics at much higher levels of abstraction.[..]

The key issue in achieving interoperability [...] is a *contract* between interoperating parties defining a set of agreed-on common features.” *

* La notion d'interopérabilité renvoie de façon générale à l'aptitude de divers systèmes autonomes à rassembler des composants et à travailler en collaboration. Dans la plupart des cas cela signifie échanger des informations significatives. Dans le domaine informatique, il y a plusieurs aspects différents de l'interopérabilité comme :1) des applications indépendantes A et B travaillent sur une même machine et sous un même système d'exploitation (interopérabilité grâce à un même matériel). 2) Une application A

L'interopérabilité entre des systèmes implémentés A et B signifie donc que le système A peut travailler avec des « connaissances pour A » qui sont fournies par « B » et vice versa. La problématique de l'interopérabilité renvoie à ce qui tient lieu d'interface entre des systèmes pour leur permettre de se communiquer des connaissances. La différence avec le partage de connaissances ne provient pas uniquement du fait que l'on parle de systèmes implémentés mais aussi du fait qu'il s'agit de communiquer des connaissances : A ne connaît de B que ce que B veut bien lui communiquer. Il interprète les données communiquées en s'appuyant sur le « contrat ».

1.2.1.2 Le transfert de connaissances

Une notion clé : la connaissance

Les notions de partage, de réutilisation et de communication de connaissances, sont problématiques à cause de la nature un peu mystérieuse de ce qu'on appelle « connaissance ». Qu'est-ce que la connaissance? Comment stocker la connaissance, c'est-à-dire comment avoir de la mémoire « efficace », autrement que dans un cerveau humain? Comment communiquer des connaissances d'un agent à un autre, c'est-à-dire comment apprendre à un agent une partie de ce que sait un autre agent?

La connaissance est comme un signal : elle ne se consomme pas comme des ressources matérielles. Nous voyons dans la suite que, comme un signal, la connaissance ne prend pas sa valeur de façon intrinsèque mais dans un processus global de « transfert » de ressources vers un agent. Nous utilisons le mot agent sans faire référence à la technologie des Systèmes Multi-Agent mais seulement dans son acception courante.

D'après ce qui précède la notion de connaissances stockées renvoie à l'existence d'un agent qui peut se l'approprier et agir avec, c'est-à-dire donner le statut de connaissances à ce qui est stocké.

Des données aux connaissances

Il y a plusieurs statuts par lesquels passent des données stockées avant de devenir des connaissances pour un agent. Nous préférons dire statut plutôt qu'état parce que cela souligne le fait que les nuances ne sont pas intrinsèques mais contextuelles. Dans la littérature, cf. [Schreiber et al 99], trois statuts sont traditionnellement distingués : données, informations, connaissances. Ces termes renvoient à des notions proches, qui sont aussi désignées indistinctement en utilisant le terme information (au singulier) pour les auteurs de [Schreiber et al. 99] et dans la suite de ce mémoire, ou le terme connaissance (au singulier).

Afin d'obtenir une vision plus claire de ce que nous nommons connaissances nous passons en revue ces statuts.

Statut de ***ressources stockées*** : les données sont stockées quelque part.

Analogie du signal : un signal est émis par un émetteur.

Analogie de la ressource matérielle : des pâtes sont mises en paquet.

lit des données écrites par une application B (interopérabilité grâce à un même format). 3) Une application A communique avec une application B par leurs processus (interopérabilité grâce à un protocole de communication). A côté des problèmes techniques, il y a aussi des problèmes d'interopérabilité plus abstraits. Le principe essentiel en interopérabilité est la définition d'un contrat entre les parties communiquant, qui définit un ensemble de caractéristiques d'accord. (tda)

Statut de **données** : un agent, que nous appelons le récepteur, sait que ce stock existe.

Analogie du signal : un signal est perçu par un récepteur.

Analogie de la ressource matérielle : un acheteur passe devant le rayon du magasin où se trouve posé le paquet, et voit le paquet.

Statut d'**informations** : un agent, que nous appelons l'interpréteur, sait interpréter les données, expliciter le sens qui a été représenté dans les données.

Analogie du signal : un signal est décrypté par un traducteur et traduit en mots du langage naturel.

Analogie de la ressource matérielle : un acheteur comprend que c'est un paquet de pâtes en le lisant sur la boîte ou en voyant les pâtes à l'intérieur.

Statut de **connaissances** : un agent, que nous appelons le raisonneur situe l'information par rapport à ce qu'il sait d'autre, et est en passe d'utiliser cette information pour agir. Nous considérons qu'une action préliminaire à toute autre est le raisonnement de l'utilisateur : de l'information stockée devient des connaissances pour un utilisateur lorsqu'il peut mener un raisonnement dessus dans des termes proches de ceux de son travail.

Analogie du signal : un signal amène un agent à penser à agir. Par exemple, le feu est rouge je dois freiner et arrêter ma voiture.

Analogie de la ressource matérielle : un acheteur pense qu'il sait préparer le contenu du paquet, qu'il aime les pâtes et que s'il achète ce paquet il pourra se préparer un dîner rassasiant rapidement.

Pour notre travail, nous détaillons le processus global de transfert de connaissances de ressources vers un utilisateur en explicitant les processus intervenant à chaque changement de statut (données, informations et connaissances) Figure 13.

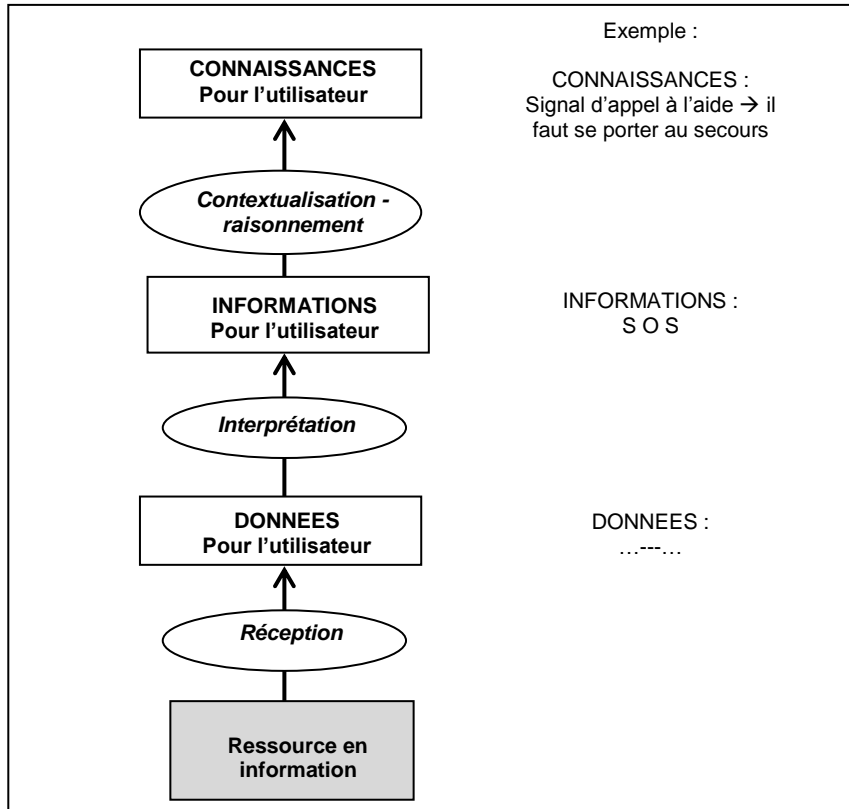


Figure 13 : Le transfert de connaissances depuis des données stockées vers un utilisateur. L'exemple du signal est repris et traduit de [Schreiber et al. 99]. [Figure 13 : Knowledge transfer from stored data to a user. The signal example is translated after [Schreiber et al. 99]]

Notons que les mêmes ressources peuvent devenir des connaissances différentes en fonction des façons dont sont tenus les rôles de récepteur, interpréteur, et raisonneur.

Nous traduisons les contextes évoqués en 1.2.1.1 dans les schémas suivants :

- la réutilisation de connaissances d'un acteur par un autre acteur, Figure 14,
- l'interopérabilité de composant, Figure 15.

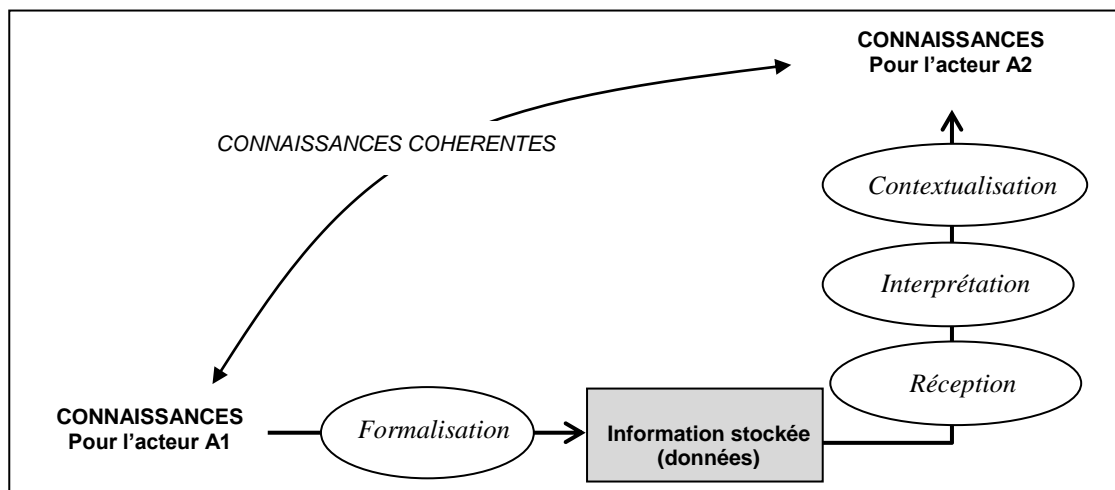


Figure 14 : La réutilisation de connaissances d'un acteur A1 par un acteur A2. [Figure 14 : Knowledge reuse between an actor A1 and an actor A2.]

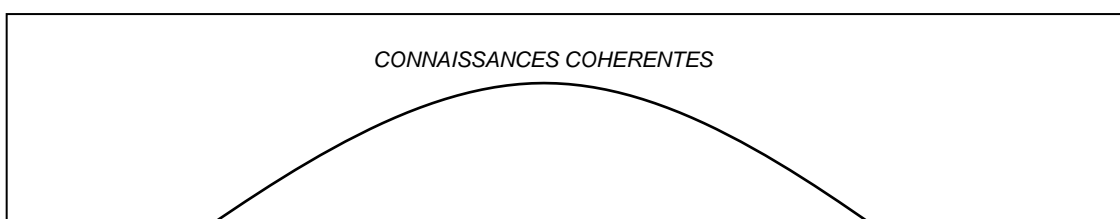


Figure 15 : L'interopérabilité entre des systèmes S1 et S2 : les systèmes peuvent échanger des connaissances via leurs interfaces. [Figure 15 : Interoperability between system S1 and system S2 : they can exchange knowledge through their interfaces.]

1.2.1.3 L'accès

Nous définissons l'accès d'un utilisateur à des ressources de la façon suivante : l'évaluation par l'utilisateur de l'information (données, informations ou connaissances) qu'il pourrait obtenir à partir des lots qui peuvent être extraits des ressources, en réponse à ses besoins. C'est un processus d'exploration des transferts d'information (données, informations ou connaissances) qui peuvent se produire depuis des lots pouvant être extraits des ressources vers l'utilisateur.

Ce processus peut se faire à l'aveuglette, c'est-à-dire en passant systématiquement en revue les données, les interprétations qui peuvent en être faites et les contextualisations possibles. Ceci est rarement possible. Un accès plus intelligent consiste en une exploration où ne retient, à chaque étape, que les possibilités intéressantes dans le contexte de l'utilisateur (cf. Figure 16). Par exemple cela peut se faire en se focalisant sur l'information que l'on s'attend à recevoir des ressources.

Nous nous intéressons à l'accès d'un utilisateur à une ressource défini comme : le processus d'exploration, focalisée par le contexte de l'utilisateur, des transferts de données, informations ou connaissances, qui pourraient se produire depuis des lots pouvant être extraits cette ressource vers l'utilisateur.

Analogie de la ressource matérielle : imaginer devant un article, de quel plat il peut être un ingrédient et dans quel menu.

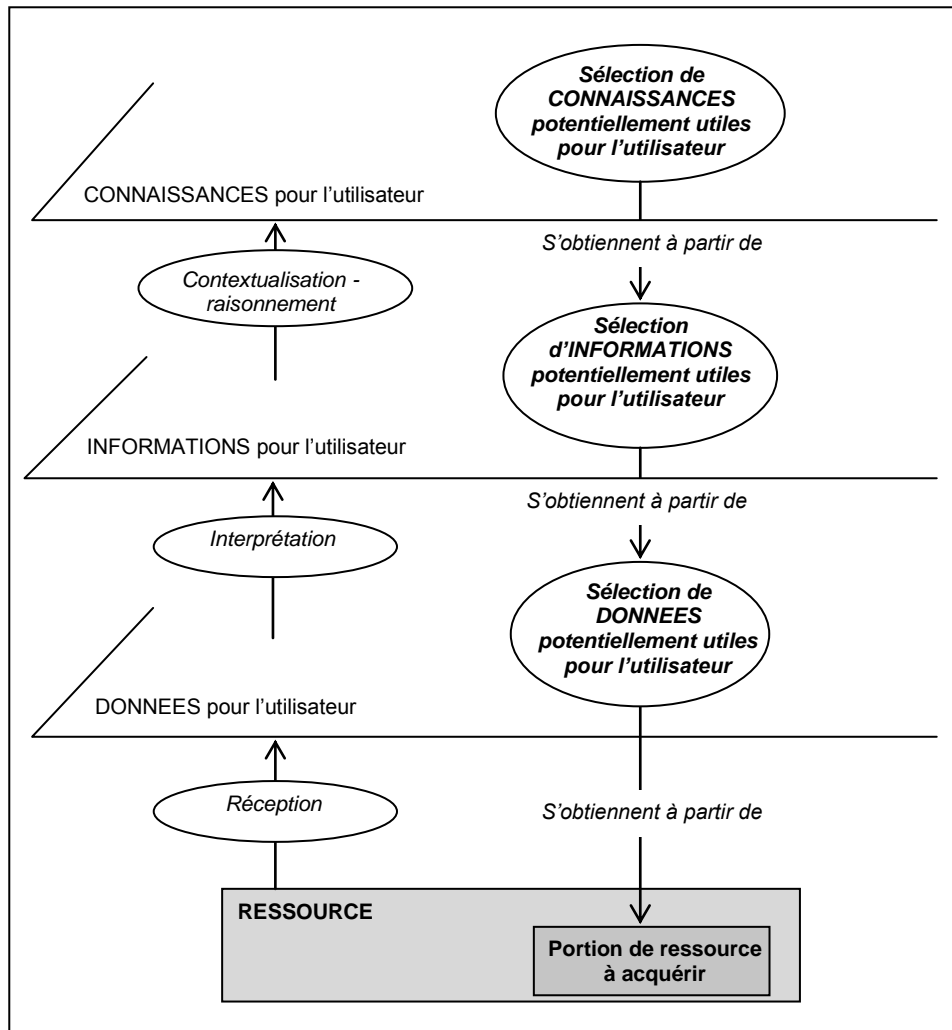


Figure 16 : Notre définition de l'accès à des données stockées : une exploration guidée des connaissances dérivables de ces données. [Figure 16 : Our definition of access to stored data : a guided exploration of knowledge that can be derived from the data.]

Un nouveau processus s'ajoute à ceux décrits Figure 13 pour former l'accès. Il s'agit de la sélection des ressources utiles c'est-à-dire nécessaires et suffisantes pour satisfaire le besoin de l'utilisateur, cf. Figure 16. Cette sélection s'appuie sur trois considérations, qui correspondent assez bien à la manipulation de connaissances, la manipulation d'informations, et la manipulation de données :

- Les ressources doivent apporter une connaissance.
- Les ressources peuvent être plus ou moins faciles à manipuler, cela peut coûter cher.
- La valeur intrinsèque des données (qualité, fraîcheur...) influe également sur le résultat.

Ce processus d'accès est conduit par l'utilisateur seul ou aidé d'un service. En anticipant sur la suite, on peut se poser les questions suivantes :

- Comment créer un stock tel que ce qu'il contient prenne rapidement le statut de connaissances pour l'agent auquel il est destiné ?
- Cet agent, le raisonneur, devra-t-il tenir lui-même le rôle du créateur de stock, du récepteur, du traducteur ou peut-il être aidé ?
- Si le stock est déjà créé, comment construire un service facilitant l'accès d'un utilisateur à ce stock ?

1.2.2 Aider l'accès

Dans cette section, nous nous appuyons sur la définition de l'accès établie précédemment, pour définir la notion d'aide à l'accès.

Cette notion d'aide à l'accès est très théorique car il s'agit d'une notion abstraite définie à partir de notions comme la connaissance. Mais il existe de nombreux cas particuliers de cet objectif générique qui renvoient à des démarches plus pratiques comme un mécanisme de parcours des ressources (comme l'ordre alphabétique), un wizzard automatisant des tâches répétitives dans l'utilisation d'un logiciel, un outil de mesure permettant de réduire la complexité des ressources pour en extraire une information, du conseil d'un expert dans le domaine de la manipulation des ressources.

Dans cette section, nous décomposons la notion générique d'aide à l'accès en sous-objectifs pour nous donner un point de vue simple sur ces diverses démarches pratiques. Pour effectuer cette décomposition, nous déclinons trois contextes d'aide à l'accès, du plus simple au plus général. Chaque contexte met l'accent sur des problèmes particuliers auxquels doit répondre une démarche d'aide à l'accès.

1.2.2.1 Faciliter l'accès d'un utilisateur à l'exploitation d'une ressource

Le premier contexte est celui où un utilisateur veut accéder à une ressource. Dans ce contexte, l'objectif générique d'aide à l'accès se décompose en trois sous-objectifs. Faciliter l'accès, dans ce contexte restreint, revient à contribuer à l'un ou l'autre de ces sous-objectifs. Notons que ces contributions sont rarement indépendantes.

Construire l'espace des transferts

L'utilisateur peut avoir besoin d'aide pour effectuer les processus du transfert décrits Figure 13, dans les cas suivants :

- l'utilisateur ne sait pas que les ressources existent,
- l'utilisateur ne sait pas où elles sont,
- l'utilisateur ne sait pas comment les manipuler, ou à quelles fins les utiliser.

Sélectionner l'information utile

Un service peut aider l'utilisateur à sélectionner l'information utile -c'est-à-dire nécessaire et suffisante pour satisfaire son besoin- parmi l'information dérivable des lots pouvant être extraits des ressources. Nous utilisons le terme information au singulier car il peut s'agir en pratique de données utiles, informations utiles ou connaissances utiles, selon le niveau auquel l'utilisateur est capable d'effectuer la sélection. L'aide à la sélection est cruciale dans les cas de saturation. La saturation est un phénomène inéluctable : lorsque l'information à traiter est en grande quantité, il y a un phénomène de saturation (cf. Figure 17).

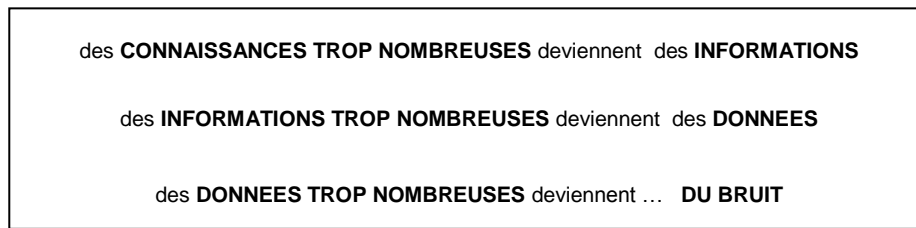


Figure 17 : La saturation des niveaux données/informations/connaissances. [Figure 17 : Saturation of the levels data/information/knowledge]

Ce sous-objectif est plus facilement atteint lorsque l'utilisateur a « une idée a priori » de ce qu'il veut faire avec les ressources.

Traduire la sélection d'information utile en une sélection de lot à acquérir

Un service peut aider l'utilisateur à distinguer, parmi la ressource initiale globale, le lot qu'il convient d'acquérir pour obtenir l'information utile pour lui. Cela revient à traduire la sélection d'information utile en une sélection de lot à acquérir duquel peut être dérivée cette information utile.

Ce sous-objectif est très important il évite à l'utilisateur de devoir considérer tour à tour chaque lot pouvant être extrait des ressources, pour chaque lot définir l'information qui peut en être dérivée, et retenir le lot permettant de dériver l'information utile. Ce dernier accès est l'accès "à l'aveuglette". Pour éviter "l'accès à l'aveuglette", il faut posséder une inversion de la relation entre les données stockées et les connaissances de l'utilisateur, cf. Figure 18.

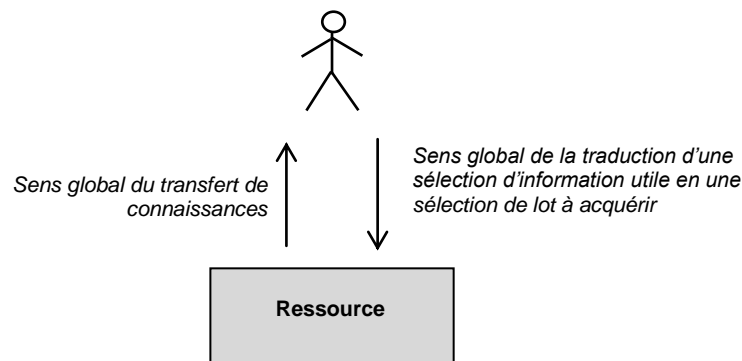


Figure 18 : L'accès correspond à une inversion de la relation ressources-utilisateur par rapport au transfert de connaissances des ressources vers l'utilisateur. [Figure 18 : Access corresponds to reversion of the relationship resources-user in knowledge transfer.]

Cet objectif est un peu réciproque de la construction de l'espace de transfert de connaissances puisqu'il revient à inverser la relation intervenant lors de ce transfert. Cette inversion permet que la sélection des connaissances ou informations, utiles pour l'utilisateur se propage en la sélection de données utiles à acquérir depuis la ressource.

La Figure 19 illustre la combinaison des 3 sous-objectifs.

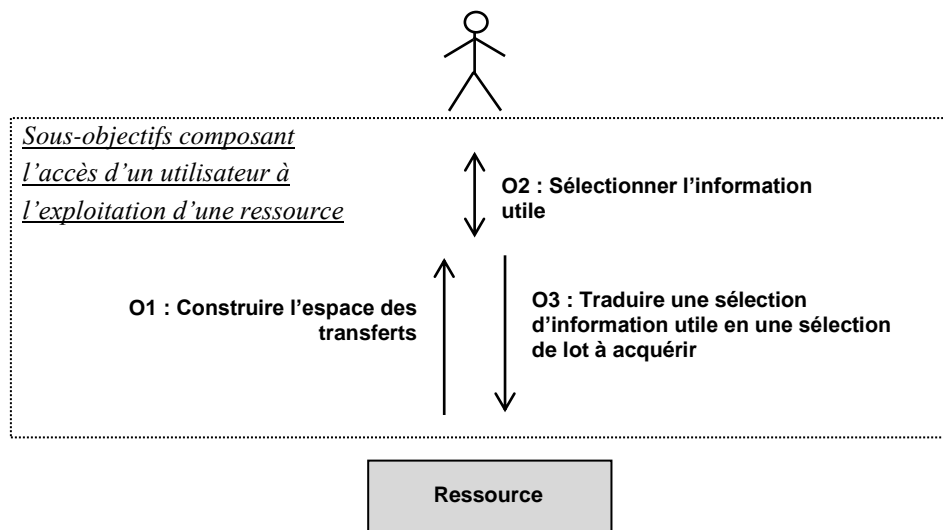


Figure 19 : Sous-objectifs composant l'accès d'un utilisateur à l'exploitation d'une ressource. Faciliter cet accès revient à contribuer à l'un ou l'autre de ces sous-objectifs. [Figure 19 : Objectives composing a user's access to the usage of a resource. To enhance this access means to contribute to one or another.]

1.2.2.2 Faciliter l'accès d'un utilisateur à l'exploitation de plusieurs ressources

Le deuxième contexte est celui où un utilisateur veut accéder à plusieurs ressources comme s'il s'agissait d'une seule ressource. Ceci veut dire d'une part que les processus de réception, interprétation, contextualisation doivent être le plus possible regroupés. D'autre part, cela signifie surtout que les sélections des connaissances potentiellement utiles chez l'une et l'autre ressources, cf. Figure 16, doivent être confondues. Notons que les connaissances extractibles de plusieurs ressources peuvent être plus élaborées que la somme des connaissances extractibles de chaque ressource.

Ceci est particulièrement complexe quand les sources sont hétérogènes, ou physiquement distribuées. Les hétérogénéités possibles de sources d'information sont nombreuses, [Siméon 99] :

- La forme de représentation, qui diffère en fonction de l'application initiale ou de l'outil qui a servi à créer la source : documents, tables relationnelles, etc.
- Le format de stockage, soit encore la syntaxe de la source.
- La sémantique des données : les métadonnées, la structure.
- Les opérations disponibles sur les données : création, mise à jour, interrogation.
- Le protocole ou l'interface qui permet de manipuler la source : communiquer et accéder.

La Figure 20 illustre la décomposition de l'objectif d'aide à l'accès dans le contexte où il y a plusieurs ressources.

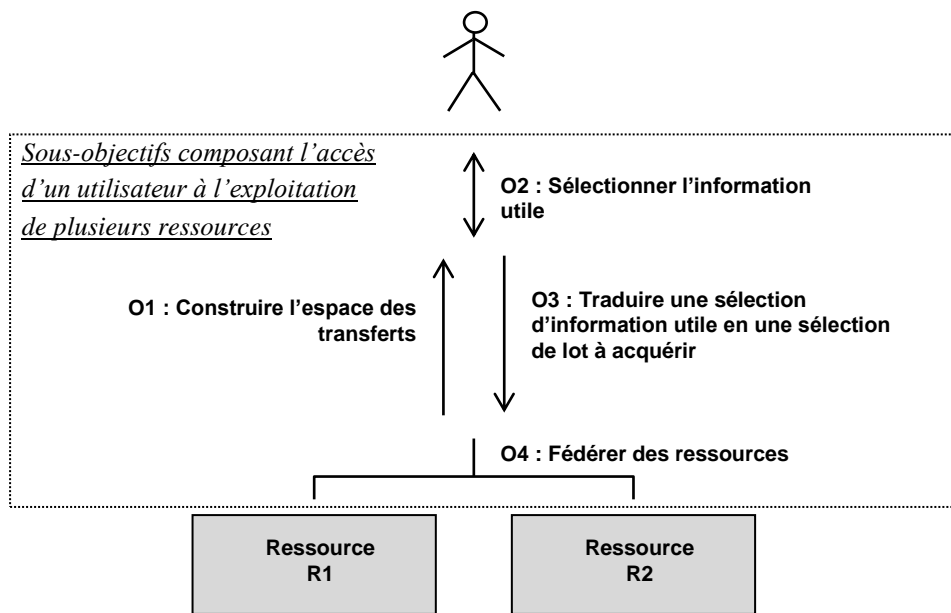


Figure 20 : Sous-objectifs composant l'accès d'un utilisateur à l'exploitation de plusieurs ressources. Faciliter cet accès revient à contribuer à l'un ou l'autre de ces sous-objectifs. [Figure 20 : Objectives composing a user's access to the usage of several resources. To enhance this access means to contribute to one or another].

1.2.2.3 Faciliter l'accès de plusieurs utilisateurs à l'exploitation de ressources

Des utilisateurs divers peuvent vouloir accéder à une même ressource sans devoir être aidés de la même façon. Ceci est de plus en plus le cas comme le note [Sheth 99] : le contexte de la spécialisation croissante du travail accompagné du besoin croissant de réutiliser et d'analyser les données conduit à la création d'informations et de connaissances qui peuvent être réutilisées et partagées. Ce contexte, à la différence du précédent, ne rend pas le problème d'accès plus complexe mais rend le problème d'aider l'accès plus complexe. En effet il faut adapter l'aide à l'utilisateur.

La Figure 20 illustre la décomposition de l'objectif d'aide à l'accès dans le contexte le plus général où plusieurs utilisateurs doivent accéder à des ressources.

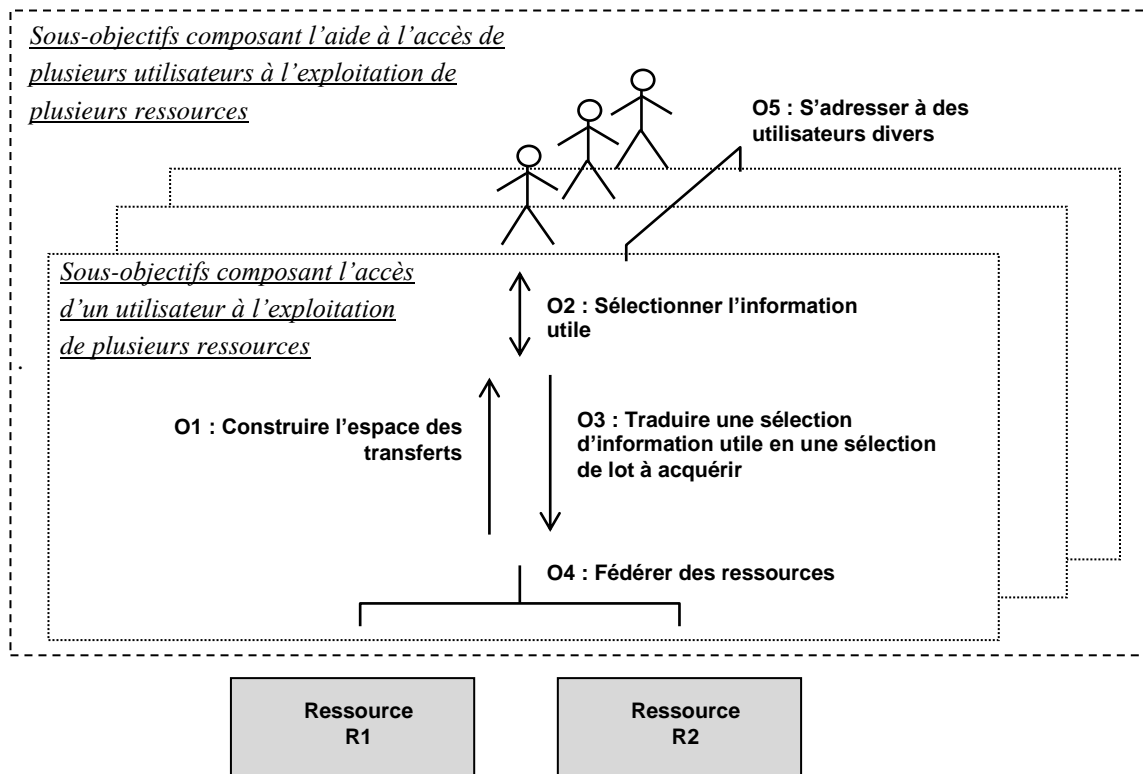


Figure 21 : Sous-objectifs composant l'aide à l'accès de plusieurs utilisateur à l'exploitation de plusieurs ressources. Mettre en place cette aide revient à contribuer à l'un ou l'autre de ces sous-objectifs. [Figure 21 : Objectives composing users' access to the usage of several resources. To enhance this access means to contribute to one or another]

1.3 Bilan

La présentation du contexte de l'accès à l'information géographique a montré que l'accès à l'exploitation de ressources en information géographique soulève des problèmes variés. De fait, cette problématique est trop vague pour pouvoir l'analyser et envisager une contribution. Nous avons formalisé la notion d'accès dans le cas général de ressources en information non nécessairement géographiques et nous avons décomposé cet objectif générique en sous-objectifs.

Nous pouvons constater alors que le cas de l'aide à l'accès aux données géographiques est un cas particulièrement épineux du cas général.

Une difficulté majeure est la construction de l'espace des connaissances qu'un utilisateur peut obtenir depuis des données géographiques. Il faut entre autres savoir quel est le contenu des données, quels traitements appliquer aux données brutes, et savoir interpréter le résultat.

De plus, l'accès à des ressources en information géographique implique souvent l'accès à plusieurs ressources hétérogènes. Si nous reprenons les hétérogénéités énumérées par [Siméon 99] (cf. en 1.2.2.2 : la forme, le format de stockage, la sémantique des données, les opérations disponibles sur les données, le protocole), nous constatons que les données géographiques cumulent toutes les hétérogénéités possibles.

L'hétérogénéité des ressources dans toutes ces dimensions rend difficile l'évaluation des critères de sélection énumérés en section 1.2.1.3.

Enfin, de même que dans le cas général, les démarches visant à faciliter l'accès aux données géographiques doivent traiter, outre les difficultés qui viennent d'être énumérées, le problème de l'hétérogénéité des utilisateurs de données géographiques.

Les structures en place pour faciliter l'accès aux données géographiques ne peuvent donner accès à ces données en tant que ressources en information puisqu'elles ne lèvent pas tous ces obstacles.

Le problème essentiel à nos yeux est la communication entre les modèles et représentations de l'information géographique utilisés par les producteurs de données et les autres acteurs de l'information géographique. Cette notion, spécifiquement difficile dans le domaine de l'information géographique, est liée à la problématique générale de l'accès à des informations stockées.

Le domaine scientifique abordant cette problématique est l'Intelligence Artificielle. Aussi nous faisons d'abord un état de l'art du cas général de l'aide à l'accès à des ressources en information, dont le point culminant sera les derniers résultats d'IA. Nous reviendrons ensuite, riches des enseignements de ce premier état de l'art, au cas géographique.

2 ANALYSE DE L'ÉTAT DE L'ART

La partie précédente expose la problématique abordée par ce travail : aider des utilisateurs à accéder à des ressources en information géographique. Nous avons d'abord présenté des éléments importants touchant à l'accès aux ressources en information géographique. Puis nous avons présenté la notion d'aide à l'accès en mettant en valeur l'axe *manipulation de connaissances* et non l'axe *information géographique* de notre problème initial.

Cette partie analyse les démarches abordant cette problématique, qu'elle soit générale ou limitée au cas de l'information géographique.

Nous faisons dans un premier temps abstraction de notre domaine d'information pour nous focaliser sur la notion d'aide à l'accès. Cet objectif générique d'aide à l'accès a été décomposé dans la partie précédente, section 1.2.2, en 5 sous-objectifs. Toute démarche contribuant à l'un de ces sous-objectifs entre dans notre état de l'art du cas général ; ces sous-objectifs forment en quelques sortes les dimensions possibles de ce premier volet. D'après la remarque faite dans le dernier bilan, l'obstacle principal que nous souhaitons aborder est le fossé entre les concepts des utilisateurs et ceux des BDG. Cet obstacle est traité par l'aide à la construction de l'espace des transferts. Nous nous intéresserons donc surtout aux démarches qui développent de façon conséquente cette fonctionnalité et nous structurons l'état de l'art dans le cas général selon cette dimension, en distinguant trois classes (cf. 2.1.2) : les démarches facilitant l'accès aux données, aux informations ou aux connaissances. Il s'agit des démarches qui facilitent l'exploration, focalisée par un contexte d'utilisateur, de l'ensemble des données, informations ou connaissances, que l'on pourrait acquérir depuis une ressource. La partie 2.1.3 détaille les travaux d'aide au raisonnement formel qui entrent dans la catégorie d'aide à l'accès aux connaissances.

A l'issue de cet état de l'art sur les techniques d'aide à l'accès, nous faisons un bilan de leurs intérêts dans notre contexte. Cela nous permet de restreindre le cadre technique de notre travail à l'aide au raisonnement formel.

Nous tenons compte des conclusions de ce premier volet de l'état de l'art et de cette délimitation du cadre technique de notre travail pour élaborer un deuxième volet portant sur le cas de l'information géographique. Ce deuxième volet étudiera donc les formalisations existantes de l'information géographique.

2.1 Etat de l'art des techniques d'aide à l'accès

2.1.1 Généralités

2.1.1.1 Solution générique

L'objectif correspondant au cas général, non spécifiquement géographique, de notre problème initial est de faciliter l'exploration, focalisée par le contexte de l'utilisateur, des données, informations ou connaissances qui pourraient être transférées de lots de ces ressources vers l'utilisateur. Cet objectif a été décomposé en sous-objectifs :

O1 : Construire l'espace des transferts

O2 : Sélectionner l'information utile

O3 : Traduire une sélection d'information utile en une sélection de lot à acquérir

O4 : Fédérer des ressources

O5 : S'adresser à divers utilisateurs

Toute démarche contribuant à atteindre un de ces objectifs entre dans notre état de l'art. Ces objectifs forment la caractérisation fonctionnelle d'une solution générique. Nous traduisons cette caractérisation en composants fonctionnels d'une architecture générique décrivant toute solution entrant dans notre état de l'art. Cette traduction est synthétisée sur la Figure 22.

La description des ressources

Un composant essentiel est la description des ressources.

Il contribue de fait à O1 car faciliter les processus de réception, interprétation, contextualisation et raisonnement sur les ressources, revient à décrire de façon plus ou moins riche les ressources :

- Désigner l'existence des ressources et des lots qui peuvent en être émis (les ressources deviennent des *données*),
- Intégrer des connaissances d'interprétation (les ressources deviennent des *informations*),
- Intégrer des connaissances de mode d'emploi (les ressources deviennent des *connaissances*).

Ce composant contribue également à O4 (la fédération de ressources hétérogènes) quand un même langage est utilisé pour décrire plusieurs ressources.

La description des besoins d'utilisateurs

De façon symétrique, un service d'aide possède une description des besoins d'utilisateurs, ou plus exactement un langage dans lequel est décrit le besoin de l'utilisateur. Cela peut être le même langage que celui servant à décrire les ressources ou un autre langage, dédié à la description de besoins, indépendamment des ressources. Ce composant contribue à O1 et à O5.

Le mode d'interaction avec l'utilisateur

Un autre élément essentiel d'une solution est le mode d'interaction avec l'utilisateur permettant de construire la description de l'utilité attendue, c'est-à-dire du besoin de l'utilisateur. Cette interaction peut être une navigation de l'utilisateur dans une description des besoins ou dans une description des ressources. Cela peut aussi être l'expression d'une requête de l'utilisateur.

Ce composant contribue de fait à O2. Il contribue aussi à O5, la prise en compte d'une variété de contextes et comportements d'utilisateurs.

Le mécanisme de mise en adéquation d'une expression de besoin en une requête de ressources

Un composant nécessaire pour éviter "l'accès à l'aveuglette" (cf. 1.2.2.1) est un mécanisme de mise en adéquation d'une description de besoin et d'une description de ressources. Ce composant correspond à O3.

L'indexation des ressources

Enfin, de même que le système possède une interface avec les utilisateurs, il possède une interface avec les ressources. L'indexation contribue à O1 de façon minimale, c'est-à-dire qu'elle désigne l'existence des ressources. Par ailleurs, O2 (l'aide à la sélection d'information utile) n'est possible que si la description des ressources est associée à des index qui inversent la relation (ressources, contenu). Enfin cette indexation contribue aussi à O4 puisqu'elle permet de regrouper des pointeurs sur des ressources physiquement distinctes.

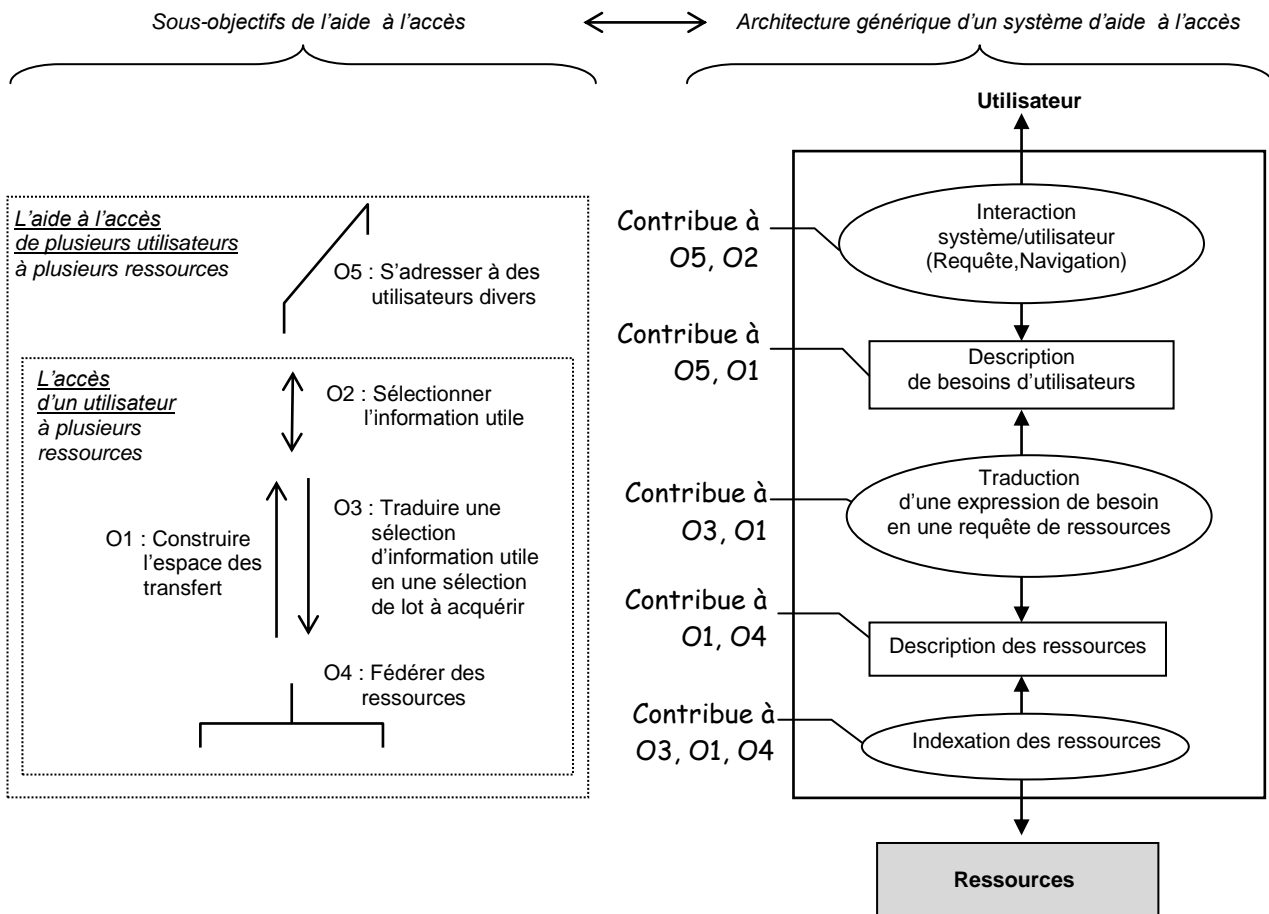


Figure 22 : Architecture générique d'un système permettant à divers utilisateurs d'accéder à diverses ressources : les éléments techniques et leur contribution aux diverses fonctionnalités caractérisant une solution. [Figure 22 : Generic architecture of a system allowing different users to access different resources : technical components and their contribution to the several functionalities characterizing a solution].

2.1.1.2 Un dilemme classique : expressivité et maniabilité

Dès qu'on ne se situe plus dans le langage naturel et que l'on cherche un mode d'expression plus formel, en vue d'implémenter un système, il y a un compromis à trouver entre : la richesse de description possible et l'efficacité du raisonnement que l'on peut mener sur ces descriptions. C'est ce que nous verrons plus tard à propos des langages de description d'information.

Dans le cas d'un système facilitant l'accès à des ressources, la richesse du langage correspond à :

- la possibilité d'avoir une bonne interface avec les ressources -décrire de nombreuses ressources différentes en les distinguant les unes des autres-
- la possibilité d'avoir une bonne interface avec les utilisateurs -décrire de nombreux besoins différents en les distinguant les uns des autres-.
- la possibilité d'intégrer des connaissances d'interprétation et de contextualisation.

Les raisonnements que l'on peut mener sur les descriptions faites dans le langage correspondent à :

- la possibilité d'organiser un mode de parcours de la description (méta-niveaux d'index etc.),
- la possibilité d'automatiser la traduction de l'expression d'un besoin en une sélection de ressources.

Le dilemme se traduit alors comme suit. Lorsqu'un langage permet de désigner, en les distinguant les unes des autres, de nombreuses ressources différentes, alors il est difficile de construire dans ce langage des raisonnements permettant de déterminer l'utilité des ressources et de les indexer.

Lorsqu'un langage permet de décrire, en les distinguant les uns des autres, de nombreux besoins différents alors il est difficile de mener dans ce langage des raisonnements permettant de caractériser la ressource la plus utile pour répondre à un besoin donné, et pour construire une indexation au niveau de l'expression des utilités.

Lorsqu'on construit une indexation efficace, c'est-à-dire qui se parcourt rapidement, cela se fait au détriment de la richesse d'expression des besoins ou de description des ressources –il y a eu introduction d'arbitraire au moment du choix des critères d'indexation et donc perte d'information-.

Tout ce qui nécessite un langage riche pose dès lors problème. C'est le cas lorsque les éléments décrits, ressources ou besoins, sont hétérogènes. C'est également le cas lorsque les ressources décrites sont nombreuses : il faut alors un niveau de détail fin pour distinguer des ressources les unes des autres.

De même, tout ce qui nécessite beaucoup de raisonnement pose problème. C'est le cas lorsque les ressources sont « loin » de leur utilisation : les processus du transfert de connaissances sont complexes.

En réponse à ce dilemme, plusieurs langages sont parfois combinés : un langage de description des ressources proche des ressources et un autre permettant de décrire les utilités attendues. Le premier langage fournit des « poignées » permettant d'attraper les ressources. Le deuxième langage organise ces « poignées » pour que l'utilisateur fasse rapidement son choix.

2.1.1.3 Structure de l'état de l'art des techniques

L'obstacle principal dans le cas géographique, c'est-à-dire l'aide à l'accès d'utilisateurs à l'exploitation de BDG, est le fossé entre les diverses représentations de l'espace géographique, construites par les utilisateurs et les producteurs de BDG. Nous nous concentrons tout particulièrement sur le sous-objectif O1. Nous classons donc les services d'aide suivant l'axe du transfert de connaissances. C'est ainsi que nous distinguons :

- les services d'aide à l'accès aux données,
- ceux d'aide à l'accès aux informations,
- et ceux d'aide à l'accès aux connaissances.

Les services d'aide à l'accès aux données (resp. aux informations, aux connaissances) sont en fait des démarches qui facilitent l'exploration de l'espace des données (resp. des informations, des connaissances) qui pourraient être transférées depuis des lots extraits des ressources.

Nous détaillons ci-dessous chacune de ces catégories de services.

Un service d'aide à l'accès aux données permet en général à l'utilisateur de sélectionner une donnée qu'il recherche, comme les coordonnées d'une personne dans un annuaire organisé par ordre alphabétique, les « pages blanches ».

Un service d'aide à l'accès aux informations permet à l'utilisateur de sélectionner une ressource sur son contenu en informations : par exemple « les pages jaunes » organisent une exploration de coordonnées de professionnels aux niveaux des informations de type « profession exercée ». Un autre exemple est celui de la médiathèque où l'on peut rechercher un document, indépendamment de sa forme (image, son, texte) en fonction d'un thème.

Un service d'aide à l'accès aux connaissances permet à l'utilisateur de sélectionner une ressource en fonction de ce qu'il veut faire avec. Un exemple serait des pages jaunes dans lesquelles l'index serait « réparer des

chaussures », « construire une maison », etc. Un exemple effectif est le guide touristique dans lequel les données sur une ville sont organisées de la façon suivante : « où dormir ? », « où manger ? », « où sortir ? ».

La construction d'un service d'aide dépend de plusieurs facteurs (aidant ou limitant).

Un premier facteur est l'existence d'un fossé important entre les connaissances des utilisateurs et l'expertise de manipulation des ressources. Dans ce cas il est nécessaire de centrer l'aide sur la fonctionnalité O1, c'est-à-dire d'associer aux ressources des connaissances sur leur contenu et leurs modes d'emploi.

Un autre facteur est l'existence d'un langage de désignation des ressources tel que les descriptions construites dans ce langage sont facilement manipulables pour traduire un besoin d'utilisateur dans une telle description. On choisit alors de construire une aide centrée sur les descriptions dans ce langage. Par exemple, pour ranger des articles dans un magasin, l'ordre des prix des articles est un ordre simple mais qui n'est pas très exploitable pour l'accès du client aux articles utiles.

Un dernier facteur, symétrique du précédent, est l'existence d'un langage de description des besoins tel que les descriptions de besoins dans ce langage s'organisent facilement et permettent une interaction simple avec un utilisateur. Dans l'exemple du rangement d'articles dans un magasin, une indexation possible est par exemple par type d'utilité car ces types s'organisent facilement –ce qui sert à se nourrir, ce qui sert à se distraire...-.

2.1.2 L'aide à l'accès à l'information (données, informations, connaissances)

2.1.2.1 L'aide à l'accès aux données

Sont catégorisés dans notre état de l'art comme des services d'aide à l'accès aux données, les services (d'aide à l'accès) tels que leur contribution à O1 se limitent à signifier l'existence de la ressource et à désigner des lots qui peuvent en être émis. Ces services sont utiles pour signifier à un utilisateur que les ressources existent et où elles sont, et également pour automatiser le parcours des lots qui peuvent être émis lorsqu'ils sont trop nombreux. Le fonctionnement de tels services est possible lorsque l'utilisateur connaît la description, dans le langage utilisé par le service pour décrire les ressources, de la ressource qu'il recherche.

L'aide à l'accès à des données la plus élémentaire consiste à regrouper physiquement des pointeurs sur les données. Une étape ultérieure est de combiner ces listes de pointeurs avec des techniques de filtrage pour gérer la sélection de la ressource utile : méthodes de gestion de requête, fichiers inversés, signatures, quadrees [Siméon 95]. Sur Internet, les moteurs de recherche les plus élémentaires sont une aide à l'accès aux données. Ils collectent automatiquement les ressources, grâce à un module qui tourne en permanence, les indexent par les mots lus dans leur fichier et proposent à l'utilisateur de formuler une requête du type « groupe de mots-clés ».

Une aide à l'accès aux données efficace consiste en un principe a priori de construction des ressources : l'utilisation d'un langage structuré pour organiser les données, associé à des opérations de sélection qui peuvent être menées sur les descriptions faites dans ce langage. Par exemple, l'algèbre relationnelle sert de base à de nombreuses techniques d'optimisation de l'accès aux données dans des bases de données structurées.

Une base de données est construite pour stocker et organiser un nombre important de données en vue d'automatiser des traitements dessus. Il existe un schéma conceptuel de données qui permet d'interpréter les

données mais les langages de requête de données sont plus proches du schéma logique et donc nous considérons que les SGBD sont une aide à l'accès aux données plus qu'à l'accès à l'information.

Les SGBD sont une aide à l'accès aux données et non aux informations car l'utilisateur doit connaître le schéma conceptuel de données pour comprendre ce que les données peuvent lui apporter et exprimer son besoin d'information utile.

Les techniques de BD et SGBD répondent toutefois à divers contextes de l'accès à des données. Nous citons ci-dessous les définitions correspondantes, issues de [Gardarin et Valduriez 90].

« *Base de données distante* (remote database) : base de données située sur un ordinateur autre que le ordinateur de l'utilisateur et accédée grâce à des commandes de communication spécifiées par l'utilisateur. »
Les exploitations rendues possibles dans cet accès sont généralement la consultation des données, les opérations de modification étant souvent centralisées.

« *Base de données répartie* (distributed database) : ensemble de bases de données coopérantes, chacune résidant sur un site différent, vu et manipulé par l'utilisateur comme une seule base de données centralisée.

[..] *Une base de données répartie homogène* est une base de données répartie obtenue en divisant une base de données en un ensemble de bases de données locales, chacune étant gérée par le même SGBD. »

[..] *Une base de données répartie hétérogène* est une base de données répartie obtenue en intégrant dans une base de données unique un ensemble de bases de données locales gérées par des SGBD différents. »

« *Base de données fédérée* (federated database) : ensemble de bases de données faiblement couplées et autonomes qu'un utilisateur peut manipuler à l'aide d'un langage spécifique.

[..] Les bases de données locales ne sont pas intégrées en une seule base mais plutôt interopérables, dans la limite où elles peuvent échanger des données et les manipuler ensemble au moyen d'un langage multibases.

[..] *Langage multibase* (multidatabase language) : Langage qui permet la définition de données multibases, la définition des dépendances inter-bases, et la manipulation de données multibases. »

[Gardarin et Valduriez 90] soulignent que l'intérêt des BD réparties hétérogènes et des BD fédérées est de fournir une indépendance au SGBD. A ce niveau, le système fournit une aide proche de l'accès aux informations puisque l'utilisateur n'accède pas aux données dans leur formalisme d'origine mais dans celui qui est utile pour son application. La plupart des BD ne sont actuellement pas organisées en BD réparties ou fédérées, du moins dans le domaine de l'information géographique. Si de tels systèmes sont attractifs, leur réalisation est très complexe.

Les réseaux, tels que Internet sans tenir compte du Web, peuvent être considérés comme des systèmes d'aide à l'accès aux données puisqu'ils permettent l'accès physique à diverses données. Mais dans notre contexte, l'intérêt d'Internet réside dans l'apport du Web et nous en parlons dans la section suivante car il s'agit d'aide à l'accès à l'information.

2.1.2.2 L'aide à l'accès aux informations

Les entrepôts de données et la fouille de données

Des démarches visent à mieux organiser les bases de données pour en dériver de l'information. Il s'agit du processus global *Knowledge Discovery In Databases* (KDD). Notons que le terme *knowledge* renvoie à une acception plus générale du terme connaissance que celle que nous avons introduite en 1.2.1.2. Cette acception est en fait celle du terme *information* (utilisé au singulier).

Un processus élémentaire du KDD consiste à construire un entrepôt de données (en anglais, *datawarehouse*) c'est à dire : « une collection de données orientées sujet, intégrées, non volatiles et historiées, organisées pour le support du processus d'aide à la décision », selon la définition utilisée dans [Lefébure et Venturi 98].

Un autre processus du KDD est la phase d'extraction des connaissances à partir de ces entrepôts. Cette phase s'appelle fouille de données, plus connue sous le terme anglais *data mining* (DM). Elle est définie comme : « la découverte de nouvelles corrélations, tendances et modèles par le tamisage d'un large volume de données » [Lefébure et Venturi 98]. [Zeitouni et Yeh 00] classent les méthodes du DM en deux catégories : l'analyse exploratoire et la prédiction. Ces méthodes exploitent les acquis des statistiques lorsque les données sont des valeurs numériques et ceux de l'intelligence artificielle lorsque les données sont des valeurs symboliques.

Dans le contexte d'aide à l'accès, les techniques de KDD participent au transfert de connaissances des données vers l'utilisateur, par contre elles n'offrent pas de mécanismes d'aide à la sélection d'information utile.

Les infrastructures d'informations

Les infrastructures d'informations sont l'équivalent, à un niveau de description d'informations, du regroupement de pointeurs sur des données : il s'agit de regroupement de descriptions de ressources. Elles sont construites à des niveaux nationaux ou international. L'infrastructure la plus célèbre est le Web. La construction d'infrastructures d'information est liée à deux techniques :

- La première consiste en un système de pointeurs sur les ressources, sur le Web ce sont les URL (Uniform Resource Locator, adresse et protocole d'accès d'une ressource sur le Web).
- La suivante consiste en un mode plus ou moins homogène de description de différentes ressources en terme de contenu. Ces descriptions permettent de situer l'accès d'un utilisateur au niveau des informations contenues dans la ressource.

Plusieurs services d'aide à l'accès sont associés aux infrastructures d'information. La plus couramment utilisée sur le Web est le moteur de recherche. Ceux-ci se divisent en deux grandes catégories : les moteurs de recherche élémentaires –les robots- et les répertoires validés par l'homme. Nous considérons que les robots appartiennent à la catégorie d'aide à l'accès aux données puisqu'ils permettent d'exprimer des requêtes désignant un mot contenu dans le document même indépendamment du sens que ce mot peut avoir dans le document. Par contre, les répertoires tiennent compte de ce sens et constituent donc une aide à l'accès à des informations.

Un autre service consiste à centraliser l'accès à plusieurs ressources qui sont susceptibles d'être requêtées dans des contextes proches : infrastructures sur des informations techniques, encyclopédies générales, réseaux de recherche, etc. Outre l'identification des sources d'informations pertinentes, ces démarches se concrétisent par la création d'un portail d'entrée sur ces sources. Dans ce domaine, l'information géographique est souvent utilisée comme une clé d'accès, en particulier le nom d'un lieu auquel le contenu de la ressource est associé. Par exemple, le système GIPSY analyse dans un document les désignations de noms de lieux pour indexer ensuite automatiquement ce document dans un gazetier¹ [Larson 96]. Il ne s'agit pas là d'accès à des données géographiques mais du recours à l'information géographique pour organiser un accès. [Laurini and Thompson 92] introduisent la métaphore d'« hypermap » pour structurer l'accès à de nombreuses informations grâce à leurs localisations.

Enfin, le « courtage » (*brokering* en anglais) d'informations ne consiste pas seulement à créer un portail unique pointant sur plusieurs ressources mais aussi à offrir un accès unique à ces ressources comme si elles étaient une seule ressource. Cela signifie qu'un utilisateur exprime un besoin non pas dans autant de langages qu'il existe de ressources mais dans un langage unique. Le service de courtage propage alors la requête vers chaque ressource.

¹ Le terme gazetier est utilisé à la place du mot anglais *gazetteer*, et de façon équivalente à l'expression *index géographique*.

Une étape délicate dans les infrastructures d'information est de construire le langage de description et de décrire effectivement les ressources. Dans la suite de cette section, nous présentons les langages dédiés à la représentation et à la manipulation d'informations. Nous présentons ensuite les langages spécifiquement dédiés à la description de ressources en information : les métadonnées.

Les langages de description d'informations

Il existe des langages qui ne sont pas dédiés à stocker des données et à optimiser des opérations de sélection mais plutôt à construire des ressources, et à optimiser des opérations d'interprétation sur ces ressources.

Les logiques de description sont par exemple souvent utilisées pour construire des bases de connaissances en IA. La structure proposée par de tels langages ne vise pas à permettre des opérations de sélection, il ne s'agit pas de langage complet comme le langage relationnel utilisé dans les SGBDR, mais elle permet de nombreuses inférences que la structure relationnelle ne permet pas. Il s'agit de modèles dits semi-structurés. Ces modèles sont beaucoup employés pour représenter des sources hétérogènes [Siméon 99].

Un autre exemple entrant dans cette catégorie de notre état de l'art des langages de description d'information est le langage HTML, *HyperText Markup Language*, utilisé pour construire une ressource Web. Les documents HTML ont une structure explicitée sous forme de balises. Cette structure est dédiée à l'affichage de ces pages HTML dans des navigateurs Web. Le langage XML, *eXtensible Markup Language*, est un langage standard d'échange d'informations, conçu pour le Web mais pouvant être utilisé dans d'autres contextes [W3C/XML/WG 00]. C'est un langage à balises comme HTML. Alors que HTML propose une structure dédiée à la présentation du document, XML ne propose pas une structure unique mais des mécanismes d'écriture de structure. Une structure XML est exprimée sous forme d'une DTD, *document type definition*, ou d'un XMLSchema. XML est utilisé pour structurer un document en vue de manipulations autres que sa présentation sur le Web. DAML, *DARPA Agent Markup Language*, est un langage étendant XML pour décrire plus simplement les liens entre objets.

Enfin, dans le domaine de l'informatique, les langages orienté-objet sont des langages que nous classons dans cette section car ils sont plus intéressants dans une perspective d'expressivité que la technique relationnelle. Par contre ces langages n'offrent pas les propriétés d'accès rapide aux données liées aux opérations possibles dans l'algèbre relationnelle. Notons de plus que ces langages ont une fonctionnalité de fédération de ressources. Un langage orienté-objet permet en effet de fédérer des objets différents en créant une classe regroupant leurs caractéristiques communes et en représentant chaque objet comme une instance de cette classe. Il permet également de fédérer des classes ayant des éléments communs –variables ou méthodes- en créant une sur-classe dont elles héritent. Il permet également de fédérer des classes non pas sur leur contenu mais sur leur manipulation en les faisant hériter d'une classe abstraite : les noms des méthodes sont définies au niveau de cette classe abstraite mais leur contenu est spécifié dans chaque sous-classe.

Les langages et modèles présentés ci-dessus sont instanciés « à la main » pour construire ou décrire des données. Il existe aussi des modèles de dérivation d'informations associés à des fonctions de mesure spécifiques qui dérivent les informations des données pour valuer le modèle. Par exemple, les techniques de modélisation et de mesure floue permettent d'avoir accès à une information en étant moins dépendant des valeurs mêmes mais plutôt de l'interprétation qu'on en fait. Les modèles statistiques, eux, permettent de réduire un grand ensemble de données à des comportements et valeurs significatives. Ces modèles sont exploités dans les techniques de *data mining*.

Les langages de description de ressources, les métadonnées

Il existe également des langages qui servent non pas à construire une ressource en information mais à décrire des ressources disponibles. Les descriptions des ressources sont généralement appelées métadonnées.

Les métadonnées sont des « données relatives à d'autres données et destinées à supporter des traitements impliquant ces autres données » [Role 99]. L'auteur ajoute que « même si elle est utilisée dans un grand nombre d'articles, il semble ne pas exister de définition vraiment formelle et précise de cette notion, ce qui est sans doute dû à sa grande généralité. » La plupart du temps, le langage utilisé pour construire des métadonnées est un langage champ-valeur. Pratiquement, remarquons qu'une métadonnée sur une ressource est :

- une donnée qui porte une information sur la ressource qui ne peut pas être retrouvée dans la ressource même,
- ou une donnée qui porte une information qui peut être retrouvée dans la ressource, mais cela permet d'exploiter cette information au niveau de la métadonnée sans avoir à manipuler la ressource elle-même.

[Kashyap 97] propose une classification des types de métadonnées :

- les métadonnées dépendantes du contenu (ex : la langue employée dans un document),
- les métadonnées descriptives du contenu (ex: les thèmes abordés dans un document),
- les métadonnées indépendantes du contenu (ex : la date de création d'un document).

Une autre classification porte sur les communautés utilisant les métadonnées, [Role 99], Figure 23.

Communauté	Utilisation
Acteurs du Web	<ul style="list-style-type: none"> - créer des index permettant d'effectuer des recherches documentaire précises (métadonnées du Dublin Core utilisées dans des robots de recherche) - bloquer l'accès à certains sites (métadonnées PICS utilisées dans des programmes de filtrage de l'information) - authentifier un document - gérer les droits
Bases de données	<ul style="list-style-type: none"> - représenter des correspondances entre attributs et contribuer à l'interopérabilité entre bases hétérogènes - générer dynamiquement des requêtes SQL à destination de bases dont le schéma est a priori inconnu - identifier les ressources intéressantes à consulter - associer à une base de données multimédia des métadonnées permettant d'effectuer une corrélation entre des données de types différents - associer à des bases de données accessibles en mode client/serveur des métadonnées permettant à des clients ne connaissant pas a priori les caractéristiques de ces bases de les interroger
Edition électronique	<ul style="list-style-type: none"> - structurer de vastes corpus de textes
Informatique "décisionnelle" et informatique de gestion	<ul style="list-style-type: none"> - gérer des entrepôts de données
Objets distribués	<ul style="list-style-type: none"> - fournir des informations sur les méthodes de façon à permettre une invocation dynamique de ces dernières

Figure 23 : Communautés utilisant des métadonnées, d'après [Role 99]. [Figure 23 : Communities using metadata, after [Role 99]]

Ces ressources étant décrite en termes d'informations, les langages utilisés pour construire des métadonnées peuvent être des langages de description d'information tels que ceux décrits dans la section précédente. Par ailleurs, pour associer des métadonnées à une ressource existante, il est parfois possible de s'appuyer sur le langage qui a servi à construire la ressource. Ces notions sont résumées sur le tableau Figure 24.

Ainsi, une métadonnée naturelle d'un document XML est sa DTD. Cette DTD peut également servir à désigner des portions du document, correspondant à certaines balises, comme étant des métadonnées. De même, pour décrire un document HTML, certains moteurs de recherche se réfèrent à la DTD de HTML et décrivent un document HTML par des portions particulières de ce document –i.e. les zones de textes- figurant entre les balises titre, les balises résumé, ou encore entre les balises indiquant des liens hypertexte. Il constitue alors des index associant à un mot les URL des pages Web dans la description desquelles figure ce mot-clé [Chartron 97]. L'utilisateur exprime une requête par mots-clés que le moteur compare aux indexes. C'est ce qui est désigné comme *text retrieval* dans [Guarino et al. 99]. Les auteurs ajoutent que des travaux s'appuyant sur l'analyse des

cooccurrences, l'analyse morphologique et les racines des mots permettent de donner de bonnes performances à ces systèmes. Les limites de tels systèmes restent le bruitage de la description et la pauvreté de l'expression de requêtes.

Les métadonnées jouent un rôle important dans la fédération des ressources sur un domaine grâce à la mise en place de standards de métadonnées dédié à ce domaine. Le plus générique, c'est-à-dire adapté à une grande variété de ressources, préconisé par le consortium W3C, est le RDF, Resource Description Framework. Ce langage standardise la description de certaines propriétés d'une ressource comme l'auteur ou la date de création.

Le Dublin Core est un standard célèbre dédié aux documents textuels [Weibel et al. 95]. Cette démarche, entreprise en mars 1995 lors d'une conférence en Ohio, veut corriger les faiblesses des descriptions de documents HTML construites classiquement par les moteurs de recherche. Le Dublin Core définit des balises HTML META particulières : le *Sujet* (thème de l'objet), le *Titre*, l'*Auteur*, l'*Editeur*, les *Autres agents*, la *Date* de publication, le *Type* (par exemple roman ou poème), le *Format* informatique, l'*Identifiant*, une *Relation* (quand le document fait partie d'un document plus général), la *Source* (lien vers un document de même contenu intellectuel mais précédent), la *Langue*, les *Couvertures spatiale et temporelle*.

Ce standard peut être utilisé pour enrichir un document HTML textuel en y ajoutant ces balises, ou pour créer un document HTML comportant ces balises et comportant une référence à un document textuel non HTML, comme un livre papier.

Ressource	Document HTML	Document HTML respectant le DC	Livre papier	Document XML	Document XML écrit avec une DTD comportant des balises spécifiques de métadonnées
Métadonnées	Mots figurant dans les sections <TITLE> et <HEAD> du document	Valeurs des balises du DC dans la ressource	Document HTML comportant les balises du DC et leurs valeurs pour le livre, et une référence vers l'emplacement du livre	DTD utilisée pour écrire le document	Valeurs des balises spécifiques
Standard de Métadonnées utilisé	Aucun	Dublin Core (DC)	DC	Aucun	La DTD

Figure 24 : Divers mécanismes de description de documents. [Figure 24 : Varied mechanisms to describe documents on the Web.]

Le principal problème des métadonnées est que la construction de ces descriptions pour les ressources existantes est une entreprise coûteuse. Or l'efficacité des systèmes de recherche s'appuyant sur des métadonnées reposent sur le principe que les producteurs de ressources acceptent des métadonnées de leurs ressources, conformes au standard de métadonnées. L'enjeu est tel que les courtiers d'information n'hésitent pas à la prendre en charge [Guarino et al. 99].

Lorsque des métadonnées sont employées pour gérer l'accès à des ressources, il est nécessaire qu'elles permettent de décrire suffisamment précisément ces ressources, qu'elles donnent la possibilité de distinguer des ressources les unes des autres et de tenir compte de cette distinction au moment où on cherche la ressource la plus utile pour répondre à un besoin. Lorsque les ressources sont nombreuses, cela nécessite une assez grande spécificité des métadonnées, c'est-à-dire qu'elles vont prendre leur sens dans un domaine limité. Pour être efficace, la recherche d'information basée sur les métadonnées fonctionne souvent sur un domaine donné. Ainsi, il existe des métadonnées spécifiques à des domaines, par exemple la précision de données scientifiques ou les droits d'accès. Le *Warwick Framework* améliore le standard du Dublin Core en définissant des paquets de métadonnées dédiés à la description de certains types de ressource [Lagoze et al. 96]. Le Dublin Core, regroupé dans le paquet DC, est dédié au catalogage de documents textuels. Concrètement, le titre est codé comme : <META NAME="DC.Title" CONTENT="Presentation">.

