

# *Projet SHERPA*

*Modèles et bases de connaissances à objets*

*Rhône-Alpes*

THÈME 3A

*R* *apport*  
*d'Activité*

1999



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>3</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>4</b>
<b>3</b>	<b>Fondements scientifiques</b>	<b>7</b>
3.1	L'utilisation de la classification conceptuelle pour modéliser . . . . .	7
3.2	La construction collaborative de modèles . . . . .	7
3.3	La visualisation et la diffusion des bases . . . . .	8
3.4	La résolution de problèmes et les modèles de tâches . . . . .	9
3.5	La construction et l'exploitation de modèles de systèmes dynamiques . . . . .	10
<b>4</b>	<b>Domaines d'applications</b>	<b>10</b>
4.1	Panorama . . . . .	10
4.2	Génomique . . . . .	10
4.3	Mémoire d'entreprise . . . . .	11
4.4	Aide au diagnostic médical . . . . .	12
4.5	Modélisation et simulation en sciences sociales . . . . .	13
<b>5</b>	<b>Logiciels</b>	<b>13</b>
5.1	Panorama . . . . .	13
5.2	AMIA . . . . .	14
5.3	APIC . . . . .	14
5.4	AROM . . . . .	14
5.5	FUSIN . . . . .	15
5.6	GNA . . . . .	16
5.7	POWERTASKS et IMAGENE . . . . .	16
5.8	Protocole de CO4 . . . . .	17
5.9	TROEPS . . . . .	17
<b>6</b>	<b>Résultats nouveaux</b>	<b>18</b>
6.1	L'utilisation de la classification conceptuelle pour modéliser . . . . .	18
6.1.1	Construction de taxonomies à partir d'instances . . . . .	18
6.1.2	Classification conceptuelle pour l'extraction de connaissances . . . . .	19
6.2	La visualisation et la diffusion des bases . . . . .	20
6.2.1	Recherche d'informations sur le Web . . . . .	20
6.2.2	XML et objets . . . . .	20
6.2.3	Visualisation de l'information . . . . .	21
6.3	La résolution de problèmes et les modèle de tâches . . . . .	21
6.3.1	Résolution de problèmes répartis . . . . .	21
6.3.2	Comparaison de graphes de tâches . . . . .	22
6.3.3	Pilotage de bibliothèques de composants . . . . .	22
6.4	La construction et l'exploitation de modèles dynamiques . . . . .	22
6.4.1	Simulation sous contraintes . . . . .	22

6.4.2	Problèmes d'optimisation combinatoire sous contraintes ensemblistes . . .	23
6.4.3	Construction assistée de modèles par assemblage de fragments . . . . .	23
6.4.4	Analyse qualitative et semi-quantitative des systèmes dynamiques complexes . . . . .	24
<b>7</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>25</b>
7.1	Génie . . . . .	25
7.2	PROSPER . . . . .	25
7.3	Xerox Research Centre Europe (XRCE) . . . . .	26
7.4	PSA Peugeot Citroën . . . . .	26
<b>8</b>	<b>Actions régionales, nationales et internationales</b>	<b>27</b>
8.1	Actions nationales . . . . .	27
8.1.1	Action de recherche coopérative ESCRIRE . . . . .	27
8.1.2	Action « Gestion de modèles et de connaissances - Application à la ville » du programme CNRS « Modélisation et simulation numérique » .	28
8.1.3	Action de recherche coopérative LIRE . . . . .	28
8.1.4	OFTA . . . . .	29
8.1.5	CEA-INRIA . . . . .	29
8.1.6	Projet « Bioinformatique des interactions moléculaires et des réseaux régulateurs à l'échelle des génomes » . . . . .	30
8.2	Actions européennes . . . . .	30
8.2.1	Programme d'actions intégrées franco-néerlandais VAN GOGH . . . . .	30
8.2.2	Projet européen EMG-NET . . . . .	31
8.2.3	Collaboration franco-bulgare (programme INCO) . . . . .	31
8.2.4	Coopération scientifique et technique lusitano-française : modélisation d'un serveur de cas médicaux . . . . .	31
8.3	Actions internationales . . . . .	32
8.3.1	Université du Québec à Montréal . . . . .	32
<b>9</b>	<b>Diffusion de résultats</b>	<b>32</b>
9.1	Animation de la Communauté scientifique . . . . .	32
9.2	Enseignement . . . . .	33
9.3	Participation à des colloques, séminaires, invitations . . . . .	33
<b>10</b>	<b>Bibliographie</b>	<b>35</b>

---

## 1 Composition de l'équipe

### Responsable scientifique

François Rechenmann [DR Inria]

### Assistante de projet

Françoise de Coninck

### Personnel Inria

Gilles Bisson [CR]

Jérôme Euzenat [CR]

Hidde de Jong [CR]

Danielle Ziébelin [CR (MdC UJF, en détachement à l'INRIA jusqu'au 1er octobre 1999)]

### Personnel de l'université Pierre Mendès-France (UPMF)

Jérôme Gensel [MdC]

Michel Page [MdC]

Daniel Bardou [MdC, depuis le 1<sup>er</sup> octobre 1999]

Sylvie Robert [IR (mi-temps) jusqu'en juin 1999]

### Collaborateur extérieur

Philippe Genoud [MdC université Joseph Fourier (UJF), Grenoble]

### Ingénieurs experts

Christophe Bruley

**Chercheurs doctorants**

Jean-Christophe Bottraud [salarié IBM]

Alexandre Faure [CIFRE (PSA Peugeot Citroën)]

Sueli Ferreira da Cunha [boursière CAPES, Brésil, thèse soutenue le 18 janvier 1999]

Jean-Marc Gabriel [allocataire Ministère, thèse soutenue le 25 juin 1999]

Camille Parisel [allocataire Ministère]

Thibaut Parmentier [allocataire Ministère]

Denys Proux [CIFRE (XRCE : Xerox Research Centre Europe)]

Vincent Thornary [allocataire Ministère]

Petko Valchev [ATER UJF, thèse soutenue le 15 décembre 1999]

**Chercheur post-doctorant**

Irena Koprinska [Académie des Sciences, Sofia, Bulgarie (jusqu'au 31 janvier 1999)]

**Stagiaires CNAM**

Véronique Chabanis [jusqu'au 20 novembre 1999]

Martine Maume [jusqu'au 15 septembre 1999]

Alain Uginet [jusqu'au 31 mars 1999]

Véronique Dupierris [depuis le 2 novembre 1999]

**2 Présentation et objectifs généraux**

**Mots clés :** AMIA, APIC, argumentation, AROM, base de connaissances, biologie moléculaire, catégorisation, classification conceptuelle, CO4, composants, contraintes, coopération, diagnostic médical, électromyographie, extraction d'information, FUSIN, génome, génomique, GNA, IMAGENE, indexation, langage de modélisation algébrique, mémoire d'entreprise, mémoire technique, modèle à objets, modèle de tâches, modélisation, modélisation de systèmes dynamiques, MYOSYS, pilotage de bibliothèques de composants, points de vue, relations, représentation de connaissances, résolution de problèmes, révision, santé, similarité, simulation, système à base de connaissances, TROEPS, typage, World Wide Web, XML.

Les systèmes à base de connaissances reposent sur la recherche d'une séparation explicite entre l'expression des connaissances et les mécanismes d'exploitation de ces connaissances

pour la résolution de problèmes. Cette séparation est réputée apporter des avantages décisifs quant aux possibilités, d'une part de faire évoluer plus aisément les connaissances, d'autre part de fournir des justifications sur les solutions obtenues et des explications sur le processus de résolution. Ces avantages sont particulièrement appréciables lorsque les connaissances à modéliser sont non consensuelles car, d'une part, il est difficile de les expliciter en une seule fois et le processus de construction de la base de connaissances est par essence incrémental, d'autre part, les solutions qu'elles permettent d'obtenir doivent être argumentées.

Cependant le développement de bases de connaissances peut se concevoir en dehors de toute préoccupation de résolution de problèmes. En effet, une base de connaissances permet de rendre explicite un modèle de domaine, en particulier scientifique ou technique. Dès lors, le bénéfice d'un processus de modélisation réside plus dans l'accroissement de la compréhension des entités et des phénomènes modélisés que dans l'obtention du modèle lui-même. Cet accroissement de compréhension résulte de l'explicitation et de la caractérisation des entités impliquées, ainsi que de leurs interrelations. Par exemple, la modélisation des connaissances dans un domaine comme la biologie moléculaire vise d'abord, à travers l'effort d'explicitation, de structuration et d'organisation des connaissances, à obtenir une meilleure compréhension des entités et de leurs interactions. Certes, la base résultante peut disposer de capacités prédictives et calculatoires pertinentes, mais sa seule existence justifie sa construction.

Concrètement, les bases de connaissances scientifiques et techniques rassemblent et structurent des connaissances de plusieurs natures : descriptives sur les entités du domaine concerné, fonctionnelles sur le comportement de ces entités, et prescriptives sur les moyens de compléter et caractériser ces descriptions. De plus, ces connaissances formalisées doivent être associées à des connaissances terminologiques et textuelles destinées à les justifier et à les commenter.

Dans ce contexte, les thèmes de recherche du projet Sherpa s'articulent autour de la *conception de modèles de connaissances à objets* adaptés aux exigences des bases de connaissances scientifiques et techniques, ainsi que de l'étude des *interactions avec ces modèles* lors des phases de conception et d'utilisation. À un niveau plus général, l'objectif scientifique du projet est de définir des environnements d'aide à la découverte, à la gestion et à l'exploitation des connaissances. Sur cette base, le projet est organisé autour des cinq axes de recherche suivants :

- l'utilisation de la classification conceptuelle pour modéliser ;
- la construction collaborative de modèles et de bases de connaissances ;
- la diffusion des bases de connaissances ;
- la résolution de problèmes et les modèles de tâches pour le pilotage de code ;
- la construction automatisée de modèles dynamiques.

Ces axes de recherche ne sont bien évidemment pas disjoints et plusieurs travaux effectués par le projet relèvent simultanément de deux d'entre-eux. Voici une vision plus détaillée du projet.

Tout formalisme à objets repose sur la construction de hiérarchies de classes qui constituent la structure pivot de la base de connaissances. Leur élaboration correspond à une formalisation du domaine qui n'est généralement pas une tâche aisée. Il est donc intéressant de fournir une

aide pour leur construction. C'est le but des travaux du projet portant sur la classification conceptuelle (ou catégorisation) : ils visent à construire ces hiérarchies de classes à partir d'un ensemble d'instances.

Toutefois, même si de tels outils aident efficacement la modélisation, il est généralement nécessaire d'effectuer plusieurs itérations pour parvenir à des résultats satisfaisants. Au cours de ces itérations, il faut donc mettre en place une collaboration entre le système de classification et l'utilisateur. Une telle approche n'est toutefois pas universelle car, outre le fait que l'on ne fait pas toujours appel au processus de classification, dans de nombreux cas, la construction d'une base de connaissances (ou d'une mémoire technique) nécessite l'intervention de plusieurs personnes sur des sites distants. Dans ce contexte, il s'agit de mettre en place un processus incrémental et collaboratif prenant en compte la gestion des conflits pouvant apparaître lorsque des utilisateurs modifient simultanément la base. Effectuées autour du système TROEPS, les contributions du projet ont permis de constituer deux bases de connaissances dans les domaines de la biologie du développement et des mémoires d'entreprise.

Le Web est aujourd'hui le moyen privilégié pour diffuser et acquérir des connaissances, c'est donc tout naturellement que le système TROEPS fait collaborer les différents concepteurs de la base à l'aide de ce média. C'est également le cas du système AROM qui offre des outils pour la consultation de bases de connaissances à partir d'un navigateur Web. Le système FUSIN, quant à lui, vise plus spécifiquement à définir des mécanismes interactifs de présentation et de recherche d'informations. L'objectif de ce travail est, entre autres, de concevoir des interfaces d'acquisition d'instances facilitant la mise en œuvre de processus de classification conceptuelle. Notons que d'autres travaux effectués par le projet ont porté également sur le thème de la présentation des bases de connaissances à l'utilisateur. Cette visualisation s'effectue via des interfaces spécialisées, par exemple dédiées à la représentation de connaissances génomiques, ou encore dans le cadre de recherches plus fondamentales visant à concevoir de nouveaux moyens d'exploration et d'interaction, à l'aide de représentations tridimensionnelles.

Si, comme il a été déjà dit, la constitution des modèles de connaissances peut constituer une fin en soi, il est néanmoins intéressant de mettre en œuvre ces modèles pour effectuer des résolutions de problèmes. Le projet Sherpa a acquis un savoir-faire important dans la construction des environnements coopératifs d'aide à la résolution de problèmes tels que POWERTASKS. Ces environnements reposent sur des modèles de tâches permettant de décrire des schémas de résolution de problèmes par décomposition récursive en sous-problèmes plus simples. Là encore, les travaux du projet s'orientent vers des modes de résolution utilisant le Web comme média de communication. De la sorte, il devient possible de travailler de manière transparente avec des bibliothèques de composants développées et maintenues dans des environnements différents. Ces travaux sont actuellement appliqués dans le projet européen EMG-NET.

