

Institut Gaspard-Monge  
Laboratoire d'informatique  
UMR 8049

Université de Marne-la-Vallée  
E.S.I.E.E.  
C.N.R.S.

Rapport scientifique  
2001 — 2004  
(novembre 2004)



# Table des matières

<b>1</b>	<b>Présentation du laboratoire</b>	<b>7</b>
1.1	Politique scientifique . . . . .	8
1.2	Activités communes . . . . .	9
1.3	Formation doctorale . . . . .	9
1.4	Recrutements et perspectives de croissance . . . . .	10
1.5	Conseil du laboratoire . . . . .	12
1.6	Liste des membres permanents du laboratoire . . . . .	13
<b>2</b>	<b>Algorithmique</b>	<b>15</b>
2.1	Composition . . . . .	15
2.2	Thèmes de recherche . . . . .	16
2.3	Résultats . . . . .	18
2.3.1	Algorithmique du texte . . . . .	18
2.3.2	Automates, codage et dynamique symbolique . . . . .	22
2.3.3	Algorithmique génomique . . . . .	32
2.3.4	Programmation générique et réseaux . . . . .	40
2.4	Activités . . . . .	43
2.4.1	Contrats . . . . .	43
2.4.2	Diffusion . . . . .	45
2.4.3	Travaux éditoriaux et organisation de colloques . . . . .	48
2.4.4	Collaborations . . . . .	48
2.4.5	Visiteurs . . . . .	49
2.4.6	Activités doctorales . . . . .	49
2.4.7	Thèses et habilitations . . . . .	50
2.5	Responsabilités . . . . .	51
2.6	Références bibliographiques . . . . .	51
<b>3</b>	<b>Combinatoire algébrique et calcul symbolique</b>	<b>65</b>
3.1	Composition . . . . .	65
3.2	Thèmes de recherche . . . . .	66

---

3.3	Résultats . . . . .	67
3.3.1	Fonctions symétriques non commutatives, fonctions quasi-symétriques . . . . .	67
3.3.2	Algèbres de Hopf combinatoires . . . . .	68
3.3.3	Généralisations du monoïde plaxique et de Robinson-Schensted-Knuth . . . . .	69
3.3.4	Polynômes de Jack et de Macdonald . . . . .	69
3.3.5	Algèbres de Hecke affines . . . . .	70
3.3.6	Tableaux de rubans . . . . .	70
3.3.7	Théorie des invariants et information quantique . . . . .	70
3.3.8	Polynômes de Schubert et de Grothendieck . . . . .	71
3.3.9	Algèbres de Lie libres . . . . .	71
3.3.10	Théorie des automates . . . . .	71
3.3.11	Combinatoire classique . . . . .	71
3.3.12	Calculs dans le centre de l'algèbre du groupe symétrique . . . . .	72
3.3.13	Combinatoire énumérative . . . . .	73
3.3.14	Hyperdéterminants, hyperpfaffiens et intégrales multiples . . . . .	73
3.3.15	Applications diverses . . . . .	74
3.3.16	Logiciels . . . . .	74
3.4	Activités . . . . .	75
3.4.1	Contrats . . . . .	75
3.4.2	Diffusion . . . . .	75
3.4.3	Collaborations . . . . .	77
3.4.4	Activités doctorales . . . . .	77
3.4.5	Thèses et habilitations . . . . .	77
3.5	Références bibliographiques . . . . .	78
<b>4</b>	<b>Informatique linguistique</b>	<b>85</b>
4.1	Composition . . . . .	85
4.2	Thèmes de recherche . . . . .	86
4.3	Résultats . . . . .	87
4.3.1	Applications . . . . .	87
4.3.2	Bibliothèques de ressources linguistiques . . . . .	89
4.3.3	Extension des ressources . . . . .	91
4.4	Activités . . . . .	92
4.4.1	Contrats . . . . .	92
4.4.2	Diffusion . . . . .	92
4.4.3	Collaborations . . . . .	93
4.4.4	Activités doctorales . . . . .	93
4.4.5	Thèses et habilitations . . . . .	93
4.5	Références bibliographiques . . . . .	94

---

<b>5</b>	<b>Géométrie discrète et imagerie</b>	<b>105</b>
5.1	Composition . . . . .	105
5.2	Thèmes de recherche . . . . .	106
5.3	Résultats . . . . .	106
5.3.1	Topologie discrète . . . . .	106
5.3.2	Opérateurs topologiques et traitement d'images . . . . .	114
5.3.3	Morphologie mathématique et applications du traitement d'images . . . . .	123
5.3.4	Géométrie algorithmique et géométrie discrète . . . . .	129
5.3.5	Compression d'image . . . . .	132
5.4	Activités . . . . .	135
5.4.1	Formation doctorale . . . . .	135
5.4.2	Participation à la vie scientifique . . . . .	136
5.4.3	Coopérations . . . . .	138
5.4.4	Contrat . . . . .	138
5.5	Références bibliographiques . . . . .	138
<b>6</b>	<b>Signal et communications</b>	<b>145</b>
6.1	Composition . . . . .	145
6.2	Thèmes de recherche . . . . .	146
6.3	Résultats . . . . .	147
6.3.1	Communications numériques . . . . .	147
6.3.2	Séparation de sources . . . . .	152
6.3.3	Théorie de l'information . . . . .	156
6.3.4	Analyse en ondelettes 2D . . . . .	158
6.4	Activités . . . . .	161
6.4.1	Contrats . . . . .	161
6.4.2	Diffusion . . . . .	163
6.4.3	Collaborations . . . . .	164
6.4.4	Activités doctorales . . . . .	165
6.4.5	Thèses et habilitations . . . . .	165
6.4.6	Rayonnement . . . . .	166
6.5	Références bibliographiques . . . . .	166
	<b>Données complémentaires</b>	<b>175</b>
	Thèses et habilitations . . . . .	175
	Rapports internes . . . . .	179
	Séminaires . . . . .	182
	Moyens et environnement . . . . .	188



# Chapitre Premier

## Présentation du laboratoire

**L**e laboratoire d'informatique de l'Institut Gaspard-Monge (IGM) a été créé en 1992. Il développe des recherches en informatique fondamentale et ses applications. Il a pu devenir productif très rapidement et même essaimer vers d'autres laboratoires d'informatique grâce à des transferts de travaux et de chercheurs opérés notamment à partir de l'université Paris 7. Les thèmes de départ incluent l'informatique théorique et le traitement de la langue naturelle. À ces thèmes originaux se sont ajoutés l'imagerie, la géométrie discrète et, plus récemment, le traitement du signal. Le lien commun entre tous ces thèmes est l'utilisation de méthodes formalisées pour atteindre un objectif de description exacte des phénomènes.

Du point de vue administratif, le laboratoire a été créé en 1992 comme équipe d'accueil de doctorants par la Direction de la Recherche et des Études Doctorales. Il a ensuite été équipe postulante du CNRS en 1994, renouvelée en 1996, puis confirmée comme UPRES-A en 1998. Depuis 2002, le laboratoire a le statut d'UMR.

Le laboratoire est constitué de cinq équipes identifiées par des thèmes de recherche qui sont à la base des projets décrits plus loin. Ces équipes sont les suivantes :

- Algorithmique ;
- Combinatoire algébrique et calcul symbolique ;
- Informatique linguistique ;
- Géométrie discrète et imagerie ;
- Signal et communications.

À celles-ci s'ajoute l'équipe « Simulacres, images, sons et arts relais » qui fait actuellement partie du laboratoire mais qui doit le quitter à la fin du contrat quadriennal actuel (décembre 2005).

Les activités, projets, collaborations, et la production de chacune de ces équipes sont décrits dans les chapitres suivants du rapport.

## 1.1 Politique scientifique

Pendant la dernière période les thématiques de recherche du laboratoire se sont affirmées pour donner la configuration actuelle en équipes. Hormis l'accueil récent d'une équipe en traitement du signal, les thèmes des autres équipes étaient présents au début du contrat précédent et la politique scientifique a eu pour but de les renforcer, sans chercher à en ajouter de nouveaux.

L'orientation scientifique pour la nouvelle période confortera l'orientation retenue et se traduira par un élargissement de la surface des thèmes de recherche des équipes existantes. Le potentiel de croissance du laboratoire, dû en particulier à son pouvoir d'attraction, au déficit en enseignants-chercheurs en informatique à l'université et à la politique de recrutement à l'ESIEE, sera utilisé pour des recrutements de qualité capables de s'intégrer dans les équipes et de les enrichir. Une attention particulière sera portée aux chercheurs susceptibles d'impulser des recherches de nature appliquée.

À titre d'exemple, les postes universitaires en informatique qui sont destinés au laboratoire ont pour profil les thèmes de recherche des équipes. Un fléchage sur l'informatique linguistique a aussi été décidé pour aider au renforcement d'une équipe possédant peu de permanents.

Par ailleurs, la politique de campus favorisée par le Polytechnicum de Marne-la-Vallée encourage les discussions avec des établissements comme l'ENPC et pourrait être profitable au laboratoire. À terme on peut envisager un regroupement de la plupart des chercheurs du campus ayant une activité dans le domaine des sciences et technologies de l'information et de la communication.

La recherche du laboratoire comporte globalement deux grandes orientations dans lesquelles s'inscrivent les travaux de toutes les équipes. La première est celle de l'informatique théorique et de la combinatoire. Elle comprend l'activité ancienne sur la combinatoire des mots et le codage qui alimente des travaux algorithmiques, la combinatoire algébrique en liaison avec des questions de calcul formel spécialisé, et la géométrie discrète qui est utilisée en analyse d'images.

La seconde orientation porte sur le traitement symbolique et statistique de signaux. En partant des signaux élémentaires jusqu'à des éléments plus complexes, elle comprend l'algorithmique du texte avec ses aspects combinatoires, le traitement statistique du signal et ses liens avec le codage de source ou de canal, le traitement d'images, l'analyse algorithmique des séquences moléculaires et le traitement de la langue naturelle.

La volonté de combiner une recherche de nature fondamentale avec le développement de logiciels prototypes trouve une traduction dans les activités des équipes. L'algorithmique de texte est étroitement associée à des études sur le traitement des séquences biologiques moléculaires en amont de la bioinformatique (il n'y a pas d'exploitation massive des données). L'activité en algorithmique est aussi associée à des travaux sur la programmation générique. L'équipe de combinatoire algébrique est fortement impliquée dans l'écriture de bibliothèques spécialisées de calcul formel pour le logi-

ciel `mupad`. Le logiciel `unitex` inclut un vaste ensemble de connaissances précises sur le lexique et la syntaxe de plusieurs langues naturelles. Les recherches en géométrie discrète sont utilisées en traitement d'images. Enfin, le traitement du signal conduit au développement de différents algorithmes d'estimation en vue d'améliorer les performances des systèmes de communication et de résoudre efficacement des problèmes de séparation et de restauration de signaux.

## 1.2 Activités communes

La fédération des équipes se réalise dans plusieurs activités et moyens communs :

- le séminaire hebdomadaire du laboratoire (actuellement le mardi après-midi) qui est l'occasion d'accueillir des chercheurs extérieurs au laboratoire ;
- le service de pré-publications qui sert pour la diffusion rapide des résultats des chercheurs sous la forme de rapports de recherche ;
- la politique d'équipement informatique du laboratoire concertée et commune aux équipes, et qui s'appuie sur le réseau de l'université ;
- le serveur informatique de courrier `monge.univ-mlv.fr` qui a pour rôle principal la communication entre chercheurs et accueille le serveur Web du laboratoire : <http://igm.univ-mlv.fr/LabInfo/>.

## 1.3 Formation doctorale

Le laboratoire est une des équipes d'accueil principales du DÉA « Informatique Fondamentale et Applications ». Celui-ci constituera à partir de 2005 la deuxième année du master recherche « Informatique ».

Pendant la période considérée 25 membres du laboratoire ont obtenu leur doctorat, et 5 autres ont été habilités à diriger des recherches.

Le DÉA fait partie de l'École doctorale « Information, Communication, Modélisation, Simulation » (ICMS). Il accueille en moyenne une vingtaine d'étudiants chaque année. Ceux-ci proviennent, pour plus de la moitié, de formations extérieures au campus.

Les établissements co-habilités à délivrer le diplôme sont :

1. École Nationale des Ponts et Chaussées (ENPC) (correspondant : Renaud Kériveren) ;
2. École Supérieure d'Ingénieurs en Électrotechnique et Électronique (ESIEE) (correspondant : Gilles Bertrand) ;
3. Université de Marne-la-Vallée (Marie-Pierre Béal, directrice du DÉA).

Le DÉA est composé d'un tronc commun qui se décline ensuite en six filières. Ces filières sont les suivantes :

– *Images et cinéma*

Cette filière est orientée vers la synthèse d'images. Ses objectifs sont d'une part la création d'images réalistes de grande qualité, et d'autre part la réalisation d'images de synthèse en mouvement. La nouvelle orientation de cette filière en fait une formation unique en région parisienne.

– *Imagerie 3D et environnements virtuels*

Cette filière traite de l'ensemble des problèmes intervenant dans l'analyse et le traitement informatique des images réelles. Une attention particulière est portée aux modèles, aux algorithmes et aux architectures matérielles spécialisées.

– *Logiciels des réseaux*

Cette filière embrasse l'ensemble des aspects logiciels, et notamment les applications réparties, la transmission multi-média, le routage, la conception de moteurs de recherche ou la sécurité.

– *Traitement des génomes*

Cette filière forme des chercheurs pour le traitement informatique des génomes, en liaison avec des organismes de recherche des sciences de la vie. L'informatique, et en particulier l'algorithmique est indispensable à l'analyse des génomes complets qui sont en cours de séquençage. Notre équipe d'algorithmique est à la pointe de ces développements.

– *Langue naturelle et représentation des connaissances*

Cette filière forme des chercheurs dans le domaine de la documentation automatique, du traitement de corpus et dans les applications vers le multimédia. Les aspects linguistiques du Web, notamment pour les moteurs de recherche, sont considérés. Le laboratoire a une position en pointe dans ces domaines.

– *Automates et combinatoire*

Cette filière est à la poursuite de la formation dans le domaine fondamental qui a fait le succès de l'école française d'informatique théorique. Elle traite des algorithmes et systèmes formels, des automates, qui sont liés au traitement de la langue naturelle, des données textuelles, ainsi que le traitement du génome. La formation dans le calcul symbolique est orientée vers les applications dans des domaines variés comprenant le calcul scientifique.

Nous formons dans ces domaines des étudiants participant aux recherches conduites dans le domaine fondamental et aussi dans le domaine industriel.

## 1.4 Recrutements et perspectives de croissance

Le laboratoire compte environ cent chercheurs dont à peu près la moitié sont des membres permanents. Après une croissance initiale assez forte (le nombre de cher-

cheurs a doublé entre 1997 et 2000) les effectifs du laboratoire continuent d'augmenter régulièrement de façon mesurée.

Au cours de la période 2001-2004 le nombre de professeurs et directeurs de recherche est passé de 12 à 17 et celui de maîtres de conférences et chercheurs de 20 à 28. L'accueil de l'équipe « Géométrie discrète et imagerie » de l'ESIEE en 1999, du « Laboratoire d'automatique documentaire et linguistique » (LADL) en 2001 et de l'équipe « Signal et communications » de l'Université en 2004 a largement contribué à cette augmentation.

Le départ programmé de l'équipe « Simulacres, images, sons et arts relais » entraînera le départ d'un professeur, de trois maîtres de conférences et de dix doctorants.

La disparition de Maurice Gross nous a tous affecté. Il avait participé à la création de l'unité en 1992, participait régulièrement à nos activités et venait juste de nous rejoindre avec l'équipe qu'il animait à Jussieu.

Parmi d'autres départs on peut noter ceux d'Olivier Carton (MC), d'Anne Micheli (PRAG) et Laurent Marsan (ATER) qui ont respectivement été nommés professeur à l'université Paris 7, maître de conférences à l'université Paris 7 et maître de conférences à l'université de Versailles. Celui de Marie-France Sagot qui a obtenu un poste de chercheur à l'INRIA mais qui maintient des liens étroits avec l'équipe d'informatique génomique. Ou encore celui de Cyril Allauzen (AMN) qui a obtenu un emploi chez AT&T Research après y être entré en stage post-doctoral.

Les arrivées sur la période ont largement compensés les départs. Ils proviennent de recrutements d'enseignants dans les établissements et au CNRS, du rattachement de chercheurs, et du rattachement d'enseignants d'autres universités.

Plus précisément, les recrutements ont été :

- deux chargés de recherche CNRS :
  - Julien Clément en 2001 dans l'équipe « Algorithmique » ;
  - Yukiko Kenmochi en 2004 dans l'équipe « Géométrie discrète et imagerie » ;
- une ingénieure de recherche CNRS chargée du développement de logiciels scientifiques pour l'équipe « Combinatoire algébrique et calcul symbolique » : Teresa Gomez-Diaz en 2002 ;
- huit maîtres de conférences à l'université :
  - Venceslas Biri en 2004 dans l'équipe « Simulacres, images, sons et arts relais » ;
  - Julien Cervelle en 2002 dans l'équipe « Algorithmique » ;
  - Rémi Forax en 2003 dans l'équipe « Algorithmique » ;
  - Jean-Gabriel Luque en 2000 dans l'équipe « Combinatoire algébrique et calcul symbolique » ;
  - Christophe Morvan en 2003 dans l'équipe « Algorithmique » ;

- Cyril Nicaud en 2001 dans l'équipe « Algorithmique » ;
- un professeur assistant à l'ESIEE :
  - Lilian Buzer en 2004 dans l'équipe « Géométrie discrète et imagerie » ;
- deux professeurs associés à l'ESIEE :
  - Laurent Najman en 2003 dans l'équipe « Géométrie discrète et imagerie ».
  - Hugues Talbot en 2004 dans l'équipe « Géométrie discrète et imagerie » ;

Au cours de la période concernée, le laboratoire a également accueilli :

- l'équipe « Signal et communications » de l'université en 2004. Celle-ci était constituée de :
  - deux professeurs : Philippe Loubaton et Jean-Christophe Pesquet ;
  - deux maîtres de conférences : Antoine Chevreuil et Christophe Vignat ;
- un directeur de recherche CNRS : Michel Lassalle en 2002 dans l'équipe « Combinatoire algébrique et calcul symbolique » ;
- un chargé de recherche CNRS : Jean-Christophe Novelli en 2002 dans l'équipe « Combinatoire algébrique et calcul symbolique » ;
- un ingénieur d'études CNRS : Christian Leclère en 2001 dans l'équipe « Informatique linguistique »
- une maître de conférences dans l'équipe « Informatique linguistique » : Catherine Gallardo de l'université Paris 10 en 2004.

Notons également la nomination en tant que professeur à l'université de Marie-Pierre Béal en 2002 et de Gilles Roussel en 2004.

Le nombre de doctorants de l'ordre d'une quarantaine est en augmentation.

Enfin, l'équipe administrative et technique a reçu le renfort de :

- de Nathalie Bély en 2001, ingénieur d'études CNRS ;
- de Nelly Muntz-Berger en 2004, technicienne d'administration contractuelle.

## 1.5 Conseil du laboratoire

Le conseil du laboratoire est représentatif des équipes constituant le laboratoire au moment de son élection. L'équipe « Signal et communications » qui l'a rejoint depuis, n'est donc pas représentée à l'heure actuelle.

Le conseil du laboratoire est constitué de :

- Julien Allali (élu) ;
- Frédérique Bassino (élue) ;
- Gilles Bertrand (nommé) ;
- Julien Cervelle (élu) ;
- Maxime Crochemore, directeur (membre de droit) ;
- Jacques Désarménien, directeur de l'Institut Gaspard-Monge (membre de droit) ;

- Teresa Gomez-Diaz (élue) ;
- Éric Laporte (élu) ;
- Sylvain Michelin (nommé) ;
- Jean-Christophe Novelli (nommé) ;
- Gilles Roussel, directeur adjoint (élu) ;
- Jean-Yves Thibon (nommé) ;
- Marc Zipstein (élu).

## 1.6 Liste des membres permanents du laboratoire

Chercheurs et enseignants-chercheurs :

- Frédérique Bassino, maître de conférences ;
- Marie-Pierre Béal, professeur ;
- Nicolas Bedon, maître de conférences ;
- Jean Berstel, professeur ;
- Gilles Bertrand, professeur ESIEE ;
- Lilian Buzer, professeur assistant ESIEE ;
- Julien Cervelle, maître de conférences ;
- Antoine Chevreuil, maître de conférences ;
- Julien Clément, chargé de recherche CNRS ;
- Michel Couprie, professeur associé ESIEE ;
- Maxime Crochemore, professeur ;
- Jacques Désarménien, professeur ;
- Etienne Duris, maître de conférences ;
- Isabelle Fagnot, maître de conférences, Université Paris VII ;
- Rémi Forax, maître de conférences ;
- Catherine Gallardo, maître de conférences, Université Paris X ;
- Florent Hivert, maître de conférences ;
- Yukiko Kenmochi, chargée de recherche CNRS ;
- Tita Kyriacopoulou, PAST ;
- Éric Laporte, professeur ;
- Alain Lascoux, directeur de recherche CNRS ;
- Michel Lassalle, directeur de recherche CNRS ;
- Christian Leclère, ingénieur d'études CNRS ;
- Philippe Loubaton, professeur ;
- Jean-Gabriel Luque, maître de conférences ;
- Serge Midonnet, PAST ;
- Christophe Morvan, maître de conférences ;
- Laurent Najman, professeur associé ESIEE ;
- Cyril Nicaud, maître de conférences ;

- Jean-Christophe Novelli, chargé de recherche CNRS ;
- Dominique Perrin, professeur ;
- Laurent Perroton, professeur associé ESIEE ;
- Jean-Christophe Pesquet, professeur ;
- Pierre-André Picon, professeur ;
- Christiane Rayssiguier, professeur ;
- Giuseppina Rindone, maître de conférences ;
- Gilles Roussel, professeur ;
- Hugues Talbot, professeur associé ESIEE ;
- Jean-Yves Thibon, professeur ;
- Christophe Vignat, maître de conférences.

Personnels administratifs et techniques :

- Nathalie Bely ;
- Andrée Impérial ;
- Éric Llorens ;
- Nelly Muntz-berger ;
- Teresa Gomez-Diaz ;
- Patrice Hérault.

# 2

## Algorithmique

### 2.1 Composition

#### Membres

- Frédérique Bassino, maître de conférences
- Marie-Pierre Béal, professeur
- Nicolas Bedon, maître de conférences
- Jean Berstel, professeur, responsable
- Julien Cervelle, maître de conférences
- Julien Clément, chargé de recherche CNRS
- Maxime Crochemore, professeur
- Étienne Duris, maître de conférences
- Isabelle Fagnot, maître de conférences, Université Paris VII
- Rémi Forax, maître de conférences
- Serge Midonnet, PAST
- Christophe Morvan, maître de conférences
- Cyril Nicaud, maître de conférences
- Dominique Perrin, professeur
- Christiane Rayssiguier, professeur
- Giuseppina Rindone, maître de conférences
- Gilles Roussel, professeur

#### Membres associés

- Marie-France Sagot, DR INRIA, UMR 5558 Biométrie et Biologie Évolutive
- Dominique Revuz, maître de conférences
- Marc Zipstein, maître de conférences

**Doctorants, ATER et post-doctorants**

- Julien Allali, ATER
- Pirro Bracka, ATER
- Julien Carcenac, doctorant CIFRE
- Christophe Deleray, ATER
- Gabriele Fici, boursier, doctorant en co-tutelle
- Francesca Fiorenzi, post-doctorante depuis mars 2003, bourse CNR puis bourse Marie-Curie
- Gautier Loyauté, allocataire de recherche
- Edouardo Moreno, boursier, doctorant en co-tutelle
- Benoît Olivieri, allocataire de recherche
- Jérôme Petazzoni, allocataire de recherche
- Pierre Peterlongo, allocataire de recherche
- Chloé Rispal, ATER

**Anciens membres**

- Cyril Allauzen, doctorant jusqu'en janvier 2002
- Philippe Blayo, doctorant jusqu'en janvier 2003
- Olivier Carton, maître de conférences jusqu'en septembre 2002
- Vincent Le Maout, doctorant jusqu'en juillet 2003
- Laurent Marsan, doctorant jusqu'en avril 2002
- Marie-France Sagot, jusqu'en septembre 2001
- Marina Zelwer, doctorante jusqu'en 2003
- Michel Nguyen The, ATER jusqu'en 2003

## 2.2 Thèmes de recherche

L'équipe travaille dans les domaines principaux suivants :

1. Algorithmique du texte
2. Automates, codage et dynamique symbolique
3. Informatique génomique
4. Programmation générique

Les deux premiers domaines sont reliés entre eux par la nature des méthodes utilisées qui relèvent de la théorie des automates et de la combinatoire des mots. Ils forment ce que l'on peut appeler le traitement symbolique du signal. Les applications liées à ce domaine sont la compression de texte, le codage de canaux contraints et, plus généralement, les questions qui touchent à l'algorithmique des suites de symboles.

Les problèmes qui sont traités comportent une part d'algorithmique classique : description de nouveaux algorithmes, avec analyse asymptotique de leurs performances,

et amélioration d'algorithmes existants. Leur description utilise les modèles standards de machines, à savoir le modèle « Random Access Machine » (RAM) pour le calcul séquentiel, et le modèle « Parallel Random Access Machine » (PRAM) pour le calcul parallèle. Les problèmes comportent aussi une partie plus formelle proche de l'algèbre qui est relative à la description par automates de langages particuliers.

L'étude structurelle du codage de l'information est complétée par une étude asymptotique fine de la complexité des algorithmes, notamment en prenant comme référence les mots de Lyndon. Les travaux sur la localisation de motifs et la compression se sont appuyés sur des structures d'automates développées dans l'équipe et sur la notion de mots interdits dans les textes.

Les recherches de l'équipe d'informatique génomique portent de façon générale sur le traitement algorithmique et statistique des séquences moléculaires considérées comme suites de symboles. Les sujets traités portent sur :

- l'identification de motifs,
- la localisation de motifs connus, régularités,
- la comparaison de séquences,
- l'indexation de séquences.

Les questions abordées comprennent :

- le développement de nouveaux algorithmes et d'applications,
- l'analyse de leur complexité.

Les traitements des génomes ont permis la mise au point d'algorithmes d'alignement dans leurs aspects théoriques et pratiques, le développement du logiciel UTOPIA pour la prédiction de gènes chez les eucaryotes et du logiciel SMILE pour l'inférence de motifs structurés, la modélisation et la comparaison des structures secondaires d'ARN au moyen d'une structure de multi-graphe appelée RNA-MiGAL.

Le premier axe de recherche de l'équipe de programmation générique concerne le développement d'outils et d'algorithmes pour la programmation générique, c'est-à-dire facilitant la réutilisation de code sans modification. En particulier, nous nous intéressons aux multi-méthodes et à la programmation par aspects en vue de leur mise en œuvre autour du langage Java. Pour cela, deux approches sont développées : l'une purement applicative permettant de conserver l'environnement standard de programmation Java et l'autre par paramétrage d'une machine virtuelle. Une implémentation en Java d'une machine virtuelle Java est achevée. Son architecture par composants permet de modifier simplement son comportement. Ces changements peuvent s'opérer au démarrage mais également au cours de l'exécution grâce à une API réflexive étendue.

L'autre thème émergent dans l'équipe concerne le développement de logiciels pour les réseaux. Deux sujets sont actuellement étudiés : la cohérence des données réparties sur le Web dans le contexte des applications pair-à-pair et l'utilisation, pour le routage, du déplacement des éléments mobiles d'un réseau. Un algorithme d'ordonnancement et de routage pour la communication dans un réseau de robots épars a été implanté dans un réseau de robots Lego MindStorm. Un protocole pour le déploiement multicast de

fichiers de grande taille et à grande échelle a été développé.