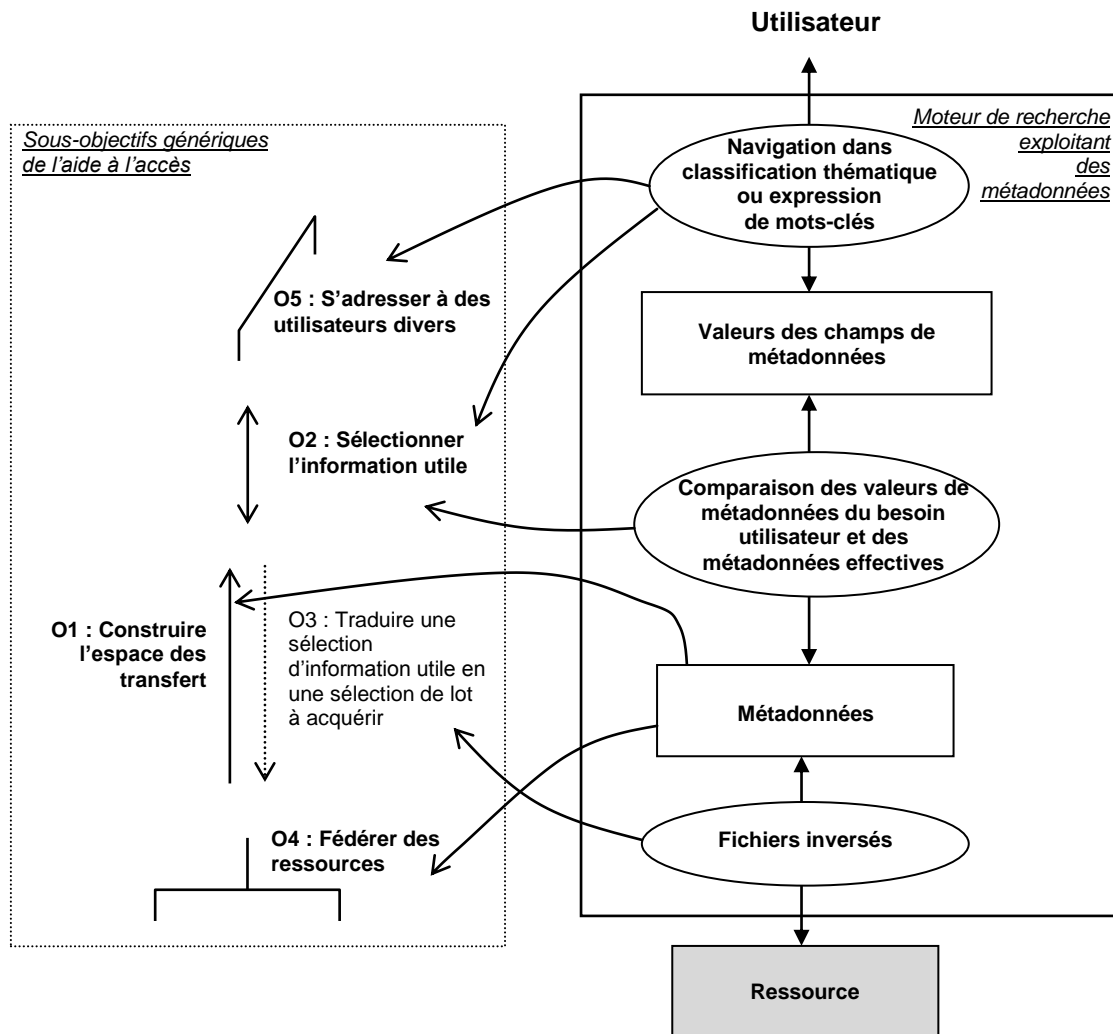


Figure 25 : Techniques mises en œuvre pour aider l'accès d'utilisateurs à des informations dans les moteurs de recherches utilisant des métadonnées. [Figure 25 : Users' access techniques in search engines relying on metadata.]

Les modules d'interprétation de requêtes

Il est possible d'utiliser des bases de données spécifiques pour interpréter des requêtes d'utilisateurs, c'est-à-dire pour relier les mots utilisés dans la requête avec les termes utilisés dans la description des ressources, par exemple les valeurs de métadonnées. Une telle base de données est un thesaurus, qui associe des mots à des concepts et explicite des relations entre mots, comme la synonymie, ou entre concepts, comme la composition ou la généralisation. Un thesaurus peut comporter des définitions de mots comme un dictionnaire, mais il comporte essentiellement des informations permettant de situer un mot par rapport à d'autres.

Un thesaurus célèbre sur le Web est Wordnet développé au Cognitive Science Laboratory de Princeton University [Miller 98]. Il est utilisé dans le prototype de [Guarino et al. 99] comme module d'interprétation de requêtes. Wordnet est une base de données lexicographique portant sur la langue anglaise et disponible gratuitement. Cette base est actuellement étendue en une base multilingue, EuroWordnet [Vossen et al. 97]. Wordnet comporte des mots, et des ensembles de mots synonymes, les synsets, associés à des concepts. Il comporte également des informations supplémentaires comme les types des concepts (noms, verbes, adjectifs) et des relations entre mots ou concepts. Ces relations sont :

- la synonymie ou l'antonymie entre mots,

- les relations hiérarchiques entre concepts, par exemple un pigeon est plus spécifique qu'un oiseau,
- les relations partitives entre concepts, par exemple le nerf optique fait partie du système nerveux, Paris est inclus dans la France.

Pour être efficaces, de tels modules d'interprétation doivent fonctionner sur des thèmes spécifiques. Dans l'architecture de bibliothèque virtuelle décrite par [Papazoglou and Hoppenbrouwers 99], des ressources sont indexées par des termes et des bases de connaissances sont utilisées pour traiter une requête d'utilisateur selon trois niveaux :

- Les thèmes, par exemple "micro-économie", "macro-économie", "économie de développement".
- Les concepts, par exemple des concepts associés au thème "micro-économie" sont "modèle de marché", "économie industrielle", "économie des ménages".
- Les termes, par exemple des termes associés au concept "modèle de marché" sont "monopole", "oligopole".

L'utilisation de tels modules d'interprétation permet aux utilisateurs d'exprimer leurs requêtes dans leurs propres syntaxes et contribue donc à O5, l'aide à l'accès d'utilisateurs divers. Notons que la constitution de dictionnaires peut servir à partager une syntaxe mais aussi à partager une conceptualisation. La construction de bases de connaissances servant à partager des connaissances sur un domaine donné sera détaillée dans la section 2.1.3.2 portant sur les démarches d'aide au raisonnement formel.

2.1.2.3 L'aide à l'accès aux connaissances

Les démarches d'aide à l'accès aux connaissances sont les démarches qui permettent que des ressources s'intègrent dans un raisonnement mené par l'utilisateur. Différents types de raisonnements sont distingués en IA [Haton et al. 91]. Ils font écho aux activités mentales de l'homme que les chercheurs en IA ont voulu reproduire. Une aide à l'accès à des connaissances peut se faire en favorisant l'un ou l'autre raisonnement sur des ressources.

Le raisonnement par analogie est un raisonnement très répandu chez l'homme, mais difficile à reproduire. Il repose sur l'usage de métaphore. Une métaphore est une correspondance entre un domaine source –sur lequel on sait raisonner- et un domaine cible –sur lequel on ne sait pas raisonner-. Le résultat de raisonnements effectués sur des éléments du domaine source, et qui appartient lui-même à ce domaine source, est mis en correspondance avec le résultat de raisonnements *similaires* qui seraient effectués avec les éléments du domaine cible correspondants aux éléments du domaine source sur lesquels ont été mené le premier raisonnement. La plupart des structures que nous utilisons sont issues de métaphores [Kuhn and Frank 91]. Le domaine source privilégié est celui de nos expériences sensorielles et motricielles comportant des concepts et relations universels –surface, lien, chemin, proche, loin, partie, tout, ...- [Mark 92].

L'usage de métaphore est donc un mode privilégié d'accès au raisonnement sur un domaine, le domaine cible. L'utilisation d'un domaine source accessible à la plupart assure grandement O5. L'usage d'une métaphore présente pourtant un défaut majeur. Rigoureusement elle permet à l'utilisateur de ne mener que les raisonnements prévus dans la construction de la métaphore. Or il est tentant de mener des raisonnements courants sur le domaine source, non prévus dans la construction de la métaphore et qui ne correspondent à rien ou, et c'est là le danger, à un raisonnement faux dans le domaine cible.

Le raisonnement géométrique est étudié par les recherches sur la reconnaissance et la manipulation d'objets physiques, typiquement le traitement d'images, la planification des mouvements d'un robot, la vision par

ordinateur. Nous n'avons pas trouvé de démarche entrant dans cette catégorie, hormis peut-être la production de schémas, graphes, cartes.

Un autre type de raisonnement est le raisonnement par généralisation et abstraction, également très répandu chez l'homme. Les classifications thématiques, résultant souvent d'abstractions dans les métadonnées, par exemple la classification de mots-clés de Yahoo, permettent à un utilisateur d'effectuer ce type de raisonnement.

Certains systèmes construisent, au fur et à mesure de requêtes d'utilisateurs, des classes de consultations typiques. Lorsqu'une personne utilise un système possédant de telles classes de consultations, le système détermine la classe dans laquelle il peut ranger ce nouvel utilisateur. Il répond alors à l'utilisateur en tenant compte de ce contexte qu'il a reconstruit, c'est-à-dire de la classe de consultation typique à laquelle appartient l'utilisateur. Par exemple, le système indique alors à l'utilisateur des ressources qu'il n'a pas demandées mais qui « typiquement devraient l'intéresser ».

Le site suivant applique des techniques de filtrage collaboratif pour conseiller un utilisateur sur les films à aller voir : <http://www-poleia.lip6.fr/~fconseil/>. Ce conseil se fonde soit en regardant les films proches de ceux que l'utilisateur est allé voir, soit en regardant les films vus par les utilisateurs qui lui ressemblent, soit en mixant les deux.

Les services d'aides reposant sur les capacités de raisonnement par généralisation et abstraction d'utilisateurs sont très efficaces pour O4, la fédération de ressources. Par contre, au sein d'une classe de la granularité la plus fine, l'utilisateur doit lui-même sélectionner les ressources utiles parmi celles appartenant à cette classe.

Le raisonnement le mieux maîtrisé en IA est le raisonnement formel. Il s'agit de « la manipulation syntaxique de structures symboliques à l'aide de règles, dans le cadre d'une certaine sémantique» [Haton et al. 91]. La sémantique est une fonction qui interprète ces structures symboliques et leurs manipulations.

Dans un contexte d'aide à l'accès, le raisonnement est le processus d'accès. Les structures symboliques manipulées dans ce raisonnement peuvent être les ressources mêmes ou des descriptions des ressources, ou de descriptions des besoins d'utilisateurs.

Les techniques d'aide au raisonnement formel permettent de construire les structures symboliques explicitant le processus d'accès. Elles permettent aussi d'implémenter des composants qui réalisent des raisonnements formels (plus exactement, procéduraux), et de répartir un même raisonnement global entre plusieurs composants tels que chacun effectue la part du raisonnement qui correspond le mieux à ses compétences.

Les techniques s'appuyant sur le raisonnement formel sont extrêmement intéressantes dans notre travail, nous leur consacrons la section suivante, 2.1.3.

Le raisonnement procédural est un cas particulier du raisonnement formel, où toutes les connaissances, la façon de les utiliser et la conduite du raisonnement sont figées. Les interfaces entre composants implémentés ont pour objectif de permettre le raisonnement procédural d'un composant à l'aide de connaissances provenant d'un autre composant. L'aide au raisonnement procédural entre composants renvoie à la problématique de l'interopérabilité introduite section 1.2.1.1. Comme le souligne [Sheth 99]: « Any attempt to give a broad survey of information systems interoperability is difficult and complex for several reasons, including different levels of requirements, the variety of approaches, the number of technical areas involved, the large literature, and the large number of example systems »*. Nous ne cherchons pas à analyser en détail l'état de l'art des contributions à l'interopérabilité mais nous rappelons quelques résultats importants dans ce domaine.

Cinq stratégies de base sont classiquement envisagées pour atteindre l'interopérabilité [Paepcke et al. 98][Abel et al. 99]:

* Toute tentative pour résumer l'interopérabilité de systèmes d'information est difficile et complexe pour plusieurs raisons qui comprennent : les divers niveaux d'interopérabilité, la variété des approches, le nombre de domaines techniques impliqués, la littérature abondante et le nombre important d'exemples. (tda)

- Les standards forts qui imposent une homogénéité à divers niveaux de l'architecture, comme par exemple ISO802, TCP/IP, DOS, Z39-50, HTTP.
- La famille de standards communiquant entre eux. On essaye d'organiser les standards en niveaux communicants correspondant aux relations effectives entre les éléments qu'on cherche à normaliser. C'est le cas par exemple avec OSI.
- L'approche du médiateur externe, qui fonctionne comme un interprète entre composants. Il s'agit de l'équivalent de la fédération de bases de données pour des informations moins structurées : il peut s'agir d'un moteur de recherche sur des pages HTML, d'un médiateur de données semi-structurées.
- L'utilisation d'un langage commun de communication entre des modules, grâce à une spécification respectée par tous les modules (cf. KIF, KQML, SETL).
- La fonctionnalité mobile, par exemple les applets java, les éléments en langage LISP.

[Sheth 99] constate une évolution dans les démarches en interopérabilité, selon la façon dont sont traitées les notions de distribution, d'autonomie et d'hétérogénéités.

La 1^{ère} génération s'arrête environ en 1985. La distribution est réduite au contexte d'un système interne à une entreprise. La plupart des systèmes de bases de données de cette période ont traité le choix de leur autonomie (on peut consulter [Sheth 99] pour les différents niveaux d'autonomie d'un système) en fonction des contraintes de mise à jour.

La 2^{ème} génération court de 1985 à 1995 environ. Avec Internet, la distribution d'un système peut se faire entre plusieurs entreprises. Par contre les systèmes de cette génération ne se sont pas préoccupés des notions de mise à jour et des problèmes précis d'autonomie.

La 3^{ème} génération a commencé vers 1996. On ne parle plus de système distribué mais de système global, les utilisateurs finaux ne voient pas la nature distribuée de ce qu'ils exploitent. Les enjeux de cette 3^{ème} génération comprennent les modes de visualisation, l'utilisation d'une variété de moyens de communication.

Cette 3^{ème} génération gère plus l'interface avec les utilisateurs finaux. Dans cette mesure elle se dégage du raisonnement procédural pour se rapprocher du raisonnement formel. Finalement, l'enjeu ultime de l'interopérabilité entre acteurs numériques est qu'un acteur humain puisse les faire travailler ensemble, leur assigner un objectif.

2.1.3 L'aide au raisonnement formel

Cette section présente des avancées essentielles de l'Intelligence Artificielle dans la formalisation du raisonnement et son automatisation. Nous pensons que ces travaux sont intéressants dans notre contexte car l'étude de l'automatisation du raisonnement formel doit permettre de comprendre jusqu'à quel point ce raisonnement est maîtrisé et révèle le contrôle que l'on possède sur les connaissances intervenant dans ce raisonnement. Or la difficulté de l'aide au raisonnement est justement de contrôler les connaissances intervenant dans ce raisonnement.

2.1.3.1 Généralités

La formalisation du raisonnement

Une formalisation élémentaire d'un raisonnement consiste à expliciter le fil du raisonnement en question dans des structures symboliques et des opérations sur ces structures, c'est-à-dire un modèle de résolution de problème.

Cela nécessite de décrire les états du domaine du problème (par exemple, la position de pions sur un jeu de dames), les actions possibles pour passer d'un état à un autre (les manières possibles de bouger un pion), les décisions prises pendant le raisonnement (la stratégie du joueur de dames et son application au choix d'une action).

Rappelons que dans notre contexte nous ne voulons pas aider un raisonnement particulier mais un ensemble de raisonnements particuliers : les processus d'accès de chaque utilisateur potentiel. Dans ce contexte spécifique, un modèle efficace doit convenir pour décrire tous ces raisonnements. Lorsque l'ensemble de ces raisonnements est trop grand pour être explicitement décrit, le modèle doit décrire explicitement certains raisonnements génériques et posséder des connaissances de spécification qui permettent d'adapter le modèle à chaque raisonnement particulier de cet ensemble (le processus d'accès d'un utilisateur particulier).

Les premières expériences de formalisation du raisonnement formel avaient pour but de faire raisonner des machines. Les limites des systèmes ainsi construits ont rapidement conduit à la conclusion suivante : l'homme a recours à d'autres connaissances que celles décrivant directement le problème pour raisonner. Il s'agit de *l'expertise*. L'expertise dans un domaine est un ensemble de connaissances permettant de résoudre les problèmes de ce domaine.

D'autres principes ont ensuite été mis en évidence. L'expertise d'un domaine ne se réduit pas à des connaissances déclaratives. Le contrôle de la résolution d'un problème fait également partie de cette expertise, c'est-à-dire qu'il est spécifique au problème. Cela rend extrêmement complexe la traduction de l'expertise d'un homme dans un formalisme car on ne possède a priori pas de structure générique de cette expertise permettant de guider cette traduction.

En réponse à cela, l'acquisition des connaissances vise à élaborer des méthodes pour reconstruire formellement le savoir d'un expert. Elle se situe entre deux pôles : le pôle de l'expert humain, étudié par le génie cognitif, et le pôle de la machine, étudié par le génie logiciel [Le Roux 94].

Les points intéressants des formalismes mis en place par l'acquisition des connaissances sont :

- le traçage du raisonnement,
- la possibilité de faire des systèmes multi-experts c'est-à-dire dont la connaissance est organisée en plusieurs modules d'expertise participant à des phases différentes du raisonnement,
- la possibilité de faire de l'IA distribuée c'est-à-dire de répartir l'activité de raisonnement non pas entre un homme et une machine mais entre plusieurs machines.

Ces principes correspondent à un recentrement des objectifs de l'IA des machines vers le raisonnement de l'homme. Comme le rappelle [Aussenac 89], les enseignements de l'IA ont rapidement remis en question le principe « tout comprendre à partir d'une théorie puis tout automatiser ». La démarche de construction d'un système à base de connaissances repose maintenant davantage sur la définition de « comment l'intelligence va être partagée entre un *agent-machine* et un *agent-humain* ». L'auteur de [Aussenac 89] renvoie à l'affirmation de G. Boy : « l'optimum pour le système ne se situe pas au degré d'automatisation maximum, mais à un point d'automatisation qui optimise le résultat de son utilisation par l'*agent-humain* . »

Aider les hommes à raisonner peut aussi signifier « aider les hommes à raisonner ensemble ». Pour reprendre la formulation de G. Boy, nous pouvons dire que la formalisation de connaissances se focalise non seulement sur le partage de l'intelligence entre un agent-humain et un agent-machine mais également sur le partage de l'intelligence entre plusieurs agents-humains. Un troisième pôle de l'acquisition des connaissances, outre le pôle de l'expert humain et celui de la machine, est donc le pôle de la communauté d'experts ou d'hommes. Ce pôle renvoie essentiellement à une réduction des différences de subjectivité entre hommes pour atteindre un

consensus. Le consensus est atteint lorsqu'il n'y a plus d'incohérences entre les connaissances possédées par les uns et les autres, et non lorsque les hommes possèdent tous les mêmes connaissances.

Une nouvelle catégorisation des connaissances : le « QUOI » et le « COMMENT »

Parmi d'autres, [Gomez and Benjamins 99] énoncent un principe majeur en acquisition des connaissances : "problem solving methods and ontologies can be seen as complementary reusable components to construct knowledge systems from reusable components."*

Ce principe est double. 1) Il y a deux catégories de connaissances, et un composant réutilisable doit appartenir à l'une de ces catégories. 2) Ces catégories sont complémentaires : toute connaissance peut être décomposée en une partie de connaissances de résolution de problème, et une partie de connaissances ontologiques.

Cette séparation en deux catégories, qui veut dépasser les limites de l'ancienne séparation déclaratif-procédural, est souvent appelée le *QUOI* et le *COMMENT* [Le Roux 94] :

- le *QUOI* correspond à la description d'un domaine d'intérêt, ce sont les connaissances factuelles d'un domaine,
- le *COMMENT* porte sur les modes de manipulation du *QUOI* pour produire de nouvelles connaissances venant enrichir ce *QUOI*. Il s'agit des connaissances de résolution de problème.

Cette séparation est justifiée par le *problème d'interaction* : "the interaction problem states that representing knowledge for the purpose of solving a problem is strongly affected by the nature of the problem and the inference strategy to be applied to the problem"* [Gomez and Benjamins 99]. Les représentations qui mélangent du contenu avec des structures dédiées à une manipulation donnée peuvent rendre difficile la réutilisation du contenu décrit dans une autre manipulation.

La prise en compte de ce problème d'interaction est intéressante dans un contexte d'aide à l'accès à des ressources car elle se traduit par la séparation, au sein d'une description, du contenu des ressources de leur manipulation. Cela permet qu'une même ressource puisse être associée à plusieurs manipulations possibles, c'est-à-dire puisse répondre à divers besoins. Cela permet aussi d'appliquer à des ressources variées une même expertise de manipulation.

2.1.3.2 La modélisation axée sur le *QUOI* : les ontologies de domaine

Les modèles de type *QUOI* visent à décrire, dans des expressions symboliques, des parties du monde. Ces modèles, appelés ontologies, sont utilisés pour construire des descriptions très générales ou dans les limites d'un domaine d'intérêt. Dans un contexte d'aide à l'accès, ils sont utilisés pour fédérer des ressources ou pour construire des modules d'interprétation de requêtes d'utilisateurs. Ces techniques ont été décrites dans la section 2.1.2.2. Nous revenons dans cette section sur les apports de l'acquisition des connaissances sur la construction d'ontologies.

* Les méthodes de résolution de problèmes et les ontologies peuvent être considérés comme des composants complémentaires réutilisables pour construire des composants de connaissances qui puissent être partagés. (tda)

* Le problème d'interaction signifie que la représentation de connaissances en vue de résoudre un problème est fortement affectée par la nature du problème et de la stratégie d'inférence employée. (tda)

La notion d'ontologie

Le fondement théorique de ces démarches est la notion d'ontologie en philosophie. Le mot ontologie semble un peu compliqué sans doute parce qu'il a plusieurs acceptions –en philosophie, en métaphysique, en intelligence artificielle...- et qu'il n'est pas entré dans le langage courant. Nous rappelons rapidement son sens en philosophie : « branche des sciences métaphysiques qui étudie la nature et les propriétés et relations essentielles de ce qui est en tant que tel, ou des principes et causes d'existence » (définition du dictionnaire Webster). La métaphysique est elle-même définie comme « la science de l'existant en référence à ses conditions abstraites et universelles, à distinguer de la science de l'existence déterminée et concrète ».

Pratiquement, la notion d'ontologie renvoie à l'expérience commune de l'homme et à une catégorisation universelle du monde. En effet, face au monde réel, l'homme se livre à des abstractions. Il s'agit pour lui de produire des objets pour représenter le monde par des informations qu'il peut conserver, manipuler dans un raisonnement et communiquer. La première activité d'abstraction est la discrimination assortie de la catégorisation puis de la classification. L'idée est de décrire le monde en termes de concepts abstraits suffisamment isolés les uns des autres pour que l'on puisse dans la plupart des cas décider si un objet du monde réel « est » une instance de tel ou tel concept. Par exemple, pour décrire un paysage à quelqu'un, il est simple de recourir au concept « arbre » -renvoyant ainsi à une image typique de ce concept chez l'interlocuteur- plutôt que de décrire chaque arbre du paysage un à un avec ses caractéristiques particulières. Cet isolement des concepts abstraits les uns par rapport aux autres ne signifie pas que le sens d'un concept est intrinsèque et indépendant des autres.

La notion d'ontologies utilisées en IA renvoie à des formalisations produites pour approcher cette catégorisation absolue désignée par ontologie en philosophie. [Chandrasekaran et al. 98] proposent une définition détaillée du terme ontologie en IA.

« In AI, the term [ontology] has largely come to mean one of two related things :

- A representation vocabulary, typically specialised to some domain or subject matter. More precisely, it is not the vocabulary as such that qualifies as an ontology, but the conceptualisations that the term in the vocabulary are intended to capture. [...] Identifying such terms –and the underlying conceptualisations—generally requires careful analysis of the kinds of objects and relations that can exist in the domain.[...]
- Occasionally, a body of knowledge describing some domain, typically a common sense knowledge domain, using such a representation vocabulary. » *

Nous considérons ici qu'une ontologie est la conceptualisation sous-jacente à un vocabulaire. En cela nous adoptons le point de vue de [Gruber 93] très répandu dans les systèmes d'information : « une ontologie est la spécification explicite d'une conceptualisation ». Cette définition correspond au premier type d'ontologies présenté par [Chandrasekaran et al. 98]. La base d'instances associée à une telle ontologie correspond au deuxième type.

[Chandrasekaran et al. 98] rappellent que le premier intérêt des ontologies est de clarifier la structure de la connaissance. L'analyse ontologique d'un domaine de connaissance doit précéder la représentation de ces connaissances et permettre d'éviter les incohérences lors de la construction de la base de connaissances. [Guarino 98] puis [Guarino and Welty 00] définissent des principes essentiels pour construire des ontologies saines et exploitables. Cela consiste à gérer rigoureusement les notions d'identité et d'unité représentées dans l'ontologie. Les auteurs de [Chandrasekaran et al. 98] donnent un exemple d'incohérence liée à une mauvaise

* En IA, le terme ontologie a actuellement de façon générale l'une ou l'autre des significations suivantes, qui sont liées : 1) Le vocabulaire d'une représentation, typiquement spécifique à un domaine ou à un problème. Plus précisément, l'ontologie n'est pas le vocabulaire lui-même mais la conceptualisation sous-jacente.[...] Identifier ces concepts demande généralement une analyse minutieuse des types d'objets et de relations qui peuvent exister dans le domaine. 2) Un corpus de connaissances décrivant un domaine, typiquement un domaine de sens commun, en utilisant un tel vocabulaire. (tda)

analyse ontologique. Cet exemple provient d'une base de connaissances portant sur un domaine où il y a des personnes, dont certaines sont étudiants, d'autres enseignants ou autres employés. Une ontologie simple a d'abord été utilisée où les classes étudiants, employés, enseignants, homme ou femme, étaient représentées comme des *sortes-de* humains. Des problèmes ont surgi liés au fait que des étudiants pouvaient aussi être parfois des employés, ou encore cesser d'être étudiant. Une deuxième analyse ontologique a alors montré que les classes étudiants, employés, et enseignants étaient plutôt des *rôles* tenus par les humains, et non des sous-classes d'humains comme les classes hommes et femmes.

Même si les ontologies sont classiquement axées sur la modélisation du *QUOI*, elles sont rarement indépendantes du *COMMENT*. [Guarino 97] caractérise les ontologies selon leur niveau de dépendance d'une tâche ou d'un point de vue particulier :

- les ontologies les plus générales contenant des concepts comme le temps, l'espace, l'action, l'événement,
- les ontologies de domaine posant le vocabulaire employé dans un domaine,
- les ontologies de tâche décrivant une activité générique,
- les ontologies d'application qui spécialisent les deux précédentes en indiquant le rôle joué par les entités du vocabulaire dans l'activité.

[Chandrasekaran et al. 98] analysent les ontologies de haut niveau présentes dans la littérature, comme CYC de [Lenat and Guha 90], pour faire une synthèse de ce qui est communément admis : Il y a des *objets* dans le monde, ils ont des *propriétés* qui peuvent prendre des *valeurs*. Les objets peuvent exister en ayant diverses *relations* les uns avec les autres. Les propriétés et relations peuvent changer avec le *temps*. Il y a des *événements* qui se produisent à différentes *dates temporelles*, il y a des *processus* auxquels participent des objets et qui se produisent dans le temps. Le monde et ses objets peuvent être dans différents *états*. Des événements peuvent causer d'autres événements comme *effets*. Les objets peuvent avoir des *parties*. Et ainsi de suite.

Parmi les ontologies célèbres, CYC est un effort de construction d'une ontologie universelle de sens commun qui se poursuit depuis une quinzaine d'années. Des efforts portent actuellement sur l'aide à la construction d'ontologies. Ontolingua est un éditeur d'ontologies proposé par l'université de Stanford. Le terme « ontolingua » renvoie en fait au langage utilisé pour construire les ontologies. Ce projet met également en place un serveur d'ontologies.

Les ontologies sont utilisées dans plusieurs contextes, comme la transmission d'un concept, la traduction d'un concept d'un modèle dans un autre. Cela permet souvent de fédérer des points de vue et des représentations. Dans une optique d'aide à l'accès, les ontologies sont utiles pour fédérer des sources d'informations portant sur un même domaine et pour construire des modules d'interprétation de requêtes.

Les outils de manipulation d'une ontologie sont la classification d'un concept ou d'une instance ainsi que la détermination d'identité entre deux concepts. Notons que les nomenclatures sont des ontologies trop simples pour de tels outils. Le monde réel représenté y est décrit par une structure hiérarchique de classes d'objets, où les liens entre ces classes sont des liens de subsumption, « est une sorte de ». Ce type de représentation, trop simple, ne peut pas se partager facilement. Il ne s'y trouve pas des liens importants entre les classes qui permettraient par exemple d'exprimer des correspondances d'une nomenclature à l'autre. Cela se comprend facilement sur un exemple décrit Figure 26. La nomenclature de l'utilisateur, un propriétaire de camions qui roulent à des vitesses différentes sur les routes selon leur type rA' ou rB', ne peut intégrer la nomenclature d'une carte routière qui distingue trois types rA, rB et rC de routes. Si l'utilisateur calcule des itinéraires avec cette carte il ne sait pas à quelle vitesse ses camions peuvent rouler sur les routes rB, puisqu'il ne sait pas si elles sont dans rA' ou dans rB', et donc il ne peut pas optimiser ses itinéraires.

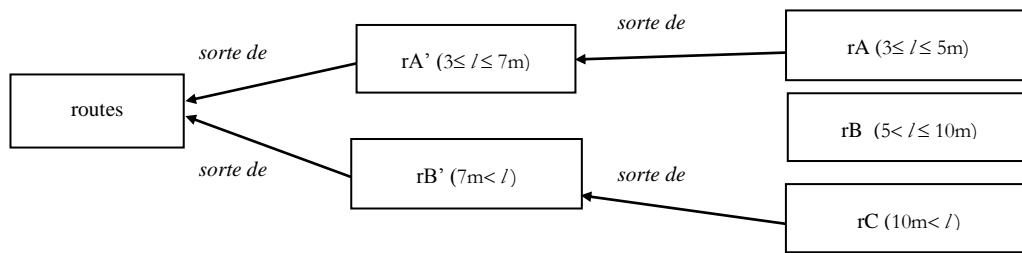


Figure 26 : Un exemple d'incompatibilité de nomenclatures. [Figure 26 : Example of incompatibility between nomenclatures.]

Une démarche importante est OIL, Ontology Inference Layer, mis en place dans le cadre européen IST [Fensel et al. 00]. Alors qu'Ontolingua permet de représenter des ontologies et de les échanger sur le Web, OIL permet en plus de les manipuler sur le Web. OIL reprend diverses techniques de représentation qui chacune apporte une caractéristique intéressante à OIL :

- les *logiques de description* permettent de représenter des connaissances de façon optimale pour conduire des inférences sur ces connaissances,
- les *langages de frame* (proches de l'orienté-objet) possèdent une expressivité plus importante que les logiques de description.

OIL a été intégré dans le langage DAML pour former DAML+OIL [Connolly et al. 01]. Cela deviendra probablement une recommandation W3C pour l'écriture d'ontologies sur le Web.

2.1.3.3 Les modèles axés sur le COMMENT

La description d'outils

La modélisation du COMMENT peut être focalisée sur le développement d'outils et de méthodes, comme les systèmes de règles ou la logique floue [Chandrasekaran et al. 98]. Les systèmes multi-agents sont un modèle de COMMENT très en vogue.

Dans le contexte du partage de connaissances sur le Web, en parallèle aux infrastructures d'informations, des tentatives d'infrastructures de traitements sont expérimentées. Ces efforts s'appuient sur des langages de description des ressources Web fournissant ces services. Le langage Web Service Description Language, WSDL, basé sur XML, permet ainsi de décrire un service en précisant son nom, les méthodes qui peuvent être invoquées, les paramètres et les types d'entrée et de sortie de ses méthodes [Christensen et al. 01]. Lorsqu'un utilisateur veut utiliser un tel service, il exprime une première requête pour obtenir auprès du site serveur la description WSDL de ce service. Puis, à l'aide de cette description, il construit une requête faisant appel à l'application de ce service. Cette requête est adressée au site serveur et exécutée sur ce site serveur. Le résultat est ensuite envoyé à l'utilisateur. Cette description se limite à répondre à la question *avec quoi?* car il ne rattache pas la description du service à des méthodes de résolution de problème décrivant dans quel contexte utiliser ce service.

Une aide plus sophistiquée consiste à donner accès à la composition et l'interopération de services. Des normes de dialogue, comme RosettaNet, sont associées aux normes d'infrastructure pour permettre aux services d'échanger des informations [Crusson et Girard 01]. Le langage DAML+OIL supporte maintenant DAML-S Web Service Markup Language qui permet d'automatiser plusieurs manipulations des services Web :

- la recherche d'un service sur le Web,
- l'invocation d'un service sur le Web,

- la composition et l'interopération de services sur le Web.

DAML-S ne s'appuie pas uniquement sur un langage de description de services mais aussi sur la construction d'une ontologie de processus élémentaires. Cette ontologie doit être suffisamment exhaustive pour que tout service décrit se compose de ces processus. La description d'un service par DAML-S est alors structurée en trois éléments :

- Le profil du service, c'est-à-dire ce qu'il fait. Il s'agit de caractéristiques comme les types d'entrée et de sortie, les préconditions et postconditions.
- Le modèle du service, c'est-à-dire comment il fonctionne. Il s'agit de la décomposition du service selon les processus de l'ontologie.
- Le support du service, c'est-à-dire comment y accéder. Il s'agit essentiellement de la définition des protocoles de communication qui peuvent être employés pour invoquer ce service.

Un exemple d'ontologie de traitements est la base de connaissances construite par [Borgo et al. 97] dans leur prototype de moteur de recherche en ligne de composants orienté-objet Ontoseek. Cette ontologie porte sur des composants orienté-objet disponibles en ligne et décrit ces composants par les tâches qu'ils permettent d'accomplir comme « ouvrir une fenêtre pop-up avec un message dedans ». Ce moteur de recherche permet à l'utilisateur d'exprimer son besoin de ressource en décrivant ce que fait ce composant et non pas ce qu'il est, soit encore, pour reprendre les éléments de description de DAML-S, en décrivant le profil de ce composant. Plus encore, Ontoseek utilise un vocabulaire proche de celui de l'utilisateur pour décrire ce profil.

Dans un contexte d'aide à l'accès, les démarches ne se cantonnent plus à décrire des outils mais intègrent d'autres connaissances. Cela peut être des connaissances de mode d'emploi comme veut le faire DAML-S. Cela peut aussi être des connaissances permettant à l'utilisateur de s'exprimer dans son vocabulaire et non dans celui des ressources, comme dans Ontoseek.

La description d'utilisations : les modèles à base de tâches et de rôles

La modélisation du *COMMENT* peut aussi s'attacher à couvrir entièrement un raisonnement, de l'énoncé de l'objectif à l'exploitation d'outils pour atteindre cet objectif. Les connaissances modélisées sont globalement les connaissances permettant de répondre aux trois questions suivantes :

Pourquoi ? : quel est l'objectif poursuivi dans la manipulation ?

Comment ? : quelles sont les stratégies et les méthodes pour manipuler les objets ?

Avec quoi ? : quels sont les outils utilisés ?

C'est bien vers de tels modèles que tendent les démarches d'aide à l'accès à des outils présentées dans la section précédente. Dans ce même contexte de diffusion d'outils sur le Web, certaines démarches visent d'ailleurs à formaliser des connaissances de mode d'emploi plus avancées que la simple interopération. Elles décrivent des processus métier qui enchaînent des services. Dans ce domaine il n'existe pas encore de norme, mais des propositions sont faites, comme par exemple BPML (Business Processus Management Language) basé sur XML [Crusson et Girard 01].

L'utilisation de tels langages permet de fournir à des utilisateurs de ressources des applications toute faites, décrites comme une séquence des services à exécuter. Toutefois, elle ne permet pas d'exprimer un besoin et de le confronter à des ressources pour sélectionner effectivement les ressources nécessaires et concevoir l'utilisation particulière de ces ressources qui répondra au besoin de l'utilisateur.

Dans le contexte de l'aide à l'accès, il est intéressant d'associer à la description de ressources une description d'un *COMMENT* couvrant tout un raisonnement, c'est-à-dire les *pourquoi*, *comment*, et *avec quoi*. Un tel modèle fait effectivement le lien entre un besoin d'utilisateur et les ressources utiles pour y répondre, comme représenté Figure 27. L'expression du besoin se fait en termes de

réponse à la question « pourquoi ? ». Le modèle fait le lien entre ce « pourquoi ? » et des connaissances de type « avec quoi ? ». Ces dernières connaissances désignent justement la sélection de ressources utiles. L'utilisateur n'a donc pas besoin de connaître les ressources, ni les outils de traitements de ces ressources, ni les modes d'emploi associés. Il lui suffit de s'exprimer en termes d'objectif poursuivi.

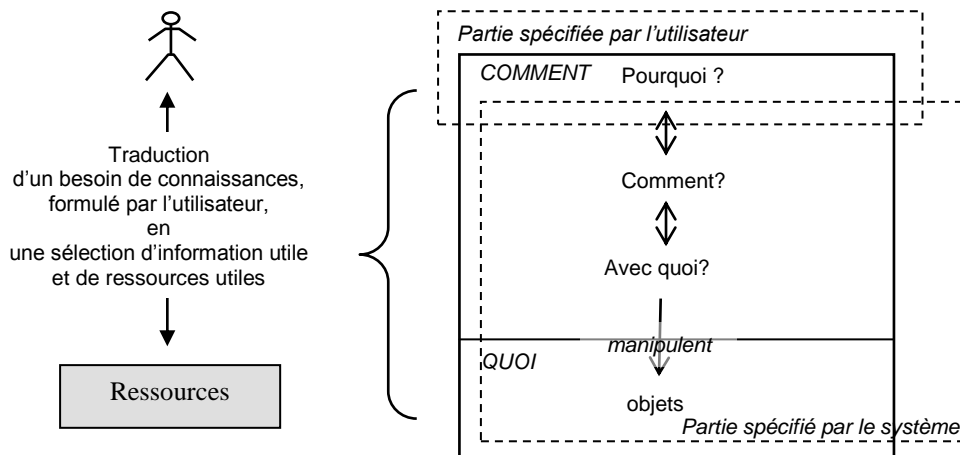


Figure 27 : Contribution d'un modèle intégrant des connaissances de type QUOI et COMMENT à l'accès d'utilisateurs à l'exploitation de ressources. [Figure 27 : Contribution of a model integrating WHAT and HOW knowledge to users' access.]

Une notion essentielle dans une modélisation globale du *COMMENT* est celle de rôle telle qu'elle est utilisée dans [Marcus 88]. Il s'agit d'attribuer aux éléments modélisés dans un domaine d'intérêt un rôle dans la résolution de problèmes portant sur ce domaine, par exemple un critère de classification, une hypothèse. Des méthodes de modélisation sont mises en place pour construire des modèles en s'appuyant sur cette notion de rôle, comme SALT de [Marcus and McDermott 89] qui s'appuie sur les rôles de la méthode générique propose-and-revise.

Une notion complémentaire est celle de tâche. Une tâche consiste à déterminer des connaissances « de sortie » à partir de connaissances « d'entrée ». Plusieurs types de tâches génériques ont été identifiés à partir de travaux en psychologie cognitive, comme calculer, comparer, généraliser. De telles tâches peuvent être spécifiées pour décrire des tâches spécifiques à un domaine, comme le diagnostic médical ou la configuration d'un ordinateur. De façon générale, la résolution d'une tâche ne correspond pas à la résolution d'un problème particulier mais plutôt d'une famille de problèmes, par exemple la famille des calculs de trajets en métro.

L'utilisation d'une description de tâches et de rôles présente de multiples avantages :

- La description de rôle pour décrire indirectement le contenu manipulé permet d'éviter le problème d'interaction présenté précédemment (2.1.3.1).
- La description de tâches intègre des réponses aux *pourquoi*, *comment* et *avec quoi*. En cela, elle permet l'expression d'un besoin, et la traduction de ce besoin en une sélection de ressources et d'un mode d'emploi de ces ressources.
- Une même tâche décrit une famille de problèmes particuliers. La représentation de connaissances de spécification d'une tâche permet à un utilisateur d'utiliser une tâche décrite pour concevoir une résolution de son problème particulier, c'est-à-dire une utilisation précisément adaptée à son besoin.

Un exemple : KADS

Nous détaillons ci-dessous un exemple de modèle qui intègre les notions de tâches et de rôles. Initié en 1983, le projet européen ESPRIT IT KADS, suivi de KADS II, met en place une méthodologie qui couvre le cycle de vie d'un système à base de connaissance [Schreiber et al. 99]. Cette méthodologie permet entre autres de construire un modèle d'expertise à base de tâches. Comme le souligne [Le Roux 94], KADS fait figure de standard européen dans la modélisation du raisonnement et d'autre part présente une flexibilité exceptionnelle, c'est-à-dire la possibilité d'utiliser ce même modèle pour modéliser divers savoir d'experts. Par contre, KADS n'offre pas de modèle d'implémentation de ce modèle conceptuel. [Le Roux 94] cite à ce propos une étude comparative : « Alors que KADS est la seule approche procurant un langage général de modélisation, c'est aussi la seule dont les modèles ne soient pas opérationnels. Ce sont là deux faces d'une même médaille. [...] Il est évident que KADS est beaucoup plus flexible que les autres approches, cela étant dû à son total manque de sémantique opérationnelle. »

Le modèle d'expertise KADS est catégorisé en trois parties qui communiquent :

- 1- Les *Tâches* sont des problèmes généraux pour lesquels on possède des méthodes de résolution. Une méthode de résolution d'une tâche peut indiquer une décomposition d'une tâche complexe en sous-tâches ou, si c'est une tâche primitive, invoquer un élément de la couche des inférences.
- 2- Les *Inférences* sont les opérateurs permettant de manipuler les objets du problème, ce sont les briques élémentaires du raisonnement.
- 3- Le *Domaine* décrit la connaissance spécifique au domaine : c'est un peu l'équivalent du modèle objet d'un système, enrichi des instances, c'est-à-dire la structure de la base de connaissances factuelles et la base de connaissances même.

Une tâche est décrite comme un but à atteindre. Il n'y a pas de façon générique universelle de dire comment s'énonce un but. Certains buts s'énoncent uniquement à l'aide du vocabulaire proposé dans le domaine. D'autres buts intègrent dans leur énoncé ce que Chandrasekaran appelle des « termes d'attitude » : souhaité, à éviter, expliquer. A chaque tâche correspondent des méthodes pour la réaliser, c'est-à-dire des successions d'actions à mener pour atteindre l'objectif. Il peut y avoir plusieurs méthodes pour atteindre un même objectif.

Quand une tâche comporte plusieurs méthodes de résolution, une stratégie lui est associée qui désigne comment choisir une méthode en fonction du contexte de réalisation de la tâche.

Les entrées et sorties des tâches et des inférences sont spécifiées fonctionnellement, sous la forme de *rôles*. Les rôles sont des variables qui prennent leurs valeurs dans le domaine.

Notons que le terme « domaine » peut prêter à confusion car l'expression « connaissances d'un domaine » désigne couramment l'ensemble des connaissances des experts d'un domaine. Dans KADS, le *domaine* n'est qu'une partie des « connaissances d'un domaine ». Les *tâches*, *inférences* et *rôles* font aussi partie de ces connaissances.

En définitive, le modèle d'expertise KADS propose une structure pour le *COMMENT*, et un mécanisme pour rattacher ce *COMMENT* à un *QUOI* mais il ne propose pas de structure spécifique pour le *QUOI*. Ceci est logique car le but d'une structuration est de supporter un type de raisonnement, voire d'optimiser des manipulations : les connaissances *QUOI* doivent pouvoir être structurées diversement selon le raisonnement que l'on souhaite effectuer dessus. Cette structuration se fait par l'intermédiaire des rôles. A une tâche correspond un ensemble de rôles qui forment le vocabulaire de base de cette tâches. Lorsque la tâche est mise en œuvre, le fait de valuer les divers rôles avec les éléments du domaine correspond à structurer dynamiquement le domaine en vue du raisonnement mené dans la tâche. Cela est résumé sur la Figure 28.

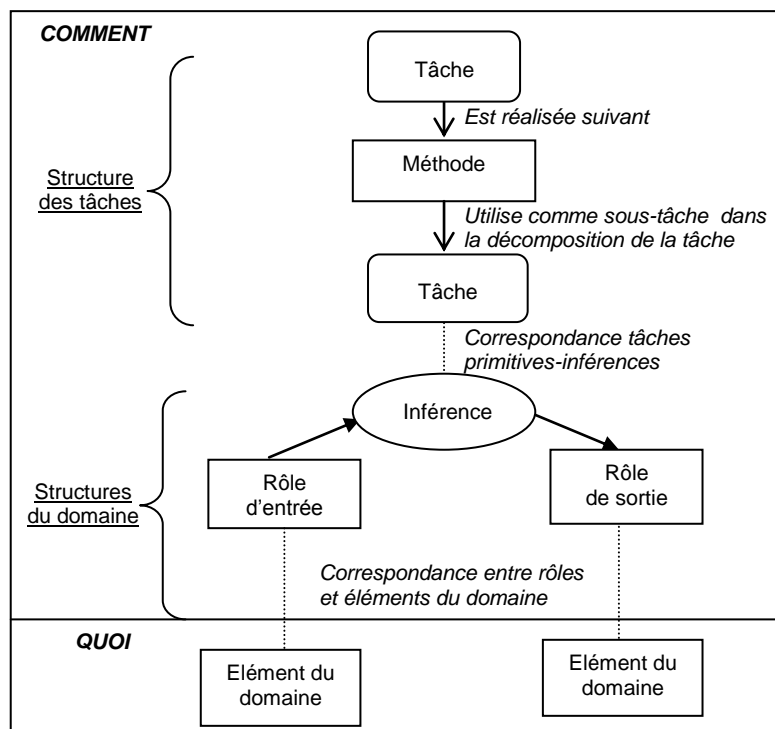


Figure 28 : Structuration des éléments du modèle d'expertise KADS. Le COMMENT est structuré par les méthodes c'est-à-dire la spécialisation des tâches et également leur décomposition en sous-tâches. Le QUOI peut être structuré différemment grâce aux rôles qui concrétisent les points de vue des tâches sur le domaine. [Figure 28 : Elements structure in a KADS expertise model. The HOW is structured by methods, i.e. the specialisation and decomposition of tasks into sub-tasks. The WHAT may be structured in different ways depending on roles making up for tasks dedicated views on the domain.]

Un modèle d'expertise KADS est généralement utilisé pour résoudre des problèmes particuliers. La résolution passe dans un premier temps par l'identification d'une tâche du modèle proche du problème. Ensuite la méthode de cette tâche du modèle est appliquée pour instancier, les concepts du domaine décrivant son objectif en tenant compte des spécificités du problème particulier. Dans l'exemple de l'affectation de bureaux, décrit Figure 29, la détermination de la tâche doit fournir les bonnes instances de liens entre employés et bureaux.

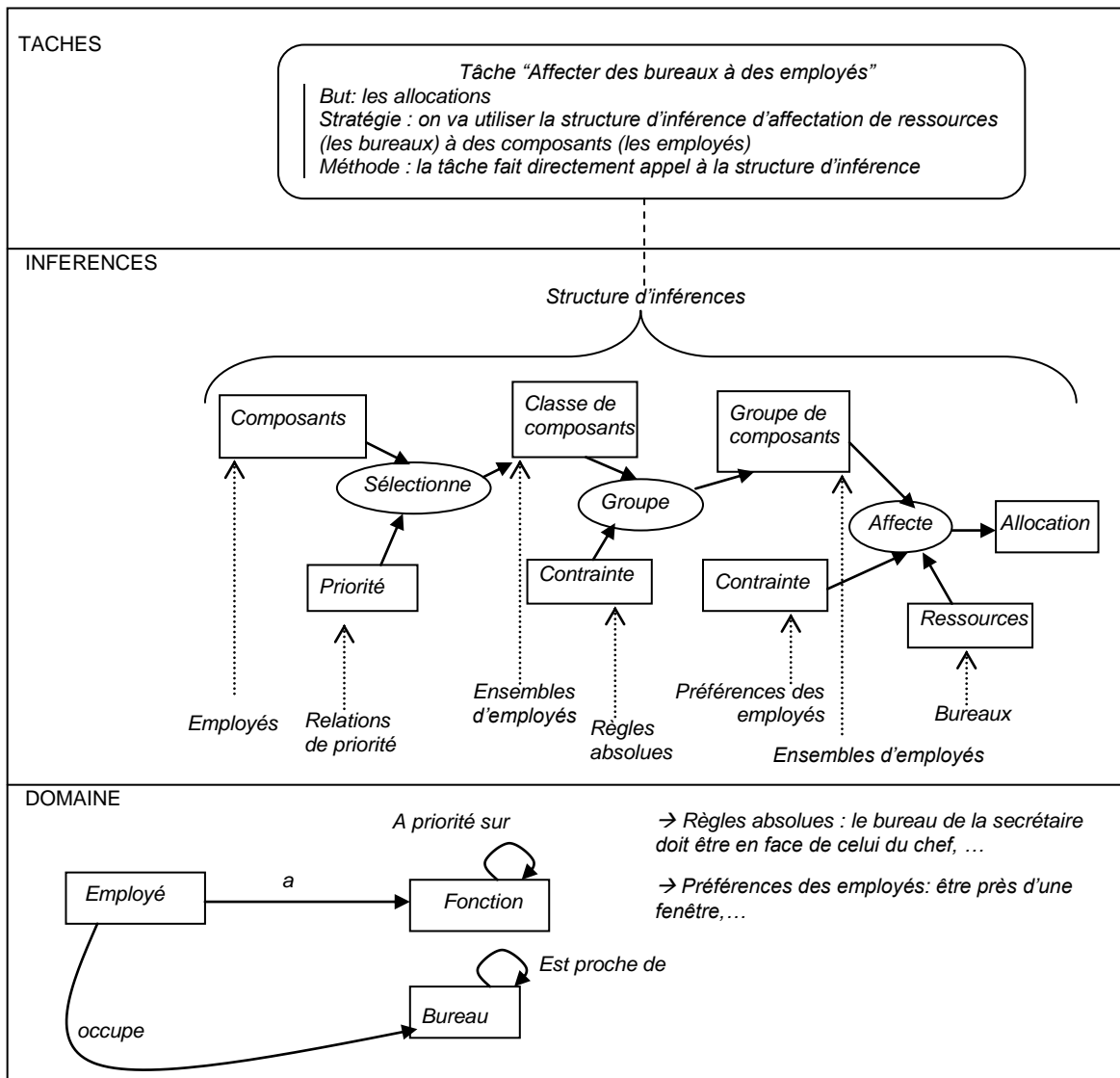


Figure 29 : Exemple d'utilisation du modèle KADS : la tâche d'affectation de bureaux. [Figure 29 : An example of usage of the KADS model to describe office rooms assignment.]

Les environnements de résolution coopérative de problèmes

Les modèles intégrant des connaissances de types *QUOI* et *COMMENT* formalisent des connaissances complexes et sont généralement délicats à utiliser. Dans une optique d'aide à l'accès, ils ne peuvent être directement exploités par les utilisateurs concernés; la construction de modèles décrivant les connaissances de types *QUOI* et *COMMENT* à propos de ressources (par exemple les modèles à base de tâches) est donc indissociable de la construction de systèmes particuliers qui servent d'interface entre les utilisateurs et de tels modèles.

De tels systèmes sont des environnements de résolution coopérative de problèmes. Un tel environnement permet de façon générale de résoudre un problème en permettant à l'utilisateur de participer à la résolution du problème. Dans le cas de l'aide à l'accès, le problème à résoudre est "comment répondre au besoin particulier d'information géographique de l'utilisateur" et sa résolution doit s'appuyer sur la description d'une tâche connue proche du problème de l'utilisateur.

L'exploitation du modèle se fait généralement en alternant des phases de planification et de résolution. La planification consiste à déterminer quelles actions effectuer pour résoudre le problème particulier de l'utilisateur et la résolution consiste à effectuer ces actions.

Un système exploitant un modèle de connaissances *QUOI* et *COMMENT* sur des ressources peut soit proposer une aide à la planification, c'est-à-dire aider un utilisateur à déterminer les actions qu'il doit effectuer, soit une aide globale à la résolution qui comprend les actions elles-mêmes. Dans un contexte d'aide à l'accès, l'aide à la planification consiste à aider l'utilisateur à comprendre comment les ressources peuvent répondre à son besoin. L'aide à la résolution consiste à aider l'utilisateur à répondre à son besoin avec les ressources.

L'architecture de tels systèmes est souvent une architecture emboîtée, comme décrit Figure 30. Cela signifie que le modèle *QUOI-COMMENT* que l'on veut exploiter grâce au système est intégré dans le *QUOI* du modèle *QUOI-COMMENT* du système.

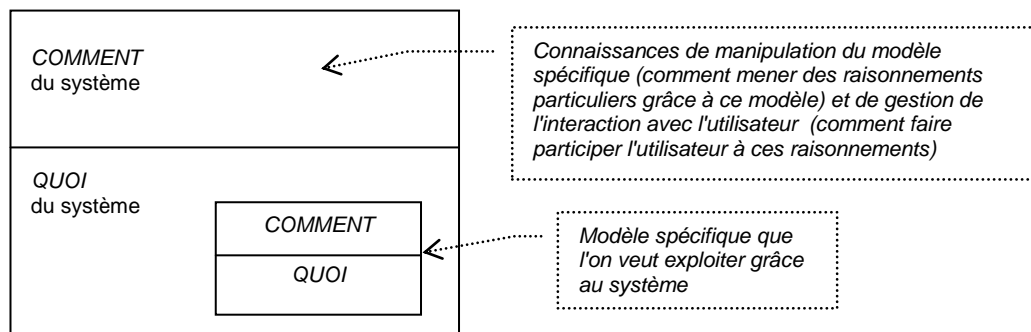


Figure 30 : Architecture d'un environnement de résolution coopérative de problèmes exploitant un modèle de connaissances *QUOI*-et-*COMMENT*. [Figure 30 : Architecture of a cooperative problem-solving environment using a *WHAT* and *HOW* knowledge model]

Le type de modèle utilisé dans le système LISA de [Delouis et Krivine 95] pour construire une telle architecture est un modèle à base de buts, méthodes et domaine. Les buts et méthodes sont utilisés pour décrire des connaissances de type *COMMENT*, spécifiques à un domaine d'intérêt donné ou spécifiques au système, et le domaine décrit les objets manipulés. Ce modèle permet d'associer à un but plusieurs méthodes pour l'atteindre. Dans ce système, l'utilisateur participe de deux façons : il peut participer à la résolution de son problème en donnant des valeurs à certains buts intermédiaires désignés par la méthode choisie, et il peut également participer à la planification en choisissant une méthode.

Le modèle SCARP proposé par [Willamowski 92] permet une coopération plus riche. L'utilisateur peut intervenir ponctuellement ou rediriger totalement le processus en cours.

C'est un modèle à base de tâches. Les tâches du domaine sont organisées en hiérarchies : hiérarchie des entrées/sorties et hiérarchie des méthodes. Cela permet une planification hiérarchique. La coopération peut alors se situer à divers niveaux d'abstraction. Le système possède des mécanismes sophistiqués de classification des entités manipulées dans le domaine qui lui permettent, en fonction de spécification des entrées et sorties par l'utilisateur d'aller classer le problème de l'utilisateur dans les tâches connues du système. Ces mécanismes sont en fait liés au langage SHIRKA utilisé pour représenter les connaissances *QUOI* du domaine d'intérêt.

De plus, l'auteur introduit des *tâches de visualisation* (dans le *COMMENT* du système), qui permettent au système de communiquer à l'utilisateur des connaissances relativement exhaustives sur le processus de planification et résolution en cours. Le système possède également des connaissances de contextualisation qui lui permettent de vérifier que les décisions de l'utilisateur sont applicables. Ces connaissances de contextualisation sont décrites au niveau de chaque tâche du domaine d'intérêt grâce à une sur-couche de SHIRKA proposée par [Willamowski 92].

DTSM (*Dynamic Selection of Tasks and Methods*) est un environnement de construction de systèmes à base de connaissances s'appuyant également sur les notions de tâches et de méthodes. Il supporte la construction d'un modèle dédié au problème de l'utilisateur et, au fur et à mesure de la modélisation, une opérationnalisation de ce modèle qui permet de l'affiner [Trichet et al. 00]. Les auteurs utilisent les graphes conceptuels pour représenter les tâches, les méthodes et le domaine.

Une tâche est représentée comme une règle produisant à partir d'un graphe conceptuel (GC) un autre GC. Le premier GC est appelé hypothèse de la règle, il représente le contexte d'activation de la tâche c'est-à-dire les conditions nécessaires pour pouvoir la réaliser. Les concepts de ce GC décrivent les entités manipulées dans le cas traité. Le GC conclusion de la règle représente les objectifs de la tâche.

Une méthode est également représentée comme une règle. Le GC hypothèse représente le contexte de sélection de la méthode. Les concepts de ce GC décrivent les ressources manipulées dans la méthode. Le GC conclusion représente les résultats produits par la méthode.

De façon similaire à [Willamowski 92], [Clouard et al. 02] proposent une aide au développement d'applications de traitement d'image, qui s'appuie sur un modèle de tâches décrivant, de façon générique, de telles applications à plusieurs niveaux d'abstractions, décrits Figure 31 :

- Le niveau *intentionnel* est le niveau du problème de l'utilisateur. Les tâches y sont décrites avec un vocabulaire intentionnel comme "isoler les objets du fond" ou "former les objets par fusion des régions". Les connaissances de détermination des tâches sont, à ce niveau, des connaissances stratégiques qui décomposent le problème en sous-problèmes.
- Le niveau *fonctionnel* est le niveau décrivant les méthodes du traitement d'image. Les tâches représentent des fonctionnalités du traitement d'image, comme la "détection de contour" ou la "classification des pixels".
- Le niveau *opérationnel* est le niveau des algorithmes. Il associe aux fonctionnalités décrites dans le niveau fonctionnel des algorithmes possibles comme "la classification de pixels basée sur une maximisation de la variance inter-classes".

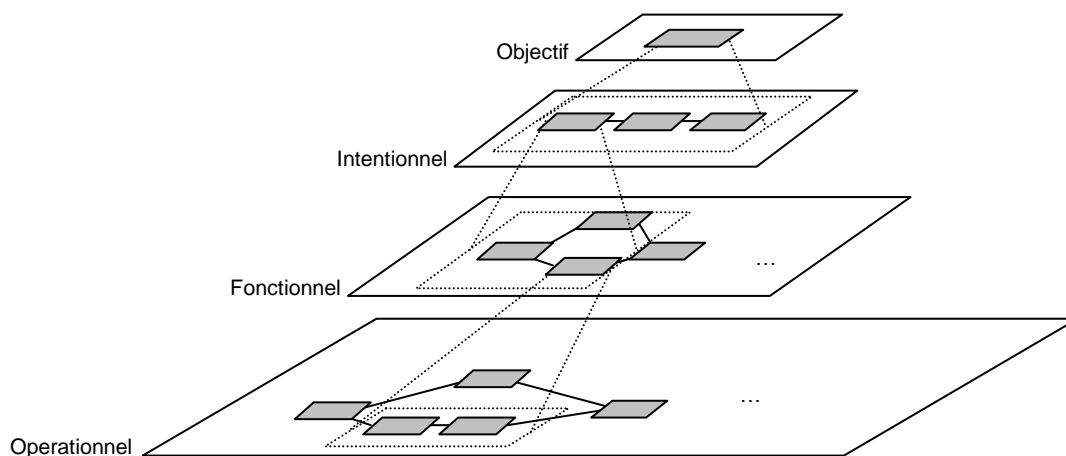


Figure 31 : Organisation en divers niveaux d'abstraction des tâches du modèle de [Clouard et al. 92]. Chaque niveau est l'expression d'une solution complète sous forme de graphe. [Figure 31 : The different levels in the tasks model of [Clouard et al. 92]. At each level, a complete solution is represented by a graph.]

Le projet Atelier Logiciel, mené dans le même cadre que les travaux de [Clouard et al. 02], est un exemple d'environnement de résolution coopérative de problèmes permettant à des utilisateurs, non experts en traitement d'image, d'accéder à l'exploitation de ressources en traitements d'image [Ficet 99]. Les ressources sont des bouts de codes de traitement d'image qui peuvent être paramétrés et assemblés pour constituer un programme complet de traitement d'image. Les utilisateurs des ressources, auxquels ce système est destiné, sont des experts de divers domaines (médical, industriel) travaillant sur des images. Le système utilise un modèle spécifique à base de tâches : le modèle Tâche-Méthodes-Outils. Les outils sont les bouts de code de traitement d'image. Les

tâches sont des tâches génériques de traitement d'image, qui sont associées à des méthodes les décomposant en sous-tâches et, ultimement, en outils. L'expression du besoin d'un utilisateur s'appuie sur la spécification d'une tâche parmi celles du modèle. Les tâches sont décrites avec un vocabulaire du domaine général de l'image, qui est supposé familier aux utilisateurs, comme « détecter les petits objets brillants sur l'image ». La tâche « extraire les objets d'une région » est par exemple décrite Figure 32.

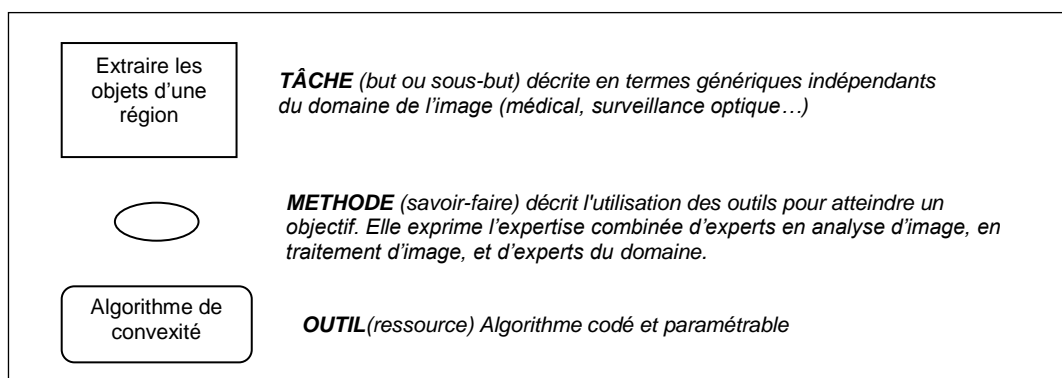
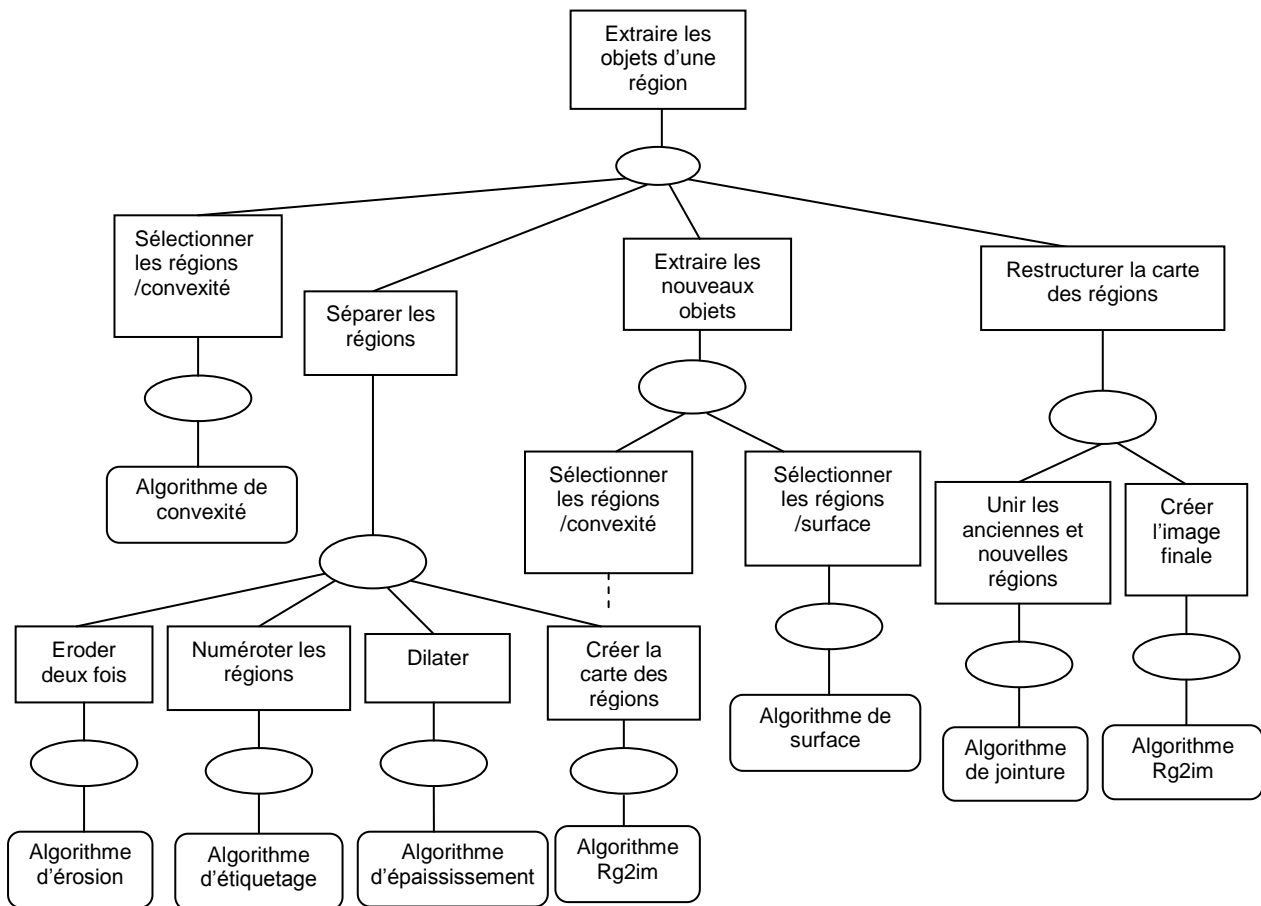


Figure 32 : Architecture Tâche-Méthode-Outil (schéma synthétisé d'après [Ficet 99]) [Figure 32 : Task-Method-Tool architecture (synthetic schema after [Ficet 99])]

Nous retenons que l'utilisation d'un modèle à base de tâches dans un contexte d'aide à l'accès nécessite de recourir à un environnement de résolution coopérative de problèmes exploitant ce modèle pour aider l'utilisateur à résoudre son problème d'accès.

De plus, cela nécessite de représenter plusieurs types de tâches. Le cheminement entre l'expression d'un besoin et l'exploitation de ressources repose en effet sur les réponses aux questions *pourquoi?*, *comment?*, et *avec quoi?*.

Certaines tâches –dites tâches intentionnelles- sont plus axées sur le *pourquoi*, c'est-à-dire l'expression du besoin.

D'autres tâches –dites tâches fonctionnelles- sont plus axées sur le *comment*. Elles entrent dans la décomposition de tâches intentionnelles.

D'autres tâches enfin sont dites tâches opérationnelles quand elles décrivent une étape élémentaire de résolution de problème, c'est-à-dire le *avec quoi*.

2.1.4 Bilan des techniques d'aide à l'accès et de leur intérêt dans notre contexte

Ce premier volet de l'état de l'art a étudié les techniques d'aide à l'accès et leurs contributions aux divers aspects de cette problématique générale. Il est important de situer notre cadre de travail parmi ces techniques. Certaines contributions à l'aide à l'accès à l'information géographique, bien qu'efficaces dans leur optique, ne nous intéresseront pas si elles ressortent d'une piste technique différente de la nôtre.

Les services d'aide à l'accès aux données s'appuient sur le regroupement de pointeurs sur des données, et sur l'utilisation de langages structurés pour stocker les données. Ils contribuent généralement efficacement à O2 (sélectionner l'information utile) et O3 (traduire une sélection d'information utile en une sélection de lot à acquérir) grâce à des opérations sur les indexes de données. Ils servent surtout à organiser des ressources contenant de nombreuses données et à automatiser des traitements automatiques sur ces données, qui ne sont pas des traitements de dérivation d'information mais de gestion de données (sélection, édition, modification...).

Par contre ces services ne contribuent pas au processus d'interprétation et de contextualisation de ces données. C'est pourquoi nous n'explorons pas davantage cette piste.

Les services d'aide à l'accès aux informations servent essentiellement à fédérer des ressources portant sur un même domaine d'information. Ces services reposent sur un langage permettant de décrire de façon homogène différentes ressources en termes de contenu, et éventuellement d'informations descriptives telles que le protocole de constitution. De tels langages se présentent souvent comme des standards de métadonnées.

Les ontologies de domaine sont utilisées dans des applications visant à diffuser des ressources sur le Web, en étant exploitées par des modules d'interprétation de requêtes. Cette aide est considérée dans notre état de l'art comme une aide à l'accès à des informations.

Dans le cas de l'information géographique, nous avons vu qu'il n'existe pas de description unifiée de cette information.

Catégorie d'aide :	Aide à l'accès aux données	Aide à l'accès aux informations	Aide à l'accès aux connaissances
Démarches appartenant à cette catégorie :	<ul style="list-style-type: none"> - Listes - Langages structurés avec opérateurs de sélection - SGBD 	<ul style="list-style-type: none"> - Entrepôts et fouille de données - Infrastructures d'informations, métadonnées, - Langages semi-structurés, et orienté-objet - Modèles et fonctions de dérivation d'information - Standards de métadonnées 	<ul style="list-style-type: none"> - Aide aux raisonnements (métaphores, visualisation, classifications,...) - Aide au raisonnement formel : modélisation d'ontologies et de connaissances de résolution de problème
Contribution à O1 (transfert de connaissances):	Désigne l'existence des ressources	Décrit le contenu des ressources	Décrit l'utilisation des ressources
Contribution à O2 (sélectionner l'information utile) :	Requête de données	Requête de métadonnées	Requête d'utilité attendue
Contribution à O3 (traduire une sélection d'information utile en une sélection de lot à acquérir):	Traduction d'une requête en portion exacte de ressource comportant les données voulues	Traduction d'une requête en ressource globale à acquérir, ou portion comportant entre autres les données voulues	Traduction d'une requête en ressource candidate
Contribution à O4 (fédérer des ressources):	Faible	Importante, surtout sur un même domaine	Possible sur des ressources portant sur des domaines variés
Contribution à O5 (s'adresser à divers utilisateurs):	Faible	Moyenne	Importante

Figure 33 : bilan des démarches contribuant à l'accès d'utilisateurs à des ressources en information. [Figure 33 : Sum up of approaches contributing to users' access to information resources.]