Enfin, la modélisation de domaines dans lesquels les entités et leurs inter-relations évoluent dans le temps nécessite d'introduire la description de la dynamique d'un système dans les modèles de connaissances : c'est typiquement le cas en biologie du développement où il faut modéliser les interactions entre gènes. Fort de l'expérience acquise ces dernières années, le projet développe un nouveau modèle à objets, implémenté dans le système AROM, qui intègre un langage algébrique permettant de décrire la dynamique d'un système et d'effectuer des simulations.

## 3 Fondements scientifiques

### 3.1 L'utilisation de la classification conceptuelle pour modéliser

La classification d'instances et de classes dans les formalismes à objets s'appuie sur une organisation hiérarchique des classes : le graphe de spécialisation. Celui-ci reflète la vision particulière que détient un spécialiste du domaine à modéliser, et il constitue de ce fait la structure pivot de la base de connaissances à objets. La construction du graphe de spécialisation est cependant une tâche complexe qui exige un effort de formalisation important. Dès lors, il est utile de fournir au concepteur de la base de connaissances une aide pour sa construction.

Il est ainsi intéressant de détecter automatiquement, au sein d'un ensemble d'instances représentatives de la discipline étudiée, un certain nombre de groupes pertinents qui pourront constituer des classes. Cette problématique se rattache typiquement au domaine de la classification conceptuelle, encore appelée catégorisation, dont les objectifs sont d'une part, de détecter, au sein d'un ensemble de données, des groupements de données à forte similarité, et, d'autre part, de « résumer » ces regroupements sous la forme de classes hiérarchisées, symboliquement caractérisables à l'aide d'un ensemble de propriétés à la fois homogènes et contrastées. En simplifiant et en réduisant la complexité des données d'un problème, ces résumés facilitent la compréhension, l'analyse du problème même et, par conséquent, ils aident la modélisation du domaine sous-jacent.

Les travaux effectués jusqu'à présent en classification conceptuelle se rattachent aux domaines de l'apprentissage automatique et de l'analyse des données. Or, dans ces deux approches, les langages de représentation utilisés sont beaucoup moins structurés que ceux développés dans le domaine des représentations à objets. Ce thème a donc pour objectif d'étendre les méthodes existantes au formalisme à objets tout en préservant les vertus d'explicabilité qui caractérisent cette approche.

### 3.2 La construction collaborative de modèles

Les réseaux informatiques permettent non seulement la diffusion de modèles (sous forme active ou passive), mais également leur élaboration à distance. Dans ce contexte, le projet Sherpa développe des travaux sur les outils informatiques s'appuyant sur la structure de la connaissance pour faire collaborer plusieurs intervenants durant les phases de construction et d'exploitation des modèles.

La construction incrémentale et collaborative d'une base de connaissances (ou d'une mémoire technique) fait intervenir simultanément plusieurs personnes sur des sites distants. Chacune est susceptible de soumettre des descriptions de classes d'objets interreliées pour les intégrer dans la base en cours de construction. Les échanges entre la base et les sites qui l'alimentent doivent être gérés afin d'assurer que le contenu de la base de connaissances croisse de manière consensuelle.

Le protocole de collaboration CO<sub>4</sub>, conçu au sein du projet Sherpa, permet à plusieurs utilisateurs de soumettre à une base de connaissances consensuelle des éléments de connaissance. Il s'appuie sur la métaphore de la soumission d'un article à un journal scientifique : les contributions sont distribuées à l'ensemble des souscripteurs de la base qui doivent répondre s'ils l'acceptent, la rejettent ou proposent des modifications. Le protocole gère la collecte des

réponses et la décision qui est produite. Il gère de la même manière l'arrivée d'un nouveau souscripteur et permet récursivement à une base consensuelle de souscrire à une autre base.

La mise en œuvre complète d'un système de construction collaboratif de bases de connaissances nécessite d'apporter une aide aux utilisateurs dans la gestion des conflits qui peuvent apparaître lorsqu'ils modifient la base. En effet, lorsqu'une base de connaissances est construite de manière incrémentale, chaque nouvel ajout peut être contradictoire avec le contenu de la base. Il faut donc modifier la base afin de préserver sa consistance. Cette problématique se rapproche de celle de la révision dans les langages logiques, dont le principal inconvénient est la complexité qui ne permet pas l'implémentation. L'objectif est justement de définir une révision, dans le cadre des représentations de connaissances par objets, qui puisse être implémentée, notamment en tenant compte des particularités de ces langages.

Dans ce contexte, il faut travailler avec un langage doté d'une sémantique définie de manière formelle et disposant d'un système déductif syntaxique correct et complet par rapport à cette sémantique. L'inconsistance y est caractérisée syntaxiquement, ce qui permet de la détecter en se basant sur les propriétés de localité du langage : contrairement aux langages logiques classiques, l'inconsistance ne se propage pas à l'ensemble de la base, elle ne permet pas de tout déduire. Par ailleurs, la notion de base révisée satisfait les postulats classiques de la révision, en particulier la minimalité, principe selon lequel il faut perdre le moins de connaissances possibles. La minimalité peut, de plus, s'interpréter intuitivement selon la relation d'ordre entre les classes. Cependant, la minimalité n'est pas un critère suffisant pour obtenir une base unique. Un algorithme est donc à mettre en œuvre pour obtenir toutes les bases révisées minimales suite à un ajout inconsistant ; il est interactif afin de maîtriser la complexité inhérente à la révision. Concrètement, un tel algorithme a été développé dans le système TROEPS et permet donc, lors de l'édition d'une base de connaissances, de proposer à l'auteur d'une modification fautive une aide dans la révision de la base.

### 3.3 La visualisation et la diffusion des bases

Le problème de l'interface avec l'utilisateur est capital pour l'acceptation et l'exploitation des systèmes qui mettent en œuvre de grandes quantités de connaissances. Très tôt le projet Sherpa s'est attaché à fournir des interfaces graphiques au-dessus des modèles et langages développés pour la représentation des connaissances. Cependant, les développements conjoints de l'informatique et des réseaux de communication et le rôle de plus en plus important qu'ils jouent dans le cycle de l'information, ont profondément bouleversé la manière dont nous accédons et utilisons l'information. C'est dans ce contexte que le projet Sherpa poursuit des recherches dans deux directions complémentaires : la mise à disposition et l'exploitation, via des serveurs HTTP, de bases de connaissances et la visualisation d'informations à l'aide de nouveaux paradigmes.

Ainsi, les systèmes TROEPS et AROM se comportent comme des serveurs de connaissances accessibles depuis un client HTTP. Ils permettent de naviguer dans la structure de la connaissance, de l'éditer ou encore de soumettre des requêtes structurées. D'autres travaux visent à définir des modèles de documents qui facilitent l'interaction avec l'utilisateur et la recherche d'informations sur le Web.

Néanmoins, les systèmes actuels ne répondent pas toujours aux besoins des utilisateurs qui

restent démunis pour appréhender de grandes masses d'informations. La transposition de la représentation de l'information dans le domaine visuel, et plus particulièrement la construction et l'exploitation de représentations graphiques tridimensionnelles, constituent un élément de réponse à ces difficultés. Cependant, alors que la visualisation d'objets réels s'appuie sur les informations géométriques qui leur sont intrinsèques, la visualisation d'informations nécessite de donner une matérialisation visuelle à des entités qui ne possèdent pas nécessairement une géométrie, et une interprétation de cette matérialisation afin de mettre en avant certaines de leurs propriétés (internes ou organisationnelles). Il est donc important de connaître les caractéristiques tant sémiologiques que cognitives du système de signes et de composantes sur lesquelles sont construites les représentations graphiques.

### 3.4 La résolution de problèmes et les modèles de tâches

Les modèles de tâches permettent de décrire des schémas de résolution de problèmes sous la forme d'une décomposition récursive du problème initial en sous-problèmes plus simples. Cette décomposition se termine lorsqu'on obtient des problèmes élémentaires solubles par l'exécution d'un fragment de code. L'utilisation d'un modèle de tâches sur un problème donné s'effectue de la manière suivante : à chaque pas de décomposition, le problème courant, représenté par une instance de tâche, est tout d'abord *caractérisé* par sa classification dans une hiérarchie de tâches de même catégorie, puis il est *décomposé* selon les prescriptions de la tâche qui a été sélectionnée. Il en résulte donc une analyse particulière en sous-problèmes et le cycle caractérisation-décomposition se répète sur chacun de ces sous-problèmes. En outre, la décomposition est réalisable de manière coopérative : ainsi, les utilisateurs peuvent à tout moment arrêter une décomposition, modifier la description du problème et comparer plusieurs schémas de résolution entre eux. Les modèles de tâches sont ainsi bien adaptés à la formalisation et l'exploration d'un processus de résolution quelconque.

Ces modèles de tâches sont mis en œuvre dans des environnements coopératifs d'aide à la résolution de problèmes. Le projet Sherpa a ainsi développé, de 1994 à 1996, en collaboration avec la société Ilog, l'environnement POWERTASKS. Ultérieurement, une version répartie de cet environnement a été réalisée : elle permet d'exécuter les méthodes associées aux tâches élémentaires sur les sites mêmes où elles sont développées et maintenues. Pour l'utilisateur, ce processus reste totalement transparent : tout se passe comme si le schéma de décomposition et les méthodes élémentaires étaient présents sur son site. L'avantage d'une telle approche est double. D'une part, on élimine les problèmes de portabilité des applications car elles restent sur les sites où elles ont été développées et d'autre part, il devient envisageable de paralléliser certains traitements.

Toutefois, la mise en œuvre d'une exécution répartie nécessite la définition de mécanismes de localisation des compétences et l'élaboration de processus de synchronisation permettant aux tâches réparties d'interagir entre les différents sites. Ces deux points entraînent évidemment des modifications structurelles et opérationnelles du modèle de tâches classique. Par ailleurs, l'utilisation d'un environnement réparti oblige à prendre en considération des contraintes temporelles et de disponibilité des ressources matérielles et logicielles. En d'autres termes, les stratégies de résolution doivent tenir compte de la puissance de calcul des machines, de leur charge de travail et de la disponibilité des compétences et des données.

### 3.5 La construction et l'exploitation de modèles de systèmes dynamiques

La représentation de connaissances scientifiques et techniques concerne fréquemment des domaines où les entités et leurs interrelations évoluent dans le temps. C'est typiquement le cas en biologie du développement pour la modélisation des interactions géniques qui mettent en jeu des réseaux complexes d'activations et d'inhibitions. L'objectif poursuivi par le projet Sherpa est d'obtenir une automatisation, ou tout au moins une assistance, pour les différentes tâches du processus de modélisation et de simulation de systèmes dynamiques. Ces tâches concernent la construction, la révision et la validation du modèle, la réalisation des simulations, l'analyse des résultats de simulation et la réutilisation de sous-modèles.

Dans cette optique, le projet Sherpa étudie l'interaction entre les connaissances symboliques et numériques qui caractérisent les systèmes dynamiques. Les premières décrivent les entités du domaine d'application, mais également des relations entre ces entités qui ne mettent pas directement en jeu des quantités numériques (choix de modélisation, conditions de validité des équations...). Les secondes décrivent les équations et les contraintes liant les propriétés numériques des entités du domaine (variables du modèle) dans un formalisme mathématique. Le projet propose des formalismes permettant de représenter de manière homogène ces deux types de connaissances ainsi que des mécanismes capables de les exploiter, par exemple, pour réaliser de la simulation sous contraintes ou effectuer une construction automatique de modèles à partir de fragments de modèles.

## 4 Domaines d'applications

### 4.1 Panorama

Le projet Sherpa cultive trois grands domaines d'expérimentation : la génomique, les mémoires d'entreprise et l'aide au diagnostic médical. Un thème plus récent concerne l'aide à l'élaboration et à l'utilisation de modèles de systèmes dynamiques, en particulier dans les domaines socio-économiques.

### 4.2 Génomique

**Résumé :** *De plus en plus de séquences de génomes complets sont maintenant disponibles, en particulier provenant d'organismes bactériens. Ces « textes » doivent être interprétés afin d'identifier des objets biologiquement significatifs et d'en comprendre le fonctionnement. Le rôle de l'informatique est ici prépondérant, tant en ce qui concerne les méthodes d'analyse de séquences et de modélisation du fonctionnement du génome que pour la représentation et la gestion des connaissances que ces processus exploratoires produisent.*

La macromolécule d'ADN est le support matériel de l'hérédité et la mémoire du processus d'évolution : son génome. Sous la forme d'une chaîne d'acides nucléiques, elle contient l'information nécessaire au développement et au maintien de tout organisme vivant. Certains fragments de la molécule d'ADN jouent un rôle particulièrement important : ce sont les gènes, qui codent les protéines. Ces dernières, qui assurent la plupart des fonctions enzymatiques et de structure

dans les cellules, sont également des macromolécules constituées par l'enchaînement de vingt acides aminés différents.