La période écoulée a vu la parution de nombreux livres écrits par des membres de l'équipe, à la fois au niveau licence qu'au niveau maîtrise (recherche).

**Perspectives** L'étude structurelle des codages de l'information va être poursuivie. Elle va s'orienter vers des aspects plus algorithmiques, en renforçant la recherche de la programmation effective pour les algorithmes et les descriptions structurelles.

Les techniques de compression conservative porteront davantage sur les structures de données que sur les textes.

Concernant les thématiques en algorithmique génomique deux axes principaux peuvent être dégagés : une approche « multiple » de l'inférence de la structure secondaire des ARN et du filtrage des séquences biologiques pour leur alignement, et une recherche de structures de données adéquates afin de diminuer les complexités de leur utilisation dans les questions de comparaison de séquences.

La machine virtuelle Java va permettre une étude pratique simple du multi-polymorphisme, de la programmation par aspects, ou des éléments de vérification. Les problèmes de performance rencontrés amènent à étudier une recherche plus fondamentale autour de modèles de plaquage de composants.

Le contrôle des mouvements sera utilisé pour la simplification d'autres protocoles de communication dans les réseaux ad-hoc. Une architecture générique pour le développement rapide de serveurs performants utilisant une architecture par événements est à l'étude.

## 2.3 Résultats

### 2.3.1 Algorithmique du texte

Les activités en algorithmique du texte portent sur plusieurs sujets. Le premier, et le plus classique, concerne le problème de la localisation de motifs. Les résultats sont décrits en partie dans cette section pour ceux qui abordent des questions génériques ou des applications au traitement des séquences musicales. Pour l'autre partie des recherches sur ce sujet qui sont issues de questions de bioinformatique les résultats sont décrits dans la partie « Informatique génomique ». Le second sujet sur lequel nous avons continué à travailler est celui de la compression conservative de données.

L'algorithmique du texte est un domaine vivant pour lequel existent deux conférences internationales annuelles, *Combinatorial Pattern Matching* (CPM) depuis 1990 et *String Processing and Information Retrieval* (SPIRE) créée plus récemment. Les méthodes algorithmiques qui y sont développées sont spécifiques et reposent à la fois sur des propriétés combinatoires des mots et des structures de données adéquates. Les recherches en combinatoire et algorithmique au sein de l'équipe se renforcent donc l'une et l'autre.

Plusieurs publications de synthèse ont été réalisées pendant la période concernée. Tout d'abord, le livre *Algorithmique du texte* de M. Crochemore, C. Hancart et T. Lecroq [131] qui est une introduction aux méthodes et problèmes de base du domaine et qui est destiné à des étudiants avancés. Sa traduction en anglais est en cours. Ensuite le livre *Jewels of Stringology* de M. Crochemore et W. Rytter [133] qui est une version profondément modifiée d'un livre précédent devenu épuisé. D'autres synthèses ont paru comme chapitres d'ouvrages collectifs : la première de A. Apostolico et M. Crochemore [123] porte sur la recherche de motifs simples et structurés tout en abordant des thèmes classiques en bioinformatique ; la seconde, due à M. Crochemore et T. Lecroq [132], combine des méthodes simples pour la localisation de motifs et pour la compression de textes ; enfin la troisième de M. Crochemore [129] se concentre sur les structures de données pour la réalisation d'index complets et constitue un chapitre du livre *Applied Combinatorics on Words* édité par J. Berstel et D. Perrin et qui est mentionné par ailleurs.

**Localisation de motifs** La localisation d'un mot donné est un sujet qui fait partie du domaine d'expertise de l'équipe et sur lequel les résultats nouveaux se font rares. Néanmoins, suite à une évaluation fine de l'algorithme dû à Apostolico et Giancarlo que nous avons obtenue précédemment, nous avons pu décrire cette approche dans un cadre intégré (M. Crochemore, C. Hancart et T. Lecroq [36]).

Utilisant la modélisation par automate de méthodes de localisation de mot, M. Crochemore et V. Stefanov ont obtenu un procédé pour calculer de façon exacte le temps moyen d'apparition d'un mot dans un texte aléatoire.

C. Allauzen, M. Crochemore et M. Raffinot [61] ont continué d'explorer les possibilités offertes par la structure dite d'oracle des facteurs (ou oracle des suffixes) pour décrire un algorithme de localisation de mot à la « Boyer-Moore ». La structure permet une description simple sans que l'efficacité de la méthode en pâtisse. Le langage reconnu par un oracle de suffixes vient récemment d'être caractérisé par A. Mancheron et C. Moan (laboratoire LINA, 2004).

La localisation de motifs bidimensionnels dans des images discrétisées, qui a fait l'objet de recherches antérieures dans l'équipe, a été poursuivie en considérant la possibilité de rotation du motif. A. Amir, A. Butman, M. Crochemore, G. Landau et M. Schaps [62, 1] ont obtenu un algorithme qui prolonge et améliore des travaux récents de Fredriksson *et al.* ; il a un temps d'exécution de  $O(m^3n^2)$  et ils ont montré qu'il existe  $O(m^3)$  rotations différentes qu'il semble nécessaire de considérer pour faire la recherche.

**Structures pour l'indexation** En dehors de l'utilisation de la structure d'oracle des facteurs, les travaux sur les structures de données pour l'algorithmique du texte ont porté sur leur encombrement mémoire. Les principes généraux utilisés pour réduire l'espace qu'elles occupent reposent sur l'identification de sous-structures et sur un as-

pect plus technique de compaction introduit par McCreight en 1976 pour les arbres de suffixes. En fait, lorsque les structures sont vues comme des automates, la réduction revient à la minimisation des automates en plus de leur compaction. Cette vision est décrite par M. Crochemore [34] où il est montré en particulier que les deux opérations de compaction et minimisation, non seulement peuvent être appliquées simultanément, mais aussi commutent.

Nous avons aussi considéré l'implantation réelle de ces automates, et plus précisément du plus réduit d'entre eux, l'automate compact des suffixes. J. Holub et M. Crochemore [108] ont montré qu'il est possible d'en obtenir une représentation qui n'utilise que 4 ou 5 mots-mémoire par lettre du texte suivant les informations que l'on souhaite conserver pour l'utilisation ultérieure de l'automate. Ce résultat est équivalent à celui obtenu pour l'implantation d'un tableau de suffixes considéré à juste titre comme la structure la plus économique en espace mémoire. À titre de comparaison, les meilleures implantations d'arbres de suffixes, telles que celles décrites par S. Kurtz, utilisent couramment 7 à 8 mots-mémoire par lettre du texte.

Enfin, M. Crochemore, B. Melichar et Z. Troníček [102, 44] ont décrit des algorithmes de construction de l'automate des sous-mots de plusieurs textes et considéré le cas des langages rationnels. Ils ont en particulier rectifié un résultat de R. Baeza-Yates sur l'évaluation de son algorithme.

**Traitement de données musicales** Les travaux sur ce sujet sont l'objet d'une collaboration principalement avec T. Lecroq du laboratoire LIFAR de Rouen et de l'équipe de C. Iliopoulos de King's College London. De plus, ils bénéficient de l'expertise de deux musiciens, E. Cambouropoulos (Université de Thessalonique) et T. Crawford (City University, London). C'est un sujet prometteur car il touche aussi bien à l'analyse de la musique qu'à des questions de production automatique et de plagiat.

La recherche d'information et de régularités dans les données musicales fait appel à des algorithmes de localisation de motifs et à des structures d'indexation. De façon élémentaire, la musique est supposée être représentée comme une suite de symboles au moyen du codage MIDI. La musique polyphonique est, elle, représentée par plusieurs suites dont les symboles respectifs sont synchronisés. Les symboles sont en réalité des entiers (de 0 à 127) et on considère usuellement que la distance entre deux d'entre eux est leur différence (en valeur absolue). On peut alors considérer des motifs approchés, pour lesquels chaque distance entre des symboles alignés est bornée par une constante, en ajoutant éventuellement une borne supplémentaire sur la somme totale de toutes ces distances. On en déduit alors une notion d'occurrence d'un motif dans une séquence musicale selon ces deux critères ( *$\delta$ -matching* et  *$\gamma$ -matching*).

Divers algorithmes reposant sur les notions ci-dessus ont été obtenus par M. Crochemore, C. Iliopoulos et Y.J. Pinzon avec l'aide d'autres collègues. Après les premiers algorithmes décrits avec T. Lecroq [92], nous avons considéré la détection de répétitions avec E. Cambouropoulos et L. Mouchard [28], puis l'enchaînement de répétitions évolutives

[38]. D'un point de vue technique, nous avons évalué des heuristiques pour réduire les temps de calcul des algorithmes précédents avec T. Lecroq, W. Plandowski et W. Rytter [38], et utilisé des mots-mémoire pour réaliser une implantation efficace avec G. Navarro [95].

**Compression de données** Les travaux en compression de données ont porté sur la compression de texte, la compression d'automates et une étude combinatoire sous-jacente à un algorithme de compression populaire.

Les résultats obtenus sur la compression de texte suivent la méthode de compression par anti-dictionnaire (DCA) introduite par M. Crochemore, F. Mignosi, A. Restivo et S. Salemi en 1999. Cette méthode repose sur la détection de mots qui n'apparaissent pas dans le texte à compresser et qui sont minimaux sous cette condition au sens de l'ordre facteur : ce sont des anti-facteurs et leur ensemble constitue un anti-dictionnaire. Des travaux de nature plus combinatoire sur les mots interdits sont décrits par ailleurs.

M. Crochemore et G. Navarro [101] ont modifié la méthode en utilisant une notion d'anti-facteur approché. Celle-ci nécessite la mémorisation d'informations supplémentaires pendant le codage du texte pour en permettre une décompression exacte, mais le résultat global est que le taux de compression tend à être un peu plus faible qu'avec la compression originelle. La taille du texte compressé est de l'ordre de 30 à 50% de la taille du texte initial sur des données standards.

Une autre extension de la méthode de compression DCA a été considérée par M. Crochemore, C. Epifanio, R. Grossi et F. Mignosi [90] pour compresser des automates. La technique initiale a dû être légèrement étendue pour ce faire. Le codage s'apparente à une auto-compression de l'automate, et l'utilisation de l'automate (existence d'un mot dans le langage reconnu) se réalise directement sur la structure compressée sans perte de temps.

M. Crochemore, J. Désarménien et D. Perrin [35] ont exploré les propriétés combinatoires des permutations qui interviennent dans la transformation de Burrows et Wheeler (1994) qui est à la base du logiciel de compression `bzip`. Les permutations sont aussi voisines de celles qui apparaissent dans les tableaux de suffixes utilisés en indexation. Les propriétés mises en évidence et qui s'appuient sur un résultat de Gessel et Reutenauer (1993) ont en particulier comme conséquence l'existence d'algorithmes linéaires immédiats pour le calcul de la transformation et de son inverse ou celui des permutations intermédiaires. Certains de ces résultats font appel aux algorithmes découverts en 2003 pour le classement en temps linéaire des suffixes d'un mot (quand l'alphabet est un intervalle entier de longueur bornée).

**Analyse dynamique des arbres digitaux** Un arbre digital (souvent appelé « trie ») est une structure d'arbre permettant de stocker des chaînes de caractères de façon efficace et donc par exemple adaptée à la représentation de dictionnaires.

En collaboration avec Brigitte Vallée et Philippe Flajolet, Julien Clément a mené

à bien l'étude de deux généralisations [33]. La première concerne la structure d'arbre elle-même et considère qu'un nœud de l'arbre n'a pas un accès direct à chacun de ses fils.

La deuxième généralisation concerne le cadre de l'analyse en moyenne lui-même. En effet toute analyse en moyenne nécessite de définir un modèle probabiliste sur les données. Ici on est amené à considérer, pour produire les chaînes de caractères à stocker dans le trie, une source de caractères. Les sources prises en compte pour l'analyse des tries étaient essentiellement des sources dites « sans mémoire », où chaque caractère est produit indépendamment des précédents, et les sources Markoviennes qui, pour émettre un caractère, tiennent compte d'un nombre fini de caractères précédents. L'introduction de sources dynamiques issues de la physique statistique permet d'englober de telles sources ainsi que bien d'autres (comme la source en fraction continue).

### 2.3.2 Automates, codage et dynamique symbolique

#### Codes

**Groupes syntaxiques d'un code préfixe fini** Un nouveau résultat combinatoire a permis de résoudre une question posée depuis longtemps concernant les codes finis. Le résultat principal démontré dans [57] est le suivant : les groupes syntaxiques d'un code préfixe à  $k$  mots sont, soit cycliques, soit inclus dans un groupe symétrique de degré au plus  $k - 1$ . Il s'agit d'une amélioration du résultat obtenu par Schützenberger (en utilisant le théorème du point critique de Césari-Vincent et Duval) qui donne une borne  $k$  au lieu de  $k - 1$ . Une conséquence est le fait que les groupes syntaxiques d'un code préfixe à trois mots sont tous cycliques, ce qui était conjecturé depuis longtemps (sur cette conjecture, voir <http://www.liafa.jussieu.fr/~jep/Problemes/Schreier.html>).

#### Automates et dynamique symbolique

Après de nombreuses années de gestation, le livre « Infinite Words : automata, semigroups, logic and games » de D. Perrin et J.-É. Pin est finalement sorti [140]. C'est un volume qui a l'ambition de rester comme une référence des résultats de la théorie des automates sur les mots infinis créée par Büchi dans les années 60 et considérablement développée depuis, avec des applications dans les domaines de la vérification notamment. Le livre constitue aussi une présentation de l'école française de théorie des automates. En particulier l'utilisation de semigroupes finis, introduits par Schützenberger, en est un élément central. Les travaux plus récents portant sur les variétés de semigroupes, les liens avec la logique ou les jeux sont traités en détail. Le livre a pour but de servir de guide pour les jeunes chercheurs abordant ce domaine ainsi que de référence des travaux les plus pointus difficilement accessibles dans les publications éparpillées dans la littérature.

**Dynamique symbolique et codage** Ce domaine concerne l'étude des propriétés de suites bi-infinies de symboles acceptées par un graphe étiqueté, ou automate, où tous les états sont terminaux et initiaux (on parle alors d'un système sofique). Ces systèmes sont utilisés comme modèle de canaux contraints (suites de bits sur un disque magnétique ou optique par exemple) et beaucoup de systèmes de codage pour canaux contraints sont traduits en circuits permettant de passer efficacement d'un canal à l'autre.

Les invariants de conjugaison permettent de différencier les systèmes. Il existe beaucoup d'invariants, souvent algébriques, pour des systèmes soifiques particuliers, dits de type fini, mais peu pour les systèmes soifiques généraux. De nouveaux invariants combinatoires ont été obtenus pour les systèmes soifiques [70, 71] en étudiant des propriétés du monoïde syntaxique ou sofique. Ces invariants sont également des invariants pour l'équivalence faible des systèmes soifiques. Ces résultats ont été obtenus par Marie-Pierre Béal, Francesca Fiorenzi et Dominique Perrin.

L'étude des blocs interdits minimaux et de leur complexité a permis d'obtenir des invariants dynamiques combinatoires différents de celui donné par l'entropie, qui compte, elle, les blocs pouvant apparaître comme facteur [14]. Cet invariant dynamique a été étendu partiellement aux systèmes dynamiques de dimension 2 ou plus [13]. Un algorithme quadratique de calcul des mots interdits minimaux d'un système sofique, ainsi qu'un algorithme linéaire de calcul de ces mots pour un ensemble fini d'échantillons, sont décrits dans [12].

Un algorithme de codage pour les canaux contraints soifiques est présenté dans [5]. La méthode fournit aussi une nouvelle optimisation pour ce type de codage lorsque le canal est de type fini. L'élaboration d'un circuit de codage et de décodage est exponentiel, mais ces circuits fonctionnent ensuite à taux de transmission constant et le décodage a la propriété de ne pas propager les erreurs survenues lors du stockage.

Gabriele Fici a étudié les mots interdits minimaux dans le cadre de la préparation d'une thèse de doctorat, en co-tutelle franco-italienne. Il a obtenu, en collaboration avec des chercheurs de l'université de Palerme, un algorithme linéaire de reconstruction d'une séquence à l'aide d'un nombre fini d'échantillons de cette séquence satisfaisant une condition de compatibilité. Les précédents algorithmes linéaires de ce type supposaient connue la taille du plus long bloc interdit minimal du mot à reconstruire. Il a présenté ces travaux aux Journées Montoises de Liège en 2004.

Il a par ailleurs étudié les mots interdits minimaux périodiques, notion introduite par Bruce Moision et Paul Siegel. On considère les facteurs d'un mot pouvant apparaître à certaines positions modulo une période fixée  $T$ , du mot. On regarde les blocs interdits minimaux pour ces occurrences. Il existe un algorithme linéaire de construction d'un système sofique défini par une liste de mots interdits minimaux périodiques pour chaque occurrence modulo une période donnée. Cet algorithme est utilisé pour calculer des sous-systèmes d'un système contraint tels que les bits qui apparaissent à certaines positions régulières sont libres. Ainsi, le changement d'un ou plusieurs de ces bits (de 1 à 0 ou inversement) ne modifie pas le fait que la séquence satisfait toujours les

contraintes de spectre du canal. Ces bits libres peuvent alors être utilisés pour corriger des erreurs. Ce travail (en cours de soumission) a été réalisé par Marie-Pierre Béal, Maxime Crochemore et Gabriele Fici.

Une nouvelle notion d'équivalence faible entre systèmes dynamiques symboliques de type fini a été introduite dans [15].

**Topologie pour les automates cellulaires** La dynamique des automates cellulaires a été étudiée principalement dans le cadre des systèmes dynamiques en plongeant l'ensemble des configurations dans la topologie de Cantor. Dans cette topologie, l'un des automates cellulaires les plus simples, le décalage des configurations, est catégorisé chaotique. C'est pourquoi d'autres topologies ont été étudiées.

Le défaut le plus flagrant de la topologie de Cantor est d'accorder une importance trop grande aux cellules centrales. C'est pourquoi la topologie de Besicovitch, basée sur le taux de différences entre les configurations a été introduite et étudiée. Dans [77, 26], on caractérise les points fixes et périodiques d'un automate, vecteurs de stabilité. On a ainsi montré que tout automate possède soit un unique point fixe et aucun autre point périodique, et dans ce cas l'espace complet des configurations s'écrase vers ce point, soit un nombre non dénombrable de points périodiques. Ce résultat traduit ainsi une stabilité inhérente des automates cellulaires. Un second résultat fut la construction d'un ensemble de configurations qui soit à la fois dense et stable, dans la topologie de Besicovitch.

Enfin, un dernier résultat prouve qu'il n'existe pas d'automate transitif dans la topologie de Besicovitch, et donc qu'il n'y a pas non plus d'automates cellulaires expansifs, ergodiques ou fortement transitifs. Ce résultat permet de statuer que les automates les plus chaotiques, dans le cadre de la topologie de Besicovitch, sont les automates cellulaires sensibles aux conditions initiales (on peut en trouver plusieurs exemples dans la littérature).

**Topologie algorithmique** L'analyse précédente montre que dans une topologie qui n'accorde pas une valeur excessive à certaines positions, les automates cellulaires paraissent plus stables. Pour étayer cette intuition dans [88] une nouvelle topologie est introduite, basée sur la complexité algorithmique (ou complexité de Kolmogorov). L'étude du comportement des automates cellulaires dans ce nouvel espace a prouvé que les automates cellulaires ne peuvent pas modifier continûment l'information présente dans une configuration. Ou bien, ils en détruisent une portion non négligeable, ou bien, ils la conservent. Ils ne peuvent en aucun cas l'augmenter. Ainsi, le chaos observé dans le comportement d'un automate cellulaire ne peut provenir que de la configuration initiale.

**Automates de sable** Les tas de sables sont modélisés par des systèmes dynamiques dans [89]. Une topologie sur les tas de sables permet de décrire leur comportement, et

un modèle d'automates de sables permet de montrer qu'ils généralisent tout modèle de tas de sable basé sur des interactions locales. Les propriétés des automates de sables vis-à-vis de la surjectivité et de l'injectivité ont été étudiées. Les propriétés qui étaient vraies pour les automates cellulaires ne sont, pour la plupart, plus vraies pour les automates de sable. Par exemple, on peut trouver un automate de sable injectif qui n'est pas inversible. Enfin, la décidabilité de la nilpotence a été prouvée être indécidable. De plus, la preuve de ce résultat montre la capacité des automates de sable à simuler une machine de Turing sur une portion finie de la configuration (mais non limitée en hauteur).

**Théorie de l'information** Frédérique Bassino, Julien Clément et Cyril Nicaud s'intéressent à la généralisation du codage de Huffman à un ensemble infini de symboles ainsi qu'à des problèmes d'estimation d'erreurs au cours de la transmission de données. S'appuyant sur des travaux de Gallager, Van Voorhis et Mordecai Golin, ils ont étudié des distributions de probabilités décrites à l'aide de séries rationnelles. Les résultats obtenus portent sur des séries  $\mathbb{N}$ -rationnelles ayant un unique pôle. L'étude de telles distributions et de la description de leur codage optimal sont un premier pas vers des sources plus complexes du point de vue de l'analyse (lois distribuées comme une fonction zêta, ou la loi de Zipf à entropie infinie, par exemple).

### Langages et automates

D. Perrin a rédigé un article [139] sur les automates et les langages formels pour l'Encyclopédie *Storia della Scienza*.

**Automates sur les ordres linéaires** Les automates sur les ordinaux sont une généralisation des automates sur les mots infinis (de longueur  $\omega$ ) à des types d'ordres plus grands. Büchi les a introduits pour prouver la décidabilité de logiques sur les ordinaux dénombrables. Dans son travail de thèse, Nicolas Bedon a introduit des structures algébriques adaptées à la reconnaissabilité de tels langages. Ces structures algébriques permettent une présentation unifiée de la reconnaissabilité des langages de mots finis, de longueur  $\omega$  et d'ordinalité supérieure. Une adaptation du théorème des variétés d'Eilenberg sur les langages de mots finis aux langages de mots dont les supports sont des ordinaux dénombrables permet de classer les langages par les propriétés algébriques de leur algèbre syntaxique. En particulier, Nicolas Bedon a étendu le théorème de Schützenberger, McNaughton et Papert sur les mots finis aux cas des ordinaux de rangs finis [18] : les langages définis par des expressions régulières « sans étoile » sont exactement ceux définis par des formules de logique monadique du premier ordre, c'est-à-dire dont l'algèbre syntaxique ne contient pas de groupes non triviaux. Il a également étendu ce résultat [17] aux mots d'ordinalité dénombrable quelconque, et redonné une preuve simple du résultat de Büchi établissant l'équivalence entre les langages de mots

d'ordinalité dénombrable reconnaissables par automates et ceux définissables par des formules de logique monadique du second ordre de l'ordre linéaire. Chloé Rispal et Olivier Carton [121] viennent récemment d'étendre la reconnaissabilité algébrique aux langages de mots dont le support est un ordre linéaire dénombrable ne contenant pas de sous-ordre dense. Ils contiennent en particulier le cas des ordinaux dénombrables. Cette reconnaissance par structure algébrique peut être utilisée en particulier pour montrer la fermeture par complémentation des langages reconnaissables, résultat que Chloé Rispal et Olivier Carton avaient déjà obtenu par une autre méthode dans le cas des rangs finis [85, 30].

Chloé Rispal et Nicolas Bedon travaillent actuellement sur la définition de tels langages par des formules de logique du second ordre monadique, pour montrer que les reconnaissabilités par de telles formules et par les algèbres sont équivalentes. Ils ont obtenu un théorème des variétés, et vont adapter le théorème de Schützenberger, McNaughton et Papert pour les ordres linéaires dénombrables dispersés.

**Graphes infinis** Comme pour les familles de langages formels, il existe une hiérarchie de graphes infinis. En 2000, Christophe Morvan a défini la famille des graphes rationnels [115]. Cette famille est obtenue en codant les sommets d'un graphe par des mots et en définissant les arcs à l'aide d'une relation rationnelle (sur les mots). Cette famille constitue une extension de nombreuses familles précédemment étudiées, tels que les graphes des automates à piles de Muller et Schupp, les graphes équationnels de Courcelle.

Lorsqu'on définit un graphe, une question naturelle est de caractériser le langage défini par les étiquettes des chemins dans ce graphe (la trace d'un graphe). Ainsi une famille de graphes définit naturellement une famille de langages. Par exemple, il est bien connu que les traces des graphes finis sont les langages rationnels. Dans le cas des graphes infinis, de nombreuses familles caractérisent les langages algébriques. D'autres familles plus générales caractérisent de façon naturelle les langages récursivement énumérables. En 2001, Christophe Morvan avec Colin Stirling ont démontré que les graphes rationnels caractérisaient les langages contextuels [116]. Chloé Rispal a étendu ce résultat à la sous-famille des graphes rationnels synchronisés [120]. Cette famille est définie par des transducteurs lettre-à-lettre suivis de relations reconnaissables.

Plus récemment, Chloé Rispal et Christophe Morvan ont donné une présentation homogène à leurs travaux [56].

**Langages XML** XML est maintenant devenu un standard très répandu pour l'échange de documents. La description des documents suit des règles grammaticales strictes qui sont en fait des grammaires context-free généralisées d'un type particulier. Berstel et Boasson [73, 21] ont étudié les grammaires correspondantes, appelées grammaires XML, et les langages qu'elles engendrent. Les résultats principaux de cette étude sont deux descriptions structurelles des langages XML, et un résultat d'indécidabilité assez

surprenant : il est prouvé qu'il est décidable que l'ensemble de ses surfaces est fini pour un langage context-free contenu dans un langage de Dyck, mais indécidable s'il est régulier.

Cette étude des langages XML par des outils de grammaires context-free a été placée dans un contexte plus général dans l'article [126], où le lien est fait entre ces grammaires et des grammaires déjà étudiées il y a une quarantaine d'années, en particulier par McNaughton et Knuth. Là aussi, la différence première vient du fait que dans les articles anciens, les règles sont en nombre fini, alors que nous admettons des ensembles réguliers de règles. Ceci rend certaines preuves plus difficiles, mais donne aussi des résultats plus généraux, comme certaines propriétés de fermeture. Il serait intéressant de replacer cette étude dans le cadre des langages d'arbres engendrés par des « hedge grammars » qui sont très en vogue dans la description des documents XML.

**Décimation** Dans un travail présenté à FCT [74] et soumis à publication dans une version longue, Jean Berstel, Luc Boasson, Olivier Carton, Bruno Pettazzoni et Jean-Éric Pin prolongent et complètent l'étude des opérations de décimation des langages réguliers. Ils caractérisent les suites qui préservent tous les langages rationnels comme étant exactement les suites drup, c'est-à-dire différentiellement résiduellement ultimement périodiques. Ces suites sont en fait très naturelles, et apparaissent aussi dans d'autres problèmes. Il y a beaucoup de telles suites, ce qui montre des propriétés de clôture inattendues des langages réguliers.

**Minimisation d'automates** Jean Berstel et Olivier Carton prouvent dans l'article [75] que la borne en  $O(n \log n)$  de l'algorithme de Hopcroft peut être atteinte dans certains cas : le choix libre des classes qui servent à raffiner une partition peut être exploité pour ralentir l'exécution de l'algorithme. La construction utilise des mots de De Bruijn pour décrire les états terminaux des automates considérés.

**Dénombrement des automates** Beaucoup de travaux ont concerné les dénombrements de différents types d'automates, mais aucun résultat précis n'est connu concernant une des classes d'automates les plus importantes : les automates déterministes complets et accessibles. En se basant sur les travaux de thèse de Cyril Nicaud, où est établie une bijection entre ces automates et des objets combinatoires plus simples à manipuler, Frédérique Bassino et Cyril Nicaud ont obtenu un encadrement précis du nombre de tels automates avec  $n$  états. Ceci constitue une première étape nécessaire pour aborder l'analyse en moyenne d'algorithmes manipulant les automates déterministes.