Les services d'aide à l'accès aux connaissances décrivent des ressources pour se rapprocher d'une représentation utilisée dans le raisonnement de l'utilisateur et l'expression de son besoin. Les divers raisonnements que l'utilisateur peut effectuer correspondent à autant de pistes techniques pour contribuer à son accès à des ressources.

L'aide au raisonnement formel est la plus riche dans notre contexte. Les services entrant dans cette catégorie contribuent à O1 de façon optimale car ils donnent accès à une manipulation des ressources. Ils contribuent également à O4 (fédérer des ressources) grâce aux travaux sur le partage et la réutilisation de connaissances. Par contre, ils ne sont valables que pour un certain type d'utilisateurs : ceux aptes à manipuler les structures symboliques proposées par le service. Par contre, les services fondés sur l'aide au raisonnement formel supposent que les utilisateurs sont aptes à manipuler les structures symboliques proposées pour son raisonnement. Les environnements de résolution coopérative de problèmes sont indissociables de ces modèles car ils servent d'interface entre l'utilisateur et le modèle.

Nous définissons notre domaine technique comme l'aide au raisonnement formel sur l'information géographique, étant données ses représentations actuelles.

Il est important de souligner que des démarches visant à aider d'autres raisonnements que le raisonnement formel sont entreprises avec succès dans le domaine de l'information géographique. Ainsi l'aide au raisonnement géométrique, fournie par les systèmes de visualisation cartographique de données, a une part importante dans l'accès d'utilisateurs. La visualisation d'une carte est un moyen d'accès aux connaissances géographiques traditionnel et familier à tous les utilisateurs. Bien plus : « Language is our primary means of communication, yet humanity is divided into between three and four speech communities, and in most cases it is impossible for members of two different speech communities to understand each other. [...] There are problems of communication even within speech communities [...] pictures present no such problems, and a picture of a lone protester in Tiananmen square is instantly meaningful to everyone.[...] Maps fall somewhere between these two

extremes. » (préface de [Vckovski 98] par Michael Goodchild)*. Les techniques de visualisation ont évolué et il est important d'exploiter les nouvelles possibilités offertes au dialogue entre un homme et un écran. Des recherches visent actuellement à favoriser le procédé de reconstruction de connaissances par l'utilisateur lors de sa lecture de la carte [MacEachren et al. 99b] [MacEachren and ICA 98]. [MacEachren et al. 99a] veulent intégrer la visualisation géographique avec les méthodes de découverte de connaissances dans les bases de données. [Andrienko and Andrienko 99] combinent de même les techniques de visualisation et d'analyse des données.

Par ailleurs, l'usage de langages visuels et de métaphores est un élément important dans la mise en place d'interfaces de dialogue d'utilisateurs avec des SIG. [Hubert 01a] propose un état de l'art de ces techniques. [Kuhn and Frank 91] étudient aussi l'application de métaphores spatiales pour décrire les outils SIG. Ils définissent des concepts de base invariants par la métaphore : le concept de base étant le « contenant », puis « centre/périphérie ».

Ces démarches sont complémentaires de la nôtre. Elles portent sur le langage de communication entre un système possédant un modèle de connaissances et un utilisateur possédant son propre modèle de connaissances, Figure 34. Or nous nous intéressons au modèle de connaissances du système, les concepts et la grammaire qu'il utilise. Le dialogue effectif entre l'utilisateur et un système d'accès à l'information géographique, fait l'objet d'une thèse au COGIT, menée par Frédéric Hubert depuis novembre 1999 [Hubert 01a][Hubert 01b]. Cette recherche est complémentaire de la nôtre. Elle s'attache à l'interface proprement dite entre le système de formalisation des connaissances que nous construisons et un utilisateur, et sur les interactions constituant le dialogue entre le système et l'utilisateur.

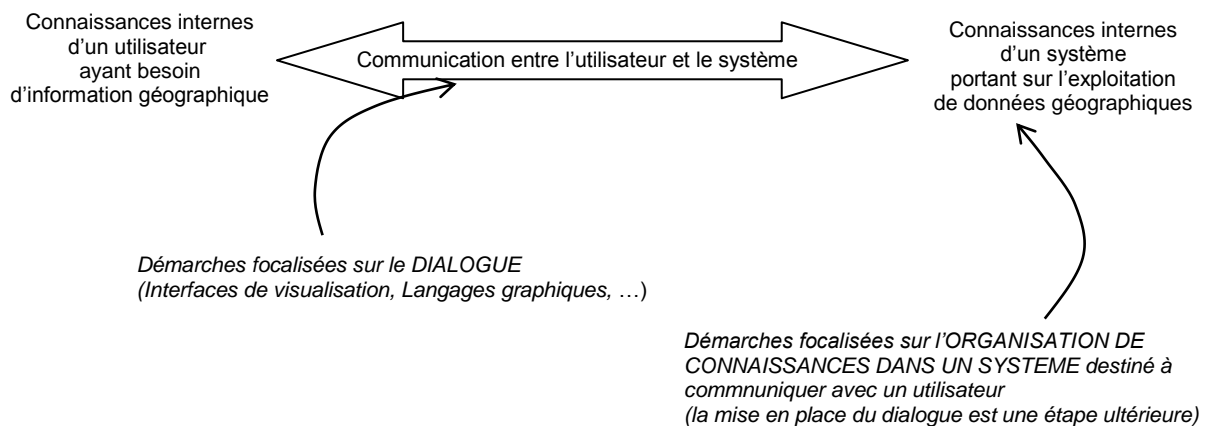


Figure 34 : Deux focalisations possibles pour l'aide à l'accès à des connaissances : le dialogue entre un système d'aide et l'utilisateur, et l'organisation de connaissances dans un système d'aide. [Figure 34 : Possible focuses in enhancing access to knowledge : the communication between an assisting system and a user, and the organisation of knowledge into a system.]

Concernant l'aide à l'accès, une formalisation efficace repose sur la distinction et l'intégration de deux catégories de connaissances, le *QUOI* et le *COMMENT* (*pourquoi, comment, avec quoi*). Une telle formalisation s'appuie par exemple sur la modélisation de tâches et de rôles.

Dans la section suivante, nous revenons à notre problématique initiale c'est-à-dire l'accès à l'information géographique mais nous ne faisons pas un état de l'art des démarches d'aide à l'accès à l'information géographique. Cette section 2.2 est en fait un premier pas dans notre

* Le langage est notre premier mode de communication. L'humanité est pourtant divisée entre environ quatre communautés de langage et il est impossible à des membres de deux communautés différentes de se comprendre entre eux. [...] Il y a même des problèmes de communication au sein de ces communautés. [...] Les images ne posent pas de tels problèmes et une photo d'un manifestant isolé sur la place de Tiananmen est immédiatement comprise par chacun. [...] Les cartes se situent quelquepart entre ces deux extrêmes.(tda)

approche : nous avons choisi une approche technique et nous évaluons son applicabilité. L'approche technique choisie est la construction d'un modèle respectant le principe énoncé dans le paragraphe précédent et l'exploitation de ce modèle par un environnement de résolution coopérative de problèmes.

Dans la section suivante, nous analysons l'existence d'une formalisation distincte et intégrée des connaissances de type *QUOI* et *COMMENT* (*pourquoi, comment, avec quoi*) portant sur des données géographiques.

2.2 Etat de l'art des formalisations des QUOI et COMMENT géographiques

2.2.1 Généralités

Dans cette section, nous cherchons à évaluer l'existence d'une formalisation distincte et intégrée des connaissances de type *QUOI* et *COMMENT* (*pourquoi, comment, avec quoi*) portant sur des données géographiques.

Une première section 2.2.2 détaille les démarches visant à améliorer les formalisations utilisées pour construire des BDG, c'est-à-dire le *QUOI* géographique informatique. Ces démarches abordent le problème de l'accès, à la base, en voulant améliorer les ressources mêmes.

Nous examinons ensuite les formalisations axées sur le *COMMENT*, c'est-à-dire la manipulation de données géographiques. De façon générale, une formalisation est axée sur la description d'une ou l'autre de ces catégories. Mais une difficulté du principe énoncé ci-dessus est d'opérer cette distinction entre le *QUOI* et le *COMMENT*. Aussi les démarches sont généralement axées sur l'une de ces catégories, mais comportent une part de l'autre catégorie.

La première question "pourquoi manipuler des données géographiques?" renvoie à l'énoncé d'objectifs d'application SIG. Les formalisations intéressantes dans ce contexte sont étudiées section 2.2.3. Cette section nous amène à étudier les formalisations du *QUOI* géographique des hommes qui sont utiles pour l'expression de besoins d'utilisateurs, comme représenté Figure 35.

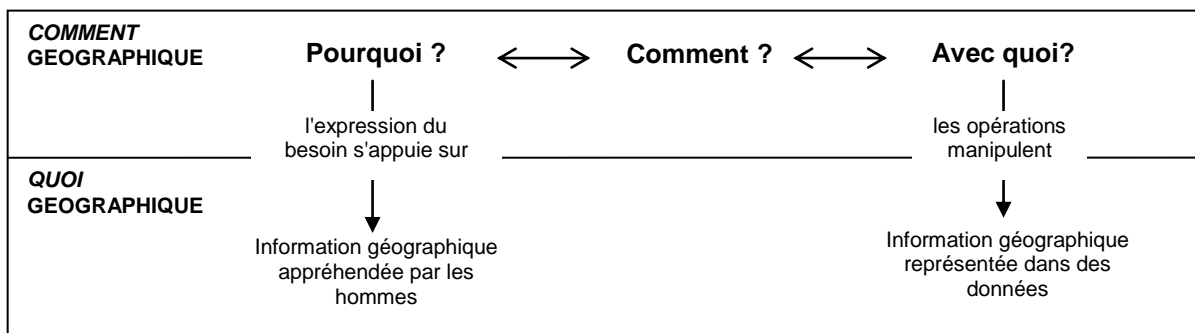


Figure 35 : Liens entre les catégories de connaissances du *COMMENT* et celles du *QUOI*, dans le cadre de l'exploitation de données géographiques. [Figure 35 : Links between the different categories of *WHAT* and *HOW* knowledge in the context of the exploitation of geographic data.]

La question "comment manipuler des données géographiques?" renvoie aux méthodes des applications SIG. Les formalisations correspondantes sont étudiées dans la section 2.2.4.

La question "avec quoi?" renvoie aux outils informatiques manipulant les données géographiques. Les formalisations correspondantes sont étudiées dans la section 2.2.5.

Ces distinctions au sein du *COMMENT* ne correspondent pas à une catégorisation tranchée des démarches. C'est typiquement le cas des démarches en interopérabilité qui, après s'être attachées à la communication entre systèmes –avec quoi-, se préoccupent actuellement de l'exploitation par des utilisateurs des modules interopérables –comment-. Nous essayons dans la mesure du possible de ne pas présenter plusieurs fois la même démarche mais plutôt de l'intégrer dans la section dont elle nous semble le plus ressortir. Pour cela, nous considérons globalement que la réponse au *comment* implique de formaliser la manipulation des outils pour que les hommes les comprennent, et la réponse au *avec quoi* implique d'informatiser la manipulation des connaissances.

2.2.2 Le QUOI

Cette section examine quelles représentation de l'information géographique peuvent être intégrées dans des systèmes informatiques. Les grands modèles de données géographiques actuels, et les métadonnées ont été présentés dans la section 1.1.1. Nous présentons ici les recherches qui veulent améliorer le lien entre les données et le monde réel soit en intervenant au niveau du processus de représentation de ce réel, soit en manipulant les représentations disponibles.

| Ces démarches contribuent à l'accès en s'attachant à améliorer les ressources elles-mêmes.

2.2.2.1 La représentation générale de l'information géographique

Comme nous l'avons vu dans la première section de ce mémoire, 1.1.1, les modèles des BDG sont très éloignés des modèles construits par l'homme pour ses raisonnements. Il n'existe pas de BDG construite en s'appuyant sur des concepts géographiques de sens commun. Les BDG ne stockent ainsi explicitement presque aucune relations spatiales entre objets, hormis parfois la topologie. Or ces relations sont une notion primordiale dans l'appréhension que se fait l'homme de l'espace géographique. [Corona and Winter 01] évaluent cet écart dans le cas particulier d'un modèle cognitif de navigation, celui de [Timpf et al. 92], et les données géographiques exploitables pour faire du calcul d'itinéraire. Ils mettent en évidence le besoin d'intégrer davantage des concepts utilisateurs dans les BDG. Les travaux de [Hangouët 98] soulignent aussi l'écart actuel entre l'information géographique informatisée et l'information géographique de sens commun. Pour réduire cet écart, l'auteur propose d'explicitier la notion de phénomène géographique, qui est « l'ensemble lié de deux parties :

- 1- l'objet réel homologue de données qui sont dans la base de données,
- 2- le plongement de cet homologue dans une explicabilité géographique. » L'explicitabilité géographique signifie pour [Hangouët 98] : « qui a trait au tracé, à la marque sur le terrain, à l'apparence (état) ou à l'apparition (processus de formation) dans le paysage pour la vue qu'on veut en faire. »

Rappelons que l'information géographique actuellement contenue dans les BDG et SIG est statique, 2D, voire 2D ½ et discontinue. Dans les années 90, de nombreux travaux ont cherché à améliorer cette représentation de l'information géographique, en dehors de tout contexte applicatif.

Certaines visent à intégrer le spatio-temporel dans les modèles de données et de métadonnées [Claramunt et al. 94], [Spéry et al. 99], [Gervais 00]. Par exemple, [Chomicki and Revesz 99] proposent une nouvelle structure pour stocker des informations sur un phénomène géographique dans une base de données. Cette structure est l'objet spatio-temporel. Elle est composée d'un objet géographique, d'un objet temporel et d'une transformation continue donnant l'image de l'objet géographique à tout temps appartenant à l'objet temporel.

D'autres se concentrent sur la représentation du 3D, c'est le cas de [De la Losa 00].

D'autres visent à intégrer la notion de flou de représentation c'est-à-dire à gérer les notions d'incertitude, de résolution, et de discrétisation [Goodchild and Gopal 98], [Goodchild and Jeansoulin 98], [Faiz 96].

Certaines approches exploitent la théorie des algèbres pour obtenir une meilleure représentation de la notion de localisation dans les BDG, comme la « ROSE algebra » de [Guting and Schneider 95], l'algèbre des localisations vagues de [Bittner 99].

Des propositions sont également avancées pour une représentation de la nature continue de l'information géographique [Laurini and Gordillo 00]. Elles associent à une représentation discrétisée d'un phénomène les mécanismes d'estimation des valeurs qui ne sont habituellement pas explicitées dans la représentation.

D'autres travaux reviennent en amont du processus de représentation de l'espace géographique et proposent d'intégrer explicitement dans la représentation informatique de l'espace géographique des éléments de l'appréhension cognitive de l'espace géographique par l'homme. Les auteurs de [Mennis et al. 00] rappellent les principes fondamentaux de telles démarches :

- décomposer l'information géographique selon trois dimensions : le temps, l'espace, le thème. Ces dimensions correspondent à notre représentation cognitive décomposée en « quand ? où ? quoi ? ».
- représenter divers niveaux de l'information géographique : les données observées et les connaissances dérivées. Cette organisation est celle de notre cognition : nos sens perçoivent de l'information, et nous dérivons des schémas qui vont guider notre cognition, cf. Figure 36.

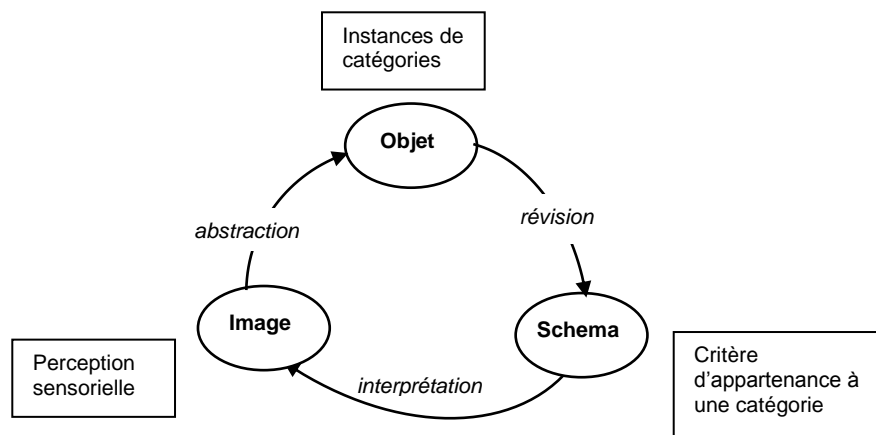


Figure 36 : Le cycle de l'acquisition et du raffinement cognitifs de connaissances géographiques (repris de [Mennis et al. 00].) [Figure 36 : Life cycle of geographic knowledge cognitive acquisition and refinement (after [Mennis et al. 00])]

Des travaux sont également réalisés pour fédérer des représentations portant sur des territoires contigus ou sur un même territoire.

Les travaux entrepris dans le cadre de projet européens comme COMMUTER (Esprit Project 22362, Co-operative Multidimensional Multimedia and Topological Elements Reintegration) visent à créer une base européenne à partir de bases nationales.

D'autres travaux proposent des mécanismes de dérivation et communication entre des représentations différentes d'un même territoire. Ils reposent sur des techniques d'appariement de représentations. L'appariement est défini comme « une technique consistant à établir une correspondance entre des objets géographiques symbolisant la même partie du monde réel mais provenant de jeux de données géographiques différents » [Lemarié 96].

L'appariement met en place des mécanismes pour reconnaître que deux représentations désignent le même objet. Ces mécanismes s'appuient sur des caractérisations particulières des objets représentés : un identifiant attribué par l'homme (numéro d'une route...), les localisations de l'objet dans les représentations, les formes de l'objet dans les représentations, ses relations avec d'autres objets dans les représentations. Il existe également des démarches sur l'appariement sémantique qui comparent non pas les formes géométriques mais les attributs [Rodriguez and Egenhofer 99]. Dans sa thèse, [Devoegele et al. 98] propose un processus d'intégration et d'appariement de BDG qui conduit à une BDG intégrée multi-représentations centralisée. Un point important est l'extension du langage de correspondance par exemple en créant des clauses d'appariement géométrique des données permettant de stocker les liens produits par le processus d'appariement entre les représentations initiales.

Plus récemment, de nombreux travaux visent à concevoir des modèles de bases multi-représentations pour rendre compte des différentes résolutions significatives de l'information géographique utilisées dans les BDG actuelles. Par exemple le projet européen Murmur (MultiRésolution MultiReprésentation) du programme IST a pour objectif la définition un modèle de données décrivant au niveau conceptuel des données spatio-temporelles sous plusieurs représentations, assurant la cohérence de ces représentations multiples [Spaccapietra 99][Motet 99]. Murmur veut également mettre en place des interfaces utilisateurs pour la définition et la manipulation de ces bases spatio-temporelles.

2.2.2.2 Des représentations dédiées à des contextes applicatifs

D'autres travaux enrichissent le *QUOI* géographique informatisé non pas en se concentrant sur la nature des objets géographiques, c'est-à-dire dans le cadre du *QUOI*, mais plutôt en se situant dans le cadre d'un *COMMENT* spécifique.

Les modèles proposés dans des contextes applicatifs spécifiques ont pour objectif de dériver de l'information utile des données ou de fédérer les diverses représentations des données utilisées dans cette communauté. Autrement dit, ces modèles prennent place dans un contexte de dérivation d'informations ou d'intégration de données. Nous détaillons l'exemple de la navigation.

La communauté du transport regroupe trois catégories : les gestionnaires du réseau, les utilisateurs militaires, les utilisateurs civils. Cette communauté utilise des systèmes de référence linéaire pour positionner des « objets » (par exemple des panneaux de signalisation), et des « événements » (par exemple un embouteillage) le long d'« objets appartenant à un réseau de navigation » organisés en réseau (par exemple un tronçon d'autoroute). Ce positionnement consiste à indiquer une distance depuis un sommet du réseau (un carrefour) en direction d'un autre sommet. En 1996 un workshop a mis en place un élément clé pour le partage d'informations entre ces catégories : un système de référence linéaire unifié [Fletcher et al. 98]. Les auteurs rappellent la définition d'un « unified linear datum » :

« ...the collection of objects which serve as the basis for locating the linear referencing system in the real world. The datum relates the database representation to the real world and provides the domain for transformations among linear referencing system and among geographic representations. The datum consist of a connected set of anchor sections that have anchor points at their junctions and termini. No application attributes are assigned to the datum. » *

Le « Datum » répond à un double objectif. L'un est d'intégrer dans une représentation unique (les sections d'appuis et points d'appui) les différentes représentations spatiales utilisées dans les diverses bases de données de cette communauté. L'autre est de proposer une représentation exploitable par les différentes méthodes de référencement linéaire utilisées dans les applications transport. D'autres travaux proposent des modèles plus

* ...l'ensemble des objets sur lesquels est fondée la localisation du système de référencement linéaire dans le monde réel. Le datum fait le lien entre une représentation de BD et le monde réel et il fournit au domaine d'application des mécanismes de transformation entre le système de référence linéaire et les représentations géographiques. Le datum consiste en un ensemble de sections d'ancrage reliées, de points d'ancrage aux jonctions de ces sections et à leurs extrémités. Il n'y a pas d'attribut applicatif dans le datum.(tda)

précis pour décrire des « événements » dans un système de référence linéaire, en distinguant des événements ponctuels, comme un feu qui passe au rouge, des événements linéaires, comme un ralentissement de circulation le long d'un tronçon, et des événements surfaciques, comme une zone urbaine [Dueker and Butler 98].

Des représentations dédiées à d'autres contextes sont également proposées comme celles visant à automatiser la généralisation cartographique. Elles visent à faciliter le processus de changement de niveau de granularité de données géographiques. On peut se reporter à [Ruas 99] pour un panorama de ces travaux. L'auteur propose également un modèle de généralisation. Dans ce modèle, les entités devant être cartographiées sont représentées par des objets qui possèdent chacun :

- des variables décrivant son état dans le cours du processus de généralisation,
- des contraintes sur ces variables (par exemple une contrainte de taille),
- des méthodes pour mesurer le niveau de satisfaction des contraintes,
- des méthodes pour corriger l'état en cours en fonction de la contrainte qui doit être satisfaite.

Cette formalisation fait le lien entre la description d'un contenu et une manipulation spécifique de ce contenu, la généralisation cartographique automatique. Elle définit ce qui est considéré comme des *rôles* dans notre approche.

Améliorer les ressources en information géographique pour qu'elles soient plus accessibles à des utilisateurs est une vaste entreprise. De nombreuses démarches y participent, en se focalisant sur la réalité représentée dans ces données ou en se situant dans un contexte de focalisation par le *COMMENT*. Dans ce dernier cas, les formalisations désignent plus des rôles qu'un contenu.

2.2.3 Le Pourquoi

Cette section détaille les démarches contribuant à formaliser des besoins en information géographique, c'est-à-dire des objectifs conduisant à construire des applications SIG.

Nous détaillons tout d'abord, dans la section 2.2.3.1, ce qui fait la complexité de telles démarches, c'est-à-dire la difficulté de formaliser l'information géographique perçue par les hommes.

Nous décrivons ensuite les représentations formelles du *QUOI* géographique des hommes. Elles prennent essentiellement place dans deux contextes :

- La formalisation de connaissances partagées par tous les hommes dans leur expérience courante, décrite section 2.2.3.2,
- Les modèles construits par des communautés d'utilisateurs dédiés à des activités spécifiques, décrits section 2.2.3.3.

Une dernière section, 2.2.3.4, décrit une démarche isolée qui formalise des résultats d'application SIG. Cette section contribue à la formalisation du *pourquoi* dans la mesure où elle formalise des objectifs liés à l'exploitation de données géographiques.

2.2.3.1 Les interactions de l'homme et du monde

Dans cette section, nous détaillons les interactions de l'homme avec le monde qui permettent d'expliquer ce que les hommes peuvent connaître en information géographique, ce qui est constitutif de leur subjectivité. Parmi la littérature, nous choisissons la classification de [Frank 01] non parce qu'elle est la plus juste mais parce qu'elle nous semble la plus facile à comprendre. L'auteur organise les représentations de l'information géographique selon le niveau d'interaction de l'homme avec le monde auquel elles se situent :

- *La réalité indépendante de l'homme.* Dans cette réalité, il n'y a pas d'objets, mais selon Frank, des dimensions continues (l'espace et le temps) et des champs $f(x,y,z,t)$ qui ne sont pas toujours continus. Cette réalité est soumise aux lois de la physique.
- *L'observation du monde physique.* Cette réalité est l'observation de propriétés de points de l'espace, depuis un système localisé, à un moment donné. L'espace et le temps ne sont pas perçus de façon équivalente : « maintenant » a le même sens pour tous les agents, et « ici » non. A ce stade, interviennent les notions de résolution et d'erreur d'observation. Ces erreurs ne sont pas subjectives. Ces observations peuvent être stockées dans des données.
- *Les mondes d'objets.* Frank considère que l'homme crée des objets pour maintenir des propriétés uniformes invariantes dans le temps. Cette construction est liée aux interactions possibles de l'homme avec ces objets.
- *La réalité sociale.* Le quatrième niveau de l'interaction de l'homme avec le monde est celui des conventions, comme par exemple les langages, et les objets institutionnels comme les billets de banque.
- *Les mondes d'agents cognitifs.* Le dernier niveau de l'interaction de l'homme avec le monde implique les croyances de cet homme, ses méthodes d'observation, sa communication avec d'autres agents. Nous rattachons à ce niveau les interactions liées à une activité spécifique de l'homme. Nous avons vu que des objets sont construits dans l'expérience courante de l'homme, qui forment des mondes d'objets. A ce niveau des objets sont également construits mais dans le contexte d'expertises spécifiques. C'est à ce niveau qu'intervient une interaction de l'homme avec le monde autre que percevoir le monde, l'observer et appuyer ses actions sur cet espace : mesurer les conséquences de ses actions dans l'espace géographique. Les mondes virtuels appartiennent également à ce type de réalité.

Il nous semble important de compléter cette typologie en explicitant des liens particuliers entre ces niveaux d'interaction de l'homme avec le monde. Ces liens particuliers sont les interactions mêmes entre ces niveaux. Par exemple, l'observation considérée par Frank semble être restreinte à l'activité de mesure, mais elle est composée également de l'activité de définition d'un protocole de mesure. La définition de ce protocole de mesure fait appel aux formalisations créées au niveau des mondes d'objets. Il s'agit en fait d'un cycle : l'observation du monde physique prend en considération les objets créés pour décrire ce monde. Ce cycle est celui de la cognition : les stimuli sont perçus par nos sens et sont organisés par notre cognition en fonction de schémas préétablis, cf. Figure 36 page 91.

Par ailleurs, la constitution d'objets renvoie au concept clé anglais d'*affordance* introduit par Gibson, et beaucoup repris depuis, pour décrire l'environnement: "I have described environment [...]. But I have also described what the environment affords animals, mentioning the terrain, shelters, water, fire, objects, tools, other animals, and human displays. How do we go from surfaces, is there information for the perception of what they afford. If so, to perceive them is to perceive what they afford" [Gibson 79]*. Ce terme d'*affordance* peut être traduit par "contribution à une capacité", ou encore "rôle dans certaines interactions". Les affordances sont construites pour désigner des objets lors d'interactions économiquement intéressantes pour l'homme qui crée ces objets. Cette notion a été reprise dans de nombreux travaux visant à formaliser la représentation de l'espace géographique par l'homme, comme ceux de [Jordan et al. 98] détaillées plus loin. La constitution d'objets décrivant le monde réel est ainsi extrêmement liée aux interactions de l'homme avec le monde physique.

* J'ai décrit l'environnement[...] Mais j'ai aussi décrit ce que l'environnement offre aux animaux, ce qui inclut le terrain, les abris, l'eau, le feu, les objets, les outils, les autres animaux et les représentations de l'homme. Comment allons-nous d'une surface à une autre ? Y a-t-il une information pour traduire la perception de ce qu'elles offrent ? Si oui, percevoir une surface revient à percevoir ce qu'elle offre. (tda)

Enfin, la réalité sociale existe en fonction des interactions entre individus de la société. Elle existe donc parce qu'il y a un monde d'agents cognitifs.

Ces interactions entre les divers contextes et activités de représentation sont essentielles. En particulier elles impliquent la propagation de facteurs d'une représentation à l'autre : l'intervention de la subjectivité humaine, l'intervention de conventions et les erreurs de mesure. Ces facteurs ne sont pas explicités alors qu'ils sont essentiels pour comprendre une représentation. Ces interactions expliquent aussi pourquoi il existe une variété de représentations renvoyant à des subjectivité d'individus ou de groupes.

Nous essayons de donner une idée de la complexité de ces interactions en en schématisant quelques unes sur la Figure 37.

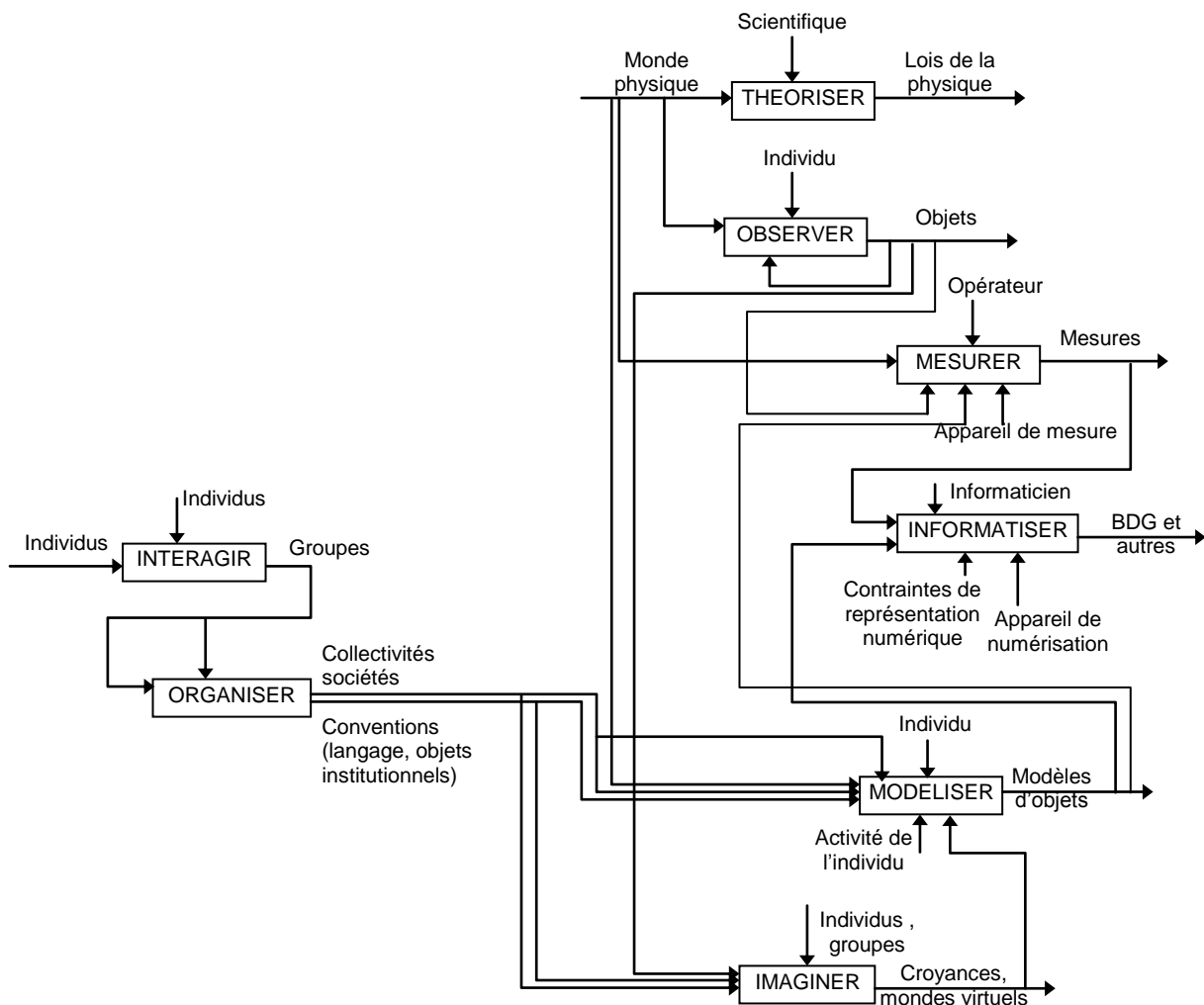


Figure 37 : Proposition de schéma de type SADT montrant les interactions entre les activités d'appréhension et de formalisation de l'espace géographique par les hommes. Ce schéma n'est pas exhaustif. Son objectif n'est pas de poser une théorie mais de faire comprendre la complexité de ces interactions. [Figure 37 : Proposed SADT schema showing interactions between apprehension and formalisation processes of geographic space by men. This is not an exhaustive schema. It does not aim at grounding a theory but rather at highlighting the complexity of such interactions.]

2.2.3.2 Les représentations de l'espace géographique partagées par tous

Certaines démarches cherchent à identifier des corpus de connaissances partagées par tous dans le domaine de l'information géographique.

La démarche du *National Center for Geographic Information and Analysis* (NCGIA) sur les modèles formels des mondes géographiques de sens commun veut expliciter une « géographie naïve » [Egenhofer and Mark 95] [Mark et al. 97]. La méthode mise en place consiste à s'appuyer sur les affirmations que nous pouvons énoncer couramment comportant des termes géographiques. C'est une démarche ambitieuse et il est difficile de prévoir quand elle s'achèvera. Elle couvre en effet un domaine très large, ce qui nuit à sa formalisation.

D'autres recherches se restreignent à des domaines limités -ce que l'on cherche à décrire- ou à des contextes limités -ce qui nécessite la description-.

Un domaine limité est par exemple la désignation de localisations. [Denand 00] analyse et structure les expressions utilisées dans le langage naturel pour répondre à la question « où ? ». Ces expressions se présentent sous la forme « préposition locative + groupe nominal ». Le groupe nominal est une entité, le « site » de la localisation. La préposition locative, comme « à », « au Nord de », désigne une relation spatiale entre l'entité à localiser, la « cible », et le « site ». L'auteur souligne que la difficulté majeure lors de la construction d'une réponse à la question « où ? » est de choisir le site le mieux adapté à la cible que l'on veut localiser. Par exemple la réponse « dans le salon » à la question « où est le stylo ? » est moins pertinente que la réponse « sur la table du salon ». Par contre si la question porte sur une personne, la réponse se limite généralement à la pièce où se trouve la personne.

[Gonçalves 99] étudie également les modes de désignation de lieux par l'homme. L'auteur veut créer une interface de dialogue entre un utilisateur et un système possédant un jeu de données géographiques. Le dialogue qu'elle cherche à mettre en place est le suivant : l'utilisateur désigne un lieu, en pratique un secteur d'une ville, et le système affiche une carte de ce lieu. La difficulté de cette approche consiste à formaliser la désignation d'un lieu par un utilisateur. Comme [Denand 00] l'auteur rappelle que la désignation d'un endroit repose sur le choix d'un référentiel, et que ce choix dépend de l'entité que l'on cherche à localiser. [Gonçalves 99] distingue par ailleurs trois éléments importants de qualification d'un lieu : l'orientation (ex : « à gauche de.. »), la distance (ex : « loin de »), la topologie (ex : « inclus dans »). Il s'agit des types de relations spatiales qui peuvent être utilisées pour construire une référence par rapport à une entité site. L'auteur propose un état de l'art extrêmement détaillé des formalismes, qualitatifs, rendant compte des connaissances cognitives d'orientation, de distance et de topologie de l'homme.

Les travaux de [Jordan et al. 98] portent sur la modélisation de la localisation dans l'expérience courante de l'homme. Ils insistent sur le fait qu'une localisation ne se réduit pas à un ensemble de relations de l'objet localisé avec d'autres objets. La localisation doit également intégrer son sens dans le contexte d'une action de l'homme, c'est à dire les *affordances* qu'elle propose à l'homme. Par exemple, un restaurant est un lieu où l'homme peut se nourrir, entretenir des relations sociales, etc. Les auteurs proposent de distinguer six facettes dans la modélisation d'une localisation :

- les objets physiques de la localisation (à diverses résolutions),
- les activités que l'homme a dans cette localisation,
- les anecdotes associées à cette localisation,
- les représentations symboliques de cette localisation (les auteurs citent l'exemple de « the Big Apple » comme symbole de la ville de New York),
- les facteurs socio-économiques et culturels,
- la catégorie de cette localisation (par exemple : une grande ville, un restaurant huppé, un coin perdu).

Les auteurs ajoutent que l'effort de modélisation nécessaire pour représenter les *affordances* est énorme.

Cette démarche de [Jordan et al. 98] est très intéressante car elle veut intégrer des notions proches des rôles dans la modélisation d'un domaine. Elle amorce le lien entre le *QUOI* représenté et un *COMMENT* s'appuyant sur ce *QUOI*. Par contre, elle manque de formalisation.

D'autres démarches étudient également les relations spatiales perçues par l'homme. Ainsi [Mathet 00] analyse dans le langage les relations entre les objets comme se croiser, longer, contourner.

La représentation cartographique fournit une description globale de l'espace géographique, synthétisant les relations spatiales pertinentes pour les lecteurs. Le langage cartographique reposant sur des conventions de communication relativement universelles [Bertin 67]. Ce langage est accessible en lecture à un grand nombre de personnes. [Cian et al. 89] et [Brun et al. 87] soulignent ainsi l'abondance des connaissances de sens commun concernant la lecture de cartes. Par contre ce langage n'est pas accessible à tous en écriture ; le dessin d'une bonne carte requiert une certaine expertise.

La notion de « carte cognitive » est introduite par [Kuiper 78] pour décrire la représentation que l'homme construit de l'espace qui l'entoure au fur et à mesure qu'il l'explore. De même qu'une carte traditionnelle, cette représentation renvoie non pas à quelques relations spatiales mais plutôt à une description englobant toutes les relations spatiales pertinentes pour la personne construisant la représentation, dans le contexte de son déplacement.

Faisant suite à la « carte cognitive », des travaux se sont attachés à décrire un modèle cognitif de navigation. La navigation est une activité courante connue de tous les hommes. Elle est étudiée dans l'optique de concevoir des systèmes informatiques proposant une aide à la navigation adaptée aux besoins des utilisateurs. La formalisation de connaissances de navigation est également intéressante car le fait de dédier la représentation à une activité permet d'obtenir une représentation relativement structurée, et, parallèlement à cela, le fait d'étudier une activité extrêmement courante permet d'obtenir un certain consensus.

[Timpf et al. 92] distinguent ainsi trois types de tâches dans l'ensemble des tâches participant à l'activité de prévoir un voyage et de suivre un itinéraire : les tâches de planification, les tâches de guidage, les tâches de conduite. Les auteurs associent, à chaque groupe de tâches, un modèle d'objets dédiés à ces tâches.

Les objets des tâches de planification sont : les autoroutes décrites par des nœuds et des arêtes, les lieux par lesquels passent l'autoroute, et les intersections qui sont les places par lesquelles passent plus d'une autoroute.

Les objets des tâches de guidage sont : les autoroutes décrites comme des objets bi-directionnels comportant un ensemble d'entrées et de sorties, les entrées d'autoroute (lieux auxquels un véhicule peut entrer sur l'autoroute), les sorties d'autoroute, les jonctions (lieux auxquels un véhicule peut passer d'une sortie d'autoroute à une entrée sur une autre autoroute) et les échangeurs (agrégats de jonctions).

Les objets des tâches de conduite se réduisent à une autoroute, celle sur laquelle se trouve le véhicule, décrite par : la voie courante, la voie de dépassement, les bretelles d'entrées, les bretelles de sortie, la bande d'arrêt d'urgence.

Ce modèle est exploité par les auteurs pour implémenter un prototype d'aide à la navigation plus proche des concepts des utilisateurs.

[Claramunt et Mainguenaud 96] distinguent également plusieurs catégories de connaissances de navigation : les connaissances procédurales, les connaissances topographiques, et l'espace perçu. Les connaissances procédurales renvoient aux tâches de guidage de [Timpf et al. 92], les connaissances topographiques renvoient aux tâches de planification, et l'espace perçu aux connaissances de conduite. Les auteurs insistent sur la nécessité de distinguer plusieurs niveaux de résolution dans la représentation de connaissances de navigation. Ils distinguent ainsi des vues spatiales à divers niveaux de résolution. Par exemple une vue peut correspondre à la navigation d'un aéroport à une ville et une autre vue à la navigation au sein de la ville pour aller à une adresse particulière. Les auteurs introduisent le « collage spatial », un mécanisme de passage d'une vue spatiale à une autre qui s'appuie sur des « connexions spatiales ». Le collage spatial, concrétisé par la connexion, associe à une localisation dans un système pertinent dans la première vue spatiale une localisation dans un système de

référence pertinent dans la deuxième vue spatiale. La connexion peut être par exemple une sortie d'autoroute. Dans une première vue spatiale, cette sortie a une référence spatiale dans le système de référence linéaire des autoroutes. Dans une deuxième vue spatiale, cette même sortie a une référence dans le système du réseau routier local.

Les formalisations présentées ci-dessus se trouvent finalement à la limite de la séparation de notre état de l'art entre « le *QUOI* des hommes » et « le *COMMENT* des hommes » :
Elles décrivent des représentations construites par l'homme de l'espace qui l'entoure.
Elles explicitent les actions de l'homme, dans son activité de navigation, et adaptent la représentation de données géographiques aux besoins de ces actions.

Pour conclure cette section, nous remarquons qu'il existe des démarches proposant des formalisations consensuelles. Plus une démarche se limite à une activité spécifique du sujet qui doit raisonner avec cette formalisation, plus elle produit des résultats formels. Les résultats de ces démarches sont extrêmement intéressants, eu égard à la désignation d'objets géographiques. Par contre ils ne suffisent pas à construire un modèle d'aide à l'accès car ils n'intègrent pas l'expression d'un besoin de connaissances dérivées des données.

2.2.3.3 Les représentations de l'espace géographique dédiées à des communautés d'utilisateurs

L'espace géographique peut également faire l'objet de représentations formelles dans le cadre d'activités spécifiques. Ces contextes sont ce qu'on appelle des « communautés d'utilisateurs ». Le cadre de telles communautés est nettement plus restreint que celui de l'ensemble d'une société et est davantage propice à l'élaboration d'une formalisation d'une réalité géographique perçue. Le consortium OpenGIS veut identifier les communautés d'utilisateurs de SIG et expliciter les connaissances de chaque communauté [Open GIS TC 98]. Une communauté est caractérisée par un « project world », c'est-à-dire les concepts utilisés dans la communauté pour décrire le monde réel, et par les métadonnées qui doivent être attachées à des BDG pour leur exploitation par cette communauté.

Dans le contexte d'une application SIG, il peut être nécessaire de fédérer les points de vue de diverses communautés devant participer à un même projet. Par exemple, [Major 99] étudie le cas d'une application de gestion territoriale à laquelle participent diverses communautés (aménagement, urbanisme, transport, patrimoine). Pour l'auteur, une conception d'application SIG territoriale passe par la représentation d'« objets pivots » permettant de passer d'un modèle d'une communauté à un autre, comme le « bâtiment » et également par la représentation d'« objets intermédiaires », termes généraux utilisés dans les discours sans être attribués à des acteurs particuliers.

En France, un organisme a été créé en 1985 pour centraliser les besoins en information géographique : le Conseil National de l'Information Géographique (CNIG) qui « contribue par ses études, avis, ou propositions, à promouvoir le développement de l'information géographique et à améliorer les techniques correspondantes, en tenant compte des besoins exprimés par les utilisateurs publics ou privés. » En particuliers, le CNIG a adopté une nomenclature d'échange associée à la norme française de transfert de données géographiques EDIGéo (Echange de données informatisées dans le domaine de l'information géographique). Cette nomenclature est mise au point par la Commission EDIGéo du Conseil National de l'Information Géographique et elle est actualisée par les utilisateurs dont la DGI-cadastre, l'IGN, le SHOM (Service Hydrographique et Océanographique de la Marine française), l'Ordre des Géomètres Experts, etc. Au sein de l'IGN, un projet en cours vise entre autres à identifier

les diverses nomenclatures nécessaires pour permettre à un utilisateur d'exprimer son besoin en des termes métiers [Cébelieu 01].

Les représentations consensuelles construites dans le cadre de communautés d'utilisateurs peuvent être exploitées dans des modules d'interprétation de requêtes d'utilisateur. Par exemple, dans le cadre d'un module de découverte de ressources, [Bishr 97] s'appuie sur une ontologie du domaine routier pour construire un traducteur sémantique des requêtes.

Nous constatons ici comme dans la section précédente, que les formalisations se limitent à la représentation de domaine d'application, voire à la fédération de telles représentations. L'application de ces travaux permet l'expression de besoins de données mais non l'expression de besoins de dérivation de connaissances.

2.2.3.4 La formalisation de résultats d'applications

La démarche de [Mitchell 99] contribue à l'expression de besoins de dérivation de connaissances. Plus exactement, ses travaux portent sur la description des produits qui peuvent être construits à l'aide de données géographiques. Ils se situent dans un contexte restreint qui est la production de cartes avec le SIG ArcInfo. La démarche de l'auteur est de se poser les questions suivantes : que veut-on connaître avec la carte ? comment faire la carte pour cela ?

Il identifie six tâches génériques d'analyse spatiale qui se décomposent en sous-tâches réalisables à l'aide des fonctionnalités de ce SIG. Ces six tâches de haut niveau sont :

- Cartographier la localisation d'un phénomène (pour analyser des patterns)
- Cartographier des localisations (homogènes) de quantités (pour comparer des places ou comprendre des relations)
- Cartographier la densité (pour repérer les places ayant une certaine concentration) entre places
- Cartographier des activités ou des phénomènes sous la forme d'aires délimitées (pour suivre une activité ou comparer des places)
- Cartographier ce qui est à une distance ou à un coût de parcours d'un objet
- Cartographier le changement

Cette démarche, encore relativement isolée, est extrêmement intéressante dans notre contexte parce qu'elle explicite un résultat produit. Par contre elle est limitée car elle ne considère pas la diversité des applications géographiques comme décrire un territoire, cartographier un risque, construire un équipement d'aménagement du territoire. De plus, elle n'est pas exploitable par des utilisateurs qui ne savent exprimer leur besoin qu'à un niveau abstrait et non en termes de produits dérivés. Les tâches décrites sont des tâches fonctionnelles et non intentionnelles. Enfin, le résultat est informel et donc ne peut être exploitée par un système.

2.2.4 Le Comment

Nous ne donnons pas ici un état de l'art exhaustif du vaste domaine de la conception de projets SIG. Nous cherchons plutôt à dégager des points clés relativement à la manipulation de connaissances. Dans la conception d'une application SIG, il existe deux axes extrêmement importants :

- la modélisation des données et le choix de la représentation,
- la structuration du travail en processus effectués par des composants logiciels.

2.2.4.1 La modélisation

La phase de représentation est cruciale dans un projet SIG, plus exactement la représentation de la situation dans l'espace telle qu'elle est manipulée par l'utilisateur. [Lopes et al. 97] considèrent que, du point de vue des utilisateurs, cela consiste à choisir entre une représentation par champ et une représentation par objets. Plusieurs facteurs influent sur le choix d'une représentation.

Un facteur essentiel est l'activité de l'utilisateur. Les représentations d'un même domaine peuvent varier beaucoup en fonction de l'utilisation escomptée des données. [Berry 93] insiste sur « *appropriate is best* » : il n'existe pas de meilleure structure de données géographiques, si une structure convient alors c'est la meilleure. Pour illustrer cela prenons l'exemple de la représentation du relief. Si l'objectif est de faire une carte des pentes, la représentation privilégiée sera du maillé, c'est-à-dire un modèle numérique de terrain. Dans une telle représentation, une altitude est associée à chaque portion élémentaire de surface de la terre. Si l'objectif est de calculer des pentes, par exemple pour déterminer un tracé de route, la représentation privilégiée sera le vecteur, c'est-à-dire des courbes de niveau. Cela dit, au sein d'une communauté, il est important de choisir une représentation intégrative en particulier pour la notion de localisation.

Le choix d'un référentiel spatial pour le projet, c'est-à-dire le système de référence associé éventuellement à des objets de référence, doit tenir compte de la diversité éventuelle des sources de données. Ces données devant être rattachées dans le projet au même référentiel.

Comme le rappellent les auteurs de [Laurini et Milleret-Raffort 93], le coût de l'acquisition des données est bien plus important que celui des logiciels. Cela incite à consacrer beaucoup de soins au choix de la représentation et de l'acquisition des données (réutilisation de données géographiques disponibles ou numérisation de nouvelles données).

La représentation peut également varier, moins radicalement, en fonction des opérations prévues sur les données. Cette dernière variation ne concerne pas uniquement la partie spatiale du modèle. Par exemple, dans une optique de cartographie, il peut être souhaitable de regrouper les classes thématiques proches en une classe générique pourvue d'un attribut qui permet de distinguer les classes initiales. Cela facilite la formation de catégories à cartographier. Un autre exemple est la représentation des routes. Si des opérations de calcul d'itinéraire sont prévues il sera utile de représenter la topologie du réseau. Si il n'y a pas d'autres opérations que la cartographie ce sera inutile.

Par ailleurs, comme le souligne [CNIG 97], le choix du SIG peut conduire à simplifier la structure de données géographiques utilisées dans le projet. En effet, les modèles de données proposés par les producteurs de données comportent parfois des structures complexes qui ne peuvent pas toujours être correctement traduites dans de nombreux SIG du commerce. Les producteurs fournissent alors des structures simplifiées, dites « descendues », de leurs données.

Il existe ainsi des principes guidant le choix d'un modèle de données lors de la réalisation d'une application SIG. Ces principes ne sont ni véritablement organisés, ni suffisamment spécifiques pour suffire au choix d'une représentation. Ils désignent des critères importants, sans expliciter comment les prendre en compte.

2.2.4.2 La structuration de l'application

Il existe une structuration classique d'une application SIG revenant fréquemment dans la littérature [Maguire and Dangermond 91][Giordano et al. 94]. Le tableau Figure 38 présente une synthèse des étapes généralement proposées et des fonctionnalités SIG correspondantes.

Etape	Fonctionnalités
Chargement des données	Saisir Transférer Editer
Intégration des données	Structurer, Restructurer, Apparier, Corriger
Analyse	Requêter Généraliser, Analyser
Présentation du résultat	Cartographier Création de rapport

Figure 38 : Etapes à passer en revue lors de l'analyse fonctionnelle interne et fonctionnalités correspondantes qui doivent être remplies par les composants SIG. [Figure 38 : Internal functional analysis stages in the conception of a GIS project.]

Cette structuration est extrêmement vague, elle peut être exploitée dans la conception très globale d'une application mais non dans la conception des manipulations à effectuer sur des données pour dériver une réponse à un besoin.

La démarche d'Ormeling présente plus d'intérêt dans notre contexte car il s'attache davantage à décrire la manipulation de données [Ormeling 98]. L'auteur propose une typologie de tâches géographiques très générale et transversale aux domaines d'application, et leur associe des outils nécessaires, Figure 39.

Classe de Tâches	Expertise utilisée	Domaine associé
Tâche du 1 ^{er} ordre : Rattache un phénomène à la surface de la terre	Décrire / identifier	Localisation spatiale (site, classes de situations) Distance Distribution spatiale (groupe)
Tâche du 2 ^{ème} ordre : Décrit les relations entre les objets formant le phénomène	Relier	Association spatiale (verticale ou horizontale) Interaction spatiale Système spatial (région)
Tâche du 3 ^{ème} ordre : Rattache divers phénomènes entre eux ou à une troisième dimension	Interpréter	Structures spatiales

Figure 39 : Classification très générale de tâches géographiques de [Ormeling 98]. [Figure 39 : Very general geographic tasks classification proposed by [Ormeling 98].]

Cette classification est faite non pas en fonction du type d'objectif poursuivi mais plutôt en fonction de la manipulation générale d'objets géographiques et de lieux. Elle ne s'adresse par aux utilisateurs finaux mais plutôt aux chercheurs en analyse spatiale.

Une démarche plus proche de l'utilisateur est celle de [Mitchell 99]. Il associe aux tâches cartographiques énumérées section 2.2.3.4 leurs décompositions en fonctionnalités importantes et les caractéristiques du domaine qui vont influencer dans le choix des fonctionnalités décomposant une tâche. Il n'utilise pas de formalisme particulier, nous traduisons son discours dans un formalisme de tâches du type KADS, Figure 40.

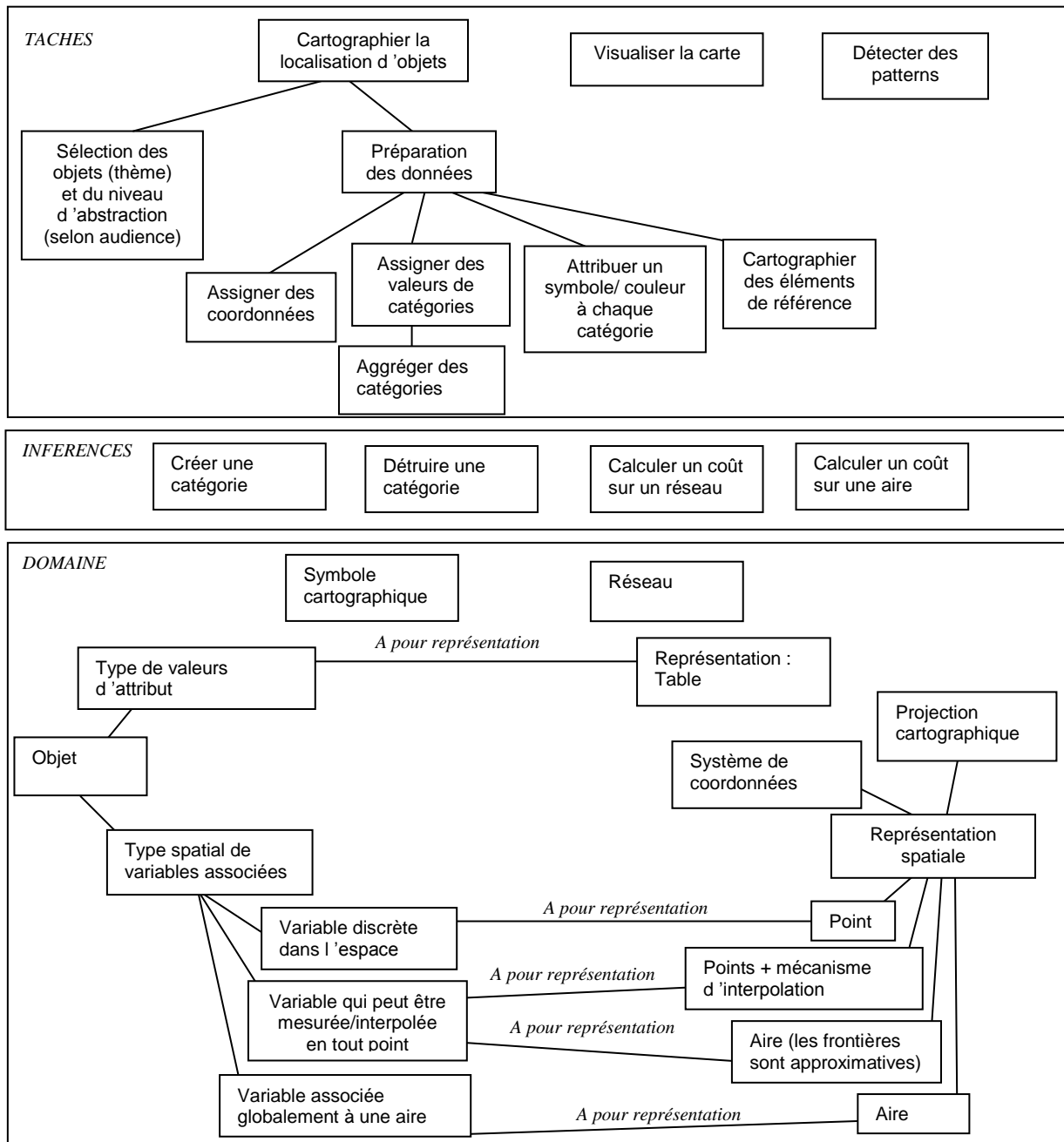


Figure 40 : Description d'une tâche cartographique, analysée par [Mitchell 99], et traduite dans un formalisme tâche-inférence-domaine. Les types de valeurs d'attributs, non représentés sur la Figure, sont les suivants : catégorie, rang, montant, ratio. [Figure 40 : Description of a mapping task, analysed by [Mitchell 99] and translated in a task-inference-domain formalism. The attributes value can be of different types (not represented on the schema : category, rank, amount, ratio.)]

Cette démarche est extrêmement intéressante car elle identifie diverses catégories de connaissances : les objectifs, les outils, le choix des outils en fonction des objectifs, les objets manipulés. Par contre elle n'est pas suffisamment formalisée pour être exploitée autrement que dans un manuel d'emploi, ce qui est d'ailleurs l'objectif de l'auteur.

Concernant la structuration d'une application SIG, la plupart des démarches se contentent de décrire de façon générique les étapes d'un projet SIG mais ne détaillent pas la dérivation d'un résultat à partir de données. Il existe peu de connaissances de spécification de telles descriptions,

hormis des indications informelles de facteurs à prendre en compte. La démarche la plus intéressante est celle de [Mitchell 99] qui explicitent les diverses catégories de connaissances. Mais ces connaissances sont disponibles sous forme de documents textuels et ne peuvent être opérationnalisées dans un système d'aide.

2.2.5 Le Avec Quoi

Les démarches de formalisation de la manipulation de données géographiques par des systèmes visent à optimiser des requêtes sur des données géographiques ou à automatiser des traitements, par exemple d'analyse spatiale. Les principaux traitements SIG ont été présentés en section 1.1.2.3. Nous montrons ici des démarches appartenant davantage au domaine de la recherche.

D'autres démarches se focalisent sur le choix de mécanismes formels. Elles se fondent sur la théorie des algèbres, ou sur les approches fonctionnelles. Les algèbres se concentrent sur la notion d'opérateur et les propriétés d'opérateurs appliqués alors que les approches fonctionnelles se concentrent sur le typage de variables et la construction de types abstraits. D'autres démarches se focalisent sur des contextes d'exploitation, comme le traitement de l'incertitude ou l'analyse spatiale.

2.2.5.1 La création d'outils

Parmi les démarches participant à la formalisation du *Avec Quoi*, nous comptons les démarches qui s'attachent à améliorer les outils de manipulation de données géographiques. Elles contribuent à l'accès, de la même façon que les travaux visant à améliorer les ressources mêmes.

Le data mining spatial

L'évolution des SIG et BDG vers une situation où les données spatiales sont de plus en plus intégrées aux données non spatiales, [Zeitouni et Yeh 00] et [Bédard et al. 00], permet d'envisager d'appliquer des mécanismes de *data mining* sur ces bases intégrées. Cela nécessite de représenter les relations spatiales, grâce à des graphes de voisinage, ou des matrices [Zeitouni et Yeh 00].

Notons qu'avant les travaux spécifiques du *data mining spatial*, des auteurs ont souligné le besoin d'un nouveau type d'analyse spatiale. [Openshaw 91] fait ainsi un bilan des outils existant focalisé sur l'analyse spatiale et ses méthodes. Il critique l'emploi de méthodes de type inférentiel qui vont à l'encontre de la tradition « exploratoire » de l'analyse spatiale. Il propose une approche où la description est un but en soi : *exploratory geographical analysis* et énumère les outils à appliquer aux diverses variables géométriques pour mener cette analyse. Nous reprenons Figure 41 ses propositions en les organisant en domaine/outil.

Outil	Plus proche voisin Méthodes quadratiques	Analyse de réseaux Méthodes « theoretic » Dimension fractale Détection d'arêtes	Mesures de forme Autocorrélation spatiale Régression spatiale Régionalisation Interaction spatiale Modélisation de location-allocation	Traitement d'image Cartographie bayésienne
Domaine	Point	Ligne	Aire	Surface

Figure 41 : Classification de [Openshaw 91] et sa proposition d'outils génériques à fournir pour la description de patterns par des méthodes d'analyse exploratoire. [Figure 41 : [Openshaw 91] classification and proposition for generic tools needed to describe patterns in spatial exploratory analysis.]

Les travaux du *data mining spatial* permettent d'enrichir les ressources. Les travaux de l'analyse spatiale exploratoire associent à un type de manipulation -l'analyse spatiale- les outils et variables nécessaires pour cette analyse.

La "map algebra"

Une approche originale est celle de Tomlin et Berry qui utilisent une algèbre pour formaliser une méthodologie générale pour analyser et synthétiser des données géographiques [Tomlin 91]. La formalisation est censée être valable sur données raster et vecteur, mais est décrite sur la structure raster. Ce formalisme, détaillé également dans [Berry 87], a pour variable la *map*. Les données manipulées sont organisées en couches, ou image à deux dimensions, les *maps*. Toute opération consiste à créer une nouvelle couche à partir d'ancienne(s). Les traitements raster disponibles sont décrits comme autant d'opérateurs appliqués à la variable *map*, et structurés en algèbre. Une couche est décrite par des voisinages chez [Berry 87], et par des formes, des distances et des directions chez [Tomlin 91]. [Tomlin 91] classe les opérations (de transformation de données en information) en fonction de leur effet sur les locations et [Berry 87] en fonction de leur signature, cf. Figure 42.

Une limite importante de ces travaux est l'interprétation de la *map* produite par l'application d'opérations, qui n'est pas aussi simple que la lecture d'une carte classique. En fait, l'application des opérateurs ne correspond pas à des étapes de construction d'une cartographie mais plutôt à des étapes de l'analyse spatiale.

Cette approche est à la limite du *avec quoi* et du *comment* mais la véritable formalisation porte surtout sur le *avec quoi* : les opérations sur les *map*.

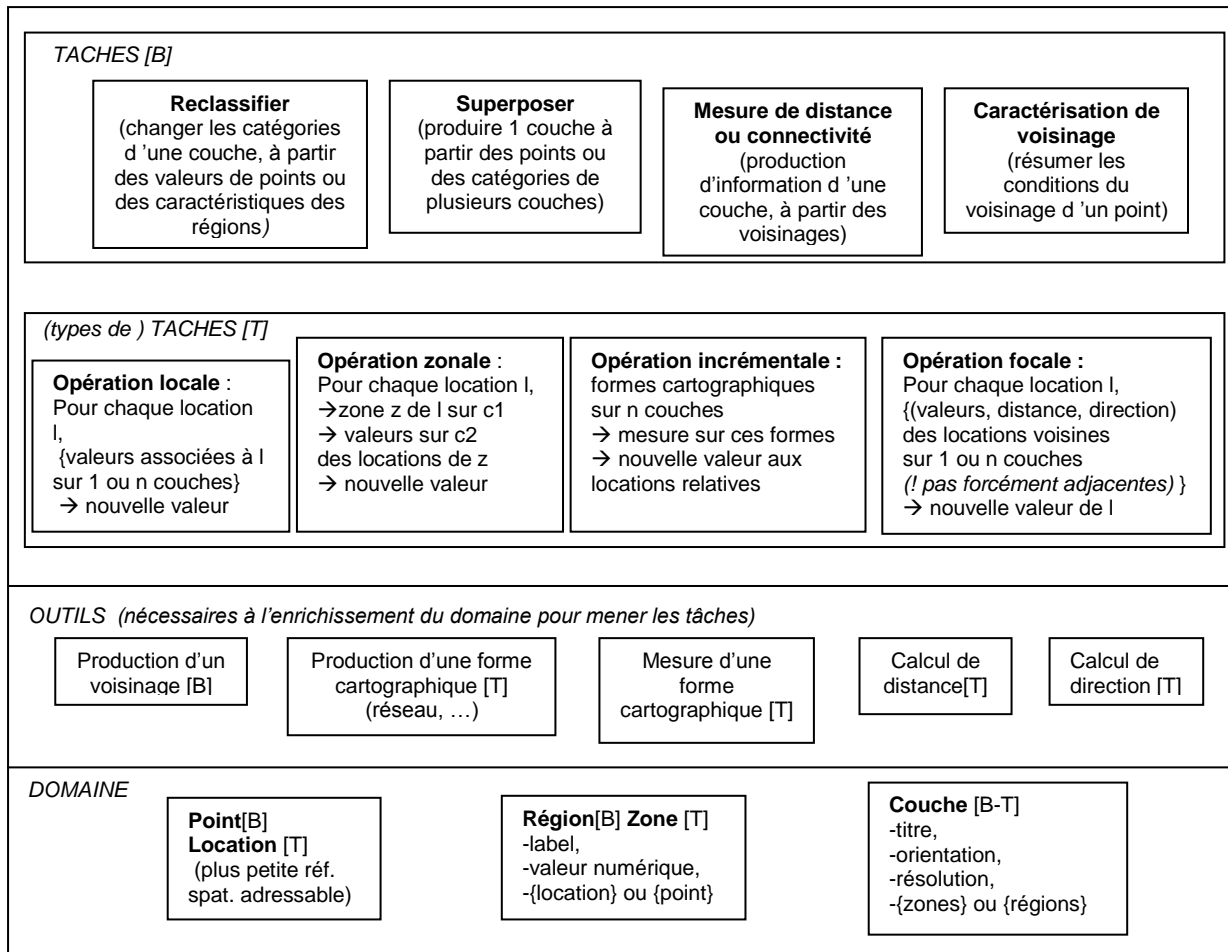


Figure 42 : Bilan des connaissances de la Map Algebra ([B] : éléments introduits par Berry, [T] : éléments introduits par Tomlin, [B-T] : éléments qui semblent être les mêmes chez Berry et Tomlin), transcrites ici dans un formalisme tâches-inférences-domaine. [Figure 42 : Synthesis of elements in the Map Algebra ([B] : elements introduced by Berry ; [T] : elements introduced by Tomlin ; [B-T] : elements that seem to be similar for both Berry and Tomlin) translated in a formalism task-inference-domain.]

Les approches fonctionnelles

D'autres approches se concentrent plus sur les notions de structures et de changement de structures.

[Chrisman 99] appelle ces changements les transformations. Il explique l'intérêt de cette notion en cartographie. Le point de départ de la cartographie est de fait une transformation : la projection. Le paradigme transformationnel, valable pour les projections et la généralisation, est, selon l'auteur, extrapolable à l'ensemble de la cartographie. Il existe diverses transformations cartographiques comme faire un buffer autour d'une route ou convertir des courbes de niveaux en TIN (triangulation). L'auteur souligne qu'une classification ne doit pas s'attacher uniquement aux mécanismes ou à la syntaxe des opérations mais aussi à la sémantique. Pour classifier une transformation de façon aisément interprétable, Chrisman distingue deux composants d'une transformation :

- la manipulation de l'espace pour créer un voisinage,
- la manipulation des attributs par une règle de combinaison. Chrisman regroupe ces règles en 3 catégories : les règles de dominance qui sélectionnent une valeur sur certains critères, les règles de contribution qui composent un résultat final à partir de toutes les valeurs, les règles d'interaction qui utilisent non seulement les valeurs mais des comparaisons par paires de ces valeurs.

Parmi les approches s'attachant à la description et la manipulation de structures, une démarche de grande ampleur est HBDS, *Hypergraph Based Data Structure*, regroupant un modèle et une méthode de modélisation. HBDS tient une place importante dans le domaine pétrolier, un champ d'application des SIG précurseur en terme de modélisation. HBDS a été introduit par [Bouillé 77]. Ce modèle est très proche de l'orienté-objet et a évolué parallèlement à cette technique. L'objectif de l'auteur est de proposer une méthode pour modéliser des données en ne se concentrant ni sur des contraintes de stockage ni sur des objectifs immédiats de manipulation. Une faiblesse de HBDS est que ce modèle n'est pas très répandu dans le domaine des SIG.

Les atouts principaux de HBDS sont un langage graphique extrêmement clair, et une bibliothèque de structures et de mécanismes permettant d'envisager des résolutions de problèmes variées, mettant en œuvre des techniques diverses, à partir d'un même jeu de données. Ce dernier atout est primordial dans notre contexte, aussi nous détaillons les éléments de ce modèle, reprenant l'exposé de [Bouillé 96].

Les concepts statiques de base de HBDS sont les types abstraits de données, TAD. Ces types sont regroupés en diverses catégories. Les éléments du squelette sont : la *classe*, l'*attribut de classe*, le *lien entre deux classes*. Les réalisations du squelette sont : l'*objet*, l'*attribut d'objet*, le *lien entre objets*. Les extensions sont : l'*hyperclasse*, l'*hyperattribut*, l'*hyperlien*, l'*embryon*, le *prototype* et la *structure*. Un catalogue de structures géographiques réutilisables, comme le réseau, est proposé.

Des éléments dynamiques sont ajoutés à ces types statiques. Les plus élémentaires sont les *conditions*, les *actions*, et les *règles*. Un moteur d'inférence est inclus dans HBDS, le SES, structured expert system. HBDS comporte également le concept de *contrainte*.