Le contenu informationnel d'un brin d'ADN peut être décrit par une séquence des lettres A, C, G et T, qui symbolisent les bases des quatre acides nucléiques : adénine, cytosine, guanine et thymine. Un génome peut ainsi être abstrait sous la forme d'un « texte » écrit dans un alphabet de quatre lettres pouvant faire l'objet de traitements informatiques. Il faut en effet interpréter ces « textes », c'est-à-dire localiser les gènes, comprendre les mécanismes qui régulent leur traduction en protéines, identifier et caractériser ces protéines, rechercher des similarités entre séquences d'espèces différentes, etc. Cette démarche exploratoire fait un usage intensif de l'informatique, tant pour ses possibilités de stockage, de gestion des données et de calcul, que de représentation d'objets complexes, de leurs relations et de leurs comportements.

Ainsi, le génome joue un rôle central dans le déroulement des processus biologiques fondamentaux comme le métabolisme et le développement. La protéine produite par un certain gène peut catalyser une réaction dans une voie métabolique, par exemple, ou réguler l'expression d'un autre gène. Afin de comprendre ces processus, les réseaux d'interactions entre les gènes, leurs produits et d'autres substances biochimiques doivent être identifiés et étudiés. Étant données la taille et la complexité de ces réseaux, l'utilisation d'outils informatiques s'avère indispensable pour leur analyse. Deux axes de recherche sont actuellement explorés dans le projet. D'une part, nous nous intéressons à la simulation des systèmes d'interactions décrits par des formalismes qualitatifs ou semi-quantitatifs. Cette simulation permet à la fois de tester les modèles biologiques et de prédire des phénomènes qui pourront faire l'objet de validations « in vitro ». D'autre part, pour aider à la conception de tels modèles nous cherchons à mettre en œuvre des outils d'analyse linguistique capables de faire de l'extraction de connaissances à partir des bases documentaires dont disposent les biologistes.

Dans ce contexte, le projet Sherpa travaille en étroite collaboration avec plusieurs laboratoires de biologie moléculaire (Laboratoire de Biométrie, Biologie Évolutive (UMR 5558), université Claude Bernard, Lyon ; équipe d'Antoine Danchin à l'institut Pasteur de Paris et Atelier de Bio-Informatique (ABI) de Paris 6 ; laboratoire de génétique et physiologie du développement de Marseille) sur le développement de bases de connaissances sur les génomes et d'environnements d'aide à l'analyse de séquences.

### 4.3 Mémoire d'entreprise

**Résumé :** *Une mémoire d'entreprise est techniquement vue comme l'association de connaissances formelles et de documents plus ou moins structurés. Afin de permettre de gérer les dimensions sociales et historiques de cette association, il convient de concevoir, de développer et d'expérimenter des outils de construction incrémentale et coopérative de telles mémoires techniques.*

Le terme « mémoire d'entreprise » participe de la problématique de la gestion des connaissances (« knowledge management ») ; il désigne des systèmes informatiques susceptibles d'accroître les capacités de capitalisation et de gestion des connaissances et des expériences au sein des entreprises. On parlera ici plutôt de « mémoire technique » destinée à recevoir la connaissance technique utilisée par les ingénieurs de l'entreprise. Cette mémoire se doit d'être vivante

parce qu'elle est utilisée et enrichie de manière quotidienne. Elle doit être, en outre, cohérente et intelligible. Notre approche de la mémoire technique est nourrie de notre expérience de la construction de bases de connaissances. Trois principes la gouvernent :

- les mémoires techniques seront d'autant plus utiles qu'elles seront exprimées dans un langage formel permettant d'effectuer un certain nombre de vérifications et de manipulations ;
- toutes les informations ne pouvant s'exprimer de manière formelle, il est nécessaire de lier cette connaissance aux documents informels qui la concernent (annotations, bibliographie, lexique) ;
- les « rédacteurs » de cette mémoire technique doivent être aidés dans cette démarche.

Les deux premiers principes justifient les développements (lexique, déploiement sur le réseau) et les études effectuées autour de notre système de représentation de connaissances TROEPS. Celui-ci contribue à atteindre les objectifs d'intelligibilité en fournissant un modèle de la mémoire clair et accessible par un client HTTP quelconque. En outre, le contenu est relié directement, par le même moyen, à des bases de données, des pages de documentation et des lexiques. Le troisième principe nécessite le développement de recherches et d'outils nouveaux pour la révision et la construction coopérative de bases de connaissances, qui permettent de renforcer la cohérence des mémoires ainsi mises en œuvre. Des expériences de mémoire d'entreprise sont actuellement menées en collaboration avec les sociétés STMicroelectronics, Dassault Aviation, Dassault Systèmes et PSA Peugeot Citroën.

#### 4.4 Aide au diagnostic médical

**Résumé :** *L'électromyographie est une technique de diagnostic des maladies neuromusculaires. Elle utilise des protocoles d'examen lourds et parfois douloureux pour le patient. Il est donc pertinent d'aider un praticien à choisir les examens les plus appropriés pour une situation donnée.*

Les maladies neuromusculaires, souvent de nature dégénératives, sont encore très mal connues. Certaines sont dues à des facteurs génétiques (c'est le cas des myopathies), d'autres à des neurotoxiques ou à des raisons inconnues comme dans le cas de la maladie de Creutzfeld-Jacob. L'étude de ces maladies, via les protocoles d'électro-myographie, impose au patient un suivi régulier et fréquent, ainsi que des examens pénibles et coûteux. Il est donc important de fournir aux praticiens des outils informatiques d'aide à la décision leur permettant, d'une part d'assurer un suivi optimal de chaque patient, et, d'autre, part de mieux comprendre, à l'échelle des populations concernées, les évolutions de ces maladies.

En effet, les études épidémiologiques apportent une meilleure connaissance des paramètres neurologiques et de leur corrélation avec des connaissances anatomiques, physiologiques ou cliniques ; elles apportent également une meilleure caractérisation de l'évolution, de l'évolutivité et de la chronicité de ces maladies. Dans ce contexte, la recherche médicale peut tirer grandement partie des technologies basées sur Internet qui autorisent, par exemple, un accès aisé à des dossiers de patients ayant contracté des pathologies neurologiques rares telles que la maladie

de Creutzfeld-Jacob, l'artériosclérose amyotrophique, la maladie de Lambert-Eaton, ainsi que certaines myopathies ou neuropathies. Par ailleurs, les experts confrontés au diagnostic de ces maladies souhaitent souvent le concours de confrères afin de comparer leurs analyses et de permettre ainsi un meilleur suivi du patient. Dès lors, la mise à disposition d'outils collaboratifs, tels ceux définis au sein du projet Sherpa, apporte une aide appréciable dans ce dialogue.

## 4.5 Modélisation et simulation en sciences sociales

**Résumé :** *L'adjonction à des modèles mathématiques de connaissances explicitant les choix de modélisation et les hypothèses de validité des équations, favorise leur réutilisation dans des contextes différents de ceux de leur élaboration. Cette approche s'avère particulièrement féconde pour la modélisation et la simulation en sciences sociales.*

Pour permettre la construction de modèles réutilisables, deux approches sont envisageables. La première consiste à proposer des outils permettant de développer des familles de modèles génériques qui puissent être adaptés à différents contextes d'utilisation. La seconde consiste à construire une bibliothèque de fragments de modèles réutilisables pouvant être sélectionnés et assemblés par des algorithmes appropriés.

La première approche est expérimentée en collaboration avec l'ADEME (Agence pour le Développement et la Maîtrise de l'Énergie) sur la prévision de la demande en énergie. Il s'agit de prévoir la consommation à long terme en énergie en fonction de différents scénarios politiques, économiques et techniques. La famille de modèles développée dans le cadre de cette application (MEDEE) est de grande taille. Elle se compose de plus de 1000 équations qui sont implantées grâce au logiciel AMIA. Cette application est actuellement utilisée par la Communauté Européenne.

La seconde approche donne actuellement lieu à une expérimentation menée en collaboration avec les géographes de l'équipe P.A.R.I.S (CNRS URA 1243) sur la dynamique des systèmes de villes. Il s'agit de construire une bibliothèque de fragments de modèles qui, à partir d'hypothèses sur les migrations entre les lieux d'habitation, permette de construire un modèle approprié pour simuler l'émergence des villes et des systèmes de villes.

## 5 Logiciels

### 5.1 Panorama

Les logiciels du projet Sherpa ont été historiquement développés sur plate-forme Unix, certains dans l'environnement de programmation Talk de la société Ilog, d'autres en langage C. La tendance actuelle est toutefois d'utiliser le langage JAVA pour les nouvelles implémentations (c'est le cas en particulier du système AROM), les plates-formes PC étant visées tout autant que les plates-formes Unix. L'autre tendance, illustrée par le système TROEPS et également par le système AROM, est de permettre l'utilisation des logiciels à travers tout navigateur Web.

## 5.2 AMIA

**Participant** : Michel Page [Correspondant].

Développé en collaboration avec l'université Pierre Mendès France de Grenoble, AMIA est un logiciel pour la modélisation et la simulation à base de connaissances. L'objectif visé à travers ce logiciel est double. Le premier est d'améliorer l'intelligibilité et la réutilisabilité des modèles numériques par l'apport et l'exploitation de connaissances de nature symbolique. Ces dernières sont utilisées pour :

- exprimer les propriétés et les relations non-numériques entre les composants du système modélisé ;
- formaliser les hypothèses sous-tendant les équations d'un modèle et rendre ainsi explicite le contexte de validité de ce modèle ;
- construire différentes variantes d'un modèle et formaliser le choix des variantes adaptées en fonction du contexte de simulation.

Ces connaissances, ainsi que les équations composant le modèle, sont exprimées dans un langage de modélisation algébrique, un formalisme d'écriture des équations proche des notations algébriques utilisées en mathématiques. AMIA dispose de mécanismes de simulation directe et inverse adaptés à ce type de langage. Actuellement, AMIA est utilisé dans des projets industriels, notamment par l'ADEME (Agence pour le Développement et la Maîtrise de l'Énergie), pour la prévision de la demande à long terme en énergie.

## 5.3 APIC

**Participant** : Gilles Bisson [Correspondant].

APIC est une interface graphique qui permet aux biologistes de visualiser les données génomiques sous la forme de cartes de différentes natures (génétique, physique...) et de naviguer entre ces cartes. Elle a été développée en C++ en utilisant la bibliothèque IlogViews. APIC est utilisée par IMAGtextscene afin d'afficher les résultats produits par plusieurs méthodes d'analyse et de les comparer.

## 5.4 AROM

**Participants** : Michel Page [Correspondant], Jérôme Gensel, Christophe Bruley.

Basé sur les compétences acquises par le projet Sherpa en représentation des connaissances par objets (TROEPS) et en simulation de systèmes dynamiques (AMIA), AROM a pour vocation d'être un système de gestion de bases de connaissances intégrant des techniques de modélisation à base d'objets et de relations et exploitant la puissance d'expression d'un langage algébrique. C'est d'ailleurs la signification de l'acronyme d'AROM: Associer Relations et Objets pour Modéliser.

Dans la plupart des systèmes de représentation de connaissances par objets, les liens entre objets sont décrits par le biais d'attributs qui référencent d'autres objets. Il existe cependant

deux cas importants dans lesquels ces systèmes n'offrent pas de solution satisfaisante pour la représentation des relations :

- lorsque les objets sont liés par des relations d'arité supérieure à 2,
- lorsque les relations comportent des attributs qui les caractérisent.

Dans AROM, un système de représentation des connaissances basé sur des objets et des relations est proposé. Les relations y sont considérées au même niveau que les classes. Elles peuvent être d'arité quelconque et comporter des attributs au même titre que les classes. Les relations sont, comme les classes, organisées en hiérarchies, matérialisant une spécialisation croissante.

Une première version du système AROM est disponible. Le système AROM offre des outils permettant aux concepteurs de construire des bases de connaissances et de réaliser des applications manipulant ces bases de connaissances. Totalement écrit en JAVA 1.2, le système est actuellement constitué des éléments suivants :

- un noyau, qui assure la gestion des différents éléments du modèle de représentation des connaissances. Ce noyau intègre également les mécanismes liés au langage algébrique et aux types. Le noyau offre une API JAVA qui permet, à partir d'une application, de créer, consulter, et modifier une base de connaissances AROM.
- un éditeur de bases de connaissances permettant la conception interactive de bases de connaissances au moyen d'un éditeur de graphes adoptant des notations graphiques voisines de celles d'UML (Unified Modeling Language).
- un module de consultation permettant l'accès à une base de connaissances et l'édition des instances à partir d'un navigateur Web. S'appuyant sur la technologie JSP (*Java Server Pages*), ce module génère dynamiquement des pages HTML qui permettent de parcourir le contenu d'une base de connaissances.