**Séries génératrices de langages réguliers** Marie-Pierre Béal et Dominique Perrin ont caractérisé les séries génératrices des tailles des mots d'un langage régulier sur un

alphabet à  $k$  lettres [72, 16]. Cette caractérisation est calculable et on peut effectivement construire un langage régulier sur un alphabet à  $k$  lettres en partant de la série, même si la complexité est exponentielle. Avec Frédérique Bassino, ils ont obtenu une version régulière du théorème de Kraft-McMillan, qui caractérise les suites génératrices des arbres  $k$ -aires, en caractérisant les suites génératrices des arbres  $k$ -aires réguliers (c'est-à-dire qui ont un nombre fini de sous-arbres) [2]. Un article de synthèse sur les suites génératrices de langages réguliers a été publié [64].

D. Perrin a de plus rédigé un exposé de synthèse sur les problèmes de combinatoire énumérative des mots [138] qui reprend aussi en partie ces résultats.

**Algorithmes sur les transducteurs** Marie-Pierre Béal et Olivier Carton se sont intéressés à l'algorithmique sur les transducteurs. En particulier, ils ont décrit deux algorithmes permettant de déterminer des transducteurs sur les mots infinis. L'un prend en entrée un transducteur dont tous les états sont terminaux et produit en sortie un automate du même type mais qui est déterministe tout en réalisant les mêmes calculs [68, 9]. L'autre s'applique au cas plus général des transducteurs où une condition de répétition de type Büchi ou Muller est imposée aux états terminaux. Il calcule en sortie un transducteur déterministe du même type [10]. Ces travaux sur les transducteurs sur les mots infinis étendent ceux de Christian Choffrut sur la détermination des transducteurs de mots finis. Les conditions de détermination sont calculables.

Avec Christophe Prieur et Jacques Sakarovitch, M.-P. Béal et O. Carton ont donné des algorithmes de décision polynomiaux pour tester certaines propriétés de transducteurs sur des mots finis, comme la fonctionnalité et la séquentialité [69, 11].

Un algorithme de calcul de l'automate préfixe d'un automate étiqueté par des mots (éventuellement vides) a été obtenu en [8]. L'automate préfixe a le même graphe que l'automate initial mais les lettres sont poussées au maximum des états terminaux vers les états initiaux. Cet algorithme étend un algorithme de Mehryar Mohri valide pour les automates sans cycle d'étiquette vide. Le calcul de l'automate préfixe est une étape préliminaire dans le calcul d'un transducteur minimal de mots finis. Un des champs d'application pour la minimisation des transducteurs est le traitement des langues naturelles et le traitement de la parole.

Marie-Pierre Béal et Olivier Carton ont aussi décrit un processus de synchronisation de transducteurs réalisant une application à fenêtre glissante entre mots bi-infinis qui utilise des techniques de base d'éclatement d'états connues en dynamique symbolique [67, 7].

## Combinatoire des mots

Le grand événement, dans le domaine de la combinatoire des mots, est la parution du volume 2 et l'achèvement du volume 3 de la série de livres publiés sous le nom de plume de Lothaire. Le volume 2, « Algebraic Combinatorics on Words » [135], présente

en près de 500 pages des développements récents depuis la parution du livre « Combinatorics on Words » de 1983. Il a reçu un excellent accueil. Depuis, la rédaction du volume compagnon, intitulé « Applied Combinatorics on Words » [136], et qui regroupe des contributions sur les applications, en algorithmique, en linguistique, en biologie, en mathématique, de la combinatoire des mots, a progressé, et le manuscrit est chez l'éditeur. Sur près de 600 pages, les sujets évoqués sont décrits dans le style propre aux autres volumes de la série, c'est-à-dire de façon didactique, avec des développements complets, bien appropriés à l'apprentissage de base. Divers logiciels, et notamment un ensemble de programmes qui couvrent l'ensemble des algorithmes exposés dans le premier chapitre, écrit par Berstel et Perrin, sont disponibles à l'adresse <http://igm.univ-mlv.fr/~berstel/Lothaire/ProgrammesJava/BibliJava/>.

Jean Berstel et Juhani Karhumäki [24] présentent un « tutoriel » sur la combinatoire des mots qui résume certains résultats récents et dresse une liste de problèmes ouverts.

**Analyse en moyenne en combinatoire des mots** Depuis les recrutements en 2001 de Julien Clément et Cyril Nicaud à l'IGM, s'est mis en place un groupe de travail « Mots aléatoires ».

Du point de vue de la modélisation, la notion de « mot aléatoire » permet d'appréhender aussi bien les séquences biologiques, les développements en chiffres issus d'un système de numération ou encore les mots d'un code en théorie de l'information. Pourvu que le modèle aléatoire sur les données soit suffisamment réaliste, la complexité en moyenne fournit une bonne estimation du comportement réellement observé d'un algorithme. Ainsi la connaissance des propriétés combinatoires et probabilistes des mots permet de concevoir des algorithmes plus efficaces en moyenne, ou encore de mieux comprendre le comportement réel d'un algorithme d'après l'étude dans un modèle aléatoire.

Notre premier objet d'étude a été les mots de Lyndon [65, 66, 3]. En combinatoire des mots, ces objets sont centraux car ils permettent de décrire les propriétés de cyclicité. Ainsi de nombreux algorithmes de combinatoire des mots sont basés sur une décomposition en mots de Lyndon. Ils interviennent aussi de manière effective pour les calculs dans les algèbres de Lie affines en combinatoire algébrique. C'est également un ensemble de mots aux propriétés très particulières qui posent de nouvelles questions quant à leur analyse. Les résultats obtenus éclairent sous un jour nouveau les mots de Lyndon en proposant une étude probabiliste. De nouveaux résultats concernent également la factorisation standard (qui joue un rôle important pour des calculs dans les algèbres de Lie affines).

La venue pendant quatre mois de la doctorante sud-africaine Margaret Archibald (dans le cadre d'un programme doctoral) a permis de progresser dans l'étude de la structure d'arbre binaire de recherche avec clés répétées (en collaboration avec Julien Clément et Frédérique Bassino). Cette étude vise à obtenir une meilleure compréhension de l'algorithme de tri rapide (quicksort) lorsque les données à trier comprennent beau-

coup de valeurs répétées. Cet algorithme est présent dans la plupart des bibliothèques standards en programmation et est universellement employé.

**Mots sturmiens** Une façon de voir les mots Sturmien est sous forme de discrétisation d'une demi-droite de pente irrationnelle, c'est-à-dire une demi-droite d'équation  $y = \alpha x + \rho$ , avec  $0 < \alpha < 1$  irrationnel et  $0 \leq \rho < 1$ . On appelle mots Sturmien *caractéristiques* ceux qui correspondent à l'équation  $y = \alpha x + \alpha$ . Ces mots sont les seuls mots qui sont prolongeables à gauche de deux manières différentes avec une lettre de telle façon qu'ils restent Sturmien. J. Berstel and P. Séébold ont montré que si un mot Sturmien caractéristique est morphique, i.e. invariant par une substitution non triviale, alors, on peut lui ajouter une ou deux lettres à gauche de sorte qu'il reste Sturmien et morphique. Yasutomi a prouvé que ce sont les seules adjonctions possibles et qu'on ne peut pas non plus ôter les premières lettres d'un tel mot de façon à ce qu'il reste morphique. Isabelle Fagnot [48] a donné une autre preuve, élémentaire de ce résultat.

Jean Berstel a donné une formule compacte pour le nombre de représentations d'un entier en base de Fibonacci, et plus généralement pour les systèmes de numération d'Ostrowski [19]. Jean Berstel et Laurent Vuillon ont étudié les codages des rotations [25]. Jean Berstel présente dans [20] un exposé synthétique de quelques propriétés caractéristiques des mots de Sturm et donne, pour l'extension aux mots d'Arnoux–Rauzy ou épisturmien les énoncés correspondants, quand ils existent.

**Combinatoire du mélange** L'opération du mélange (« shuffle ») associe à deux mots un ensemble de mots, leur mélange. La question à laquelle répondent Jean Berstel et Luc Boasson dans [22] concerne la réciproque : étant donné un ensemble fini de mots, décider s'il est le mélange de plusieurs mots, et décrire les ensembles de mots dont il est le mélange. Les auteurs prouvent qu'il n'existe essentiellement qu'un seul ensemble de générateurs pour le mélange d'un ensemble fini. Le problème reste ouvert pour les langages réguliers.

**Combinatoire des mots dans les systèmes de type fini** Dans sa thèse en cours d'achèvement (sous la direction de Dominique Perrin, en co-tutelle avec l'université du Chili), Eduardo Moreno a étudié un domaine nouveau : la combinatoire des mots (mots de Lyndon, cycles de De Bruijn,...) dans les systèmes de type fini. Il s'agit de chercher quelles sont les propriétés des mots qui restent vraies lorsqu'on impose un ensemble fini des blocs interdits. Ce type de contrainte apparaît en particulier lorsqu'on travaille dans ce qu'on nomme des canaux contraints. Il a obtenu de nombreux résultats, tant de nature combinatoire, qu'algorithmique. Il a en particulier caractérisé les systèmes de type fini pour lesquels il existe en toute longueur des cycles de De Bruijn. Il a aussi obtenu une généralisation de l'algorithme de Fredericksen et Maiorana pour construire ces cycles (voir [113] et [114]).

**Ensembles inévitables** Une série de contributions diverses a permis de résoudre un problème extrémal en combinatoire des mots. Le résultat final, publié dans [32], donne une méthode pour trouver un système de représentants des classes de conjugués de mots de longueur donnée qui forme un ensemble inévitable. Ce résultat est le point d'arrivée d'une série de rebondissements. Tout d'abord, après son annonce dans divers séminaires, les auteurs ont appris qu'une autre méthode, totalement différente, avait été obtenue par J. Mykkeltveit en 1972. Ensuite, la question avait fait l'objet récemment de plusieurs communications. Notamment, C. Saker et P. Higgins (University of Essex) avaient les premiers soulevé la question en liaison avec une erreur trouvée dans l'ouvrage collectif *Combinatorics on Words* de 1983 et remontant à un article de Schützenberger de 1965. Des résultats partiels et des résultats expérimentaux de dénombrement avaient fait aussi l'objet de publications [54]. Le résultat final ouvre la voie à de nombreux développements possibles et, notamment celui de l'étude des ensembles inévitables minimaux tels que ceux qui ont été trouvés. Des essais par programmes ont été faits, en particulier par D. Knuth (voir son programme UNAVOINDABLE2 à l'adresse <http://www-cs-faculty.stanford.edu/>).

**Numérations non-standard** La représentation usuelle des nombres par leur développement en base entière  $q$  en utilisant les chiffres de l'ensemble  $\{0, \dots, q - 1\}$  a été généralisée dans plusieurs directions soit dans l'optique d'accélérer les calculs, soit dans une perspective de modélisation.

D'abord, l'ensemble des chiffres utilisés peut être modifié. Par exemple, les  $(q, d)$  systèmes de numération utilisent des chiffres de l'ensemble  $\{d, d + 1, \dots, q + d - 1\}$  où  $-q < d < 1$ , les représentations signées symétriques utilisent l'ensemble de chiffres  $\{-q/2, \dots, q/2\}$  où  $q$  est un entier pair. On obtient ainsi des représentations utilisant peu de chiffres non nuls, ce qui permet d'accélérer les calculs et peut être utilisé en cryptographie pour des calculs relatifs aux courbes elliptiques.

Frédérique Bassino et Helmut Prodinger ont étudié des propriétés statistiques et asymptotiques des  $(q, d)$  systèmes de numération [4]. Les résultats obtenus portent sur les moments et leurs comportements asymptotiques ainsi que sur la valeur en moyenne du minimum de  $n$  nombres tirés au hasard lorsqu'un nombre arbitraire de chiffres sont absents. Un article présentant des résultats analogues sur les représentations signées symétriques est en cours de rédaction. Ces recherches ont été menées dans le cadre d'un projet de coopération CNRS-NRF.

L'autre généralisation des systèmes de numération classique consiste à remplacer la base entière par un nombre réel  $\beta$  plus grand que 1. Dans ce cas, les chiffres utilisés sont les entiers positifs plus petits que  $\beta$ . Frédérique Bassino a caractérisé les propriétés des  $\beta$ -développements dans le cas où la base  $\beta$  est un nombre de Pisot cubique [63].

Les ensembles de Meyer ont été introduits sous le nom de « quasicristaux » dans le but de généraliser les structures cristallines. Ils interviennent dans certaines structures a périodiques rencontrées en physique également qualifiées de « quasicristaux ». Shigeki

Akiyama, Frédérique Bassino et Christiane Frougny [59] ont étudié les ensembles de Meyer associés aux représentations en base  $\beta$ . Ils ont donné un algorithme permettant la construction d'un ensemble minimal de translations permettant d'obtenir ces ensembles de Meyer à partir des représentations des  $\beta$ -entiers.

L'étude des systèmes de numération non standard fait appel à un large éventail de techniques comme la combinatoire des mots, la théorie des automates, la dynamique symbolique, la combinatoire analytique et la théorie des nombres.

### 2.3.3 Algorithmique génomique

Plusieurs chercheurs de l'équipe travaillent sur les questions de décidabilité, de complexité et de conception d'algorithmes pour le traitement des séquences biologiques moléculaires (ADN, ARN, protéines). Cette activité bénéficie de l'expertise de Christiane Rayssiguier, spécialiste de biologie moléculaire qui participe activement à la formation doctorale sur ces questions.

Des collaborations avec Jean-Loup Risler et Bernard Prum ont permis d'animer une partie de la communauté nationale du domaine et se sont concrétisées par plusieurs éditions dont : « Genome and Informatics » numéro spécial de la revue *Computers and Chemistry* [141].

Un article de synthèse sur les questions de localisation et d'extraction de motifs dans les séquences a été rédigé par Maxime Crochemore et Marie-France Sagot et est en cours de parution [134].

**Alignements** Un des problèmes de base dans le domaine de l'algorithmique pour la bioinformatique est la conception d'algorithmes de comparaison de séquences biologiques moléculaires par alignement. De très nombreuses applications en bioinformatique utilisent ces algorithmes comme procédé central. Il est donc essentiel de développer des méthodes efficaces en temps de calcul mais surtout en espace mémoire. La plupart d'entre elles fonctionnent en temps et espace quadratique relativement aux longueurs des séquences à comparer. Maxime Crochemore, Gad M. Landau et Michal Ziv-Ukelson [100, 43] ont conçu une méthode qui optimise la méthode vis-à-vis des deux critères simultanément sans aucune condition sur les coûts initiaux des opérations élémentaires de Levenshtein. L'algorithme a un temps de calcul qui est sous-quadratique. Plus précisément, celui-ci est  $O(hn^2/\log n)$  pour des séquences de longueur  $n$ ,  $h$  désignant leur entropie topologique ( $0 \leq h \leq 1$ ). Ce résultat est meilleur que le résultat précédent ( $O(n^2/\log n)$ ) dû à Fischer et Paterson (1974) qui ne fait pas intervenir l'entropie. De plus, l'algorithme admet en entrée des séquences compressées par l'algorithme LZ77 de Lempel et Ziv (1977). Le résultat provient de l'utilisation de la technique « diviser pour régner » appliquée à des sous-problèmes de tailles différentes provenant de la décomposition LZ77. La technique appliquée de façon équilibrée permet de retrouver le résultat de Fischer et Paterson.

Une recherche plus effective a été menée pour vérifier la possibilité de réduire le temps réel de calcul d'un alignement par utilisation de vecteurs binaires représentés par des mots-mémoires. Comme il est usuel de le faire, la vérification a commencé sur le problème du calcul de la plus longue sous-séquence commune à deux mots qui est une simplification de l'alignement. Maxime Crochemore, Costas S. Iliopoulos et Yoan J. Pinzon [97] ont montré que l'accélération est possible pour les algorithmes de Hirschberg (qui n'utilise qu'un espace mémoire linéaire pour travailler) et de Hunt-Szymanski (qui est à la base de la commande `diff` d'UNIX). La recherche a été étendue à la localisation de motifs à trous par Maxime Crochemore, Costas S. Iliopoulos, Yoan J. Pinzon et Wojciech Rytter [39].

**Systemes de recombinaison (splicing systems)** Les systèmes de recombinaison ont été introduit par T. Head en 1987; ils s'inspirent du mécanisme de recombinaison de l'ADN. Une variante de ce système a été développée par Paun. L'idée principale est, étant donné 2 mots (linéaires ou circulaires), de les couper et de les recoller entre eux pour obtenir un troisième mot. Les règles qui régissent ses opérations sont des quadruplets de facteurs qui doivent apparaître à l'endroit où on va couper et recoller.

Il était déjà connu que des langages non rationnels pouvaient être obtenus avec ce système en partant d'un ensemble initial fini (Siromoney *et al.*, 1992). Isabelle Fagnot a prouvé que l'on réussissait aussi à obtenir des langages non algébriques, et plus généralement, qu'à partir d'un langage dépendant du contexte et d'un ensemble fini de règles, on restait dans le domaine des langages dépendant du contexte. Par ailleurs, elle a trouvé une condition liée à la longueur des règles qui préserve le caractère algébrique d'un langage.

**Algorithmique sur les graphes** La détection d'ensembles de gènes proches dans plusieurs génomes permet d'identifier des gènes orthologues. Le calcul d'une partition maximale de  $n$  gènes situés sur  $m$  chromosomes linéaires et telle que sur chaque chromosome, les gènes de chaque partie sont espacés d'une distance au plus  $\delta$ , est possible en temps  $O(mn \log n \log \delta)$ . Pour  $\delta$  petit et dans le cas de deux chromosomes, on obtient ainsi un algorithme en  $O(n \log n)$  [6]. Cet algorithme, au delà des applications possibles dans le domaine de la génomique, présente une technique de partitionnement « à la Hopcroft » qui travaille sur plusieurs partitions en même temps, les pivots d'une partition cassant les parties des autres partitions [6]. Il s'agit d'une instance particulière d'un problème plus général d'algorithmique sur les graphes qui consiste à calculer les composantes connexes communes à deux (ou plusieurs) graphes qui partagent les mêmes sommets. La complexité est calculée de façon amortie en utilisant l'argument des parties pointées dû à A. Cardon et M. Crochemore.

**Bases de motifs à jokers** L'alignement d'un nombre quelconque de séquences est un problème NP-complet. L'importance des applications a engendré beaucoup de re-

cherches sur les heuristiques utilisables pour contourner la difficulté. Certaines d'entre elles se fondent sur une notion d'accroches ou de segments conservés à partir desquels un alignement multiple peut être développé. C'est autour de ce thème que Parida *et al.* (2000) ont introduit la notion de motifs maximaux et non-redondants qu'ils ont appliqués non seulement à l'alignement mais également à l'extraction de connaissances dans la fouille de données. Les motifs sont des mots qui sont écrits sur l'alphabet des séquences étendu avec un joker qui s'apparie avec tout autre symbole, et qui possède au moins deux occurrences. Les motifs maximaux et non-redondants forment une base à la manière de celle d'un espace vectoriel. Parida *et al.* ont aussi conjecturé que la base d'une séquence de longueur  $n$  ne contient pas plus de  $3n$  mots. Nous avons produit un contre-exemple de séquences possédant une base avec un nombre quadratique de motifs. Ce travail en collaboration entre Nadia Pisanti, Maxime Crochemore, Roberto Grossi et Marie-France Sagot [162, 119] a été l'occasion de revoir les définitions initiales. Il a permis d'introduire une notion de motif tuilant dont la base associée ne contient pas plus de  $n$  motifs, et de décrire un algorithme de construction qui s'exécute en temps  $O(n^2 \log n \log A)$  et dont l'optimalité est en discussion.

Maxime Crochemore, Costas S. Iliopoulos, Manal Mohamed et Marie-France Sagot [94] ont examiné la recherche de motifs (au sens précédent) particuliers. Ils ont conçu un algorithme qui localise des motifs de longueur maximale ne contenant qu'un seul bloc de jokers de taille fixe. La méthode n'est pas immédiate et utilise les arbres des suffixes de la séquence et sa renversée, ainsi qu'une préparation de l'un d'eux pour le calcul optimal de l'ancêtre commun le plus proche de deux nœuds donnés.

À la suite, Maxime Crochemore, Raffaele Giancarlo et Marie-France Sagot [91] ont étendu la méthode à d'autres types de motifs permettant de contrôler les trois segments des motifs au moyen de morphismes alphabétiques ou de relations spécifiques.

**Outil pour l'analyse comparative des graphes d'interactions entre protéines** Des techniques modernes de cribles double-hybride permettent de détecter des interactions directes entre protéines. L'hypothèse biologique commune est que si deux protéines partagent un grand nombre d'interactions communes, elles sont fonctionnellement reliées. Par ailleurs, Il est possible de repérer des liens de similarité (ou orthologie) entre protéines de deux espèces différentes. Comme les protéines qui interagissent entre elles ont tendance à être présentes ou absentes de manière groupée dans un génome donné, la recherche des sous-graphes d'interactions entre protéines qui sont conservés via les liens d'orthologie d'une espèce vers une autre nous permettent d'obtenir des informations phylogénétiques.

Isabelle Fagnot, Gaëlle Lalandais et Stéphane Vialette [105] ont utilisé un formalisme à base de graphe pour trouver si un complexe de protéines donné est conservé. Ce problème s'est révélé être polynomial s'il y a au plus deux orthologues par protéine et NP-complet pour trois. Néanmoins, trois traductions vers d'autres problèmes connus sont proposés dans le but d'obtenir des solutions effectives raisonnables. Ils proposent

également trois extensions naturelles de ce problème.

**Prédiction de gènes chez les eucaryotes** Philippe Blayo [147] a réalisé sa thèse sous la direction de Maxime Crochemore et de Marie France Sagot sur ce sujet. Le problème de la prédiction de gènes d'organismes eucaryotes, gènes composés de parties introniques qui seront épissées et de parties exoniques qui seront assemblées en un seul morceau et traduites en protéines, a été approché de façons diverses par les différents laboratoires qui se sont intéressés au problème. La recherche dans ce domaine a débuté par une reconnaissance des gènes par contenu uniquement puis s'est ajoutée une reconnaissance par signal (sites d'épissage donneurs et accepteurs, sites de branchement). Puis des approches de plus en plus intégrées sont apparues avec assemblage des exons et comparaison de ces assemblages à des séquences déjà connues déposées dans les banques de données. Mais un certain nombre de séquences correspondant à des ORF (Open Reading Frame) demeurent « orphelines » *i.e.* aucune séquence qui leur serait homologue n'a été repérée dans les banques.

Ph. Blayo a développé un algorithme qui était le premier à comparer deux séquences d'ADN contenant à la fois des séquences codantes (exons) et non-codantes (introns et séquences intergéniques) et qui utilisait une approche d'assemblage des exons codants (Gene Prediction as Exon Assembly). Dans cette approche, les régions codantes sont considérées comme non-indépendantes les unes des autres et donc, un gène, comme une suite d'exons ordonnés composant les pièces d'un assemblage. La technique utilisée est la programmation dynamique utilisant un alphabet étendu des codons dont l'ATG initiateur et également les signaux d'épissage AG et GT. Le modèle du gène utilisé ici a l'avantage de tenir compte d'un plus grand nombre de contraintes biologiques (notamment la compatibilité de phase de lecture) que les modèles qui considèrent les gènes comme suite d'exons indépendants. L'analyse algorithmique réalisée est poussée puisqu'une version linéaire (et non quadratique) en espace et quadratique en temps est proposée.

Un programme, « Utopia » a été implémenté basé sur le travail algorithmique. Il constitue un outil pour réaliser une prédiction de gènes présents dans deux séquences génomiques. Ce programme présente l'avantage d'être générique au sens où aucune supposition n'est faite sur les organismes auxquels appartiennent ces séquences. D'autre part, l'approche plus générique permet de traiter des séquences contenant plusieurs gènes (et non un seul) à condition qu'ils soient dans le même ordre et la même orientation sur les deux séquences traitées. De plus, les codons interrompus par un intron sont correctement traités. L'évaluation de l'algorithme sur un ensemble de séquences biologiques tests a été réalisé. Globalement les résultats obtenus sont très proches de ceux obtenus par l'algorithme Pro-Gen, ce qui est normal car les problèmes traités par les deux algorithmes sont très voisins. Pro-Gen est également fondé sur le principe de programmation dynamique ; il vise à optimiser un critère. Ce programme permet de comparer deux séquences et de trouver dans chaque séquence une chaîne d'exons

les plus similaires entre eux au niveau protéique. Utopia est clairement plus performant que Pro-Gen dans les situations où des inexactitudes expérimentales introduisent des décalages de phase de lecture. Utopia s'avère cependant plus adapté à un travail manuel de raffinement d'annotation qu'à l'analyse automatique de grandes séquences génomiques car les temps de calcul deviennent lourds au-delà d'environ 15 Kbases.

Ce travail a été obtenu grâce à des collaborations avec C. Mathé, T. Schiex, P. Rouzé et M.-F. Sagot [137, 27].

**Inférence de motifs structurés** Laurent Marsan, qui est maintenant maître de conférences à l'université de Versailles-Saint Quentin, a réalisé sa thèse [152] sous la direction de Maxime Crochemore et de Marie-France Sagot sur l'inférence de motifs structurés (algorithmes et outils appliqués à la détection de sites de fixation dans les séquences génomiques).

Il s'est intéressé à la détection de sites de fixation de facteurs de transcription dans des séquences d'ADN. Il a développé une nouvelle représentation de ces sites en introduisant la notion de modèles structurés et en développant des algorithmes combinatoires exacts pour leur détection. Les facteurs de transcription qui régulent l'expression des gènes se fixent sur certains sites en amont des gènes. La conformation spatiale de certains facteurs est telle qu'ils se fixent sur plusieurs sites simultanément ; de plus, il est souvent nécessaire qu'il y ait des interactions entre différents facteurs se fixant sur des sites différents pour que l'expression soit correctement régulée. Il existe donc une association entre divers sites situés à des distances contraintes l'un de l'autre.

Le travail est basé sur celui de M.-F. Sagot qui avait développé un algorithme d'extraction de modèles simples communs à plusieurs séquences ; le modèle est un motif consensus soumis à un critère de quorum, le quorum indiquant le nombre minimum de séquences qu'un modèle doit recouvrir pour être valide. Les séquences traitées sont stockées dans une structure d'arbre des suffixes. L'arbre des suffixes est une structure de données classique en algorithmique du texte permettant d'indexer les suffixes d'un mot. Dans l'étude réalisée par L. Marsan, les sites de fixation peuvent être composés d'un nombre quelconque de motifs séparés par des « spacers » de longueur variable et, éventuellement, très longs. D'où la notion de modèle structuré. Les algorithmes doivent donc permettre l'extraction de consensus souples c'est-à-dire composés d'un nombre quelconque de boîtes séparées par des « spacers » de longueur comprise dans un intervalle quelconque, sans limitation théorique sur le nombre de substitution de bases possibles dans les boîtes ni sur la longueur des boîtes. Un arbre plus léger que l'arbre des suffixes a été utilisé. En effet, on fixe la longueur maximale des modèles à extraire. Ainsi, lors d'extraction de modèles dont la longueur est comprise entre  $kmin$  et  $kmax$ , un arbre de tous les facteurs de longueur comprise entre ces deux bornes suffit. Un tel arbre (*i.e.* l'arbre des  $k$ -facteurs) a été défini par J. Allali et M.-F. Sagot [156]. Une nouvelle structure d'indexation des facteurs d'un texte basée sur l'arbre des suffixes a donc été proposée pour ce qui concerne les applications d'extraction de motifs dans

les séquences biologiques où les motifs intéressants sont souvent courts ou composés de parties courtes à des distances contraintes les unes des autres. L'arbre des suffixes est un arbre dont les arcs sont étiquetés par des mots tel que tout chemin depuis la racine à une des feuilles de cet arbre épelle un des suffixes du mot. Cette structure peut être construite en temps linéaire en la taille du texte et occupe un espace mémoire lui aussi linéaire en la taille du texte, ce qui en fait une structure d'indexation efficace utilisée dans de nombreux domaines tels que la bioinformatique, la compression de données, etc. L'arbre des facteurs de longueur  $k$ , aussi appelé l'arbre des  $k$ -facteurs est en fait un arbre des suffixes tronqué. C'est-à-dire que tout chemin depuis la racine jusqu'à une feuille de l'arbre épelle un facteur du mot de longueur au plus  $k$ . En partant de l'algorithme de Ukkonen pour la construction de l'arbre des suffixes, un algorithme permettant de construire un arbre des facteurs en un temps linéaire en la longueur du texte indexé est proposé. Bien que théoriquement, le gain mémoire ne puisse être prouvé dans le pire des cas (sauf pour de petites valeurs de  $k$ ), l'arbre des facteurs se montre en pratique économique en espace mémoire. L'un des codages les plus performants de l'arbre des suffixes, proposé par S. Kurtz, a été adapté afin de construire l'arbre des  $k$ -facteurs. En pratique des gains mémoire substantiels peuvent ainsi être obtenus comme par exemple :  $-15\%$  pour l'indexation du chromosome 14 de l'homme (87 mégabases) avec une valeur de  $k$  égale à 15.  $-13\%$  en moyenne pour un arbre des 20-facteurs sur du texte structuré tel que le code source d'un programme.