L'hyperclasse *processus* permet de modéliser l'aspect dynamique de phénomènes. Ses attributs sont : l'état, l'horloge, le corps, le signal d'arrêt du corps. La modélisation de la dynamique d'un phénomène se fait en plaçant la classe décrivant ce phénomène dans l'hyperclasse *processus* et en écrivant le code informatique correspondant à son corps. Le moteur HBDS dédié à la gestion de processus permet de réaliser des simulations complexes.

Par ailleurs, un ADT peut avoir un *comportement* générant des *actions* quand des *événements* se produisent. Le couple (*événement – action*) est appelé *réflexe* et l'ensemble des *réflexes* d'un composant forme son *comportement*. Un *réflexe* est lié à un ADT particulier : son *déclencheur*. Le *neurone* est une hyperclasse munie d'un attribut *comportement*. Cet élément est dédié à être exploité dans des mécanismes d'apprentissage automatique. Les résolutions de problèmes que cette technique supporte efficacement sont les simulations de propagation dans un réseau et de diffusion.

Enfin, des *foncteurs* sont associés à ce modèle. Ce sont des fonctions qui prennent leur valeur dans une multiplicité de types. Toute manipulation de données, modélisées en HBDS, se traduit comme l'application de foncteurs, ou encore, graphiquement, par un chemin à l'intérieur d'un graphe HBDS. Un intérêt des foncteurs est que l'expression d'une requête est proche, en structure, du détail de cette requête en langage naturel.

HBDS est finalement un excellent langage de description d'un domaine et de mécanismes de manipulation des éléments de ce domaine. Il intègre de nombreux mécanismes. Par contre, il ne relie ces éléments ni à des objectifs d'utilisation (le *pourquoi*), ni à des enchaînements de foncteurs permettant d'atteindre des objectifs (le *comment*).

2.2.5.2 Les descriptions normalisées d'outils

Nous présentons rapidement les démarches en interopérabilité de l'information géographique car elles doivent conduire à de meilleurs outils et à une description normalisée de ces outils qui facilite l'accès d'utilisateurs à ces outils.

Les premières avancées en interopérabilité sont dus à des standards faits de fait comme DXF, SDTS, des formats propriétaires de SIG très utilisés [Clément et al. 97]. Des familles de standard ont ensuite été proposées par des organismes de normalisation (CEN/TC287, ISO/TC211). La principale approche actuellement est menée par le

consortium OpenGIS réunissant des vendeurs de SIG, des vendeurs de données, des universitaires et des utilisateurs [OpenGIS TC 98]. Son objectif est de construire un ensemble de spécifications, OGIS, pour implémenter les interfaces de composants manipulant ou stockant de l'information géographique. Ces spécifications doivent se construire comme un compromis entre les fournisseurs et utilisateurs de diverses techniques. En pratique, cette démarche prend essentiellement place dans le domaine commercial. [Vckovski 98] souligne alors qu'une telle démarche visant à mettre en place l'interopérabilité la plus complète n'est pas un processus naturel. En effet elle peut avoir pour effets :

- des difficultés pour cacher des technologies propriétaires,
- un frein à l'adoption de nouvelles technologies,
- un frein au turn-over des versions de logiciels imposé par les vendeurs pour vendre plus.

Dans ce contexte, la définition d'outils -plus exactement des interfaces de modules- n'obéit ainsi pas nécessairement à la logique de proposer de meilleurs outils.

Dans une démarche comme OpenGIS, il faut souligner que le résultat effectif consiste surtout à permettre la mise en place de cercles d'interopération entre des acteurs qui communiquent déjà entre eux (les difficultés de cette communication ayant amené ces acteurs à travailler sur les outils et formalismes en jeu pour améliorer cette communication). Or il est également important d'encourager le développement de telles situations quand elles ne sont pas encore établies, de permettre l'établissement de nouveaux objectifs exigeant des situations d'interopérabilité. Dès lors, les systèmes et acteurs concernés ne sont pas tous connus à l'avance ; il s'agit en fait non plus de compatibilité en cercle mais de « réseau » ouvert, c'est-à-dire une interopérabilité ouverte des ressources. Il existe des utilisateurs potentiels que l'on ne connaît pas encore et qu'il ne faut pas exclure de l'interopérabilité à venir. De plus, il n'est probablement pas souhaitable que l'accès aux données ne se fasse que via une couche SIG. Ces SIG ne connaissent actuellement pas toute la richesse de certaines bases de données géographiques. Il est important de préserver une distinction entre les fournisseurs de contenu et les fournisseurs d'outils.

Une description des services disponibles sur les données géographiques est mise en place dans le cadre de tels travaux de normalisation et d'interopérabilité. OGIS distingue des grandes catégories de services (distribués) spécifiquement géographiques. Chaque service est spécifié par ses interfaces. Chaque catégorie se décompose elle-même en plusieurs sous-catégories de services.

- l'accès (interfaces pour localiser et retirer de l'information géographique, pour gérer le contenu d'une bibliothèque..)
- la transformation de coordonnées
- l'annotation
- la manipulation d'images
- la manipulation d'objets géographiques (création, contrôle de qualité, analyse, présentation)
- l'exploitation d'images (photogrammétrie)
- l'analyse d'objets géographiques (dérivation d'information quantitative pour une application à partir d'objets)
- la généralisation d'objets (simplification de la visualisation)
- la génération de cartes images
- la présentation (génération de sortie pour affichage à l'écran ou impression)
- la synthèse d'images
- le traitement de formes géométriques dans les images (par exemple la production d'orthophotographies)
- extraction d'information (d'images scannées ou issues de la télédétection)
- gestion des symboles
- compréhension d'image.

ISO fournit de plus un modèle très générique pour décrire un service : les services metadata, Figure 43.

Mais ce modèle de métadonnées ISO n'est pas documenté sur des services existants. De plus, il décrit les signatures des services, c'est-à-dire les types d'entrée et de sortie, mais l'interprétation du service n'est contenue que dans la sous-catégorie à laquelle il appartient, et la description de l'instance. Il n'y a donc pas d'information sur la combinaison de services dans une même application.

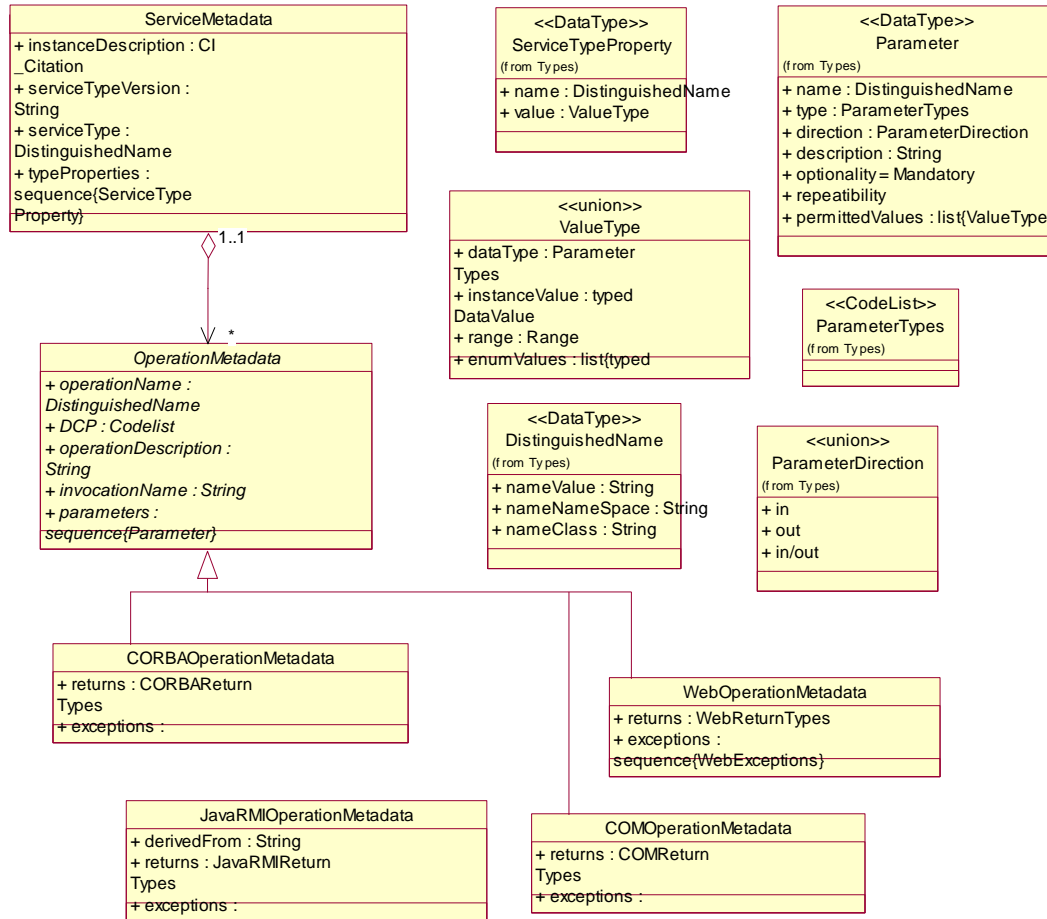


Figure 43 : Modèle ISO de métadonnées pour les services. [Figure 43 : ISO metadata model to describe services.]

Dernièrement, [OpenGIS 01] propose un nouveau standard pour l'échange d'objets géographiques, basé sur XML : Geographic Markup Language (GML). La dernière version publique de GML, GML2, ne comporte pas des éléments comme les relations de topologie mais la version suivante, GML3, doit proposer un ensemble plus exhaustif d'éléments servant à décrire des objets géographiques. Ce n'est pas une norme internationale mais ce sera bientôt un standard de fait. GML est une avancée importante dans le domaine du transfert de données géographiques. Il contribue à l'aide à l'accès dans la mesure où un fichier GML est un fichier XML et est donc lisible par un utilisateur. Par contre il ne décrit pas les manipulations possibles des données transférées.

Dans le même contexte, Virtual GIS vise à décrire les fonctionnalités d'analyse spatiale communes à tous les SIG [Albrecht 96]. L'objectif est de permettre aux utilisateurs de manipuler ces fonctionnalités sans se soucier des problèmes de formats et de codage de ces fonctionnalités dans les SIG. Les outils identifiés par l'auteur sont regroupés en catégories correspondant au contexte typique dans lequel ces outils sont utilisés, Figure 44.

Chercher	Analyse locationnelle	Analyse de terrain	Distribution/voisinage	Analyse spatiale	Mesures
Interpolation Recherche thématique Recherche spatiale (re-)classification	Buffer Corridor Overlay Thiessen/Voronoi	Slope/aspect Catchement/basins Drainage/network Intervisibilité	Coût/diffusion Proximité Plus proche voisin	Analyse multivariée Pattern/dispersion Centralité/connectivité	Mesures

Figure 44 : Les opérations SIG analytiques universelles d'[Albrecht 96]. [Figure 44 : Universal GIS analysis operations of [Albrecht 96]]

La description de VGIS ne peut être exploitée que par des utilisateurs connaissant le principe de fonctionnement de ces outils. Cela permet une manipulation transparente des outils et non d'aider des utilisateurs à comprendre comment exploiter ces outils.

De nombreuses démarches proposent des descriptions normalisées des composants. Cela aide l'accès dans la mesure où cela permet de faire intervenir le plus tard possible dans la conception de l'application les considérations de formats d'échange de données entre les composants. Ces composants peuvent être décrits à très haut niveau, les services, ou alors à un niveau plus proche de la manipulation des données, les fonctionnalités SIG.

Actuellement, de plus en plus de travaux voient l'avenir du SIG comme un ensemble de modules de services, certains offrant des fonctionnalités de type SIG, d'autres offrant par exemple des fonctionnalités d'analyse statistique, interopérables. [Abel et al. 99] s'intéressent aux applications de simulation et proposent une syntaxe de description de ressources hétérogènes, bases de données et fonctions de dérivation, dédiée à des requêtes portant sur la simulation. [Günther and Müller 98] appliquent également les techniques MMM (Method ManageMent) au domaine de l'information géographique et des composants SIG. Ces démarches vont dans le sens de faciliter l'accès d'utilisateurs aux outils interopérables en les regroupant en services.

2.2.6 Bilan

Ce deuxième volet de l'état de l'art avait pour objectif d'évaluer l'application au cas de l'information géographique du principe suivant : la modélisation distincte et intégrée de deux catégories de connaissances, le *QUOI* et le *COMMENT* (*pourquoi, comment, avec quoi*).

Nous remarquons qu'il existe davantage de formalisations axées sur le « *QUOI* » que sur le « *COMMENT* ».

Les démarches visant à décrire un *QUOI* –des hommes ou informatisé- prennent toujours place dans un contexte de recherche de consensus ou d'opérationnalisation, contextes qui conduisent toujours à la recherche de formalisation. Le fait qu'il existe de nombreuses formalisation du *QUOI* pose problème car ces formalisations ne sont pas liées les unes aux autres.

Une partie des démarches axées sur la description du *COMMENT*, les démarches visant à décrire la manipulation de l'information géographique par les hommes, ne recherchent ni un consensus sur la définition d'éléments ni à apprendre à un système à exploiter des connaissances. Elles recherchent plutôt à expliciter une expertise de manipulation de données géographiques et la conception d'applications SIG et à l'expliquer à des utilisateurs. Dans ce contexte, les auteurs ne recherchent pas d'emblée la formalisation, car elle est souvent perçue comme un obstacle à la compréhension. Les auteurs préfèrent ainsi généralement une énumération de termes commentés en langage naturel qu'un modèle structuré. Le *pourquoi* et le *comment* sont d'ailleurs très peu formalisés. Une formalisation globale devrait comporter non seulement les opérations sur les données géographiques, mais encore les méthodes d'utilisation de ces opérations pour construire des applications, et également les questions surgissant dans des domaines d'activité et nécessitant la construction d'application.

Par ailleurs, il existe de fortes interactions entre les *QUOI* et *COMMENT* géographiques. La non explicitation de ces interactions nous semble être à l'origine d'un problème majeur : la difficulté de tracer les interactions entre les activités d'appréhension, de formalisation et de représentation de l'information géographique aux divers stades où elles se produisent.

Pourtant, dans le contexte de l'information géographique, faire la distinction entre la nature d'un objet – l'élément du domaine- et la fonction d'un objet manipulé –le rôle- est particulièrement pertinent. L'information géographique renvoie à une même réalité et donc il est important d'explicitier des caractéristiques de cette réalité –dans le domaine-. L'information géographique est aussi un outil destiné à de multiples utilisations, à divers raisonnements. Pour chaque raisonnement, il existe une structuration privilégiée de cette information qui est décrite dans l'ensemble des rôles associés au raisonnement en question.

En définitive, eu égard à la formalisation d'un *QUOI* et *COMMENT* des données géographiques, certaines catégories de connaissances sont modélisées et d'autres ne le sont pas. Les catégories modélisées sont par ailleurs peu reliées les unes aux autres. Il existe des briques permettant d'envisager la construction d'un tel modèle mais il faut encore relier ces briques les unes aux autres et les compléter.

La piste technique mise en évidence à l'issue de la section 2.1 est donc prometteuse. Elle n'a pas encore été explorée mais, pour autant, il existe des travaux qui vont dans son sens. Nous fondons sur elle notre contribution présentée dans la partie suivante, ainsi que sur la prise en compte des contributions de fait à cette piste que nous avons relevées dans la section 2.2.

3 NOTRE APPROCHE

La partie précédente a permis d'une part de définir une technique efficace d'aide à l'accès, la construction d'un modèle du *QUOI* et *COMMENT* des ressources, et d'autre part d'évaluer l'applicabilité de ce principe en passant en revue les catégories d'un tel modèle qui ont déjà fait l'objet d'une formalisation.

Dans cette partie, nous présentons notre approche.

La section 3.1 explique les principes généraux de cette approche. Nous nous concentrons sur la construction d'un modèle intégrant le *QUOI* et le *COMMENT* des données géographiques. Pour que ce modèle contribue effectivement à l'accès d'utilisateurs, nous lui associons un système de résolution coopérative de problème qui sache exploiter ce modèle pour faciliter l'accès d'utilisateurs aux ressources en information géographique. Ce système doit également supporter un autre processus : l'intégration de nouvelles connaissances dans le modèle comme un nouveau jeu de métadonnées, un nouveau traitement, une nouvelle méthode.

Nous définissons dans cette section ce modèle grâce à une grammaire Backus Naur Form.

Nous construisons dans la section 3.2 les premiers éléments d'une base de connaissances associée à notre modèle. Cela permet d'évaluer la capacité de ce modèle à expliciter les connaissances de raisonnements géographiques. Cela permet également de démontrer en quoi il aide à distinguer les connaissances de type *QUOI* des connaissances de type *COMMENT*.

La section 3.3 présente une démarche de prototypage évaluant les fonctionnalités d'aide à l'accès associées à ce modèle.

3.1 Définition de notre contribution

Cette section met en place notre approche en tirant partie des conclusions des deux volets de notre état de l'art.

3.1.1 Principes généraux

Proposition

L'aide à l'accès que nous voulons mettre en place consiste en une architecture de métadonnées qui associe à la description du contenu des jeux de données la description du mode d'emploi de ces jeux. Cette architecture doit faire les liens entre des besoins d'utilisateurs, et des contenus associés à leurs modes d'emploi appropriés pour répondre aux besoins.

Il y a deux cas d'utilisation d'une telle architecture, qui correspondent à deux situations d'accès différentes. Dans le cas 1, l'utilisateur possède une ressource spécifique et souhaite connaître les connaissances qu'il peut en dériver, c'est-à-dire ses modes d'emploi possibles. L'accès se limite donc au transfert de connaissances, la sélection d'information utile étant laissée de côté. Dans le cas 2, l'utilisateur a un besoin de connaissances géographiques et souhaite connaître quelle exploitation de quelles ressources peut lui apporter ces connaissances, c'est-à-dire quelles ressources acquérir et quel mode d'emploi utiliser pour ces ressources. Il s'agit là d'un véritable accès : traduire un besoin en une sélection de ressources répondant à ce besoin. Notre travail se concentrera sur ce deuxième cas qui est le plus urgent à résoudre.

La technique d'aide à l'accès utilisée est l'aide au raisonnement formel analysée dans l'état de l'art 2.1.2.3. Aussi, quel que soit le cas d'utilisation envisagé, cas 1 ou cas 2, l'aide à l'accès se décompose en deux parties décrites ci-dessous :

- Le modèle sur lequel le raisonnement formel constituant l'accès peut être réalisé –dans l'absolu-. L'architecture de métadonnées que nous proposons repose sur un modèle de description des BDG dont la nouveauté par rapport aux formalismes vus dans l'état de l'art est d'opérer la distinction entre les connaissances de type *QUOI* et celles de type *COMMENT*, et d'explicitier les liens entre ces deux types de connaissances. Ceci repose sur la modélisation de notions spécifiques : les tâches et les rôles.
- Un environnement de résolution coopérative de problèmes répartissant ce raisonnement entre lui-même et l'utilisateur.

Première composante de l'aide à l'accès : un modèle

Nous détaillons ci-dessous en quoi un modèle décrivant des connaissances de type *QUOI* et *COMMENT* sur les ressources supporte le raisonnement formel constituant l'accès (dans le cas 1 ou cas 2).

Dans le cas 1, le raisonnement d'accès est la détermination des connaissances que l'on peut dériver de ressources données. Il s'appuie sur les éléments suivants du modèle : les descriptions de ressources, les liens entre les ressources et les opérations disponibles sur elles, les liens entre les opérations et les méthodes qui les utilisent, les liens entre les méthodes et les objectifs qu'elles permettent d'atteindre.

Dans le cas 2, le raisonnement d'accès est la traduction d'un besoin de connaissances, formulé par l'utilisateur, en une sélection d'information et de ressources. Il s'appuie sur les éléments suivants du modèle : un concept d'utilisateur est lié à des objectifs énoncés à l'aide de ce concept, ces objectifs sont associés à des méthodes permettant de les atteindre, les méthodes sont reliées aux outils SIG et aux données qu'elles mettent en œuvre. Cette aide est illustrée sur la Figure 45. Le modèle comprenant ces structures est détaillé section 3.1.2.

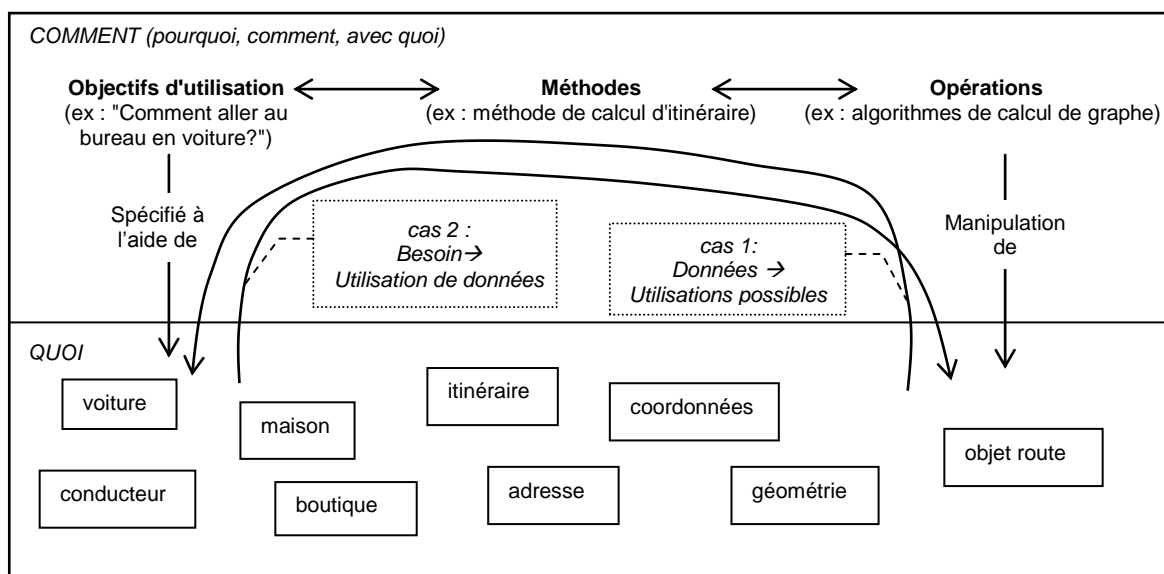


Figure 45 : Contribution d'un modèle intégrant des connaissances géographiques de type *QUOI* et *COMMENT* à l'accès d'utilisateurs à l'exploitation de données géographiques. [Figure 45 : Contribution of a model integrating *WHAT* and *HOW* geographic knowledge to users' access to the exploitation of geographic data.]

La mise en place de cette première composante de l'aide consiste à définir ce modèle, c'est-à-dire des éléments permettant de décrire des jeux de données géographiques en intégrant des connaissances de type *COMMENT* (*pourquoi, comment, avec quoi*) et *QUOI*. Les éléments génériques de ce modèle sont décrits section 3.1.2. La section 3.2 spécifie ensuite cette description générique pour décrire effectivement des raisonnements géographiques.

Deuxième composante de l'aide à l'accès : un environnement de résolution coopérative de problèmes

Une deuxième partie de l'aide consiste à répartir ce raisonnement entre l'utilisateur et un environnement de résolution coopérative de problèmes. Nous détaillons ci-dessous comment cette répartition est faite.

Dans le cas 1, l'utilisateur contribue au raisonnement global en décrivant les données qu'il possède. Le système expert détermine les informations et connaissances dérivables de ces données par l'application de traitements. Le système doit apporter à l'utilisateur des connaissances sur le sens et le mode d'emploi des données qu'il possède.

De plus, le système doit prendre en compte la possibilité de combiner les données que possède l'utilisateur à d'autres données. Ce cas est illustré Figure 46.

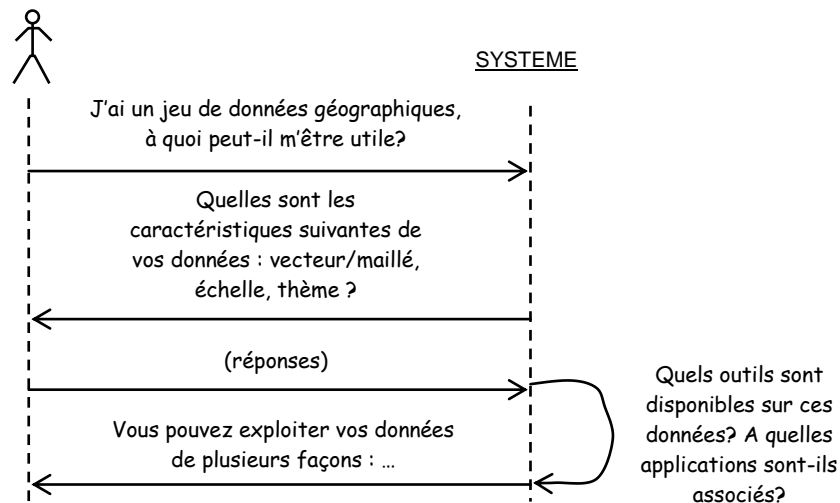


Figure 46 : Type d'aide à l'accès dans le cas 1. [Figure 46 : Scenario of users' access in Case 1.]

Dans le cas 2, l'utilisateur exprime son besoin de connaissances. Le système détermine quelles données et quelle manipulation de ces données peuvent apporter une réponse à ce besoin. Ce cas est illustré Figure 47.

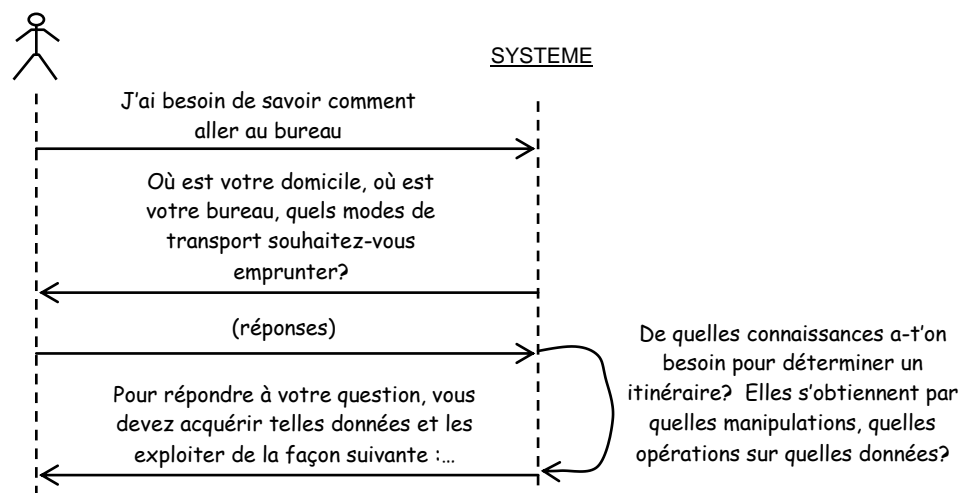


Figure 47 : type d'aide à l'accès dans le cas 2. [Figure 47 : Scenario of users' access in Case 2.]

Nous ne détaillerons pas davantage le cas 1, notre analyse portera sur le cas 2.

Pour mettre en place cette deuxième composante de l'aide, il faut implémenter dans un système la base de connaissance associée à notre modèle de description des jeux de données géographiques, et définition des fonctions d'aide à l'accès –dans le cas 2- que l'on peut associer à cette base de connaissances. Le système possédant cette base de connaissances et ces fonctions est en fait un environnement de résolution coopérative de problèmes. Sa construction nécessite :

- de coder les mécanismes correspondant à la contribution du système au raisonnement global d'accès (ces mécanismes exploitent la base de connaissance),
- d'établir un mode de communication entre l'utilisateur et le système permettant de construire le raisonnement global à partir de leurs contributions.

Ce système est présenté en section 3.1.3. L'implémentation d'un prototype est décrite dans la section 3.3.

3.1.2 Définition du modèle

Nous décrivons dans cette section notre modèle de description des jeux de données géographiques intégrant les notions clés de tâches, méthodes, et rôles présentées précédemment section 2.1.3.3. Ces notions sont plus proches du raisonnement de l'homme que des "raisonnements" de systèmes informatiques. Elles ne peuvent pas être décrites facilement dans des structures manipulables par une machine comme des propositions logiques ou des objets java. Ceci explique d'ailleurs que les méthodes classiques d'ingénierie de logiciel ne supportent pas encore de tels modèles.

Il est donc important de définir un premier modèle explicitant ces notions, avant même de proposer une implémentation de ce modèle tenant compte des contraintes informatiques. Cette section est dédiée à la présentation du modèle d'expertise. Le modèle d'implémentation sera présenté en section 3.3.

Nous introduisons dans cette section les éléments du modèle, les connaissances géographiques qu'ils permettent de décrire, et ce que ces connaissances signifient dans un système d'aide à l'accès.

3.1.2.1 Les tâches

Une tâche est un problème que l'on sait résoudre. Rappelons qu'une tâche ne correspond pas à un cas précis de problème résolu (comme : aller en métro le 6 mars 2000 à 15h de la place de la Bastille à la Place de Clichy à Paris) mais plutôt à une famille de problèmes (comme : aller d'un endroit de Paris à un autre en métro). La description d'une tâche peut être spécifiée pour décrire comment résoudre un problème précis qui est un cas particulier du problème général décrit dans la tâche. Aussi, les éléments que nous utilisons pour décrire une tâche doivent tenir compte de la nécessité de décrire une famille de problèmes particuliers.

La description d'une tâche se décompose en :

- la description du problème posé (facette déclarative de la tâche),
- la description de la résolution du problème (facette opérationnelle de la tâche).

La facette déclarative d'une tâche est décrite dans notre modèle par des entrées et des sorties qui sont des variables génériques pouvant être spécifiées. Parmi les sorties de la tâche, il y a souvent une sortie principale qui est l'objectif de la tâche et il peut y avoir d'autres sorties qui sont des connaissances produites par la tâche sans être l'objectif véritable de la tâche, on les omet ou on les appelle sorties secondaires.

Il existe deux catégories de tâches selon que la facette opérationnelle varie ou non avec le contexte de réalisation de la tâche, c'est-à-dire en fonction de la spécification des entrées et sortie de la tâche. On distingue les tâches complexes qui décrivent un raisonnement complexe dont la résolution varie en fonction du contexte et les tâches primitives qui décrivent un raisonnement dont la résolution ne varie pas. Cette distinction renvoie à un choix de modélisation. Il n'existe pas de limite franche dans la réalité entre des tâches complexes et des tâches primitives.

La personne construisant un modèle choisit de détailler une tâche en la modélisant comme une tâche complexe, ou de ne pas la détailler en la modélisant comme une "boîte noire", une tâche primitive.

Notons que la spécification d'une tâche pour décrire un problème précis se fait toujours en spécifiant la facette déclarative de la tâche et en propageant cette spécification à une autre spécification de cette facette ou à une spécification de la facette opérationnelle.

Dans un contexte d'aide à l'accès, une tâche est utile car elle représente une manipulation générique de ressources. Une tâche dont les deux facettes, déclarative et opérationnelle, sont spécifiées décrit une utilisation particulière de données géographiques.

3.1.2.2 Les tâches complexes

Eléments de représentation

Une tâche complexe est une tâche dont la facette opérationnelle dépend du contexte de réalisation de la tâche. Après spécification, la facette opérationnelle est un plan d'exécution, c'est-à-dire une structure de tâches primitives qui ne comporte pas de choix. Il est rarement possible de décrire tous les plans d'exécutions associés à une tâche complexe. La facette opérationnelle, appelée méthode, est composée d'un plan générique non détaillé (c'est-à-dire qui peut comporter des sous tâches complexes et des choix) et de connaissances permettant de guider la spécification de la tâche. Ces connaissances sont représentées sous forme de règles. Les prémisses de telles règles sont toujours une spécification d'un rôle et les conclusions soit une spécification d'un rôle soit une spécification du plan.

Ces éléments sont résumés sur la figure Figure 48.

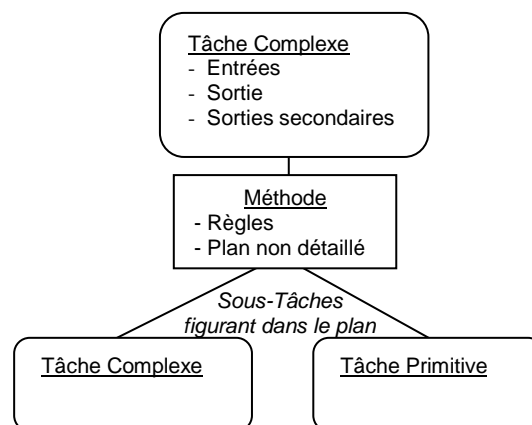


Figure 48: Eléments explicites représentant une tâche complexe et notations graphiques.

Sans détailler le processus de spécification d'une tâche complexe, qui est décrit plus loin 3.1.3, nous pouvons faire quelques remarques concernant les connaissances de spécification d'une tâche complexe. Le plan d'exécution d'une tâche spécifiée est obtenu à partir du plan non détaillé, en résolvant les choix de ce plan, et en remplaçant les tâches complexes par leurs plans non détaillés pour résoudre les choix apparaissant éventuellement dans ses plans non détaillés, jusqu'à ce qu'il n'y ait plus ni tâches complexes ni choix dans le plan.

Les connaissances de spécification du plan d'une tâche complexe ne sont pas toutes décrites au niveau de cette tâche, elles sont également décrites dans les méthodes des tâches complexes entrant dans les décompositions successives de cette tâche. Globalement, les connaissances de spécification d'une tâche sont :

- des connaissances de résolution des choix du plan non détaillé. Elles sont représentées dans des règles de la méthode de cette tâche,
- des connaissances de résolution des choix des plans non détaillés des sous-tâches complexes apparaissant dans le plan de cette tâche, et dans les décompositions successives du plan. Elles sont représentées dans des règles des méthodes de toutes les sous-tâches de la tâche principale.

Par contre, les règles de spécification d'une tâche complexe sont les mêmes quelle que soit la tâche dans le plan de laquelle elle puisse figurer.

Notons qu'une tâche complexe est associée finalement à deux types de résultats. La sortie de la tâche (le résultat attendu) et la sortie produite par le plan d'exécution. Ces sorties ne sont pas identiques. Le résultat attendu renvoie au besoin d'un utilisateur et la sortie du plan détaillé renvoie à un résultat d'application, qui est une forme de réponse au résultat attendu. Une difficulté dans la construction d'un tel modèle est que la sortie produite par le plan détaillé soit effectivement une réponse acceptable au résultat attendu. Si le plan est exécuté, une autre difficulté est que le résultat obtenu soit effectivement la sortie prévue dans le plan.

Les tâches complexes géographiques

La notion de tâche complexe géographique est utilisée pour décrire des applications géographiques génériques, c'est-à-dire un objectif général lié à l'exploitation de données géographiques et une méthode générale pour atteindre cet objectif. Une tâche complexe géographique doit pouvoir être spécifiée pour décrire une méthode de résolution de cas particuliers de ce problème générique. Prenons l'exemple du calcul d'itinéraire. La méthode générale de calcul d'itinéraire consiste à construire un graphe pondéré et à effectuer un calcul d'optimisation sur ce graphe. L'objectif peut être spécifié en précisant par exemple un lieu de départ, un lieu d'arrivée, des étapes souhaitées, un ou plusieurs véhicules, des contraintes de coût en temps ou financier, des souhaits sur les portions d'itinéraires. En fonction de ces spécifications, les éléments constituant le graphe sont spécifiés : les entités constituant ses arêtes (tronçons routiers, lignes de métro,...), les entités constituant ses sommets (ville, bâtiment remarquable, centre commercial,...), les poids.

Il est important d'explicitier des tâches transversales aux domaines d'applications, comme déterminer une localisation, rédiger une carte, mesurer une proximité ou changer de systèmes de coordonnées.

Dans un contexte d'aide à l'accès, les tâches complexes géographiques sont utiles de plusieurs façons :

- Le besoin d'information géographique d'un utilisateur est reformulé dans le système comme une sortie de tâche complexe géographique.
- La réponse au besoin est exprimée comme un plan d'exécution d'une tâche complexe, c'est-à-dire une suite d'actions à effectuer pour dériver un résultat répondant au besoin reformulé.

3.1.2.3 Les tâches primitives

Éléments de représentation d'une tâche primitive

Certaines tâches, dites primitives, ne sont pas décomposables. Leur méthode est un mécanisme figé. Ce sont des tâches dont le point de vue déclaratif peut être spécifié mais pas le point de vue opérationnel, hormis la variation de paramètres.

Les tâches primitives sont généralement utilisées pour décrire des traitements de données. Dans ce cas, les sorties sont des jeux de données et les entrées sont des jeux de données et les valeurs des paramètres du traitement.

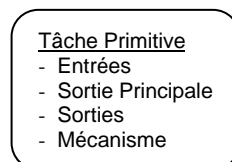


Figure 49: Représentation graphique d'une tâche primitive.

Les tâches primitives géographiques

Les tâches primitives géographiques décrivent des fonctions de dérivation d'information. Il peut s'agir d'un traitement sur des jeux de données géographiques, d'une opération de sélection de données ou encore de l'application d'opérateurs sur des modèles conceptuels, en particulier les règles permettant de passer d'un élément d'une nomenclature à un élément plus spécifique.

Un exemple de tâche primitive géographique est la détermination de l'enveloppe d'une ville grâce à l'application d'un algorithme de croissance morphologique à un ensemble de bâtiments. L'entrée nécessaire est les objets bâtiments, le mécanisme est l'algorithme de croissance morphologique et la sortie obtenue est une surface correspondant à l'extension spatiale de la ville.

Dans l'exemple précédent de la tâche du calcul d'itinéraire, une tâche primitive entrant dans la décomposition de cette tâche est un algorithme de calcul de graphe. Le corps de cette tâche primitive ne peut être modifié, hormis la définition de paramètre selon le type d'optimisation souhaité (vitesse, coût).

Dans un contexte d'aide à l'accès, les tâches primitives géographiques sont utiles pour décrire, dans le plan d'exécution proposé en réponse, les ressources SIG à acquérir.

3.1.2.4 Rôles et domaine

Éléments de représentation des rôles et du domaine

La notion de rôle est moins intuitive que celle de tâche. Les entrées et sorties des tâches sont décrites par des rôles. Ce sont des variables qui prennent leurs valeurs dans le domaine, c'est-à-dire dans le Quoi. Il est possible de comparer cela à l'écriture d'une pièce de théâtre qui se fait non pas en utilisant des noms d'acteurs mais en utilisant des noms de personnages. Ces personnages sont des rôles qui peuvent être tenus par divers acteurs.

Un rôle est composé d'un nom et d'un ensemble d'éléments du domaine qui décrit les valeurs que peut prendre ce rôle dans des réalisations de la tâche. Un ensemble d'éléments du domaine peut être décrit de deux façons, par son extension ou par son intension. Son intension est représentée comme un ensemble d'éléments du domaine qui sont les représentations génériques des éléments de l'ensemble.

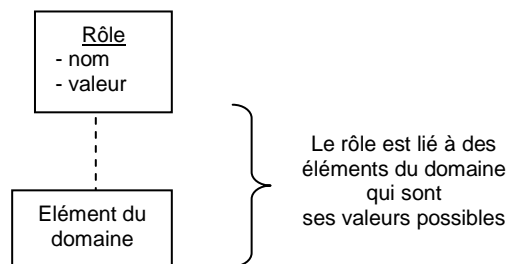


Figure 50 : Représentation d'un rôle.

Les rôles et le domaine décrivent le vocabulaire des tâches à divers niveaux.

Les rôles décrivent les structures manipulées par des applications, ils explicitent la fonction des objets manipulés, comme les sommets et arêtes d'un graphe, les classes et les critères d'une classification. Un rôle est décrit par le nom de la variable, et son ensemble de candidats c'est-à-dire la désignation des éléments du domaine sur lesquels cette variable prend sa valeur. L'ensemble des candidats a une valeur par défaut très générale, représentée au niveau du rôle. La spécification d'un rôle consiste à préciser cet ensemble. Cela peut consister à restreindre son extension, ou à spécifier son intension.

Le Domaine décrit la connaissance spécifique aux objets manipulés : les concepts et les instances de ces concepts. Les éléments du domaine sont le contenu manipulé dans une application correspondant à un cas particulier d'une tâche, i.e. des valeurs prises par les rôles dans des applications particulières.

Les rôles géographiques

Les rôles sont le vocabulaire spécifique des tâches. Par exemple, la description de connaissances de calcul d'itinéraire s'appuie sur des termes spécifiques comme « point de départ », « point d'arrivée », « dessertes », « temps de parcours », « véhicule », « moyenne ». Ces termes sont des rôles et ils appartiennent au *COMMENT*.

Les rôles concrétisent par ailleurs le lien du *COMMENT* et du *QUOI* en exprimant à quels éléments du domaine (le *QUOI*) renvoie le vocabulaire de la tâche. Par exemple, le terme véhicule renvoie aux termes : voiture, bus,

vélo, fleuve. Ces derniers termes ne sont pas des rôles car ils ne sont pas spécifiques à la tâche, ce sont des éléments du domaine. Les connaissances de valuation des rôles, c'est-à-dire quels éléments du domaine peuvent le remplir, sont un élément essentiel de notre système. Elles peuvent consister en des mécanismes de sélection désignant dans les diverses BDG les objets ou lots de données correspondant à ces rôles. Reprenons l'exemple de la tâche de calcul d'itinéraire, dans le cas où le véhicule est un convoi militaire spécial. Il est nécessaire de déterminer quels objets vont remplir le rôle d'arête du graphe, Figure 51. Il s'agit des entités sur lesquelles ce véhicule peut rouler c'est-à-dire à des contraintes de largeur de la chaussée et de capacité à supporter un certain poids. Cela peut se traduire en écrivant, pour chaque BDG, une expression désignant les éléments de cette base qui remplissent les conditions.

Rappelons que cette distinction entre les rôles appartenant au *COMMENT* et le domaine représentant le *QUOI* est essentielle. Elle se traduit par les propriétés suivantes Figure 51:

- Un même élément du domaine peut être candidat pour remplir des rôles différents selon la tâche en cours. Par exemple, dans un calcul d'itinéraire, un tronçon de route est un arc de graphe. Dans un calcul de nuisance, il est un foyer de pollution. Dans une planification de rénovation des chaussées il est une surface. Cela permet la réutilisation de mêmes ressources dans diverses tâches.
- De plus divers éléments du domaine peuvent être manipulés dans une tâche de façon indistincte, c'est-à-dire sans se préoccuper de distinguer ces éléments entre eux si cela n'est pas nécessaire pour la tâche, par exemple des ponts et tronçons de route dans une tâche de calcul d'itinéraire. Cela permet de réutiliser une tâche dans des cas particuliers où les objets manipulés ne sont pas les mêmes.

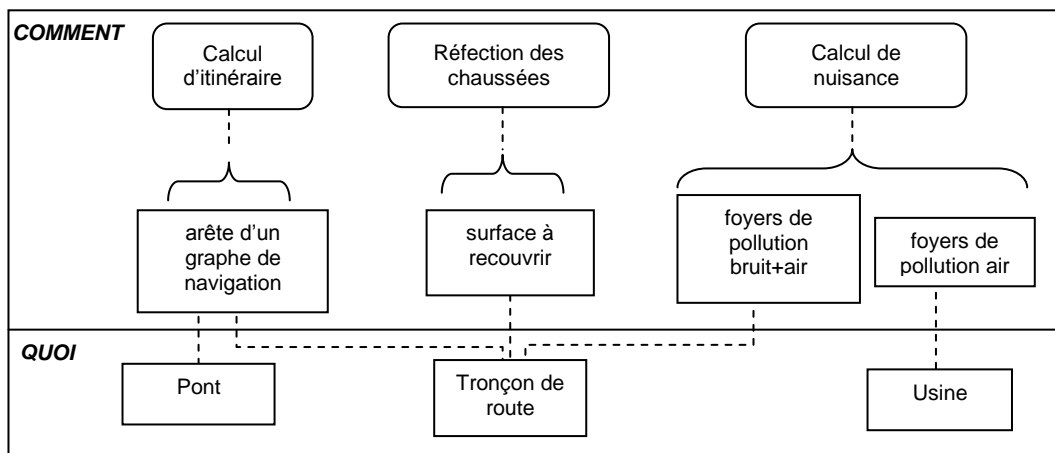


Figure 51 : Le *QUOI* est structuré en fonction du *COMMENT* grâce aux rôles. [Figure 51 : The *WHAT* is structured depending on the *HOW*, thanks to roles.]

Le domaine géographique

Pour chaque tâche, le domaine contient les valeurs particulières de ses rôles dans les applications correspondant à des cas particuliers de cette tâche. Comme annoncé sur la Figure 35, cela implique de représenter dans le domaine deux grandes catégories d'objets :

- Les objets permettant de spécifier un résultat attendu. Ces objets prennent souvent place dans le contexte d'activités de l'homme comme se déplacer, passer ses vacances, observer. Par contre, ils ne sont pas toujours explicités dans les données géographiques, comme par exemple une ville, une montagne, une région montagneuse, une navigation, une cartographie. Etant nécessaires au raisonnement de l'utilisateur, ils doivent être présents dans le modèle pour l'aider à exprimer son besoin.
- Les objets permettant de décrire des entrées nécessaires ou des résultats obtenus : des jeux de données, extraits des BDG ou dérivés par application de traitements. L'aide que notre système apporte aux utilisateurs ne consiste pas à dériver des connaissances géographiques à la demande, mais à fournir des plans de

dérivation de connaissances. Cette aide ne nécessite pas de manipuler des données géographiques mais elle s'appuie sur la manipulation de métadonnées. C'est pourquoi le domaine ne comporte pas de BDG mais plutôt des descriptions des BDG. Il faut prévoir par ailleurs de représenter aussi des bases de données liées à des activités géographiques et apportant de l'information utile dans des applications géographiques, comme des données INSEE ou des données géologiques.

Nous ne pouvons pas délimiter un domaine de l'information géographique. Cela est possible lorsque l'expertise renvoie à un « micro-monde » relativement isolé et complet tel que sa description comprend toutes les connaissances nécessaires à la résolution de problèmes surgissant dans ce « micro-monde ». Dans le cas de l'information géographique, il n'est pas envisageable de construire un tel « micro-monde » en raison du réseau complexe de relations entre les objets du monde réel. Seule une description extensible est possible. Cette description doit pouvoir être croisée avec divers modèles métiers.

La Figure 52 résume les éléments structurés dans un modèle des tâches géographiques.

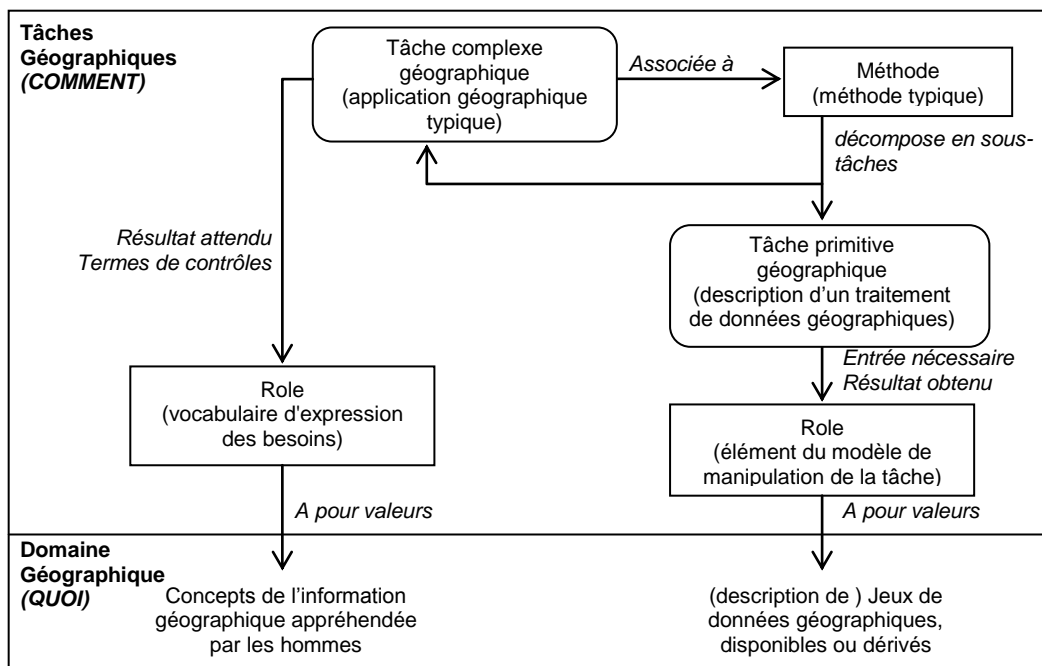


Figure 52 : Connaissances touchant à l'information géographique structurées par notre modèle de tâches. [Figure 52 : Geographic information elements that can be structured by our tasks model.]

Dans un contexte d'aide à l'accès, la modélisation du domaine et des rôles géographiques est utile de plusieurs façons :

- Le domaine décrit, entre autres, les jeux de données que l'utilisateur peut exploiter. Il s'agit du Quoi géographique informatique.
- Les rôles des tâches primitives géographiques apparaissant dans le plan d'exécution proposé en réponse ont pour valeur des éléments du Quoi géographique informatique. Ils désignent les jeux de données que l'utilisateur peut utiliser pour dériver l'information qu'il souhaite.
- Le domaine décrit également les termes utilisés dans les représentations de l'espace géographique que se font des utilisateurs. Il s'agit du Quoi géographique des hommes.
- Les rôles des tâches complexes géographiques désignent les termes pertinents pour décrire le besoin et le contexte de l'utilisateur. Ces termes prennent leur valeur dans le Quoi géographique des hommes.

3.1.2.5 Un langage de description de tâches

Nous proposons un langage permettant de décrire sans ambiguïtés les connaissances importantes de notre modèle. Ce langage ne cherche pas à décrire des tâches de façon exhaustive mais à expliciter des connaissances suffisantes dans notre contexte et à les structurer de façon à ce qu'elles soient exploitables. L'exploitation des descriptions de tâches géographiques ne consiste pas dans notre contexte à réaliser ces tâches mais à déterminer la tâche représentant le mieux le besoin de l'utilisateur et la réponse à ce besoin.

Ce langage est défini à l'aide de la grammaire Backus-Naur Form étendue. La Figure 53 présente notre grammaire. Nous la commentons ci-dessous.

Une *tâche* est définie comme une tâche complexe ou une tâche primitive.

Une *tâche complexe* est définie par les éléments suivants :

- son nom, suivi éventuellement d'une description,
- sa sortie qui est un rôle,
- ses entrées éventuelles qui sont des rôles,
- ses sorties secondaires éventuelles qui sont des rôles
- sa méthode.

Une *tâche primitive* est définie par une description sommaire ou détaillée. Sa description détaillée est proche de celle d'une tâche complexe à la différence près qu'elle n'a pas de méthode mais un mécanisme. Un *mécanisme* est défini par son nom.

Il s'agit là de la description de tâches génériques. Il est possible de décrire des spécifications de tâches génériques. Une *tâche spécifiée* est décrite par le nom de la tâche générique, et par les rôles spécifiés, c'est-à-dire les rôles génériques associés aux ensembles spécifiés. Soulignons que la description d'une tâche complexe générique n'explicite pas nécessairement tous les rôles de cette tâche. Certains rôles, dont on ne peut définir de façon générique les candidats a priori, ne sont décrits qu'à partir du moment où ils sont spécifiés.

Un *rôle* est défini par un nom suivi d'un ensemble qui décrit les valeurs candidates du rôle.

Un *ensemble* peut être défini par un nom, par une intension, par une extension, ou comme l'union de deux ensembles. Il peut aussi être défini comme les valeurs d'une propriété d'un élément.

Une définition intensive d'un ensemble consiste à donner des conditions d'appartenance à cet ensemble. Elle s'appuie sur un ou plusieurs éléments génériques du domaine, par exemple *Int(entité)* s'appuie sur l'élément du domaine *entité* pour désigner l'ensemble de tous les éléments du domaine qui sont des entités.

Une définition extensive d'un ensemble consiste à énumérer ses éléments. C'est donc une énumération d'éléments du domaine.

Un *élément du domaine* est défini par un nom et éventuellement des propriétés. Une propriété est définie par son nom et l'ensemble de ses valeurs. Par exemple, l'élément nommé "entité" a une propriété nommée "est identifiée par" qui a pour valeurs possibles l'ensemble défini par l'intension "identifiant sémantique".

Un élément du domaine peut aussi être désigné par un ensemble (tel qu'il est l'élément définissant l'intension de cet ensemble), ou par un rôle (tel qu'il est l'élément définissant l'intension des candidats).

Dans la description courante d'une tâche, un même mot, comme "système de référence" renvoie souvent soit à un élément du domaine, soit à un ensemble (qu'il définit intensionnellement), soit à un rôle (dont il définit intensionnellement les candidats). Dans notre langage, la distinction entre ces trois notions est clairement établie car ce sont des types différents. Par exemple le rôle "système de référence" dans la tâche de localisation renvoie aux systèmes de références décrits dans le domaine, ou à une cartographie, ou à une navigation. L'ensemble "système de référence"

renvoie aux systèmes de références décrits dans le domaine qui peuvent être des systèmes directs, des systèmes linéaires, des systèmes d'adresses postales ou des systèmes administratifs.

Une *méthode* est définie par :

- sa description éventuelle en langage naturelle,
- des règles éventuelles de proposition de spécification,
- des règles éventuelles de propagation de spécification,
- des règles éventuelles de stratégie, c'est-à-dire de modification du plan,
- un plan non détaillé qui est une structure de tâches.

Une structure de tâche est une structure classique telle que celles utilisées dans les langages procéduraux. Les procédures élémentaires sont ici des tâches spécifiées. Une structure de tâches est composée de tâches spécifiées, ou de structures de tâches, et d'un terme d'enchaînement qui peut être "ou", "et", "tant que", "si"..."alors".. "sinon". Les expressions booléennes utilisées dans les "si" et "tant que" peuvent être des valeurs exprimées par un utilisateur ou des tests effectués sur des rôles, et plus précisément sur leurs candidats.

Une structure de tâche possède également éventuellement un vocabulaire. Le vocabulaire d'une structure sert à désigner des rôles importants définis dans les tâches composant cette structure. Un terme du vocabulaire est défini par un nom et un pointeur vers des rôles.

Un pointeur vers des rôles peut être de deux types. Ce peut être un pointeur direct vers des rôles appartenant à des sous-tâches de la structure. Ce peut être un pointeur indirect qui désigne non pas des rôles mais des termes parmi les vocabulaires des sous-structures.

Les règles de proposition ou de propagation de spécification sont des *règles de spécification* définies par une spécification de rôle (prémisse) qui implique une ou plusieurs spécifications de rôles (conclusion).

Une *spécification de rôle* est définie par le nom du rôle spécifié suivi éventuellement d'une *spécification d'ensemble* (portant sur les candidats du rôle).

Une *spécification d'ensemble* est définie par un préfixe suivi d'un ensemble.

Cette description est sommaire mais elle est suffisante à ce stade. Lors de l'implémentation, il sera nécessaire de structurer davantage ces notions de spécification d'ensemble, spécification de rôle et règles de spécification.

Les *règles de stratégie* sont définies par une spécification de rôle (prémisse) qui implique une action sur le plan (conclusion).

Une *action sur le plan* consiste à remplacer une structure par une autre c'est-à-dire généralement résoudre un choix, par exemple remplacer une structure du type " OU(structure1, structure2, structure3,..) " par " structure1 ". Soulignons que le plan n'est pas modifié uniquement par les *actions sur le plan*. Les règles de propagation de spécification agissent également sur le plan puisqu'elles peuvent spécifier un rôle apparaissant dans des tâches du plan.

Vocabulaire terminal : Nom : chaîne de caractère identifiant une tâche, un rôle, un ensemble, un élément du domaine ; Description : texte en langage naturel ; Nombre : un entier positif ou "n"; ChoixUtilisateur : valeur booléenne

```

<Tâche>::= <Tâche Complexe> | <Tâche Primitive>
<Tâche Complexe>::= TC Nom [(Description)]
    sortie : <Rôle>
    [entrée : <Rôle> {;<Rôle>}]
    [sortie secondaire : <Rôle>{;<Rôle>}]
    méthode : <Méthode>
<Tâche Primitive>::= <Tâche Primitive Détaillée>| TP(Description)
<Tâche Primitive Détaillée>::= TP Nom [(Description)]
    sortie : <Rôle>
    [entrée : <Rôle> {;<Rôle>}]
    [sortie secondaire : <Rôle>{;<Rôle>}]
    mécanisme : <Mécanisme>
< Tâche spécifiée >::=Nom{/<Rôle> = <Ensemble>|<PointeurRôle>}
<Mécanisme> ::= Nom
<Rôle>::= Nom [(Ensemble)]
<Ensemble>::= Ens Nom | Int(<Elt du Domaine>[<Elt du Domaine>]) | Ext(<Elt du Domaine>[<Elt du Domaine>]) | <Ensemble>+<Ensemble> | <Intervalle>uplet de <Ensemble> | "Nom" de <Elt du Domaine>
<Intervalle>::= Nombre[. Nombre]
<Elt du Domaine>::= Nom [(propriétés : <Propriété>{;<Propriété>})] | Elt.(<Ensemble>)|Elt.(<Role>)
<Propriété>::="Nom" <Ensemble>
<Méthode> ::= [Description]
    [Propos spécif : <RèglesSpecif>[<RèglesSpecif>]]
    [Propag spécif : <RèglesSpecif>[<RèglesSpecif>]]
    [Stratégie: <RèglesStratégie>[<RèglesStratégie>]]
    [Plan non détaillé : <Structure de tâches>]
<Structure de tâches >::= [structure Nom]
    [Vocabulaire : Nom[(PointeurRôle)]{<u>Nom[(PointeurRôle)]}]
    [<Tâche spécifiée> | ET(<Structure de tâches >[<Structure de tâches >]) | OU(<Structure de tâches >[<Structure de tâches >]) | TANT QUE (<Expression booléenne>)
    REPETE (<Structure de tâches >) | SI <Expression booléenne> ALORS <Structure de tâches > SINON <Structure de tâches > FINSI ]
    [{<Structure de tâches>}]
<Expression booléenne>::= <Exp bool simple>|<Exp bool simple>^<Exp bool simple>|<Exp bool simple>v<Exp bool simple>
<Exp bool simple>::= <TestRole>|<ChoixUtilisateur>
<TestRole>::=Nom<SpécificationEnsemble>
<PointeurRôle>::= Nom [dans <Tâche> | dans <Structure de tâche>]
<RèglesSpecif>::= <SpécificationRole>→<SpécificationRole>[<SpécificationRole>]
<RèglesStratégie>::= <SpécificationRole>→<ActionPlan>[<ActionPlan>]
<SpécificationRole>::= Spécification de Nom [(SpécificationEnsemble)]
<SpécificationEnsemble>::= <Préfixe><Ensemble>
<Préfixe>::= =|≠|+|=|=|>|<|∅

```

Figure 53: Définition du vocabulaire et de la grammaire de notre langage de description de tâches, à l'aide du langage Extended Backus-Naur Forms. [Figure 53: Definition of the vocabulary and grammar of our tasks description language, with the extended BNF language.]

Dans une optique d'aide à l'accès à l'information géographique, ce langage de description permet de :

- Stocker une expertise de manipulation que ne possèdent pas les utilisateurs et qui est nécessaire pour concevoir une application. Cette expertise consiste à déterminer pour une tâche son modèle de manipulation (les rôles) et quel contenu peut être manipulé avec ce modèle (les ensembles de valeurs associés aux rôles).
- Expliciter des connaissances permettant de reformuler un besoin d'utilisateur et de préciser des éléments du contexte de cet utilisateur, dans la mesure où ils sont pertinents pour la détermination d'une réponse au besoin.

La section 3.1.3 présente les principes de l'exploitation d'une base de connaissances structurée avec notre langage. L'implémentation de ce langage fera l'objet de la section 3.3.

3.1.3 Un système d'aide à l'accès

Une deuxième composante de notre contribution, complémentaire du modèle présenté dans la section précédente, est un système aidant l'utilisateur à exploiter une base de connaissances structurée selon ce modèle pour résoudre son problème d'accès. Ce système est un système de résolution coopérative de problème.

Nous présentons dans cette section l'architecture globale de notre système et ses fonctionnalités.

Architecture globale du système

Rappelons qu'un système à base de connaissances est globalement organisé en une base de connaissances, les connaissances statiques, et un moteur, les connaissances dynamiques, manipulant cette base, comme décrit Figure 54. Dans notre cas, l'ensemble du modèle des raisonnements géographiques est inclus dans la base de connaissances du système.

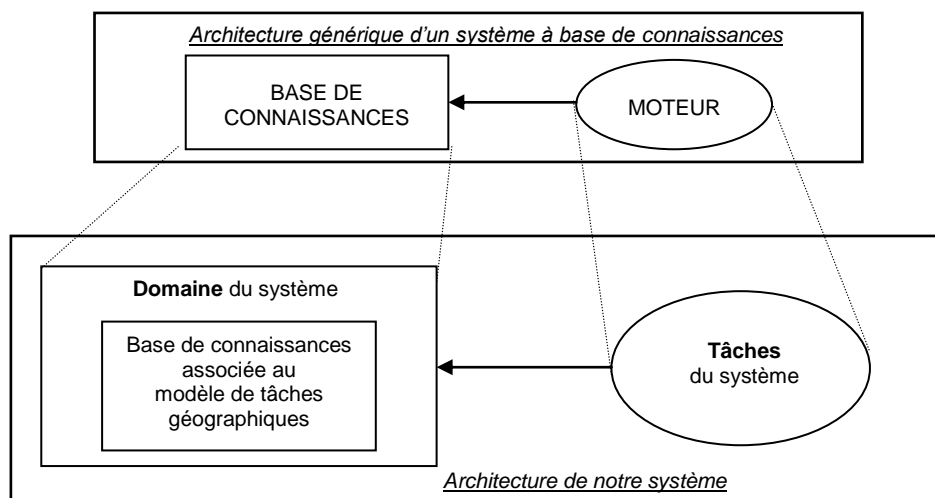


Figure 54 : L'architecture de notre système. [Figure 54 : Architecture of our system]

Les tâches géographiques ne font pas partie de la dynamique de raisonnement du système. Cette dynamique est représentée dans d'autres tâches, les tâches du système, qui permettent d'accéder à

la base de tâches géographiques. Les tâches géographiques sont finalement des éléments du domaine du système.

Fonctionnalités du système

Rappelons que parmi les cas d'accès introduits précédemment section 3.1.1. nous nous restreignons au cas décrit Figure 47, où un utilisateur a un besoin de connaissances géographiques et souhaite savoir quelles données acquérir et comment les utiliser pour dériver ces connaissances. Le système doit aider l'utilisateur à exprimer son besoin, lui répondre en retour quelles ressources il doit exploiter, et de quelle façon, pour obtenir une réponse à son besoin.

L'exploitation du modèle par le système pour remplir ces fonctionnalités ne consiste donc pas à réaliser une tâche géographique modélisée mais à spécifier une tâche géographique telle que sa facette déclarative représente le besoin de l'utilisateur et son plan détaillé représente la réponse à ce besoin.

Ce processus de spécification est la détermination d'une tâche complexe. Il consiste à construire un plan d'exécution de cette tâche, à partir de son plan détaillé, en itérant deux processus : remplacer les tâches complexes par leur plan non détaillé jusqu'à ce qu'il ne reste plus de tâche complexe, et éliminer les choix apparaissant éventuellement dans les plans non détaillés participant à cette décomposition, Figure 55.

C'est donc un processus de spécification cohérente des facettes déclaratives et opérationnelles d'une tâche géographique et des tâches géographiques entrant dans ses décompositions successives.

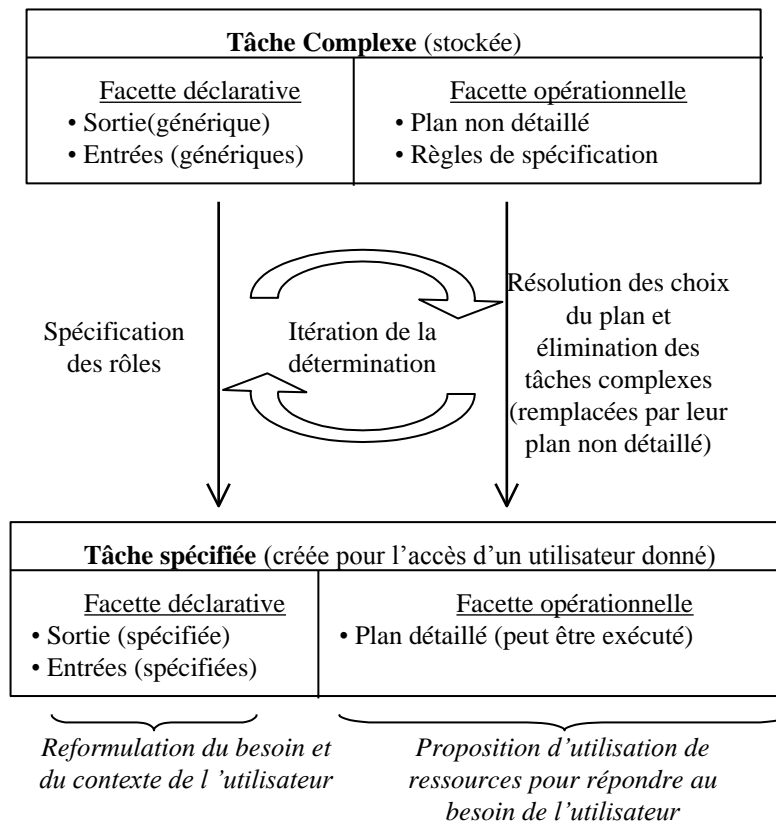


Figure 55: Processus de détermination d'une tâche complexe conduisant à une tâche représentant le besoin de l'utilisateur et une réponse à ce besoin.[Figure 55: Determination of a complex task leading to a specified task representing the user's need and an answer to this need.]

Ce processus est réparti entre l'utilisateur et le système. L'utilisateur, spécifie en fonction de son propre contexte, la sortie et les entrées de la tâche en cours de spécification. Pour faire cela, il est aidé par le système. Le système construit peu à peu le plan d'exécution de la tâche en suivant le principe schématisé Figure 55.

Nous décrivons de façon graphique les tâches du système correspondant à ces fonctionnalités Figure 56.

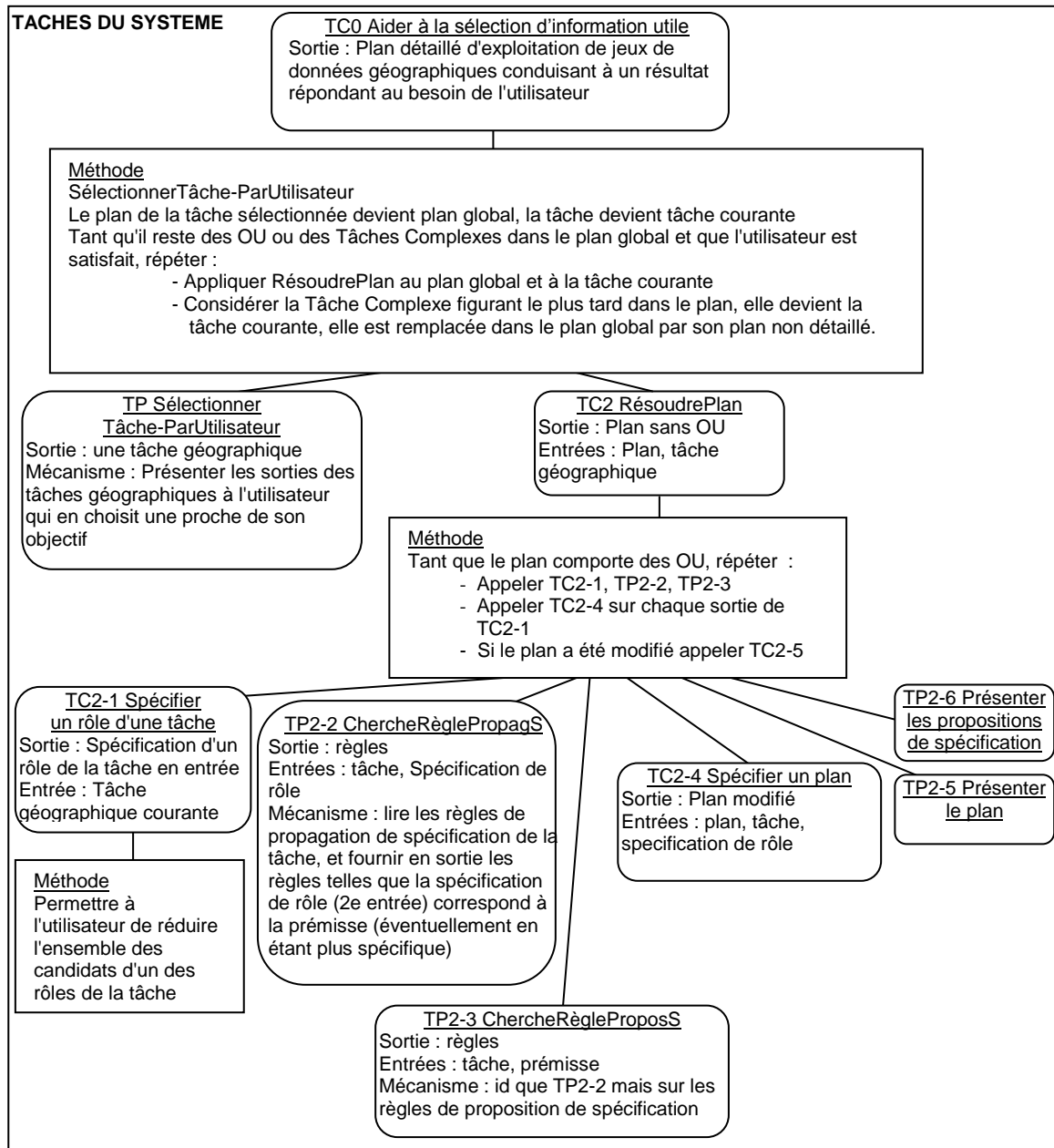


Figure 56 : Exploitation du modèle de tâches géographiques pour l'aide à la sélection d'information utile. [Figure 56 : Exploitation of the geographic tasks model to enhance user's selection of useful information.]