Les informations concernant le système AROM sont disponibles à l'URL suivant : <http://www.inrialpes.fr/sherpa/pub/arom>

## 5.5 FUSIN

**Participant** : Gilles Bisson [Correspondant].

Aujourd'hui, de nombreux outils permettent de diffuser des informations en connectant un gestionnaire de bases de données (Oracle, Access, Filemaker...) avec des générateurs de sites (FrontPage, HomePage...). Toutefois, l'interaction que l'on obtient reste souvent faible : l'utilisateur remplit un formulaire, le soumet et obtient une nouvelle page, voire une notification d'erreur si le formulaire est mal rempli. Il s'agit donc d'une communication où les protagonistes interviennent à tour de rôle. Via les possibilités offertes par HTML 4.0, les Applets et JavaScript, il est possible d'améliorer ce comportement, mais la conception du site devient alors complexe. Le système FUSIN qui est développé au sein du projet permet de créer facilement des sites Web autorisant la saisie et la recherche interactive et contextuelle d'informations. Ce système est particulièrement adapté à la gestion des sites de type « support-client ».

Dans FUSIN, le concepteur du site a juste besoin de décrire, à l'aide d'un langage déclaratif, les comportements qu'il souhaite obtenir en termes de navigation entre les pages, de vérification de cohérence ou de formatage des informations. Lors de la connexion au site, l'affichage dans le navigateur du client est entièrement contrôlé par une Applet JAVA qui construit dynamiquement le code HTML de la page courante ainsi que le code JavaScript permettant d'implémenter les comportements souhaités. Grâce à cette approche, il devient possible de guider efficacement les utilisateurs du site dans les opérations de saisie ou de recherche d'informations par l'intermédiaire de questionnaires dont le contenu change dynamiquement au fur et à mesure de l'introduction des données. Par ailleurs, comme FUSIN permet d'explicitier la sémantique des connaissances manipulées dans le site, la recherche d'information qu'il met en œuvre est basée sur la similarité de contenu et non plus sur la seule présence de mots-clefs comme dans le cas des moteurs de recherche classiques.

Le système FUSIN repose sur les technologies standard (HTML, XML, JAVA) et peut être facilement mis en œuvre dans des environnements divers. Les informations concernant ce système sont accessibles via <http://nangaparba.inrialpes.fr/index.html>.

## 5.6 GNA

**Participants :** Hidde de Jong [Correspondant], Michel Page.

Le logiciel Genetic Network Analyzer (GNA), qui est en cours de développement, a pour but d'analyser automatiquement des modèles de réseaux de régulation entre gènes. Les réseaux sont modélisés par une classe d'équations différentielles « par morceaux » qui est une famille de modèles souvent utilisés à cet effet. Les modèles sont qualitatifs, dans le sens où les valeurs numériques des variables d'état et des paramètres ne sont pas connues. Par contre, le signe des variables et des paramètres est connu, de même que certaines (in)égalités imposant un ordre entre eux. GNA trouve les points d'équilibre dans différentes régions de l'espace d'état et trace les trajectoires du système d'une région à l'autre. Les algorithmes reprennent des travaux de biomathématiciens sur la classe de modèles mentionnée ci-dessus et des idées provenant du domaine de raisonnement qualitatif en informatique. GNA est implémenté en JAVA.

## 5.7 POWERTASKS et IMAGENE

**Participants :** François Rechenmann, Danielle Ziébelin [Correspondant].

L'environnement POWERTASKS a été construit dans le cadre d'un transfert technologique auprès de la société Ilog. Cet outil générique de résolution de problèmes a été conçu autour d'un modèle d'objets et de tâches issu des expériences précédentes du projet. Ainsi, deux environnements précédents avaient permis l'élaboration d'une dizaine de systèmes dans des domaines aussi variés que l'analyse de données, le traitement du signal, l'interprétation de courbes en automatique, la résolution d'équations aux dérivées partielles, le diagnostic médical et l'analyse de séquences génomiques. Le système POWERTASKS dispose d'un moteur de résolution associant à chaque tâche les méthodes possibles et d'une interface graphique qui permet à l'utilisateur :

- de visualiser toutes les étapes d'une stratégie au cours de son exécution ;

- de spécifier les paramètres et les choix nécessaires à cette exécution ;
- de reprendre une exécution à n'importe quel endroit, ce qui autorise l'utilisateur à émettre, tester et valider plusieurs hypothèses en parallèle ;
- de modifier une stratégie en modifiant ou ajoutant des sous-tâches.

POWERTASKS a été diffusé dans plusieurs laboratoires de biologie, notamment l'Atelier de Bio-Informatique (ABI) de l'université Pierre et Marie Curie (Paris 6) et l'institut Pasteur de Paris. Ces deux équipes ont ainsi conçu le système IMAGtextscene, qui aide un biologiste lors de l'analyse de séquences génomiques. IMAGtextscene est en particulier destiné à l'annotation de génomes bactériens. Il permet de représenter et de gérer de manière uniforme les connaissances biologiques produites au cours de l'analyse d'un génome (gènes, signaux de régulation...), ainsi que les connaissances méthodologiques attachées à cette analyse, à savoir les méthodes d'analyse proprement dites et les conditions de leur mise en œuvre. En outre, IMAGtextscene dispose de l'interface APIC développée au sein du projet Sherpa. Elle permet d'afficher, sur le même écran, l'ensemble des résultats produits par une ou plusieurs méthodes (objets biologiques, courbes de prédiction, etc.) afin de les comparer.

Le système IMAGtextscene a été utilisé pour le projet de séquençage du génome de la bactérie *B. subtilis*. Il a en particulier permis d'expérimenter différentes stratégies de détection des erreurs de séquençage (erreurs de cadre de lecture) et a ainsi contribué à identifier les régions contenant de telles erreurs.

## 5.8 Protocole de CO4

**Participant** : Jérôme Euzenat [Correspondant].

La construction coopérative d'un serveur de connaissances permet à plusieurs utilisateurs de modifier une base de connaissances. L'architecture CO<sub>4</sub> est une infrastructure permettant d'éditer une base TROEPS (dite consensuelle) entre plusieurs intervenants. Elle comprend des serveurs de connaissance TROEPS ainsi qu'un protocole de soumission de connaissances à une base.

Le protocole de CO<sub>4</sub> a été implémenté (en C) sous la forme d'une bibliothèque. Celle-ci dispose d'outils de manipulation des éléments du protocole (soumissions à examiner, votes, etc.) accessibles depuis un client HTTP quelconque. Ce protocole peut être connecté à différents types d'applications et a été connecté à TROEPS de manière à pouvoir soumettre directement la connaissance contenue dans les bases. La version 1.0 du protocole de CO<sub>4</sub>, ainsi que sa documentation, sont disponibles en `ftp://ftp.inrialpes.fr/pub/sherpa/logiciels/co4/current.tar.gz`

## 5.9 TROEPS

**Participant** : Jérôme Euzenat [Correspondant].

TROEPS est un système de gestion de bases de connaissances dans lequel les connaissances sont principalement représentées sous forme d'objets. Parmi les points originaux, il intègre

une notion de points de vue permettant de définir plusieurs hiérarchies de classes sur le même ensemble d'objets. TROEPS sert de support à diverses applications et a constitué la base sur laquelle de nouvelles recherches sont effectuées. La version 1.3 de TROEPS est mise à disposition sur le réseau. Elle intègre des développements effectués ces dernières années :

- Un système de types extensible est intégré. C'est sur lui que s'appuient les opérations de vérification de types, de classification et de catégorisation (voir ci-dessous).
- Un éditeur de bases de connaissances au travers du protocole HTTP a été développé. En plus de permettre la modification des bases de connaissances à partir d'un client HTTP, il permet la protection par mot de passe de ces bases, ainsi qu'un paramétrage des serveurs (éditables/non éditables).
- Une interface vers le langage XML permet d'importer et d'exporter des bases ou fragments de bases de connaissances TROEPS ainsi que d'agir sur celles-ci (voir 6.2.2 « XML et objets »).
- Un lexique indexe les noms des entités de la base de connaissances et les relie aux définitions des termes correspondants. Il est structuré à l'aide de plusieurs relations telles que synonymie, antonymie ou hyperonymie.
- Un environnement de catégorisation (ou « classification conceptuelle ») permet de construire de manière automatique des hiérarchies de classes à partir des objets disponibles (voir 6.1.1 « Construction de taxonomies à partir d'instances »). Il fournit des descriptions synthétiques des classes obtenues.
- La connexion au protocole de CO<sub>4</sub> permet à plusieurs agents de construire des bases de connaissances consensuelles. Depuis l'interface HTTP de TROEPS il est possible de soumettre tout ou partie d'une entité à la base consensuelle et d'intégrer des éléments venant de l'extérieur.
- Enfin, un système de révision permet de guider l'utilisateur pour corriger les erreurs détectées et un premier outil d'argumentation permet de discuter les différentes révisions entre plusieurs utilisateurs.

## 6 Résultats nouveaux

### 6.1 L'utilisation de la classification conceptuelle pour modéliser

#### 6.1.1 Construction de taxonomies à partir d'instances

**Participants :** Petko Valtchev, Jérôme Euzenat [Correspondant].

D'une manière générale, les algorithmes de classification conceptuelle cherchent à établir des zones de forte densité d'individus dans l'espace de représentation des connaissances. Une approche classique consiste à définir une mesure de ressemblance entre les éléments individuels de cet espace. Dans le cas des objets, la conception d'une mesure de ressemblance se heurte à des difficultés liées à la structure complexe des objets.

Le problème nouvellement résolu de manière convaincante est la prise en compte de relations circulaires entre objets. Tout d'abord, celle-ci entraîne des difficultés dans le calcul de la similarité qui sont résolues par le calcul des solutions des systèmes d'équations définissant la similarité. On a montré que, malgré la présence des constructeurs de collection dans la structure des objets, ce système trouvait une solution unique que l'on pouvait obtenir par des algorithmes classiques.

Le second problème est celui de l'obtention de descriptions de classes qui reflètent cette circularité. Le découplage de la phase de regroupement des objets en classes et de la phase de description des classes proprement dites a permis d'obtenir des taxonomies de classes en co-définition ce qui est unique dans les algorithmes de construction de taxonomies.

### 6.1.2 Classification conceptuelle pour l'extraction de connaissances

**Participant** : Gilles Bisson [Correspondant].

Dans le domaine de l'extraction d'information à partir de corpus textuels, les approches basées sur la recherche d'indices superficiels (techniques issues de l'extraction de connaissances) ou à base de mots-clés (techniques issues de la recherche documentaire) n'obtiennent que des résultats limités dès lors que les connaissances à extraire apparaissent dans des contextes syntaxiques mal définis. Pour lever ces limitations, il est nécessaire d'effectuer une analyse plus poussée des fragments du corpus en se basant sur des ressources linguistiques de type thésaurus ou ontologies qui sont spécifiques au domaine étudié et donc généralement lourdes à acquérir. Dans ce travail, nous cherchons à construire automatiquement de telles ressources par l'utilisation de méthodes de classification conceptuelle. La classification est effectuée à partir d'un texte déjà découpé et étiqueté à l'aide d'un analyseur syntaxique sous la forme de triplets (<verbe><lien><nom>). Par exemple, on pourra avoir les triplets comme (<inhibe><COD><gène>) ou (<contrôle><SUJET><protéine>). Le but est d'apprendre les «schémas de sous-catégorisation» des verbes pour le domaine étudié, c'est-à-dire d'apprendre, pour chaque verbe, la liste des noms et des liens qui peuvent s'y appliquer. Les noms sont organisés sous la forme de hiérarchies conceptuelles.

L'originalité de cette recherche est double. D'une part, dans le domaine de la linguistique, les approches existantes ne cherchent généralement pas à construire des hiérarchies mais plus simplement des partitions de mots. D'autre part, les structures à apprendre ne sont pas des hiérarchies strictes mais plus généralement des DAGs du fait des phénomènes de polysémie. Dès lors les algorithmes de classification ascendants hiérarchiques (CAH) ne sont pas directement utilisables. Nous travaillons donc sur une famille de méthodes basée sur l'utilisation de plusieurs opérateurs d'agrégation de classes en parallèle. Comme dans toute recherche par gradient, il faut alors définir des critères de contrôle des opérateurs. Sur ce dernier point, nous travaillons avec un critère de sélection des individus basé sur le «degré d'inclusion», qui est de granularité plus fine que la «distance» telle qu'elle est utilisée en classification et en analyse de données.

Ce travail a conduit à la construction d'un prototype, le système Mo'K, qui est un atelier permettant d'implémenter et de comparer facilement différentes métriques entre les mots et différentes stratégies de classification multi-opérateurs. Ce travail est effectué en collaboration avec l'équipe Inférence et Apprentissage du laboratoire de recherche en informatique (LRI) de

l'université de Paris-sud.

## 6.2 La visualisation et la diffusion des bases

### 6.2.1 Recherche d'informations sur le Web

**Participants** : Gilles Bisson [Correspondant], Alain Uginet, Danielle Ziébelin.

Pour améliorer l'efficacité de la recherche d'informations dans les sites Web thématiques, il semble nécessaire de l'articuler autour d'une base de connaissances définissant un ensemble prédéfini de critères normalisés et typés. L'avantage d'une telle approche est double. D'une part, la normalisation des critères de recherche permet de guider efficacement l'utilisateur dans la formulation de sa requête par l'intermédiaire de questionnaires dont le contenu peut changer dynamiquement au fur et à mesure de la saisie des informations. D'autre part, en explicitant la sémantique des critères par le typage, il devient possible d'effectuer des recherches plus souples basées sur la similarité des informations et non plus leur stricte égalité.