Les algorithmes développés par Laurent Marsan pour détecter les sites de fixation des facteurs de transcription ont été testés sur des données biologiques afin de cerner les capacités et les limites des modèles et algorithmes développés. Comparativement aux outils existants, ils s'avèrent capables de traiter de plus grandes quantités de données et d'extraire des signaux plus faiblement représentés grâce au gain de spécificité apporté par les modèles structurés.

L. Marsan a développé à partir de ces algorithmes un outil dénommé SMILE pour « Structured Motifs Inference and Evaluation ». SMILE est un outil d'inférence permettant d'extraire des modèles simples ou structurés et d'estimer ensuite la signification des modèles obtenus au moyen de différentes mesures de signification. De même que les algorithmes qu'il utilise, SMILE est, lors de sa mise au point, le seul outil permettant l'inférence de modèles structurés composés d'un nombre quelconque de boîtes sur un alphabet quelconque et sans limitation sur les caractéristiques structurelles des modèles. Cependant, sur des données biologiques, lorsqu'on a peu de connaissance sur le modèle à extraire, SMILE requiert de progresser en tâtonnant par extractions et évaluations successives afin de cerner les caractéristiques structurelles les plus pertinentes. Son utilisation diffère donc radicalement de celle des outils d'inférence les plus couramment employés tel MEME, Gibbs Sampler ou Motif Sampler basés sur une approche d'optimisation, approche qui choisit de privilégier l'optimisation des critères de signification pour guider l'inférence. SMILE vient donc en complément des outils d'inférence basés sur une approche d'optimisation ; il permet de décrire le motif de façon plus fine et

l'approche développée est également moins sensible au bruit. Mais, malgré la possibilité d'exprimer des fourchettes de longueur pour les modèles et les « spacers », certains paramètres tel le quorum sont délicats à estimer lors d'inférence dans un contexte où l'on n'a pas de connaissance *a priori*.

Le travail a fait l'objet de publications avec A. Vanet, A. Labigne et M.-F. Sagot [58, 55].

**Modélisation et comparaison des structures secondaires d'ARN** Julien Allali (doctorant depuis 2001) termine sa thèse sous la direction de Maxime Crochemore et de Marie-France Sagot sur ce sujet. Les ARN, qu'ils soient ARN messagers ou non, ont un rôle très important dans le fonctionnement d'une cellule. Ces dernières années ont d'ailleurs permis de découvrir l'importance grandissante de ce rôle. Ces ARN qui sont majoritairement sous forme d'un simple brin, se replient dans l'espace et ce repliement est indispensable à leur activité. Le repliement dans l'espace de la molécule d'ARN qui est une séquence linéaire de nucléotides (A,C,G et U) est basé sur l'appariement des nucléotides deux à deux créant ainsi une série d'hélices. Julien Allali a débuté par une modélisation des structures. Le but est de mettre en évidence des régions proches d'un point de vue structural et l'approche est basée sur l'idée d'avoir un gradient de granularité dans la comparaison des structures. La modélisation consiste en plusieurs arbres liés entre eux par des relations d'abstraction. Ainsi, l'on dispose d'un arbre de plus haut niveau permettant de voir la structure de façon macroscopique et d'un arbre de plus bas niveau dans lequel on distingue chaque paire de bases. Cette structure a été nommée RNA-MiGaL (RNA Multiple Graph Layers).

La comparaison des structures RNA-MiGaL nécessite un algorithme de comparaison d'arbres. Pour cela on commence par comparer les arbres de plus haut niveau puis on passe aux arbres plus détaillés en exploitant les informations collectées par la comparaison précédente, et ce jusqu'à l'arbre de plus bas niveau. Un algorithme d'édition basé sur les trois opérations standard : le renommage, l'insertion et la délétion a été communément utilisé dans la littérature pour des comparaisons de structures secondaires d'ARN. Cependant, comme cet algorithme montre des limitations, de nouvelles opérations d'édition permettant le regroupement de nœuds ou d'arcs dans un même arbre ont dû être introduites. Cet algorithme de comparaison d'arbres par programmation dynamique qui permet de comparer deux structures MiGaL est en cours de finition. Parallèlement, un algorithme efficace permettant la comparaison multiple de RNA-MiGaL est en cours de développement dans le but de faire du « clustering » automatique ainsi que de l'inférence de modèles pour un ensemble d'ARNs d'une même famille. La perspective de pouvoir retrouver les appariements entre bases se formant lors du repliement de l'ARN par cette nouvelle approche spécifique à RNA-MiGaL est également envisagée.

La publication de ce travail est en cours ; il a fait l'objet d'un premier article avec M.-F. Sagot [60].

**Détection et localisation des événements de recombinaison entre génomes au cours de l'évolution**

La fluidité des génomes au cours de l'évolution est une notion maintenant bien répandue. Des événements divers se sont produits tels que insertions/délétions, inversions, duplications, transpositions modifiant la structure des génomes. Marina Zelwer, doctorante, s'est intéressée au problème des conséquences quand on fait des analyses phylogénétiques des événements de recombinaison qui se sont produits au cours de l'évolution, notamment ceux entre gènes paralogues (issus d'une duplication de gènes au sein d'un génome). Elle a développé pour localiser dans les génomes les points de recombinaison une méthode appelée DRUID (Determining Recombinations Using an Incongruence Detector) basée sur une approche par fenêtre glissante. Malheureusement, pour des raisons personnelles, cette étudiante a abandonné son travail de thèse au bout de deux ans.

**Une approche multiple de l'inférence de la structure secondaire des ARN**

Benôit Olivieri, doctorant depuis 2003, réalise sa thèse sous la direction de Maxime Crochemore et de Marie-France Sagot sur ce sujet. Certains problèmes d'inférence de structure spatiale telle l'identification de petits motifs structuraux dans l'ARN peuvent être abordés par des extensions des approches utilisées pour l'inférence de motifs. Les boîtes d'un motif multiple (correspondant à un motif structural dans l'ARN) vérifient alors une certaine relation entre elles. L'inférence de structures spatiales complètes d'ARN pose des problèmes plus complexes. De nombreux ARN adoptent en effet une structure spatiale composée d'un arrangement complexe de palindromes qui se succèdent et/ou s'imbriquent. Une approche comparative est développée en se basant sur certaines hypothèses concernant la nature d'une structure d'ARN conservée. L'hypothèse principale est qu'au sein d'une famille d'ARN (qui est un regroupement de séquences d'ARN connues pour avoir la même fonction), le nombre et l'arrangement des palindromes ainsi que quelques séquences de petite taille sont conservés. De plus, la position de ces segments conservés par rapport à l'arrangement des palindromes est conservée. L'objectif initial est donc de rechercher les éléments conservés en séquence et en structure de façon simultanée.

Ce travail est en cours et bénéficie d'ores et déjà de l'expertise au sein de l'équipe sur les structures de données associées au traitement des séquences.

**Techniques multiples de filtrage des séquences biologiques appliquées à l'inférence de motifs et de gènes**

Pierre Peterlongo, doctorant depuis 2003, sous la direction de Maxime Crochemore et de Marie-France Sagot travaille sur ce sujet. Le but est de proposer un algorithme efficace permettant le filtrage de séquences d'ADN afin d'accélérer ensuite l'alignement de ces séquences. Les alignements locaux multiples ne sont en effet pas possibles lorsque l'on dispose d'un grand nombre de longues séquences. Il est donc très intéressant de pouvoir utiliser une technique de filtrage de données permettant de limiter la taille des séquences à aligner. L'algorithme de filtrage doit

détecter des facteurs présents dans toutes les séquences (ou présents dans un quorum de séquences) avec un minimum de degré de similarité et qui feront potentiellement partie de l'alignement local tandis que tous les autres ensembles possibles de sous-séquences n'en feront pas partie.

La similarité utilisée ici est basée sur la distance de Hamming (i.e. le nombre de substitutions permettant de transformer l'une dans l'autre). Le travail comporte deux étapes : d'une part la détection et la formalisation mathématique des critères de filtrage et d'autre part l'application de ces critères aux séquences d'ADN. La première partie concernant la détermination des critères mathématiques de filtrage est terminée. La seconde phase (l'application algorithmique de ces critères) conduit notamment à détecter des motifs d'un type particulier appelés motifs à trous (Gapped Motifs) qui sont des motifs composés d'un ensemble de boîtes de taille fixée espacées par des trous de longueur bornée. Ce travail est dans la continuité de l'étude des motifs avec jokers exposée plus haut.

**Motifs dans un texte aléatoire** Dans une collaboration avec Mireille Régnier et Mathias Vandenberghe [157], Julien Clément s'intéresse à l'étude du calcul effectif du nombre d'occurrences attendu d'un motif dans le génome étant donné un modèle aléatoire. Ici, on considère un motif généralisé puisque ce motif peut constituer un ensemble de mots (en considérant une certaine notion de voisinage) et même contenir des trous de longueur variable.

Ce travail a pour application l'aide à l'extraction de motif. Les algorithmes sont actuellement implantés ou en cours d'implantation dans la bibliothèque `quickscore` (voir <http://pauillac.inria.fr/algo/online.html>).

### 2.3.4 Programmation générique et réseaux

**Multi-polymorphisme** Contrairement aux langages fonctionnels, le langage Java ne dispose pas de constructions spécifiques destinées au filtrage des types structurés. Cette caractéristique rend peu lisible et difficilement maintenable le code de certaines implantations, comme les parcours de graphes ou d'arbres. Le multi-polymorphisme et son implantation sous la forme de multi-méthodes est le mécanisme orienté-objet qui paraît le mieux adapté pour combler ce manque. Il permet de sélectionner la méthode à appeler en fonction du type dynamique des arguments de l'appel, à la manière du filtrage (`match`) fonctionnel. Malheureusement, en Java, ce mécanisme n'est pas disponible et les algorithmes existants ont été développés pour des langages dont le système de types est clos. L'approche proposée par Rémi Forax, Etienne Duris et Gilles Roussel pour introduire les multi-méthodes en Java a été de fournir un paquetage, `JMMF`, qui permet de simuler les multi-méthodes en utilisant le mécanisme de réflexivité du langage. Le développement de ce paquetage s'est appuyé sur deux nouveaux algorithmes de sélection de méthode [53] qui intègrent les aspects dynamiques de Java. D'un point

de vue pratique, l'utilisation de ce paquetage facilite le développement et la maintenance des logiciels sans trop pénaliser les performances par rapport aux approches *ad-hoc* existantes qui sont moins portables. Ces performances peuvent certainement être encore améliorées en utilisant des approches par génération de code.

**Vérification d'interfaces graphiques** Si de nombreux outils existent déjà pour construire l'aspect visuel des interfaces, il n'existe pas pour l'instant de méthodologie ou d'outil largement acceptés permettant de spécifier et de vérifier le comportement logique des interfaces graphiques. L'idée à l'origine de ce travail a été de considérer les suites d'événements valides, produits lors de l'utilisation d'une interface graphique, comme les phrases d'un langage formel caractérisant la logique de cette interface. Cette constatation a tout d'abord conduit Jean Berstel, Stefano Crespi Reghizzi, Pierluigi San Pietro et Gilles Roussel à proposer un formalisme de spécification [76, 23], appelé VEG, basé sur les techniques classiques de compilation et adapté au développement d'interfaces graphiques. Dans ce formalisme, le comportement de chaque composant est décrit par une grammaire. En réutilisant les techniques classiques de compilation, ils ont développé des outils qui, à partir de cette spécification, produisent un code source Java. Il ne reste plus qu'à lier ce code avec la partie visuelle et les bibliothèques sémantiques pour compléter l'interface graphique.

L'utilisation d'un formalisme basé sur les grammaires leur a également permis d'interfacer ces outils avec un moteur de vérification automatique basé sur la technique de *model checking*. Il est ainsi possible de vérifier automatiquement des propriétés d'interblocage et d'accessibilité. Des propriétés plus complexes peuvent également être vérifiées moyennant la spécification de formules de logique temporelle par l'utilisateur.

**Machine virtuelle ouverte** Bien que les techniques liées au développement de langages de programmation (production de compilateur, développement de machine virtuelle) soient bien maîtrisées, ces outils sont mal adaptés au développement incrémental des langages par ajouts successifs, ou à l'adaptation de ces langages à de nouvelles contraintes. De nombreux travaux récents dans ce domaine le prouvent puisque chacun d'eux a conduit au développement complet d'un des éléments de la chaîne de compilation.

Cette constatation a amené Christophe Deleray, Nicolas Bedon, Etienne Duris, Rémi Forax et Gilles Roussel à mettre en place un projet de longue haleine qui a pour but d'étudier le développement d'une chaîne complète de production de logiciel (le compilateur et la machine chargée de l'exécution) permettant l'introduction incrémentale de nouvelles propriétés, fonctionnalités ou caractéristiques dans un langage. Ils se sont d'abord penchés sur la machine virtuelle dans le but de la rendre modulaire et ouverte [104, 103] afin de permettre des extensions statiques et dynamiques directement écrites en Java. Corosol est un premier prototype de machine virtuelle ayant ces propriétés. Son architecture est découpée en composants fonctionnels simples et de gros grain

qui correspondent aux divers éléments de la machine : tas, pile, etc.... Il est possible d'accéder ou de modifier ces composants au cours d'une exécution *via* une interface d'introspection étendue. Ces caractéristiques permettent de modifier le comportement de la machine virtuelle de façon incrémentale, tout en préservant sa portabilité. Malheureusement, les performances de cette machine virtuelle la cantonne à une utilisation de recherche. Une des perspectives de travail actuelle est de minimiser le surcoût induit par l'empilement de machines virtuelles en réutilisant au maximum la machine sous-jacente.

**Routage dans un réseau de robots** Le problème du routage des messages dans un réseau est un problème algorithmique qui est toujours largement étudié. Pirro Bracka, Serge Midonnet et Gilles Roussel étudient le problème particulier du routage de messages dans un réseau constitué de robots mobiles et plus précisément le problème de la communication asynchrone en mode *ad-hoc* dans ce réseau, c'est-à-dire dans un mode où les messages sont acheminés de robot en robot sans utiliser d'infrastructures fixes. Dans ce contexte très large, l'idée est de forcer le déplacement des robots pour simplifier le routage et assurer l'acheminement des messages dans le réseau. Quelle que soit la topologie de réseau, ils ont prouvé de façon constructive [82, 83], en utilisant un formalisme à base d'automates, qu'il est toujours possible de trouver un ordonnancement des déplacements des robots qui assure que tous les robots pourront communiquer les uns avec les autres dans un temps borné. Une fois l'ordonnancement des mouvements fixé, cette approche permet de ramener le routage des messages à un problème de routage statique pour lequel il existe de nombreux résultats. Outre l'ordonnancement des mouvements et le routage des messages, ils ont également développé une stratégie pour gérer la tolérance aux pannes et pour améliorer le choix des ordonnancements.

**Réplication de données à grande échelle** Actuellement, la plupart des organisations disposent d'un nombre important de postes de travail et ils ont besoin d'outils performants leur permettant de déployer rapidement des systèmes ou des logiciels. Une solution intéressante lorsqu'un grand nombre de postes doit être installé ou mis à jour simultanément consiste à utiliser le principe de multicast IP. Malheureusement, ce protocole est intrinsèquement non fiable, ce qui est problématique pour ce type d'application. Jérôme Petazzoni a développé un protocole très simple de déploiement en multicast fiable, rapide et qui permet un passage à l'échelle [118]. Celui-ci a été implanté dans le logiciel Deplika qui a été testé pour le déploiement simultané d'une centaine de machines.

**Tolérance aux fautes temporelles** La thématique étudiée est la prise en compte des fautes temporelles dues à une sur-utilisation de ressources (faute) ou à un dépassement d'échéance (défaillance) dans un système CORBA. La tolérance aux fautes et le temps-réel sont deux aspects de la qualité de service qui interviennent dans CORBA à travers

deux spécifications qui s'ignorent : Real-Time CORBA et Fault Tolerant CORBA. Choisir un objet temps-réel plutôt qu'un objet non temps-réel peut avoir des conséquences négatives sur la tolérance aux fautes d'une application. Dans un système CORBA tolérant aux fautes, un composant est considéré comme défaillant lorsqu'il n'interagit plus avec son environnement. Dans ce travail un composant est considéré comme fautif lorsqu'il ne respecte plus ses contraintes temporelles (coût initial, période, échéance). Nous proposons une solution pour la prévention des fautes dans le cas des applications à longue durée d'interactivité (applications composées de tâches périodiques [107, 79, 111]). Dans ce cas, une analyse de faisabilité et une réservation de ressources sont envisageables. Nous proposons également une solution pour la détection de surcharge dans le cas des applications à courte durée d'inter-activité (applications composées de tâches aperiodiques [78]). Dans ce cas, des détecteurs de surcharge seront implantés globalement (au niveau du système) ou localement (associés à chaque tâche). Nous introduisons deux nouvelles politiques pour la tolérance aux fautes des servants au sein d'un système CORBA : Provisioning and Overload Detection. Politiques implantées au sein d'un nouvel adaptateur d'objets : le FTPOA. Une application pourra alors décider de la valeur de ces deux règles. La validation de ce travail est faite par la réalisation du système SORBET. Ce système est implanté sur une machine virtuelle Java Temps-réel conforme à la nouvelle norme RTSJ et embarque les politiques précédemment citées. Certains paramètres qui influencent le déterminisme du système ont été identifiés comme le coût d'activation d'un thread, la variation de la période pour les applications périodiques et le coût des changements de contexte. Ces différents paramètres ont été modélisés et intégrés dans l'analyse théorique de faisabilité ainsi qu'au sein des détecteurs de surcharge [112].

## 2.4 Activités

### 2.4.1 Contrats

- Maxime Crochemore est coresponsable du projet ECOS : *Pattern matching and application*, 1999–2001, Chili ;
- Projet Bioinformatique CNRS-INRA-INRIA-INSERM *Régulation, Synténie et Pathogénicité — Algorithmes et expérimentations*. Partenaires : Institut Pasteur, Institut de Biologie Physico-Chimique de Paris et Institut Gaspard Monge. Coordinateurs Anne Vanet et Marie-France Sagot (2000-2002) ;
- *Algorithmique pour la Bioinformatique*, Action Spécifique du département STIC du CNRS ;
- NATO Grant : *String Algorithms*, 2000-2002 ;
- Maxime Crochemore est coresponsable du réseau *Traitement et analyses de séquences*, action du MENRT, 2000-2002 ;
- Maxime Crochemore est coresponsable du projet VINCI : *Algorithmes et struc-*

- tures de données pour l'extraction de motifs dans les séquences biologiques*, 2001 ;
- Maxime Crochemore est coresponsable de l'Action Spécifique du département STIC du CNRS *Algorithmes pour la Bioinformatique*, 2001-2003 ;
  - Projet de Bioinformatique inter EPST 2002 : *Algorithms for modelling search and Inference problems in Molecular Biology* ;
  - Action Spécifique du département STIC du CNRS *Algorithmes et séquences*, 2002-2003 ;
  - Bourse de la Wellcome Trust Fondation accordée aux équipes du King's College à Londres, Université de Marne-la-Vallée et Institut Pasteur jusqu'en 2003 (échange de chercheurs en vue de collaborations en particulier pour l'étude de la combinatoire des mots et l'élaboration d'algorithmes permettant de traiter certains problèmes en biologie) ;
  - Royal Society *Pattern inference in computational molecular biology*, pour 2002-2003 ;
  - Action Spécifique du département STIC du CNRS : *Indexation de texte et découverte de motifs*, 2003-2004 ;
  - L'IGM est aussi impliquée dans l'Action Spécifique CNRS/STIC dont les premières journées, intitulées *Systèmes dynamiques et modélisation en algorithmique* ont eu lieu les 28 et 29 novembre 2002, à Paris ;
  - ACI du Ministère de la Recherche : *Nouvelles Interfaces des Mathématiques : Mathematical and Algorithmical aspects of biochemical and evolutionary networks*, 2004-2007 ;
  - Contrat MathSTIC pour l'année 2001–2002, intitulé *Les mots : de la combinatoire à la dynamique symbolique*, dirigé par Valérie Berthé (Université de Montpellier) et Jean Berstel ;
  - Contrat MathSTIC pour l'année 2002–2003, intitulé *Numération non-standard, substitutions, pavages et quasi-cristaux*, dirigé par Pierre Arnoux et Frédérique Bassino ;
  - Projet de coopération CNRS/CNR (Italie) en 2001 et 2002 dirigé par Jean Berstel ;
  - Projet de coopération CNRS/NHRF (Grèce) en 2001 et 2002 ;
  - Projet de coopération CNRS/JSPS (Japon) en 2003 ;
  - Projet de coopération CNRS/FNRS (Afrique du Sud) en 2003 et 2004 dirigé par Frédérique Bassino ;
  - Action Spécifique du département STIC du CNRS *Systèmes dynamiques et modélisation en algorithmique* en 2002-2003 ;
  - Frédérique Bassino, Julien Clément et Cyril Nicaud font partie du groupe de travail ALEA du G.D.R. ALP (avec une rencontre qui a lieu tous les ans et dure une semaine) ;
  - Julien Clément est membre de l'Action Spécifique *Indexation de texte et découverte de motifs* et de l'ACI NIM 2004 *DynamicAL : Interface mathématiques et informatique, Dynamique et Algorithmique* (Responsables : Véronique Maume, Brigitte

Vallée). Il est également membre associé au Projet Algorithmes de l'INRIA ;

- Frédérique Bassino est membre de l'ACI NIM 2004 *Nouvelles techniques en numération*.

## 2.4.2 Diffusion

**Livres** La période couverte par ce rapport a été riche en publication de livres de diffusion de connaissances, tant au niveau des livres d'enseignements (actuelle licence ou master professionnel) que pour la recherche (master recherche).

- *Algorithmique du texte*, Vuibert 2001, de Maxime Crochemore, Christophe Hancart et Thierry Lecroq.
- *Algebraic Combinatorics on Words*, Cambridge University Press, édité par Jean Berstel et Dominique Perrin qui sont aussi parmi les auteurs.
- *Jewels of Stringology*, de Maxime Crochemore et Wojciech Rytter, World Scientific, 2002.
- *Infinite Words, Automata, Semigroups, Logic and Games*, de Dominique Perrin et Jean-Eric Pin, Academic Press, 2004.
- *Applied Combinatorics on Words*, Cambridge University Press, édité par Jean Berstel et Dominique Perrin qui sont aussi parmi les auteurs (sous presse).
- *Java et Internet*, de Gilles Roussel, Étienne Duris, Nicolas Bedon, Rémi Forax ; Deuxième édition entièrement refondue, Vuibert, 2002.

## Logiciels

- Corosol : <http://igm.univ-mlv.fr/~cdeleray/>
- JMMF : <http://igm.univ-mlv.fr/~forax/works/jmmf/>
- VEG : <http://www.elet.polimi.it/upload/campi/veg/>
- Smile : <http://igm.univ-mlv.fr/~marsan/smile.html>
- Utopia : <http://igm.univ-mlv.fr/~blayo/utopia-front.html>

**Conférences** Marie France Sagot a donné des cours dans la *School on Algorithms and Combinatorics* en Mars 2001 à Cerea, Brésil.

Marie-France Sagot et Maxime Crochemore ont organisé la conférence JOBIM (*Journées ouvertes : Biologie, Informatique et Mathématiques*) et du séminaire *Algorithmique et Biologie*.

Maxime Crochemore a donné les conférences :

- *Approximate String matching in musical sequences* (Prague stringology club, Prague 2001) ;
- *Speeding up Hirschberg and Hunt-Szymanski LCS algorithms* (Symposium on string processing and information retrieval, Laguna de San Rafael, Chile 2001) ;
- *Algorithmique génomique* (Comité IMPG, Paris, 15–16 janvier 2002) ;

- *Alignement sous-quadratique* (Action AlBio du CNRS, Montpellier, 11–13 mars 2002) ;
- *A subquadratic sequence alignment algorithm for unrestricted cost matrices* (ROSA'2002, Rouen, 6–7 juin 2002) ;
- *Alignment* (University of Leicester, 24 janvier 2003) ; *Indexing and Compressing, Repeats, Alignments* (Lipari School, 14-25 juillet 2003) ;
- *Bases of repeated motifs in texts* (22e Colloque international Grammaires et lexiques comparés, Thessaloniki, Grèce, 17-21 septembre 2003) ;
- *Periodicities and Algorithms on Words, Indexing and Compressing Texts, Repeats* (51e Séminaire Lotharingien de Combinatoire, Bertinoro, Italie, 22-24 septembre 2003) ;
- *Pattern Matching and Text Indexing, Repeated Motifs in Sequences* (Université de Milan Bicocca, Italie, 13-14 novembre 2003) ;
- *Index compacts* (Rencontres de l' AS Indexation de texte et découverte de motifs du CNRS, Montpellier, 20-21 novembre 2003) ;
- *Subquadratic alignment* (Université technique de Lisbonne, 27 février 2004) ;
- *Burrows-Wheeler Transform and Permutations* (Caesarea Rothschild Institute, Université de Haïfa Israël 10 mai 2004).

Julien Allali a donné les conférences :

- *L'arbre des suffixes tronqués* (Séminaires IGM-jeunes chercheurs, 5 Juin 2002) ;
- *RNA secondary structure modelling and comparison* (Projet INTER EPST, Lyon, 31 Mars 2003) ;
- *MiGaL : Modélisation et algorithmique des structures d'ARNs* (AS STIC, Modélisation et algorithmique des structures d'ARN, Bordeaux, 9 Octobre 2003 et Université de Versailles, 28 Novembre 2003) ;
- *Novel tree edit operations for RNA secondary structure comparison* (4th Workshop on Algorithms in Bioinformatics, 15 Septembre 2004).

Pierre Peterlongo a donné les conférences :

- *Indexation et recherche de motifs* (Journée d'action spécifique du CNRS, 28 Mai 2004) ;
- *Présentation de la structure de l'arbre des bi-facteurs* (Séminaires IGM – jeunes chercheurs, 22 Juin 2004) ;
- *A first approach to finding common motifs with gaps* (Prague Stringology Conference, 1er Septembre 2004).

Chloé Rispal a donné des conférences :

- à l'École jeunes chercheurs en algorithmique et calcul formel, Université de Marne-la-vallée, le 02/04/2003 ;
- aux Journées montoises d'informatique théorique, Université de Liège, Belgique, le 10/09/2004.