Nous avons proposé un langage de description de tâches dédié à la description de tâches géographiques. Ce langage peut être utilisé pour décrire des tâches du système. Son utilisation a alors pour objectif d'explicitier des tâches du système pour les réaliser ensuite. En particulier nous ne faisons pas figurer des "OU" dans les plans non détaillés des tâches décrites. Pour réaliser une tâche du système décrite dans notre langage, il faut suivre le plan non détaillé et dès qu'un rôle est produit regarder dans la méthode si cela conduit à la spécification d'autres rôles.

Dans la description de tâches du système, les termes décrits comme <Elt du Domaine> peuvent être des éléments intervenant dans la description de tâches géographiques non pas comme éléments du domaine mais comme <Tâche>, <Méthode>, <Rôle>, et autres termes. Une description de tâche géographique est décrite comme un <EltDomaine> du système de la façon suivante :

Tâche géo (**propriétés** : "sortie" **Int**(Rôle géo) ;
 "entrées" **Int**(Rôle géo [{;Rôle géo}])
 "méthode" **Int**(Méthode))

Une méthode géographique est définie de façon similaire en prenant les mots utilisés dans la grammaire BNF pour décrire ces champs et en définissant pour chacun une propriété. Pour décrire les règles nous explicitons les propriétés "prémisse" et "conclusion".

RègleSpecif géo (**propriétés** : "prémisse" **Int**(SpécificationRole géo);
 "conclusion" **Int**(SpécificationRole géo [{;SpécificationRole géo}]))
 RègleStratégie géo (**propriétés** : "prémisse" **Int**(SpécificationRole géo);
 "conclusion" **Int**(ActionPlan [{;ActionPlan}]))
 SpécificationRole géo (**propriétés** : "rôle" **Int**(Rôle géo);
 "préfixe" **Int**(Préfixe);
 "ensemble" **Int**(Ensemble géo))

Une structure de tâches géographiques est définie de la même façon comme un <Elt du Domaine>:

Structure de tâches géo (**propriétés** : "nom" **Int**(Nom);
 "vocabulaire" **Int**(SpécificationRole géo);
 "terme enchainement" **Int**(Terme[{;Terme}]);
 "sousStructures" **Int**(structure de tâches géo[{;structure de tâches géo}])
 "sousTâches" **Int**(tâches géo[{;tâches géo}]))

Dans la suite, nous détaillons les tâches complexes du système Figure 56 mais nous ne détaillons pas les tâches primitives.

TC TC0 (Aider à la sélection d'information utile)

sortie : **Int**(plan géo détaillé)

méthode : **Propag spécif** : **Spécification de** "sortie" de SélectionnerTâcheParUtilisateur → **Spécification de** plan géo global (= "plan non détaillé" **de** **Elt** ("sortie" de SélectionnerTâcheParUtilisateur)), **Spécification de** tâche géo courante (= "sortie" **de** SélectionnerTâcheParUtilisateur) ;

Spécification de "sortie" de SélectTcDsPlan → **Spécification de** tâche géo courante (= "sortie" **de** SélectTcDsPlan) ;

Plan non détaillé : **Vocabulaire** : plan géo global, tâche géo courante

ET(SélectionnerTâcheParUtilisateur/entrée=**Int**(Tâches géographiques);

TANT QUE ((plan géo global **Int**(Plan géographique/terme enchainement **Ext**(OU))) **v**(plan géo global **Int**(Plan géographique/sousTâches **Int**(Tâches primitives géographiques)))

REPETE (**ET** (RésoudrePlan/plan géo=plan géo global/tâche géo =tâche géo courante ;

SélectTcDsPlan/plan géo=plan géo global;

SubstitutDsPlan/plan géo=plan géo global;tâche géo=tâche géo courante))

)

TC RésoudrePlan (TC2 sur Figure 56, Résoudre le plan d'une tâche, i.e. : enlever les choix qui apparaissent)

sortie : plan géo

entrée : plan géo, tâche géo

méthode : **Propag spécif** : **Spécification de "sortie" de SpecifRôlesTâches** → **Spécification de** spécification géo courante (= "sortie" de SpecifRôlesTâches) ;

Spécification de "sortie" de ChercheRèglePropagS → **Spécification de** règle de propagation géo (= "sortie" de ChercheRèglePropagS) ;

Spécification de "sortie" de ChercheRèglePropoS → **Spécification de** règle de proposition (= sortie de ChercheRèglePropoS);

Spécification de "sortie" de SpecifiePlan → **Spécification de** spécification courante(="sortie" de OteUnSingletonA(Ensemble = "candidats" de spécifications propagées)) ;

Spécification de la sortie de AppliqueRègle → **Spécification de** spécifications propagées (= "sortie" de AppliqueRègle)

Plan non détaillé : **Vocabulaire** : spécification géo courante, règle de propagation géo, règle de proposition géo, spécification géo propagée

TANT QUE (plan géo global ⊂ Int(Plan géographique/terme enchainement ⊃ Ext(OU)))

REPETE (**ET** (SpecifRôlesTâches /tâche géo= tâche géo courante ;

ChercheRèglePropagS/tâche géo= tâche géo courante/prémisse = spécification géo courante ;

ChercheRèglePropoS/tâche géo= tâche géo courante/prémisse = spécification géo courante ;

SpecifiePlan /plan géo=plan géo global/tâche géo=tâche géo courante/spécification=spécification géo courante ;

AppliqueRègle/règle=règle de propagation géo;

TANT QUE (spécification propagée géo ≠ ∅);

REPETE (SpecifiePlan /plan géo=plan géo global/tâche géo=tâche géo courante/spécification=spécification géo courante ;);

PresentePlan/plan géo=plan géo global;

Presenterègles/règles=règles géo de proposition;

))

La tâche SpecifRôlesTâche consiste à spécifier la facette déclarative de la tâche géographique courante. Le système aide l'utilisateur en lui fournissant un vocabulaire dans lequel il est capable de préciser la sortie de la tâche en question, voire les entrées. Cette aide renvoie à deux mécanismes différents.

Prenons l'exemple d'un utilisateur dont le besoin est de répondre à la question "où est Paris?". Le système fournit un vocabulaire à l'utilisateur pour spécifier ce qu'est "Paris" : une ville identifiée par le mot « Paris » dans une nomenclature. Cette aide consiste à utiliser un modèle du domaine suffisamment riche pour qu'il puisse servir d'ontologie entre le système et l'utilisateur.

Un autre exemple est celui d'un utilisateur qui ne sait pas préciser ce que veut dire "où est" c'est-à-dire le format de localisation qu'il attend, et qui a déjà spécifié que l'entité qu'il voulait localiser est une ville. Le système peut alors l'aider en interprétant le fait que l'utilisateur recherche la localisation d'une ville et en proposant de remplacer "où est" par "dans quel région?", "dans quel pays?". Cette aide consiste à appliquer une règle :

"l'entité à localiser est un élément d'une nomenclature compositionnelle (la nomenclature administrative) donc une localisation peut consister à donner un élément de plus haut niveau dans la nomenclature". Cette règle est une heuristique.

TC SpécifRôlesTâche (TC2-1 sur Figure 56, Spécifier les rôles d'une tâche)

sortie : Spécification de rôle

entrée : tâche géo

méthode : **Plan non détaillé** : **Vocabulaire** : rôles à spécifier ("entrées" de tâche géo, "sortie" de tâche géo) ; ensemble courant

TANT QUE (rôles à spécifier $\neq\emptyset$) **REPETE** (**ET** (TP(candidats de Ensemble courant = "sortie" de OteUnSingletonA/Ensemble = "candidats" de rôles à spécifier); SpécifieEnsemble/Ensemble= ensemble courant;))

La tâche SpécifieEnsemble se fait de la façon suivante. Si l'ensemble a une extension, la tâche propose de réduire cette extension. Si l'ensemble n'a pas d'extension ou que la proposition a été refusée, la tâche affiche les propriétés de l'élément définissant l'intension de l'ensemble et propose de spécifier les valeurs de ces propriétés. Nous ne la détaillons pas ici pour ne pas alourdir cette section.

TC SpécifiePlan (tâche TC2-4 sur Figure 56, Spécifier un plan géographique)

sortie : plan

entrée : plan, tâche, spécification de rôle

méthode : **Propag Spécif** : **Spécification de** "sortie" de RechercheRègleStratégie → **Spécification de** actions (= "conclusion" de "sortie" de RechercheRègleStratégie);

Plan non détaillé : **Vocabulaire** : actions ; action courante

ET(RechercheRègleStratégie/tâche=tâche/prémisse=spécification de rôle

TANT QUE (actions $\neq\emptyset$)

REPETE (**ET**(

TP(OteUnSingletonA/Ensemble = candidats de actions/sortie= action courante);

AppliqueAction/plan=plan/action=action courante;

))

Pour chacune des fonctionnalités prévoyant une contribution de l'utilisateur, il est nécessaire de définir des interfaces pour permettre la communication entre l'utilisateur et le système. Nous verrons ces éléments lors de la présentation du prototype, section 3.3.

3.2 Spécialisation et instanciation du modèle

Nous utilisons dans cette section le modèle de tâches, mis en place dans la section précédente, pour modéliser des connaissances de raisonnements géographiques particulières. L'enjeu de cette étape est d'évaluer la capacité de notre modèle à décrire, distinguer et relier des catégories *QUOI* et *COMMENT* parmi les connaissances touchant à l'information géographique.

L'opérationnalisation de ce modèle sera détaillée dans la section 3.3 qui présente le prototype du système expert exploitant ce modèle de connaissances. Toutefois, lorsque nous introduisons de nouveaux éléments du modèle, nous détaillons comment ils peuvent être exploités par un système pour aider l'accès d'utilisateurs.

3.2.1 Spécification du domaine géographique

De façon générale, la construction d'un modèle d'expertise de tâches ne se fait pas en modélisant un à un les divers niveaux : tâches complexes, tâches primitives, domaine. Il y a un va et vient entre les modélisations de chaque niveau. Quand on cherche à construire le modèle d'une tâche de haut niveau, comme la configuration d'un système informatique ou d'un emploi du temps, le niveau qui initie la modélisation est le niveau des tâches. Dans notre cas, le modèle construit n'est pas dédié à une tâche spécifique mais à un domaine spécifique. Aussi, avant de modéliser des tâches géographiques, nous nous attachons à rendre compte des spécificités de ce domaine.

Rappelons que le domaine géographique est une représentation du *QUOI* géographique. Les spécificités du *QUOI* géographique ont été montrées lors de la définition de la problématique, 1.1.1, et lors de l'état de l'art sur les formalisations, 2.2.3 et 2.2.2. Nous avons vu que les données géographiques ne sont pas une copie de la réalité, mais la trace des diverses activités de représentations –au sens large-. En particulier, elles se distinguent de la réalité par divers choix rarement explicités dans la BDG. Ces choix peuvent être liés au contexte de la représentation : situation dans l'espace, capacité d'observation et activité du sujet construisant la représentation. D'autres choix découlent de la difficulté de formaliser l'espace géographique : définir des limites d'objets dans un espace continu, discrétiser des formes géométriques pour pouvoir les numériser. Enfin, il existe des interactions complexes, et implicites, entre les diverses activités d'appréhension et représentation de l'espace géographique, Figure 37. Ces interactions impliquent la propagation d'un choix d'une représentation dans une autre.

Il est essentiel que le domaine de notre modèle de tâches géographiques respecte la diversité des représentations et explicite ce qui permet de comprendre une représentation. Cela implique :

- De décrire le contexte d'une représentation, les facteurs permettant de comprendre les choix de cette représentation,

- De fournir la possibilité de retrouver les liens éventuels entre diverses représentations, généralement les traces de dépendances entre les activités de représentation.

Répondre à ces exigences est un domaine de recherche complexe et dépasse le cadre de notre travail. Nous nous limitons à éviter d'aller à l'encontre de ces deux principes sur les cas que nous avons étudiés.

Nous introduisons dans cette section les premières spécifications d'éléments, appartenant au domaine de notre modèle des raisonnements géographiques, et illustrant les principes énoncés ci-dessus.

3.2.1.1 Rendre compte du contexte de l'utilisateur

Une première spécificité du Quoi géographique est la diversité des représentations que se construisent des utilisateurs à propos de l'espace géographique. Ces représentations sont généralement différentes de celles utilisées dans les BDG. Nous pensons qu'il est pertinent de décrire certains éléments du contexte du sujet construisant la représentation, la situation de l'homme dans l'espace géographique, pour rendre compte de spécificités de la représentation construite.

Nous adoptons une représentation « mécanique » pour représenter l'homme dans l'espace géographique. En effet les principes de la mécanique ont pour objectif de décrire l'espace en le fragmentant en systèmes et de décrire les interactions de ces systèmes entre eux. Cela semble adapté à notre contexte. De plus ces principes sont presque aussi largement adoptés que les modèles mathématiques.

Les éléments de notre modèle, décrit Figure 57, sont les suivants.

L'utilisateur est une entité particulière. Une entité a un comportement décrit par des états, et des changements d'états. L'état peut être absolu, par exemple un véhicule dont le moteur est cassé, une rivière asséchée. Il peut aussi être relatif comme la localisation dans l'espace ou une relation avec une autre entité.

On définit une activité, comme voyager, vivre ou faire ses courses, par un ensemble de comportements composant cette activité. Une activité est associée à une échelle pertinente pour analyser cette activité.

Une activité est soumise à des contraintes, pesant sur les comportements (états ou changements d'état) composant cette activité. Parmi ces contraintes, certaines sont des lois universelles et d'autres sont des conventions. Une loi universelle est par exemple qu'un corps se déplace de façon continue dans l'espace en fonction du temps, qu'une voiture se déplace sur le réseau routier, qu'un bateau se déplace sur l'eau. Des exemples de lois conventionnelles sont les restrictions de circulation.

Ces éléments permettent de représenter des connaissances sur des comportements et des activités typiques d'entités typiques, comme le déplacement d'une personne ou d'un véhicule, la répartition de végétaux ou de minéraux dans l'espace.

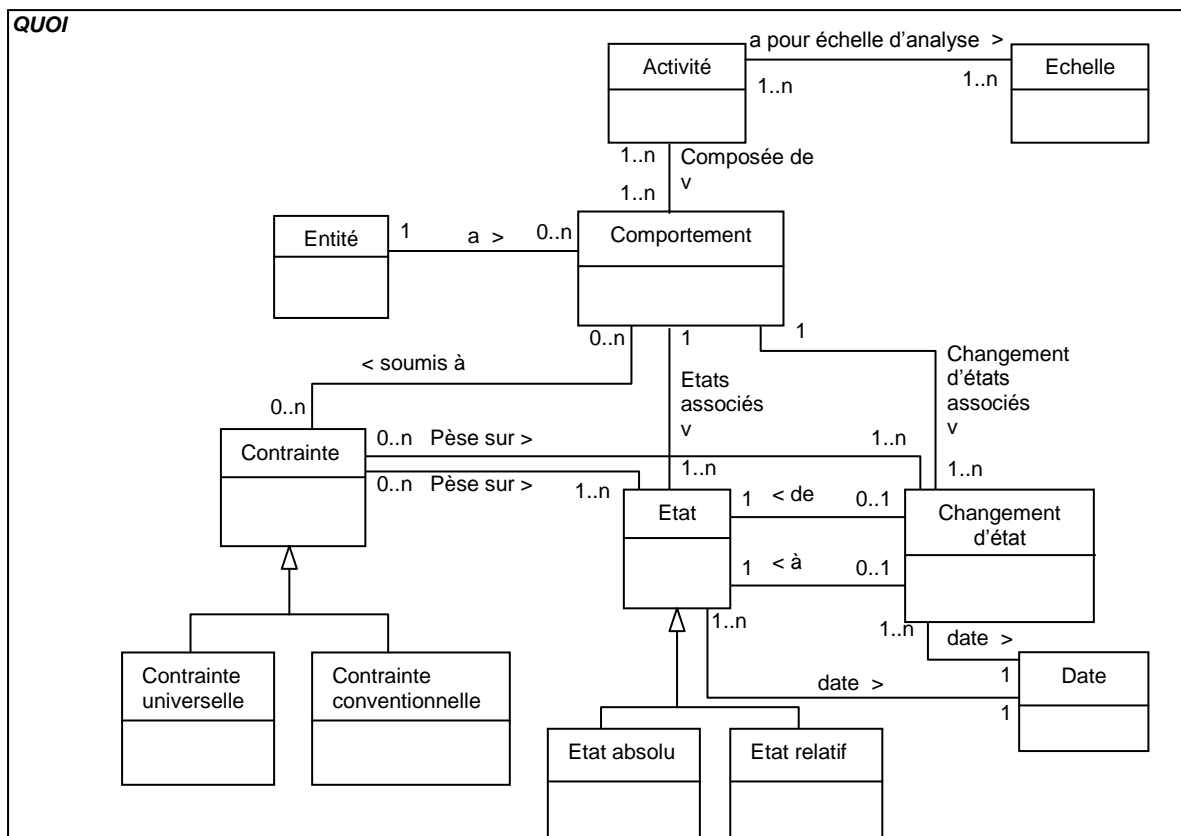


Figure 57 : Eléments du domaine géographique introduits pour expliciter ce qui influe les représentations construites par l'homme de l'espace géographique qui l'entoure. [Figure 57 : Elements of the geographic domain specified to represent factors interfering the process of a man representing the geographic space around him.]

Ces connaissances permettent au système de replacer un besoin d'utilisateur dans son contexte grâce aux connaissances qu'il possède à propos de l'activité conduisant l'utilisateur à exploiter les données géographiques. Il peut dès lors savoir quelle information géographique doit posséder l'utilisateur en considérant les contraintes pesant sur cette activité et les objets nécessaires pour vérifier ces contraintes et les optimiser.

C'est typiquement le cas d'une aide à la navigation. Si nous naviguons sans contrainte géographique, c'est-à-dire si nous pouvions nous « télétransporter », nous n'aurions pas besoin de carte pour déterminer nos itinéraires. Le système peut aussi posséder les connaissances permettant de déterminer l'échelle pertinente d'analyse du besoin de l'utilisateur.

3.2.1.2 Intégrer les deux grandes catégories de représentations de l'espace géographique

Il y a deux grandes représentations de l'espace, que ce soit chez les utilisateurs ou dans les BDG. Il est important de ne pas arbitrer entre ces représentations. Il est possible que, dans le contexte d'un besoin spécifique, le contenu en information recherché par l'utilisateur ne soit pas disponible dans son type de représentation mais soit disponible dans l'autre type.

La représentation par objets renvoie à l'observation d'entités. Elle mène à la construction de BDG vecteur.

La représentation par champs renvoie à l'observation de portions d'espace. Un champ est une fonction dont la variable est un type de portion élémentaire d'espace. Cette représentation mène à la construction de BDG maillées.

Pour ne pas privilégier une représentation sur l'autre, c'est-à-dire une observation sur l'autre, lorsque nous détaillons des éléments attachés à la notion d'entité, nous cherchons si ces éléments sont également attachables à la notion de lieu. Les notions précédentes d'activité et de comportement, Figure 57, s'étendent par exemple aussi au cas du lieu comme représenté Figure 58. Un exemple de loi universelle observée par l'état d'un lieu est que deux objets matériels ne sont pas au même endroit en même temps. Notons que cette loi est vraie dans le monde réel, c'est-à-dire le monde dans lequel l'utilisateur effectue son activité, mais elle n'est pas vraie dans le monde des BDG. En effet, lorsqu'une base est référencée dans un système de dimension 2, les objets étant à la verticale l'un de l'autre dans le monde réel occupent ensemble un même lieu dans la BDG, comme une portion de pont et une portion de rivière.

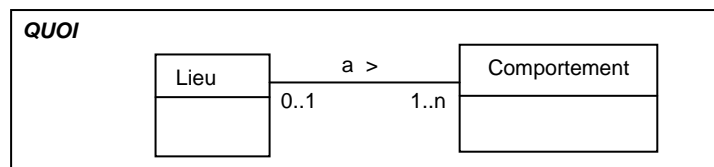


Figure 58 : Un lieu a un comportement soumis à des contraintes, de même qu'une entité. [Figure 58 : A place is associated to a behaviour.]

Par ailleurs, il est important d'explicitier les liens entre le lieu et l'entité. Le principal lien est la notion de situation dans l'espace. Rappelons qu'elle est au cœur des raisonnements géographiques et que c'est une notion très complexe à représenter, 1.1.1.2. Nous proposons une représentation de la situation dans l'espace lorsque nous analysons la tâche de localisation dans la section 3.2.4.1.

Nous reprenons par ailleurs les grandes lignes du modèle de [Laurini et Gordillo 00] pour corréliser les représentations informatiques par objets et celles par champ. Par soucis de généralité, nous introduisons un nouvel élément : la grandeur géographique, qui caractérise un objet et peut avoir pour valeur un champ continu ou des valeurs de ce champ. Par exemple, un objet ville peut être caractérisé par sa température dont les valeurs sont des mesures de ce champ effectuées à des points contenus dans la ville. Un objet fleuve peut être caractérisé par son débit, champ continu qui a pour domaine l'espace occupé par le fleuve.

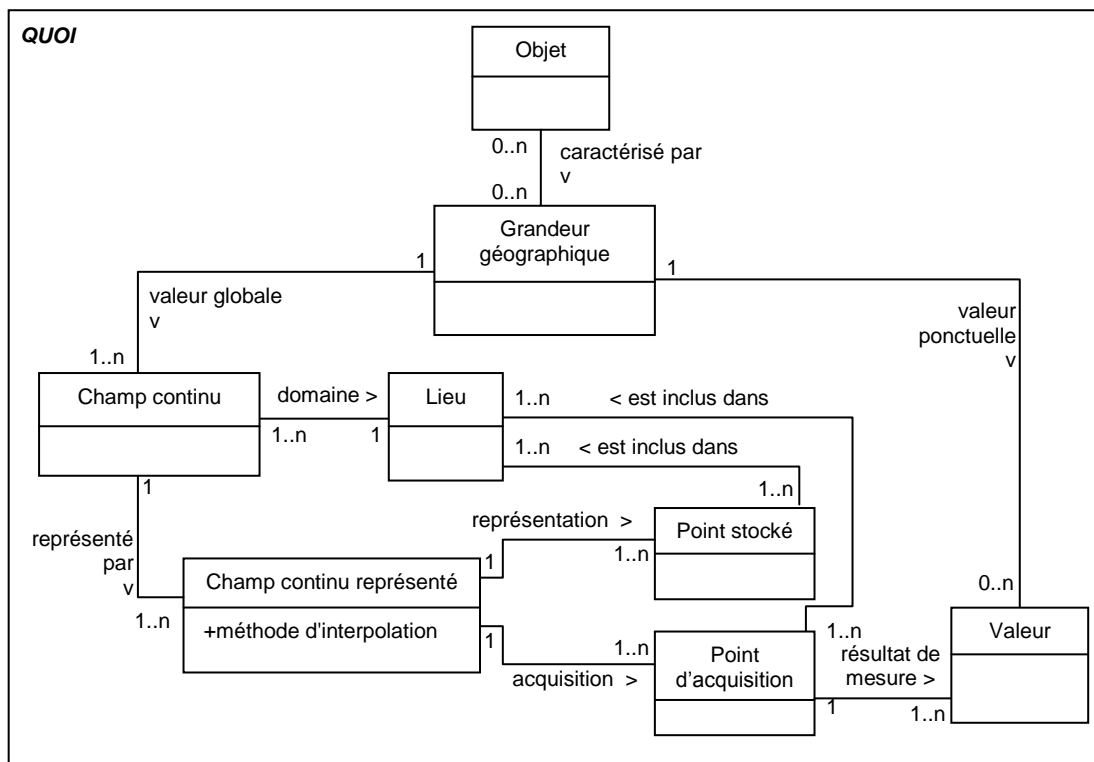


Figure 59 : Eléments de structuration du domaine géographique visant à relier une représentation par champ à une représentation par objets. [Figure 59 : Elements of geographic domain aiming at linking a field representation and an object representation.]

L'introduction de ces éléments dans notre modèle a pour objectif de permettre à un utilisateur, raisonnant sur un modèle représentant l'espace géographique par champs, de savoir quelles données géographiques structurées en objets –données vecteur- peuvent lui apporter des informations utiles. Et vice versa. Pour cela, le système peut s'appuyer sur une base de connaissances comportant les principales grandeurs géographiques et les objets créés en discrétisant les valeurs de cette grandeur. Par exemple, la grandeur occupation du sol est associée aux objets forêt, zone urbaine.

3.2.1.3 Décrire les lots de données disponibles

Enfin, le domaine doit représenter les jeux de données en s'appuyant sur des descriptions disponibles et non en proposant une nouvelle description. En d'autres termes, il doit comporter des métadonnées classiques.

Notons qu'il est important que la notion de métadonnée ne soit pas une catégorie d'élément mais plutôt une relation entre un élément et un autre élément pour lequel il est métadonnée, Figure 60. Par exemple, dans le projet européen LaClef, la base administrative pan-européenne Seamless Administrative Boundaries for Europe, SABE, est utilisée comme métadonnée d'information de référence spatiale d'autres bases de données géographiques [LaClef 01].

Nous avons souligné, section 1.1.3.1, que les métadonnées portaient trop souvent sur un regroupement de données –une base ou un jeu transféré- et non sur des portions qui peuvent être acquises à l'intérieur de ce regroupement. Pour qu'elles remplissent un rôle de *poignées* pour aller chercher des lots de données à l'intérieur d'un regroupement, nous associons à la relation « métadonnée de » une structure d'inférences spécifique. Cette structure d'inférences a pour entrée globale une relation de métadonnée comme « le schéma S est métadonnée du

jeu de données J ». Elle restreint cette relation en : « le schéma S' est métadonnée du jeu de données J' », où S' est un sous-schéma de S.

La relation « restreint » peut exister entre un sous-schéma et un schéma, entre une couverture et une autre couverture dans laquelle la première est incluse.

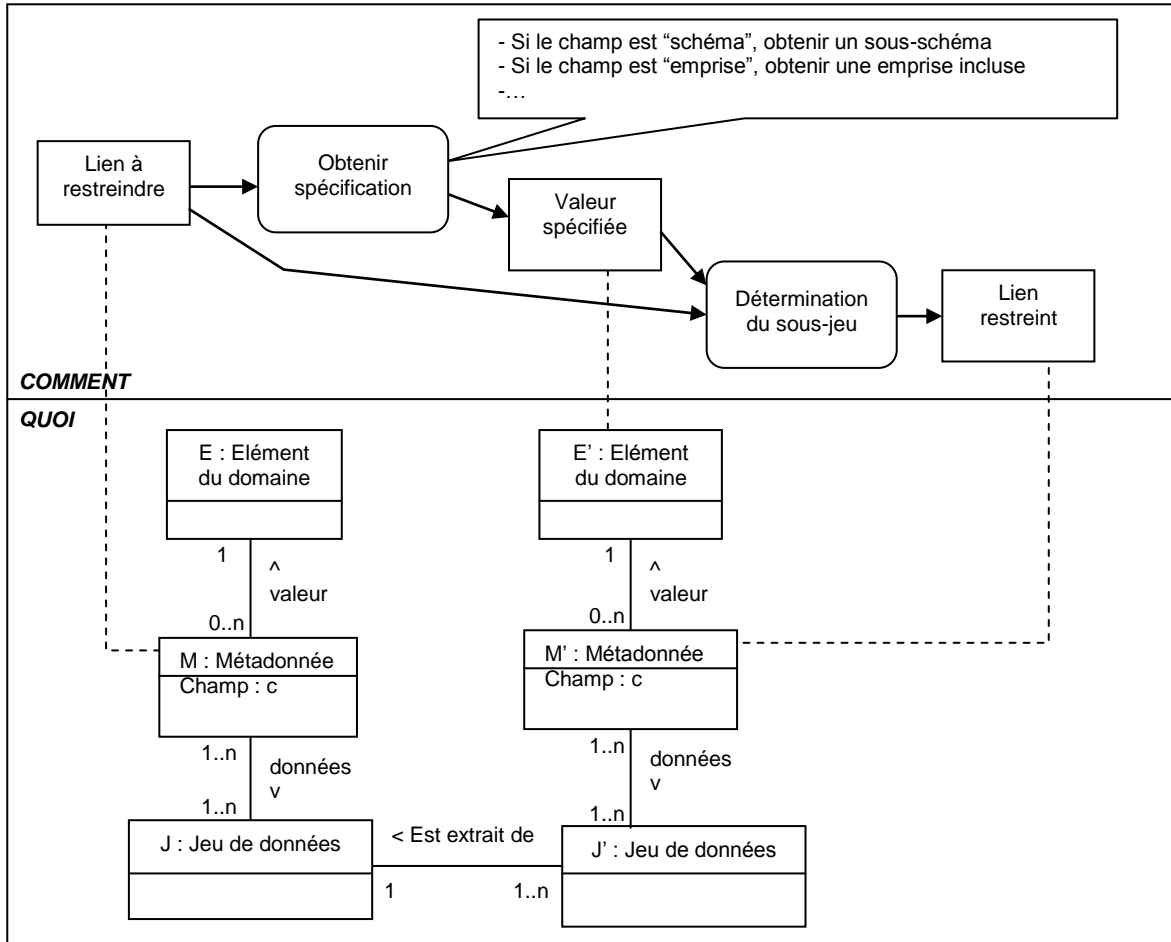


Figure 60 : Association d'inférences aux métadonnées, permettant d'exploiter les métadonnées pour sélectionner des lots de données à l'intérieur d'un jeu. [Figure 60 : Inferences linked to metadata, allowing to use metadata to select data inside a data set.]

3.2.2 Méthode d'analyse de tâches géographiques

Après avoir spécifié le domaine dans la section précédente, nous cherchons dans cette section à formaliser des raisonnements fondamentalement géographiques à l'aide des notions de tâches, méthode, rôle, domaine.

Méthode d'analyse d'une tâche

L'analyse est composée d'une analyse intentionnelle de la tâche, c'est-à-dire axée sur l'explicitation d'un objectif, et d'une analyse fonctionnelle, c'est-à-dire axée sur l'explicitation d'une méthode de résolution.

Dans l'analyse intentionnelle, nous nous concentrons sur l'analyse du vocabulaire de la tâche et sur la distinction de :

- ce qui décrit le résultat attendu (la sortie) et ce qui décrit un résultat obtenu (la sortie d'un plan détaillé possible),
- ce qui est un terme de contrôle important pour spécifier la méthode choisie (les entrées).

Les entrées et sorties d'une tâche renvoient aux éléments qui varient en fonction des contextes possibles de réalisation de la tâche et qui sont pertinents dans la description de la tâche; la valeur de ces éléments à des conséquences sur le plan détaillé proposé à l'utilisateur.

Outre cette distinction, un choix important consiste à savoir, pour chaque terme, s'il renvoie à un rôle ou à un élément du domaine. La limite entre les deux n'est pas franche. C'est sur elle que repose pourtant d'une part la distinction entre le *QUOI* et le *COMMENT*, et d'autre part le lien entre ces deux catégories de connaissances. La description courante d'une application ne respecte pas toujours cette limite. Nous décrivons parfois trop spécifiquement une application et le contenu effectivement manipulé par cette application, c'est-à-dire des objets particuliers du domaine. Une difficulté de la modélisation consiste donc à identifier et nommer des termes qui sont spécifiques à la résolution des applications étudiées.

Pour chaque rôle ainsi défini, il faut ensuite déterminer ses valeurs possibles c'est-à-dire les éléments du domaine désignés par ce rôle. En particulier, il est important de distinguer les valeurs dans le *QUOI* des hommes des valeurs dans le *QUOI* des systèmes.

Dans l'analyse fonctionnelle, nous concentrons sur l'analyse des étapes de la détermination et des tâches primitives correspondant à la décomposition totale de la tâche. Il est alors particulièrement important d'identifier les entrées et sorties des tâches primitives.

Lors de la modélisation d'une tâche, et plus particulièrement des connaissances de spécification de cette tâche, il est extrêmement important de savoir si ces connaissances de spécification se rattachent à la tâche principale ou à une tâche figurant dans le plan. Cette distinction est très importante pour la flexibilité de la base de connaissance et plus exactement :

- la réutilisation de tâches pour décrire le plan de nouvelles tâches, cela demande qu'une tâche décrite puisse s'adapter à une variété de contextes d'invocation.
- la modification de la méthode d'une tâche qui figure dans plusieurs plans, sans pour autant modifier ces plans dans lesquels elle figure.

3.2.3 La tâche d'appariement

Nous analysons dans un premier temps la tâche d'appariement. C'est une tâche pour laquelle nous possédons des connaissances des diverses catégories de notre modèle. L'analyse de cette tâche est essentiellement fonctionnelle.

Nous utilisons dans cette section notre langage de description de tâches pour décrire la tâche d'appariement. Cette description n'est pas destinée à permettre la réalisation effective d'une tâche d'appariement, par exemple à automatiser cette tâche. Elle doit être exploitée pour permettre d'intégrer les méthodes et outils de l'appariement dans les plans détaillés proposés aux utilisateurs, lorsque ces connaissances de dérivation sont utiles pour répondre à leur besoin.

3.2.3.1 Vocabulaire général

L'appariement consiste à établir des correspondances entre les données de deux jeux de données. Une correspondance entre données signifie qu'elles représentent le même phénomène du monde réel, comme un bâtiment, un nœud routier. Le processus d'appariement a été exploité et perfectionné dans de nombreux travaux du laboratoire COGIT. [Lemarié 96] propose un état de l'art des mécanismes connus d'appariement géométrique de données. [Badard 00] améliore le processus général d'appariement de deux jeux de données et l'utilise dans le contexte de la mise à jour. Ce processus a ensuite été intégré dans un projet de développement mené à l'IGN, Carto2001 [Lemarié and Badard 01]. Par ailleurs, [Bel Hadj Ali 01] étudie l'appariement d'objets surfaciques, dans un contexte de contrôle qualité.

Pour analyser cette tâche nous n'avons pas repris les descriptions de processus de ces travaux pour les traduire dans notre formalisme. Nous avons préféré expliciter la connaissance d'un expert en nous appuyant sur le modèle générique des tâches géographiques, en l'occurrence nous avons travaillé avec Cécile Lemarié.

Contextes

Avant d'analyser la tâche d'appariement nous rappelons dans quel contexte elle prend place. Cette tâche n'est pas une fin en soi, elle est toujours invoquée dans le contexte d'une autre tâche.

C'est une sous-tâche de la tâche de comparaison d'un jeu de données à un autre jeu. L'appariement prend place au début de la tâche de comparaison pour définir ce qui peut être comparé à quoi. A l'issue de cette étape, la comparaison des jeux se fait en comparant, correspondance par correspondance, la représentation d'un phénomène d'un jeu à l'autre.

La comparaison prend elle-même place dans divers contextes. Il peut s'agir de comparer deux jeux, portant sur une même réalité mais issus de bases différentes, pour contrôler la qualité d'une base à l'aide d'une base considérée comme de meilleure qualité, comme dans les travaux de [Bel Hadj Ali 01]. Il peut aussi s'agir de comparer deux jeux portant sur une même réalité et issus de différentes version d'une même base pour détecter des évolutions de la réalité représentée dans la base, comme dans les travaux de [Badard 00].

L'appariement est aussi une sous-tâche de la tâche de correction d'un jeu de données J à partir de la description des corrections d'un jeu de données référence JR. C'est dans ce contexte que l'appariement est étudié dans les travaux de [Badard 00] et intégré dans un projet de développement [Lemarié and Badard 01]. Cette tâche est appelée intégration des corrections décrites sur un jeu dans un autre jeu. La description des corrections est appelée "différentiel", ou encore journal des corrections entre le jeu de référence non corrigé et le jeu de référence corrigé. Elle associe à des données du jeu de référence représentant un phénomène l'évolution de la représentation du phénomène dans la base : cela peut être une destruction, un déplacement, une modification d'attribut.

L'intégration des corrections dans le jeu J nécessite une première étape d'appariement entre le jeu de référence non corrigé et le jeu J. La description du différentiel, associée aux liens de correspondance, sert de clé d'accès aux données du jeu J devant être corrigées et leur associe les corrections qu'elles doivent subir.

Nous détaillons Figure 61 le cas de la mise à jour d'une base de données à l'aide d'une autre qui utilise toutes ces tâches.

Résultats attendu et obtenu

Cette première description générale de la tâche d'appariement introduit un vocabulaire spécifique. Nous reprenons ce vocabulaire en l'associant à ses valeurs dans le domaine géographique : la Figure 62 décrit les rôles ainsi définis.

Cette tâche n'est pas symétrique et nous parlons d'appariement d'un jeu de référence et d'un jeu comparé.

La sortie est un ensemble de correspondances entre les données de chaque jeu. Le plan détaillé décrit comment produire de telles correspondances. Une correspondance de données se présente comme une association entre des objets du premier jeu et des objets du deuxième jeu. Cette association n'est pas symétrique puisqu'elle est le résultat d'un mécanisme non symétrique. Nous désignons par « amont » de la correspondance les données du jeu de référence et par « aval » les données du jeu comparé. L'attribut cardinalité de la correspondance se présente comme un couple (i,j) où i désigne le nombre d'objets du jeu référence participant à la correspondance et j le nombre d'objets du jeu comparé. Une correspondance de cardinalité $(1,0)$ désigne un objet du jeu de référence qui n'a pas de correspondant dans le jeu comparé. Nous utilisons le terme correspondance bancaire pour désigner une correspondance où i ou j vaut 0. Un objet participant à une correspondance non bancaire est dit apparié.

Notons aussi qu'une correspondance peut être établie entre n objets du jeu de référence et m objets du jeu comparé.

Méthode générale

La méthode générale de la tâche d'appariement repose sur la notion de variation de données. Il s'agit de l'écart entre les représentations. L'idée générale est que si deux ensembles de données représentent un même phénomène du monde réel alors la taille de cet objet dans une représentation doit être proche de la taille du même objet dans l'autre représentation, de même que sa position par exemple. Une variation est quantifiée en s'appuyant sur des caractéristiques du phénomène dont les représentations dans l'un et l'autre jeu peuvent être mesurées et comparées. De telles caractéristiques sont généralement la forme, la position, la description sémantique.

L'appariement s'appuie sur une analyse de variations entre les représentations pour établir les correspondances aux "points" où ces variations sont minimales. La notion de "points" renvoie ici aux objets : une correspondance est établie entre un ou plusieurs objets d'une représentation et un ou plusieurs objets d'une autre représentation.

La notion de "minimisation" est complexe. Il n'est pas simple de comparer les variations entre elles pour sélectionner les plus faibles : écarts entre les définitions de catégories et de relations entre catégories, écarts entre les représentations de géométrie, écarts entre les localisations, etc.

Le corps de la méthode d'appariement consiste en l'exploration de l'ensemble des correspondances candidates (i.e. paire de lot d'objets du jeu de référence et lot d'objets du jeu comparé) combinée à l'évaluation des variations de ces correspondances. Globalement, cette méthode produit puis restreint des sous-ensembles de correspondances candidates qui minimisent les variations de caractéristiques particulières : celles qui permettent le mieux de discriminer des objets d'une représentation entre eux et celles qui sont censées varier le moins d'une représentation à l'autre.

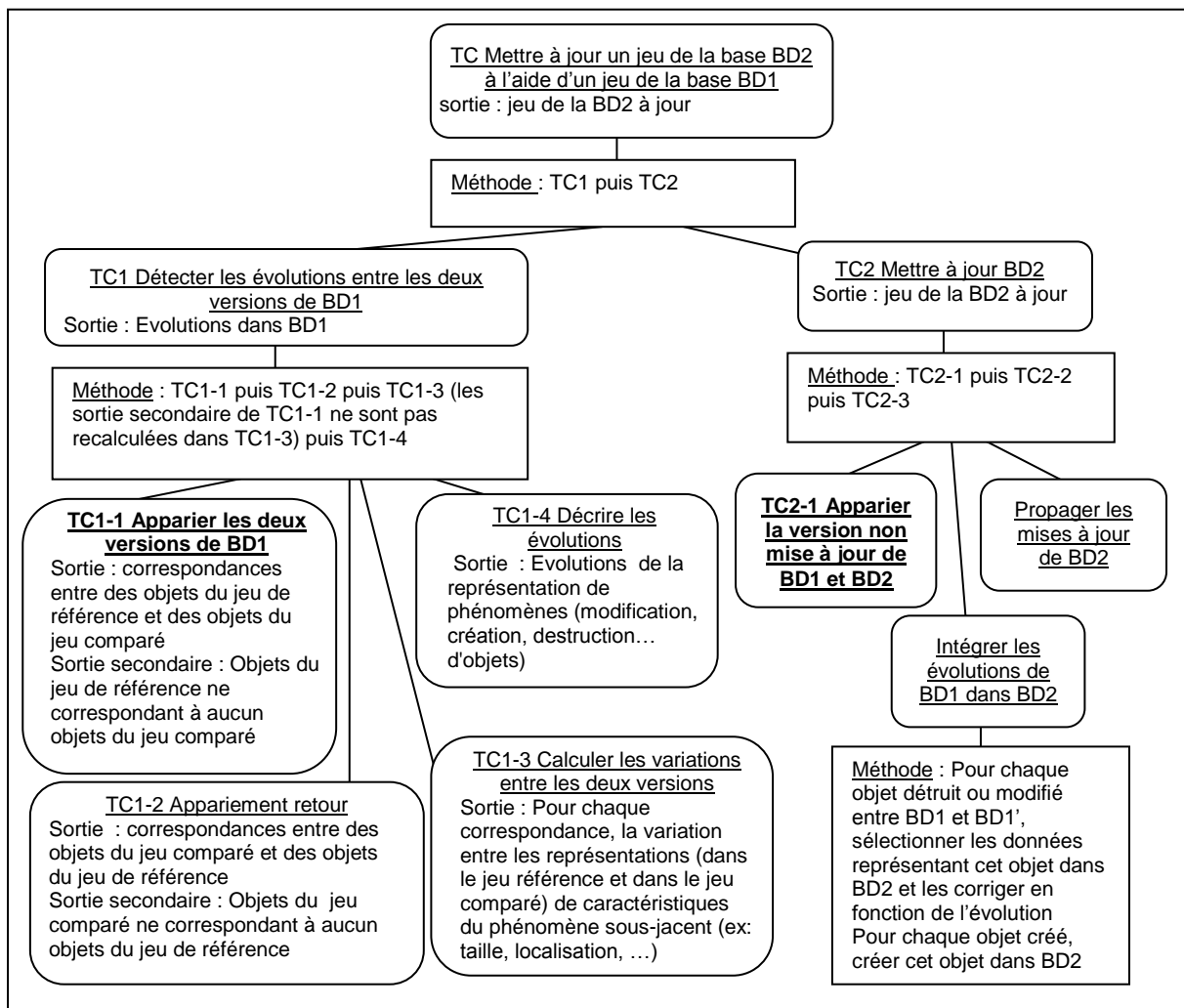


Figure 61 : Structure générale d'une tâche de haut niveau, « mettre à jour une base à l'aide d'une autre », qui s'appuie sur des appariements (les tâches en gras). [Figure 61 : General structure of a high level task, « to update a DB with the help of another », which relies on data-matching (bold titled tasks).]

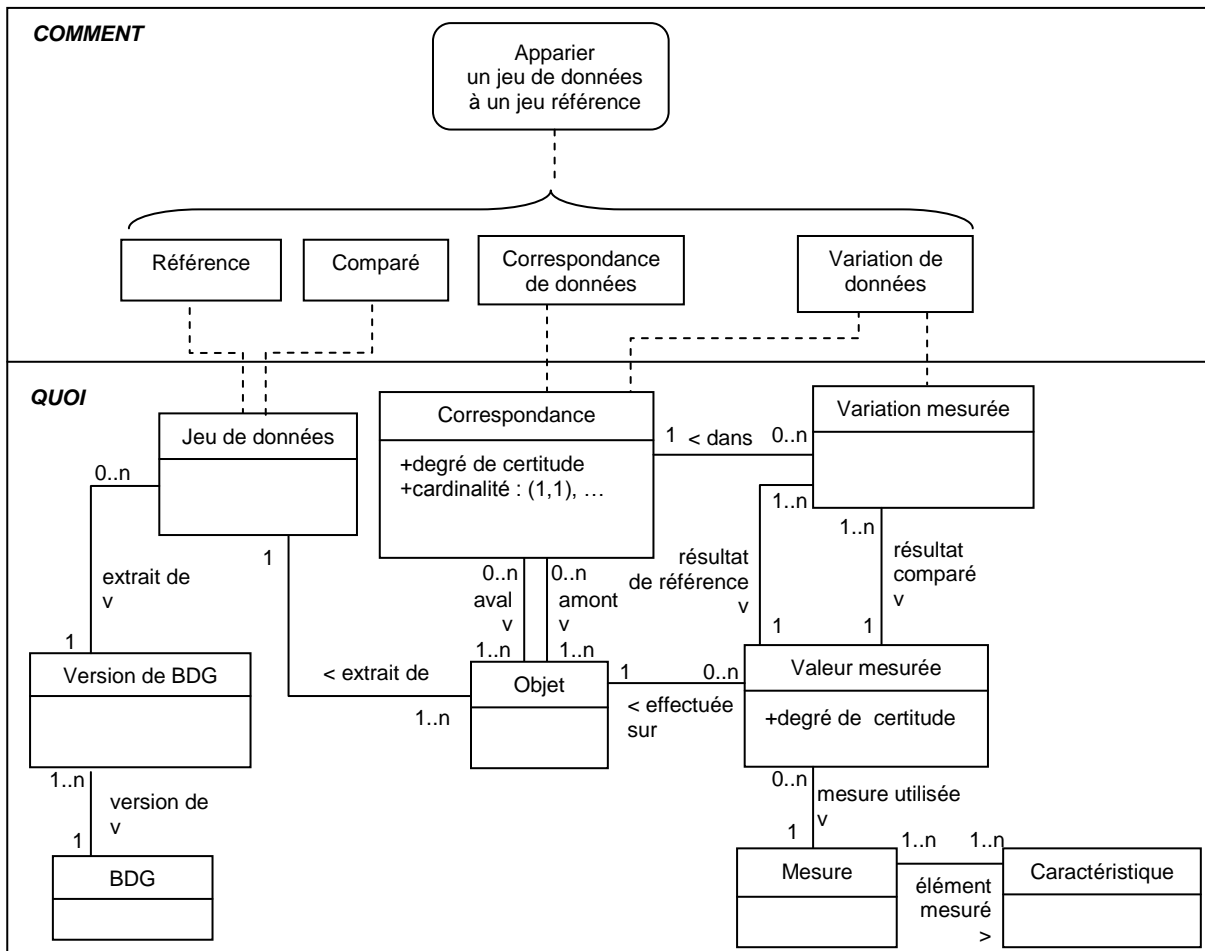


Figure 62 : Principaux rôles de la tâche d'appariement, et leurs valeurs. [Figure 62 : Main roles in the data-matching task, and their values.]

3.2.3.2 Détail de la méthode

Nous détaillons maintenant la méthode utilisée pour appairer deux jeux de données, Figure 64. Nous nous contentons de détailler l'établissement de correspondances, de type (1,n), entre un jeu de référence et un jeu comparé. C'est-à-dire des correspondances où l'amont appartient au jeu de référence et l'aval appartient au jeu comparé. Cet appariement est suivi d'un « appariement retour » qui analyse les correspondances pour produire les correspondances où l'amont appartient au jeu comparé et l'aval appartient au jeu comparé.

Apparier les schémas

Une première étape consiste à appairer, «à la main», les schémas. Les rôles importants à l'issue de cet appariement sont :

- Les correspondances de classes.
- Les attributs discriminants invariants: Il s'agit des attributs qui existent dans chaque représentation –ils sont mis en correspondance-, tels qu'au sein de chaque représentation ils prennent des valeurs différentes pour des objets différents –ils sont discriminants dans chaque représentation-, et tels que leur valeur est la même

d'une représentation à l'autre –ils sont invariants-. Cela peut être par exemple l'attribut "nom de route" entre deux bases respectant la même nomenclature.

- Les relations invariantes : Il s'agit des relations qui existent dans chaque représentation –elles sont mises en correspondance-, et telles que les instances de ces relations existent indépendamment de la représentation choisie. Autrement dit, l'appariement devient un morphisme entre les jeux de données munis de ces relations. Si des objets du jeu de référence sont liés par cette relation –ou plus précisément cette relation telle qu'elle est représentée dans le jeu de référence- et qu'ils ont des correspondants dans le jeu comparé, alors ces derniers objets sont également liés par cette relation –ou plus précisément cette relation telle qu'elle est représentée dans le jeu comparé-. De telles relations peuvent exister au sein d'une même classe, comme la topologie, et ou entre deux classes, comme la composition. Nous les appelons respectivement relations invariantes intra-classe et relations invariantes inter-classes.

Notons que dans cet "appariement à la main" des schémas, la notion de variation n'est pas utilisée de la même façon qu'elle le sera dans l'appariement de données. Seules les variations nulles, i.e. les invariances, sont prises en compte, les autres ne sont pas étudiées. Ces notions sont décrites Figure 63.

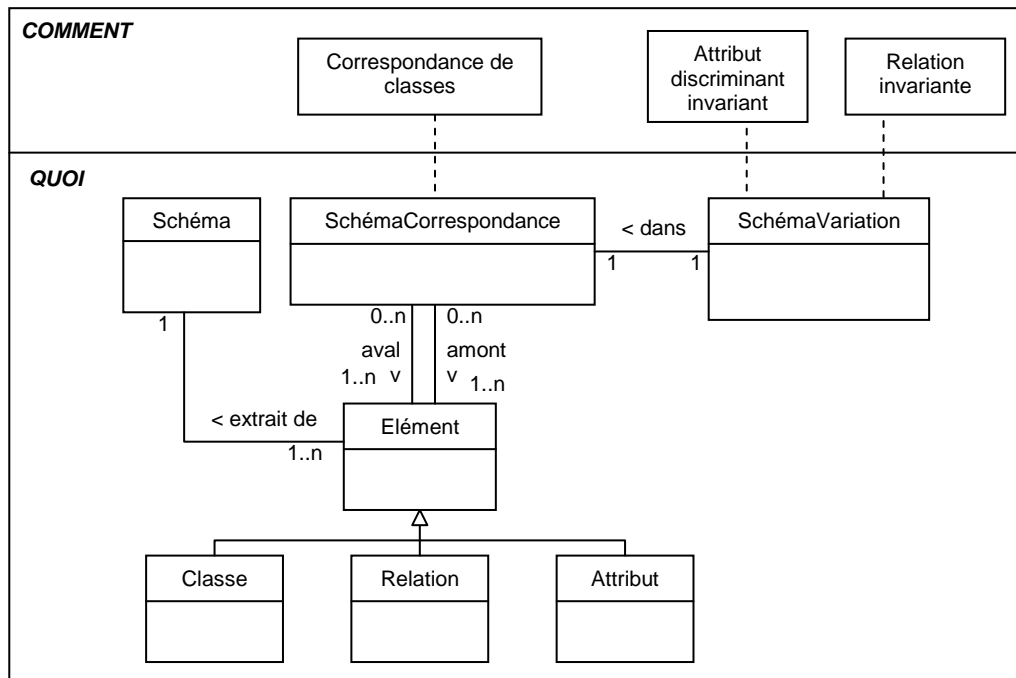


Figure 63 : Rôles importants associés à la comparaison de schémas. [Figure 63 : Important roles associated to schemas comparison.]

Les connaissances de valuation des rôles qui doivent être stockées dans notre système sont par exemple pour chaque BDG, la liste des attributs discriminants "dans l'absolu", la liste des éléments invariants "dans l'absolu".

Planifier l'appariement décomposé des jeux

L'appariement de schéma est utilisé pour décomposer l'appariement de données en décomposant les jeux à appairer selon les classes mises en correspondance : des données ne peuvent être mises en correspondance que si les objets structurant ces données appartiennent à des classes mises en correspondance lors de l'appariement des schémas. L'appariement de données est ainsi décomposé en composantes qui correspondent à des restrictions dans les jeux à appairer.

L'appariement de schéma est également utilisé pour planifier l'appariement des composantes. Cette planification obéit à un souci de fiabiliser et simplifier l'appariement, c'est-à-dire veiller à ce que l'appariement des données produise des correspondances les plus certaines possibles et éviter d'utiliser des processus coûteux (algorithmes géométriques complexes). Les appariements restreints aux composantes ne sont pas indépendants et l'ordre dans lequel ils sont effectués influe sur la complexité du calcul et la fiabilité du résultat. Lorsque des composantes ont été appariées, l'appariement des composantes suivantes peut s'appuyer sur les correspondances établies. Par exemple, si les objets routes ont été appariés, alors l'appariement des tronçons de route s'appuie sur les correspondances entre routes : un tronçon de route est comparé non pas avec tous les tronçons de route du jeu comparé mais plutôt avec les tronçons entrant dans la décomposition de la route appariée avec la route à laquelle lui-même appartient.

Cette planification s'appuie sur des relations d'ordre entre appariement de composantes, que nous appelons relations d'ordre entre correspondances de classes. Notons que ces relations peuvent parfois être concurrentes. Elles sont les suivantes :

- Les classes portant un attribut discriminant invariant sont appariées en premier. En effet chacune de ces composantes s'apparie de façon fiable et simplement, sans recourir aux résultats des appariement d'autres composantes.
- Lorsqu'il existe une relation invariante inter-classes, les correspondances produites lors de l'appariement d'une classe sont utilisées pour simplifier l'appariement de l'autre classe et il vaut mieux appairer en priorité celle dont l'appariement nécessite le moins d'algorithmes complexes ou celle dont le résultat est le plus fiable.
- Si une classe participe à un grand nombre de relations invariantes inter-classes il est intéressant de l'appairer avant les autres classes participant à ces relations.

Il existe également des règles pour évaluer la complexité et la fiabilité de l'appariement d'une composante. La complexité dépend des algorithmes utilisés. Si une classe porte une relation invariante intra-classe alors son processus d'appariement est simplifié en exploitant cette relation, comme nous le verrons par la suite.

La planification se traduit par la production d'un graphe orienté de correspondances de classes. Ce graphe de correspondances de classes est proposé à la validation de l'utilisateur. L'utilisateur peut décider de ne pas appairer certaines classes, dès lors que cela n'a pas de conséquence sur l'appariement d'autres classes.

Appairer les données

La dernière étape de l'appariement consiste à prendre une à une les correspondances de classes, dans l'ordre de parcours du graphe indiqué, et établir les correspondances de données entre objets appartenant aux classes de la correspondance de classes.

Pour chaque correspondance, l'appariement de données se fait selon les étapes suivantes :

- Sélection d'un objet parmi les objets de référence. La sélection se fait en suivant les critères suivants. Les objets portant des valeurs d'attributs discriminant invariant sont sélectionnés en premier. Les objets participant à une relation invariante avec un objet précédemment apparié (une relation intra-classe avec un

- objet de la même classe ou une relation inter-classes avec un objet d'une autre classe déjà apparié) sont sélectionnés ensuite. La sélection se poursuit selon ces critères ou aléatoirement s'ils ne s'appliquent pas.
- Sélection des objets à comparer à cet objet. Rappelons que le jeu comparé est déjà restreint à ce stade aux objets appartenant à la classe appariée à la classe de l'objet de référence. La sélection se fait alors en restreignant encore ce jeu comparé quand cela est possible. C'est possible lorsque l'objet sélectionné est en relation invariante RAmont avec un objet OAmont apparié avec un objet OAval. Dès lors, le jeu à comparer est restreint à l'ensemble des objets du jeu comparé qui sont liés par la relation correspondante RAval avec l'objet Oval.
 - Évaluation et minimisation de la variation de données entre l'objet de référence et chaque objet comparé sélectionné. S'il ne reste qu'un objet dans le jeu restreint il est mis en correspondance avec l'objet sélectionné. S'il reste plusieurs objets une dernière étape est mise en œuvre : l'utilisation d'algorithmes géométriques pour discriminer l'objet dont les caractéristiques de position ou de forme sont les plus proches de celles de l'objet sélectionné.

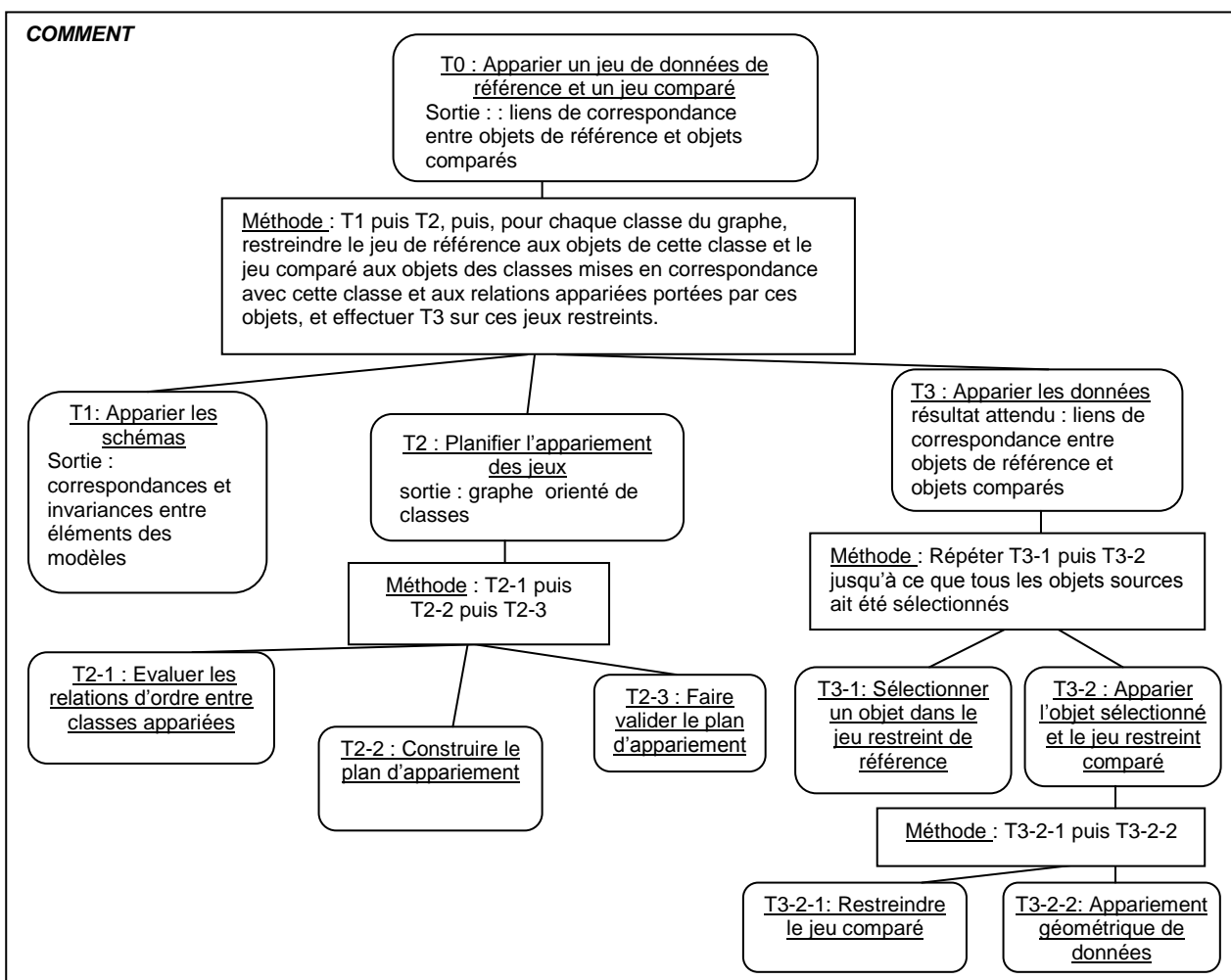


Figure 64 : Décomposition générique de la tâche d'appariement. [Figure 64 : Generic decomposition of the data-matching task.]

Nous utilisons maintenant notre langage de description de tâches pour décrire la tâche d'appariement.

TC T0 (appairer un jeu de données de référence à un jeu comparé)

Sortie : liens de correspondance

Entrée : jeu de référence, jeu comparé

Méthode : **Propag Spécif** : **Spécification de "sortie"** de AppairierSchéma → **Spécification de** correspondances de classes (= "sortie" de TC1), **Spécification de** éléments invariants (= "sortie secondaire" de TC1) ;
Spécification de "sortie" de TC2 → **Spécification de** graphe (= "sortie" de TC2) ;
Spécification de "sortie" de ParcoursGraphe → **Spécification de** classe courante de référence (= "sortie" de ParcoursGraphe) ;
Spécification de classe courante de référence → **Spécification de** classes courantes comparées (= "correspondances de classe" de **Elt.**(classe courante de référence)) ;
Spécification de "sortie" de RestreindreJeu/jeu=jeu de référence → **Spécification de** jeu de référence restreint (= "sortie" de RestreindreJeu/jeu=jeu de référence + éléments appariés de référence) ;
Spécification de "sortie" de RestreindreJeu/jeu=jeu comparé → **Spécification de** jeu comparé restreint (= "sortie" de RestreindreJeu/jeu=jeu comparé + éléments appariés comparés) ;
Spécification de "sortie" de TC3 → **Spécification de** liens de correspondance (+="sortie" de TC3) , **Spécification de** éléments appariés de référence (+="sortie" secondaire éléments appariés référence de TC3), **Spécification de** éléments appariés comparés (+="sortie secondaire" éléments appariés comparés de TC3) ;

Plan non détaillé : **Vocabulaire** : correspondances de classes, éléments invariants, graphe de classes, jeu de référence restreint, jeu comparé restreint, classe courante de référence, classes courantes comparées, éléments appariés de référence, éléments appariés comparés

ET(AppairierSchéma/référence="schéma" de jeu de référence/comparé="schéma" de jeu comparé ;
 TC2/référence= schéma du jeu de référence/ correspondances =correspondances de classes/invariances=éléments invariants;
 ParcoursEtMarqueGraphe/graphe=graphe de classe;
TANT QUE(classe courante de référence≠∅)
REPETER (**ET** (
 RestreindreJeu/jeu=jeu de référence/classes=classe courante de référence;
 RestreindreJeu/ jeu=jeu comparé/classes=classes courantes comparées;
 TC3/ référence=jeu de référence restreint/comparé=jeu comparé restreint/correspondances=liens de correspondances;
 ParcoursEtMarqueGraphe/graphe=graphe de classe;
)))

La tâche T1 d'appariement de schéma se fait pour l'instant en lisant les spécifications. Lorsqu'on travaille sur un nombre limité de BD et de types de jeux, le résultat de cette tâche peut être stocké une fois pour toute au lieu d'être recalculé pour chaque tâche d'appariement.

TC TC2 (Planifier l'appariement de jeux)

Sortie : graphe orienté de classes

Entrée : référence, correspondances, invariances

Méthode : **Plan non détaillé** : **ET** (**TP**(Evalue critères d'ordre entre classes);

TP(Construit relations globales d'ordre entre classes);

TP(Construire le graphe);
TP(Faire valider graphe);
TANT QUE (utilisateur pas satisfait)
REPETER (**TP**(Faire valider graphe) ; **TP**(Modifier graphe)))

TC TC3 (Apparier des données)

Sortie : liens de correspondance entre objets de référence et objets comparés

Entrée : jeu de données de référence, jeu de données comparé, graphe de classes

Méthode : **Plan non détaillé** : **Vocabulaire** : classe courante, jeu restreint de référence, jeu restreint comparé, élément de référence, éléments à comparer

TANT QUE (jeu restreint de référence $\neq\emptyset$)

REPETE(**ET**(

ParcoursEtMarqueGraphe/graphe=graphe de classes/ sortie=classe courante;

FiltreJeu/jeu=jeu de référence/classe filtre=classe courante/sortie= jeu restreint de référence;

FiltreJeu/jeu= jeu comparé, classes filtre= classes en correspondance avec la classe courante/sortie=jeu restreint comparé;

TP(sélectionne un objet du jeu restreint de référence en fonction des critères de sélection)/sortie=élément de référence;

TP(sélectionne des objets du jeu restreint comparé en fonction des correspondances déjà établies et des relations invariantes intra-classe)/sortie=éléments à comparer;

liens de correspondance+= "sortie" de **TC**(Appariement géométrique de l'élément de référence aux éléments à comparer);

))

La tâche primitive `ParcoursEtMarqueGraphe` consiste à parcourir un graphe orienté, à sélectionner le premier nœud non marqué rencontré, et à le marquer pour ne pas le re-sélectionner la fois suivante.

3.2.3.3 Exploitation du modèle de l'appariement

Notre modèle de tâches a permis de formaliser le processus d'appariement. Ce modèle n'est pas dédié à un jeu de données, grâce à l'utilisation de rôles pour décrire les objets manipulés. Il peut donc être adapté à divers contextes en spécifiant les valeurs de ces rôles.

Cette formalisation est intéressante dans un contexte d'aide à l'accès car le processus d'appariement est un processus intervenant dans plusieurs contextes d'exploitation de données géographiques.

Un premier type de besoin est celui où l'utilisateur possède deux jeux et doit réaliser un appariement. Un tel besoin surgit par exemple quand un utilisateur possède deux jeux de données portant sur des zones géographiques différentes et ayant une zone de recouvrement, et qu'il veut réaliser une jointure de ces jeux. Il surgit aussi quand un utilisateur souhaite détecter des évolutions entre deux versions d'une même base. Le modèle permet d'automatiser la construction du graphe de classes d'appariement. Pour cela, il faut stocker une base de connaissances associant à un type d'objet géométrique la complexité algorithmique de la mesure de caractéristiques de cet objet. Le système utilise cette base de connaissances pour évaluer la complexité algorithmique a priori de chaque appariement de classe, puis il utilise la liste d'attributs et relations invariants relative aux jeux sur lesquels sont réalisés l'appariement pour modifier ces complexités et ordonner les classes.

Un autre type de besoin est celui où un utilisateur possède un jeu de données et a besoin d'acquérir un autre jeu auquel apparier son propre jeu. Par exemple il peut vouloir évaluer l'actualité ou la qualité de ses données en appariant son jeu à un jeu de référence dont l'actualité et la qualité sont satisfaisantes.

Le modèle permet de répondre à l'utilisateur si son jeu peut être apparié, avec quel autre jeu, et comment. Le système s'appuie sur ses connaissances sur la tâche d'appariement pour guider la sélection d'un jeu de référence parmi les BDG connues du système. Pour cela il s'appuie sur la spécification par l'utilisateur des attributs discriminants figurant dans son propre jeu et des classes portant de nombreuses relations de composition ou de topologie. Outre désigner un jeu de référence auquel l'utilisateur peut apparier son jeu, le système peut aussi lui décrire une méthode générale d'appariement.

Un autre besoin auquel répond l'appariement est la réalisation d'une représentation multi-échelles à partir de deux bases de données, portant sur une même réalité mais à des échelles différentes. Les objets à mettre en correspondance dépendent des bases initiales et de l'exploitation que l'utilisateur veut faire de la représentation multi-échelles finale. Ce sont par exemple des nœuds routiers qui peuvent être représentés à une échelle par des entités complexes comportant les bretelles d'accès et les ponts, et à une autre échelle par un simple symbole ponctuel positionné à l'intersection des routes.

En définitive, pour aider un utilisateur à spécifier précisément son besoin d'apparier des jeux de données, il faudrait modéliser les contextes dans lesquels l'appariement est utile c'est-à-dire des tâches faisant appel à l'appariement dans leur plan non détaillé.

3.2.4 La tâche de localisation

Nous modélisons à présent des tâches demandant une analyse intentionnelle poussée, c'est-à-dire qui renvoient à un vocabulaire ambigu et à une variété de besoins effectifs. Cela nous permet de nous concentrer sur la notion d'expression de besoins d'utilisateurs. Les interrogations fondamentales présentées en section 1.1.2.1 sont : la localisation « Où est-ce ? », l'adressage « Qu'est-ce qu'il y a là ? », la détection de proximité « Qu'est-ce qu'il y a proche de ça ? », le calcul d'itinéraire. Cette section se concentre sur les deux premières questions et la section suivante sera consacrée à la troisième question.

Nous nous concentrons sur la tâche de localisation car d'un point de vue déclaratif, elle est symétrique de la tâche d'adressage : les termes de lieu et d'entité y jouent des rôles symétriques.

La localisation est une tâche très générique qui renvoie à une variété de contextes. Connaître ces contextes est indispensable pour pouvoir envisager d'aider automatiquement un utilisateur à planifier sa réponse à une question du type « où est ce ? ». Il est donc important d'analyser en priorité l'aspect déclaratif qui renvoie à l'expression du besoin de l'utilisateur, avant d'aborder la planification de la réponse à ce besoin.

3.2.4.1 L'aspect déclaratif

Les principaux rôles

La localisation vise à déterminer le lieu tel qu'une entité donnée *s'y trouve*. La sortie est donc un lieu et une entrée est l'entité. L'adressage vise à déterminer les entités qui *se trouvent* à un lieu donné. La sortie est donc des entités et une entrée est le lieu.