Le système FUSIN (FUZZY Search withIN sight) satisfait ces différents objectifs en offrant des outils permettant de construire et d'exploiter, parallèlement aux documents existants sur un site Web, un modèle assurant une indexation bi-directionnelle entre ces documents et la base de connaissances. La recherche d'informations repose entièrement sur les mécanismes des mesures de similarité qui ont été mis au point au sein de l'activité de catégorisation du projet ([1]). Le champ d'application de FUSIN concerne principalement les sites devant disposer de mécanismes d'interrogation souples et s'adressant à des publics ayant des niveaux d'expertise très divers ; les sites de support produit/client en sont un bon exemple. Ce système est expérimenté dans le cadre du réseau EMG-NET qui regroupe un ensemble de médecins européens travaillant en électromyographie.

### 6.2.2 XML et objets

**Participant** : Jérôme Euzenat [Correspondant].

La connaissance formalisée n'est pas forcément destinée à être centralisée dans un système de gestion de bases de connaissances. Elle peut s'insérer au sein de documents (par exemple pour en exprimer le contenu) ou être échangée entre bases (par exemple pour la construction coopérative de bases de connaissances). À cette fin, l'utilisation du langage XML (« eXtensible Markup Language ») s'impose par sa nature même de métalangage. Nous avons donc considéré les différentes possibilités d'encodage d'un langage de représentation de connaissance par objets en XML et examiné les lacunes d'XML pour la représentation de connaissances. Ces réflexions ont débouché sur la conception de descriptions de types de documents (DTD) pour TROEPS ainsi que pour les logiques de description. La DTD de TROEPS a été utilisée dans le cadre du programme Génie.

Mais la description de la syntaxe ne permet qu'un échange déjà prévu entre deux langages. Nous avons donc été amenés à proposer la définition de « description de sémantique de documents » à lier aux DTD afin de permettre l'interopérabilité sans préjuger du langage utilisé.

### 6.2.3 Visualisation de l'information

**Participants :** Christophe Bruley, Philippe Genoud [Correspondant].

À la difficulté qu'il peut y avoir à représenter visuellement un grand nombre d'informations, une réponse consiste à déformer l'espace de représentation afin d'obtenir une représentation visuelle, dense, mais localement intelligible (vue « œil de poisson », arbres hyperboliques, mur perspectif). Les représentations graphiques tridimensionnelles s'inscrivent dans la droite ligne de ces techniques de déformation. La déformation dans ce cas n'est autre que celle induite par la projection perspective qui présente l'avantage d'être naturellement interprétée par le système perceptif humain. Bien que de très nombreuses expérimentations aient été effectuées dans des domaines variés, il n'existe que peu de travaux étudiant de manière générale les questions que soulève l'usage de la troisième dimension : quels types d'information peuvent être transcrits par les composantes de l'espace de représentation tridimensionnel, quelles constructions graphiques permettent de traduire la structure de l'espace d'information, quelles techniques d'interaction permettent de « naviguer » dans la représentation graphique de l'espace d'information et de manipuler efficacement les données représentées.

La thèse de Christophe Bruley a permis de caractériser les représentations graphiques de l'information et de mettre en évidence les différents éléments qui interviennent : l'espace de représentation (espace géométrique et composantes de la graphique associée à celui-ci), le point de vue adopté sur l'espace d'information, les constructions graphiques traduisant la catégorie à laquelle appartient le point de vue retenu, le cas échéant le mode de représentation, et enfin l'emploi ou non de techniques de déformation de la représentation. Cette approche, en explicitant les éléments contribuant à l'élaboration d'une représentation graphique, permet d'envisager de nouvelles représentations ou d'adapter des représentations existantes à des domaines d'application autres que ceux pour lesquels elles furent initialement conçues.

## 6.3 La résolution de problèmes et les modèle de tâches

### 6.3.1 Résolution de problèmes répartis

**Participants :** Thibaut Parmentier, Danielle Ziébelin [Correspondant].

Ce travail s'inscrit dans le cadre des environnements coopératifs distribués. Cette étude vise non seulement à mettre en place l'environnement réparti `POWERTASKS`, mais aussi à prendre en compte des contraintes temporelles et des disponibilités de ressources. Ainsi le modèle de tâches a été étendu afin d'intégrer les spécificités de résolutions parallèles ou concurrentes. De plus, l'algorithme de résolution a été également modifié en y intégrant des paramètres d'évaluation sous contraintes de ressources. À terme, cet algorithme permettra d'évaluer le coût et le gain apportés par une méthode envisagée dans un processus de résolution. Cette évaluation se fera à partir d'une analyse multicritères permettant de mesurer la performance de la méthode, basée sur des critères tels que la précision, le temps d'exécution, la place mémoire utilisée, la disponibilité des données nécessaires ou encore la nécessité d'interactions avec l'utilisateur.

### 6.3.2 Comparaison de graphes de tâches

**Participants** : Sueli Ferreira da Cunha, Danielle Ziébelin [Correspondant].

Ce travail s'intègre dans l'exploitation des connaissances multi-sites. Il vise le développement d'outils de comparaison de stratégies de résolution, décrits par différents concepteurs, mais portant sur un même type de problème et pilotant une même bibliothèque de méthodes. Afin de comparer ces différentes résolutions d'une même tâche, un modèle de représentation canonique des graphes de décomposition et de spécialisation de tâches a été défini en utilisant les structures de E-graphes (graphes orientés, sans circuit ni sommet isolé, possédant un seul sommet d'entrée et un seul sommet de sortie). Des opérations de concaténation et d'alternance ont été introduites ; les propriétés d'associativité et de distributivité de la concaténation par rapport à l'alternance ont été vérifiées. Ce travail qui s'est déroulé dans le cadre de la thèse de Sueli Ferrera, s'est concrétisé par l'implémentation d'un module de recherche de formes canoniques pour les différents modèles de tâches développés au sein du projet. Ce module est associé à une évaluation des dissemblances : ainsi, les différences des deux stratégies comparées sont identifiées et des modifications sont proposées à l'utilisateur afin de lui permettre une unification éventuelle.

### 6.3.3 Pilotage de bibliothèques de composants

**Participants** : Martine Maume, Danielle Ziébelin [Correspondant].

Historiquement, les tâches élémentaires issues des phases de décomposition et de spécialisation, permettent l'exécution de méthodes faisant généralement appel à des bibliothèques de programme. Avec le développement actuel d'Internet et l'émergence des architectures à base de composants logiciels (COM, JavaBeans, CORBA), nous sommes conduits à réévaluer notre modèle de représentation de connaissances afin d'exploiter ces nouvelles technologies. Ainsi nos travaux effectués dans le cadre d'un mémoire CNAM étendent la notion de tâches élémentaires pour le pilotage de bibliothèques de programmes réparties sur plusieurs sites à la notion de composants logiciels répartis et interopérables. Cette redéfinition des tâches élémentaires s'intègre aux travaux menés conjointement sur la spécification d'un nouveau modèle de tâches s'appuyant sur le système de représentation des connaissances AROM, l'objectif étant de permettre l'intégration efficace dans des environnements de résolution de problèmes de composants logiciels disponibles sur le réseau.

## 6.4 La construction et l'exploitation de modèles dynamiques

### 6.4.1 Simulation sous contraintes

**Participants** : Michel Page, Jérôme Gensel [Correspondant], Véronique Chabanis.

Le projet Sherpa conduit des travaux sur la simulation à base de connaissances et a implémenté des algorithmes de simulations directes et inverses qui agissent sur des modèles comportant des équations et des connaissances symboliques sur ces équations (conditions de validité des équations...). Ces algorithmes reposent sur des techniques alliant théorie des graphes et

satisfaction de contraintes. Ils ont été implantés dans le système AMIA. Les recherches en cours portent sur l'amélioration du module de satisfaction de contraintes sur intervalles qui gèrent les équations des modèles en tant que contraintes. Notamment, les règles de maintien de la consistance locale, qui visent à réduire la taille d'un problème de satisfaction de contraintes, doivent être optimisées. L'optimisation repose sur des tests de positionnement relatif des intervalles représentant les domaines des variables contraintes.

#### 6.4.2 Problèmes d'optimisation combinatoire sous contraintes ensemblistes

**Participants :** Vincent Thornary, Jérôme Gensel [Correspondant].

Pour des raisons de compacité d'expression, certains des modèles traités en simulation portent sur des variables ensemblistes. La simulation sous contraintes nécessite alors d'être capable, d'une part de représenter des variables ensemblistes, et, d'autre part, de proposer des techniques de résolution adaptées à ces variables ensemblistes. Le travail réalisé ici porte donc sur ces deux aspects. En premier lieu, une représentation *hybride* des domaines des variables ensemblistes a été proposée. Elle allie les avantages des deux représentations traditionnellement adoptées en programmation par contraintes ensemblistes : énumération et intervalle d'ensembles. La représentation hybride permet, sous contrôle de seuils, de renseigner le système de résolution sur les valeurs qui ne peuvent pas faire partie de la solution d'une contrainte (inconsistance locale), et donc d'accélérer la recherche de solutions en évitant les instanciations inconsistantes de variables. Concernant la résolution de problèmes de satisfaction de contraintes ensemblistes, les techniques classiques de résolution exhaustives proposées par la programmation par contraintes ne conviennent pas pour les modèles de grande taille. En revanche, les méthodes de recherche locale de solutions (recuit simulé, méthode « tabou ») donnent des résultats satisfaisants sur des problèmes d'optimisation combinatoire de grandes tailles. Le travail réalisé consiste à adapter deux méthodes de recherche locale de solutions (le recuit simulé et la méthode « tabou ») à la représentation hybride proposée. L'idée est d'exploiter l'information fournie par cette représentation dans la détermination de la fonction de voisinage sous-jacente à chacune de ces méthodes.

#### 6.4.3 Construction assistée de modèles par assemblage de fragments

**Participants :** Camille Parisel, Michel Page [Correspondant], Hidde de Jong.

La construction automatisée de modèles peut être envisagée comme un processus de sélection et d'assemblage de sous-modèles élémentaires prédéfinis appelés fragments. Ces fragments modélisent une fraction du domaine d'application (un phénomène, une équation, un composant physique...) et sont organisés en une bibliothèque. Les fragments reposent souvent sur des hypothèses à différents niveaux qui sont partiellement, voire totalement, contradictoires. L'objectif de la construction automatisée de modèles par assemblage de fragments est de déterminer dans la bibliothèque l'ensemble de fragments pertinents pour constituer un modèle qui soit complet, cohérent et le plus simple possible pour répondre à une question donnée. Les travaux existants sur la construction de modèles proposent une approche automatique de la construction de modèles : une fois la bibliothèque de fragments constituée, le modélisateur

fournit une description de son système et obtient, si la bibliothèque est correcte, un modèle de ce système. Dans la pratique, cette approche manque de flexibilité dès lors que la bibliothèque s'avère incomplète ou inadéquate pour une portion du modèle.

Nous proposons une approche semi-automatisée. Le modélisateur peut faire appel à la bibliothèque pour concevoir certaines parties du modèle et construire par lui-même d'autres parties sous le contrôle du système de composition qui assure la cohérence globale du modèle. Un formalisme de description de fragments de modèles et les spécifications du moteur de composition sont en cours d'élaboration.

En collaboration avec les géographes de l'équipe P.A.R.I.S (CNRS URA 1243), et avec le soutien du programme « Modélisation et simulation numérique » du CNRS, des expérimentations sont menées sur la dynamique des systèmes de villes. Il s'agit, à partir d'un semis initial de lieux d'habitation et de différentes hypothèses sur les migrations entre ces lieux d'habitation, de simuler l'émergence des villes et des systèmes de villes.

#### 6.4.4 Analyse qualitative et semi-quantitative des systèmes dynamiques complexes

**Participants :** Hidde de Jong [Correspondant], Michel Page.

L'analyse mathématique des systèmes complexes est souvent rendue difficile par le fait que la connaissance du système est incomplète. La modélisation des réseaux régulateurs géniques en biologie est un bon exemple. Même si l'influence d'un gène sur l'expression d'un autre peut être caractérisée comme positive ou négative, une précision de cette interaction est souvent difficile à donner. Lorsque les valeurs numériques des paramètres ou les formes exactes des fonctions ne sont pas connues, il est possible de recourir à des modèles qualitatifs ou semi-quantitatifs. En mathématiques et en informatique, des techniques ont été développées pour raisonner sur ce type de modèles.

Le projet Sherpa travaille sur le développement de méthodes pour l'analyse qualitative et semi-quantitative des systèmes dynamiques complexes, en particulier les réseaux d'interactions géniques. Une étude sur les différentes techniques susceptibles d'être mises en œuvre est en cours de réalisation. En outre, une méthode pour l'analyse qualitative d'une classe d'équations appropriée à la modélisation des systèmes régulateurs géniques est actuellement développée. La méthode trouve les points d'équilibre dans différentes régions de l'espace d'état et trace les trajectoires du système d'une région à l'autre. Un logiciel basé sur ces idées (GNA) est en cours de réalisation.