Julien Clément a participé aux manifestations suivantes :

- Séminaire du projet POLKA (organisé par G. SCHAEFFER) au LORIA en 2001 ;

- Séminaire de l’université d’Evry (séminaire Statistique et Génome organisé par B. PRUM) en 2001 ;
- Séminaire général du laboratoire d’informatique de l’Université de Marne-la-Vallée (organisé par F. BASSINO) en 2001 ;
- Séminaire de l’Université de Créteil en avril 2001 (organisé par A. DURAND) ;
- Journées internationales d’analyses d’algorithmes *AofA* (Analysis of Algorithms) à Strobl (Autriche) en 2002 ;
- Rencontre conjointe au réseau *String Algorithmics* de l’OTAN et au groupe de travail *Algorithmique des séquences* du GDR ALP (Algorithmique, Langage et Programmation) du département STIC du CNRS en 2002 ;
- Groupe de travail *Arbres Aléatoires et Algorithmes* en 2003 à Versailles ;
- Rencontres ALEA en 2003 à Marseille ;
- Séminaire et groupe de travail d’algorithmique du GREYC à Caen en 2004 ;
- Action Spécifique *Indexation de texte et découverte de motifs* à Nantes en mai 2004.

Francesca Fiorenzi a donné les conférences suivantes :

- *Dynamique symbolique et invariants*, au colloque : La question de la modélisation en sciences humaines : mathématiques et informatique, EHESS, Paris, novembre 2003 ;
- *Minimal forbidden patterns of multi-dimensional shifts*, aux Journées Montoises d’Informatique Théorique, Montpellier, septembre 2002 ;
- *Semi-strongly irreducible shifts*, au colloque : Words, Palerme, septembre 2001.

Marie-Pierre Béal a été responsable du thème « Systèmes dynamiques symboliques » pour l’École jeunes chercheurs en algorithmique et calcul formel, Lille 2003, et a donné un cours d’introduction à la dynamique symbolique.

Christophe Deleray a donné les conférences :

- *MyJVM : a 100% pure Java parameterizable Java virtual machine* (PPPJ, Kilkenny 18 juin 2004) ;
- *Corosol une JVM modulaire paramétrable à la volée* (LMO, Lille 15 mars 2004).

Jérôme Petazzoni a donné les conférences :

- *Deplika : a scalable multicast tool for Unix system replication* (CST, Cancun le 20 mai 2003) ;
- *BUL : Block device in Userland - Un terrain d’expérimentation générique pour le stockage de données* (GRM Paris le 18 décembre 2003).

Pirro Bracka a donné les conférences :

- *Routage dans un réseau de robots* (Algotel, Mèze le 24 mai 2002) ;
- *Scheduling and Routing in an ad-hoc network of robots* (CST, Cancun le 20 mai 2003).

Serge Midonnet a donné les conférences :

- *Aperiodic Invocations Admission in Real-Time CORBA* (IEEE/Softcom, Split le 11 Octobre 2004) ;

- *An Adaptative Allowance For Failure Prevention Of Fixed Priority Scheduled Real-Time Systems* (IASTED/PDCS, Cambridge le 11 Novembre 2004).

### 2.4.3 Travaux éditoriaux et organisation de colloques

- Dominique Perrin est éditeur pour les journaux *Theoretical Computer Science* (Elsevier), *Advances in Applied Mathematics* (Elsevier), *Semigroup Forum* (Springer) et *International Journal of Algebra and Computation* (World Scientific).
- Jean Berstel est membre du comité de rédaction de la revue RAIRO Informatique théorique et applications. Il a également été membre du comité de programme des colloques : Words'03 et DLT'04.
- Marie-Pierre Béal a été membre du comité de programme du colloque CIAA'2002.
- Maxime Crochemore est membre du comité éditorial de *Theoretical Computer Science* (depuis 1993) et de *Computational Biology and Chemistry* (depuis 2004). Il a également été co-président du comité de programme de CPM'2003 et membre des comités de programme de PSC'2001, PSC'2002, PSC'2003, PSC'2004, de JO-BIM'2001, de CIAA'2001, CIAA'2002, de SPIRE'2001, SPIRE'2002, SPIRE'2003, SPIRE'2004, de ICALP'2001, de ISAAC'2001 et de MFCS'2002.
- L'IGM est organisateur de l'*Ecole Jeunes Chercheurs* du GDR ALP en Algorithmique et Calcul Formel qui s'est tenu à l'Université de Marne-la-Vallée du 31 mars au 4 avril 2003.
- L'IGM a été co-organisateur des séminaires Graphes, Réseaux et Modélisation en 2003 et 2004 et des journées du même nom les 17 et 18 décembre 2003.
- Julien Cervelle a organisé, à l'IGM, un *Workshop Automates Cellulaires*, les 26 et 27 mai 2004 qui a réuni 14 participants.
- Julien Clément est membre du comité d'organisation des 7<sup>e</sup> Rencontres Internationales d'Analyse d'Algorithmes qui se sont déroulées à Tatihou (Normandie) du 3 au 10 juillet 2001. Ces rencontres rassemblent environ une soixantaine de chercheurs.
- Gilles Roussel était membre du comité d'organisation des séminaires inter-site « Graphe, Réseau et Modélisation » en 2002-2003.

### 2.4.4 Collaborations

- Équipe d'algorithmique de C.S. Iliopoulos au sein du département d'informatique du King's College of London. Un projet commun est financé par la Wellcome Trust Fondation ;
- M. Crochemore en tant que membre de l'IMPG (Informatique, Mathématique, Physique pour la Génomique) anime avec B. Prum et J. L. Risler des rencontres *Traitement et analyse des séquences* ;
- *Algorithmes et séquences*, projet à l'intérieur du GDR ALP : Automate, Language,

Programmation ;

- J. Holub et B. Melichar de la Czech Technical University de Prague (République Tchèque) sur algorithmique du texte et méthode d'indexation ;
- M.F Sagot (UMR 5558, Université de Lyon-1, Directeur C. Gauthier) ;
- S. Crespi-Reghizzi et P. San Pietro (Politecnico di Milano) ;
- A. Viola (Montevideo, Uruguay) sur des questions de théorie de l'information ;
- H. Prodinger en Afrique du Sud (Witwatersrand University, Johannesburg dans le cadre d'un accord de coopération NRF/CNRS) ;
- M. Régnier (INRIA) en bioinformatique ;
- S. Akiyama Université de Niigata au Japon (coopération dans le cadre d'un accord CNRS/JSPS) ;
- F. Mignosi, A. Restivo, M. Sciortino (Université de Palerme) ;
- M. Raffinot (CR CNRS) ;
- O. Carton, C. Prieur (Liafa, Paris 7) ;
- J. Sakarovitch (DR CNRS, ENST Paris) ;
- A. Bergeron (Professeur UQAM, Montréal).

### 2.4.5 Visiteurs

Nous avons eu la visite de

- Filippo Mignosi (Université de Palerme) ;
- Antonio Restivo (Université de Palerme) ;
- Roberto Grossi (Université de Parme) ;
- Juhani Karhumäki (Université de Turku) ;
- Raffaele Giancarlo (Université de Palerme).
- Accueil de Margaret Archibald, doctorante, durant 4 mois à l'IGM dans le cadre d'un programme doctoral CNRS/FNRS (Afrique du Sud).

### 2.4.6 Activités doctorales

DEA Informatique Fondamentale et Applications : la Bioinformatique est enseignée dès le tronc commun dans un cours appelé *Informatique du génome* par Marie-France Sagot, Christiane Rayssiguier et, en 2004, Marie-Pierre Béal.

Plusieurs cours sont donnés dans la filière *Traitement des génomes* :

*Algorithmique des séquences* par Maxime Crochemore (et Thierry Lecroq) ;

*Introduction à la biologie et à l'évolution moléculaire* par Christiane Rayssiguier ;

*Comparaison de macromolécules et inférence de motifs* par Marie-France Sagot ;

*Recherche de motifs et évolution* par Olivier Gascuel (jusqu'en 2002), Christophe Hancart et Pierre Darlu (depuis 2002).

Plusieurs cours sont donnés pour la filière *Logiciels des réseaux* :

*Logiciels des réseaux* par Gilles Roussel ;

*Programmation d'applications réparties* par Serge Midonnet ;  
*Recherche d'information* par Dominique Revuz.

## 2.4.7 Thèses et habilitations

### Habilitations

- Marie-Pierre Béal, sous le titre *Codage symbolique*, à l'Université de Marne-la-Vallée, le 5 janvier 2001. Depuis, Marie-Pierre Béal a été nommée professeur à l'IGM.
- Olivier Carton, sous le titre *Automates et mots infinis* à l'Université de Marne-la-Vallée, le 14 décembre 2001. Depuis, Olivier Carton a été nommé professeur au LIAFA, Université Paris 7.
- Gilles Roussel, sous le titre *Grammaires et automates comme outils pour le développement logiciel*, à l'Université de Marne-la-Vallée, le 8 décembre 2003. Depuis, Gilles Roussel a été nommé professeur à l'IGM.

### Thèses

- Cyril Allauzen, *Combinatoires sur les mots et recherche de motifs*, 29 janvier 2001.
- Simone Bentolila, *La logique et le vivant ; les formalismes de représentation des connaissances en biologie*, 24 Juin 2002.
- Philippe Blayo, *Une approche comparative combinatoire pour la prédiction de gènes chez les eucaryotes*, 6 janvier 2003.
- Rémi Forax, *Les multi-méthodes en Java*, 5 décembre 2001.
- Laurent Marsan, *Inférence de motifs structurés : algorithmes et outils appliqués à la détection de sites de fixation dans des séquences génomiques*, 8 avril 2002.
- Claude Martineau, *Compression de textes en langue naturelle*, 7 décembre 2001.
- Vincent Le Maout, *Expérience de programmation générique sur des structures non-séquentielles : les automates*, 1 juillet 2003.
- Chloé Rispal, *Automates sur les ordres linéaires : complémentation*, 7 décembre 2004.

### Thèses en cours

- Julien Allali, depuis septembre 2001, *Modélisation et comparaison des structures secondaires d'ARN*.
- Pirro Bracka, depuis septembre 2001, *Routage dans un réseau de robots*.
- Julien Carcenac, depuis septembre 2003, *Systèmes de recherche et de filtrage pour le Web*.
- Christophe Deleray, depuis septembre 2001, *Adaptation incrémentale de langages de programmation*.
- Gabriele Fici, depuis octobre 2002, en thèse de co-tutelle Palerme - Marne-la-Vallée, *Mots interdits minimaux et application*. Co-encadrant Filippo Mignosi.

- Gautier Loyauté, depuis septembre 2004, *Outils génériques pour le développement de serveurs non bloquants*.
- Edouardo Moreno, depuis septembre 2001, *Secuencias y grafos de De Bruijn en lenguajes con restricciones*.
- Benoît Olivieri, depuis Septembre 2003, *Une approche multiple de l'inférence de la structure secondaire des ARN*.
- Jérôme Petazzoni, depuis septembre 2001, *Outils génériques pour le stockage distribué de données*.
- Pierre Peterlongo, depuis Septembre 2003, *Techniques multiples de filtrage des séquences biologiques appliquées à l'inférence de motifs et de gènes*.

## 2.5 Responsabilités

- Frédérique Bassino est membre du CNU 27 ;
- Marie-Pierre Béal a été membre nommée du CNU 27 entre 1999 et 2001 ;
- Maxime Crochemore est membre du Conseil scientifique de l'université, il a été membre du Comité national de la recherche scientifique et il directeur scientifique adjoint du département STIC du CNRS ;
- Dominique Perrin est président du Polytechnicum et directeur de l'ESIEE ;
- Christiane Rayssiguier est membre de la Mission Scientifique Technique et Pédagogique du Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche.

## 2.6 Références bibliographiques

### Articles de revues

- [1] A. Amir, A. Butman, **M. Crochemore**, G. M. Landau, et M. Schaps. Two-dimensional pattern matching with rotations. *Theoret. Comput. Sci.*, 314 :173–187, 2004. Full version of [62].
- [2] **F. Bassino**, **M.-P. Béal**, et **D. Perrin**. A finite state version of the Kraft-McMillan theorem. *SIAM J. Comput.*, 30(4) :1211–1230, 2000. (electronic).
- [3] **F. Bassino**, **J. Clément**, et **C. Nicaud**. The standard factorization of Lyndon words : an average point of view. *Discrete Math.*, 2004. À paraître.
- [4] **F. Bassino** et H. Prodinger.  $(q, \delta)$ -numeration systems with missing digits. *Monatshefte fuer Mathematik*, 141(2) :89–99, 2004.
- [5] **M.-P. Béal**. Extensions of the method of poles for code construction. *IEEE Trans. Inform. Theory*, 49(6) :1516–1523, 2003.
- [6] **M.-P. Béal**, A. Bergeron, S. Corteel, et M. Raffinot. An algorithmic view of gene teams. *Theoret. Comput. Sci.*, 320(2-4) :395–418, 2004.

- 
- [7] **M.-P. Béal** et **O. Carton**. Asynchronous sliding block maps. *Theor. Inform. Appl.*, 34(2) :139–156, 2000.
- [8] **M.-P. Béal** et **O. Carton**. Computing the prefix of an automaton. *Theor. Inform. Appl.*, 34(6) :503–514, 2000.
- [9] **M.-P. Béal** et **O. Carton**. Determinization of transducers over finite and infinite words. *Theoret. Comput. Sci.*, 289(1) :225–251, 2002.
- [10] **M.-P. Béal** et **O. Carton**. Determinization of transducers over infinite words : the general case. *Theory Comput. Syst.*, 37(4) :483–502, 2004.
- [11] **M.-P. Béal**, **O. Carton**, C. Prieur, et J. Sakarovitch. Squaring transducers : an efficient procedure for deciding functionality and sequentiality. *Theoret. Comput. Sci.*, 292(1) :45–63, 2003. Selected papers in honor of **J. Berstel**.
- [12] **M.-P. Béal**, **M. Crochemore**, F. Mignosi, A. Restivo, et M. Sciortino. Computing forbidden words of regular languages. *Fund. Inform.*, 56(1-2) :121–135, 2003. Special issue on computing patterns in strings.
- [13] **M.-P. Béal**, **F. Fiorenzi**, et F. Mignosi. Minimal forbidden patterns of multi-dimensional shifts. *International Journal of Algebra and Computation*, 2004. À paraître.
- [14] **M.-P. Béal**, F. Mignosi, A. Restivo, et M. Sciortino. Forbidden words in symbolic dynamics. *Adv. in Appl. Math.*, 25(2) :163–193, 2000.
- [15] **M.-P. Béal** et **D. Perrin**. A weak equivalence between shifts of finite type. *Adv. in Appl. Math.*, 29(2) :162–171, 2002.
- [16] **M.-P. Béal** et **D. Perrin**. On the generating sequences of regular languages on  $k$  symbols. *J. ACM*, 50(6) :955–980, 2003.
- [17] **N. Bedon**. Logic over words on denumerable ordinals. *Journal of Computer and System Sciences*, 63(3) :394–431, Nov. 2001.
- [18] **N. Bedon**. Star-free sets of words on ordinals. *Information and Computation*, 166(2) :93–111, May 2001.
- [19] **J. Berstel**. An exercise on Fibonacci representations. *Theor. Inform. Appl.*, 35(6) :491–498, 2002. A tribute to Aldo de Luca.
- [20] **J. Berstel**. Recent results on extensions of Sturmian words. *Internat. J. Algebra Comput.*, 12(1-2) :371–385, 2002. International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory (Lincoln, NE, 2000).
- [21] **J. Berstel** et L. Boasson. Formal properties of XML grammars and languages. *Acta Inform.*, 38(9) :649–671, 2002.
- [22] **J. Berstel** et L. Boasson. Shuffle factorization is unique. *Theoret. Comput. Sci.*, 273(1-2) :47–67, 2002. WORDS (Rouen, 1999).
- [23] **J. Berstel**, S. Crespi Reghizzi, P. S. Pietro, et **G. Roussel**. A scalable formal method for design and automatic checking of user interfaces. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 2004. À paraître, 49 pages.
- [24] **J. Berstel** et J. Karhumäki. Combinatorics on words—a tutorial. *Bull. European Assoc. Theoret. Comput. Sci.*, 79 :178–228, 2003.

- 
- [25] **J. Berstel** et L. Vuillon. Coding rotations on intervals. *Theoret. Comput. Sci.*, 281(1-2) :99–107, 2002. Selected papers in honour of M. Nivat.
- [26] F. Blanchard, **J. Cervelle**, et E. Formenti. Some results about the chaotic behavior of cellular automata. *Theoret. Comput. Sci.*, 2004. À paraître.
- [27] **P. Blayo**, P. Rouzé, et **M.-F. Sagot**. Orphan gene finding - An exon assembly approach. *Theoret. Comput. Sci.*, 290 :1407–1431, 2003.
- [28] E. Cambouropoulos, **M. Crochemore**, C. S. Iliopoulos, L. Mouchard, et Y. J. Pinzon. Algorithms for computing approximate repetitions in musical sequences. *International Journal of Computer Mathematics*, 79(11) :1135–1148, 2002.
- [29] **O. Carton** et M. Michel. Unambiguous Büchi automata. *Theoret. Comput. Sci.*, 297 :37–81, 2003.
- [30] **O. Carton** et **C. Rispal**. Complementation of rational sets on scattered linear orderings of finite rank. *Theoret. Comput. Sci.*, 2004. À paraître (special issue of LATIN'04 (full version of [85])).
- [31] **J. Cervelle** et B. Durand. Tilings : recursivity and regularity. *Theoret. Comput. Sci.*, 310(1-3) :479–488, Jan. 2004.
- [32] J.-M. Champarnaud, G. Hansel, et **D. Perrin**. Unavoidable sets of constant length. *International Journal of Algebra and Computation*, 14 :241–251, 2004.
- [33] **J. Clément**, P. Flajolet, et B. Vallée. Dynamical sources in information theory : a general analysis of tries structures. *Algorithmica*, 29 :307–369, 2001. (special issue).
- [34] **M. Crochemore**. Reducing space for index implementation. *Theoret. Comput. Sci.*, 292(1) :185–197, 2003.
- [35] **M. Crochemore**, **J. Désarménien**, et **D. Perrin**. A note on the Burrows Wheeler transformation. *Theoret. Comput. Sci.*, 2004. À paraître.
- [36] **M. Crochemore**, C. Hancart, et T. Lecroq. A unifying look at the Apostolico-Giancarlo string-matching algorithm. *Journal of Discrete Algorithms*, 1(1) :37–52, 2003.
- [37] **M. Crochemore**, C. Iliopoulos, M. Korda, et J. Reid. A failure function for multiple two-dimensional pattern matching. *Combinatorial Mathematics and Combinatorial Computing*, 35 :225–238, 2000.
- [38] **M. Crochemore**, C. S. Iliopoulos, T. Lecroq, Y. J. Pinzon, W. Plandowski, et W. Rytter. Occurrence and substring heuristics for  $\delta$ -matching. *Fundamenta Informaticae*, 56(1,2) :1–21, 2003.
- [39] **M. Crochemore**, C. S. Iliopoulos, C. Makris, W. Rytter, A. Tsakalidis, et K. Tsihlias. Approximate string matching with gaps. *Nordic Journal of Computing*, 9(1) :54–65, 2002.
- [40] **M. Crochemore**, C. S. Iliopoulos, et Y. J. Pinzon. Computing evolutionary chains in musical sequences. *Electronic Journal of Combinatorics*, 8(2), 2001. [http://www.combinatorics.org/Volume\\_8/v8i2toc.html](http://www.combinatorics.org/Volume_8/v8i2toc.html).
- [41] **M. Crochemore**, C. S. Iliopoulos, et Y. J. Pinzon. Speeding-up Hirschberg and Hunt-Szymanski LCS algorithms. *Fundamenta Informaticae*, 56(1,2) :89–103, 2003. Full version of [97].

- 
- [42] **M. Crochemore**, C. S. Iliopoulos, Y. J. Pinzon, et J. Reid. A fast and practical bit-vector algorithms for the longest common subsequence problem. *Information Processing Letters*, 80(6) :279–285, 2001.
- [43] **M. Crochemore**, G. M. Landau, et M. Ziv-Ukelson. A sub-quadratic sequence alignment algorithm for unrestricted cost matrices. *SIAM Journal of Computing*, 32(6) :1654–1673, 2003. Full version of [100].
- [44] **M. Crochemore**, B. Melichar, et Z. Troníček. Directed acyclic subsequence graph - Overview. *Journal of Discrete Algorithms*, 1 :255–280, 2003.
- [45] **M. Crochemore**, F. Mignosi, A. Restivo, et S. Salemi. Data compression using anti-dictionaries. *Proceedings of the I.E.E.E.*, 88(11) :1756–1768, 2000. Special issue *Lossless data compression* edited by J. Storer.
- [46] **M. Crochemore** et V. Stefanov. Waiting time and complexity for matching patterns with automata. *Information Processing Letters*, 87(3) :119–125, 2003.
- [47] A. Dartois, **F. Fiorenzi**, et P. Francini. Sandpile group on the graph  $\mathcal{D}_n$  of the dihedral group. *European J. Combin.*, 24(7) :815–824, 2003.
- [48] **I. Fagnot**. A little more about morphic sturmian words. *Theor. Inform. Appl.*, 2004. À paraître.
- [49] **I. Fagnot** et L. Vuillon. Generalized balances in Sturmian words. *Discrete Appl. Math.*, 121(1-3) :83–101, 2002.
- [50] **F. Fiorenzi**. The Garden of Eden theorem for sofic shifts. *Pure Math. Appl.*, 11(3) :471–484, 2000.
- [51] **F. Fiorenzi**. Cellular automata and strongly irreducible shifts of finite type. *Theoret. Comput. Sci.*, 299(1-3) :477–493, 2003.
- [52] **F. Fiorenzi**. Semi-strongly irreducible shifts. *Adv. in Appl. Math.*, 32(3) :421–438, 2004.
- [53] **R. Forax**, **E. Duris**, et **G. Roussel**. A multi-method design and implementation for Java. *IEEE Transactions on Software Engineering (TSE)*, 2004. À paraître, 35 pages.
- [54] G. Han et **D. Perrin**. Ensembles inévitables. *Séminaire Lotharingien de Combinatoire*, 47, 2002. Article B47e, 16pp.
- [55] **L. Marsan** et **M.-F. Sagot**. Algorithms for extracting structured motifs using a suffix tree with application to promoter and regulatory site consensus identification. *J. of Comput. Biol.*, 7 :345–360, 2001.
- [56] **C. Morvan** et **C. Rispal**. Families of automata characterizing context-sensitive languages. *Acta Inform.*, 2004. À paraître.
- [57] **D. Perrin** et **G. Rindone**. Syntactic groups. *Bulletin of the Belgium Mathematical Society*, 10(5) :749–759, Dec. 2003.
- [58] A. Vanet, **L. Marsan**, A. Labigne, et **M.-F. Sagot**. Inferring regulatory elements from a whole genome. An application to the analysis of genome of *helicobacter pylori*  $\sigma_{80}$  family of promoter signals. *J. Mol. Biol.*, 297 :335–353, 2000.

## Actes de conférences

- [59] S. Akiyama, **F. Bassino**, et C. Frougny. Automata for arithmetic meyer sets. In *LATIN'04*, volume 2976 of *Lecture Notes in Comput. Sci.*, pages 252–261. Springer, 2004.
- [60] **J. Allali** et **M.-F. Sagot**. Novel tree edit operations for RNA secondary structure comparison. In *Proceedings of the 4th Workshop on Algorithms in Bioinformatics*, 2004.
- [61] **C. Allauzen**, **M. Crochemore**, et M. Raffinot. Efficient experimental string matching by weak factor recognition. In A. Amir et G. Landau, editors, *CPM'2001, Combinatorial Pattern Matching (Jerusalem, 2001)*, volume 2089 of *Lecture Notes in Comput. Sci.*, pages 51–72. Springer-Verlag, 2001.
- [62] A. Amir, A. Butman, **M. Crochemore**, G. M. Landau, et M. Schaps. Two-dimensional pattern matching with rotations. In R. Baeza-Yates, E. Chavez, et **M. Crochemore**, editors, *Combinatorial Pattern Matching*, volume 2676 of *Lecture Notes in Comput. Sci.*, pages 17–31. Springer-Verlag, 2003.
- [63] **F. Bassino**. Beta-expansions for cubic Pisot numbers. In *LATIN'02*, volume 2286 of *Lecture Notes in Comput. Sci.*, pages 141–152. Springer, 2002.
- [64] **F. Bassino**, **M.-P. Béal**, et **D. Perrin**. Length distributions and regular sequences. In J. Rosenthal et B. Marcus, editors, *Codes, systems, and graphical models (Minneapolis, MN, 1999)*, volume 123 of *IMA Vol. Math. Appl.*, pages 415–437, New York, 2001. Springer.
- [65] **F. Bassino**, **J. Clément**, et **C. Nicaud**. The average lengths of the factors of the standard factorization of Lyndon words. In M. Ito et M. Toyama, editors, *Developments in Language Theory*, volume 2450 of *Lecture Notes in Comput. Sci.*, pages 307–318. Springer, 2003. 6th International Conference, DLT, Kyoto, Japan, September 18-21, 2002.
- [66] **F. Bassino**, **J. Clément**, et **C. Nicaud**. Lyndon words with a fixed standard right factor. In J. I. Munro, editor, *SODA*, pages 646–647. SIAM, 2004. Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004.
- [67] **M.-P. Béal** et **O. Carton**. Asynchronous sliding block maps. In *Developments in language theory (Aachen, 1999)*, pages 47–59, River Edge, NJ, 2000. World Sci. Publishing.
- [68] **M.-P. Béal** et **O. Carton**. Determinization of transducers over infinite words. In *ICALP 2000*, volume 1853 of *Lecture Notes in Comput. Sci.*, pages 561–570, Berlin, 2000. Springer.
- [69] **M.-P. Béal**, **O. Carton**, C. Prieur, et J. Sakarovitch. Squaring transducers : An efficient procedure for deciding functionality and sequentiality. In G. Gonnet, D. Panario, et A. Viola, editors, *LATIN'2000*, volume 1776 of *Lecture Notes in Comput. Sci.*, 2000.
- [70] **M.-P. Béal**, **F. Fiorenzi**, et **D. Perrin**. A hierarchy of irreducible sofic shifts. In *Symposium on Mathematical Foundations of Computer Science 2004*, volume 3153 of *Lecture Notes in Comput. Sci.*, pages 611–622. Springer-Verlag, 2004.