Un terme au cœur de ces tâches et de leur symétrie est : « se trouve à ». Nous notons ce terme localisation spatiale (d'une entité dans un lieu), il correspond à l'expression situation dans l'espace que nous employons pour décrire le domaine de l'information géographique 1.1.1. La localisation spatiale est une notion essentielle à tout raisonnement prenant place dans l'espace géographique. C'est un rôle. Il existe des relations entre localisations spatiales comme l'implication ou l'incompatibilité.

Nous modélisons la localisation spatiale comme un concept, Figure 65, et non comme une relation entre une entité et un lieu.

A partir de ce choix, une structure de rôles très simple s'offre à nous pour décrire la notion de localisation Figure 65.

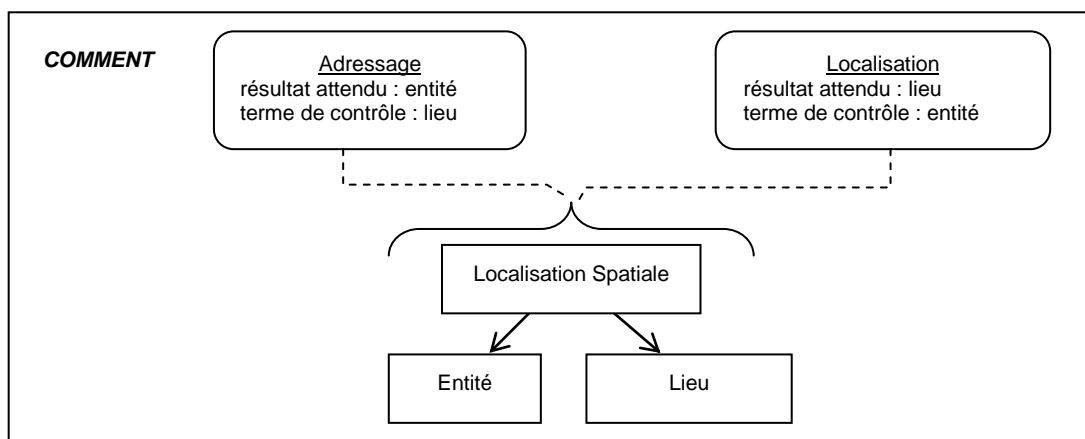


Figure 65 : Rôles servant de vocabulaire aux tâches d'adressage et de localisation. [Figure 65 : Main roles for the task of addressing and locating.]

Les valeurs du rôle « entité »

Les valeurs du rôle « entité » d'une tâche de localisation sont tout ce qui peut faire l'objet d'une tâche de localisation, c'est-à-dire un phénomène géographique : une ville, une manifestation de maladie, une voiture, une tendance culturelle, une personne, un événement historique, etc. Notons que le rôle entité correspond dans le domaine à deux notions Figure 66 : les entités utilisées dans les modèles de raisonnement des hommes (que nous notons « entité ») et les entités informatisées (que nous notons « entité de représentation »).

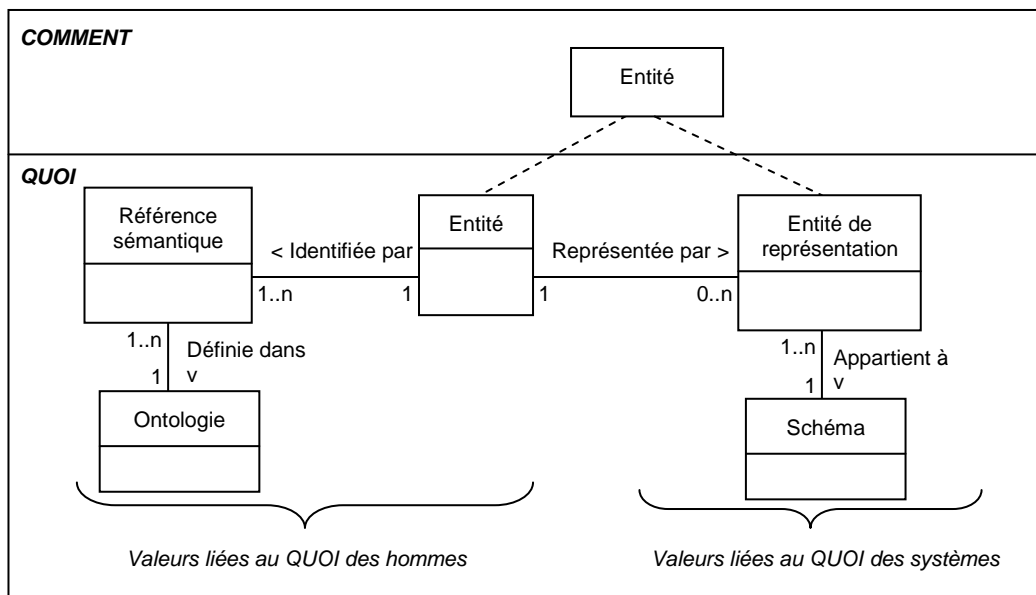


Figure 66 : Eléments du domaine associés au rôle « entité ». [Figure 66 : Domain elements associated to the role « entité ».]

Les entités connues des hommes peuvent être identifiées par une référence sémantique dans une nomenclature (ou ontologie). Ces entités peuvent être représentées dans des systèmes par des entités connues des systèmes. Ces dernières sont appelées objets géographiques ou encore objets sémantiques. Lorsqu'une entité est identifiée par une référence sémantique dans une nomenclature (par exemple un numéro de département), cette référence est souvent conservée dans une représentation de cette entité, et stockée comme un attribut pouvant servir de clé d'accès à l'objet.

La représentation privilégiée des entités est la représentation vecteur. De fait, cette représentation est associée à l'opération « où est-ce ? » [Ruas 99] p28. Elle facilite la phase de spécification de l'entrée « entité à localiser ». Au contraire, la spécification d'une entité dans une représentation maillée est plus complexe. Au mieux elle s'exprime par une expression du type : les mailles dont la valeur de l'attribut « occupation du sol » vaut « forêt ».

Les valeurs du rôle « lieu »

Une structure classique en information géographique pour la notion de lieu consiste à distinguer un système de référence spatiale et une référence dans ce système. Le lieu est une référence dans un système de référence, par exemple un jeu de coordonnées dans un système de coordonnées. Nous prenons a priori cette structure pour structurer le rôle lieu. Notons par ailleurs qu'il est possible de construire des lieux agrégés grâce aux expressions : « dans tel lieu *ou* tel lieu », « dans tel lieu *et* tel lieu ».

Pour ce qui est des lieux appréhendés par l'homme, ils peuvent être des coordonnées GPS, une ville, une région sur une carte, un concept abstrait (« dans la direction indiquée par mon doigt » etc..). Notons que dans le langage courant, des entités sont souvent utilisées pour désigner un lieu, comme une ville. Nous nous reportons à l'analyse de [Denand 00] pour formaliser ces notions de lieux ; introduite section 2.2.3.2. Il s'appuie sur les structures de phrases suivantes : « Question : pronom interrogatif + verbe locatif statique + groupe nominal
Réponse : préposition locative + groupe nominal ». Par ex : « Où est ta voiture ? - Derrière la maison. »

Les valeurs du rôle « lieu » sont ici les expressions de la forme : « la préposition locative + groupe nominal ». La préposition locative renvoie à une relation spatiale et le groupe nominal à une entité de référence, le site. Par

exemple « dans Paris » est un lieu, « Paris » est l'entité de référence, le site, et « dans » est une relation spatiale servant de référence dans le système de référence spatiale indirect centré sur Paris.

[Denand 00] et [Gonçalves 99] soulignent que la désignation d'un endroit repose sur le choix d'un référentiel, et que ce choix dépend de l'entité que l'on cherche à localiser. Il existe ainsi une relation entre entités : une entité est un site acceptable pour une autre entité si elle peut servir de référence pour la localiser.

Pour ce qui est des relations spatiales utilisées pour construire une référence par rapport à une entité site, nous retenons les mêmes relations que [Gonçalves 99] : l'orientation (ex : « à gauche de.. »), la distance (ex : « proche de »), la topologie (ex : « inclus dans »). La relation de distance ne renvoie pas toujours à une distance purement spatiale, elle peut renvoyer à une distance associée à la navigation d'un agent dans un réseau. Nous détaillons cette notion de navigation dans la section 3.2.5.

Enfin, nous considérons qu'un autre système de référence est le système de navigation d'un agent : à la question "où se trouve telle entité" il veut pour réponse l'itinéraire qu'il doit emprunter pour aller jusqu'à l'entité.

La Figure 67 propose une synthèse de ces éléments Figure 67.

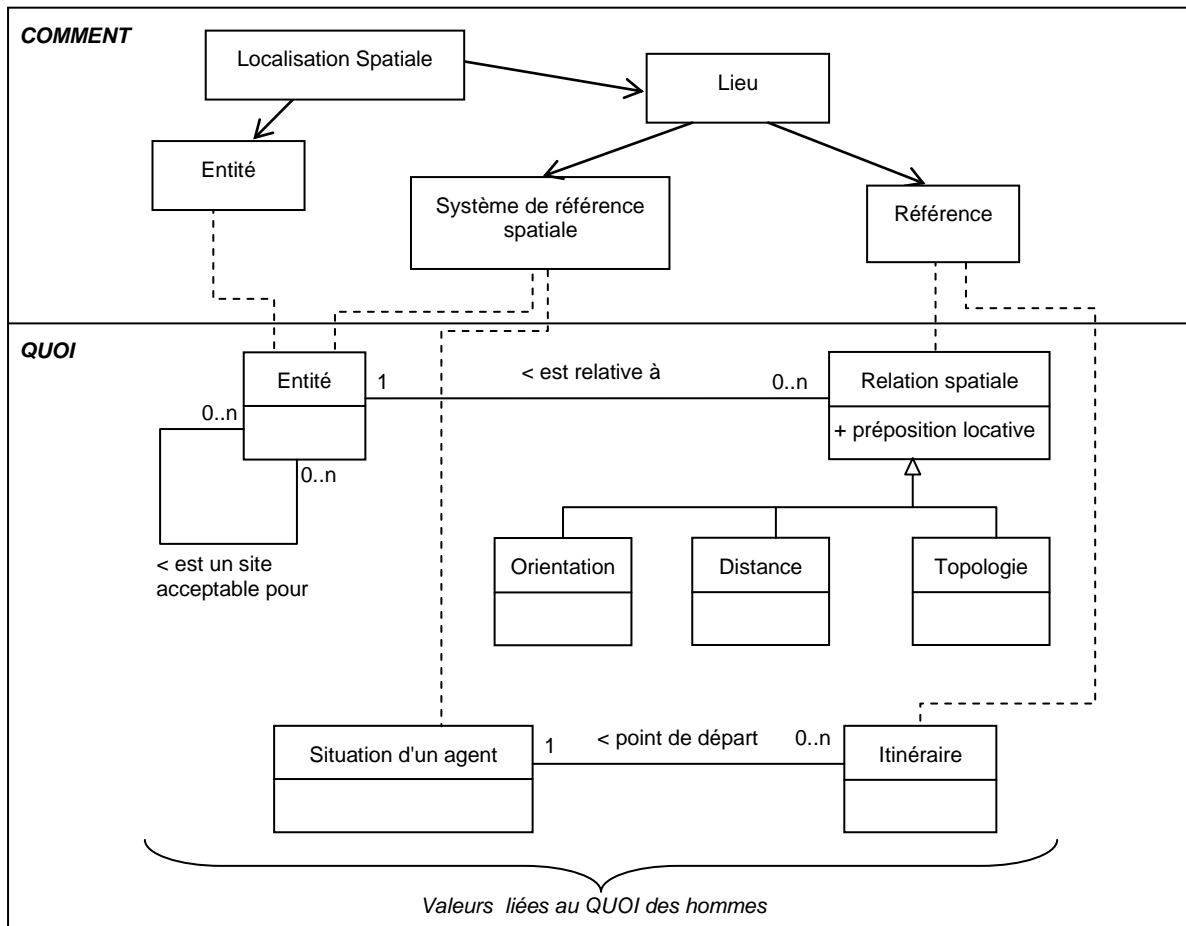


Figure 67 : Structure du rôle lieu. Dans le QUOI des hommes ces rôles renvoient à l'utilisation d'une entité de référence, dite « site », et d'une relation spatiale vis-à-vis de cette entité de référence. En fonction de l'entité à localiser, il existe des entités qui sont des sites acceptables pour la localiser. [Figure 67 : Structure of the role « lieu » (place). In the men WHAT, the roles refer to using a reference entity, the site, and a spatial relationship towards it. Depending on the entity to be located, some entity are possible sites to locate it.]

Pour ce qui est des lieux représentés dans les données, il existe deux grands types de localisations : la localisation directe et la localisation indirecte. La localisation directe correspond à l'attribution de coordonnées

dans un système de référence mathématique et la localisation indirecte correspond à un lien avec un objet géographique. Ce dernier type de localisation est celui utilisé de façon naturelle par l'homme mais également dans des applications SIG, comme dans la communauté du transport [Fletcher et al. 98]. Les principaux systèmes de références spatiales indirects sont les systèmes linéaires, et systèmes zonaux. Les systèmes linéaires ont été introduits 2.2.2.2. Les systèmes zonaux sont en fait des couches d'entités, généralement administratives, constituant une partition de l'espace. Chacun de ces systèmes indirects correspond au modèle général du lieu dans le *QUOI* des hommes en restreignant les relations spatiales. Dans la partition de l'espace la seule relation est la relation topologique d'inclusion. Dans le système linéaire le lieu est un lieu agrégé composé de : une relation de situation dans un tronçon *et* une relation de distance à un nœud du graphe.

Nous rendons compte de ces éléments Figure 68.

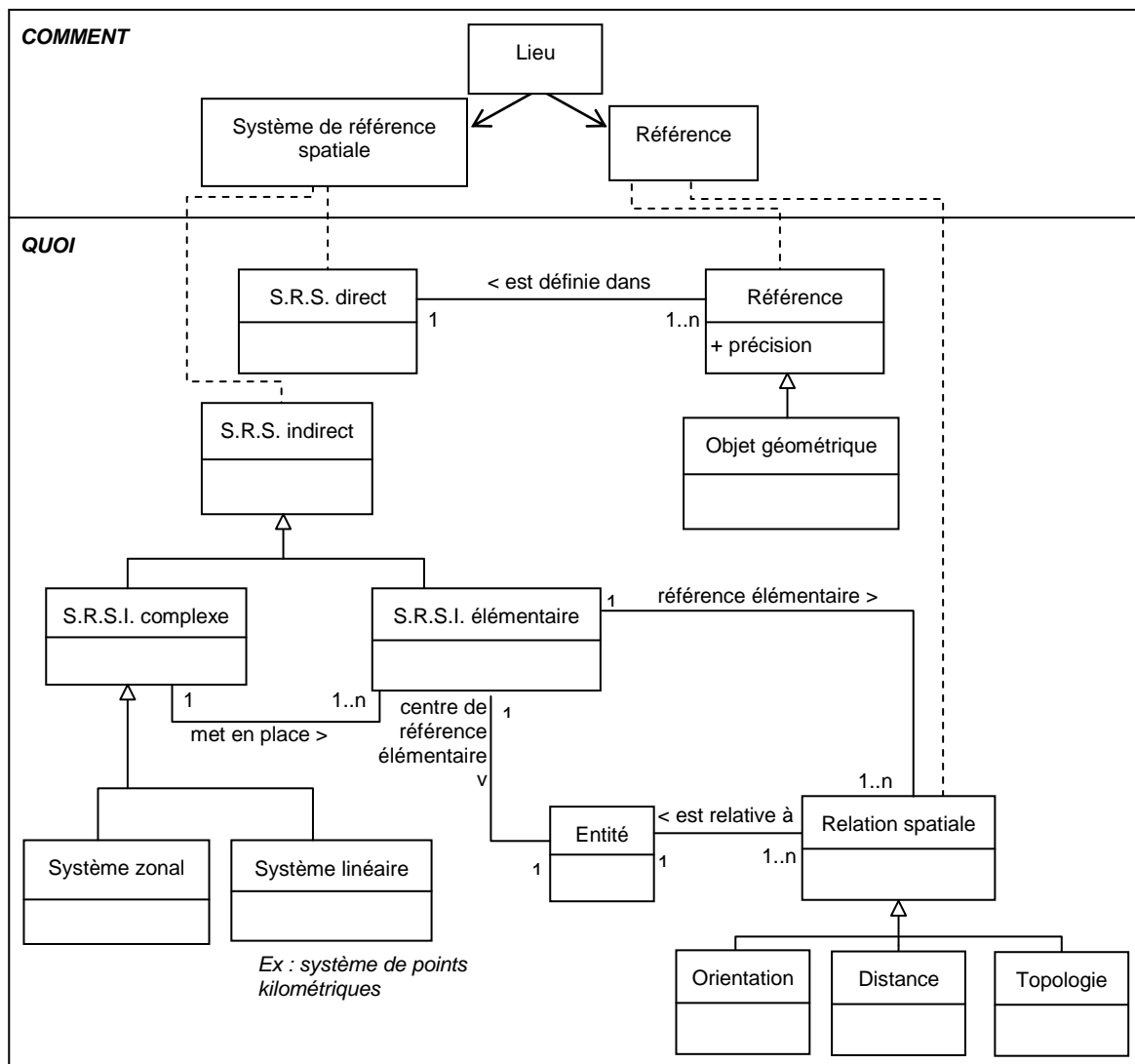


Figure 68 : Valeurs des rôles composant le lieu dans le *QUOI* informatisé. (S.R.S. : système de référence spatiale, S.R.S.I. : système de référence spatiale indirect). [Figure 68 : Values of roles composing a place in the computed WHAT. (S.R.S : spatial reference system ; S.R.S.I. : indirect spatial reference system)]

Les valeurs du rôle « localisation spatiale »

Le rôle localisation spatiale ne renvoie pas à un élément du monde réel. Il s'agit d'une notion abstraite sur laquelle s'appuient tous nos raisonnements géographiques : la notion de situation dans l'espace.

Dans le langage courant, la notion de localisation spatiale ne renvoie pas uniquement à l'association d'un lieu et d'une entité, mais aussi à des notions plus complexes. L'étude de l'ontologie de sens commun CYC, [Jarrold and Davis 96], dans le domaine de l'information géographique, a montré que certaines prépositions locatives, comme le mot anglais « above », ont des acceptions différentes selon le contexte dans lequel elles sont employées. Dans le contexte de lecture d'une carte « above » signifie « au Nord de ». Dans le contexte du vol d'un avion, « above » signifie « à la verticale de ». Une conséquence de cette remarque est que la notion de localisation spatiale renvoie parfois dans le langage courant à : « est cartographié », « vole ». Cela est lié au fait que la notion de localisation prend souvent place dans un raisonnement mené sur cette localisation. Une réponse à une tâche de localisation peut par conséquent être :

- Un lieu dans l'un ou l'autre des systèmes de référence présentés précédemment.
- Une carte quand le besoin ne consiste pas à avoir uniquement un lieu associé à l'entité mais plutôt l'ensemble des relations que l'entité entretient avec l'espace géographique. De telles relations sont appréhendées de façon privilégiée sur une carte.
- Une situation sur une trajectoire : l'avion vole au-dessus de tel pays signifie par exemple pour l'utilisateur qu'il est à mi-chemin.

Pour représenter cela, nous nous appuyons sur notre spécification du domaine géographique présentée section 3.2.1.1 qui associe à un sujet une situation pouvant être un état ou un changement d'état.

Dans notre modèle, la localisation spatiale statique est un état. La localisation spatiale dynamique est un comportement auquel est associée, comme état, une localisation spatiale statique.

Par ailleurs nous modéliserons au niveau de la tâche de localisation le fait que la sortie peut être spécifiée comme une carte ou une description de navigation.

Dans les représentations utilisées dans les bases de données géographiques, ce rôle est souvent un lien d'attribut entre un objet géographique (l'entité) et un objet géométrique (l'endroit), ou entre une maille (l'endroit) et un attribut sémantique (l'entité).

Ces éléments sont repris Figure 69.

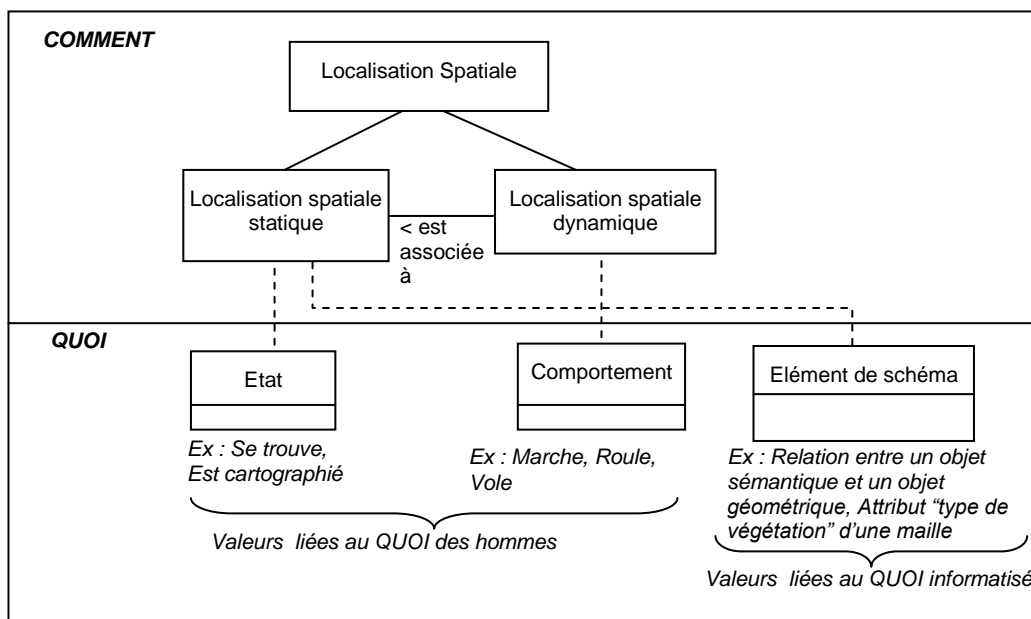


Figure 69 : Valeurs du rôle localisation spatiale. [Figure 69 : Values of the role « spatial localisation ».]

3.2.4.2 L'aspect opérationnel

La tâche de localisation est extrêmement générale. Il n'existe pas dans la littérature de méthode générique de localisation, qui puisse être spécifiée en fonction des divers contextes dans lesquels on fait appel à la tâche de localisation.

Nous définissons ici une méthode générique de localisation et les éléments de spécification de cette méthode en fonction des contextes de réalisation de cette tâche.

Méthode générique

Nous proposons de résumer la tâche de localisation en considérant que cette tâche consiste à produire une référence spatiale telle que :

- cette référence est effectivement associée à l'entité que veut localiser l'utilisateur,
- cette référence est définie dans un système pertinent pour le besoin de l'utilisateur. En effet, l'utilisateur cherche rarement les coordonnées Lambert d'une entité, mais il peut chercher une relation spatiale avec une autre entité.

Cela permet de définir une méthode générique décomposant la tâche de localisation selon deux étapes :

- La première étape est de produire une référence spatiale effectivement associée à cette entité, obtenue le plus simplement possible, et exploitable par un SIG. Notons que si l'utilisateur possède une référence spatiale de l'entité qu'il veut localiser, cette première étape est inutile.
- La deuxième étape est de produire une nouvelle référence spatiale, à partir de celle produite dans la première étape, en se plaçant dans un système de référence pertinent pour l'utilisateur. Notons que si la référence produite dans la première étape est exploitable par l'utilisateur alors cette deuxième étape est inutile.

Nous détaillons ci-dessous ces étapes. Chacune se décompose elle-même en deux sous-tâches :

- acquérir un système de référence : dans la première étape, il s'agit d'un système dans lequel on peut définir simplement une référence, et, dans la deuxième étape, il s'agit du système permettant de produire le type de référence spatiale voulu par l'utilisateur,
- produire une référence spatiale.

Première étape : produire une référence spatiale intermédiaire

Lorsque l'utilisateur possède déjà une référence spatiale, dans un système de référence spatiale direct, de l'entité qu'il veut localiser, comme des coordonnées GPS ou géographiques, la première étape est inutile.

Dans d'autres cas, l'utilisateur possède déjà une référence spatiale dans un système indirect suffisamment "universel" : système d'adresses postales, système de référence linéaire comme le réseau routier, ferroviaire ou de métro, système de partition de l'espace en zones administratives. Dans ces cas, il y a deux possibilités. Si l'étape suivante peut être accomplie en exploitant cette référence indirecte alors il n'y a pas de première étape. Si l'étape suivante a besoin d'une référence spatiale dans un système direct alors il faut effectuer un changement de système de référence. Cela se fait en acquérant une BDG permettant deux modes de positionnement : le premier étant celui de l'utilisateur, le deuxième étant un système de coordonnées géographiques.

Lorsque l'utilisateur ne possède pas de référence spatiale connue, il faut exploiter les éléments spécifiant l'entité qu'il cherche à localiser et traduire ces éléments en une référence spatiale.

Il est possible que l'entité à localiser soit représentée dans une BDG. Pour obtenir une référence spatiale d'une telle entité, il faut :

- Acquérir un jeu de données dans lequel l'entité est représentée.

- Appliquer des mécanismes pour sélectionner cette entité : cela peut être simplement l'expression d'un identifiant dans une nomenclature commune à l'utilisateur et à la BD. Cela peut être aussi l'expression de relations sémantiques ou spatiales entre cette entité et d'autres entités représentées dans la base.
- La référence spatiale est alors le résultat d'une opération de localisation élémentaire de cette entité (décrite plus loin).

Il est possible également que l'entité à localiser ne soit pas représentée dans une BDG mais soit liée à des entités représentées dans des BDG par des relations qui peuvent se traduire en des restrictions sur la référence spatiale recherchée. Il faut alors :

- Acquérir un jeu de données dans lequel les entités en relation avec l'entité de l'utilisateur sont représentées (cela peut nécessiter d'intégrer plusieurs jeux).
- Pour chaque relation, il faut sélectionner les entités concernées dans ce jeu, et effectuer une localisation élémentaire de ces entités.
- Pour chaque relation, traduire la relation sémantique en une restriction s'appuyant sur les références spatiales produites précédemment. Par exemple, si l'entité recherchée est proche d'une route, un buffer autour d'un segment représentant une référence de cette route produit une zone dans laquelle se trouve la référence spatiale recherchée.

Les éléments du contexte permettant de spécifier cette étape de la tâche sont donc :

- des éléments de localisation connus,
- des caractéristiques de l'entité recherchée, type, relations sémantiques ou spatiales avec d'autres entités,
- des restrictions sur la localisation recherchée, comme un rectangle englobant.

Deuxième étape : produire une référence spatiale exploitable par l'utilisateur

Cette deuxième étape consiste à :

- acquérir un système de référence satisfaisant l'utilisateur,
- traduire dans ce système la référence spatiale produite par l'étape précédente en une référence spatiale répondant au besoin de localisation de l'utilisateur, c'est-à-dire effectuer un changement de système de référence.

Il existe une variété de tels systèmes car il existe une variété de formats de réponse à la tâche de localisation. Une localisation recherchée peut être des coordonnées géographiques, ou un positionnement par rapport à un objet de référence, ou une situation sur une carte, ou un itinéraire depuis un lieu de référence.

L'élément du contexte permettant de préciser le format de la réponse est ce que l'utilisateur veut faire du résultat. De fait, la tâche de localisation, comme la tâche d'appariement, n'est pas une fin en soi mais prend place dans une tâche de plus haut niveau : un utilisateur a besoin d'une localisation pour mener un raisonnement dessus. Cela peut être un raisonnement mené par l'homme, ou un raisonnement mené par un système informatique.

Si le résultat doit être exploité par un homme, il est important de choisir un système de référence qui corresponde à la localisation perçue par l'homme. Les formalismes permettant de représenter sur des données géographiques les localisations perçues par l'homme sont étudiés par [Gonçalves 99]. Un tel formalisme de description "conviviale" de la localisation impose souvent un changement de représentation des objets composant le système de référence et également une étape d'abstraction des formes géométriques [Gonçalves 99].

Si le résultat doit être exploité par un système il est important que son format soit adapté à la manipulation que le système en fera.

En définitive, pour spécifier cette étape, il faut connaître la tâche que le système réalisera avec le résultat.

Nous traduisons sur un schéma général, Figure 70, la méthode générique de la tâche de localisation décrite ci-dessus. Cette méthode a été codée dans le prototype présenté dans la section 3.3. Nous détaillons dans cette section certaines tâches primitives utilisées dans cette méthode.

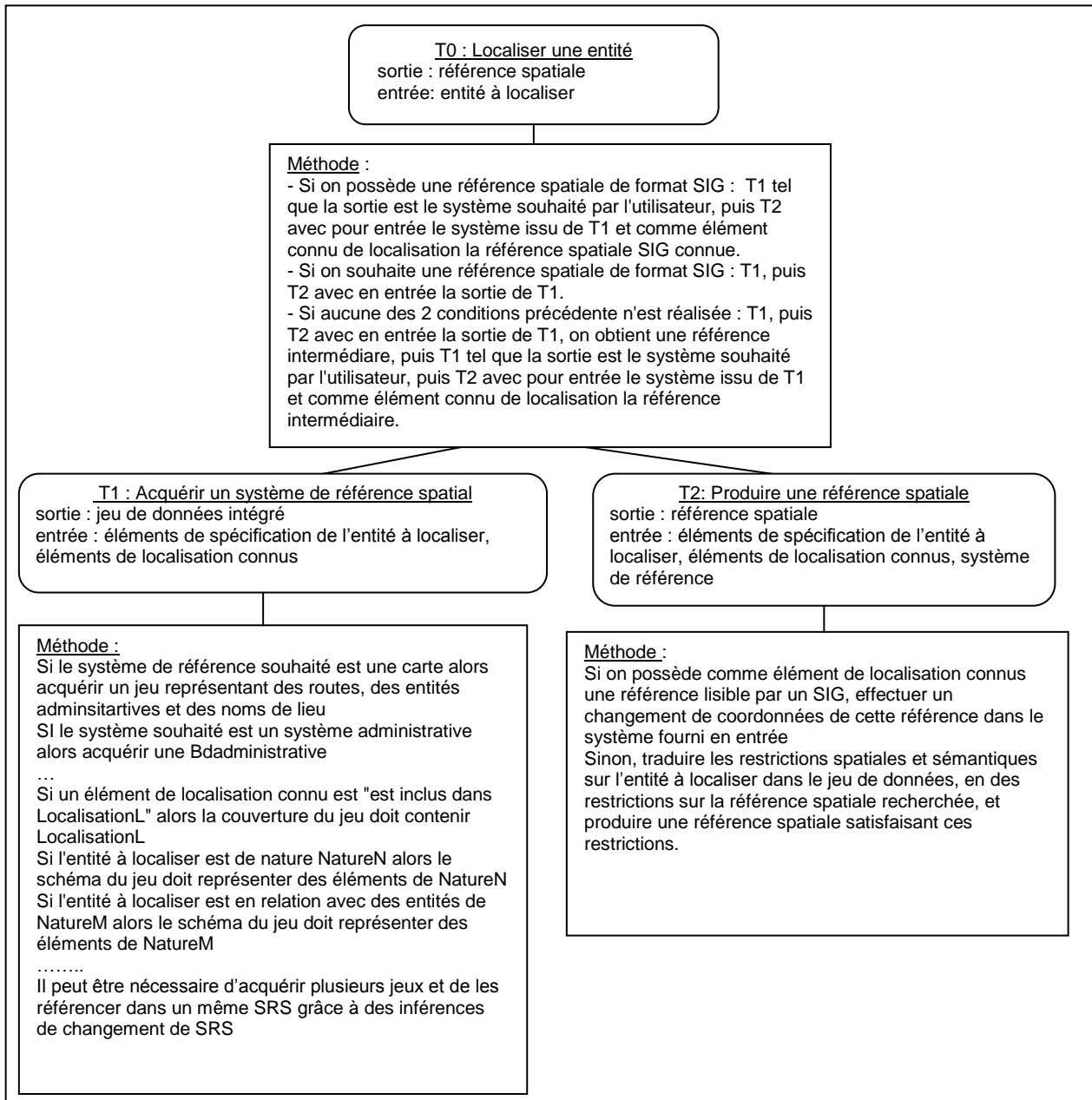


Figure 70 : Décomposition de la tâche de localisation. [Figure 70 : Decomposition of the localisation task.]

Sous-tâches intervenant dans la décomposition

Une sous-tâche essentielle est la localisation élémentaire. Elle consiste à associer, dans un jeu de données, une expression géométrique à une expression sémantique. Elle s'effectue en fonction de l'expression sémantique en entrée et du schéma de données.

Si l'expression sémantique consiste à en une sélection d'un objet, la géométrie peut être déjà représentée dans le jeu de données et associée à l'entité, comme un objet linéaire est associé à une route. Il est possible que cette géométrie ne soit pas représentée mais puisse être calculée : par exemple associer à un département une surface qui est la somme des surfaces des communes incluses dans ce département, associer une surface à une zone cultivée délimitée par des routes. Il est également possible que plusieurs lieux soient disponibles pour une entité. Par exemple une ville peut être associée à une surface, l'extension spatiale de cette ville, ou à un point, le centroïde de cette extension. Le calcul de chacun de ces lieux peut être plus ou moins complexe. Dans certaines

BDG, des géométries simplifiées sont stockées pour certaines entités, comme un rectangle englobant pour une route.

Le plan faisant appel à cette sous-tâche doit ainsi spécifier le type de lieu qui l'intéresse en s'appuyant sur des critères de simplicité du calcul (la géométrie est-elle fournie dans le jeu de données ou faut-il la calculer, et à quel prix), et sur des critères d'exploitation du résultat.

Une autre sous-tâche essentielle consiste à traduire une relation sémantique en une relation spatiale. Elle est utilisée pour approcher la référence spatiale recherchée à partir des connaissances de spécification énoncées par l'utilisateur comme : « l'entité que je recherche est un bâtiment », « le bâtiment que je recherche est près d'une route », « l'entité que je recherche est au centre de la France ». Le mécanisme de cette tâche peut être spécifié ainsi que la sortie produite. Par exemple si la relation est une relation de distance (ex : "est à une distance inférieure à 100m de la route r"), le résultat est une relation spatiale d'inclusion de la référence recherchée dans la zone produite et le mécanisme est un buffer. De plus, le buffer peut être un traitement SIG ou un traitement effectué de façon interactive. Dans ce dernier cas, l'utilisateur doit utiliser un SIG pour éditer ses données graphiquement et traduire lui-même graphiquement la relation spatiale.

Une autre sous-tâche est le changement de système de référence. Cette tâche est bien maîtrisée entre des systèmes de référence direct car elle correspond alors à des opérations de changement de coordonnées mathématiques. Il demeure toutefois une difficulté : déduire, à partir de la précision de la référence dans l'ancien système, la précision de la référence dans le nouveau système.

Lorsqu'il s'agit d'opérer des changements de systèmes de références entre systèmes dont un est indirect, la tâche est plus compliquée. Il existe des systèmes de référence spatiale indirecte dont la vocation est de permettre d'obtenir des références unifiées à partir de références dans divers systèmes. Ces systèmes sont associés à des méthodes de changement de système de référence, des autres systèmes vers le système standard.

Le « Datum » introduit par [Fletcher et al. 98] a ainsi pour rôle essentiel "to transform locations between the real-world space and data-world space(s) and to project positions between spaces", dans un contexte d'applications de navigation, Figure 71. C'est un système de référence indirect linéaire, dans lequel les références utilisées sont les coordonnées de Vonderhoe [Fletcher et al. 98].

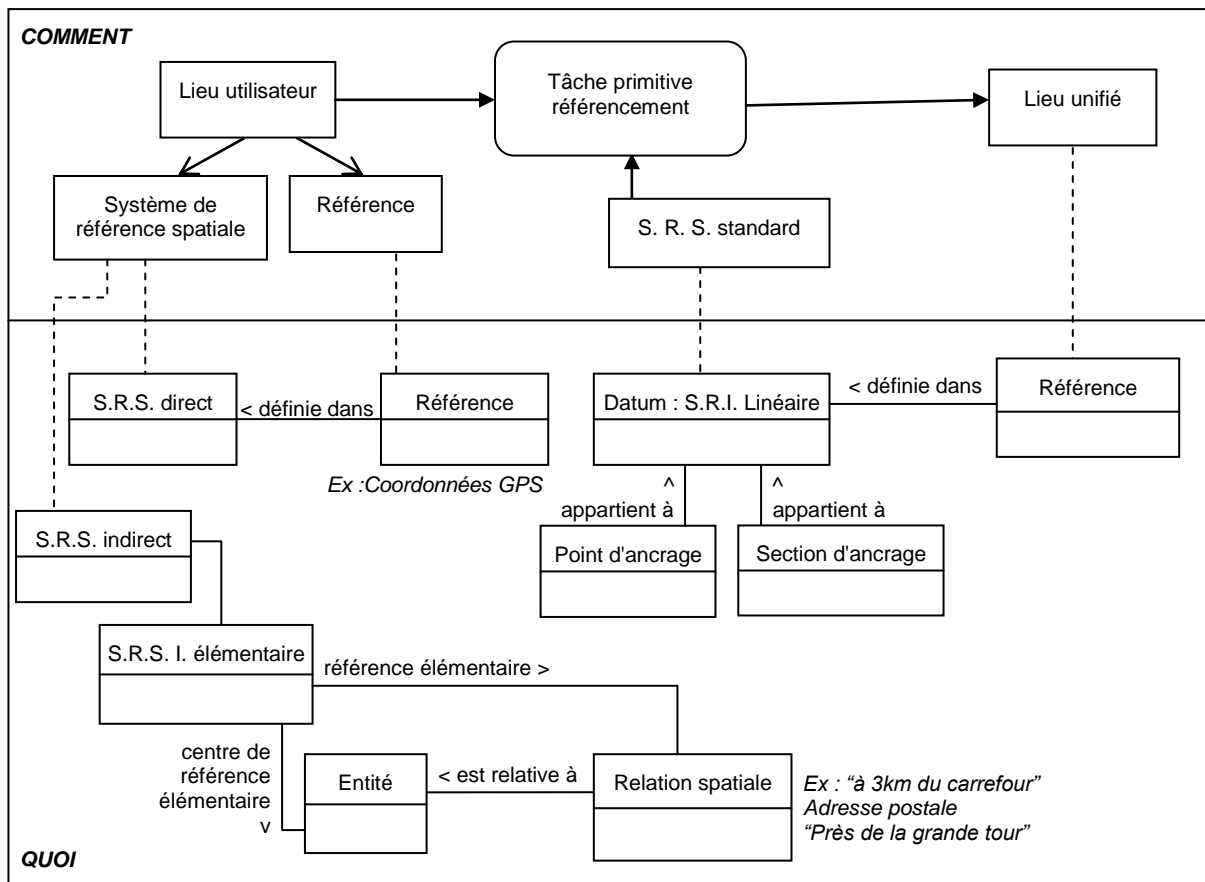


Figure 71 : Représentation des éléments servant à référencer linéairement des objets d'utilisateur dans une application de navigation (éléments repris à partir de [Bédard and Sondheim 97], [Fletcher et al. 98]). [Figure 71 : Representation of elements building a linear reference for user objects in a navigation application (after [Bédard and Sondheim 97][Fletcher et al. 98])]

La notion de « collage spatial » de [Claramunt et Mainguenaud 96] est en fait un changement de système de référence qui permet de passer d'un système de référence pertinent dans la première vue spatiale à un autre système plus pertinent dans la deuxième vue spatiale.

Description de la tâche de localisation

Nous proposons une description BNF de cette tâche, qui a fait l'objet d'une implémentation dans notre prototype. Cette description n'est pas reprise complètement car elle est relativement lourde.

TC Localisation**Sortie** : référence spatiale**Entrée** : entité à localiser, éléments de localisation connus**Méthode** : **Propag spécif** : **Spécification de sortie**(=**Int**(est définie dans srs spécifique)) → **Spécification de srs souhaité**(=**Int**(srs spécifique));**Spécification de éléments de localisation connus**(=**Int**(est définie dans srs direct))→ **Spécification de entrée "réf connue"** de la tâche produisant la ref souhaitée (=**Int**(éléments de localisation connus));**Spécification de éléments de localisation connus**(=**Int**(est définie dans srs spécifique))→ **Spécification de entrée "jeu de données"** de la tâche produisant le srs intermédiaire (=**Int**(jeu supportant le positionnement spécifique d'éléments de localisation connus)) ;**Spécification de éléments de localisation connus**(=**Int**(est incluse dans loc spécifique)) → **Spécification de entrée "jeu de données"** de la tâche produisant le srs intermédiaire (=**Int**(jeu couvrant la localisation spécifique));**Spécification de entité à localiser**(=**Int**(est une entité administrative)) → **Spécification de entrée "jeu de données"** de la tâche produisant le srs intermédiaire (=**Int**(jeu représentant des entités administratives));**Spécification de entité à localiser**(=**Int**(est proche d'une entité spécifique)) → **Spécification de entrée "jeu de données"** de la tâche produisant le srs intermédiaire (=**Int**(jeu représentant l'entité spécifique)), **Spécification de entrée "relation connue"** de la tâche produisant la ref intermédiaire (=**Int**(est proche de entité spécifique)) ;**Stratégie** : **Spécification de sortie**(=**Int**(est définie dans srs direct)) → plan non détaillé = structure2 ;**Spécification de éléments de localisation connus**(=**Int**(est définie dans srs direct))→ plan non détaillé=structure2 ;

....

Plan non détaillé : **Vocabulaire** : srs intermédiaire(srs intermédiaire **dans** Structure1, srs souhaité **dans** Structure2),ref intermédiaire (ref intermédiaire **dans** Structure1,ref souhaitée **dans** Structure2),srs souhaité(srs souhaité **dans** Structure1, srs souhaité **dans** Structure2),ref souhaitée(ref souhaitée **dans** Structure1, ref souhaitée **dans** Structure2)

OU (Structure1, Structure2)

structure Structure1 : **vocabulaire** = (srs intermédiaire, ref intermédiaire, srs souhaité, ref souhaitée)

ET(AcquérirSRS/ sortie = srs intermédiaire;

ProduireRS/srs=srs intermédiaire/ sortie=ref intermédiaire;

AcquérirSRS/ sortie = srs souhaité;

ProduireRS/srs=srs souhaité/ref connue=ref intermédiaire/ sortie=ref souhaitée)

structure Structure2 : **vocabulaire** = (srs souhaité, ref souhaitée)

ET(AcquérirSRS/sortie=srs souhaité;

ProduireRS/srs=srs souhaité/sortie=ref souhaitée)

Nous ne détaillons pas la tâche d'acquisition d'un système de référence spatiale qui est modélisée comme l'acquisition d'un jeu de données géographiques. Cette tâche devra par la suite être détaillée davantage. En particulier, elle peut être réalisée en acquérant plusieurs jeux et en les intégrant.

TC ProduireRS (Production d'une référence spatiale)**Sortie** : référence spatiale

Entrée : système de référence spatiale (**Int**(système de référence, cartographie, navigation)), référence connue

Méthode : **Propag spécif** : **Spécification de** sortie de structureK → **Spécification de** restrictions de localisation+=sortie spécifiée ;

Stratégie : **Spécification de** référence spatiale connue (=définie dans SRS direct) → S = Changement de système de coordonnées;

Spécification de relation connue (=proche de entité spécifique) → S = ET(Localisation élémentaire, Buffer(/matrice géométrique = sortie de Localisation élémentaire));

.....

Plan non détaillé : **Vocabulaire** : relations connues, restrictions de localisation, référence spatiale connue

TANT QUE((L'utilisateur n'est pas satisfait du résultat)^(L'utilisateur peut spécifier encore des rôles de la tâche de localisation)) **REPETE**(ET(S , CroisementRestriction)

structure S : **OU** (Changement de système de coordonnées/entrée=référence spatiale connue),(Localisation élémentaire), **ET**(Localisation élémentaire/entité = entité spécifique/srs = SRS , Traduction de relation sémantique en restriction de localisation))))

La tâche Traduction de relation sémantique en restriction de localisation produit des restrictions du type : est incluse dans zone spécifique, intersecte zone spécifique, n'intersecte pas zone spécifique, contient zone spécifique. Ces restrictions traduisent les relations sémantiques connues spécifiées au niveau de la tâche ProduireRS.

La tâche CroisementRestriction consiste à définir une référence spatiale dans le système SRS satisfaisant toutes les restrictions de localisation.

3.2.4.3 Exploitation du modèle de localisation

Le principal intérêt de ce modèle de la tâche de localisation concerne la mise de l'information géographique à disposition du grand public. En effet la principale question géographique que se pose le grand public est une question « où est-ce ? ». Le modèle a montré qu'il existait de nombreuses formes de réponses et qu'il était important de savoir quelle forme de « où » intéressait l'utilisateur. Il a aussi montré quelles connaissances de l'utilisateur sur l'entité à localiser exploiter pour obtenir ce « où » recherché. Ceci est illustré sur la Figure 72.

Les connaissances sur lesquelles s'appuie le dialogue Figure 72 sont :

- les connaissances des tâches de localisation, d'acquisition d'un système de référence spatiale, de production d'une référence spatiale,
- les connaissances du domaine.

Par contre, les phrases apparaissant dans cette figure ont été rédigées à la main. Elles ne correspondent pas à une interface de dialogue en langage naturel du système mais elles veulent dépeindre de façon claire les connaissances communiquées entre le système et l'utilisateur.

Les modalités de ce dialogue sont les interfaces graphiques du prototype présentées dans la section suivante.

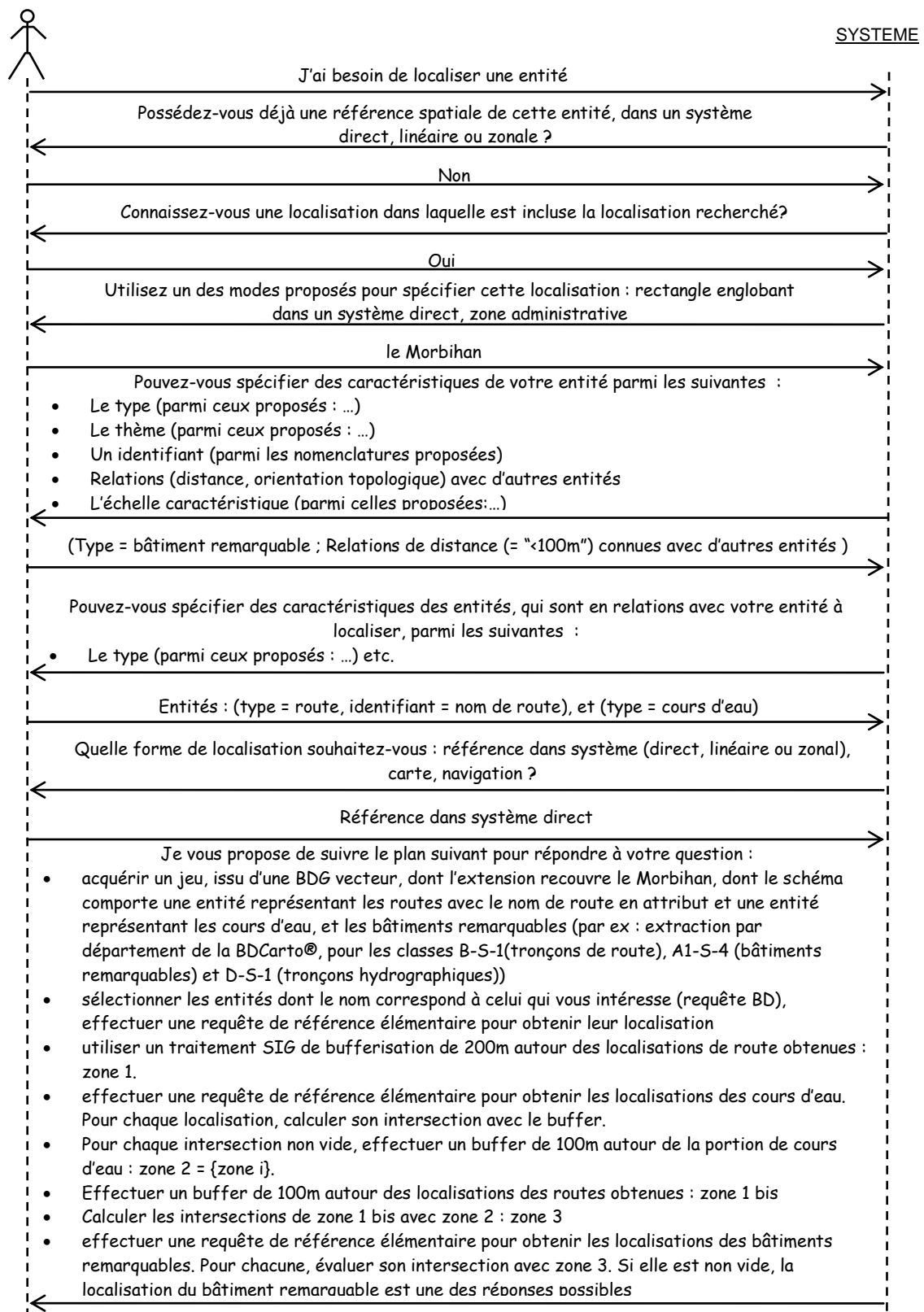


Figure 72 : Diagramme de séquence s'appuyant sur le modèle de la tâche de localisation pour aider un utilisateur à répondre à une question « où ? ». [Figure 72 : Sequence diagram describing the process of answering a user's need for answering a «where's that ? » question.]

3.2.5 La tâche de détection de proximité et de détermination de navigation

Nous analysons dans cette section la tâche de détection de proximité dont un cas particulier est le calcul d'itinéraire, soit encore la détermination d'une navigation.

Il existe de nombreux systèmes de calcul d'itinéraire et nous ne cherchons pas à analyser leurs mécanismes et à évaluer les méthodes qu'ils appliquent pour effectuer cette tâche. Par contre nous nous attachons à la spécification du besoin de calcul d'itinéraire. En effet, les systèmes actuels proposent un calcul avec peu de paramètres alors que dans la réalité, de nombreux paramètres interviennent dans la détermination d'un itinéraire optimal.

3.2.5.1 Les rôles de la détection de proximité

L'objectif de cette tâche : déterminer des objets *proches* d'un autre. La notion d'*être proche* est associée à un concept important : le voisinage. Ce concept est lui-même associé au concept de distance. Ces concepts sont purement spatiaux. Pourtant, on peut envisager une tâche de détection de proximité associée à une navigation comme par exemple « Quelle est la station d'essence la plus proche en voiture ? ». Nous utilisons le terme proximité pour englober plusieurs cas : celui du voisinage associé à une distance purement spatiale et d'autres cas comme la distance associée à la navigation d'un agent. La distance associée à une navigation est un coût en temps ou longueur.

La structure des rôles de la détection de proximité comprend une base, un but, et une proximité de la base au but, Figure 73. Ceci est lié à deux considérations. La première est que la distance servant à évaluer une proximité n'est pas toujours symétrique. Une distance purement spatiale est symétrique. Une distance associée à la navigation d'un agent est rarement symétrique : il peut être rapide de se rendre de A à B en voiture mais le retour peut être plus long selon les restrictions de circulation (sens interdits, feux). La deuxième est que la question peut se décliner en : l'objet le plus proche, les n plus proches, ceux qui sont à une distance inférieure à une certaine valeur. Ces cas particuliers ne sont pas symétriques. Par exemple si tel hôtel H est le plus proche d'une station de métro M, cela n'implique pas que M est la station de métro la plus proche de H.

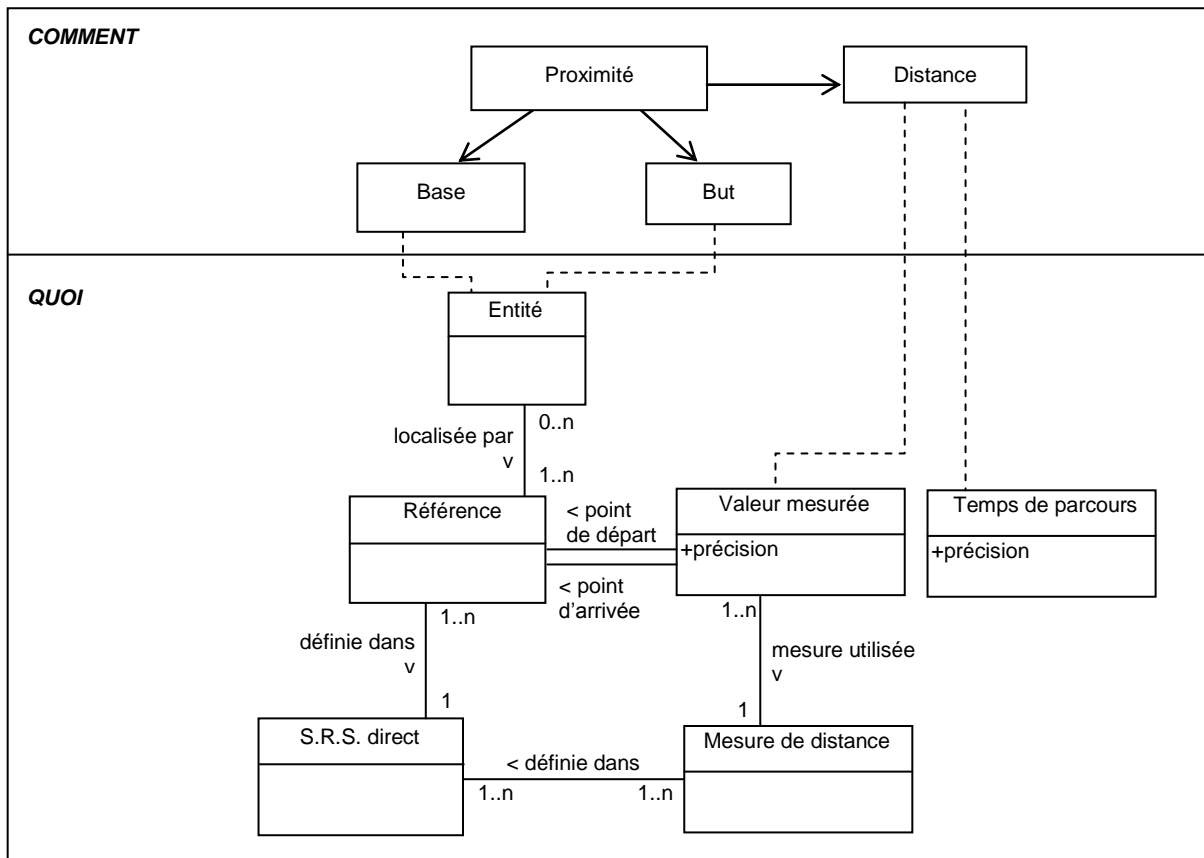


Figure 73 : Objets du domaine associés aux rôles de détection de proximité, dans le cas d'une distance définie dans un S.R.S. direct. [Figure 73 : Domain objects associated to the proximity detection roles, in the case where the distance is defined in a direct S.R.S..]

3.2.5.2 Les rôles de la détermination de navigation

Nous détaillons les rôles liés à la navigation.

De nombreux éléments sont fournis dans la littérature, soit pour formaliser des principes cognitifs de navigation, soit dans une optique de fédérer des données géographiques pour des applications de navigation, 2.2.2.1.

Temps, localisation spatio-temporelle et événement

Le temps est une notion importante pour construire et décrire une dynamique comme la navigation. La notion de temps perçue par l'homme est attachée au changement et au mouvement. Le changement est la différence entre deux états, en particuliers à des dates différentes.

Pour représenter cela, nous introduisons tout d'abord la notion de localisation spatio-temporelle : l'association entre une localisation spatiale et une référence dans un système de référence temporel.

Notons qu'une caractéristique importante d'un système de référence spatiale est sa stabilité temporelle. Par exemple, le système utilisé dans l'expression « au niveau de la twingo rouge garée là-bas », est instable. Cette caractéristique entre en jeu dans le choix d'une site admissible pour localiser une entité donnée [Denand 00].

La notion de changement est représentée par l'événement dans le monde et l'évolution dans la représentation [Claramunt et al. 94].

Enfin, le mouvement d'un sujet peut être représenté comme une suite continue de localisations spatio-temporelles de ce sujet. Ces localisations sont vues soit comme des situations statiques, des photos figées, soit comme des situations dynamiques.

Navigation, trajectoire

Nous utilisons le terme navigation au lieu de mouvement. C'est une activité particulière qui peut être représentée par notre schéma général Figure 57.

Pour nous rattacher aux éléments en place du modèle, et introduire la contrainte de continuité, nous modélisons la navigation comme une suite chronologique dans le temps de localisations spatiotemporelles d'une même entité, qui doit posséder une continuité significative dans le référentiel temporel utilisé pour caractériser les localisations.

Le mouvement respecte des contraintes comme :

- la contrainte de continuité par rapport au temps
- la contrainte de véhicule : si je suis dans une voiture et que cette voiture bouge alors je suis le déplacement de la voiture
- la contrainte de support de navigation : une entité ne peut pas bouger librement dans l'espace géographique. Par exemple un homme ne peut pas traverser un mur, ou voler au dessus d'un océan sans moyen de transport.

Le concept de trajectoire correspond à la suite de localisations spatiales associées à une navigation : il n'y a pas de notion temporelle dans une trajectoire alors qu'il y en a une dans la navigation.

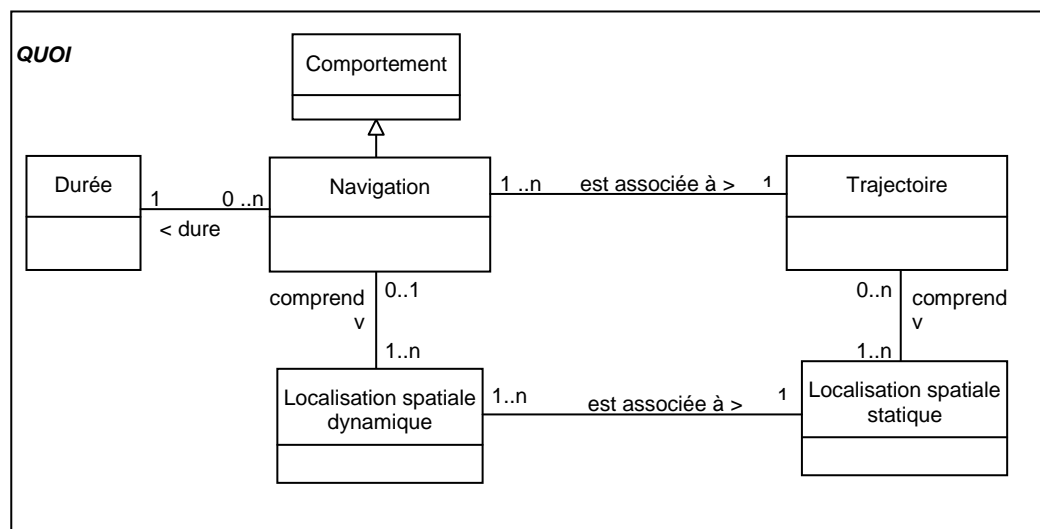


Figure 74 : Les termes de navigation et de trajectoire. [Figure 74 : The concepts of navigation and of trajectory.]

Nous introduisons ci-dessous le vocabulaire d'une navigation dans les représentations de l'homme. Ces termes sont rattachés au rôle générique « navigation », qui est défini au paragraphe précédent.

Le rôle navigation principale désigne la navigation de l'entité qui doit changer de lieu dans la tâche de navigation. C'est une navigation caractérisée par des localisations particulières : départ, destination, dessertes. L'agent de la navigation principale peut être transporté par une autre entité. Nous exprimons cela de la façon

suivante : une navigation peut être véhiculée par une autre. Dans ce cas on ne se préoccupe pas de la navigation relative de l'entité par rapport à son véhicule et on lui donne pour références spatiales et temporelles celles de son véhicule. Pour regrouper la notion de prise de véhicule, de navigation supportée par la navigation du véhicule et d'abandon de véhicule, nous utilisons le terme "véhiculement".

Notre modèle doit permettre de stocker les relations entre entités : « est un véhicule admissible de ». et par une ou plusieurs navigations véhiculantes successives.

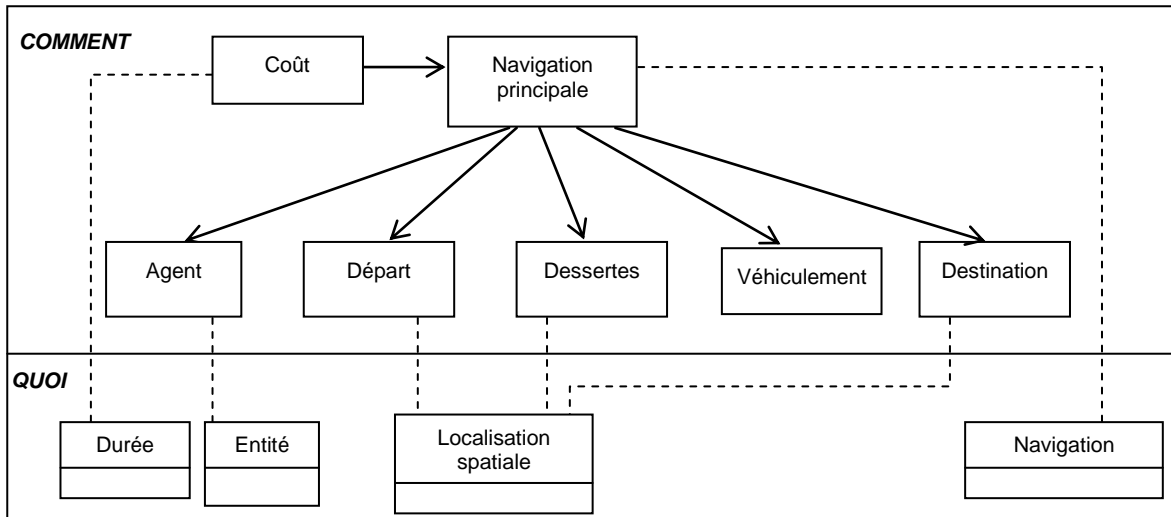


Figure 75 : Rôles associés à la spécification d'une navigation principale. [Figure 75 : Roles associated to the specification of a user's navigation.]

Un "véhiculement" est caractérisé par le véhicule mais ne se résume pas à la navigation du véhicule. Le véhicule peut être l'agent de la navigation principale lui-même (ex : quand je marche à pied) ou une autre entité (ex : une voiture, un train...). Les abandon et prise de véhicule sont des comportements de l'entité agent. Il sont soumis à des contraintes, comme « il n'est pas permis d'abandonner son véhicule au bord de l'autoroute s'il n'est pas en panne ».

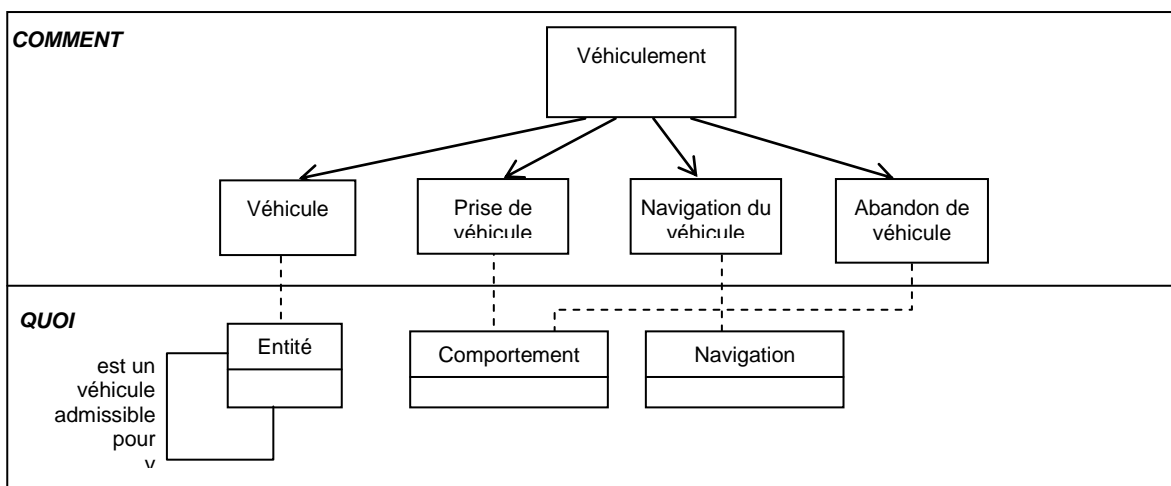


Figure 76 : Rôles associés à la spécification d'une navigation véhiculante. [Figure 76 : Roles associated to the specification of a vehiculating process.]

Par ailleurs, un agent –éventuellement un véhicule- qui n'est pas véhiculé doit être supportée. Notre modèle doit permettre de stocker des relations comme : « est un support admissible de ». Nous considérons cette relation comme une relation entre navigation : une navigation peut être supportée par une navigation « support ». Les navigations support seront généralement des navigations immobiles (typiquement une route ne se déplace pas sur la terre).

Alors que la localisation de l'agent est liée de façon constante à celle de son véhicule, il est souvent nécessaire de discriminer la localisation d'un agent ou d'un véhicule à l'intérieur de la localisation de l'entité support. Les évolutions de cette localisation (d'une entité à l'intérieur d'une entité support) sont éventuellement soumises à des contraintes de guidage (on ne peut pas faire demi-tour sur une autoroute). Nous modélisons cela en introduisant la notion de portion guidée et de prise et abandon de guide. Une portion guidée est une partie de la navigation où il n'y a pas de possibilité de changer d'itinéraire –par exemple une portion d'autoroute entre deux sorties-. Une prise et un abandon de guide sont des comportements de l'agent de la navigation, comme tourner à un carrefour, et ces comportements sont soumis à des contraintes.

En définitive, un support de navigation d'un véhicule est caractérisée par une entité support, les abandons et prises de support qui sont des localisations (accès du véhicule à l'entité support). L'itinéraire du véhicule à l'intérieur de l'entité support ne correspond pas exactement à la trajectoire du véhicule mais plutôt à ce qui suffit à caractériser la navigation principale (pour une tâche de guidage par exemple, ou de mesure de consommation d'essence etc.). Les guides de navigation d'une entité dans l'entité support sont des parties de cette entité dans lesquelles l'entité ne peut avoir qu'une trajectoire possible. Par exemple une portion d'autoroute entre deux sorties est un guide de navigation pour une voiture.

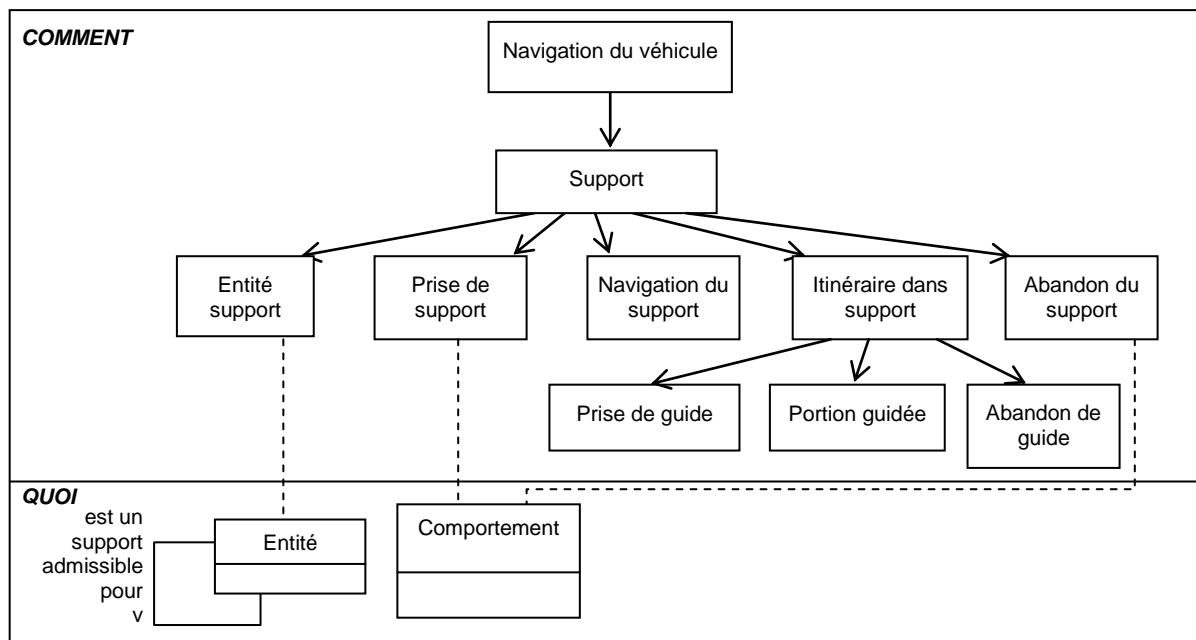


Figure 77 : Rôles associés à la spécification de la navigation d'un véhicule. [Figure 77 : Roles associated to the specification of the navigation of a vehicle.]

Des connaissances importantes à stocker dans notre système sont les connaissances de valuation des rôles :

- les relations entre entités : « est un véhicule admissible pour » et « est un support admissible pour »,
- les connaissances de satisfaction des contraintes de comportement : prise et abandon de véhicule, prise et abandon de support, prise et abandon de guide.

3.2.5.3 Exploitation du modèle de navigation

Nous n'avons pas développé la méthode générale de calcul d'itinéraire. Elle consiste à construire un graphe reflétant les caractéristiques de la navigation et utiliser un algorithme d'optimisation de parcours de graphe.

Dans notre modèle, ce graphe est défini de la façon suivante :

- Les arêtes sont des portions guidées.
- Les sommets sont des comportements de prise ou abandon de véhicule, de prise ou abandon de support et de prise ou abandon de portion guidée.

Ce modèle permet de tenir compte très largement du contexte de l'utilisateur et de nombreux facteurs entrant en jeu dans la spécification de la navigation qui l'intéresse.