Plusieurs extensions de notre approche sont envisagées, notamment la construction automatique des modèles et leur validation au moyen des données d'expression. En coopération avec Ivayla Vatcheva, de l'Université de Twente (Pays-Bas), deux méthodes pour l'analyse comparative des systèmes dynamiques ont été réalisées (CEC\* et SQCA). Les méthodes sont capables de prédire les différences dans le comportement de deux systèmes, décrits par des modèles qualitatifs et semi-quantitatifs, en raison d'une perturbation des conditions initiales ou d'une modification structurelle. L'application de SQCA à l'analyse des systèmes régulateurs géniques est poursuivie dans le cadre du Programme d'Actions Intégrées franco-néerlandais VAN GOGH.

## 7 Contrats industriels (nationaux, européens et internationaux)

### 7.1 Génie

**Participants :** Jérôme Euzenat, Danielle Ziébelin [Correspondant], Christophe Bruley, Philippe Genoud.

Le projet Sherpa participe à la phase 2 de l'action de développement Génie (Génération Nouvelle d'Ingénierie), sur le thème 3 « Mémoire technique ». Il est proposé de développer le concept de « mémoire logicielle », vue comme l'association d'une base de composants logiciels et de connaissances formelles et textuelles. La mémoire logicielle est ici considérée comme un cas particulier de mémoire technique, destinée, au delà de son intérêt intrinsèque, à expérimenter des méthodes et des outils génériques et donc adaptables à d'autres situations. Le travail concernant ce projet est mené selon deux axes :

- Le premier axe concerne le problème spécifique de l'association d'une base de connaissances à une base de composants. Les travaux menés ici ont comme cadre applicatif le développement d'un système à base de connaissances pour assister un utilisateur aérodynamicien dans sa démarche de mise en place des scénarios pour l'exploitation de codes de calculs en aérodynamique. L'objectif est d'aider l'utilisateur à choisir un scénario dans un panel en fonction du problème qu'il doit traiter. D'autres types d'aide peuvent être envisagés : aide à la construction de scénarios à partir des briques de base que représentent les actions, aide à la paramétrisation d'un scénario. La modélisation des connaissances s'effectue à l'aide du système AROM, les mécanismes d'inférences exploitant le langage algébrique et la classification. Une passerelle entre le système AROM et le formalisme UML permet d'intégrer les modélisations effectuées au sein de Dassault Aviation à partir du logiciel Rational Rose.
- Le deuxième axe concerne l'intégration de connaissances modélisées et de connaissances textuelles pour la traçabilité. La liaison entre TROEPS et XTERM, logiciel d'extraction de terminologie à partir de textes développé par Dassault Aviation, a permis d'explorer l'intérêt de l'utilisation de connaissances terminologiques dans la traçabilité des objets vers les sources textuelles qui les décrivent. Ainsi, à partir des documents, il est possible de construire une terminologie puis d'alimenter une base TROEPS par des classes correspondant aux termes. Les classes sont communiquées à TROEPS à l'aide de leurs descriptions XML. Ce faisant, les éléments TROEPS sont liés aux termes qui ont permis de les définir et permettent de remonter aux documents de spécification. Enfin, il est aussi possible d'indexer dynamiquement une base TROEPS préexistante à l'aide d'une terminologie (assurant ainsi une « traçabilité à posteriori »)

### 7.2 PROSPER

**Participant :** Jérôme Euzenat [Correspondant].

Le projet Sherpa participe au programme PROSPER du CNRS dans le cadre du projet ARDECO « Extraction, modélisation et gestion des connaissances réutilisables de conception »

qui rassemble le projet Eiffel à Rocquencourt, le laboratoire LACO (Poitiers) et le LISA (Lyon), ainsi que les sociétés Dassault Systèmes et Renault. Les objectifs du projet sont d'expérimenter des modèles et des mécanismes susceptibles de participer à la réutilisation des connaissances en conception : objets, points de vue multiples, modèles de tâches, protocole de construction coopérative, etc.

Pour cela, les mécanismes de réutilisation en conception mécanique sont étudiés parallèlement sous l'aspect informatique (Dassault, LISA, Sherpa) et sous l'aspect psycho-ergonomique (Eiffel, LACO) en collaboration avec un utilisateur (Renault). Les deux points de vue seront ensuite articulés de manière à formuler les objets informatiques (modèles, plans, contraintes...) et circonstances méthodologiques (phases de conception, objets et opérateurs manipulés...) pour lesquels une aide peut être produite.

### 7.3 Xerox Research Centre Europe (XRCE)

**Participants :** Denys Proux, François Rechenmann [Correspondant].

Le projet Sherpa collabore avec le centre européen de recherche Xerox situé à Meylan à quelques kilomètres du site de Montbonnot. Les travaux de thèse de Denys Proux, financés par une convention CIFRE, portent sur l'extraction d'informations à partir de textes : le problème est de remplir une structure de données, qui joue le rôle d'un formulaire, en interprétant le contenu de textes rédigés en langage naturel. Ce formulaire peut être implémenté comme une classe d'un modèle de connaissances à objets : remplir un formulaire revient alors à créer et compléter une instance de cette classe à partir des informations extraites du texte.

L'extraction de données sur les interactions moléculaires impliquées dans le développement de l'embryon de la drosophile constitue le premier domaine d'expérimentation de ces techniques. Une étape préalable a consisté à développer un module capable d'identifier dans des textes (commentaires extraits de la base de données FlyBase, puis résumés d'articles extraits de la base documentaire Medline) les noms de gènes et à les « étiquetter » de façon appropriée. Les travaux en cours portent sur l'extraction d'informations proprement dite. Le texte est analysé à l'aide des outils linguistiques développés au centre de recherches de Xerox (« shallow parser »). Conjointement, l'exploitation d'une ontologie du domaine des interactions moléculaires permet de construire une représentation formelle, en termes de graphes conceptuels, du contenu d'une phrase dans la mesure où elle concerne l'existence ou la description d'une ou plusieurs interactions.

### 7.4 PSA Peugeot Citroën

**Participants :** Alexandre Faure, Gilles Bisson [Correspondant], François Rechenmann.

Ce travail fait l'objet d'une convention CIFRE entre la Direction Informatique de PSA Peugeot Citroën et l'INRIA Rhône-Alpes. Il s'agit d'effectuer une modélisation du processus dit de « retour d'expérience » dans un environnement industriel. Le « retour d'expérience » consiste en une analyse systématique des activités industrielles tant en conception qu'en production. Cette analyse permet théoriquement de maintenir et d'enrichir la « mémoire d'entreprise » qui constitue le référentiel sur lequel se basent les acteurs de l'entreprise pour résoudre leurs

problèmes ou pour communiquer leurs nouvelles connaissances. Ces travaux ont permis de concevoir un modèle générique de l'activité de « retour d'expérience » qui doit être maintenant instancié pour divers métiers de la conception automobile. Ce modèle sera utile pour plusieurs activités :

- La détection de boucles de retour d'expérience existantes. Ces boucles existantes, mais qui ne sont pas toujours formellement identifiées, pourront ainsi être intégrées dans le schéma organisationnel des référentiels métier.
- L'optimisation du processus de retour d'expérience. À l'aide des guides de rédaction d'expérience, des méthodes identifiées d'analyse et des stratégies de diffusion adaptées, la communication entre les équipes se trouvera simplifiée.
- La mise en place de nouvelles boucles de retour d'expérience là où elles font défaut. En se basant sur la structure de référentiel et les activités présentées, la mise en œuvre de nouvelles capitalisations interactives sera rendue plus simple et plus rapide.

## 8 Actions régionales, nationales et internationales

### 8.1 Actions nationales

#### 8.1.1 Action de recherche coopérative **ESCRIRE**

**Participant** : Jérôme Euzenat [correspondant].

Un intranet (ou, plus généralement, l'utilisation des technologies d'Internet) est l'opportunité pour les entreprises d'accéder et de partager la connaissance bien souvent difficilement accessible sous forme documentaire. Les documents numériques et numérisés peuvent être rendus accessibles de manière standard et transparente auprès de tous les utilisateurs concernés. L'ambition, à terme, est de réaliser de véritables serveurs de connaissances permettant la recherche et la manipulation des ressources de l'entreprise. Cependant, les limites de cette approche apparaissent rapidement : l'organisation des sites se révèle une tâche coûteuse et la recherche en texte intégral peu efficace. La recherche et l'interrogation d'un site en s'appuyant sur le contenu des documents sont une nécessité et les formalismes de représentation de connaissances sont de bons candidats pour représenter ce contenu. La représentation du contenu permettra de le manipuler pour faire de la recherche par analogie, par spécialisation, par similitude, etc. Mais il existe différents formalismes de représentation de connaissances et nul ne connaît exactement leurs qualités respectives. Le but de l'ARC Ecrire consiste donc à comparer trois types de représentations de connaissances (graphes conceptuels, représentations de connaissances par objets et logiques de descriptions) du point de vue de la représentation du contenu de documents et de sa manipulation. Pour cela, elle s'appuie sur les compétences dans chacune des représentations des projets Acacia, Sherpa et Orpailleur respectivement. L'objectif de l'action consiste à comparer les apports de chacun des types de représentation pour la représentation du contenu dans les serveurs de connaissances. La mise à l'épreuve de ces différents formalismes pour le traitement d'un jeu de documents (fourni par un partenaire industriel) nécessite de mener une réflexion méthodologique sur le passage des textes à leur

représentation formelle (de façon suffisamment indépendante des formalismes employés) en lien avec le type d'accès que l'on veut avoir sur ces documents. Cette représentation formelle sera définie conjointement et introduite dans un format XML (pour « eXtensible Markup Language »). Un ensemble de requêtes défini de manière coordonnée sera évaluée dans chacun des contextes. A l'issue de ce travail, les différents formalismes seront comparés entre eux (mais aussi à la recherche en texte intégral) selon un protocole prédéfini. Celui-ci devra apprécier des critères tant qualitatifs (expressivité des requêtes, accessibilité/lisibilité des informations, etc.) que quantitatifs (temps de réponse à une requête, taux de précision/rappel des réponses, etc.). Cette évaluation proposera une grille d'analyse des avantages et inconvénients d'un langage de représentation formel vis-à-vis de la recherche d'informations sur le Web.

### 8.1.2 Action « Gestion de modèles et de connaissances - Application à la ville » du programme CNRS « Modélisation et simulation numérique »

**Participants :** Camille Parisel, Michel Page [Correspondant].

Le projet Sherpa et les géographes de l'équipe P.A.R.I.S (CNRS URA 1243) collaborent dans le cadre du projet CNRS « Modélisation et simulation numérique » sur l'analyse de la dynamique des systèmes de villes. Ce travail a pour objectifs d'étudier, d'adapter et d'expérimenter l'approche par composants dans le domaine de la construction, de la configuration et de l'interconnexion de modèles urbains. Ce processus de construction exploite des connaissances, à la fois sur le système analysé (les systèmes de villes) et sur les composants disponibles.

Un premier ensemble de composants urbains a été constitué. L'assemblage manuel de ces composants a permis de constituer différents modèles ayant donné lieu à des simulations qui sont en cours de validation. L'organisation de ces composants sous forme d'une bibliothèque devrait permettre l'assemblage automatisé des composants en vue de la construction de modèles.

### 8.1.3 Action de recherche coopérative LIRE

**Participants :** Thibaut Parmentier, Danielle Ziébelin [Correspondant].

L'action de recherche coopérative LIRE (résolution de problèmes avec L'imitation de REsources) regroupe les projets Maia (LORIA, INRIA Lorraine), Orion (INRIA Sophia Antipolis) et Sherpa (INRIA Rhône-Alpes). Elle a pour objectif de fournir des méthodes permettant de déterminer la meilleure stratégie pour résoudre une tâche donnée en fonction de contraintes qui lui sont imposées. Le résultat escompté de ces travaux est la conception de systèmes robustes ayant des capacités d'auto-adaptation face à un environnement en constante évolution. Les futurs algorithmes permettront d'évaluer le coût et le gain qui résultent de l'utilisation de plusieurs méthodes de résolution de problèmes concurrentes dans un domaine d'application donné. Cette évaluation se fera à partir d'une analyse multicritères permettant de mesurer, d'une part la performance des méthodes, par exemple, à travers leur précision, leur temps d'exécution, la place mémoire utilisée, la disponibilité des données nécessaires ou la nécessité d'interactions avec l'utilisateur, et, d'autre part d'assurer un déroulement satisfaisant des résolutions en tenant compte aussi de la puissance de calcul des machines, de leur charge de travail

et de la disponibilité des compétences et des données.

Une comparaison des différents modèles de tâches dans les divers projets a été faite. Le projet Sherpa a contribué à la réflexion sur une version répartie de la plate-forme LAMA du projet Orion. Par ailleurs, un travail de synthèse sur les modèles de tâches pour le pilotage de code a été effectué en collaboration avec Sabine Moisan du projet Orion. Le projet Maia a contribué à la réflexion sur l'intégration d'un modèle de représentation de ressources au sein du modèle de tâches POWERTASKS. Le descriptif du projet LIRE peut être consulté à l'URL suivant : <http://www.inrialpes.fr/lire/>.