- 
- [71] **M.-P. Béal**, **F. Fiorenzi**, et **D. Perrin**. The syntactic graph of a sofic shift. In V. Diekert et M. Habib, editors, *Annual Symposium on Theoretical Aspects of Computer Science 2004*, volume 2296 of *Lecture Notes in Comput. Sci.*, pages 282–293. Springer Verlag, 2004.
- [72] **M.-P. Béal** et **D. Perrin**. On the enumerative sequences of regular languages on  $k$  symbols. In H. Alt et A. Ferreira, editors, *Annual Symposium on Theoretical Aspects of Computer Science 2002*, volume 2285 of *Lecture Notes in Comput. Sci.*, pages 547–558, Berlin, 2002. Springer.
- [73] **J. Berstel** et L. Boasson. XML grammars. In *Mathematical Foundations of Computer Science 2000 (Bratislava)*, volume 1893 of *Lecture Notes in Comput. Sci.*, pages 182–191, Berlin, 2000. Springer.
- [74] **J. Berstel**, L. Boasson, **O. Carton**, B. Petazzoni, et J.-É. Pin. Operations preserving recognizable languages. In *Fundamentals of Computation Theory*, volume 2751 of *Lecture Notes in Comput. Sci.*, pages 343–354, Berlin, 2003. Springer.
- [75] **J. Berstel** et **O. Carton**. On the complexity of Hopcroft’s state minimization algorithm. In K. Salomaa, editor, *Conference on Implementation and Application of Automata (CIAA)*, Lecture Notes in Comput. Sci. Springer, 2004. À paraître.
- [76] **J. Berstel**, S. Crespi Reghizzi, P. San Pietro, et **G. Roussel**. A scalable formal method for design and automatic checking of user interfaces. In *International Conference on Software Engineering (ICSE’01)*, pages 453–462. ACM SIGSOFT, May 2001.
- [77] F. Blanchard, **J. Cervelle**, et E. Formenti. Periodicity and transitivity for cellular automata in besicovitch topologies. In B. Rován et P. Vojtás, editors, *Symposium on Mathematical Foundations of Computer Science 2003*, volume 2747 of *Lecture Notes in Comput. Sci.*, pages 228–238. Springer, 2003. 28th International Symposium, MFCS 2003, Bratislava, Slovakia, August 25-29.
- [78] L. Bouguerroua, L. George, et **S. Midonnet**. An adaptative allowance for failure prevention of fixed priority scheduled real-time systems. In *IASTED Int. Conf. on Parallel and Distributed Computing Systems (PDCS’04)*, pages 461–466, MIT Cambridge, USA, Nov. 2004.
- [79] L. Bouguerroua, L. George, et **S. Midonnet**. Task allowance for failure prevention of real-time JAVA systems. In *IASTED Int. Conf. on Parallel and Distributed Computing and Networks (PDCN’04)*, pages 375–380, Innsbruck, Austria, Feb. 2004.
- [80] L. Boukhalfa, L. George, P. Minet, et **S. Midonnet**. Router timeliness analysis in multihop network. In *11th Int. Conf. on Software, Telecommunications and Computer Networks (SoftCOM’03)*, pages 365–369, Split, Croatia, Oct. 2003. IEEE Communications Society.
- [81] L. Boukhalfa, P. Minet, L. George, et **S. Midonnet**. Mobile ad-hoc networks and qos demanding applications. In *5th IEEE Int. Conf. on Mobile and Wireless Communications Networks (MWCN’03)*, pages 11–14, Singapore, Oct. 2003. World Scientific Publisher.

- 
- [82] **P. Bracka**, **S. Midonnet**, et **G. Roussel**. Routage dans un réseau de robots. In *Quatrièmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (ALGOTEL'02)*, pages 163–170, Mèze, France, May 2002.
- [83] **P. Bracka**, **S. Midonnet**, et **G. Roussel**. Scheduling and routing in an ad-hoc network of robots. In S. Sahni, editor, *Proceedings of the International Conference of Computer Science and Technology (CST'03)*, pages 337–342, Cancun, Mexico, May 2003. ACTA Press.
- [84] **O. Carton**. Unambiguous automata on bi-infinite words. In B. Rován et P. Vojtas, editors, *Symposium on Mathematical Foundations of Computer Science 2003*, volume 2747 of *Lecture Notes in Comput. Sci.*, pages 308–317, 2003.
- [85] **O. Carton** et **C. Rispal**. Complementation of rational sets on scattered linear orderings of finite rank. In M. Farach-Colton, editor, *LATIN*, volume 2976 of *Lecture Notes in Comput. Sci.*, pages 292–301. Springer, 2004. LATIN 2004 : Theoretical Informatics, 6th Latin American Symposium, Buenos Aires, Argentina, April 5-8, 2004, Proceedings.
- [86] T. Ceccherini-Silberstein, **F. Fiorenzi**, et F. Scarabotti. The Garden of Eden theorem for cellular automata and for symbolic dynamical systems. In *Random walks and Geometry*, pages 73–108, Berlin, 2004. de Gruyter.
- [87] **J. Cervelle** et B. Durand. Tilings : recursivity and regularity. In H. Reichel et S. Tison, editors, *Annual Symposium on Theoretical Aspects of Computer Science 2000*, volume 1770 of *Lecture Notes in Comput. Sci.*, pages 491–502. Springer, 2000. 17th Annual Symposium on Theoretical Aspects of Computer Science, Lille, France, February 2000, Proceedings.
- [88] **J. Cervelle**, B. Durand, et E. Formenti. Algorithmic information theory and cellular automata dynamics. In J. Sgall, A. Pultr, et P. Kolman, editors, *Symposium on Mathematical Foundations of Computer Science 2001*, volume 2136 of *Lecture Notes in Comput. Sci.*, pages 248–259. Springer, 2001. 26th International Symposium, MFCS 2001 Mariánské Lázně, Czech Republic, August 27-31.
- [89] **J. Cervelle** et E. Formenti. On sand automata. In H. Alt et M. Habib, editors, *Annual Symposium on Theoretical Aspects of Computer Science 2003*, volume 2607 of *Lecture Notes in Comput. Sci.*, pages 642–653. Springer, 2003. 20th Annual Symposium on Theoretical Aspects of Computer Science, Berlin, Germany, February 27 - March 1, 2003.
- [90] **M. Crochemore**, C. Epifanio, R. Grossi, et F. Mignosi. A trie-based approach for compacting automata. In S. C. Sahinalp, S. Muthukrishnan, et U. Dogrusoz, editors, *Combinatorial Pattern Matching*, volume 3109 of *Lecture Notes in Comput. Sci.*, pages 145–158. Springer-Verlag, 2004.
- [91] **M. Crochemore**, R. Giancarlo, et **M.-F. Sagot**. Longest motifs with a functionally equivalent block. In A. Apostolico et M. Melucci, editors, *SPIRE*, Lecture Notes in Comput. Sci. Springer-Verlag, 2004. Report I.G.M. 2004-01.
- [92] **M. Crochemore**, C. S. Iliopoulos, T. Lecroq, et Y. J. Pinzon. Approximate string matching in musical sequences. In M. Balík et M. Šimánek, editors, *PSC'2001, Prague*

- Stringoly Club (Prague, 2001)*, pages 26–36. Czech Technical University of Prague, 2001. DC-2001-06.
- [93] **M. Crochemore**, C. S. Iliopoulos, T. Lecroq, W. Plandowski, et W. Rytter. Three heuristics for *delta*-matching : *delta*-BM algorithms. In A. Apostolico et M. Takeda, editors, *Combinatorial Pattern Matching, CPM'2002*, volume 2373 of *Lecture Notes in Comput. Sci.*, pages 178–189. Springer-Verlag, 2002.
- [94] **M. Crochemore**, C. S. Iliopoulos, M. Mohamed, et **M.-F. Sagot**. Longest repeated motif with a block of don't cares. In M. Farach-Colton, editor, *Latin American Theoretical Informatics (LATIN)*, number 2976 in *Lecture Notes in Comput. Sci.*, pages 271–278. Springer-Verlag, 2004.
- [95] **M. Crochemore**, C. S. Iliopoulos, G. Navarro, et Y. J. Pinzon. A bit-parallel suffix automaton approach for  $(\delta, \gamma)$ -matching in music retrieval. In M. Nascimento, E. de Moura, et A. Oliveira, editors, *SPIRE*, volume 2857 of *Lecture Notes in Comput. Sci.*, pages 211–223. Springer-Verlag, 2003.
- [96] **M. Crochemore**, C. S. Iliopoulos, et Y. J. Pinzon. Fast evolutionary chains. In V. Hlaváč, K. G. Jeffery, et J. Wiedermann, editors, *Sofsem 2000—Theory and Practice of Informatics*, volume 1963 of *Lecture Notes in Comput. Sci.*, pages 306–317. Springer-Verlag, 2000.
- [97] **M. Crochemore**, C. S. Iliopoulos, et Y. J. Pinzon. Speeding-up Hirschberg and Hunt-Szymanski LCS algorithms. In G. Navarro, editor, *SPIRE'2001, 8th International Symposium on String Processing and Information Retrieval (Laguna de San Rafael, Chile, 2001)*, pages 59–67. IEEE Computer Society, 2001.
- [98] **M. Crochemore**, C. S. Iliopoulos, Y. J. Pinzon, et J. Reid. A fast and practical bit-vector algorithms for the longest common subsequence problem. In L. Brankovic et J. Ryan, editors, *Proceedings of the eleventh Australasian Workshop on Combinatorial Algorithms, AWOCA'2000*, pages 75–86. University of Newcastle, NSW, Australia, 2000.
- [99] **M. Crochemore**, C. S. Iliopoulos, Y. J. Pinzon, et W. Rytter. Finding motifs with gaps. In *Proceedings of International Symposium on Music Information Retrieval (ISMIR'00)*, pages 306–317, Plymouth, Massachusetts, 2000.
- [100] **M. Crochemore**, G. M. Landau, et M. Ziv-Ukelson. A sub-quadratic sequence alignment algorithm for unrestricted cost matrices. In D. Eppstein, editor, *Proceedings of the Thirteen Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 679–688. ACM-SIAM, 2002. Rapport I.G.M. 2001-08.
- [101] **M. Crochemore** et G. Navarro. Improved antidictionary based compression. In *SCCC'02, Chilean Computer Science Society*, pages 7–13. I.E.E.E. CS Press, Nov. 2002.
- [102] **M. Crochemore** et Z. Tronicek. On the size of DASG for multiple texts. In A. Laender et A. Oliveira, editors, *String Processing and Information Retrieval, SPIRE'2002*, volume 2476 of *Lecture Notes in Comput. Sci.*, pages 58–64. Springer-Verlag, 2002.
- [103] **C. Deleray**, **N. Bedon**, **G. Roussel**, et **E. Duris**. Corosol une JVM modulaire paramétrable à la volée. In Hermès, editor, *Langages et Modèles à Objets (LMO'04)*,

- volume 10 of *L'objet*, pages 89–102. Revue des Sciences et Technologies de l'Information, Mar. 2004.
- [104] **C. Deleray**, **N. Bedon**, **G. Roussel**, **E. Duris**, et **R. Forax**. MyJVM : a 100% pure Java parameterizable Java virtual machine. In *Proceedings of the International Conference on the Principles and Practice of Programming in Java (PPPJ'03)*, pages 117–119, Kilkenny City, Ireland, June 2003. ACM SIGAPP. Extended Abstract.
- [105] **I. Fagnot**, G. Lelandais, et S. Vialette. Bounded List Injective Homomorphism for Comparative Analysis of Protein-Protein Interaction Graphs. In *Proceedings of CompBioNets 2004*, Dec. 2004. À paraître.
- [106] **R. Forax**, **E. Duris**, et **G. Roussel**. Java multi-method framework. In *International Conference on Technology of Object-Oriented Languages and Systems (TOOLS'00)*, pages 45–56, Sydney, Australia, Nov. 2000. IEEE Computer Society.
- [107] L. George et **S. Midonnet**. A real-time admission control into CORBA audio/video stream architecture. In *9th Int. Conf. on Software, Telecommunications and Computer Networks (SoftCOM'01)*, pages 537–544, Split, Croatia, Oct. 2001. IEEE Communications Society.
- [108] J. Holub et **M. Crochemore**. On the implementation of compact DAWG's. In J.-M. Champarnaud et D. Maurel, editors, *Implementation and Application of Automata*, volume 2608 of *Lecture Notes in Comput. Sci.*, pages 289–294. Springer-Verlag, 2003.
- [109] C. S. Iliopoulos, J. McHugh, **P. Peterlongo**, N. Pisanti, W. Rytter, et M.-F. Sagot. A first approach to finding common motifs with gaps. In M. Simanek et J. Holub, editors, *Prague Stringology Conference 2004 September 2004*, pages 88–97, 2004.
- [110] F. N. Kooh et **S. Midonnet**. CLAGS CORBA-based group communication and monitoring over Internet. In *13th ISCA Int. Conf. on Parallel and Distributed Computing Systems (PDCS'00)*, pages 19–26, Las Vegas, USA, Aug. 2000.
- [111] **S. Midonnet**. Aperiodic invocations admission in real-time CORBA. In *12th Int. Conf. on Software, Telecommunications and Computer Networks (SoftCOM'04)*, pages 141–145, Split, Croatia, Oct. 2004. IEEE Communications Society.
- [112] **S. Midonnet** et H. Zhao. A standard real-time application development using real-time java specification, from theory to practice. In *IASTED Int. Conf. on Software Engineering and Applications (SEA'04)*, pages 571–576, MIT Cambridge, USA, Nov. 2004.
- [113] **E. Moreno**. Lyndon words and de Bruijn sequences in a subshift of finite type. In T. Harju et J. Karhumäki, editors, *Proceedings of WORDS'03*, number 27 in TUCS General Publications, pages 400–410, Turku, Finland, August 2003. Turku Centre for Computer Science.
- [114] **E. Moreno** et M. Matamala. Minimal de Bruijn sequence in a language with forbidden substrings. In *Graph-Theoretic Concepts in Computer Science*, Lecture Notes in Comput. Sci. Springer-Verlag Heidelberg, 2004. À paraître.
- [115] **C. Morvan**. On rational graphs. In J. Tiuryn, editor, Fossacs 00, volume 1784 of *Lecture Notes in Comput. Sci.*, pages 252–266, 2000. ETAPS 2000 best theoretical paper Award.

- [116] **C. Morvan** et C. Stirling. Rational graphs trace context-sensitive languages. In A. Pultr et J. Sgall, editors, *Symposium on Mathematical Foundations of Computer Science 2001*, volume 2136 of *Lecture Notes in Comput. Sci.*, pages 548–559, 2001.
- [117] **M. Nguyen The**. Distribution of the size of simplified or reduced trees. In *Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics, pages 341–354. Birkhäuser, 2002. Versailles.
- [118] **J. Petazzoni**. Deplika : a scalable multicast tool for Unix system replication. In S. Sahni, editor, *Proceedings of the International Conference of Computer Science and Technology (CST'03)*, pages 203–207, Cancun, Mexico, May 2003. ACTA Press.
- [119] N. Pisanti, **M. Crochemore**, R. Grossi, et **M.-F. Sagot**. A basis of tiling motifs for generating repeated patterns and its complexity for higher quorum. In B. Rovan et P. Vojtáš, editors, *Symposium on Mathematical Foundations of Computer Science 2003*, volume 2747 of *Lecture Notes in Comput. Sci.*, pages 622–632. Springer-Verlag, 2003.
- [120] **C. Rispal**. The synchronized graphs trace the context-sensitive languages. In A. Kucera et R. Mayr, editors, *Electronic Notes in Theoretical Computer Science*, volume 68. Elsevier, 2003. INFINITY 2002, 4th International Workshop on Verification of Infinite-State Systems – August 24, 2002 Brno, Czech Republic.
- [121] **C. Rispal** et **O. Carton**. Complementation of rational sets on scattered linear orderings. In *DLT'2004*, Lecture Notes in Comput. Sci. Springer, 2004.
- [122] E. M. Rodrigues, **M.-F. Sagot**, et Y. Wakabayashi. Some approximation results for the maximum agreement forest problem. In M. Goemans, K. Jansen, J. Rolim, et L. Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization : Algorithms and Techniques (APPROX & RANDOM 2001)*, volume 2129 of *Lecture Notes in Comput. Sci.*, pages 159–169. Springer Verlag, 2001. Berkeley, Californie, USA.

## Collections, livres et chapitres dans les livres

- [123] A. Apostolico et **M. Crochemore**. String pattern matching for a deluge survival kit. In J. Abello, P. Pardalos, et M. Resende, editors, *Handbook of Massive Data Sets*, pages 151–194. Kluwer Academic Publishers, 2002.
- [124] R. Baeza-Yates, E. Chávez, et **M. Crochemore**, editors. *Combinatorial Pattern Matching*, volume 2676 of *Lecture Notes in Comput. Sci.* Springer-Verlag, 2003.
- [125] **F. Bassino** et V. Bruyère, editors. *Proceedings of the 8th Journées Montoises d'Informatique*, volume 8. Bulletin of the Belg. Math. Soc., 2001.
- [126] **J. Berstel** et L. Boasson. Balanced grammars and their languages. In W. Brauer, H. Ehring, J. Karhumäki, et A. Salomaa, editors, *Formal and Natural Computing*, volume 2300 of *Lecture Notes in Comput. Sci.*, pages 3–25. Springer-Verlag, 2002. Essays dedicated to G. Rozenberg.
- [127] **J. Berstel** et **D. Perrin**. Finite and infinite words. In **M. Lothaire**, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

- 
- [128] **J. Berstel** et P. Séébold. Sturmian words. In **M. Lothaire**, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
- [129] **M. Crochemore**. Structures for indexes. In **M. Lothaire**, editor, *Applied Combinatorics on Words*. Cambridge University Press, 2004.
- [130] **M. Crochemore** et L. Gaşieniec, editors. *Matching Patterns*. Hermès, 2000. Special issue of *J. Discrete Algorithms*.
- [131] **M. Crochemore**, C. Hancart, et T. Lecroq. *Algorithmique du texte*. Vuibert, 2001. 347 pages.
- [132] **M. Crochemore** et T. Lecroq. Pattern matching and text compression algorithms. In A. B. Tucker, editor, *The Computer Science and Engineering Handbook*, chapter 13, pages 1–48. CRC Press, 2004.
- [133] **M. Crochemore** et W. Rytter. *Jewels of Stringology*. World Scientific Publishing, Hong-Kong, 2002. 310 pages.
- [134] **M. Crochemore** et **M.-F. Sagot**. Motifs in sequences : localization and extraction. In M. Crabbe, M. Drew, et A. Konopka, editors, *Handbook of Computational Chemistry*. Marcel Dekker, Inc., 2002. À paraître, rapport I.G.M. 2000-13.
- [135] **M. Lothaire**. *Algebraic Combinatorics on Words*, volume 90 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002. With a preface by **J. Berstel** and **D. Perrin**.
- [136] **M. Lothaire**. *Applied Combinatorics on Words*. Cambridge University Press, 2004. À paraître, cf <http://igm.univ-mlv.fr/~berstel/Lothaire/>.
- [137] C. Mathé, T. Schiex, P. Rouzé, **P. Blayo**, et **M.-F. Sagot**. Gene finding in eukaryotes. In Q. Lu et M. Weiner, editors, *Cloning and expression technologies*, pages 27–43. Eaton Publishing, 2002.
- [138] **D. Perrin**. Enumerative combinatorics on words. In H. Crapo et G.-C. Rota, editors, *Algebraic Combinatorics and Computer Science*, pages 391–430. Springer Verlag, 2001.
- [139] **D. Perrin**. Automi e linguaggi formali. In S. Petruccioli, editor, *Storia della Scienza*, volume IX, pages 197–205. Istituto della Enciclopedia Italiana, 2003.
- [140] **D. Perrin** et J.-E. Pin. *Infinite Words, Automata, Semigroups, Logic and Games*. Elsevier, 2004.
- [141] J.-L. Risler, **M. Crochemore**, A. K. Konopka, B. Prum, et P. Rouzé, editors. *Genome and Informatics*, volume 26-5. Pergamon, 2002. Special issue of *Computers and Chemistry*.
- [142] **G. Roussel** et **E. Duris**. *Java et Internet : concepts et programmation*. Vuibert Informatique, Paris, Jan. 2000. 576 pages.
- [143] **G. Roussel**, **E. Duris**, **N. Bedon**, et **R. Forax**. *Java et Internet : concepts et programmation, Tome 1 : coté client, 2<sup>e</sup> édition*. Vuibert Informatique, Paris, Nov. 2002. 848 pages.

## Thèses et habilitations

- [144] **C. Allauzen.** *Combinatoires sur les mots et recherche de motifs.* Thèse de doctorat, Université de Marne la Vallée, Jan. 2001. Jury : **J. Berstel**, C. Choffrut, **M. Crochemore**, G. Kutcherov, F. Mignosi et J.-C. Spehner.
- [145] **M.-P. Béal.** *Codage symbolique.* Habilitation à diriger des recherches, Université de Marne la Vallée, Jan. 2001. Jury : J.-P. Allouche, G. Cohen, R. Cori, **M. Crochemore**, C. Frougny, B. Marcus, F. Morain et **D. Perrin**.
- [146] **S. Bentolila.** *La logique et le vivant ; les formalismes de représentation des connaissances en biologie.* Thèse de doctorat, Université de Marne la Vallée, 2002. Jury : **M. Crochemore**, M.-C. Maurel, **C. Rayssiguier**, **M.-F. Sagot** et D. Thieffry.
- [147] **P. Blayo.** *Une approche comparative combinatoire pour la prédiction de gènes chez les eucaryotes.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : **M. Crochemore**, **C. Rayssiguier**, T. Schiex, M. Gouy, **M.-F. Sagot** et P. Rouzé.
- [148] **O. Carton.** *Automates et mots infinis.* Habilitation à diriger des recherches, Université de Marne la Vallée, Dec. 2001. Jury : **J. Berstel**, C. Choffrut, **D. Perrin**, J.-É. Pin, P. Schupp, G. Sénizergues et W. Thomas.
- [149] **R. Forax.** *Les multi-méthodes en Java.* Thèse de doctorat, Université de Marne la Vallée, Dec. 2001. Jury : **J. Berstel**, G. Hedin, C. Queinnec, J. Malenfant, **G. Roussel** et É. Simon.
- [150] J. Grazzini. *Analyse multiéchelle d'images météorologiques : application à la détection des zones précipitantes.* Thèse de doctorat, Université de Marne la Vallée, Dec. 2003. Jury : **J. Berstel**, I. Herlin, F.-X. Le Dimet, E. Memin, J.-P. Nadal, A. Szantai, A. Turiel et H. Yahia.
- [151] **V. Le Maout.** *Expérience de programmation générique sur des structures non-séquentielles : les automates.* Thèse de doctorat, Université de Marne la Vallée, July 2003. Jury : **M. Crochemore**, T. Lecroq, **D. Perrin**, **D. Revuz** et J.-M. Rifflet.
- [152] **L. Marsan.** *Inférence de motifs structurés : algorithmes et outils appliqués à la détection de sites de fixation dans des séquences génomiques.* Thèse de doctorat, Université de Marne la Vallée, 2002. Jury : S. Dulucq, A. Viari, R. Grossi, **J. Berstel**, **M.-F. Sagot**, **M. Crochemore** et A. Vanet.
- [153] **C. Martineau.** *Compression de textes en langue naturelle.* Thèse de doctorat, Université de Marne la Vallée, Dec. 2001. Jury : **M. Crochemore**, É. Laporte, D. Maurel, G. Plateau, S. Tomi Klein et **M. Zipstein**.
- [154] **C. Rispal.** *Automates sur les ordres linéaires : Complémentation.* Thèse de doctorat, Université de Marne la Vallée, Dec. 2004. Jury : **J. Berstel**, V. Bruyère, D. Caucal, **O. Carton**, **D. Perrin**, J.-É. Pin.
- [155] **G. Roussel.** *Grammaires et automates comme outils pour le développement logiciel.* Habilitation à diriger des recherches, Université de Marne-la-Vallée, Dec. 2003. Jury : **J. Berstel**, I. Attali, S. Crespi-Reghezzi, B. Lorho et P. Minet.

---

**Divers**

- [156] **J. Allali** et **M.-F. Sagot**. The at-most  $k$ -deep factor tree. Rapport 2004-03, IGM, Université de Marne la Vallée, 2004.
- [157] **J. Clément**, M. Régnier, et M. Vandenbogaert. Combinatorial algorithms for approximate words, 2004. (cf <http://igm.univ-mlv.fr/~clementj/publications>).
- [158] **C. Deleray**, **N. Bedon**, **G. Roussel**, et **E. Duris**. Corosol : a component-based and reflexive JVM dynamically customizable. Technical report, I.G.M., Université de Marne-la-Vallée, 2004.
- [159] **G. Fici**, F. Mignosi, A. Restivo, et M. Sciortino. Fragment assembly through minimal forbidden words. Preproceedings of X Journées Montoises d'informatique théorique, Liège, 8-11 Sept., 2004.
- [160] **R. Forax**, **E. Duris**, et **G. Roussel**. Java multi-method framework. Rapport de Recherche 2000-10, I.G.M., Université de Marne-la-Vallée, Oct. 2000. 15 pages.
- [161] **R. Forax**, **E. Duris**, et **G. Roussel**. A simple dispatch technique for pure Java multi-methods. Rapport de Recherche 2001-02, I.G.M., Université de Marne-la-Vallée, Feb. 2001. 27 pages.
- [162] N. Pisanti, **M. Crochemore**, R. Grossi, et **M.-F. Sagot**. A basis for repeated motifs in pattern discovery and text mining. Technical report, Institut Gaspard-Monge, 2002. Rapport I.G.M. 2002-10.



# 3

## Combinatoire algébrique et calcul symbolique

### 3.1 Composition

#### Membres

- Jacques Désarménien, professeur
- Florent Hivert, maître de conférences
- Alain Lascoux, directeur de recherche CNRS
- Michel Lassalle, directeur de recherche CNRS
- Jean-Gabriel Luque, maître de conférences
- Jean-Christophe Novelli, chargé de recherche CNRS
- Pierre-André Picon, professeur
- Jean-Yves Thibon, professeur, responsable

#### Membres associés

- Gérard Duchamp, professeur, Université Paris XIII
- Karol Penson, professeur, Université Paris VI
- Nicolas Thiéry, maître de conférences, Université Lyon I
- Frédéric Toumazet, maître de conférences, Université Paris XIII

#### Doctorants et post-doctorants

- François Descouens, doctorant
- Lenin Arcadio García de León Rumazo, doctorant
- Michael Jöllenbeck, doctorant
- Klaas Slooten, post-doctorant
- Carmelo Vaccaro, doctorant

**Anciens membres**

- Emmanuel Briand, docteur, ATER en 2002–2003

## 3.2 Thèmes de recherche

La Combinatoire Algébrique s'intéresse aux aspects algorithmiques et combinatoires de diverses théories algébriques, par exemple, la théorie des représentations des groupes. Les applications de cette dernière étant innombrables, les problèmes intéressants ne manquent pas.

L'outil traditionnel dans ce contexte est la théorie des fonctions symétriques, et l'exemple classique est la théorie combinatoire des fonctions de Schur (les caractères des groupes unitaires), qui s'exprime en termes de tableaux de Young, et qui a donné naissance à l'algorithme de Robinson-Schensted-Knuth et au monoïde plaxique.

Nos travaux de ces dernières années ont essentiellement porté sur des généralisations des notions précédentes. Par exemple, les fonctions de Schur font partie d'une famille à deux paramètres (les fonctions de Macdonald), qui font actuellement l'objet de recherches intensives auxquelles nous avons contribué. Une autre voie de généralisation consiste à étendre les méthodes issues de la théorie des fonctions symétriques à des familles de polynômes ne possédant plus que des symétries partielles (les polynômes de Schubert, les fonctions quasi-symétriques), ou encore, non commutatifs (les fonctions symétriques non commutatives, les fonctions quasi-symétriques libres). Ceci débouche sur un autre thème, actuellement très en vogue, celui des Algèbres de Hopf Combinatoires.

Si l'idée d'exploiter la riche structure des algèbres de Hopf en combinatoire n'est pas nouvelle (elle remonte au moins aux travaux de Rota et de son école dans les années soixante-dix), c'est seulement récemment que l'on a vu apparaître les premiers exemples véritablement exploitables (algèbres de permutations, d'arbres, de tableaux). Il arrive fréquemment que la même algèbre apparaisse dans des contextes très différents. C'est le cas notamment de l'algèbre des arbres binaires de Loday-Ronco, à l'origine issue de la théorie des opérades (mathématiques pures), retrouvée en physique théorique (renormalisation de l'électrodynamique quantique), et en algorithmique par notre équipe (le monoïde sylvestre, qui est aux arbres binaires de recherche ce que le monoïde plaxique est aux tableaux). Nous avons découvert de nouveaux exemples d'algèbres de Hopf combinatoires, et dans la plupart des cas, montré comment les réaliser au moyen de polynômes non commutatifs ordinaires.

La théorie des représentations des groupes est issue de la théorie classique des invariants, qui connaît actuellement un regain d'intérêt, en particulier de par ses applications à la théorie de l'information quantique. Nous avons également apporté quelques contributions à ces questions.

Tous les travaux esquissés ci-dessus s'appuient sur une activité expérimentale in-

tense, laquelle requiert une très grande puissance de calcul, et des logiciels spécialisés - en particulier la bibliothèque Combinat de MuPAD - dont nous assurons aussi le développement.