Ce modèle donne également des pistes pour aider un utilisateur à effectuer la tâche de calcul d'itinéraire ainsi spécifiée. Les connaissances de valuation des rôles permettent à un système de déterminer quels éléments utiliser pour construire le graphe. D'autres connaissances sont nécessaires pour effectuer l'optimisation, il s'agit de la spécification de la notion de coût pour l'utilisateur. Le coût peut correspondre à un temps minimum de navigation, ou à un nombre maximal de portions d'itinéraire dans des paysages agréables, ou à un minimum de frais (billet de train, location de véhicule, essence, péages, etc.). Notons que le coût est souvent un compromis. Ceci est illustré sur la Figure 78. Les phrases apparaissant dans cette figure ont été rédigées à la main. Elles ne correspondent pas à une interface de dialogue en langage naturel du système mais elles veulent dépeindre de façon claire les connaissances communiquées entre le système et l'utilisateur.

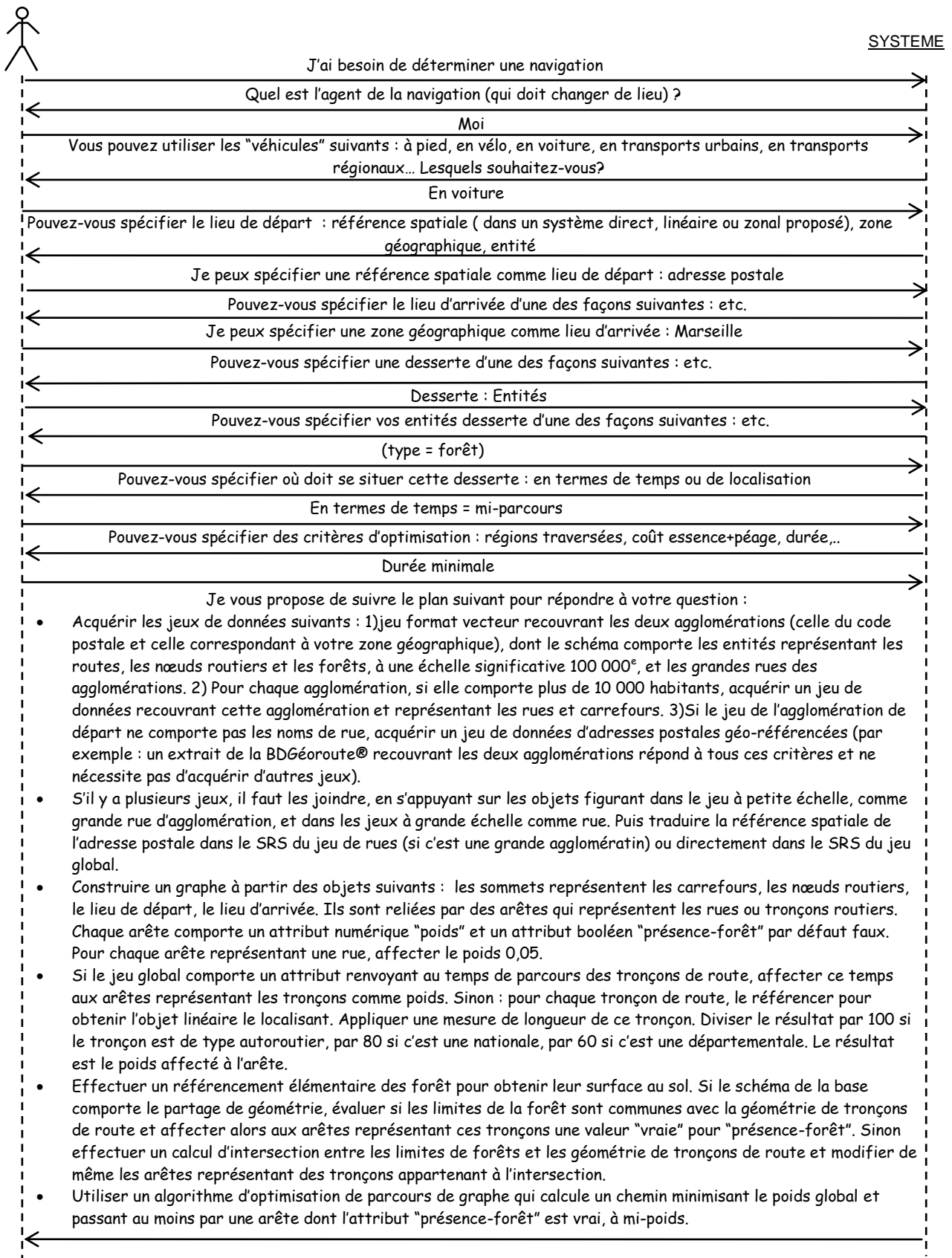


Figure 78 : diagramme de séquence s'appuyant sur le modèle de la tâche de détermination de navigation pour aider un utilisateur à déterminer un itinéraire. [Figure 78 : Sequences diagram of a users's access process relying on the navigation determination task.]

3.3 Opérationnalisation du modèle

Cette section présente un prototype implémentant le modèle présenté dans la section 3.1.2. L'enjeu est d'évaluer les fonctionnalités suivantes de notre système :

- permettre de stocker une base de connaissances correspondant au modèle des tâches géographiques. Cela nécessite d'une part de définir les éléments génériques utilisés pour stocker des tâches, plans, rôles et des éléments du domaine. Cela nécessite d'autre part de remplir la base en stockant des connaissances particulières en les décrivant à l'aide de ces éléments. Ces connaissances particulières seront reprises des sections précédentes.
- permettre de coder les fonctionnalités voulues du système expert manipulant cette base de connaissances pour aider l'accès d'un utilisateur, c'est-à-dire aider à la sélection d'information utile.

Nous choisissons de représenter notre modèle en orienté-objet car les langages orienté-objet sont des langages de représentation proches de la modélisation. Nous choisissons le langage de programmation orienté-objet java. C'est un langage simple à utiliser et qui se prête bien à la réalisation d'interfaces grâce à une bibliothèque de composants.

3.3.1 La représentation du modèle de tâches géographiques

3.3.1.1 La représentation du domaine

Un défaut de java est la non persistance des objets dans le code. Pour pallier cela, java offre divers mécanismes pour conserver des objets entre deux sessions d'une application.

Par exemple, le jdbc est une bibliothèque de composants qui permet de stocker des objets java sous forme de tables relationnelles.

Un autre mécanisme est la sérialisation qui consiste à écrire des objets dans un fichier lors d'une session. A la session suivante, il est possible de relire ces objets à condition de connaître leurs définitions de classes et que ces définitions implémentent certaines méthodes. Toutefois les références des objets ne sont pas conservées. L'intérêt de la sérialisation, outre sa facilité d'emploi, est qu'il est possible de sérialiser des objets sous forme de fichiers XML, aisément lisibles.

Dans le cadre du prototype, nous choisissons de nous appuyer sur la sérialisation pour stocker le domaine. Pour cela, nous construisons une classe générique `EltDomaine` qui permet de stocker les éléments du domaine

comme des objets de cette classe. Nous utilisons, comme référence d'un objet sérialisé, une variable particulière de cet objet : son `nom`. Notons que certains éléments spécifiques seront représentés non pas comme des instances de `EltDomaine` mais comme des instances de sous-classes d'`EltDomaine`, lorsque ces éléments ont des méthodes spécifiques utiles. Il s'agit par exemple des classes `Schéma` et `Couverture`.

Un objet `EltDomaine` est décrit par sa variable `nom`, et par des objets `Propriété`.

La classe `Propriété` comporte plusieurs variables : le nom de la propriété, son arité, la classe dans laquelle elle prend ses valeurs et l'ensemble de ses valeurs. Par exemple, l'élément "localisation" a une propriété nommée "est incluse dans" qui prend pour valeur des éléments "localisation".

Un objet `EltDomaine` possède deux propriétés particulières nommées `representation` et `produitPar`. La propriété `representation` recouvre les liens avec d'autres éléments du domaine correspondant à une représentation de cet élément, c'est-à-dire une forme plus concrète opérationnelle pour l'utilisateur. Par exemple, un système de référence spatiale peut être représenté par un jeu de données géographiques de références ou par une carte. Notons que la sortie obtenue dans le plan d'exécution d'une tâche est généralement un élément du domaine qui se trouve dans la propriété `representation` de l'élément décrivant la sortie de cette tâche.

La propriété `produitPar` recouvre les liens avec les ressources dynamiques produisant cet élément. Par exemple, l'élément `localisation` est produit par la tâche complexe de localisation, une distance est produite par une mesure de distance, un bassin versant est produit par une tâche primitive enchaînant des algorithmes spécifiques de calculs d'élévations.

La classe `Ensemble` joue un rôle clé dans les processus de spécification présentés par la suite. Comme nous l'avons introduit dans notre langage de description de tâches, un ensemble peut être défini de façon intensive ou extensive. Jusqu'ici nous avons prévu la représentation d'un type simple d'intension mais il sera possible par la suite d'en rajouter d'autres. Cela consiste à désigner un élément `e` du domaine tel que l'ensemble est l'ensemble des éléments du domaine, explicites ou non dans le domaine, qui correspondent à des cas particuliers (des spécialisations) de `e`.

L'élément peut être précisé en contraignant les valeurs de ses propriétés. Une variable de l'élément permet de signaler si certaines de ses propriétés ont été contraintes par rapport à sa définition initiale.

La notion d'intension est un intermédiaire important entre le discours de l'utilisateur et le domaine géographique du système : si l'utilisateur utilise le terme « entité », le système comprend « un élément appartenant à l'ensemble des entités que je connais dans mon domaine ». Un cas particulier est la description de jeux de données. Cela peut être des données disponibles ou dérivées. L'ensemble des données est décrit par un objet :

```
descriptionDonnées = new Intension() ;
descriptionDonnées.element = new EltDomaine("JeuDonnees") ;
descriptionDonnées.element.proprietesContraintes =
    {« echelle », « couverture », ...} ;
```

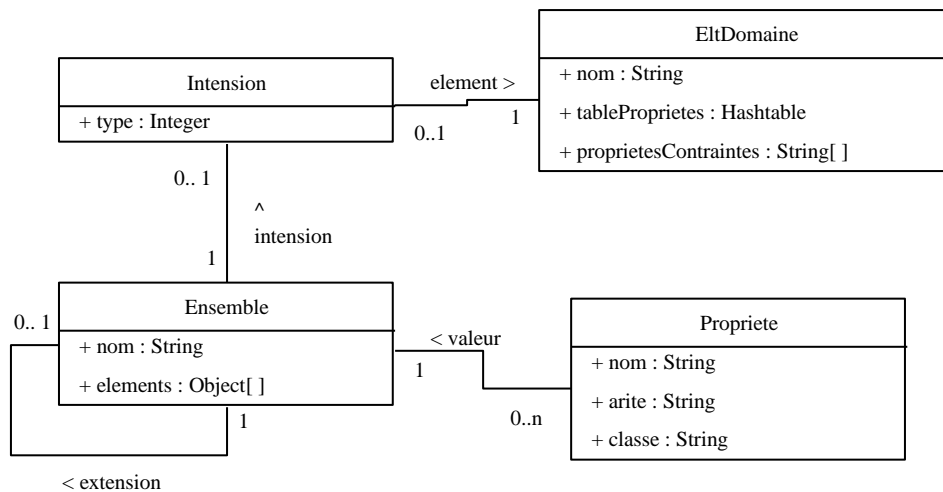


Figure 79 : Classes utilisées dans la représentation des éléments du domaine géographique.

3.3.1.2 La représentation des tâches

Concernant la représentation des tâches nous reprenons les éléments définis dans notre langage, section 3.1.2.5, ayant des caractéristiques communes comme tâche, tâche complexe, méthode, rôle, pour définir des classes décrites Figure 80 :

- la classe `Tache`, étendue par les classes `TacheComplexe` et `TachePrimitive`,
- la classe `Role`, décrite par son nom, les valeurs prises par le rôle dans le domaine (la variable candidat) et la tâche à laquelle le rôle est attaché.
- la classe `Methode`, décrite par un plan plus ou moins spécifié, qui est une structure de tâches, et des objets `Regle`.
- la classe `StructureTaches`, définie par un type, un vocabulaire, des sous-tâches et des sous-structures. Pour l'instant nous avons prévu les cas où le type de la structure est "ou" ou "et". Il nous faudra définir par la suite les autres types de termes d'enchaînement.

La variable `vocabulaireStructure` est une table de hachage dont les clés sont des mots. Ces mots sont le vocabulaire de détermination de la tâche dont la méthode a pour plan la structure. Comme dans le langage de description, les objets associés aux clés dans la table peuvent être de plusieurs types. Si la structure est complexe et se décompose en sous-structures, les objets sont des tableaux de mots, qui décrivent en fait, pour un terme du vocabulaire de la structure principale, les termes auxquels ils renvoient dans le vocabulaire de chaque sous-structure. Si la structure se décompose en sous-tâches, les objets associées aux clés sont des rôles des sous-tâches.

De plus, dans un plan, une sortie d'une sous-tâche peut être une entrée d'une autre. Il faut représenter cette dépendance entre rôles. Nous avons représenté une forme très simple de dépendance entre rôles : la variable `candidats` d'un rôle prend la valeur de la variable `candidats` de son `roleAmont` s'il en a un. Par exemple, dans la tâche de localisation, un pas consistant à acquérir un système de référence spatiale est suivi d'un pas consistant à produire une référence spatiale : la valeur de la sortie "système de référence spatiale acquis" du premier pas est la valeur de l'entrée "système de référence dans lequel produire la référence" du pas suivant. Il peut exister des formes de dépendance plus complexes, qu'il faudra par la suite représenter au cas par cas.

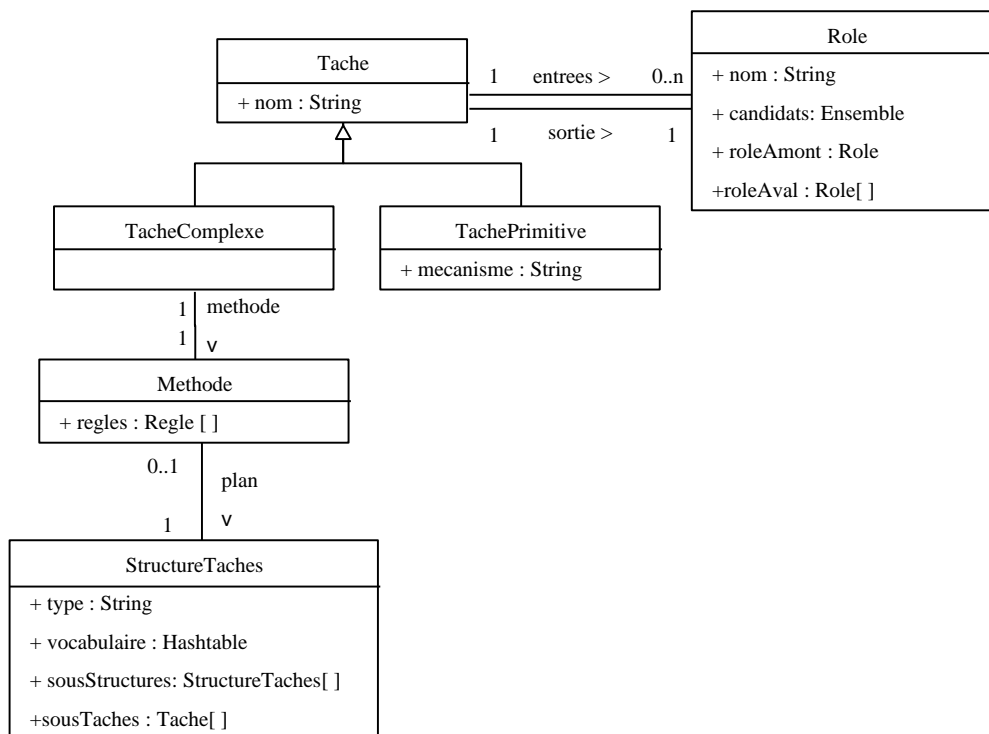


Figure 80 : Classes utilisées pour représenter les tâches géographiques.

Une contrainte majeure dans la représentation des tâches géographiques est qu'elles doivent être des instances d'une même classe et ne doivent pas comporter de connaissances procédurales spécifiques. Les connaissances procédurales de spécification de toute tâche géographique particulière doivent être décrites de façon indépendante dans une même tâche générique.

Dans la suite, nous ne présentons pas l'écriture des tâches de localisation, d'acquisition d'un système de référence et de production d'une référence spatiale, car ces écritures Java ne sont pas lisibles. Elles comportent des connaissances déclaratives qui doivent être associées à des connaissances procédurales pour prendre leur sens. Ces connaissances procédurales sont écrites dans des classes génériques indépendantes des objets décrivant les tâches géographiques.

Pour représenter les `Regle`, nous introduisons plusieurs classes.

La prémisses est décrite comme des tests. Ces tests doivent être appliqués sur un objet `SpecificationRole` décrivant la spécification qui a conduit à activer la méthode de la tâche.

Les variables de `SpecificationRole` se décomposent en des variables dédiées à l'identification du rôle spécifié et une variable dédiée à la description de la spécification même des candidats de ce rôle, de type `SpecificationEnsemble`. Nous ne détaillons pas cette dernière classe qui est spécifique à notre représentation du domaine et des ensembles.

Pour ce qui est des actions possibles, nous avons précisé dans la grammaire BNF qu'elles pouvaient être de plusieurs types. Ces actions sont décrites comme autant de variables différentes au niveau de l'objet `Regle`. Ainsi la modification de plan est décrite comme un entier qui est le numéro de la sous-structure choisie dans le tableau `sousStructures` de l'objet `StructureTaches` décrivant le plan. Les variables `propagSpecif` décrivent comme la spécification de rôle qui a conduit à appliquer les règles est propagée à la spécification d'autres rôles.

Nous ne détaillons pas ici la classe `PropagationSpecif`, elle comporte de nombreuses variables. Certaines sont dédiées à la désignation des rôles concernés par la propagation. D'autres sont dédiées à la description de la spécification même (spécification des ensembles candidats des rôles désignés), ou plus exactement de la définition de cette spécification à partir de la spécification initiale.

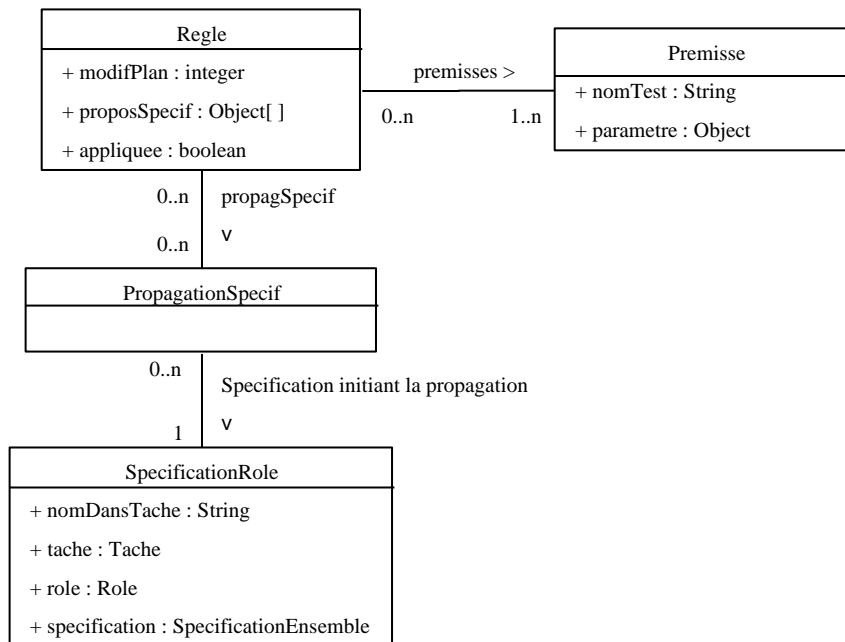


Figure 81: Classes utilisées pour représenter les connaissances de spécification géographiques.

3.3.2 La dynamique associée au modèle géographique

Il ne suffit pas de stocker des tâches géographiques, il faut encore pouvoir les manipuler dans le raisonnement de l'accès : « traduire un besoin d'utilisateur en sélection d'information utile, et en une désignation de ressources utiles pour dériver cette information. » Ce raisonnement doit être réparti entre un système expert et un utilisateur. Il est donc important de coder les fonctionnalités du système expert correspondant à sa contribution au raisonnement, présentées section 3.3.2.1, et de coder les interfaces supportant les communications nécessaires entre l'utilisateur et le système lors du raisonnement global, présentées section 3.3.2.2.

3.3.2.1 Mécanismes de spécification

Principe

Rappelons le principe de l'exploitation du modèle de tâches géographiques dans notre système. Le besoin de l'utilisateur est représenté comme une `TacheComplexe`. La réponse à ce besoin est donnée par cette même tâche, une fois spécifiée :

- la facette déclarative décrit le besoin et le contexte de l'utilisateur (la « sélection d'information utile » est exprimée dans la variable `sortie`),
- la facette opérationnelle décrit la réponse (la « sélection de ressources utiles » est représentée dans les `entree` des tâches primitives composant le plan détaillé de la tâche qui n'ont pas de `roleAMont`).

En définitive, le processus d'aide à l'accès consiste à spécifier une tâche qui représente à la fois le besoin de l'utilisateur -facette déclarative de la tâche- et la réponse à ce besoin -plan d'exécution de la tâche-

Le système possède des connaissances de spécification géographiques, et des connaissances de spécification génériques.

Les connaissances géographiques sont les règles des méthodes des tâches. Elles sont représentées de façon déclarative, et ont été décrites dans la section précédente.

Les connaissances génériques sont les connaissances de manipulation de tâches comme le déclenchement d'une règle quand un rôle est spécifié, la propagation de la spécification d'un rôle du vocabulaire d'une structure aux rôles des sous-structures ou sous-tâches, la propagation de la spécification d'un rôle aux rôles éventuels dont il est rôle amont. Ces connaissances ont été décrites dans la section 3.1.3. Nous détaillons dans la suite l'implémentation de ces connaissances.

Algorithmes utilisés

Une première étape pour spécifier le côté déclaratif du besoin consiste à représenter ce besoin comme une instance d'une classe représentant une tâche géographique connue. Le système présente les `TacheComplexe` qu'il connaît en décrivant leur `sortie`, l'utilisateur choisit un de ces objectifs géographiques génériques, et le système crée l'objet `besoinUtilisateur = Tache_i.copie()`, où `Tache_i` est la classe correspondant à l'objectif sélectionné par l'utilisateur. Le `besoinUtilisateur` devient la `tache courante` et sa méthode la `methode courante`.

Les étapes suivantes consistent à spécifier l'objet `tache courante`. Pour cela, le système propose à l'utilisateur une interface, décrite dans la section suivante, qui présente les rôles de `tache courante` et les éléments de spécification de ces rôles. La mise en place de cette interface au cours de la session se fait grâce à la lecture des rôles de `tache courante` et à la lecture des éléments du domaine impliqués dans la définition des valeurs de ces rôles. Spécifier un rôle revient à restreindre les ensembles candidats de ces rôles. Nous détaillons également dans la section 3.3.2.2, comment ces mécanismes de spécification sont effectivement accessibles à l'utilisateur, c'est-à-dire par exemple comment il lui est possible de restreindre une extension, de choisir une propriété à contraindre.

Chaque fois que l'utilisateur fait une action de spécification sur cette interface, le système construit un objet `specificationCourante` de type `specificationRole` qui reflète la spécification faite.

Ensuite, le système lit les prémisses de chaque règle `regle_j` de la méthode `methode courante`. La lecture d'une prémisses se fait en lisant le nom de la méthode de test de la classe `specificationRole` stocké dans la variable `nomTest` de la prémisses et en appliquant cette méthode sur la variable `parametre` de la prémisses. Les différentes méthodes peuvent consister à tester le nom du rôle spécifié, et la type de spécification d'ensemble faite.

Lorsque les prémisses d'une règle sont vérifiées, le système applique les actions décrites dans les variables de cette même règle.

Certaines variables permettent de modifier l'interface pour suggérer la spécification d'autres rôles à l'utilisateur. Nous ne les décrivons pas davantage ici.

Si la variable `modifPlan` est non nulle (`modifPlan=i`), le système modifie le plan de la méthode courante. Pour cela, il attribue à cette variable la valeur de sa sous-structure d'indice `i`. De plus, il réécrit le vocabulaire de cette sous-structure en lui ajoutant les termes du vocabulaire de la sous-structure précédente qui pointaient sur des éléments de cette sous-structure. Ceci est nécessaire pour que ce vocabulaire soit cohérent avec celui utilisé au niveau des règles de la méthode.

Si la variable `propagSpecif` est non nulle, le système sélectionne tour à tour chaque élément `PropagationSpecif` de ce tableau, soit `PSi`, donne à sa variable `specApropager` la valeur `specificationCourante` et applique `PSi`.

L'application d'une `propagationSpecif` se fait de la façon suivante.

Le système détermine les rôles qui doivent être spécifiés. Cela peut se faire de différentes façons :

- Si les variables `nomIndirect` et `nomDansStructure` de `PSi` sont nulles, cela signifie que le rôle à spécifier est désigné directement par son nom dans `tache courante` et la valeur de ce nom est la variable `nomDansTache` de `PSi`.
- Si la variable `nomDansStructure` de `PSi` est non nulle, cela signifie que le rôle à spécifier est désigné par son nom dans le `vocabulaireStructure` de la variable `plan de methode courante`. Si les sous-tâches du plan sont non nulles, les rôles à spécifier sont désignés directement par le terme du vocabulaire. Si les sous-structures sont non nulles, le terme du vocabulaire désigne des noms dans le vocabulaire des sous-structures et le système doit faire une itération pour obtenir les rôles à spécifier.
- Si la variable `nomIndirect` est non nulle, le système effectue le même processus que celui effectué ci-dessus avec la variable `nomDansStructure` pour obtenir les rôles désignés dans le plan par `nomIndirect`. Ces rôles ne sont pas les rôles à spécifier mais ils sont associés aux tâches auxquelles sont associés les rôles à spécifier. Ces derniers sont obtenus en sélectionnant pour chacune de ces sous-tâches le rôle nommé `nomDansTache` (variable de `PSi`) dans cette tâche.

Le système construit alors autant d'objets `SpecificationRole` qu'il y a de rôles à spécifier, en affectant à chacune, comme variable `role`, un des rôles à spécifier.

Le système construit ensuite la spécification d'ensemble qui doit être appliquée à chacun de ces rôles, c'est-à-dire la valeur des variables `specification` des objets `SpecificationRole` qu'il a construits. Il construit un objet `SpecificationEnsemble` (soit `SE`) dont les principales variables ont les mêmes valeurs que celles de la spécification d'ensemble initiale pour certaines, ou que des valeurs de variables spécifiques de `PSi`. Cette spécification d'ensemble initiale est la variable `specification` de la `specApropager`. Puis le système précise certaines variables de `SE`, selon plusieurs étapes :

- Si la variable booléenne `paramVautChemin` de `PSi` est vraie, cela signifie que la variable `parametre` de `SE` doit être déduite de l'ensemble initialement spécifié en suivant la variable `cheminBis` de `PSi`. Par exemple, si un ensemble a été spécifié comme étant défini intensionnellement par "une référence spatiale définie dans un système de référence administratif", et que le `cheminBis` vaut "définie dans" alors le `parametre` vaut l'élément "système de référence administratif". Ce processus ne consiste bien sûr pas à faire des opérations sur des chaînes de caractères comme cet exemple peut le laisser croire. Il s'agit plutôt de sélectionner par son nom une propriété d'un élément du domaine, de sélectionner la valeur de cette propriété, de sélectionner l'élément définissant éventuellement intensionnellement cet ensemble, et d'itérer cela autant de fois qu'il y a de noms de propriété dans le `cheminBis`.

- Si la variable booléenne `paramVautChoix` de `PSi` est vraie alors la variable `parametre` de `SE` vaut le choix définissant la spécification d'ensemble initiale. Par exemple, si un ensemble a été spécifié comme étant défini intensionnellement par "une entité dont la valeur est bâtiment", le choix est "bâtiment".
- Le système utilise ensuite la variable `classParam` de `PSi` pour donner le type souhaité au `parametre`.
- Le système procède de façon similaire pour la variable `element` de `SE`, en fonction des booléens `elementVautChemin` et `elementVautChoix`.

Ensuite, le système affecte aux variables `specification` de chaque objet `SpecificationRole` créé une variable `tacheAmont`. Cette variable vaut soit la tâche même associée au rôle à spécifier, soit une tâche désignée par la variable `nomIndirectTacheAmont` de `PSi`, qui à l'instar de la variable `nomIndirect`, permet de désigner une tâche par un rôle -qui a pour nom `nomIndirectTacheAmont` dans le vocabulaire de méthode courante- auquel elle est associée.

Puis, chaque objet `SpecificationRole` est appliqué.

L'application d'une `SpecificationRole` consiste dans un premier temps à initialiser la variable `role` si elle est nulle. Pour cela le système utilise les variables `nom` et `tache`.

Puis, si le rôle est non nul, le système applique la spécification d'ensemble (variable `specification`) aux candidats du rôle.

Si le rôle est nul, cela signifie qu'il n'existe pas encore dans la tâche et qu'il faut non pas le spécifier mais le créer. Dans ce cas le système crée un ensemble reflétant la spécification -méthode spécifique de la classe `SpecificationEnsemble`- et ajoute à la tâche, désignée par la variable `tache`, un rôle nommé `nom` qui a pour candidat cet ensemble.

Puis cette spécification appliquée devient `specification courante`, c'est-à-dire que le système étudie les règles de la méthode de la tâche correspondant au rôle spécifié et déclenche les règles applicables (comme nous venons de le décrire).

Enfin, si le rôle a des rôles aval, c'est-à-dire des rôles dont les variables candidats ont la même valeur que ses candidats, alors le système crée un objet `SpecificationRole` rendant compte de la spécification de chaque rôle aval et ces objets deviennent tour à tour `specification courante`.

L'application d'une `SpecificationEnsemble` peut consister à modifier un ensemble existant ou à créer un nouvel ensemble.

Avant d'appliquer une `SpecificationEnsemble`, le système réinitialise éventuellement les variables `element`, `parametre` et `ensemble` de cet objet de la façon suivante. Si les variables `paramDependant` (resp `elementDependant`, `ensembleDependant`) et `nomRoleAmont` ne sont pas nulles, le système sélectionne le rôle nommé `nomRoleAmont` de la tâche désignée par la variable `tacheAmont`. Il définit alors le `parametre` (resp `element`, `ensemble`) comme un ensemble qui vaut ces candidats.

Ensuite, le système applique la spécification en fonction de la valeur de la variable `type` de l'objet `SpecificationEnsemble`.

S'il vaut "1", cela signifie que la spécification consiste à créer un ensemble. Sa valeur est donnée par la variable `ensemble`. Si la spécification consiste à modifier un ensemble défini intensionnellement, cet ensemble créé est la valeur de la propriété désignée par la variable `chemin`.

S'il vaut "2", cela signifie que la spécification consiste à restreindre une extension en lui donnant une valeur indiquée. Cette extension peut éventuellement être la valeur d'une propriété désignée par la variable `chemin`.

S'il vaut "3", cela signifie que la spécification fait appel à une méthode spécifique. C'est le cas généralement lorsque la spécification porte sur un ensemble défini intensionnellement par l'élément du domaine "jeu de données".

S'il vaut "4", cela signifie que la spécification s'appuie sur une propriété. C'est le cas par exemple lorsqu'un ensemble est défini intensionnellement comme "référence spatiale". Une spécification possible peut être décrite par une `SpecificationEnsemble` de type 4 ayant pour `nomPropriete` "est définie dans" et pour variable `nomElt2` "système de référence administrative". Dans ce cas l'application consiste à lire la propriété "référence spatiale - est définie dans - système de référence spatiale" pour définir que l'ensemble est spécifié comme "entité administrative".

En définitive, le besoin est traduit, dans le plan, en une requête aboutie : l'intension des ressources recherchées. Notons que les utilisateurs familiarisés avec les données et traitements peuvent exprimer leur besoin en décrivant directement les ressources recherchées, par leurs intensions.

Le diagramme de séquence Figure 82 illustre ces mécanismes.

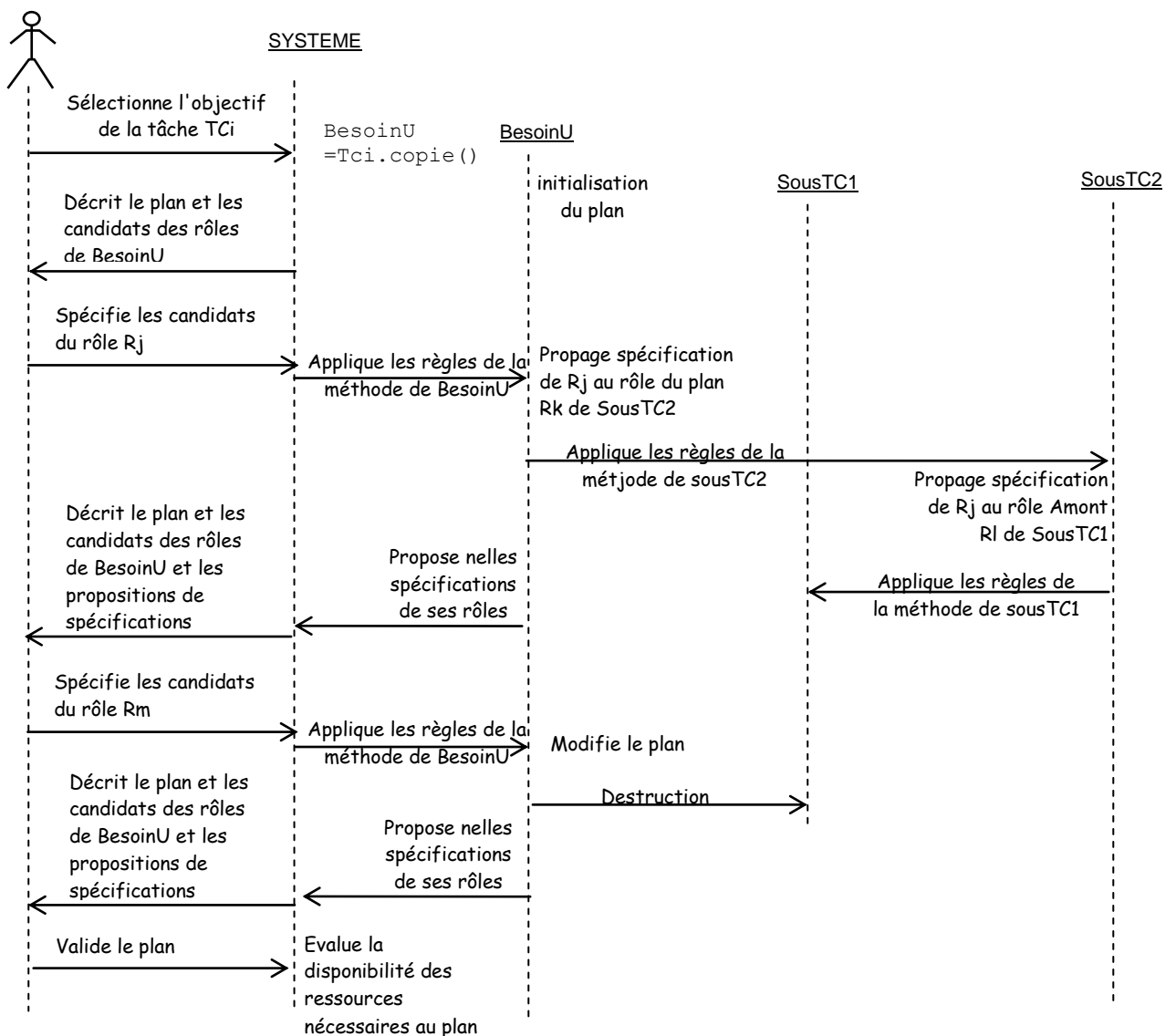


Figure 82 : Mécanismes de spécification d'un besoin d'utilisateur. [Figure 82 : Mechanisms to specify a user's need.]

3.3.2.2 L'interface

Swing est une bibliothèque java contenant des composants servant à construire des interfaces graphiques. Une grande partie des traitements du système reposeront sur des objets swing permettant à l'utilisateur de visualiser les concepts du modèle de tâches géographiques.

Notons que l'utilisateur de notre système est presque directement en contact avec le modèle de la base de connaissance de notre système. L'interface graphique affiche ces concepts plus qu'elle ne les traduit dans un langage intuitif. Nous ne cherchons pas à gérer cet aspect rigoureusement c'est-à-dire en mettant en place un véritable dialogue homme-machine faisant appel à des techniques complexes. Ceci est l'objet d'un travail de thèse actuellement en cours au COGIT et nous y revenons plus loin.

Pour compenser un peu le manque de convivialité de notre système nous guidons l'utilisateur à tout moment dans le système par une fenêtre d'aide : cette fenêtre lui donne des définitions simples des concepts employés par l'interface, lui explique ce qui se passe quand il clique sur un bouton ou sélectionne un nœud d'un arbre.

Présentation des objectifs géographiques connus

L'interface de présentation des objectifs connus est une fenêtre permettant à l'utilisateur de naviguer dans un arbre des objectifs connus du système. Il s'agit du cadre "VUE SUR LA BASE DE CONNAISSANCES DU SYSTEME" Figure 83. Cet arbre est organisé en associant à chaque objectif des objectifs qui le spécialisent. Il peut être redondant. De plus, l'utilisateur peut consulter une description de la tâche. Cette description possède des champs génériques et également des commentaires libres entrés dans la définition de la classe. Cette fenêtre est implémentée comme la classe `AfficheObjectifs`.

Spécification de la tâche décrivant le besoin de l'utilisateur

La première étape de spécification est le choix d'une tâche connue du système et qui est proche du besoin de l'utilisateur. Cela est fait à partir de la fenêtre d'ouverture de session, apparaissant, en suivant les instructions apparaissant dans le cadre orange « à partir de l'étape sélectionnée vous pouvez : ». Cette fenêtre est implémentée dans la classe `FenetreLancement`.

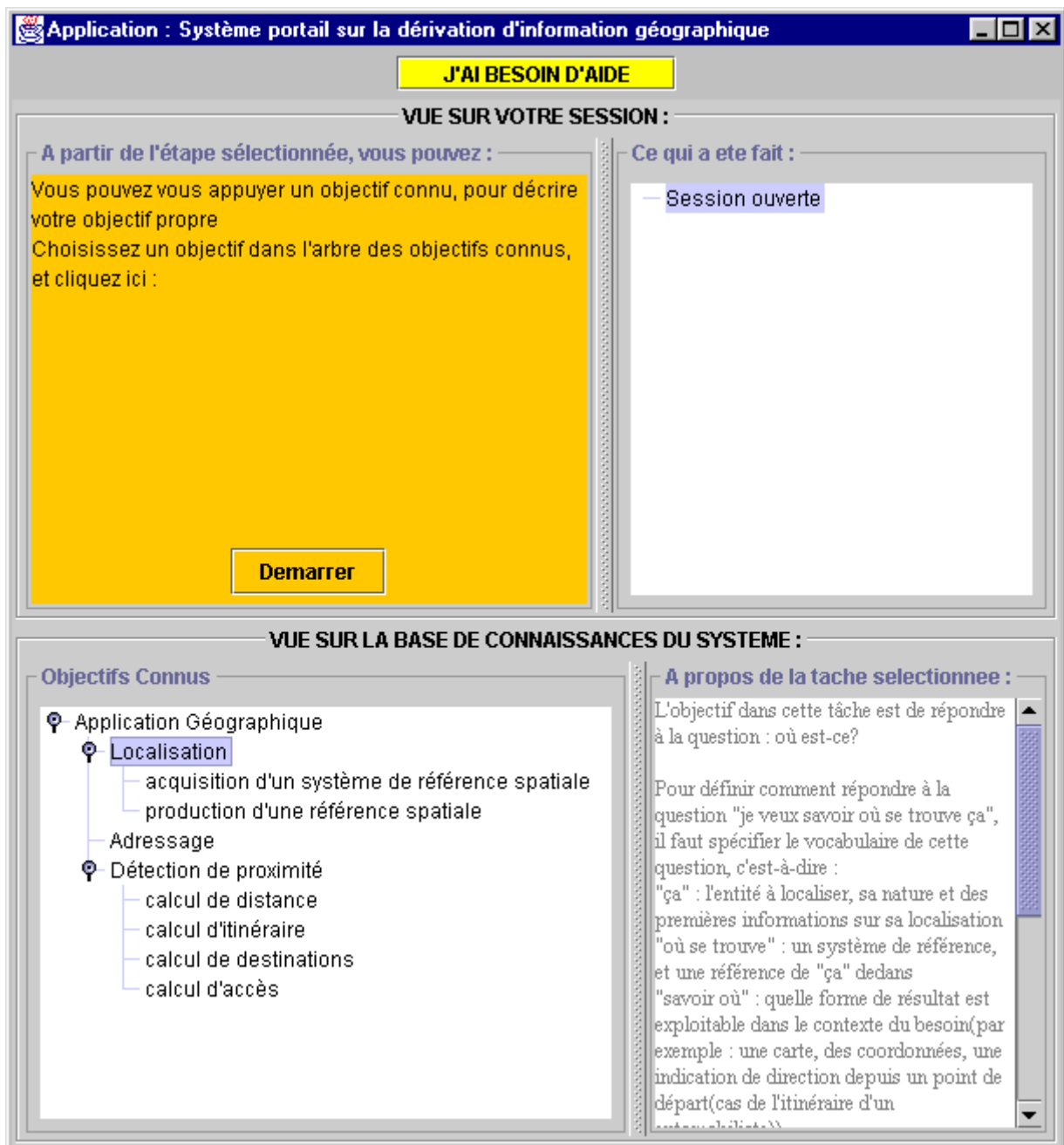


Figure 83 : Fenêtre d'ouverture d'une session. [Figure 83 : Opening session window.]

Lorsque l'utilisateur clique sur le bouton « démarrer », le système crée l'objet `besoinUtilisateur`.

Il crée aussi une fenêtre présentant automatiquement les rôles que l'utilisateur peut spécifier. Une première fenêtre, Figure 84, donne la liste des rôles de la tâche : entrées et sorties.

Pour chaque rôle, l'utilisateur peut éditer une description de sa variable `candidats`, c'est-à-dire l'ensemble des éléments du domaine remplissant potentiellement ce rôle. Cette description est également construite automatiquement lors de la session. Sur la Figure 84, l'ensemble édité est l'ensemble correspondant au rôle « localisation recherchée ». Cette description de l'ensemble `candidats` est associée à un commentaire expliquant à l'utilisateur l'importance de ce rôle dans la détermination de la tâche, et suggérant éventuellement des choix de spécification.

L'utilisateur peut décider de spécifier un rôle plutôt qu'un autre en le sélectionnant dans la liste des rôles. Il peut aussi se laisser guider par le système qui lui propose de spécifier les rôles dans un certain ordre. Notons que lorsque l'utilisateur spécifie un rôle, le système va éventuellement propager cette spécification en ajoutant une suggestion de spécification d'un autre rôle, dans la fenêtre « informations pour vous aider à spécifier ».

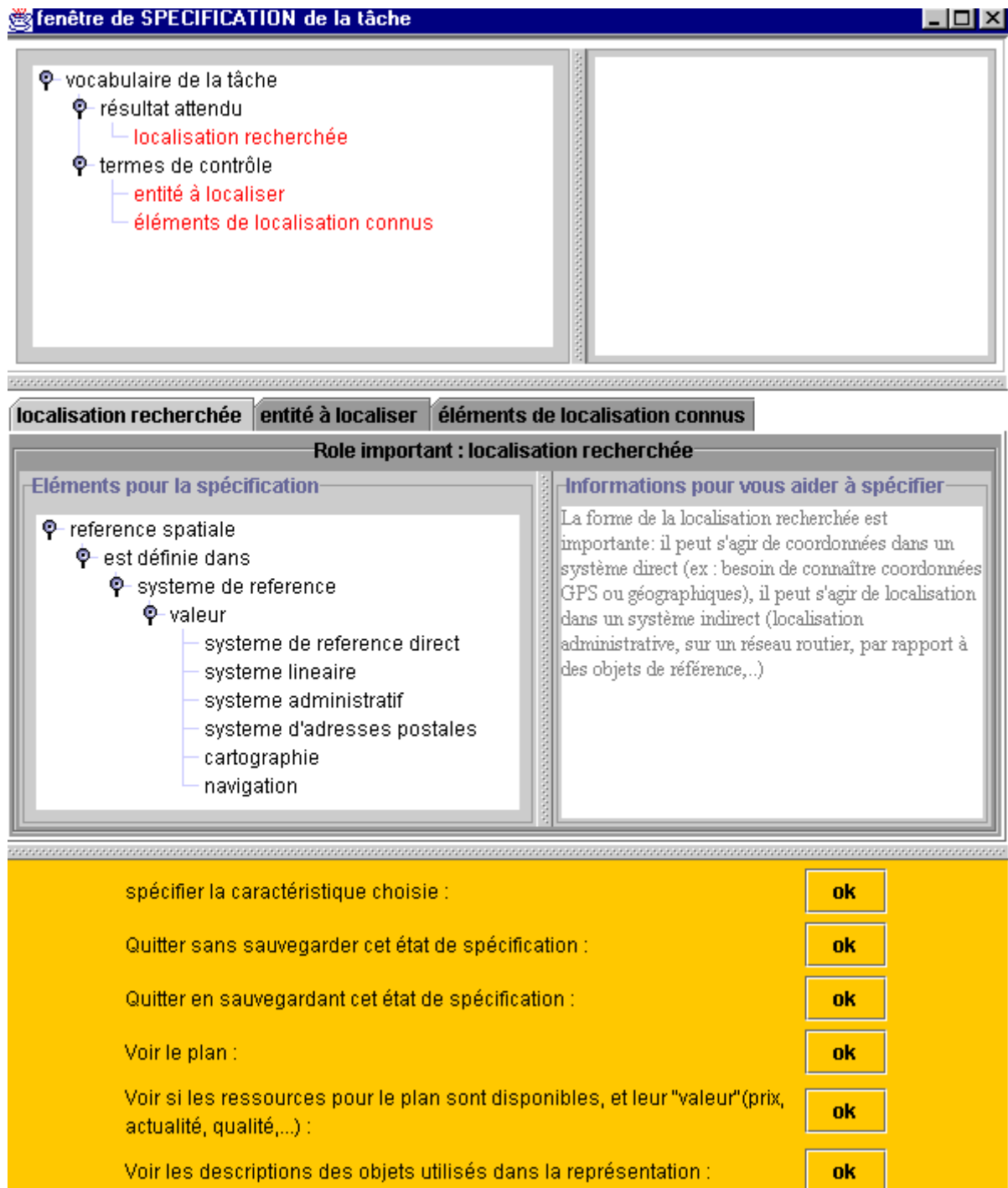


Figure 84 : Interface permettant à un utilisateur de spécifier la tâche représentant son besoin. [Figure 84 : Window allowing a user to specify the task representing his need.]

La spécification d'un ensemble peut se faire :

- soit sur une extension, en sélectionnant une ou plusieurs valeurs parmi celles comprises dans cet ensemble. Si la spécification ne peut porter que sur une valeur alors les autres valeurs ne sont plus sélectionnables par la suite.
- soit sur intension, en spécifiant une propriété de l'élément générique de l'ensemble.

Les termes nécessaires à la spécification n'apparaissent pas tous sur la fenêtre de spécification du rôle. Lorsque l'utilisateur veut spécifier un élément comme une localisation dans laquelle est contenue la localisation recherchée, ou une entité dont est proche l'entité qu'il veut référencer, une nouvelle fenêtre de dialogue s'ouvre.

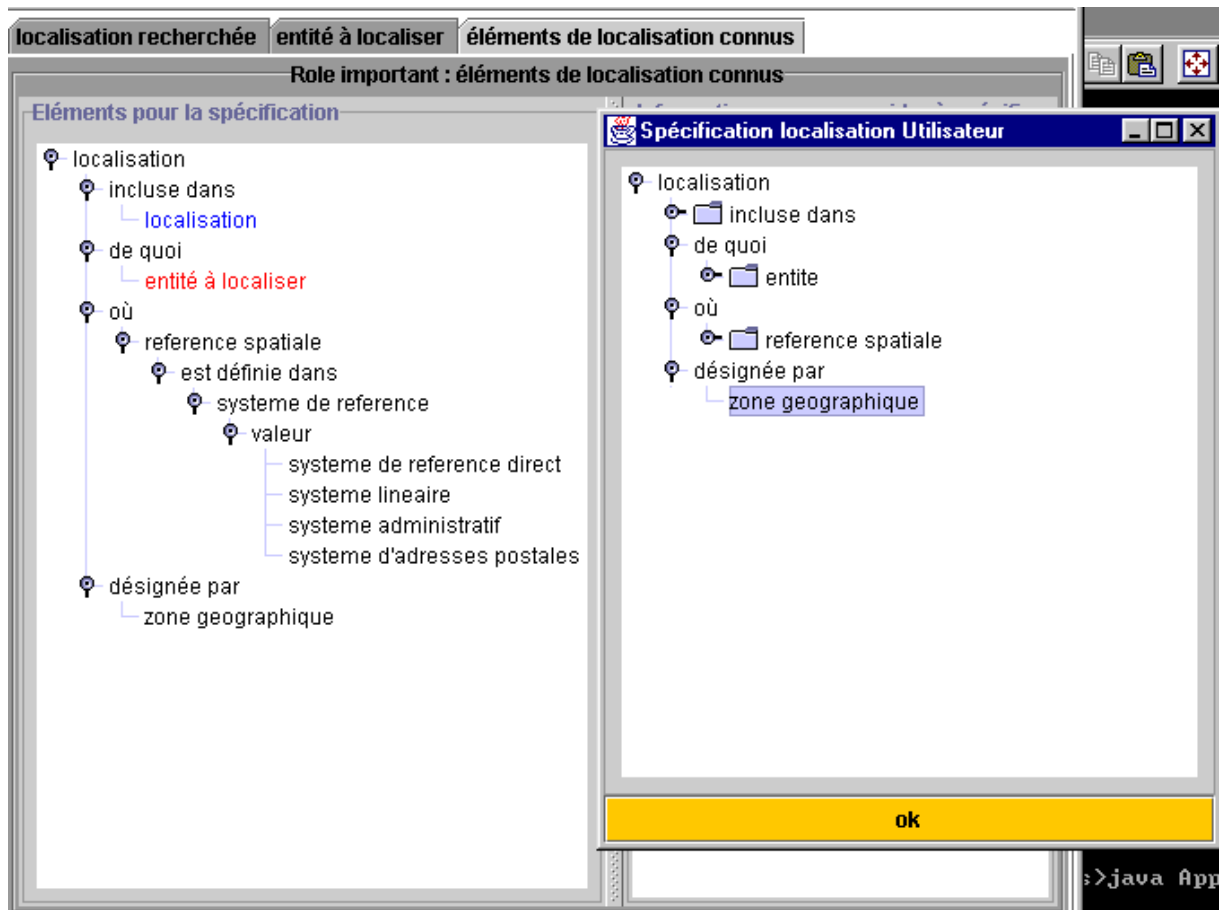


Figure 85 : Spécification de la localisation dans laquelle est incluse la localisation recherchée. Une fenêtre de dialogue s'ouvre pour poursuivre l'arbre de spécification. [Figure 85 : Specification of a localisation inside which is included the researched localisation. A dialog window opens to detail the process.]

Le système édite les spécifications faites en modifiant la description graphique de l'ensemble : les termes spécifiés apparaissent en bleu, les termes qui ne peuvent plus être sélectionnés apparaissent en gris clair, comme sur l'exemple Figure 86.

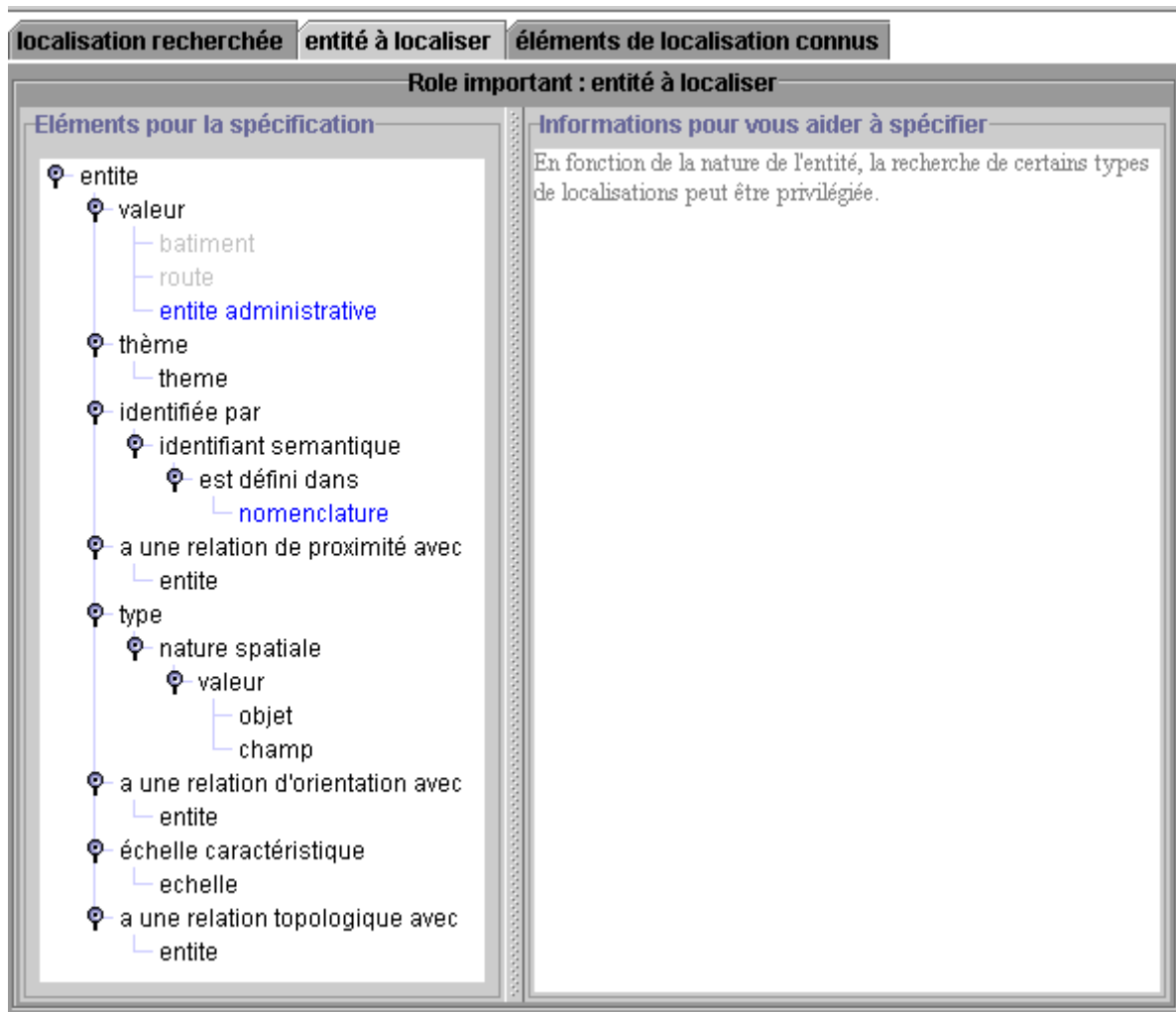


Figure 86 : Fenêtre de spécification du rôle entité à localiser de la tâche de localisation. Les valeurs de l'entité « bâtiment » et « route » ne peuvent plus être sélectionnées car l'utilisateur a spécifié qu'il s'agissait d'une entité administrative. [Figure 86 : Window for the specification of the role "entity to locate" of the localisation task. Values "bâtiment" and "route" can no more be selected since the user has already specified this entity was of type "entité administrative".]

Présentation du plan

La présentation du plan se fait sous forme textuelle, comme sur l'exemple Figure 87.

Ainsi, une tâche complexe figurant dans un plan n'est pas décrite au même niveau de détail que la tâche complexe principale (c'est-à-dire celle représentant le besoin de l'utilisateur). Son plan doit être résumé comme une chaîne de caractères de façon similaire au mécanisme d'une tâche primitive.

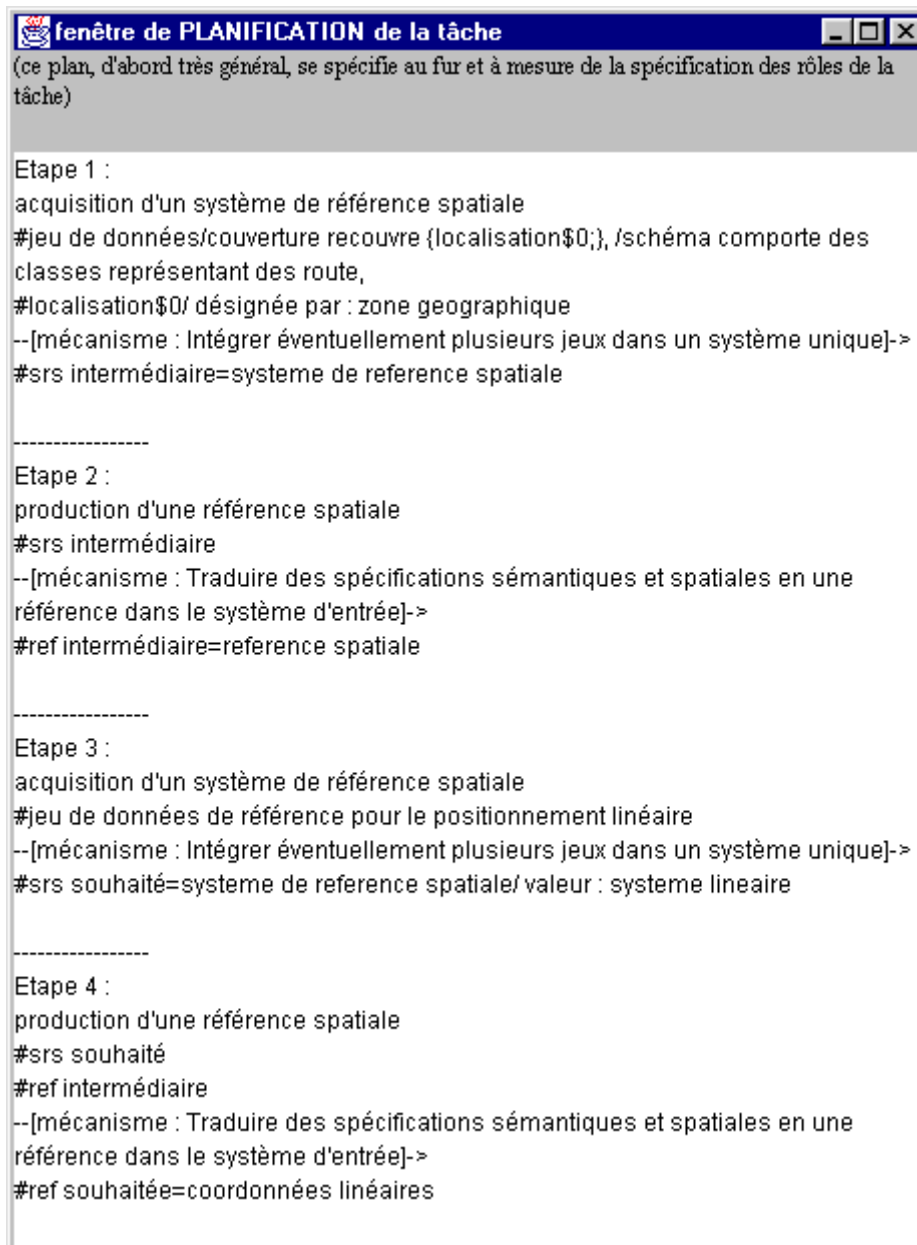


Figure 87 : Fenêtre de présentation du plan d'une tâche spécifiée. [Figure 87 : Window displaying the plan of the specified task.]

Le plan est rafraîchi lors de chaque spécification entrée par l'utilisateur. En particulier, des entrées peuvent être ajoutées au corps d'un pas pour décrire des spécifications avancées, comme "localisation\$0" sur l'exemple Figure 87. Les noms de ces rôles, en général "entite\$i", "localisation\$i", sont attribués grâce à un compteur du plan qui incrémente la variable *i*.

L'évolution du plan est sauvegardée lorsque l'utilisateur sauvegarde l'état de spécification de sa tâche de sorte que lorsque l'utilisateur souhaite revenir à un état de spécification sauvegardé il retrouve les spécifications des rôles et les spécifications du plan.

Navigation de l'utilisateur dans le processus de spécification

A tout moment du processus de spécification, l'utilisateur peut « prendre une photographie » des spécifications en l'état. Une fenêtre du système contient la mémoire de ces « photographies ». L'intérêt de cette mémoire est qu'il est à tout moment possible de revenir en arrière dans ces états des spécifications et de poursuivre la spécification à partir d'un état choisi dans les états passés.

Cette fenêtre, « vue sur votre session » est construite à partir d'une classe globale Historique dont une instance est créée à chaque session. Cette instance décrit la session et les « photographies » prises pendant cette session de la façon suivante :

- Elle utilise des instances de la classe Etape pour décrire les étapes de la session. Pour l'instant nous n'avons envisagé comme étapes que les sauvegardes de spécifications.
- Elle associe à chaque objet Etape une instance de la classe Agenda décrivant ce que l'utilisateur peut faire à partir de cette session. Pour l'instant l'agenda consiste toujours à relancer le processus de spécification.

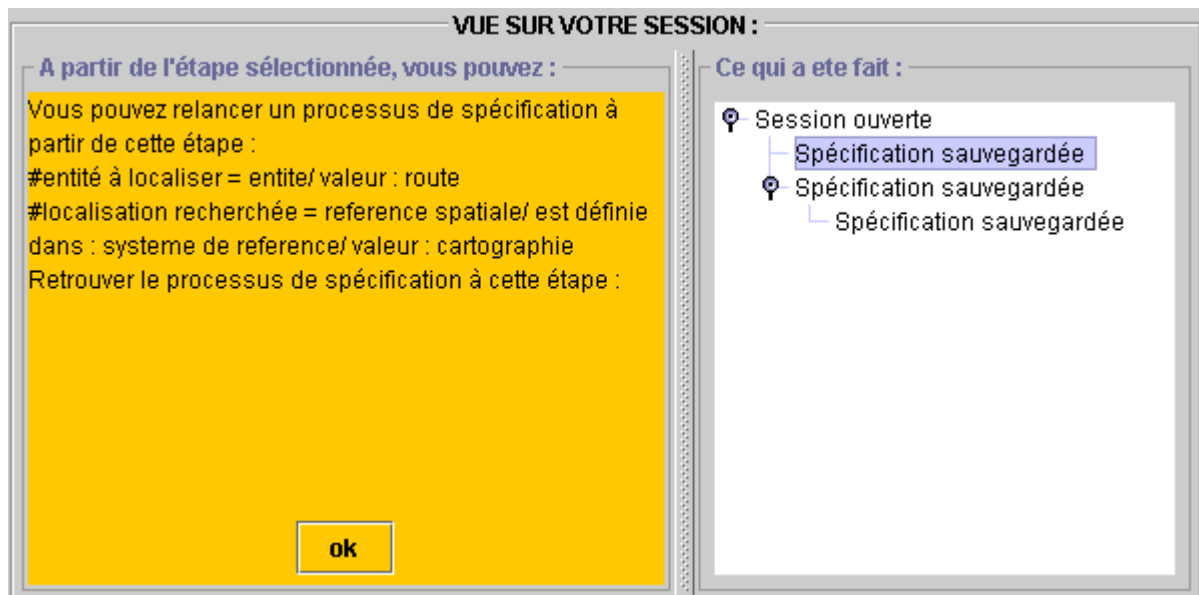


Figure 88 : Fenêtre permettant à l'utilisateur de naviguer dans divers états de spécification de son besoin. [Figure 88 : Window supporting the user's browsing of several states of specification of his need.]

La relance du processus de spécification repose sur une copie de la TacheUtilisateur (et de toutes ses variables) faite lorsque l'utilisateur décide de sauvegarder l'état des spécifications. Ainsi, à l'étape est associée une tâche reflétant exactement l'état des spécifications et dont les méthodes permettent de poursuivre le processus.

3.3.2.3 Implémentation de la tâche de Localisation

Ces structures ainsi ébauchées peuvent être exploitées de la façon suivante pour construire notre base de connaissances. Les étapes de stockage d'une tâche géographique connue sont les suivantes :

1. Une première étape est de se demander si le plan de cette tâche est toujours le même et s'il est possible de le construire simplement sans interaction complexe avec l'utilisateur. Dans ce cas il s'agit d'une instance de

TachePrimitive avec pour mécanisme le plan en question. Sinon c'est une instance de TacheComplexe, dont le plan a pour valeur le plan en question.

2. Dans le cas d'une tâche complexe, on crée une instance, Tache_i de TacheComplexe.
3. L'objectif de la tâche doit être énoncé de façon générique. Il est décrit à l'aide d'une variable ou d'une structure de variables. Des rôles correspondant aux noms de ces variables sont créés, la sortie de Tache_i est le tableau de ces rôles. Les Ensembles candidats de ces rôles sont décrits, éventuellement en ajoutant dans le domaine les instances d'EltDomaine nécessaires.

Le plan générique est décrit comme une décomposition (non nécessairement séquentielle) en sous-tâches.

Les variables du contexte amenant à spécifier le plan sont identifiées et spécifiées comme cela a été fait pour les rôles de sortie. Ces rôles sont ensuite rangés dans le tableau controle.

On écrit les règles de spécification.

4. Dans le cas d'une tâche primitive, on crée une instance de TachePrimitive et on procède de même pour spécifier ses divers rôles.

Nous détaillons dans les Figure 89 et Figure 90 un cas d'utilisation du système s'appuyant sur la tâche de localisation ainsi codée.