Dans le cadre de cette action, Danielle Ziébelin a également effectué un séjour (30 juin au 1er août) au Knowledge Modelling Group du Stanford Medical Informatics en vue de confronter l'approche PSM (Problem Solving Methods) du système Protégé 2000 au modèle de tâches développé par le projet Sherpa et de définir des collaborations ultérieures.

#### 8.1.4 OFTA

**Participant** : Danielle Ziébelin [Correspondant].

L'observatoire Français des Techniques Avancées (OFTA) a créé un groupe de travail sur le thème « Architecture de logiciels et réutilisation de composants ». L'objectif de ce groupe est d'estimer la situation actuelle en matière d'architectures logicielles et de proposer des recommandations concernant l'intégration et la réutilisation de composants. En effet, un phénomène majeur de ces dernières années semble bien être la généralisation des méthodes à objets qui se proposent de faciliter la réutilisation et l'assemblage de composants logiciels préfabriqués. Des bibliothèques de composants logiciels généraux (C++, Java Beans...) utilisés de manière générique dans des applications diverses sont maintenant disponibles. De plus, le développement d'Internet devrait accroître ce phénomène en internationalisant cette approche par composants logiciels réutilisables.

L'ambition de ce groupe de travail est de proposer des recommandations sur la modélisation formelle des architectures logicielles s'appuyant sur des composants réutilisables. Il vise également à proposer des conseils organisationnels sur les processus de développement des applications. Actuellement, le groupe de travail, après avoir réalisé une étude sur l'utilisation de composants dans des travaux de recherche universitaire et des applications industrielles, se propose de rédiger un ouvrage collectif de synthèse faisant un état de l'art du domaine.

#### 8.1.5 CEA-INRIA

**Participant** : Danielle Ziébelin [Correspondant].

Un groupe de travail formé des deux projets Sherpa et Orion (INRIA Sophia-Antipolis) et de trois services du CEA a été constitué sur le thème « Modèle de connaissances et logiciels scientifiques ». Ce groupe de travail se propose d'aborder la problématique des logiciels scientifiques selon différentes approches : génie logiciel, intelligence artificielle et gestion des connaissances. Au cours de cette première année, un bilan a permis de préciser les points forts de chacune. Les outils d'aide au pilotage de programmes scientifiques ont particulièrement été étudiés. Des techniques pour faciliter ou automatiser la réutilisation de programmes à partir

de modèles de représentation de composants et de raisonnements sont l'objet d'études menées dans ce cadre.

### 8.1.6 **Projet « Bioinformatique des interactions moléculaires et des réseaux régulateurs à l'échelle des génomes »**

**Participants** : François Rechenmann [Correspondant], Hidde de Jong, Jérôme Euzenat, Gilles Bisson.

Ce projet de recherche s'inscrit dans le programme « Génome » du CNRS et rassemble cinq équipes de recherche en biologie, biométrie et informatique, sous la direction de Bernard Jacq du laboratoire de génétique et physiologie du développement de Marseille. Il a pour thème la conception, la réalisation et l'utilisation d'outils pour l'acquisition de données sur les réseaux régulateurs, l'analyse de ces réseaux, la simulation de certains aspects de leur fonctionnement, et enfin l'étude de leur évolution. Il comporte ainsi quatre volets :

- I. acquisition assistée et semi-automatisée de données sur les interactions moléculaires, chez la drosophile puis chez la souris, à partir de documents textuels sous forme électronique ;
- II. représentation des connaissances concernant les interactions moléculaires et les réseaux régulateurs chez la drosophile et la souris ;
- III. analyse des structures des réseaux et simulation d'aspects spécifiques du fonctionnement de réseaux de régulations moléculaires ;
- IV. étude de l'évolution des réseaux régulateurs eucaryotes.

Le projet Sherpa contribue sur le premier volet à travers les travaux de Denys Proux sur l'extraction d'informations à partir de textes. Le deuxième volet est abordé par Jérôme Euzenat à travers l'extension de la base de connaissances KNIFE, développée avec TROEPS. Les travaux de recherche du troisième volet s'appuient sur les compétences du projet Sherpa en matière de modèles dynamiques, en particulier qualitatifs. Il s'agit en pratique de réaliser un atelier d'analyse et de simulation des réseaux d'interactions moléculaires, autour d'une interface graphique de visualisation et de manipulation de ces réseaux. Enfin, les mesures de similarité étudiées par Gilles Bisson seront mises en œuvre dans le quatrième volet.

## 8.2 **Actions européennes**

### 8.2.1 **Programme d'actions intégrées franco-néerlandais VAN GOGH**

**Participants** : Hidde de Jong [Correspondant], Michel Page, Camille Parisel, François Rechenmann.

Dans le cadre du programme d'actions intégrées franco-néerlandais VAN GOGH, le projet Sherpa a commencé une coopération scientifique avec le projet Plinius du Département d'Informatique à l'Université de Twente (Pays-Bas). La coopération entre Sherpa et Plinius a pour but, d'une part le développement de techniques de modélisation applicables à l'analyse

de données scientifiques et, d'autre part l'échange d'expériences obtenues en appliquant ces techniques dans les domaines de la biologie moléculaire et des sciences des matériaux. L'action est planifiée sur deux ans, du printemps 1999 au printemps 2001, et prévoit des visites entre les deux projets des chercheurs participants.

### 8.2.2 Projet européen EMG-NET

**Participants :** Danielle Ziébelin [Correspondant], Gilles Bisson.

Dans le cadre des programmes européens INCO-COPERNICUS, le projet Sherpa est coordinateur du projet EMG-NET. Ce projet implique un consortium de quinze partenaires de onze nationalités différentes. Il s'agit d'un réseau de recherche s'ouvrant largement sur les pays d'Europe Centrale et Orientale. Il a pour but de développer une plate-forme composée d'un ensemble de modules d'aide à la décision qui sont basés sur du raisonnement à l'aide de cas. Une maquette de cette plate-forme est en cours de réalisation. Elle contiendra notamment un serveur de cas, d'ores et déjà disponible. Ce serveur s'appuie principalement sur une base de 500 cas réalisés au laboratoire d'électromyographie du CHU Albert Michallon (Grenoble). Cet outil contiendra à terme les 2000 cas jugés intéressants qui ont déjà été collectés au travers du projet, et en permettra leur consultation par un navigateur Web. Une maquette de serveur de cas a été réalisée à l'aide de l'outil générique FUSIN développé au sein du projet Sherpa. Ce serveur permet la capitalisation du savoir-faire européen pour le traitement des maladies neuro-musculaires. Par ailleurs, une base de connaissances est élaborée dans le langage AROM à partir de la base de connaissances du système MYOSYS développé précédemment. Les différents modules d'aide à la décision communiqueront au travers de cette base de connaissances.

### 8.2.3 Collaboration franco-bulgare (programme INCO)

**Participants :** Irena Koprinska, Danielle Ziébelin [Correspondant].

Cette collaboration a permis la constitution d'une nouvelle base de connaissances utilisant les acquis d'une base de 500 cas qui a été constituée lors d'un précédent projet (ESTEEM). Une analyse des cas a permis de « découvrir » des relations entre concepts qui n'étaient pas explicitées jusqu'alors dans le système MYOSYS. En effet, la base de connaissances du système d'aide au diagnostic électro-myographique MYOSYS s'est constituée de manière empirique et sans prendre en compte une analyse discriminante des différentes situations basée sur une analyse des cas. Ceci entraîne des modifications fréquentes et peu productives. Une nouvelle base a été implémentée dans le langage de représentation de connaissances AROM. Elle prendra en compte les données complémentaires issues de l'analyse des cas déjà effectuée.

### 8.2.4 Coopération scientifique et technique lusitano-française : modélisation d'un serveur de cas médicaux

**Participants :** Danielle Ziébelin [Correspondant], Jérôme Gensel, Philippe Genoud.

Le projet Sherpa, le laboratoire d'EMG du CHU Albert Michallon à Grenoble, le laboratoire d'EMG Egas Moniz de Lisbonne et le Département d'Informatique de la Faculté des Sciences

et de la Technologie de l'Université Nouvelle de Lisbonne ont participé au projet européen AIM ESTEEM. Cette collaboration se poursuit dans le cadre d'un accord de coopération lusitano-française.

Le département d'informatique de l'université nouvelle de Lisbonne a mis au point un modèle qualitatif et quantitatif basé sur un modèle de contraintes et qui permet d'exprimer des dépendances causales afin de modéliser la propagation de l'influx nerveux. Il permet de prédire la propagation de l'affection sur la structure nerveuse, l'évolution et la sévérité de la maladie diagnostiquée. Par la présentation de tableaux synthétiques, ce système offre à l'utilisateur la possibilité de comparer aisément différents cas.

### 8.3 Actions internationales

#### 8.3.1 Université du Québec à Montréal

**Participants :** Petko Valtchev, Jérôme Euzenat [Correspondant].

Une collaboration avec l'équipe du professeur Robert Godin à l'université du Québec à Montréal, partiellement soutenue par le centre Jacques Cartier (Lyon), vise à comparer les méthodes et formalismes des deux projets (langages de programmation ou de représentation de connaissances) dans l'aide à la conception des graphes de spécialisation. Pour cela des recherches communes aux deux équipes (UQAM et INRIA) sont menées sur trois fronts :

- construction automatisée de graphes à partir des données ;
- réorganisation de graphes suite à l'ajout de nouvelles données ;
- étude de la bonne gestion des graphes par les programmeurs.

Cette collaboration a donné lieu au séjour d'une semaine de Petko Valtchev à Montréal préparant son séjour post-doctoral dans l'équipe de l'UQAM en liaison avec le professeur Guy Mineau de l'université Laval (Québec). Les travaux portent sur la liaison entre deux points forts des équipes : la construction de taxonomies à partir de dissimilarités (voir 6.1.1 « Construction de taxonomies à partir d'instances ») et à partir de treillis de Galois.

## 9 Diffusion de résultats

### 9.1 Animation de la Communauté scientifique

- François Rechenmann est membre de l'Editorial Board de la revue « Bioinformatics » et du comité de rédaction de la « Revue d'Intelligence Artificielle ». Il est membre du Comité de Coordination des Sciences du Vivant et membre du comité de pilotage de l'action IMPG : Informatique, Mathématiques et Physique pour la Génomique.
- Jérôme Euzenat est membre du comité de programme de l'édition 1999 du colloque « Practical applications of knowledge management » et du séminaire « Knowledge management and organizational memory » de l'IJCAI 1999, lecteur pour « IEEE Intelligent systems » (1999). Il a organisé un séminaire ouvert de l'action Ecrire à CPE Lyon sur le

thème «XML et objets» auquel ont participé une dizaine de personnes de la région lyonnaise. Il participe au club CRIN ITI «Outils et méthodes pour la mémoire d'entreprise» depuis sa création en juillet 1998.

- Petko Valtchev a été lecteur pour la revue «Computational intelligence».
- Gilles Bisson a été membre du comité scientifique de la conférence SFC 99 (Société Francophone de Classification) qui s'est déroulée les 15-17 septembre à Nancy.
- Danielle Ziébelin participe au groupe de travail de l'OFTA «Architecture de logiciels et réutilisation de composants».

## 9.2 Enseignement

Les enseignants-chercheurs membres du projet effectuent des cours sur les domaines concernés par le projet dans différentes filières des universités grenobloises. Ne sont donc cités ci-dessous que les cours dispensés par les chercheurs INRIA du projet au niveau 3<sup>e</sup> cycle.