Nous espérons disposer, dans un avenir assez proche, d'une théorie unifiée des algèbres de Hopf combinatoires, qui permettrait d'en découvrir les applications de manière plus systématique, et en autoriserait un traitement informatique plus simple et plus efficace. Nous cherchons également à obtenir pour ces algèbres (dont la structure rappelle toujours un peu celle des fonctions symétriques) des analogues de la théorie de Macdonald, ce qui déboucherait sur des applications importantes à la théorie classique, ainsi que des interprétations en termes de représentations de tours d'algèbres.

Nous continuerons l'étude des polynômes de Macdonald par d'autres méthodes, depuis les techniques d'interpolation ou de symétrisation, jusqu'à la combinatoire des tableaux de rubans.

Nous comptons renforcer l'axe « Information Quantique », en nous attaquant en particulier au problème des mesures d'intrication.

Les activités de développement logiciel seront principalement orientées vers l'étude expérimentale des représentations des algèbres associatives (et plus seulement des groupes).

## 3.3 Résultats

### 3.3.1 Fonctions symétriques non commutatives, fonctions quasi-symétriques

Dans [164], J.-C. Novelli et J.-Y. Thibon, en collaboration avec des chercheurs du LIAFA, décrivent des algorithmes combinatoires pour le calcul du pléthysme des fonctions quasi-symétriques, introduit par C. Malvenuto et C. Reutenauer.

J.-G. Luque et J.-Y. Thibon ont mis en évidence un lien entre la construction des vecteurs de Witt et les factorisations de Lazard du monoïde libre, ceci grâce à une notion de spécialisation des fonctions symétriques non commutatives associée à un code [230].

Dans [223], F. Hivert et N. Thiéry ont introduit une  $q$ -déformation non commutative de l'algèbre des fonctions symétriques qui se spécialise, pour  $q = 1$ , à l'algèbre de Steenrod rationnelle. Ils l'utilisent pour obtenir des résultats partiels sur la conjecture de Wood, concernant la description de l'anneau des co-invariants de l'action de cette algèbre sur les polynômes, et proposent de nouvelles conjectures pour sa  $q$ -déformation.

### 3.3.2 Algèbres de Hopf combinatoires

Dans [217, 170], G. Duchamp, F. Hivert et J.-Y. Thibon introduisent l'algèbre des *fonctions quasi-symétriques libres*, une algèbre de polynômes non commutatifs en une infinité de variables, qui se révèle isomorphe à l'algèbre des permutations de Malvenuto-Reutenauer. La possibilité de réaliser ainsi cette algèbre en simplifie considérablement la théorie, et permet d'en donner de nouvelles applications. On trouve à l'intérieur une sous-algèbre isomorphe à l'algèbre des tableaux standards de Poirier-Reutenauer, ce qui permet de donner une preuve extrêmement simple de la règle de Littlewood-Richardson, et de comprendre les similitudes entre l'algèbre des tableaux et l'algèbre des arbres binaires de Loday et Ronco. L'article [170] contient également des résultats fins sur les représentations des algèbres de Hecke à  $q = 0$ , en particulier la détermination explicite de leurs carquois.

Poursuivant l'étude de l'algèbre de Loday-Ronco, F. Hivert, J.-C. Novelli et J.-Y. Thibon ont montré qu'elle admettait une construction similaire à celle de l'algèbre des tableaux, qui se définit à l'aide de la correspondance de Robinson-Schensted ou du monoïde plaxique. Ils ont découvert le *monoïde sylvestre*, un analogue du monoïde plaxique associé à un analogue de la correspondance de Robinson-Schensted, dans laquelle l'image d'un mot est un couple formé d'un arbre binaire de recherche et d'un arbre tournoi de même forme (au lieu de deux tableaux de même forme, dont un standard) [180]. Ils ont ensuite appliqué ce résultat à la détermination des analogues des bases classiques de fonctions symétriques dans l'algèbre des arbres, et au calcul de leurs produits scalaires [181, 222, 178].

J.-C. Novelli et J.-Y. Thibon ont montré qu'on pouvait construire une algèbre de Hopf ayant pour base les fonctions de parking, et contenant celle de Malvenuto-Reutenauer. Ils en ont donné une réalisation en termes de bimots, et en ont identifié des sous-algèbres de Hopf basées respectivement sur les arbres binaires plans, et sur les arbres plans.

J.-C. Novelli, J.-Y. Thibon et N. Thiéry ont montré comment construire de nombreuses algèbres de Hopf commutatives basées sur divers types de graphes [206].

J.-C. Novelli et J.-Y. Thibon ont introduit de nouvelles algèbres de descentes pour les produits en couronnes, et des analogues des fonctions symétriques libres associées aux permutations colorées [259].

Dans [163], F. Hivert et J.-Y. Thibon, en collaboration avec N. Bergeron (de Toronto), montrent que l'*algèbre des pics* de Stembridge (une certaine sous-algèbre de Hopf des fonctions quasi-symétriques) est l'anneau de Grothendieck d'une certaine spécialisation de la tour des algèbres de Hecke-Clifford (dues à G. Olshanskii). Ils montrent également comment on peut obtenir simplement toutes les propriétés essentielles de l'algèbre des pics et de sa duale au moyen de la théorie des fonctions symétriques non-commutatives.

### 3.3.3 Généralisations du monoïde plaxique et de Robinson-Schensted-Knuth

F. Hivert et J.-C. Novelli (en collaboration avec J. Cassaigne, M. Espie et D. Krob) ont étudié en détail le *monoïde chinois*, un monoïde ressemblant au monoïde plaxique en ceci qu'il est présenté par des relations monomiales d'ordre 3 et admet les tableaux comme section [167].

F. Hivert, J.-C. Novelli et J.-Y. Thibon ont découvert un analogue du monoïde plaxique dans lequel les arbres binaires plans jouent le rôle des partitions, les tableaux semi-standards et standards étant remplacés par les arbres binaires de recherche et les arbres tournois (cf. section précédente).

### 3.3.4 Polynômes de Jack et de Macdonald

Les polynômes de Macdonald forment une base de l'algèbre des fonctions symétriques à coefficients rationnels en deux paramètres  $q, t$ . Ils généralisent la plupart des bases classiques de fonctions symétriques, qu'on obtient pour divers cas particuliers de  $q, t$  : fonctions monomiales, élémentaires, complètes, de Schur, de Hall - Littlewood, de Jack.

Cependant, contrairement au cas des fonctions de Schur et des polynômes de Hall - Littlewood, on ne disposait jusqu'ici d'aucune formule analytique explicite pour les polynômes de Jack et de Macdonald. On ignorait notamment leur développement sur les bases classiques, sauf pour quelques rares cas particuliers.

Dans [202] M. Lassalle a d'abord donné une solution à ce problème pour les polynômes de Macdonald qui sont indexés par une partition de longueur trois, ou dont les parts sont  $(1, 2, 3)$ .

Il a pu ensuite résoudre ce problème en toute généralité en collaboration avec Michael Schlosser (Vienne). Ils ont notamment obtenu le développement analytique explicite de tout polynôme de Macdonald sur les fonctions symétriques élémentaires.

Ils utilisent une méthode générale permettant d'inverser des matrices infinies indexées par des multi-entiers. Ils appliquent cette méthode à la matrice de Pieri. Cette matrice de changement de base est bien connue, elle a été calculée par Macdonald. Cependant on ignorait que son inverse puisse être explicitement calculé.

Ce résultat produit des conséquences intéressantes, notamment le développement analytique explicite des polynômes de Hall - Littlewood sur les fonctions élémentaires. En particulier, il donne le développement analytique explicite des fonctions monomiales sur les fonctions élémentaires, un problème qui remonte aux pères fondateurs de la théorie des fonctions symétriques (Vandermonde, Waring) [203, 261, 260].

F. Hivert, A. Lascoux et J.-Y. Thibon ont trouvé des analogues non commutatifs et quasi-symétriques de polynômes de Macdonald.

La notion de graphe de Yang-Baxter permet de donner une présentation uniforme des différentes versions des polynômes de Jack et Macdonald (symétriques ou non, ho-

mogènes ou non) et de généraliser les éléments de Cherednik [175]. Une autre approche s'appuie sur des décompositions de l'espace des tableaux de Young [186].

### 3.3.5 Algèbres de Hecke affines

J.-Y. Thibon, en collaboration avec B. Leclerc (Caen) et M. Nazarov (York) a obtenu la condition nécessaire et suffisante pour que le produit d'induction de deux modules d'évaluation des algèbres de Hecke affines de type A soit irréductible [232]. Ceci revient à analyser les singularités de certaines  $R$ -matrices.

F. Hivert, J.-C. Novelli et J.-Y. Thibon ont trouvé comment faire dégénérer à  $q = 0$  les algèbres d'Ariki-Koike, et ont montré que l'anneau de Grothendieck de la tour d'algèbres obtenue était isomorphe à l'algèbre de Mantaci-Reutenauer [250].

### 3.3.6 Tableaux de rubans

J.-Y. Thibon, en collaboration avec B. Leclerc (Caen) a montré que les  $q$ -analogues des coefficients de Littlewood-Richardson obtenus au moyen des tableaux de rubans étaient en fait des polynômes de Kazhdan-Lusztig paraboliques du groupe symétrique affine. Ceci entraîne, grâce à des travaux plus récents de Kashiwara, qu'ils sont à coefficients entiers positifs [233].

F. Descouens a obtenu un nouvel algorithme de génération des tableaux de rubans, permettant de calculer plus efficacement les polynômes de spin [216].

### 3.3.7 Théorie des invariants et information quantique

Pour les besoins de la théorie de l'information quantique, il est utile de classifier les états de systèmes quantiques finis, en particulier, les systèmes de qubits ou de qutrits, sous l'action de diverses opérations. Nous avons étudié l'action des opérations de filtrage locales sur les systèmes de 4 qubits et de 3 qutrits, qui sont les premiers cas vraiment difficiles. J.-G. Luque et J.-Y. Thibon ont tout d'abord obtenu un système complet d'invariants polynomiaux de 4 qubits [200], puis en collaboration avec E. Briand, un système complet de 170 covariants (au sens de la théorie classique des invariants), résolvant ainsi un problème dont l'étude avait été amorcée dès les années 1880, par l'astronome belge C. Le Paige [165]. Ils ont ensuite, avec la collaboration du physicien F. Verstraete, décrit géométriquement les orbites des opérations de filtrage local sur les systèmes de 3 qutrits. Cette description fait intervenir les polyèdres complexes réguliers de Shephard et la théorie des groupes de réflexions complexes [166].

### 3.3.8 Polynômes de Schubert et de Grothendieck

L'anneau des polynômes en  $n$  variables admet différentes bases linéaires motivées par la cohomologie (polynômes de Schubert) ou la  $K$ -théorie (polynômes de Grothendieck) des variétés de drapeaux, ou bien encore, intervenant comme caractères de représentations (polynômes clefs, caractères de Demazure). Les transitions entre polynômes de Grothendieck sont décrites dans [225]. Une interprétation des polynômes de Grothendieck en terme du modèle de la glace carrée est obtenue dans [252].

Les polynômes de Schubert pour les groupes de Weyl de type  $B$  et  $D$  sont considérés dans [194, 231].

Une formule de Cauchy pour les caractères de Demazure est démontrée dans [226].

L'étude plus géométrique des variétés de Schubert, en particulier la détermination de leur lieu singulier, est traité dans [184].

### 3.3.9 Algèbres de Lie libres

Dans [172], Jean-Gabriel Luque en collaboration avec G. Duchamp et E. Laugerotte donnent une généralisation du théorème du support de l'algèbre de Lie libre à certaines algèbres de Lie partiellement commutatives libres qu'ils caractérisent par leurs graphes de commutation.

Dans [219], Jean-Gabriel Luque en collaboration avec G. Duchamp caractérisent les congruences du monoïde libre qui sont compatibles avec le coproduit du produit de shuffle. Cette caractérisation dépend du semi-anneau des coefficients. Dans le cas de la caractéristique 0 ou bien lorsque ce dernier n'est pas un anneau, les seules congruences compatibles sont les commutations. Dans le cas des anneaux de caractéristique première, d'autres exemples intéressants apparaissent et donnent lieu à une classification.

### 3.3.10 Théorie des automates

Dans [173], Jean-Gabriel Luque en collaboration avec G. Duchamp, M. Flouret et E. Laugerotte explicitent la construction des produits de shuffle, d'infiltration et de Hadamard sur les automates à multiplicités. Ces constructions sont dérivées de la définition des coproduits associés. Ils caractérisent les lois produits interpolant entre le shuffle et l'infiltration et dont le coproduit reste coassociatif. Ils donnent la construction associée sur les automates.

### 3.3.11 Combinatoire classique

Notre équipe continue à traiter des thèmes de la combinatoire classique, qui permettent d'intégrer plus aisément des étudiants dans un travail de recherche : partitions [188], chemins de Motzkin [192],  $q$ -identités et séries hypergéométriques [191, 168,

174, 247, 248], polynômes orthogonaux [182], algorithme euclidien [193, 190], fonctions symétriques [189, 228, 227].

En étudiant les polynômes de Jack, M. Lassalle a été conduit à plusieurs conjectures qui se formulent comme des identités de la théorie classique des partitions. Il a démontré une de ces conjectures dans [201] en utilisant la théorie des fonctions symétriques. M. Lassalle et A. Lascoux démontrent une deuxième conjecture [189], en utilisant les méthodes remarquablement efficaces de la théorie des  $\lambda$ -anneaux.

Le problème suivant fait partie des classiques de la théorie des fonctions symétriques. Si  $f$  est une fonction symétrique et  $q$  une indéterminée, quelle est la valeur de la spécialisation  $f(1, q, q^2, \dots, q^{N-1})$ ? Ceci est connu depuis longtemps pour les fonctions élémentaires et complètes, et même les fonctions de Schur, mais pas pour les fonctions monomiales.

Dans [200] M. Lassalle a explicité la spécialisation des fonctions monomiales sur « l'alphabet »  $(a-b)/(1-q)$ . Le cas particulier  $a=1$  et  $b=q^N$  correspond à la valeur de la fonction monomiale au point  $(1, q, q^2, \dots, q^{N-1})$ , et le cas particulier  $a=1$  et  $b=0$  à sa valeur en la suite infinie  $(1, q, q^2, q^3, \dots)$ .

Il a pu donner deux formulations distinctes pour cette spécialisation, ce qui était totalement inattendu. À noter que ces résultats sont énoncés et démontrés dans le cadre de la théorie des  $\lambda$ -anneaux, ce qui produit un nouvel exemple de son efficacité.

Dans [204] M. Lassalle a pu introduire une famille de nombres entiers, totalement nouvelle, définis en modifiant légèrement la formule classique de Chu-Vandermonde.

Dans [198] ce résultat lui a permis de démontrer certaines conjectures qu'il avait formulées il y a 5 ans à propos des polynômes de Jack, et de dégager ainsi des aperçus prometteurs sur la structure combinatoire de ces polynômes. La technique de l'interpolation de Lagrange est pour cela essentielle.

Il a pu aussi présenter une nouvelle méthode, élémentaire et très rapide, pour calculer les polynômes de Jack (et en particulier les caractères du groupe symétrique). Cette méthode est implémentée sur ordinateur. Les résultats sont disponibles sur la page <http://igm.univ-mlv.fr/~lassalle/jack.html>.

Il faut souligner que l'intérêt de ces travaux n'est pas restreint à la combinatoire. Ils ont également d'importantes conséquences en théorie des probabilités. Jason Fulman les a notamment utilisés dans son récent article « Martingales and character ratios » (disponible à <http://arXiv.org/abs/math.RT/0402409>).

### 3.3.12 Calculs dans le centre de l'algèbre du groupe symétrique

On sait que le centre de l'algèbre du groupe symétrique  $S_n$  a pour dimension le nombre de partitions de  $n$ , et il est classique de l'identifier à l'espace vectoriel des fonctions symétriques homogènes de degré  $n$ . Goulden avait obtenu, en 1994, un opérateur

différentiel sur les fonctions symétriques réalisant la multiplication par la classe des transpositions pour tout  $n$ . Récemment, Frenkel et Wang avaient montré que les commutateurs de cet opérateur avec les sommes de puissances et les opérateurs différentiels associés, engendraient une représentation de l'algèbre de Virasoro. A. Lascoux et J.-Y. Thibon ont montré que plus généralement, on pouvait décrire explicitement les opérateurs différentiels implémentant la multiplication par les sommes de puissances des éléments de Jucys-Murphy, et qu'on obtenait alors par le procédé de Frenkel et Wang une représentation de l'algèbre  $W_{1+\infty}$ , l'extension centrale universelle de l'algèbre de Lie des opérateurs différentiels sur le cercle. Leurs calculs leur ont également permis de retrouver les opérateurs de Goupil-Poulhalon-Schaeffer au moyen d'intégrales matricielles, naturelles dans ce contexte [197].

### 3.3.13 Combinatoire énumérative

J.-Y. Thibon a obtenu la série indicatrice de cycles de l'ensemble des permutations unimodales, et en a donné un  $q$ -analogue, ainsi que quelques applications [206].

### 3.3.14 Hyperdéterminants, hyperpfaffiens et intégrales multiples

J.-G. Luque et J.-Y. Thibon ont observé que les célèbres formules de N.G. de Bruijn pour les intégrales multiples revenaient à calculer des déterminants ou des pfaffiens dans une algèbre de mélange. Cette remarque les a conduits à des généralisations, faisant intervenir des analogues multidimensionnels du déterminant ou du pfaffien [199]. Ils ont alors réalisé que les intégrales de Selberg et de Kaneko s'exprimaient comme des hyperdéterminants de Hankel, dont ils ont fait une étude systématique [201], permettant de simplifier ou de généraliser de nombreux résultats sur les polynômes orthogonaux, en particulier les calculs de Turániens de Karlin et Szegö. Ils ont ensuite donné une preuve purement algébrique du résultat de Selberg, reposant uniquement sur des manipulations d'hyperdéterminants [207].

## Pseudo-Permutations

Dans [215], F. Boulier, F. Hivert, D. Krob, et J.-C. Novelli ont étudié les pseudo-permutations du point de vue des applications en géométrie et en théorie des représentations.

## Multi-zêtas

Dans [220], M. Espie, J.-C. Novelli et G. Racinet ont calculé expérimentalement les dimensions de l'algèbre de Lie graduée  $DMR_0$  introduite par Racinet dans sa thèse.

Ceci a démontré jusqu'en poids 19 la conjecture de Zagier sur les dimensions de l'algèbre des multi-zêtas formelles.

### 3.3.15 Applications diverses

Dans [169], J.-Y. Thibon, en collaboration avec J.-L. Dornestteter (Nortel), D. Krob et E. Vassilieva, décrit un algorithme numérique stable pour l'évaluation du taux d'erreurs par bit dans un certain type de canal. Ce calcul est nécessaire pour effectuer des simulations numériques. L'algorithme présenté a été obtenu au moyen de la théorie des fonctions de Schur.

### 3.3.16 Logiciels

Depuis le printemps 2001, F. Hivert et N. Thiéry coordonnent le développement de `MuPAD-Combinat` [179, 249]. Il s'agit d'une plate-forme d'expérimentation pour la recherche en combinatoire algébrique. Plus précisément, `MuPAD-Combinat` est une bibliothèque libre (sous licence LGPL) pour le système de calcul formel MuPAD. Elle contient actuellement des fonctions pour manipuler les classes d'objets combinatoires usuelles (partitions, tableaux, arbres, graphes, classes décomposables définies par des grammaires, etc), les polynômes de Schubert, les groupes symétriques et sous-groupes ainsi que les automates à poids. Nous donnons aussi des outils pour construire de nouveaux objets algébriques et comme application nous donnons, entre autres, des implantations des fonctions symétriques et de leurs généralisations comme l'algèbre des arbres binaires. Notons que les applications ne se limitent pas à la combinatoire algébrique mais s'étendent également à la complexité et à la vérification automatique de programmes.

Le code représente actuellement environ 80 000 lignes de code MuPAD, 3 000 lignes de code C++. Il inclut 50 000 lignes de code C/C++ externe ainsi que quelques 450 pages de documentation. Outre les trois développeurs principaux N. Thiéry, F. Hivert, T. Gomez-Diaz, nous avons une vingtaine de contributeurs dont les principaux sont É. Laugerotte (maître de conférences, LIFAR, Université de Rouen), J.-C. Novelli (chargé de recherches CNRS, IGM, Université de Marne-la-Vallée), F. Descouens (doctorant, IGM, Université de Marne-La-Vallée), X. Molinero (professeur, Universitat Politècnica de Catalunya, Espagne) et H. Abbad (doctorante, Djillali Liabes, Université de Sidi Bel Abbes, Algérie). Le cœur de `MuPAD-Combinat` est inclus dans la bibliothèque officielle de MuPAD depuis la version 2.5.0.

La bibliothèque et toute la documentation en ligne sont librement accessibles sur <http://mupad-combinat.sourceforge.net/>. Une introduction rapide se trouve également dans [182]. Précisons enfin que du 28 juin au 3 juillet 2004 ont eu lieu à Rouen les premières rencontres des développeurs de `MuPAD-Combinat` qui ont réuni une vingtaine de personnes.

## 3.4 Activités

### Changements dans la composition de l'équipe

Jean-Christophe Novelli, CR CNRS, a été rattaché à notre équipe en 2001. Il a passé son habilitation le 20 décembre 2001. Pour l'année 2002-2003, Jean-Christophe Novelli a été mis à disposition du laboratoire franco-russe CNRS de Moscou (laboratoire commun avec l'Université Indépendante de Moscou) pour travailler avec des collègues russes en combinatoire algébrique.

En 2003-2004, Florent Hivert a succédé à Jean-Christophe Novelli au laboratoire franco-russe CNRS de Moscou.

Jean-Gabriel Luque, précédemment ATER, est resté avec nous : il a été recruté comme Maître de Conférences à l'IUT de Champs en 2001.

Michel Lassalle, DR CNRS a été rattaché à partir de Juillet 2002.

Emmanuel Briand, ATER, a été rattaché de Septembre 2002 à Septembre 2003. Il a soutenu sa thèse en Octobre 2002.

Nicolas Thiéry, Maître de Conférences à Lyon, Gérard Duchamp, Professeur à Paris 13, Karol Penson, Professeur à Paris 6, et Frédéric Toumazet, Maître de Conférences à Paris 13, ont été intégrés à l'équipe comme associés.

Trois doctorants ont été intégrés : Lenin Arcadio García de León Rumazo (Octobre 2003, co-tutelle avec Paris 7), François Descouens et Carmelo Vaccaro (Octobre 2004).

Klaas Slooten (post-doc du réseau ACE, d'Amsterdam) a été intégré en Novembre 2003 pour une durée de 14 mois.

Michael Jöllenbeck (doctorant du réseau ACE) effectue un séjour de 5 mois (Août-Décembre 2004) dans notre équipe.

#### 3.4.1 Contrats

Notre équipe fait partie du réseau européen ACE (Algebraic Combinatorics in Europe).

Nous avons participé à un projet MathStic avec Paris 7 et Caen.

#### 3.4.2 Diffusion

##### Travaux éditoriaux et organisation de colloques

- Alain Lascoux est éditeur de :
  - Advances in Mathematics ;
  - Annals of Combinatorics.
- Jean-Yves Thibon est éditeur de :
  - Journal of Algebra ;
  - Séminaire Lotharingien de Combinatoire.

- Jean-Yves Thibon a été en résidence au Isaac Newton Institute for Mathematical Sciences (Université de Cambridge) du 1 Avril au 7 Juillet 2001, dans le cadre d'un programme semestriel intitulé « *Symmetric functions and Macdonald polynomials* ».  
Il y a organisé un colloque d'une semaine sur le thème « *Applications of the Macdonald polynomials* ».
- Jacques Désarménien et Jean-Yves Thibon ont organisé la 48ème session du *Séminaire Lotharingien de Combinatoire*, du 10 au 13 Mars 2002.
- Jean-Christophe Novelli et Gleb Koshevoy ont organisé un colloque d'une semaine sur le thème « *Combinatorics, Mathematics, and Physics* » du 12 au 17 mai 2002 à Moscou.
- Jacques Désarménien, Jean-Gabriel Luque et Jean-Yves Thibon ont organisé les 50ème et 52ème sessions (23-27 Mars 2003, 28-31 Mars 2004) du *Séminaire Lotharingien de Combinatoire*.
- Jean-Yves Thibon a été membre du comité d'organisation du 24ème colloque « *Group theoretical methods in physics* », du 15 au 20 Juillet 2002 à Paris. Il y a été responsable du mini-symposium et des sessions parallèles sur les groupes quantiques.

## Conférences invitées, cours à l'étranger

Jean-Yves Thibon a été conférencier invité au colloque « *The heritage of Schur's 1901 dissertation* » (The University of Wales, Juin 2001).

Alain Lascoux, Michel Lassalle et Jean-Yves Thibon ont été conférenciers invités au colloque « *Jack and Macdonald polynomials* », ICMS, Edimbourg, Septembre 2003.

Alain Lascoux a également séjourné au Newton Institute en Avril-Mai 2001.

Il a donné de nombreuses conférences :

- Dix heures de conférences sur les « *Opérateurs combinatoires sur les polynômes* », AMS-CBMS Conference, Durham, USA (juin 2001). Un volume de 270 pages a été publié par l'AMS dans la série AMS-CBMS.
- 3 exposés sur les « *Polynômes de Schubert* » à la réunion du Groupe de Tresses, Lacanau-Bordeaux sept 02.
- Cours réguliers à l'Université Nankai de Tianjin (Chine), par périodes de deux mois, deux fois par an.

Le premier cours a eu lieu en octobre-novembre 2001 : Fonctions symétriques. Puis, les années suivantes : Représentations du groupe symétrique ; Polynômes de Schubert ; Algorithme d'Euclide et Fractions continues ; Polynômes orthogonaux ; Tableaux de Young.

- Un minicours à l'École d'été « Combinatoire des groupes et des algèbres » au C.I.R.M. (Marseille-Luminy, 5-16 juillet 2004), sur : Opérateurs combinatoires sur les polynômes.

---

Florent Hivert a été conférencier invité au colloque « *Words 2003* » (University of Turku, Finland 2003) et aussi au colloque « *Opérades 2004* » à Strasbourg, France.

## Groupe de travail

Notre équipe organise son propre groupe de travail hebdomadaire, le vendredi de 10h30 à 12h30.

### 3.4.3 Collaborations

Notre équipe a écrit des articles en collaboration avec les chercheurs suivants :

- Nantel Bergeron, sur le thème des algèbres de Hopf combinatoires,
- Gleb Koshevoy, sur des généralisations du monoïde plaxique et de l'algorithme de Robinson-Schensted-Knuth,
- Daniel Krob, notamment sur des applications de la combinatoire à l'évaluation de taux d'erreurs de transmission pour la téléphonie,
- Piotr Pragacz, sur les variétés de drapeaux pour les groupes de Weyl classiques,
- Frank Verstraete, sur le thème de l'information quantique.

Elle collabore aussi avec le groupe MuPAD, et en particulier avec SciFace GmbH, pour le développement de la bibliothèque MuPAD-Combinat.

### 3.4.4 Activités doctorales

Jacques Désarménien, Alain Lascoux et Jean-Yves Thibon ont enseigné dans le DEA « *Informatique Fondamentale et Applications* ».

Deux stages de DEA (François Descouens et Carmelo Vaccaro).

Alain Lascoux a écrit 8 articles avec ses étudiants à Tianjin.

### 3.4.5 Thèses et habilitations

#### Habilitations

- Florent Hivert, sous le titre *Combinatoire et calcul symbolique dans les algèbres de Hopf* à l'Université de Marne-la-Vallée, le 13 décembre 2004.
- Jean-Christophe Novelli, sous le titre *Applications de la combinatoire bijective et algébrique*, à l'Université de Marne-la-Vallée, le 21 décembre 2001.