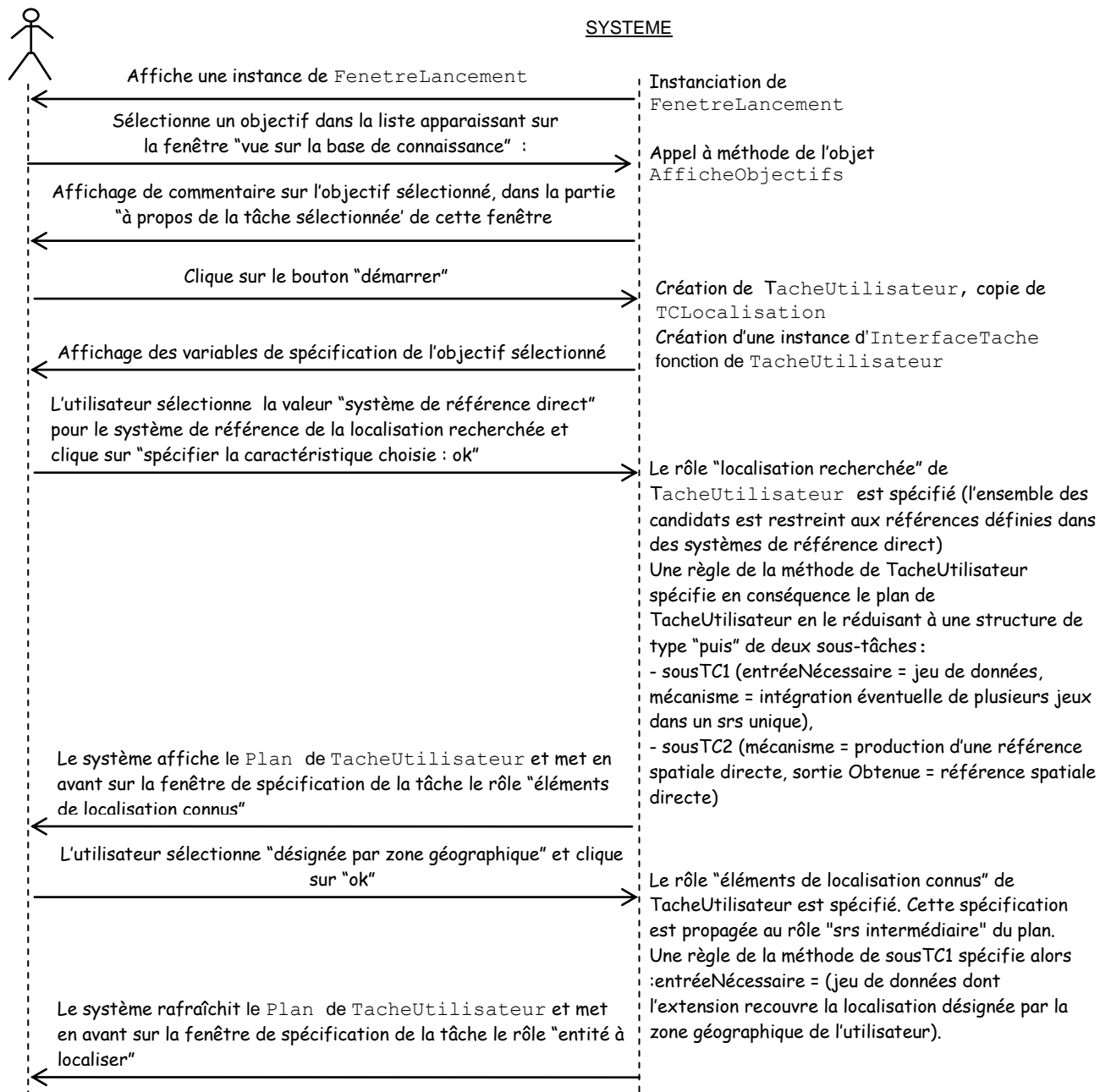


Figure 89 : Diagramme de séquence décrivant un scénario d'utilisation du système (poursuivi Figure 90). [Figure 89 : Sequence diagram describing an access relying on the localisation task (to be followed Figure 90)]

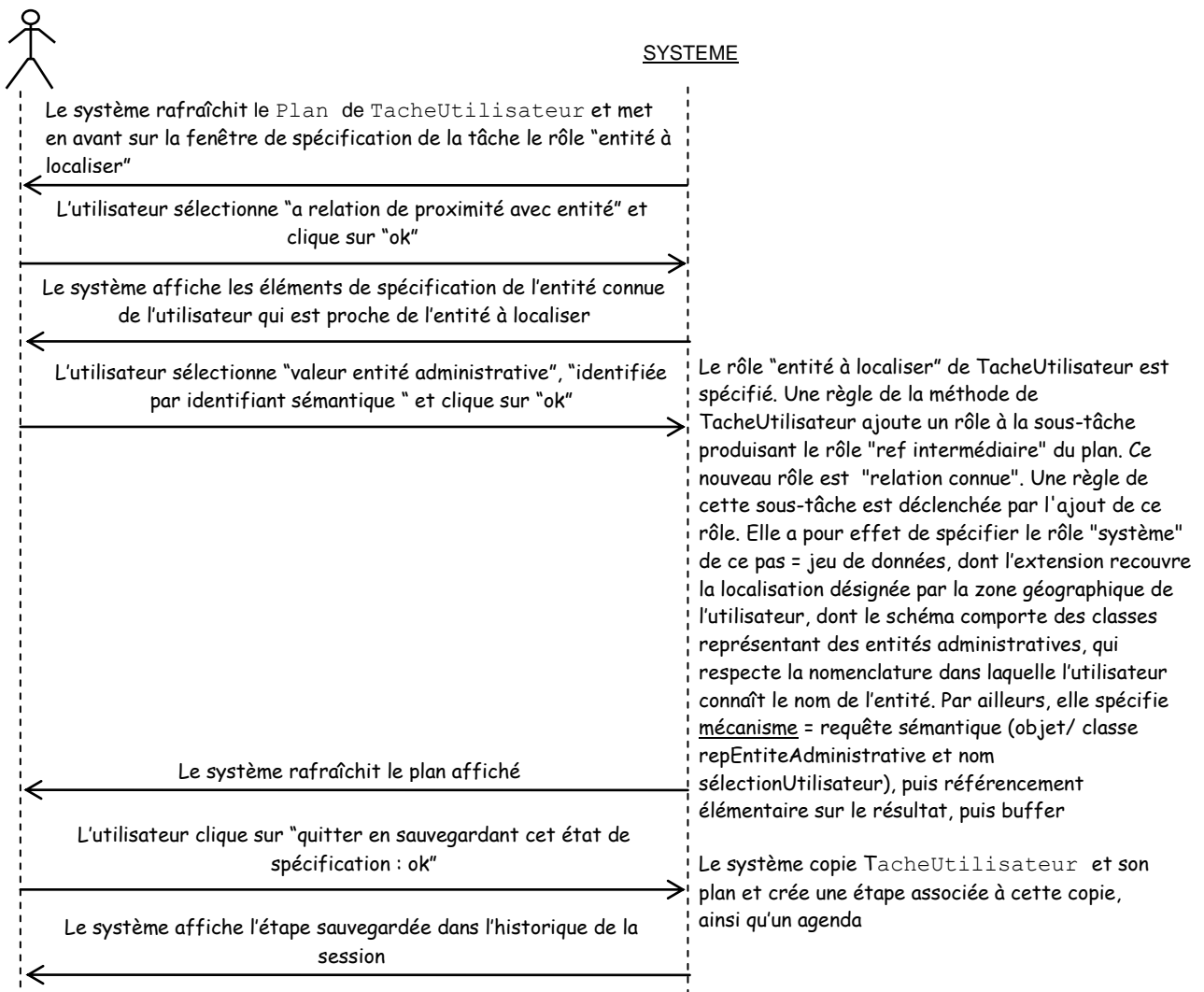


Figure 90 : Diagramme de séquence décrivant un scénario d'utilisation du système (suite de la Figure 89.)
 [Figure 90 : Sequence diagram following up the diagram Figure 89.]

3.3.3 Bilan

Les diverses sections de cette partie 3 ont présenté notre approche : ses principes fondateurs et la construction d'un prototype. A l'issue de cette section, nous devons évaluer :

- les apports de cette approche à l'accès d'utilisateurs aux données géographiques,
- l'application des principes dans le prototype.

3.3.3.1 Apports à la résolution de la problématique initiale

Pour évaluer notre approche, nous reprenons les 5 sous-objectifs identifiés lors de la définition de la problématique.

Contribution à O1 (le transfert de connaissances)

Rappelons que nous avons axé notre démarche sur ce sous-objectif O1. En principe, le service d'aide possède les connaissances permettant de désigner à un utilisateur l'existence de ressources, de décrire leur contenu et de fournir leur mode d'emploi.

Cette contribution doit être modulée en fonction de l'utilisateur. Un utilisateur peu familiarisé avec le monde des BDG et SIG aura besoin de cette aide optimale alors qu'un utilisateur plus expert n'aura besoin que de la désignation de l'existence des ressources et de leur contenu.

Contribution à O2 (l'aide à la sélection d'information utile)

L'aide à la sélection d'information utile est assurée par les mécanismes de spécification des rôles de la tâche décrivant le besoin de l'utilisateur.

Dans le cas de l'information géographique, il nous semble qu'une aide à la sélection doit consister non pas tant à discriminer de l'information utile parmi une abondance d'information mais plutôt à proposer plusieurs solutions à l'utilisateur en lui décrivant les termes d'un compromis à faire. Par exemple le système peut lui proposer comme sélection : "pour calculer une réponse à la question qui vous préoccupe, vous avez le choix entre plusieurs plans de dérivation de données. Tel plan est peu coûteux, tel plan vous assure une réponse la plus précise possible, tel plan vous permet d'utiliser des données provenant du même producteur et donc d'éviter des mauvaises surprises (d'incompatibilités non prévues), tel plan vous permet d'utiliser un seul outil logiciel...". Ce principe n'a pas été exploré dans notre thèse.

Contribution à O3 (la traduction d'une sélection d'information utile en une sélection de ressources)

La fonctionnalité O3 est assurée par les mécanismes de propagation de la spécification des rôles à la spécification du plan de la tâche décrivant le besoin de l'utilisateur. Cette fonctionnalité est en principe extrêmement avancée. Mais elle est complexe à réaliser. Effectivement, elle n'est pas mise en place une fois pour toute mais doit être précisée pour chaque nouvelle tâche. Notre apport consiste à fournir des structures génériques pour expliciter les connaissances nécessaires à cette fonction pour chaque tâche.

Contribution à O4 (la fédération de ressources)

La fonctionnalité O4 est assurée de plusieurs façons.

Une requête est confrontée à l'ensemble des ressources et non à chaque ressource. Ceci est possible car les diverses ressources sont décrites avec un même langage, celui des métadonnées.

Par ailleurs, plusieurs ressources peuvent être exploitées de concert. Il s'agit là d'une fédération optimale des ressources. Cela est possible grâce à des tâches primitives décrivant comme obtenir un jeu de données unifié à l'aide de divers jeux. Notons que cette fonctionnalité n'a de sens que si la fédération de ressources ne se limite pas à une fédération au niveau des métadonnées (i.e. au niveau de l'accès) mais consiste aussi en une fédération au niveau des données (i.e. au niveau de la manipulation).

Contribution à O5 (s'adresser à divers utilisateurs)

La fonctionnalité O5 n'a pas été suffisamment explorée. Elle dépend de plusieurs éléments.

Elle dépend de l'interface de spécification du besoin de l'utilisateur. Cette interface doit être plus conviviale et permettre un vrai dialogue entre le système et un utilisateur.

Elle dépend également du domaine, c'est-à-dire des connaissances applicatives que possède le système. Il est important d'y intégrer des ontologies de domaine applicatifs.

3.3.3.2 Evaluation du prototype

Le prototype a permis de définir des objets pour représenter les éléments tâches, plans, rôles, éléments du domaine. Il a proposé aussi une représentation du mécanisme de spécification de ce modèle pour qu'il puisse s'adapter à des besoins particuliers d'utilisateurs.

Notons cependant des limites de ce prototype.

L'interface du système permettant à un utilisateur de spécifier son besoin n'est pas suffisamment intuitive. Il est nécessaire d'ajouter à ce système une interface homme-machine plus sophistiquée.

La flexibilité de la base de connaissances du système ne peut être assurée que si on dispose d'une interface d'enrichissement de cette base. Pour l'instant nous avons proposé une méthode assez générale d'analyse des manipulations de l'information géographique connues pour les décrire à l'aide des éléments de notre modèle. Il est nécessaire que l'intégration de nouvelles connaissances dans le modèle s'appuie sur les connaissances existantes. Pour cela il faut construire une interface d'édition de la base de connaissances du système indiquant les possibilités de réutilisation des connaissances existantes dans les contextes suivants : intégration de nouveaux concepts applicatifs, de nouveaux jeux de données, intégration de nouveaux traitements sur les données disponibles, intégration de nouvelles méthodes pour des tâches déjà représentées.

Enfin, nous nous sommes concentrés dans ce prototype sur la spécification d'un besoin de connaissances géographiques. Il est également nécessaire de permettre à un utilisateur de spécifier un besoin de traitement ou un besoin de données c'est-à-dire de parcourir non plus la base d'objectifs connus, mais plutôt la base de traitements connus du système (tâches primitives), et la base de jeux de données disponibles.

Enfin, il est nécessaire de confronter le besoin de ressources, traduit en intensification de ressources par le système, aux ressources effectivement disponibles pour déterminer quelles ressources peuvent effectivement remplir les rôles voulus dans l'application, comme décrit Figure 91, c'est-à-dire permettre de monter pratiquement l'application. Nous n'avons pu mettre en place cette fonctionnalité faute de posséder suffisamment de métadonnées. Plusieurs travaux en cours à l'IGN visent à mettre en place des bases de métadonnées et leurs résultats devront être exploités dans notre système.

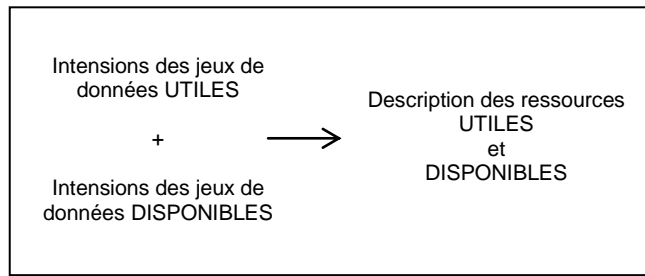


Figure 91 : Etape ultime : la reconnaissance de ressources disponibles en réponse à une intension de ressources. [Figure 91 : Ultimate stage : to reckon available resources answering a need specified as required resources.]

CONCLUSION ET PERSPECTIVES

Les données géographiques sont des ressources en information extrêmement riches qui peuvent répondre à des besoins variés grâce à diverses techniques de manipulation de ces données et de nombreux outils. L'accès d'utilisateurs à ces données est difficile, principalement pour les raisons suivantes :

- Il existe de nombreuses BDG et les utilisateurs ne connaissent pas toujours leur existence.
- Les métadonnées, qui peuvent être utilisées pour cataloguer les BDG, ne sont pas destinées aux utilisateurs. Elles restent dédiées au transfert de données ou à la gestion de données par un producteur.
- L'interprétation de la représentation utilisées dans une BDG est complexe. Il n'existe pas de représentation universelle de l'espace géographique. Les représentations utilisées pour construire les BDG sont éloignées des concepts d'utilisateurs. Ces représentations sont souvent guidées par des modèles cartographiques, qui ne nécessitent pas d'explicitement des relations spatiales entre objets géographiques, alors que les raisonnements d'utilisateurs s'appuient essentiellement sur de telles relations. Par ailleurs, les besoins de représenter la notion de situation dans l'espace conduisent souvent à utiliser des techniques mathématiques peu accessibles pour informatiser des formes géométriques.
- L'exploitation de données géographiques est également complexe et requiert une expertise spécifique, comme connaître les outils disponibles, connaître les relations d'interopérabilité entre les nombreux formats utilisés dans les BDG et les traitements, savoir interpréter les résultats d'un traitement.

Le travail présenté dans ce rapport traite ce problème en se concentrant sur ses aspects liés à la gestion de la connaissance.

Pour détailler ce problème, nous avons étudié la notion générale d'« accès à des connaissances ». Un premier apport de ce travail est de proposer une formalisation de la problématique très générale de l'accès à des ressources en information, non nécessairement géographique. Cette formalisation permet de caractériser la notion d'aide à l'accès par plusieurs objectifs décrits Figure 21 page 55 :

- O1 : construire l'espace des transferts, c'est-à-dire l'espace des connaissances dérivables des ressources,
- O2 : faire une sélection d'information utile, c'est-à-dire reformuler le besoin de l'utilisateur en un besoin d'information dérivée des ressources,
- O3 : traduire une sélection d'information utile en une sélection de lots à acquérir, c'est-à-dire propager une requête en termes d'information dérivable en une requête en termes de portions de ressources,

- O4 : fédérer des ressources, c'est-à-dire proposer à l'utilisateur un mode d'accès à diverses ressources comme s'il s'agissait d'une seule ressource,
- O5 : s'adresser à divers utilisateurs, c'est-à-dire prendre en compte une diversité de besoins et d'aptitudes à exprimer ces besoins.

Cette décomposition en sous-objectifs permet de situer, les uns par rapport aux autres, dans l'état de l'art, diverses démarches qui doivent être prises en compte dans notre travail, soit parce qu'elles prennent place dans un autre domaine d'information mais qu'elles mettent en œuvre une technique efficace méritant d'être appliquée au cas de l'information géographique (section 2.1), soit parce qu'elles contribuent effectivement à l'accès d'utilisateurs à l'information géographique (section 2.2).

Un premier volet de cet état de l'art (section 2.1) évalue les techniques actuelles pour aider l'accès tel que nous l'avons défini. Nous nous sommes concentrés tout particulièrement sur la contribution au sous-objectif O1 car un problème essentiel est selon nous l'écart entre les expressions possibles de besoins, et les descriptions de réponses à ces besoins construites avec des données géographiques. Ce premier volet classe les démarches selon leur contribution à O1 : démarches d'aide à l'accès aux données, d'aide à l'accès aux informations et d'aide à l'accès aux connaissances. Il évalue également la contribution de chaque démarche aux autres objectifs.

Les techniques les plus complètes sont les techniques d'aide au raisonnement formel mises en place par la communauté d'acquisition des connaissances. Dans notre contexte, ces techniques doivent être appliquées au raisonnement constituant l'accès : l'exploration, focalisée par le contexte de l'utilisateur, des connaissances qu'il pourrait obtenir à partir de lots pouvant être extraits des ressources. L'apport des techniques d'aide au raisonnement formel est double :

- D'une part, elles sont utilisées pour construire un modèle de connaissances dans lequel ce raisonnement global peut être mené. Ce modèle consiste en une description des ressources, des besoins, et des mécanismes permettant de relier un besoin à des ressources lui répondant.
- D'autre part, elles permettent de répartir ce raisonnement global entre plusieurs acteurs, en les faisant travailler sur un même modèle. Comme il est possible de faire raisonner formellement un composant implémenté, un de ces acteurs peut être un tel composant. Donc les techniques d'aide au raisonnement formel aident un utilisateur à raisonner en le soulageant d'une partie du raisonnement global constituant son accès.

Un principe clé de formalisation consiste à distinguer, au sein du modèle utilisé, les connaissances qui décrivent un domaine, ou *QUOI*, et celles qui décrivent sa manipulation, ou *COMMENT*. Ce principe a deux propriétés intéressantes dans le contexte de l'aide à l'accès, si on l'applique en construisant un tel modèle pour décrire des données géographiques :

- C'est une méthode pour séparer des connaissances de contenu de leurs divers modes d'emploi. Cela assure d'aider l'accès à un jeu de données sans se restreindre à un mode d'emploi donné de ce jeu. Cela permet aussi de réutiliser un même mode d'emploi sur des contenus différents. Tout cela concourt à permettre à un utilisateur d'exploiter les données avec le mode d'emploi adapté à son besoin propre.
- La description d'un *COMMENT* global comporte les connaissances de réponses aux questions *pourquoi*, *comment* et *avec quoi*. Ces connaissances permettent d'explicitement l'expression d'un besoin (grâce au *pourquoi*) et de le relier (grâce au *comment*) à une désignation de ressources utiles pour y répondre (grâce au *avec quoi*).

Un deuxième volet évalue l'application de ce principe dans le domaine de l'information géographique c'est-à-dire l'existence d'un tel modèle décrivant les données et décrivant leurs manipulations possibles.

La plupart des démarches portent sur la formalisation du *QUOI*, en intégrant une part de *COMMENT* pour guider la modélisation. Ces formalisations du *QUOI* ne sont pas unifiées.

La formalisation effective du *COMMENT* ne répond principalement qu'aux besoins d'implémentation des traitements sur des données géographiques ou interopérabilité entre traitements et données. Elle concerne le *avec quoi*. Il n'y a par contre pas de véritable formalisation des *pourquoi* et *comment*, c'est-à-dire :

- l'énoncé de problèmes sous forme d'objectifs à atteindre en des termes proches des concepts utilisateurs,
- la description de méthodes pour résoudre ces problèmes (atteindre ces objectifs).

Cet état de l'art nous permet d'envisager une contribution qui consiste à mettre en œuvre le principe de modélisation énoncé à l'issue du premier volet section 2.1, en nous concentrant tout particulièrement sur les lacunes du domaine de l'information géographique mises en évidence dans le second volet section 2.2.

Cette contribution est concrétisée par la construction d'un modèle spécifique, intégrant les notions de tâches, méthodes et rôles, pour modéliser de façon distincte ce qu'est l'information géographique, le *QUOI*, et les raisonnements qui s'appuient sur cette information, le *COMMENT*. Le principe de cette contribution est de décrire des gabarits d'utilisations de données géographiques et de permettre à un utilisateur de savoir comment répondre à son besoin d'information géographique en adaptant ces gabarits à son besoin et contexte particuliers. Ce modèle est présenté section 3.1.2. Il est formalisé par la définition d'une grammaire Backus Naur Form section 3.1.2.5.

Les apports de ce modèle sont multiples :

- Il permet de modéliser des connaissances jusqu'ici peu formalisées comme le *pourquoi* et le *comment*. Ces connaissances portent sur la description de besoins d'information et de méthodes de dérivation de telles informations.
- Il permet de modéliser le lien entre les diverses catégories de connaissances. Ce lien est essentiel dans le transfert de connaissances de données vers un utilisateur.
- Il permet enfin de modéliser des connaissances de spécification permettant l'expression d'un besoin particulier et sa traduction en une sélection de ressources, c'est-à-dire des catégories de connaissances intervenant précisément dans l'accès. Ces connaissances portent sur la spécification du besoin et sur la spécification de la méthode de réponse à ce besoin.

Ce modèle est associé à un système de résolution coopérative de problème qui opérationnalise les mécanismes d'adaptation des gabarits au problème particulier d'un utilisateur. Ce système est présenté section 3.1.3.

Afin de démontrer la capacité du modèle à représenter de façon distincte et reliée des connaissances appartenant au *QUOI* et des connaissances appartenant au *COMMENT*, il a été utilisé pour modéliser des raisonnements géographiques particuliers en section 3.2.

Comme le modèle de tâches géographiques ne porte pas sur une tâche spécifique mais sur un domaine spécifique, nous avons dans un premier temps exprimé les spécificités du domaine, section 3.2.1. Le domaine du modèle est la partie de notre modèle décrivant le *QUOI* géographique. En tant que tel, il doit rendre compte de la diversité des représentations qui sont faites de ce *QUOI*, et des facteurs permettant de comprendre une représentation. Un élément important est de représenter le sujet de l'activité conduisant à construire la représentation.

Nous avons ensuite modélisé plusieurs raisonnements géographiques.

- La modélisation de la tâche d'appariement de jeux de données, section 3.2.3, a permis de montrer en quoi le modèle de tâches permettait de décrire un processus bien connu [Lemarié 96][Badard 00][Bel Hadj Ali 01].
- La tâche de localisation est une tâche extrêmement générale pour laquelle il n'existait pas de méthode générique. Le terme « localisation » peut renvoyer à une variété de besoins et d'outils utilisés pour répondre à ces besoins. Nous proposons un modèle de tâche de localisation qui se spécifie pour couvrir une diversité de cas, section 3.2.4.

- La modélisation de la tâche de détermination d'une navigation est axée sur l'expression du besoin. Cette tâche, telle que modélisée 3.2.5, permet de spécifier de façon avancée un besoin de détermination de navigation.

La modélisation d'une tâche implique de distinguer, dans le vocabulaire utilisé dans un raisonnement géographique, ce qui renvoie à une structure dédiée à une manipulation et ce qui renvoie à la nature de l'information géographique. Les structures construites dans le cadre de raisonnement appartiennent au *COMMENT* et sont modélisées sous forme de rôles. Etablir cette distinction est une difficulté essentielle de cette modélisation et cependant un point clé pour le bon fonctionnement du modèle. Le *QUOI* géographique doit pouvoir se prêter à tous les raisonnements qui peuvent être menés dessus, c'est-à-dire qu'il doit pouvoir s'intégrer dans diverses structures. Pour cela, il est important de représenter les connaissances « de valuation des rôles », c'est-à-dire les connaissances désignant quel élément du domaine tient lieu, dans le raisonnement courant, de la structure décrite par le rôle. Nous avons indiqué au cas par cas, au cours de la section 3.2, quelles pouvaient être ces connaissances de valuation des rôles et leur mode de représentation dans le *COMMENT*. Mais nous n'avons pas formalisé en détail de telles connaissances. De fait, elles sont liées aux connaissances de spécification, qui ont été décrites mais qui n'ont pas été formalisées de façon poussée, par manque de temps.

Le modèle et la base de connaissances associée sont exploités par un système expert pour aider un utilisateur à explorer les connaissances, dérivables de données géographiques disponibles et pertinentes dans son contexte.

Un prototype de ce système expert est proposé en section 3.3. La difficulté essentielle de cette étape a été de proposer une représentation implémentable de notre modèle. Nous avons choisi le langage d'implémentation orienté-objet java pour cette étape. Les raisons de ce choix sont que le modèle orienté-objet est, parmi les langages implémentables, le plus proche des raisonnements de l'homme et que le langage java est un langage simple à utiliser.

Une attention particulière a été apportée à la définition de structures génériques. La description d'une tâche géographique particulière s'appuie uniquement sur la spécification de variables et non sur l'écriture de connaissances procédurales spécifiques.

Cette représentation est instanciée sur l'exemple de la tâche de localisation. Ce prototype montre qu'il est possible de stocker sous forme d'objets les éléments du modèle de tâches géographiques de façon à ce qu'ils puissent être manipulés automatiquement pour construire la réponse à un besoin d'utilisateur en spécifiant une tâche géographique.

En définitive, le travail présenté dans ce rapport initie une démarche. Il apporte une vue d'ensemble sur la problématique traitée, situe une contribution originale et démontre la faisabilité de cette approche. Plusieurs aspects de ce travail demandent à être poursuivis : la modélisation des tâches de localisation et détermination de navigation et l'implémentation de ces tâches dans le prototype. Par ailleurs, nous envisageons de nombreuses pistes pour compléter et étendre cette contribution.

Une première piste concerne l'interface utilisateur de ce prototype, dans un contexte d'aide à l'accès. Nous n'avons pu confronter des utilisateurs à notre prototype qui ne possède pas d'interface de dialogue homme-machine et est en l'état peu intuitif. Il est nécessaire de construire une interface de dialogue. Pour cela, nous exploiterons entre autres les résultats du travail de thèse de Frédéric Hubert au COGIT [Hubert 01a] [Hubert 01b]. Les utilisateurs auxquels nous nous adresserons dans un premier temps sont des utilisateurs experts en information géographique travaillant à l'IGN pour produire des données.

Une autre piste est l'enrichissement de la base de connaissances de ce prototype. Elle se décline en trois axes :

- les modalités d'édition de cette base pour y intégrer de nouvelles connaissances,
- l'identification des BD qui peuvent être intégrées dans la base de connaissance,
- la modélisation de connaissances n'ayant pas encore fait l'objet d'une formalisation.

Pour ce qui est du premier axe, il est important de formaliser le processus d'intégration de nouveaux éléments dans la base de connaissances du système, voire d'automatiser une partie de ce processus. Cette intégration ne consiste pas uniquement à décrire avec les éléments génériques du modèle une nouvelle connaissance (données, outil, méthode, objectif). Cela inclut aussi la propagation de l'enrichissement des connaissances ontologiques aux connaissances de résolution de problèmes et vice versa. Lorsqu'une nouvelle ressource est décrite, il faut relier cette description aux exploitations qui peuvent être faites de la ressource. Lorsqu'une nouvelle méthode est décrite, il faut la relier aux opérations et aux données qu'elle permet de manipuler.

Pour ce qui est du deuxième axe, l'identification des BD qui peuvent être intégrées dans le prototype, il s'agit essentiellement de bases de métadonnées et de bases applicatives.

Il existe actuellement peu de bases de métadonnées. Un projet de développement portant sur la diffusion des données à l'IGN a débuté en juin 2001 et une première étape consiste à acquérir, éventuellement construire, des métadonnées [Cébelieu 01].

Il est également important d'intégrer dans le prototype des connaissances utiles pour l'expression du besoin et du contexte de l'utilisation. De telles connaissances sont par exemple des nomenclatures métiers. Cela peut être aussi des connaissances d'interprétation permettant à un utilisateur de s'exprimer dans des concepts plus familiers que ceux utilisés formellement pour décrire les ressources. Par exemple, la désignation d'un lieu par l'utilisateur peut se faire grâce à des indexes géographiques (fichiers de noms de lieux associés à leurs coordonnées géographiques), grâce à une base de données administratives, ou grâce à une BDG de très petite échelle qui peut servir d'index géographique en associant des coordonnées à des objets géographiques de référence comme des fleuves, des forêts ou des sites touristiques connus.

Le dernier axe, la modélisation de nouvelles connaissances, est une perspective très riche.

Ces nouvelles connaissances sont par exemple des métadonnées d'un autre type que celles actuellement définies dans les standards. Il s'agit de connaissances sur les BDG nécessaires à la valuation des rôles.

Ce sont aussi la description de traitements SIG. Un format pertinent pour cette description est en train d'émerger, il s'agit de formats de description de services basés sur XML tels WSDL. Il sera donc utile de traduire notre grammaire BNF dans un XMLSchema et de décrire plus précisément les éléments "mécanismes" dans ce nouveau formalisme en intégrant WSDL.

Une autre perspective intéressante est de modéliser des tâches de généralisation. Cela permet de contribuer à un processus de capitalisation de connaissances aussi difficile à mettre en place que crucial. Cette problématique de la capitalisation de connaissances est également importante dans un contexte de production. A l'IGN, cette capitalisation ne concerne pas uniquement des tâches de haut niveau telles que celles figurant sur la Figure 4 page 24. Il existe à l'IGN de nombreuses données géographiques et algorithmes qui correspondent à des étapes intermédiaires de processus de production. Ces connaissances sont développées au cas par cas, de façon indépendante, mais pourraient être réutilisées dans d'autres contextes si les utilisateurs connaissaient leur existence. Il est donc pertinent d'indexer ces méthodes et données intermédiaires pour permettre leur réutilisation dans un contexte différent de celui où elles ont été produites.

Une autre perspective extrêmement intéressante à nos yeux de ce travail, mais à long terme, est d'intégrer dans le modèle les structures mises en place par les géographes pour analyser l'action de l'homme dans l'espace géographique. Un utilisateur ayant besoin de ressources en information géographique pourrait accéder à la connaissance des modèles des géographes. Actuellement il n'existe pas de démarche dans ce sens. Pourtant cela permettrait de rendre les utilisateurs de données géographiques mieux informés de leur intervention dans l'espace géographique.

Pour conclure ce mémoire, rappelons que cette thèse rend compte d'un cas d'application de principes provenant de l'acquisition des connaissances. Il a été nécessaire de reformuler en détail la problématique traitée et les apports de l'acquisition des connaissances dans ce contexte pour deux raisons :

- La démarche, de par le sujet même et le domaine technique, est nouvelle dans le laboratoire où s'est déroulée la thèse et il était important de définir une approche pertinente.
- L'acquisition des connaissances est un domaine extrêmement complexe qu'il n'était pas envisageable de s'approprier intégralement. Il a plutôt fallu s'approprier les notions intéressantes dans notre contexte.

Une telle appropriation partielle d'un domaine technique complexe est une démarche longue mais elle est aussi nécessaire à toute collaboration avec des chercheurs de cette communauté. Or le domaine de l'information géographique et des systèmes d'information géographique a beaucoup à recevoir de la communauté d'acquisition des connaissances : l'application de ses acquis en termes de partage et de réutilisation de connaissances nous semble pouvoir répondre au problème épineux et crucial de l'accès d'utilisateurs à l'information géographique sous ses représentations actuelles. Cette démarche nous permet d'envisager une solution à ce problème, et donc une aide à l'exploitation –généralement dans un contexte de prise de décision- plus efficace et plus rapide de données géographiques dans des domaines où l'expertise géographique est mineure mais dont l'information géographique est devenue néanmoins une composante du processus décisionnel. L'enjeu de ces travaux, et de ceux allant dans le même sens, est que le domaine de l'information géographique, qui doit être accessible à tous, ne soit pas un monde de BDG et SIG réservé à des experts.

GLOSSAIRE

Connaissances : éléments intégrés dans le savoir et potentiellement exploitables.

Données : information qui existe indépendamment d'une intelligence humaine, par exemple dans un ordinateur.

Formalisation : description de connaissances dans un langage leur apportant une nouvelle maniabilité.

Information : terme général renvoyant à des données, des informations, ou des connaissances.

Informations : données associées à la réalité qu'elle représentent.

Intelligence Artificielle : domaine de l'informatique traitant de la façon dont les ordinateurs peuvent reproduire ou simuler des comportements humains, en particuliers la faculté de traiter des connaissances.

Métadonnées : Le terme *métadonnée* est utilisé pour désigner de façon générale des informations relatives à des données [Role 99], destinées à être manipulées elles-mêmes comme des données.

Modélisation : Construction de structures formelles pour représenter un savoir. Dans notre rapport ce terme est souvent équivalent de celui de « formalisation ». Lorsque le terme « modélisation » est utilisé en opposition avec le terme « représentation » il renvoie à une formalisation qui peut être plus facilement maniée par un homme que par une machine alors que la représentation est destinée à être intégrée dans une machine.

Ontologie : Spécification explicite d'une conceptualisation [Gruber 93]. Ensemble de concepts et de relations entre ces concepts permettant de décrire, avec un certain consensus et une certaine exhaustivité, un domaine donné.

Profil utilisateur : Des profils utilisateurs sont des descriptions, selon certaines variables, d'utilisateurs typiques d'un système. Ils sont construits pour permettre, lorsque le système est exploité par un utilisateur réel, d'appréhender cet utilisateur comme appartenant à un profil connu et d'en déduire des informations utiles lors des interactions avec cet utilisateur.

Représentation : Ce terme est généralement employé de façon équivalente à modélisation ou formalisation. Il faut noter deux nuances paradoxales du terme « représentation ». Lorsqu'il est utilisé en opposition au terme modélisation, il renvoie à une formalisation destinée à être intégrée dans une machine (c'est en fait l'interprétation « IA » de ce terme). Par contre, dans certains contextes, il permet de renvoyer à une activité de formalisation très peu poussée « la représentation que se fait l'homme du monde qui l'entoure ». Il s'agit là de l'acception commune de ce terme, qui renvoie à une activité moins spécialisée que la modélisation.

Sémantique : lien entre un symbole et la réalité que représente ce symbole.

Structure : Ensemble d'éléments et de liens entre ces éléments.

Type : Variable caractérisé par son domaine de valeur et des opérations disponibles sur cette variable.

UML : Unified Modelling Language. Langage de modélisation orienté-objet.

BIBLIOGRAPHIE

- [Abel et al. 99] David Abel, Volker Gaede, Kerry Taylor, Xiaofang Zhou, SMART : Towards Spatial Internet Marketplaces, *Geoinformatica*, Volume 3, Number 1, March 1999, pp.141-164
- [AFIGEO 97] Etude du Marché Européen de l'Information Géographique numérique, étude lancée par le Conseil National de L'Information Géographique et l'Association Française pour l'Information Géographique, octobre 1997
- [AFIGEO 98] L'information géographique française dans la société de l'information, état des lieux et propositions d'action, étude lancée par le Conseil National de L'Information Géographique et l'Association Française pour l'Information Géographique, mai 1998
- [Albrecht 96] Jochen Albrecht, Universal GIS operations for environmental modeling, in *Proceedings of the 3rd International Conference on Integrating GIS and Environmental Modeling*, Santa Barbara, 1996
- [Andersen et al. 95] D. Andresen, L. Carver, R. Dolin, C. Fischer, J. Frew, M. Goodchild, O. Ibarra, R. Kothuri, M. Larsgaard, B. Manjunath, D. Nebert, J. Simpson, T. Smith, T. Yang, Q. Zheng, The WWW Prototype of the Alexandria Digital Library, *Proceedings of ISDL'95: International Symposium on Digital Libraries*, Tsukuba, Japan, August 1995
- [Andrienko and Andrienko 99] Andrienko, G. and Andrienko, N., Interactive Maps for Visual Data Exploration, *International Journal Geographical Information Science*, Special Issue on Visualization for exploration of Spatial data, 1999 v.13 (4), pp.355-374
- [Aussenac 89] Nathalie Aussenac, Conception d'une Méthodologie et d'un Outil d'Acquisition de Connaissances Expertes, Thèse en Informatique et Intelligence Artificielle, Université Paul Sabatier, Toulouse, octobre 1989
- [Badard 00] Thierry Badard, *Propagation des mises à jour dans une base de données multi-représentations par analyse des changements géographiques*, Thèse en Sciences de l'Information Géographique, Université de Marne la Vallée, 2000
- [Barrault et al. 01] Barrault M., Regnauld N., Duchêne C., Haire K., Baeijs C., Demazeau Y., Hardy P., Mackaness W., Ruas A., Weibel R. 2001. Integrating Multi-agent, Object-oriented, And Algorithmic Techniques For Improved Automated Map Generalization, in *Proceedings of 20th International Cartographic Conference*, Beijing, 2001
- [Bédard and Sondheim 97] Yvan Bédard, Mark Sondheim, BCRNS Conceptual Data Dictionary, version 1.0, march 1997
- [Bédard et al. 00] Yvan Bédard, Tim Merrett, Jiawei Han, Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, in *Geographic Data Mining and Knowledge Discovery*, H. Miller and

- J. Han (ed), Research Monographs in GIS series, Peter Fisher and Jonathan Raper (ed), Taylor & Francis, 2000
- [Bel Hadj Ali 01] Atef Bel Hadj Ali, *Qualité géométrique des entités géographiques surfaciques, Application à l'appariement et définition d'une typologie des écarts géométriques*, Thèse de doctorat en Sciences de l'Information Géographique, Université de Marne la Vallée, 2001
- [Berry 87] Joseph Berry, Fundamental Operations in Computer-Assisted Map Analysis, IJGIS, Vol 1, No2, 1987, pp119-136
- [Berry 93] Joseph Berry, *Beyond Mapping, concepts, algorithms and issues in GIS*, GIS World Books, USA, 1993
- [Bertin 67] Jacques Bertin, *La Sémiologie Graphique*, ed Gauthier-Villars, Paris, 1967
- [Bishr 97] Yaser Bishr, Semantic Aspects of Interoperable GIS, PhD Thesis, ITC, Netherlands, 1997
- [Bittner 99] Thomas Bittner, Rough Location, PhD Thesis, Technical University Vienna, January 1999
- [Borgo et al. 97] Stefano Borgo, Nicolas Guarino, Claudio Masolo, Guido Vetere, Using a Large Linguistic Ontology for Internet-Based Retrieval of Object-Oriented Components, Conference on Software Engineering and Knowledge Engineering SEKE'97
- [Bouillé 77] F. Bouillé, Un modèle universel de banque de données, simultanément partageable, portable et répartie, Thèse de Doctorat ès Sciences Mathématiques-Informatique, Université Paris VI, 1977
- [Bouillé 96] F. Bouillé, HBDS in GIS : overview on the model, methodology and applications, in *proceedings of the Congress on Remote Sensing for Environment – Data Processing and Interpretation*, Moscow, 1996
- [Boursier et Mainguenaud 92] Patrice Boursier, Michel Mainguenaud, Spatial Query Languages : extended SQL vs visual languages vs hypermaps, in *Proceedings of the 5th International Conference on Spatial data Handling*, SDH'92, Charleston, USA, 1992, pp.249-259
- [Brassel and Weibel 88] Kurt Brassel, Robert Weibel, A Review and Framework of Automated Map Generalization, *Int. Journal of Geographical Information Systems* 2(3), pp.229-244, 1988
- [Brun et al. 87] F. Brun, L. Buffat, P. Stivalet, C. Raphael, Approche cognitive des représentations mentales spatiales, Centre de Recherches du Service de Santé des Armées, Rapport Interne No. 6/CRSSA/PS, 1987
- [Buttenfield and Kumler 96] Buttenfield, B.P. & Kumler, M.P., Tools for browsing environmental data: The Alexandria Digital Library Interface, in *Proceedings of the Third International Conference on Integrating Geographic Information Systems and Environmental Modeling*. Santa Fe, New Mexico, January 21-26, 1996
- [Cébelieu 01] Gilles Cébelieu, *Mise en place de la base de métadonnées pour l'archivage, la consultation et la diffusion des données numériques sur le réseau*, Rapport interne de l'IGN DIFNUM/201, Saint Mandé, 2001
- [Chartron 97] Ghislaine Chartron, Repérage de l'Information sur Internet : Nouveaux Outils, Approches Bibliothéconomiques et Micro-structures, *CDI Virtuel : Utilisation des Réseaux pour l'Accès à la Communication*, CRDP de reims, avril 1997, <http://www.urfist.jussieu.fr/urfist/article.htm>
- [Chandrasekaran et al. 98] B. Chandrasekaran, J. R. Josephson, V. Richard Benjamins, The Ontology of Tasks and Methods, in *Proceedings of KAW'98*, Eleventh Workshop on Knowledge Acquisition, Modeling and Management, Alberta, Canada, April 1998
- [Cheylan 92] Jean-Paul Cheylan, Classification des fonctions de traitement dans les SIG : éléments de synthèse, Communication à la conférence SIG-GIS 1992
- [Chomicki and Revesz 99] Jan Chomicki, Peter Revesz, A geometric framework for specifying spatiotemporal objects, in *proceedings of the 6th International Workshop on Time Representation and Reasoning*, 1999, pp.41-46
- [Chrisman 99] Nicholas Chrisman, A transformational approach to GIS operations, in *International Journal of Geographical Information Science*, 1999, vol 13, n 7, pp.617-637
- [Christensen et al. 01] Erik Christensen, Greg Meredith, Francisco Curbera, Sanjiva Weerawarana, Web Services Description Language (WSDL) 1.1, 2001 <http://msdn.microsoft.com/library/?url=/library/en-us/dnwebsrv/html/wSDL.asp?frame=true>

- [Cian et al. 89] C. Cian, P. Filliat, C. Raphel, Processus cognitifs mis en jeu dans la lecture de cartes topographiques, Centre de Recherches du Service de Santé des Armées, Rapport Interne No. 3/FH/PS, 1989
- [Claramunt et al. 94] Christophe Claramunt, Marie-Hélène de Sède, Roland Prélaz-Droux, Laura Vidale, Sémantique et Logique Spatio-temporelles, revue Internationale de Géomatique, Vol. 4, No. 2, pp165-180, 1994
- [Claramunt and Mainguenaud 96] Christophe Claramunt, Michel Mainguenaud, A Spatial Data Model for Navigation Knowledge, SDH '96, 1996
- [Claval 82] Paul Claval, La nouvelle géographie, Que Sais-je ?, Paris, 1982
- [Clement et al. 97] Gilles Clement, Christian Larouche, Denis Gouin, Paul Morin, Henry Kucera, OGDI : Towards Interoperability among Geospatial Databases, SIGMOD Record, Vol. 26, No. 3, September 1997
- [Clouard et al. 02] Régis Clouard, Abderrahim Elmoataz, Marinette Revenu, Une méthodologie de développement d'applications de traitement d'images, RFIA, Angers, 2002
- [CNIG 97] Conseil National de l'Information Géographique, Le Catalogage des Données, 9^{ème} Journée Annuelle de la Recherche Géographique, Synthèse des interventions et débats, Lyon, 1997
- [Connolly et al. 01] Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein, *DAML+OIL Reference Description*, W3C Note 18, December 2001, <http://www.w3.org/TR/daml+oil-reference>
- [Corona and Winter 01] Corona, B.; Winter, S., 2001: Navigation Information for Pedestrians from City Maps. In: Konecny, M.; Friedmannova, L.; Golan, J.; Kolar, M. (Eds.), *GI in Europe: Integrative – Interoperable – Interactive*, Proceedings of the 4th AGILE Conference on Geographic Information Science, Masaryk University Brno, pp. 189-197
- [Crusson et Girard 01] Tanguy Crusson, Didier Girard, *XML Cartographie des standards pour l'entreprise*, Developpeur Référence, v1.3, juin 2001
- [Dangermond 83] J. Dangermond, A classification of software components commonly used in geographic information systems, in *Design and implementation of computer-based geographic information systems*, D. J. Peuquet, J. O'Callaghan (eds), 1983, pp.70-91
- [David 91] Benoît David, Modélisation, représentation et gestion d'information géographique, Thèse de doctorat de l'Université de Paris VI en informatique, Paris 1991
- [De la Losa 00] Arnaud de la Losa, *Modélisation de la troisième dimension dans les bases de données géographiques*, thèse de doctorat en Sciences de l'Information Géographique, Université de Marne la Vallée, janvier 2000
- [Delouis et Krivine 95] I. Delouis, J.P. Krivine, LISA : un langage réflexif pour opérationnaliser les modèles d'expertises, Revue d'Intelligence Artificielle, 9(1), Hermès, Paris, 1995, pp.53-88
- [Denand 00] Nicolas Denand, La question où : des critères pour le calcul de la réponse, in *Actes des Rencontre des Jeunes Chercheurs en Intelligence Artificielle*, Lyon, Septembre 2000
- [Denègre et Salgé 96] Jean Denègre, François Salgé, Les systèmes d'information géographique, *Que sais-je ?*, Presses Universitaires de France, Paris, 1996
- [Devogele et al. 98] Thomas Devogele, Christine Parent, Stefano Spaccapietra: On Spatial Database Integration. International Journal of Geographical Information Science 12(4): 335-352, 1998
- [DG 99] Distributed Geolibraries : Spatial Information Resources, Summary of a Workshop, Panel on Distributed Geolibraries, National Research Council, National Academy Press, Santa Barbara 1999.
- [Didier 90] Michel Didier, Utilité et Valeur de l'Information Géographique, Ed. ECONOMICA, 1990
- [Duchêne et Ruas 01] Duchêne C. et Ruas A., Généralisation de données géographiques : Présentation des résultats du projet AGENT. Bulletin d'information de l'IGN, n°72 : Bilan de la Recherche 2000, 2001
- [Dueker and Butler 98] K. J. Dueker, J. A. Butler, GIS-T Enterprise Data Model with Suggested Implementation Choices, URISA Journal, vol. 10, No. 1, Spring 1998
- [Egenhofer and Mark 95] Max Egenhofer , David Mark, Naive Geography, Andrew U. Frank, Werner Kuhn (Eds.): Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT '95, Semmering, Austria, September 1995, Proceedings. Lecture Notes in Computer Science, Vol. 988, Springer, 1995 pp.1-15

- [Faiz 96] Sami Othman Faiz, Modélisation, exploitation et visualisation de l'information qualité dans les Bases de Données Géographiques, Thèse de l'université de Paris-Sud, UFR scientifique d'Orsay, 1996
- [Fensel et al. 00] D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, M. Klein, OIL in a nutshell, in *Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000)*, R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, October 2000
- [Ficet 99] Valérie Ficet Cauchard, Réalisation d'un Système d'Aide à la Conception d'Applications de Traitements d'Images : une Approche Basée sur le Raisonnement à Partir de Cas, thèse en informatique, Université de Caen, 1999
- [Fletcher et al. 98] David Fletcher, John Espinoza, R. D. Mackoy, Stephen Gordon, Bruce Spear, Alan Vonderohe, The Case for a Unified Linear Reference System, *URISA Journal*, Vol. 10, No. 1, Spring 1998
- [Frank 01] Andrew Frank, Ontology for spatio-temporal databases, in *Spatiotemporal Databases: The Chorochronos Approach*, Lecture Notes in Computer Science, Berlin, Springer-Verlag, to appear
- [Frew et al. 95] Frew, J., L. Carver, C. Fischer, M. Goodchild, M. Larsgaard, T. Smith, and Q. Zheng, The Alexandria Rapid Prototype: building a digital library for spatial information, in *Proceedings of the 1995 ESRI User Conference Proceedings*, Environmental Systems Research Institute, Inc., Redlands, CA, May 22-25, 1995
- [Frew et al. 99] James Frew, Michael Freeston, Linda Hill, Greg Janee, Mary Larsgaard, Qi Zheng, Generic Query Meta-data for Geospatial Digital Libraries, in *Proceedings of The Third IEEE Meta-Data Conference*, April 1999, Maryland, USA
- [Gardarin et Valduriez 90] G. Gardarin, P. Valduriez, SGBD Avancés, Bases de Données Objets, Déductives, Réparties, EYROLLES, 1990
- [Gervais 00] Marc Gervais, Conception d'un manuel intelligent informatisé pour l'utilisateur de données géographiques comme support à la prise de décision, mémoire d'avant-projet de recherche, département des sciences géomatiques de l'université de Laval, Québec, octobre 2000
- [Gibson 79] J. J. Gibson, *Ecological Approach to Visual Perception*, published by Lawrence Erlbaum Associates, 1979
- [Giordano et al. 94] A. Giordano, H. Veregin, E. Borak, D. Lanter, A conceptual model of GIS-based spatial analysis, in *Cartographica*, vol 31, n4, 1994, pp.44-51
- [Gómez and Benjamins 99] Asuncion Gómez, Richard Benjamins, Overview of Knowledge Sharing and Reuse Components : Ontologies and Problem-Solving Methods, In *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999*. (V.R. Benjamins, B. Chandrasekaran, A. Gomez-Perez, N. Guarino, M. Uschold, eds.), pp. 1.1-1.15, Stockholm, 1999
- [Gonçalves 99] Marie-Rose Gonçalves, *Formalisation hybride du raisonnement spatial, Application à la consultation d'une base de données géographiques*, thèse de doctorat en informatique de l'Université de Paris XI, Orsay, France, 1999
- [Goodchild 87] M. F. Goodchild, A spatial analytical perspective on geographical information systems, in *International Journal of Geographical Information Systems*, vol 1, No2, 1987, pp.327-334
- [Goodchild 91] M. F. Goodchild, Introduction to spatial analysis, GIS/LIS 1991 workshop
- [Goodchild 95], Mike Goodchild, Report on a Workshop on Metadata held in Santa Barbara, University of California, November 1995
- [Goodchild and Gopal 98] Michael Goodchild, Sucharita Gopal, *Accuracy of Spatial Databases*, Hermes, Paris, 1998
- [Goodchild and Jeansoulin 98] Michael Goodchild, Robert Jeansoulin, *Data Quality in Geographic Information*, Hermes, Paris, 1998
- [Goodchild 99] Michael Goodchild, *The Geolibrary*, 1999
- [Gordillo et al. 96] Silvia Gordillo, Fernando Das Neves, Catalina Mostaccio, Ana Levato, Design Patterns for Geographic Information Systems, in *proceedings of Samos International Conference on Geographic Information Systems in Urban Environmental and Regional Planning*, Samos, Greece, 1996
- [Grasland 92] Claude Grasland, " Analyse des couples de lieux et modélisation en géographie ", *Géopoint* 1992, Groupe Dupont, Université d'Avignon, pp.97-105, 1992

- [Gruber 93] Thomas Gruber, Towards Principles for the Design of Ontologies Used for Knowledge Sharing, KSL 93-04, Knowledge System Laboratory, Stanford University, 1993
- [Guarino 95] Nicola Guarino, Formal Ontology, Conceptual analysis and Knowledge Representation, *International Journal of Human-Computer Studies*, 43(5/6) :p625—640, 1995
- [Guarino 97] Nicola Guarino, Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration, in M. T. Paziienza (ed.) *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Springer Verlag, 1997, pp.139-170.
- [Guarino 98] Nicola Guarino, Some Ontological Principles for Designing Upper Level Lexical Resources, 1st International conference on Language resources and Evaluation, Spain, May 1998
- [Guarino et al. 99] Nicola Guarino, Claudio Masolo, Guido Vetere, Ontoseek : Content-Based Access to the Web, IEEE 1999
- [Guarino and Welty 00] Nicola Guarino, Christopher Welty, Identity, Unity, and Individuality: Towards a Formal Toolkit for Ontological Analysis, in *the proceedings of ECAI-2000*, H. Werner (ed.), IOS Press, Berlin, 2000, pp. 219-223.
- [Günther and Voisard 97] Oliver Günther and Agnès Voisard, Metadata in Geographical and Environmental Data Management, in *Managing Multimedia Data : Using Metadata to Integrate and Apply Digital Data*, W. Klas and A. Sheth (eds), McGraw Hill, 1997
- [Günther and Müller 98] O. Günther, R. Müller, From GISystems to GIServices: Spatial Computing in the Internet Marketplace, in *Interoperability in Geographic Information Systems*, M. Egenhofer, R. Fegeas, M. Goodchild (eds.), Kluwer Academic Publishers, 1998.
- [Guting and Schneider 95] R. H. Guting, M. Schneider, Realm-Based Spatial Data Types: The ROSE Algebra, in *the VLDB Journal*, 4(2):243--286, April 1995
- [GSDI/TWG 00] GSDI Technical Working Group, *Developing Spatial Data Infrastructures : the SDI Cookbook*, v1.0, Douglas Nebert (Ed), 2000, CD-ROM
- [Hangouët 98] Jean-François Hangouët, *Approche et méthodes pour l'automatisation de la généralisation cartographique : application en bord de ville*, Thèse en Sciences de l'Information Géographique, Université de Marne la Vallée, 1998
- [Hanigan 90] Francis L. Hanigan, GIS Marketing in the 1990s, in *The GIS FORUM*, Francis L. Hanigan (ed), 1990
- [Haton et al. 91] Jean-Paul Haton, Nadjet Bouzid, François Charpillet, Marie-Christine Haton, Brigitte Lâasri, Hassan Lâasri, Pierre Marquis, Thierry Mondot, Amedeo Napoli, *Le Raisonnement en Intelligence Artificielle*, InterEditions, Paris 1991
- [Hedorfer et Bianchin 98] Markus Herdorfer, Alberta Bianchin, Un modèle Structurel pour Métadonnées, Journées Cassini, Marne la Vallée, 1998
- [Hill et al. 97] Linda L. Hill, Ron Dolin, James Frew, Randall B. Kemp, Mary Larsgaard, Daniel R. Montello, Mary-Anna Rae, Jason Simpson, User Evaluation: Summary of the Methodologies and Results for the Alexandria Digital Library, Alexandria Digital Library Project, University of California, Santa Barbara, California, 1997
- [Hubert 01a] Frederic Hubert, Conception d'une interface Web pour l'aide à la saisie des besoins en information géographique, *La Recherche à l'IGN, Bulletin d'information de l'IGN n°72*, IGN (Eds), 2001, à paraître
- [Hubert 01b] Frederic Hubert, Assistance mechanisms use for needs specifications in geographical information on the Web, in *Proceedings of the 20th International Cartographic Conference*, Beijing, august 6-10, vol 4, pp2320-2329, 2001
- [Jarrold and Davis 96] William Jarrold, Tony Davis, The CYC system as an aid for GIS, presented at the NCGIA Specialist Meeting, Texas, 1996, NCGIA Technical Report 97-2
- [Jordan et al. 98] T. Jordan, M. Raubal, B. Gartrell, M. Egenhofer, An Affordance-Based Model of Place in GIS, 8th International Symposium on Spatial Data Handling, Vancouver, Canada, T. Poiker and N. Chrisman (Ed.), July 1998, pp. 98-109
- [Kashyap 97] Vipul Kashyap, Information Brokering over heterogeneous digital data : A metadata based approach, PhD Thesis, Large Scale Distributed Information System Laboratory, Department of Computer Science, University of Georgia, October 1997

- [Kuhn and Frank 91] Werner Kuhn, Andrew Frank, A formalization of metaphors and image-schemas in user interfaces, in D. Mark and A. Frank (eds.), *Cognitive and linguistic aspects of geographic space*, pp.419-434, Kluwer academic publisher, Netherlands, 1991
- [Kuiper 78] Ben Kuiper, Modeling spatial knowledge, *Cognitive Science*, Vol. 2, 1978, pp. 129-153
- [LaClef 01] La Clef, *An Operational Model for Unlocking Public Sector GI through E-Commerce*, LaClef Final Report PUB1102-LACLEF 25035/0, WP1/MEGR/ 030 / 11, Luxembourg, 2001
- [Lagoze et al. 96] Carl Lagoze, Clifford Lynch, Ron Daniel Jr., The Warwick Framework a Container Architecture for Aggregating sets of Metadata, result of Metadata Workshop II in Warwick U.K. 1996
- [Lakoff 92] George Lakoff, The Contemporary Theory of Metaphor, *Metaphor and Thought*, Ortony A. (ed), Cambridge University Press, 1992
- [Lamy et al. 99] Lamy S., Ruas A., Demazeau Y., Jackson M., Mackaness W.A. and Weibel R., The Application of Agents in Automated Map Generalisation. In *Proceedings of 19th International Cartographic Conference*, Ottawa., 1999, vol.2, pp.1225-1234.
- [Larson 96] Ray R. Larson, Geographic Information Retrieval and Spatial Browsing, in *GIS and Libraries: Patrons, Maps and Spatial Information*, edited by Linda Smith and Myke Gluck, University of Illinois, 1996, pp. 81-124
- [Lataillade 97] Philippe Lataillade, Les Métadonnées Géographiques, Etude d'un Système de Catalogue des Données Géographiques pour la Défense, Rapport d'Etude, 1997
- [Laurini and Thompson 92] R. Laurini, D. Thompson, *Fundamentals of spatial information systems*, Academic Press, San Diego, 1992
- [Laurini et Milleret-Raffort 93] Robert Laurini, Françoise Milleret-Raffort, *Les bases de données en géomatique*, Hermes, Paris, 1993
- [Laurini et al. 98] Robert Laurini, Sylvie Servigne, Thierry Ubeda, Alain Puricelli, La Réingénierie des Données géographiques : Problématiques Particulières, Réingénierie SI 98
- [Laurini and Gordillo 00] Robert Laurini, Sylvia Gordillo, Field Orientation for Continuous Spatio-temporal Phenomena, in *proceedings of the International Workshop on Emerging Technologies for Geo-based Applications*, Ascona, Switzerland, 2000, pp. 77-101 May 22-26, 2000
- [Lefébure et Venturi 98] René Lefébure, Gilles Venturi, *Le Data Mining*, ed Eyrolles, Paris, 1998
- [Lemarié 96] Cécile Lemarié, Etat de l'art sur l'appariement, rapport de recherche de l'Institut Géographique National, DT/9600022, juillet 1996
- [Lemarié and Badard 01] Cécile Lemarié, Thierry Badard, Cartographic database updating, in *Proceedings of the 20th International Cartographic Conference*, Beijing, august 6-10, 2001, vol 2, pp.1376-1385
- [Lenat and Guha 90] D. Lenat, R. Guha, Building Large Knowledge-based Systems, Representation and Inference in the CYC project, Addison-Wesley, Massachusetts, 1990
- [Le Roux 94] Bernard Le Roux, Eléments d'une Approche Constructive de la Modélisation et de la Réutilisation en Acquisition des Connaissances, Thèse de Doctorat de l'Université de Paris IV, 1994
- [Lopes et al. 97] Juliano Lopes de Oliveira, Fatima Pires, Claudia Bauzer Medeiros, An environment for modeling and design of geographic applications, in *Geoinformatica*, vol 1, Kluwer Academic Publisher, Boston, 1997, pp.29-58
- [Luzet 98] Claude Luzet, MEGRIN's GDDD, moving to distributed metadata, EOGEO98, Salzburg, February 1998
- [MacEachren and ICA 98] Alan MacEachren, the ICA Commission on Visualization, Visualization – Cartography for the 21st century, proceedings of the Polish Spatial Information Association conference, Warsaw, Poland, May 1998
- [MacEachren et al. 99a] Alan MacEachren, Monica Wachowicz, Robert Edsall, Daniel Haug, Constructing Knowledge from Multivariate Spatiotemporal Data, *IJGIS International Journal of Geographic Information Science* 1999
- [MacEachren et al. 99b] Alan MacEachren, Menno-Jan Kraak, Edward Verbree, Cartographic issues in the design and applications of geospatial virtual environments, 19th International Cartographic Conference, Ottawa, Canada, August 1999

- [Maguire 91] D. J. Maguire, An overview and definition of GIS, in *Geographical Information Systems : Principles and Application*, edited by D. J. Maguire, M. F. Goodchild and D. Rhind (Harlows : Longmans), 1991, vol1, pp.9-20
- [Maguire and Raper 90] D. J. Maguire, J. F. Raper, An overview of GIS functionality, in *Proceedings of the Design Models and Functionality Conference*, Midlands Regional Research Laboratory, Leicester, 1990
- [Maguire and Dangermond 91] D. J. Maguire, J. Dangermond, The functionality of GIS, in *Geographical Information Systems : Principles and Application*, edited by D. J. Maguire, M. F. Goodchild and D. Rhind (Harlows : Longmans), 1991, vol1, pp.319-335
- [Major 99] Wladimir Major, Approche de la concertation territoriale par l'analyse systémique et l'analyse lexicale du discours des acteurs. Perspectives d'application aux systèmes d'information géographique, Thèse en informatique, Ecole Polytechnique Fédérale de Lausanne, 1999
- [Marcus 88] Sandra Marcus (Ed.), *Automating Knowledge Acquisition for Expert Systems*, Boston: Kluwer Academic, 1988
- [Marcus and McDermott 89] Sandra Marcus, John McDermott, Salt: A knowledge acquisition language for propose-and-revise systems, *Artificial Intelligence*, No 39, 1989, pp1--37
- [Mark 92] David M. Mark, Spatial Metaphors for Human-Computer Interaction, *Spatial Data Handling SDH'92* vol 1
- [Mark et al. 97] David Mark, Max Egenhofer, Kathleen Hornsby , Formal Models of Commonsense Geographic Worlds, NCGIA technical report 97-2 1997
- [Mathet 00] Yann Mathet, *Étude de l'expression en langue de l'Espace et du Déplacement : analyse linguistique, modélisation cognitive, et leur expérimentation informatique*. Thèse de doctorat de l'université de Caen, décembre 2000
- [Mennis et al. 00] J. L. Mennis, D. J. Peuquet, L. Qian, A conceptual framework for incorporating cognitive principles into geographical database representation, in the *International Journal of Geographical Information Science*, 14(6), 2000, pp.501-520
- [Miller 98] G. Miller, Nouns in WordNet, in C. Fellbaum (ed.), *WordNet: an electronic lexical database, Language, Speech and Communication*, The MIT Press, pp. 23-46, 1998
- [Mitchell 99] Andy Mitchell, *The ESRI Guide to GIS Analysis*, volume 1 : Geographic patterns & relationships, USA, 1999
- [Motet 99] S.Motet, Multi-représentation : Projet MurMur, Cassini'2000, La Rochelle, septembre 2000
- [OpenGIS TC 98] OpenGIS Consortium Technical Committee, The OpenGIS ® Guide : Introduction to Interoperable Geoprocessing and the OpenGIS Specification, Edited by Kurt Buehler and Lance McKee, 3rd edition, June 3, 1998
- [OpenGIS 01] OpenGIS GML Working Group, *Geography Markup Language (GML) 2.0, OpenGIS® Implementation Specification*, OGC Document Number: 01-029, Simon Cox, Adrian Cuthbert, Ron Lake, Richard Martell (eds), Feb 2001
- [Openshaw 91] S. Openshaw, Developing appropriate spatial analysis methods for GIS, in *Geographical Information Systems : Principles and Application*, edited by D. J. Maguire, M. F. Goodchild and D. Rhind (Harlows : Longmans), 1991, vol1, pp.389-402
- [Ormeling 98] Ferjan Ormeling, Environmental mapping strategies, oral presentation at Intercarto, Brno, 1998
- [Papazoglou and Hoppenbrouwers 99] M. Papazoglou, J. Hoppenbrouwers, Knowledge Navigation in Networked Digital Libraries, Keynote speech for the 11th European Workshop on Knowledge Acquisition, Modeling, and Management (EKAW'99), in *Lecture Notes in Computer Science*, LNAI 1621, Springer Verlag, Heidelberg, 1999
- [Paepcke et al. 98] Andreas Paepcke, Chen-Chuan K. Chang, Héctor Garcia-Molina, Terry Winograd, Interoperability for Digital Libraries Worldwide, communication of the ACM, vol 41, n°4, april 98
- [Raper 96] Jonathan Raper, Unsolved problems of spatial representation, in *proceedings of SDH'96*, 1996, 14.1 – 14.11
- [Rhind 01] David Rhind, Global and National Geographic Information Policies, Practice and Education in a g-business world, in *GI in Europe : Integrative, Interoperable, Interactive, proceedings of the 4th AGILE conference*, april 2001, Brno, Czech Republic

- [Riedo et Mottier 02] Marc Riedo, Vincent Mottier, Principes et concepts des langages des SIG commerciaux, dans *Langages pour les SIG - Conception, développement et IHM*, Michel Mainguenaud (ed), Information Géographique et Aménagement du Territoire, Hermès Science Publication, Paris, 2002
- [Rodriguez and Egenhofer 99] Andrea Rodriguez, Max Egenhofer, Putting Similarity Assessments into Context: Matching Functions with the User's Intended Operations, in *the proceedings of Modeling and Using Context*, CONTEXT-99, Trento, Italy, P. Bouquet, L. Serafini, P. Brezillon, and F. Castellani (eds.), Lecture Notes in Artificial Intelligence, Vol. 1688, Springer-Verlag, pp. 310-323, September 1999
- [Role 99] François Role, Panorama des travaux en cours dans le domaine des métadonnées, rapport de Recherche de l'Institut National des Recherches en Informatique et en Automatique, numéro 3628, février 1999
- [Ruas 99] Anne Ruas, Modèle de généralisation de données géographiques à base de contraintes et d'autonomie, Thèse de doctorat en Sciences de l'Information Géographique de l'Université de Marne la Vallée, 1999
- [Schreiber et al. 99] A. Th. Schreiber, J. M. Akkermans A. A. Anjewierden, R. de Hoog, N. R. Shadbolt, W. Van de Velde, B. J. Wielinga, Knowledge Engineering and Management, The CommonKADS Methodology, MIT Press 1999
- [Sheth 99] Amit Sheth, Changing Focus on Interoperability in Information Systems : from System, Syntax, Structure to Semantic, *Interoperating Geographic Information Systems*, M. Goodchild, M. Egenhofer, R. Fegeas, C. Kottman (eds). Kluwer, 1999
- [Siméon 95] Jerome Siméon, Recherche en Texte Intégral et Bases de Données Orientées-Objet , Master thesis in Computer Science, Université Nancy, France, 1995
- [Siméon 99] Jérôme Siméon, Intégration de Sources de Données Hétérogènes ou Comment Marier Simplicité et Efficacité, Thèse de l'Université Paris XI en Informatique, Paris 1999
- [Spaccapietra et al. 99] S. Spaccapietra, C. Parent, E. Zimanyi, C. Vangenot. MurMur: A Research Agenda on Multiple Representation, in *Proceedings of the International Symposium on Database Applications in Non Traditional Environments*, DANTE'99, Kyoto, Japan, 1999
- [Spéry et Libourel 98] Laurent Spéry, Thérèse Libourel, Vers une structuration des métadonnées, Journées Cassini, Marne la Vallée, 1998
- [Spéry et al. 99] Laurent Spéry, Christophe Claramunt, Thérèse Libourel, A lineage metadata model for the temporal management of a cadastre application, in *proceedings of the International Workshop on Spatio-temporal Models and Language SDTDML'99*, A. M. Tjoa, A. Cammelli and R. R. Wagner (eds.), The IEEE Computer Society, Florence, 1999, pp.466-474
- [Timpf et al. 92] S. Timpf, G. Volta, D. Pollock, M. Egenhofer, A conceptual model of wayfinding using multiple levels of abstraction, *Theories and Methods of Spatio-temporal Reasoning in Geographic Space*, A. Frank, I. Campari, U. Formentini (eds), Springer-Verlag, Italia, 1992, pp.348-362
- [Timpf et al. 96] Sabine Timpf, Martin Raubal, Werner Kuhn, Experiences with Metadata, SDH'96
- [Timpf and Frank 97] Sabine Timpf, Andrew Frank, Using Hierarchical Data Structures for Hierarchical Reasoning pp.69-83
- [Tomlin 91] C. D. Tomlin, Cartographic Modelling, in *Geographical Information Systems : Principles and Application*, edited by D. J. Maguire, M. F. Goodchild and D. Rhind (Harlows : Longmans), 1991, vol1, pp.361-374
- [Tomlin and Berry 79] C. D. Tomlin, J. K. Berry, A mathematical structure for cartographic modelling in environmental analysis, in *Proceedings of the 39th Symposium*, American Congress on Surveying and Mapping, Virginia, 1979, pp.269-283
- [Traynor and Williams 95] Carol Traynor, Marian G. Williams, Why are Geographic Information Systems Hard to Use ? proceedings of CHI'95
- [Trichet et al. 00] Francky Trichet, Michel Leclère, Christophe Choquet, Construire un Système à Base de Connaissances de type Tâche/Méthode à l'aide des Graphes Conceptuels, actes de IC, Toulouse, 2000
- [Trifona and Sharma 96] Nectaria Tryfona, Jayant Sharma, On Information Modeling to Support Interoperable Spatial Databases, CAiSE pp. 210-221, 1996
- [Uschold and Jasper 99] M. Uschold, R. Jasper, A Framework for Understanding and Classifying Ontology Applications, IJCAI-99 Workshop on Ontologies and Problem-Solving Methods, Stockholm, 1999

- [Vckovski 98] Andrej Vckovski, Interoperable and distributed processing in GIS, PhD Thesis, University of Zurich, ed. Taylor & Francis, 1998
- [Voisard et Schweppe, 98] Agnès Voisard, Heinz Schweppe, Abstraction and decomposition in interoperable GIS, *International Journal of Geographic Information Systems*, 1998
- [Vossen et al. 97] P. Vossen, W. Peters, P. Díez-Orzas, The Multilingual design of the EuroWordNet Database, in Kavi Mahesh (ed.) *Ontologies and multilingual NLP, Proceedings of workshop at IJCAI-97*, Nagoya, Japan, August 23-29, 1997
- [Weibel 91] R. Weibel, Amplified Intelligence and Knowledge-Based Systems, in Buttenfield, B.P. and McMaster, R.B. (eds.), *Map Generalization: Making Rules for Knowledge Representation*, London: Longman, 1991, pp.172-186.
- [Weibel et al. 95] Stuart Weibel, Jean Godby, Eric Miller, Ron Daniel, OCLC/NCSA Metadata Workshop Report, 1995
- [Willamowski 92] Jutta Willamowski, *Modélisation de tâches pour la résolution de problèmes en coopération système-utilisateur*, Thèse en informatique à l'Université Joseph Fourier, Grenoble, 1992
- [Worboys 96] M. F. Worboys, Metrics and topologies for geographic space, in *Advances in GIS research II, proceedings of 7th International Symposium on Spatial Data Handling*, Kraak and Molenaar (eds), Taylor and Francis, 1996, pp.365-375.
- [W3C/XML/WG 00] W3C XML Core Working Group, *W3C Recommendation: Extensible Markup Language (XML) 1.0*, Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler (eds), 2nd edition, 2000
- [Zeitouni et Yeh, 00] Karine Zeitouni, Laurent Yeh, Le Data Mining Spatial et les bases de données spatiales, in *Actes des Journées Data Mining Spatial et Analyse du Risque*, Versailles, 2000

Bénédicte Bucher, Indexer les données géographiques par leurs rôles dans la détermination de connaissances d'utilisateurs, *Bulletin d'Information de l'IGN n°70* (1999/3), Saint Mandé, France, 1999

Bénédicte Bucher, Users' access to geographic information resources : a model of tasks and roles to specify intentional uses regarding available resources, in *Geographical Domain and Geographical Information Systems*, Winter, S. (Ed.), GeoInfo Series Vol. 19. Institute for Geoinformation, Vienna University of Technology, Vienna, Austria, 2000

Bénédicte Bucher, A Model to Store and Reuse Geographic Application Patterns, in *GI in Europe : Integrative, Interoperable, Interactive, proceedings of the 4th AGILE conference*, april 2001, Brno, Czech Republic

Bénédicte Bucher, Structuring and enriching metadata to enable users' access to geographic information resources, in *proceedings of the 20th International Cartographic Conference*, Beijing, China, vol 4, pp. 2791-2797, 2001