- «La modélisation informatique des connaissances», filière «biomathématiques», maîtrise de biologie des populations et des écosystèmes, université Claude Bernard, Lyon, 14h, François Rechenmann
- «Systèmes à base de connaissances», DEA Informatique, Systèmes et Communication, université Joseph Fourier et INPG, Grenoble, 9h, François Rechenmann
- «Intelligence artificielle», DESS «Informatique double compétence», université Joseph Fourier, 30h cours, 40h TD, Jérôme Euzenat, Jérôme Gensel, Gilles Bisson
- «Sémantique des représentations de connaissance», DEA d'informatique, communication et système (ISC), université Joseph Fourier et INPG, Grenoble, 12h, Jérôme Euzenat
- «Problématique de l'apprentissage», 3e année ENTPE, Vaulx-en-Velin, 3h, Gilles Bisson
- «Techniques d'apprentissage symbolique et numérique», cycle C, CNAM Grenoble, 9h, Gilles Bisson

## 9.3 Participation à des colloques, séminaires, invitations

- Exposé de François Rechenmann «Object-based knowledge modeling: entities, relations and tasks» au 5e séminaire Algorithmique et biologie: «Bases de Connaissances - Formalisation des Données et des Démarches dans les Projets Génomes et Post-Génomes», Institut Pasteur, Paris, 22-24 mars
- Présentation par François Rechenmann des activités du projet Sherpa en bio-informatique, ESF Exploratory Workshop «Functional Genomics» à l'Institut Pasteur, Paris les 1<sup>er</sup> et 2 avril 1999

- Exposé de François Rechenmann « Classes, relations, tâches : la modélisation des connaissances génomiques » à la journée « Biométrie et génomes » à l'INRIA Rhône-Alpes le 20 mai
- Exposé de François Rechenmann « Imagene : un environnement intégré pour l'annotation et l'analyse de séquences génomiques », avec Alain Viari lors d'un séminaire LORIA, Nancy, 28 juin
- Exposé de François Rechenmann « Informatique et génome », avec Christian Gautier aux 10<sup>es</sup> Rencontres Régionales de la Recherche à Grenoble le 8 octobre
- Exposé de François Rechenmann « La modélisation informatique des connaissances génomiques » lors du séminaire du magistère d'informatique à l'université Joseph Fourier, Grenoble, 26 novembre
- Exposé de Jérôme Euzenat lors du séminaire « sport-étude » du magistère MIM de l'université Claude Bernard sur « Quelques problèmes de représentation de connaissance par objets », 9 mars 1999.
- Exposés de Petko Valtchev lors du séminaire de la cathédre « Systèmes Intelligents » de la Faculté de Mathématiques et Informatique de l'Université de Sofia (Bulgarie) le 4 janvier sur « Comparaison d'individus co-définis pour la construction de classes », lors du séminaire du département d'informatique de l'université du Québec à Montréal (Canada) le 26 février sur « Regroupement conceptuel pour l'aide à la construction de bases de connaissances à objets » et lors du séminaire du Lise Ceremade, université PARIS-IX Dauphine, le 18 février sur le même sujet.
- Hidde de Jong a présenté l'article « Semi-quantitative comparative analysis » par I. Vatcheva et lui-même lors de l'« International Joint Conference on Artificial Intelligence (IJCAI) » à Stockholm (Suède) du 31 juillet au 6 août 1999. Il a également participé au workshop « Qualitative and Model-Based Reasoning for Complex Systems and their Control », le 1<sup>er</sup> août 1999, lors de la même conférence.
- Hidde de Jong a effectué une visite à l'Université de Twente (Pays-Bas) dans le cadre du programme d'actions intégrées franco-néerlandais VAN GOGH (du 7 au 18 octobre 1999). Lors de cette visite, il a fait un exposé devant les bioinformaticiens et biomathématiciens du « Center for Biometrics » du CPRO-DLO (« Centre for Plant Breeding and Reproduction Research, Wageningen University & Research Centre »). Titre de l'exposé: « Computer-Supported Analysis of Genetic Regulation Systems: Simulation and Validation of Qualitative Models »
- Séminaire de Hidde de Jong à l'INRIA Sophia-Antipolis, lors de sa visite au projet Comore (les 17 et 18 juin 1999). Titre de l'exposé: « Computer-Supported Analysis of Gene Regulation Networks »
- Exposé de Hidde de Jong à l'Université Libre de Bruxelles (Belgique), lors de sa visite au Département de Biologie Moléculaire (le 21 juin 1999). Titre de l'exposé: « Modeling and Simulation with Incomplete Knowledge »

- Michel Page a présenté un exposé: « La construction automatisée de modèles » à l'Ecole Thématique « La modélisation de la ville 2 », Nantes, le 24 septembre 1999
- Exposé de Danielle Ziébelin et Philippe Genoud au « Knowledge Modelling Group » du « Stanford Medical Informatics » : « Task Models for Problem Solving Environments », Stanford University, Palo-Alto, USA, 16 Juillet 1999.
- Exposé de Danielle Ziébelin au groupe « Representation and Reasoning Program » du «Artificial Intelligence Center» au SRI «Task models for Problem Solving Environments», Stanford Research Institute, Palo-Alto, 4 août 1999.
- Exposé de Danielle Ziébelin au groupe de travail « Architectures de logiciels et réutilisation de composants » de l'OFTA: «Modèle de tâches pour le pilotage de bibliothèques de programmes», Paris, 9 septembre 1999.

## 10 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] G. BISSON, «Why and how to define a similarity measure for object-based representation systems», *in: Towards very large knowledge bases*, N. Mars (éditeur), IOS press, Amsterdam (NL), 1995, p. 236–246.
- [2] I. CRAMPÉ, J. EUZENAT, « Object knowledge base revision », *in: actes 13th ECAI, Brighton (UK)*, p. 3–7, août 1998.
- [3] H. DE JONG, F. VAN RAALTE, « Comparative envisionment construction: A method for the comparative analysis of dynamical systems », *Artificial Intelligence*, 115, 1999, p. 145–214.
- [4] R. DUCOURNAU, J. EUZENAT, G. MASINI, A. NAPOLI (éditeurs), *Langages et modèles à objets : état et perspectives de la recherche, collection "Didactique", 19*, INRIA, INRIA Rocquencourt (FR), 1998, <http://co4.inrialpes.fr/lmobook/>.
- [5] J. EUZENAT, C. CHEMLA, B. JACQ, « A knowledge base for D. melanogaster gene interactions involved in pattern formation », *in: Actes 5th international conference on intelligent systems for molecular biology (ISMB)*, p. 108–119, Halkidiki, Grèce, juin 1997.
- [6] J. GENSEL, « Integrating constraints in an object-based knowledge representation system », *in: Constraint processing*, M. Meyer (éditeur), *Lecture notes in computer science*, 923, p. 67–77, 1995.
- [7] C. MÉDIGUE, F. RECHENMANN, A. DANCHIN, A. VIARI, « Imagen: An Integrated Computer Environment For Sequence Annotation And Analysis », *Bioinformatics*, 15, 1999, p. 2–15.
- [8] P. VALTCHEV, J. EUZENAT, « Classification of concepts through products of concepts and abstract data types », *in: Ordinal and symbolic data analysis*, E. Diday, Y. Lechevalier, et O. Opitz (éditeurs), Springer Verlag, Heidelberg (DE), 1996, p. 3–12.
- [9] D. ZIÉBELIN, A. VILA, « Artificial Intelligence in Medecine », *in: Lecture notes in computer science*, E. Keravnou, C. Garbay, R. Baud, J. Wyatt (éditeurs), 1211, Springer, p. 343–348, 1997.

## Thèses et habilitations à diriger des recherches

- [10] C. BRULEY, *Analyse des représentations graphiques de l'information - extension aux représentations tridimensionnelles*, thèse d'informatique, université Joseph Fourier, Grenoble, juin 1999, <ftp://ftp.inrialpes.fr/pub/sherpa/theses/bruley.ps.gz>.
- [11] S. F. D. CUNHA, *Comparaison de résolutions de problèmes exprimées par un modèle de tâches*, thèse d'informatique, université Joseph Fourier, Grenoble, janvier 1999, <ftp://ftp.inrialpes.fr/pub/sherpa/theses/these-cunha.ps.gz>.
- [12] J. EUZENAT, *Représentations de connaissance : de l'approximation à la confrontation*, mémoire d'habilitation à diriger des recherches, université Joseph Fourier, Grenoble, janvier 1999, <ftp://ftp.inrialpes.fr/pub/sherpa/theses/hdr-euzenat.ps.gz>.
- [13] J.-M. GABRIEL, *Un modèle collaboratif pour le dialogue entre un système d'apprentissage empirique inductif et son utilisateur*, thèse en sciences cognitives, Institut National Polytechnique de Grenoble, Grenoble, juin 1999.
- [14] P. VALTCHEV, *Construction automatique de taxonomies dans une représentation de connaissances par objets*, thèse d'informatique, université Joseph Fourier, Grenoble, décembre 1999.

## Articles et chapitres de livre

- [15] H. DE JONG, N. MARS, P. VAN DER VET, « Computer-supported resolution of measurement conflicts: A case-study in materials science », *Foundations of Science* 4, 1999.
- [16] H. DE JONG, F. VAN RAALTE, « Comparative envisionment construction: A method for the comparative analysis of dynamical systems », *Artificial Intelligence* 115, 1999, p. 145–214.
- [17] J. EUZENAT, « La représentation de connaissance est-elle soluble dans le web ? », *Document numérique* 3, 3-4, 1999, p. 151–167.
- [18] C. MÉDIGUE, F. RECHENMANN, A. DANCHIN, A. VIARI, « Imagen: An Integrated Computer Environment For Sequence Annotation And Analysis », *Bioinformatics*, 1999.
- [19] F. RECHENMANN, « Au cœur de la génomique in silico : la modélisation informatique des données et des connaissances », *in : Développement et applications de la génomique - L'après-génome*, Académie des Sciences.
- [20] C. SANCHEZ, C. LACHAIZE, F. JANODY, B. BELLON, L. RÖDER, J. EUZENAT, F. RECHENMANN, B. JACQ, « Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database », *Nucleic Acids Research* 27, 1, 1999, p. 89–94.
- [21] P. VALTCHEV, « An algorithm for minimal insertion in a type lattice », *Computational intelligence* 15, 1, 1999, p. 63–78.

## Communications à des congrès, colloques, etc.

- [22] A. FAURE, G. BISSON, « Modeling the experience feedback loop to improve knowledge base reuse in industrial environment », *in : Actes de KAW 99, Twelfth Workshop on Knowledge Acquisition, Modeling and Management*, Banff (CA), 1999.

- [23] M. MAUME, P. GENOUD, D. ZIÉBELIN, «Pilotage de composants par l'intermédiaire d'une base de connaissances», *in: Objet 99*, Nantes, mai 1999.
- [24] M. PAGE, J. GENSEL, M. BOUDIS, «AMIA: an environment for knowledge-based discrete-time simulation», *in: Actes du AAAI 1999 Spring Symposium on Hybrid Systems and AI*, Stanford University, CA (USA), 1999, <ftp://ftp.inrialpes.fr/pub/sherpa/publications/page99a.ps.gz>.
- [25] M. PAGE, C. PARISEL, «Automated modeling in weakly-formalized domains», *in: IJCAI Workshop: Qualitative Reasoning for Complex Systems and their Control*, Stockholm, août 1999, <ftp://ftp.inrialpes.fr/pub/sherpa/publications/page99c.ps.gz>.
- [26] D. Z. THIBAUT PARMENTIER, «Distributed Problem Solving Environment Dedicated to the DNA Sequence Annotation», *in: Proceedings of the European Knowledge Acquisition Workshop (EKAW'99)*, D. F. et al. (éditeur), *Lecture Notes in Artificial Intelligence, LNA*, Springer-Verlag, mai 1999.
- [27] P. VALTCHEV, J. EUZENAT, «Une stratégie de construction de taxonomies dans les objets», *in: Actes 7èmes journées de la société française de classification*, p. 307–314, Nancy (FR), 1999.
- [28] P. VALTCHEV, «Building classes in object-based languages by automatic clustering», *in: actes 3rd International Symposium on Intelligent Data Analysis*, M. B. David Hand, Joost Kok (éditeur), *Lecture notes in computer science, 1662*, p. 303–316, Amsterdam (NL), 1999.
- [29] I. VATCHEVA, H. DE JONG, «Semi-quantitative comparative analysis», *in: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI-99*, Morgan Kaufmann.
- [30] I. VATCHEVA, H. DE JONG, «Semi-quantitative comparative analysis and its application», *in: Working Notes of the Thirteenth International Workshop on Qualitative Reasoning, QR-99*, AAAI Press, p. 231–239, Menlo Park, CA (USA), 1999.
- [31] D. ZIÉBELIN, «Distributed Problem Solving Environment», *in: Journées GRACQ: Modèles, Objets et Composants: Apports et convergence avec l'ingénierie des connaissances*, Paris, juin 1999.

## Rapports de recherche et publications internes

- [32] C. BRULEY, «UML/AROM Mapping User Guide, version 0.1», *rapport de recherche*, INRIA Rhône-Alpes, Grenoble, novembre 1999, 11p, [ftp://ftp.inrialpes.fr/pub/sherpa/logiciels/arom/userguide-UMLAROM-en-0\\_1.doc](ftp://ftp.inrialpes.fr/pub/sherpa/logiciels/arom/userguide-UMLAROM-en-0_1.doc).
- [33] P. SHERPA, «Manuel Utilisateur du Noyau AROM, version 0.1», *rapport de recherche*, INRIA Rhône-Alpes, Grenoble, juin 1999, 18p, [ftp://ftp.inrialpes.fr/pub/sherpa/logiciels/arom/userguide-kernel-0\\_1.doc](ftp://ftp.inrialpes.fr/pub/sherpa/logiciels/arom/userguide-kernel-0_1.doc).
- [34] P. SHERPA, «TROEPS 1.3 reference manual», *rapport de recherche*, INRIA Rhône-Alpes, Grenoble, décembre 1999, 150p, <http://co4.inrialpes.fr/docs/troeps-manual.html>.

**Divers**

- [35] J. EUZENAT, «Des arbres qui cachent des forêts : remarques sur l'organisation hiérarchique de la connaissance», Session posters 2e conférence du chapitre français de l'ISKO, Lyon, 1999.
- [36] A. UGINET, « Outil d'aide à la conception d'interfaces WWW et recherche d'informations par mesures de similarités », mémoire d'ingénieur CNAM, juin 1999.