### Thèses en cours

- François Descouens, depuis octobre 2004, *Tableaux de rubans et algèbre de Virasoro* sous la direction de J.-Y. Thibon.
- Lenin Arcadio García de León Rumazo, depuis décembre 2003, *Théorie de invariants et information quantique* sous la direction de J.-Y. Thibon et J.-P. Gazeau.

## 3.5 Références bibliographiques

### Articles de revues

- [163] N. Bergeron, **F. Hivert**, et **J.-Y. Thibon**. The peak algebra and the Hecke-Clifford algebras at  $q = 0$ . *J. Combin. Theory Ser. A*, 107(1) :1–19, 2004.
- [164] K. Bertet, D. Krob, M. Morvan, **J.-C. Novelli**, H. Phan, et **J.-Y. Thibon**. An overview of  $\Lambda$ -type operations on quasi-symmetric functions. *Comm. Algebra*, 29(9) :4277–4303, 2001. Special issue dedicated to Alexei Ivanovich Kostrikin.
- [165] **E. Briand**, **J.-G. Luque**, et **J.-Y. Thibon**. A complete set of covariants of the four qubit system. *J. Phys. A*, 36(38) :9915–9927, 2003.
- [166] **E. Briand**, **J.-G. Luque**, **J.-Y. Thibon**, et F. Verstraete. The moduli space of three qutrit states. *J. Math. Phys.*, 2004. À paraître, preprint : arXiv :quant-ph/0306122.
- [167] J. Cassaigne, M. Espie, D. Krob, **J.-C. Novelli**, et **F. Hivert**. The Chinese monoid. *Internat. J. Algebra Comput.*, 11(3) :301–334, 2001.
- [168] W. Chen, Q.-H. Hou, et **A. Lascoux**. An involution for the Gauss identity. *J. Combin. Theory Ser. A*, 102(2) :309–320, 2003.
- [169] **M. Crochemore**, **J. Désarménien**, et **D. Perrin**. A note on the Burrows-Wheeler transformation. *Theoretical Computer Science*, 2004. À paraître.
- [170] J. Dornstetter, D. Krob, **J.-Y. Thibon**, et E. Vassilieva. Performance analysis of demodulation with diversity—a combinatorial approach. I. Symmetric function theoretical methods. *Discrete Math. Theor. Comput. Sci.*, 5(1) :191–204 (electronic), 2002.
- [171] G. Duchamp, M. Flouret, É. Laugerotte, et **J.-G. Luque**. Direct and dual laws for automata with multiplicities. *Theoret. Comput. Sci.*, 267(1-2) :105–120, 2001. Implementing automata (Mont-Saint-Aignan, 1998).
- [172] G. Duchamp, **F. Hivert**, et **J.-Y. Thibon**. Noncommutative symmetric functions. VI. Free quasi-symmetric functions and related algebras. *Internat. J. Algebra Comput.*, 12(5) :671–717, 2002.
- [173] G. Duchamp, É. Laugerotte, et **J.-G. Luque**. On the support of graph Lie algebras. *Theoret. Comput. Sci.*, 273(1-2) :283–294, 2002. WORDS (Rouen, 1999).
- [174] F. Fomin, D. Kratsch, et **J.-C. Novelli**. Approximating minimum cocolorings. *Inform. Process. Lett.*, 84(5) :285–290, 2002.
- [175] A. Fu et **A. Lascoux**. Partition analysis and symmetrizing operators. *Journal of Combinatorial Theory Series A*, 2004. À paraître, preprint : math.CO/0404064.

- 
- [176] A. M. Fu et **A. Lascoux**.  $q$ -identities from Lagrange and Newton interpolation. *Adv. in Appl. Math.*, 31(3) :527–531, 2003.
- [177] **B. Gauthier**. Hyperg, a Maple package for manipulating hypergeometric series. *Séminaire Lotharingien de Combinatoire*, 43 :10 pp., 2000. électronique.
- [178] **F. Hivert**. Hecke algebras, difference operators, and quasi-symmetric functions. *Adv. Math.*, 155(2) :181–238, 2000.
- [179] **F. Hivert**, **J.-C. Novelli**, et **J.-Y. Thibon**. Un analogue du monoïde plaxique pour les arbres binaires de recherche. *C. R. Math. Acad. Sci. Paris*, 335(7) :577–580, 2002.
- [180] **F. Hivert**, **J.-C. Novelli**, et **J.-Y. Thibon**. Sur quelques propriétés de l’algèbre des arbres binaires. *C. R. Math. Acad. Sci. Paris*, 337(9) :565–568, 2003.
- [181] **F. Hivert**, **J.-C. Novelli**, et **J.-Y. Thibon**. The algebra of binary search trees. *Theoretical Computer Science.*, 2004. À paraître, preprint : math.CO/0401089.
- [182] **F. Hivert** et N. Thiéry. MuPAD-combinat, an open-source package for research in algebraic combinatorics. *Séminaire Lotharingien de Combinatoire*, 51 :70 pp., 2003. électronique.
- [183] Q.-H. Hou, **A. Lascoux**, et Y.-P. Mu. Continued fractions for Rogers-Szegő polynomials. *Numer. Algorithms*, 35(1) :81–90, 2004.
- [184] Q.-H. Hou, **A. Lascoux**, et Y.-P. Mu. Evaluation of some hankel determinants. *Advances in Appl. Math., Volume Robbins*, 2004. À paraître.
- [185] C. Kassel, **A. Lascoux**, et C. Reutenauer. Factorizations in Schubert cells. *Adv. Math.*, 150(1) :1–35, 2000.
- [186] C. Kassel, **A. Lascoux**, et C. Reutenauer. The singular locus of a Schubert variety. *J. Algebra*, 269(1) :74–108, 2003.
- [187] L. Lapointe, **A. Lascoux**, et J. Morse. Determinantal expression and recursion for Jack polynomials. *Electron. J. Combin.*, 7(N1), 2000.
- [188] L. Lapointe, **A. Lascoux**, et J. Morse. Tableau atoms and a new Macdonald positivity conjecture. *Duke Math. J.*, 116(1) :103–146, 2003.
- [189] **A. Lascoux**. Motzkin paths and powers of continued fractions. *Sém. Lothar. Combin.*, 44, 2000.
- [190] **A. Lascoux**. About division by 1. *Electron. J. Combin.*, N8, 2001.
- [191] **A. Lascoux**. Yang-Baxter graphs, Jack and Macdonald polynomials. *Ann. Comb.*, 5(3-4) :397–424, 2001. Dedicated to the memory of Gian-Carlo Rota (Tianjin, 1999).
- [192] **A. Lascoux**. Addition of  $\pm 1$  : Application to arithmetic. *Séminaire Lotharingien de Combinatoire*, 52, 2004.
- [193] **A. Lascoux**. Sylvester’s bijection between strict and odd partitions. *Discrete Math.*, 277(1-3) :275–278, 2004.
- [194] **A. Lascoux** et **M. Lassalle**. Une identité remarquable en théorie des partitions. *Math. Ann.*, 318(2) :299–313, 2000.

- 
- [195] **A. Lascoux** et P. Pragacz. Orthogonal divided differences and Schubert polynomials,  $\tilde{P}$ -functions, and vertex operators. *Michigan Math. J.*, 48 :417–441, 2000.
- [196] **A. Lascoux** et P. Pragacz. Jacobians of symmetric polynomials. *Ann. Comb.*, 6(2) :169–172, 2002.
- [197] **A. Lascoux** et P. Pragacz. Double Sylvester sums for subresultants and multi-Schur functions. *J. Symbolic Comput.*, 35(6) :689–710, 2003.
- [198] **A. Lascoux** et **J.-Y. Thibon**. Vertex operators and the class algebras of symmetric groups. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 283(Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 6) :156–177, 261, 2001.
- [199] **M. Lassalle**. Une identité en théorie des partitions. *J. Combin. Theory Ser. A*, 89(2) :270–288, 2000.
- [200] **M. Lassalle**. Explicitation des polynômes de Jack et de Macdonald en longueur trois. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(6) :505–508, 2001.
- [201] **M. Lassalle**. Une  $q$ -spécialisation pour les fonctions symétriques monomiales. *Adv. Math.*, 162(2) :217–242, 2001.
- [202] **M. Lassalle**. A new family of positive integers. *Ann. Comb.*, 6(3-4) :399–405, 2002.
- [203] **M. Lassalle**. Jack polynomials and some identities for partitions. *Trans. Amer. Math. Soc.*, 356(9) :3455–3476 (electronic), 2004.
- [204] **M. Lassalle** et M. Schlosser. An analytic formula for Macdonald polynomials. *C. R. Math. Acad. Sci. Paris*, 337(9) :569–574, 2003.
- [205] **J.-G. Luque** et **J.-Y. Thibon**. Pfaffian and Hafnian identities in shuffle algebras. *Adv. in Appl. Math.*, 29(4) :620–646, 2002.
- [206] **J.-G. Luque** et **J.-Y. Thibon**. Hankel hyperdeterminants and Selberg integrals. *J. Phys. A*, 36(19) :5267–5292, 2003.
- [207] **J.-G. Luque** et **J.-Y. Thibon**. Polynomial invariants of four qubits. *Phys. Rev. A* (3), 67(4) :042303, 5, 2003.
- [208] **J.-G. Luque** et **J.-Y. Thibon**. Hyperdeterminantal calculations of Selberg’s and Aomoto’s integrals. *Molecular Physics*, 102(11–12) :1351–1359, 2004. Special Issue : In Memory of Brian Garner Wybourne.
- [209] **J.-C. Novelli**, **J.-Y. Thibon**, et N. Thiéry. Algèbre de Hopf de graphes. *C. R. Acad. Sci. Paris Sér. I Math.*, 2004. À paraître.
- [210] **V. Prosper**. Factorization properties of the  $q$ -specialization of Schubert polynomials. *Ann. Comb.*, 4(1) :91–107, 2000.
- [211] **V. Prosper**. SFA, a package on symmetric functions considered as operators over the ring of polynomials for the computer algebra system MAPLE. *J. Symbolic Comput.*, 29(1) :83–94, 2000.
- [212] **V. Prosper** et **S. Veigneau**. On the palindromic reversal process. *Calcolo*, 38(3) :129–140, 2001.
- [213] **K. Slooten**. Generalized Green functions and graded Hecke algebras. *Advances in Mathematics*, 2004. À paraître, preprint : math.CO/404202.

- [214] **J.-Y. Thibon**. The cycle enumerator of unimodal permutations. *Ann. Comb.*, 5(3-4) :493–500, 2001.

## Actes de conférences

- [215] F. Boulier, **F. Hivert**, D. Krob, et **J.-C. Novelli**. Pseudo-permutations. II. Geometry and representation theory. In *Discrete models : combinatorics, computation, and geometry (Paris, 2001)*, Discrete Math. Theor. Comput. Sci. Proc., AA, pages 123–132 (electronic). Maison Inform. Math. Discrèt. (MIMD), Paris, 2001.
- [216] **F. Descouens**. Un algorithme de génération des tableaux de rubans et de calcul de polynôme de spin. In *Journées Montoises, Proceedings*, 2004.
- [217] G. Duchamp, **F. Hivert**, et **J.-Y. Thibon**. Some generalizations of quasi-symmetric functions and noncommutative symmetric functions. In *Formal power series and algebraic combinatorics (Moscow, 2000)*, pages 170–178, Berlin, 2000. Springer.
- [218] G. Duchamp, É. Laugerotte, et **J.-G. Luque**. Extending the scalars of minimization. In *SCI'2001 Fifth World Multi-Conference on Systemics, Cybernetics and Informatics*, July 2001.
- [219] G. Duchamp et **J.-G. Luque**. Congruences compatible with the shuffle product. In D. Krob, A. Mikhalev, et A. M. (Eds.), editors, *Formal power series and algebraic combinatorics (Moscow, 2000)*, pages 422–431, Berlin, 2000. Springer.
- [220] M. Espie, **J.-C. Novelli**, et G. Racinet. Formal computations about multiple zeta values. In *From combinatorics to dynamical systems*, volume 3 of *IRMA Lect. Math. Theor. Phys.*, pages 1–16, Berlin, 2003. de Gruyter.
- [221] F. Fomin, D. Kratsch, et **J.-C. Novelli**. Approximating minimum cocolorings. In *Fundamentals of computation theory (Riga, 2001)*, volume 2138 of *Lecture Notes in Comput. Sci.*, pages 118–125, Berlin, 2001. Springer.
- [222] **F. Hivert**, **J.-C. Novelli**, et **J.-Y. Thibon**. An analogue of the plactic monoid for binary search trees. In *4th International Conference on Combinatorics on Words'03*, volume 27, pages 27–35. Turku Center for Computer Science, 2003.
- [223] **F. Hivert** et N. Thiéry. Deformation of symmetric functions and the rational Steenrod algebra. In *Invariant theory in all characteristics*, volume 35 of *CRM Proc. Lecture Notes*, pages 91–125, Providence, RI, 2004. Amer. Math. Soc.
- [224] A. Kirillov et **A. Lascoux**. Factorization of Kazhdan-Lusztig elements for Grassmannians. In *Combinatorial methods in representation theory (Kyoto, 1998)*, volume 28 of *Adv. Stud. Pure Math.*, pages 143–154, Tokyo, 2000. Kinokuniya.
- [225] **A. Lascoux**. Young's representations of the symmetric group. In *Symmetry and structural properties of condensed matter*. World Scientific, 2000.
- [226] **A. Lascoux**. Alphabet splitting. In *Algebraic combinatorics and computer science*, pages 431–444. Springer Italia, Milan, 2001.
- [227] **A. Lascoux**. Ordering the affine symmetric group. In *Algebraic combinatorics and applications (Gößweinstein, 1999)*, pages 219–231, Berlin, 2001. Springer.

- [228] **A. Lascoux**. Transition on Grothendieck polynomials. In *Physics and combinatorics, 2000 (Nagoya)*, pages 164–179, River Edge, NJ, 2001. World Sci. Publishing.
- [229] B. Leclerc et **J.-Y. Thibon**. Littlewood-Richardson coefficients and Kazhdan-Lusztig polynomials. In *Combinatorial methods in representation theory (Kyoto, 1998)*, volume 28 of *Adv. Stud. Pure Math.*, pages 155–220, Tokyo, 2000. Kinokuniya.
- [230] **J.-G. Luque** et **J.-Y. Thibon**. Noncommutative symmetric functions associated with a code, Lazard factorisation and Witt vectors. In *Journées Montoises, Proceedings*, 2004.
- [231] **J.-C. Novelli** et D. Rossin. On the toppling of a sand pile. In *Discrete models : combinatorics, computation, and geometry (Paris, 2001)*, Discrete Math. Theor. Comput. Sci. Proc., AA, pages 275–285 (electronic). Maison Inform. Math. Discrèt. (MIMD), Paris, 2001.
- [232] **J.-C. Novelli** et **J.-Y. Thibon**. Free cumulants as characters of symmetric groups, and a Hopf algebra of parking functions. In *FPSAC, proceedings*, 2004.
- [233] **J.-C. Novelli** et E. Vassilieva. On the center of the quantum pseudoplactic algebra. In *Formal power series and algebraic combinatorics (Moscow, 2000)*, pages 709–718, Berlin, 2000. Springer.

## Collections, livres et chapitres dans les livres

- [234] **J. Désarménien**. Statistics on permutations and words. In **M. Lothaire**, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
- [235] J.-P. Gazeau, R. Kerner, J.-P. Antoine, S. Métens, et **J.-Y. Thibon**, editors. *GROUP 24 : Physical and Mathematical Aspects of Symmetries : Proceedings of the 24th International Colloquium on Group Theoretical Methods in Physics, Paris, 15-20 July 2002*, Institute of Physics Conference Series 173, 2003.
- [236] **A. Lascoux**. Double crystal graphs. In *Studies in memory of Issai Schur (Chevaleret/Rehovot, 2000)*, volume 210 of *Progr. Math.*, pages 95–114. Birkhäuser Boston, Boston, MA, 2003.
- [237] **A. Lascoux**. Littlewood’s formulas for characters of orthogonal and symplectic groups. In *Algebraic combinatorics and quantum groups*, pages 125–133. World Sci. Publishing, River Edge, NJ, 2003.
- [238] **A. Lascoux**. *Symmetric functions and combinatorial operators on polynomials*, volume 99 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.
- [239] **A. Lascoux**, B. Leclerc, et **J.-Y. Thibon**. The plactic monoid. In **M. Lothaire**, editor, *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.
- [240] **A. Lascoux** et P. Pragacz. Schur  $Q$ -functions and degeneracy locus formulas for morphisms with symmetries. In *Recent progress in intersection theory (Bologna, 1997)*, Trends Math., pages 239–263. Birkhäuser Boston, Boston, MA, 2000.

- [241] B. Leclerc, M. Nazarov, et **J.-Y. Thibon**. Induced representations of affine Hecke algebras and canonical bases of quantum groups. In *Studies in memory of Issai Schur (Chevaleret/Rehovot, 2000)*, volume 210 of *Progr. Math.*, pages 115–153. Birkhäuser Boston, Boston, MA, 2003.
- [242] **J.-Y. Thibon**. Lectures on noncommutative symmetric functions. In *Interaction of combinatorics and representation theory*, volume 11 of *MSJ Mem.*, pages 39–94. Math. Soc. Japan, Tokyo, 2001.

## Thèses et habilitations

- [243] **F. Hivert**. *Combinatoire et calcul symbolique dans les algèbres de Hopf*. Habilitation à diriger des recherches, Université de Marne la Vallée, 2004. Jury : N. Bergeron, C. Frougny, J.-L. Loday, M. Morvan, P. Paule, M. Petkovsek, **J.-Y. Thibon**.
- [244] **J.-C. Novelli**. *Applications de la combinatoire bijective et algébrique*. Habilitation à diriger des recherches, Université de Marne la Vallée, 2001. Jury : F. Bergeron, M. Bousquet-Melou, **M. Crochemore**, M. Delest, P. Gastin, M. Habib, D. Krob, **J.-Y. Thibon**.

## Divers

- [245] **E. Briand**. Symétriques (polynômes). Encyclopedia Universalis en ligne. À paraître, <http://www.universalis-edu.com>.
- [246] A. Fu et **A. Lascoux**. Rational interpolation and basic hypergeometric series. *Ramanujan Journal*, 2003. Soumis, preprint : math.CO/0404063.
- [247] A. Fu et **A. Lascoux**.  $q$ -identities related to overpartitions and divisor functions, 2004. Soumis, preprint : math.CO/0404065.
- [248] **F. Hivert**, **J.-C. Novelli**, et J. Thibon. Representation theory of the 0-Ariki-Koike-Shoji algebras, 2004. Preprint : math.CO/040748.
- [249] **F. Hivert** et N. Thiéry. MuPAD-Combinat. en développement.
- [250] **M. Jöllenbeck**. On the multi graded Hilbert and Poincaré series and the Gold-property of monomial rings. Preprint 2004.
- [251] D. Krob, M. Latapy, **J.-C. Novelli**, H.-D. Phan, et S. Schwer. Pseudo-permutations I : First combinatorial and lattice properties. *Advances in Applied Mathematics*, 2001. Soumis.
- [252] **A. Lascoux**. Chern and Yang through ice. *Selecta Mathematica*, 2002. Soumis.
- [253] **A. Lascoux**. Schubert et Grothendieck : Un bilan bidécennal. Séminaire Lotharingien de Combinatoire, 2003. Soumis.
- [254] **A. Lascoux**, **B. Gauthier**, R. Gugish, **V. Prosper**, B. Ung, et **S. Veigneau**.  $\mu$ -EC. non maintenu.
- [255] **A. Lascoux**, H. Lin, et A. Yang. Lattice graphs and Schubert polynomials, 2004. Soumis.

- [256] **A. Lascoux**, H. Lin, et A. Yang. Pieri formula for key polynomials, 2004. Soumis.
- [257] **A. Lascoux** et P. Pragacz. Bezoutians, Euclidean division, and orthogonal polynomials, 2004. Soumis.
- [258] **A. Lascoux**, **V. Prosper**, **J.-Y. Thibon**, **S. Veigneau**, et B. Ung. Ace. non maintenu.
- [259] **M. Lassalle**. A short proof of generalized Jacobi-Trudi expansions for Macdonald polynomials. Contemporary Mathematics, 2002. Soumis, preprint : math.CO/0402127.
- [260] **M. Lassalle** et M. Schlosser. Inversion of the Pieri formula for Macdonald polynomials. Preprint : math.CO/0402127.
- [261] **J.-C. Novelli** et **J.-Y. Thibon**. Free quasi-symmetric functions of arbitrary level. Preprint : math.CO/0405597.
- [262] **P.-A. Picon**. Two numbers close to the binomial coefficient. Séminaire Lotharingien de Combinatoire, 2000. À paraître.
- [263] **K. Slooten**. Reducibility of induced discrete series representations of affine Hecke algebras of classical type. Preprint 2004.

# 4

## Informatique linguistique

### 4.1 Composition

#### Membres

- Catherine Gallardo, maître de conférences, Université Paris X
- Tita Kyriacopoulou, PAST
- Éric Laporte, professeur, responsable
- Christian Leclère, ingénieur d'études CNRS

#### Membres associés

- Mirella Conenna, professeur, Université de Bari, Italie
- Mylène Garrigues, maître de conférences
- Michele de Gioia, Université de Trieste, Italie
- Jacqueline Giry-Schneider, ex-maître de conférences à l'Université de Paris VIII
- LIM Joon-seo, ex-lecteur de coréen à l'Université de Rouen
- Claude Martineau
- Annie Meunier, ex-maître de conférences à l'Université de Paris VIII
- NAM Jee-sun, Université coréenne des Etudes étrangères, Corée du Sud
- Alexis Neme, Université de Feira de Sant'Ana, Brésil

#### Doctorants, ATER et post-doctorants

- Olivier Blanc, doctorant
- Matthieu Constant, ATER
- Marina Dumitriu, post-doctorante
- HUH Hyun-gue, doctorant
- JIN Guang-jin, post-doctorante
- JUNG Eun-jin, doctorante
- Guénaëlle Mercier, doctorante CIFRE

- Takuya Nakamura, doctorant
- Javier Sastre, doctorant
- Tahirisoa Rakotonanahary, doctorante en co-tutelle
- Rania Voskaki, doctorante
- Stavroula Voyatzi, doctorante
- Anastasia Yannacopoulou, doctorante

### **Anciens membres**

- BAE Sun-mee, doctorante 2001-2003
- CHUNG Gohsran, doctorante 2001-2003
- Blandine Courtois, collaboratrice bénévole 2001-2002
- Catherine Domingues, doctorante 2001
- EUM Du-eun, doctorante 2001-2004
- Maurice Gross, professeur à l'Université de Paris VII, 2001
- Krit Kosawat, doctorant 2001-2003
- Mario Monteleone, doctorant 2002-2003
- Mavina Pantazara, doctorante 2001-2003
- Sébastien Paumier, doctorant 2001-2003, ATER 2003-2004
- Agata Savary-Chrobot, post-doctorante 2001-2002
- Marta Ungermanova, doctorante 2001-2003

## **4.2 Thèmes de recherche**

L'équipe a organisé ses activités autour des ressources linguistiques dont elle est dépositaire. Il s'agit de données sur les langues naturelles qui ont été formalisées en vue de leur utilisation dans des applications informatiques. Ces ressources ont été utilisées dans des applications (recherche d'informations dans des textes juridiques, levée d'ambiguïtés lexicales, détection des pronoms impersonnels, segmentation du texte écrit en thaï...). Certains logiciels développés pour ces applications sont d'intérêt général et ont reçu le statut de logiciel libre sous licence LGPL. Les innovations apportées dans ces applications ont trait aux algorithmes utilisés et à l'utilisation des modèles linguistiques sous-jacents. Un effort important a été consacré à organiser une partie des ressources linguistiques du laboratoire sous la forme de bibliothèques de données munies d'outils d'enrichissement, de validation et d'interrogation. Les plus innovantes de ces fonctionnalités sont :

- la compilation de tables de propriétés syntaxiques en graphes permettant une recherche directe dans le texte écrit,
- de nouveaux algorithmes de compression de textes et de dictionnaires,
- la notion de masque lexical permettant la reconnaissance d'un ensemble de mots du vocabulaire à travers des critères explicites.

Un système intégré d'analyse de textes écrits à l'aide de ressources linguistiques, Unitem, a été construit en tant que logiciel libre et a réuni en 2 ans une vaste communauté internationale d'utilisateurs. Un système analogue est en construction en partenariat avec un consortium industriel dans le cadre d'une plate-forme du Réseau national des technologies logicielles (RNTL), Outilex. Une partie des ressources ont été traduites dans des formats XML, ce qui a nécessité un travail de documentation et de normalisation mené dans le cadre de projets coopératifs, et une prise en compte des besoins des différents utilisateurs potentiels. Des extensions des ressources existantes ont été formalisées et mises sous des formes diverses : dictionnaires, tables de propriétés, graphes, réseaux de transitions récursifs (RTN). Les principales langues concernées sont le français (expressions de pourcentage, expressions verbales figées dans 4 variétés de français, adjectifs non prédicatifs, noms propres de lieux, constructions en *faire le...*), le coréen, l'italien, le tchèque, le grec moderne et le roumain.

La communauté internationale étant réceptive aux questions relatives aux ressources linguistiques, y compris aux ressources complexes dont nous sommes spécialistes (mots composés, grammaires lexicalisées), la politique de distribution libre de ressources linguistiques sera poursuivie. Elle doit permettre au laboratoire d'établir des collaborations avec un nouveau type d'utilisateurs : ceux qui adaptent les méthodes statistiques classiques du domaine à des ressources plus complexes que celles habituellement utilisées. Cette évolution est d'ores et déjà entamée. Elle devrait déboucher sur de nouvelles applications performantes, qui valideront la qualité de nos ressources, ou permettront d'adapter les méthodes de formalisation employées lors de leur construction.

## 4.3 Résultats

### 4.3.1 Applications

Dans cette partie nous présentons les principales applications réalisées par l'équipe. Nous introduisons ainsi la raison d'être de nos travaux plus fondamentaux, que nous exposons plus loin. Toutes ces applications correspondent à des objectifs spécifiques du traitement automatique des textes en langues naturelles. Elles utilisent des ressources linguistiques disponibles au sein de l'équipe ou fournies par nos partenaires, et font le moins possible appel à des approximations à partir de données incertaines.

Dans le cadre d'une thèse financée par le Centre scientifique et technique du bâtiment (CSTB) ont ainsi été réalisés des outils de recherche d'informations dans des textes juridiques sur la sécurité incendie [364]. À partir de connaissances précises sur les variations syntaxiques des formes susceptibles d'être cherchées par l'utilisateur, ces outils font le lien entre les formes utilisées dans les requêtes et celles présentes dans les textes juridiques. La technique utilisée est celle des graphes d'expressions.

Un travail sur la formation des traducteurs et interprètes a mis en évidence comment les outils informatiques d'accès aux lexiques terminologiques, et aux textes spécialisés, sont exploitables pour la formation initiale ou continue à la traduction spécialisée. Ce travail a bénéficié de l'expérience née de nos contacts avec deux équipes d'enseignants dans ce domaine, l'une à l'Université de Marne-la-Vallée et l'autre à l'Université Aristote de Thessalonique [330].

Une des techniques classiques de correction d'erreurs, la consultation de lexique tolérant une distance entre le mot recherché et le mot trouvé, a reçu une amélioration du point de vue algorithmique [336].

Outre ces expériences purement applicatives, nous nous intéressons également à certaines opérations fondamentales sur le texte écrit. Nous réalisons des composants de base qui effectuent ces opérations et sont utilisables dans des applications. Nous intégrons systématiquement ces composants dans le système Unitex, un système multilingues d'analyse de textes écrits à l'aide de ressources linguistiques, réalisé en 2002 par Sébastien Paumier. Unitex est constitué de logiciel libre sous la licence Lesser Gnu Public Licence (LGPL). Le système Unitex a réuni en 2 ans une communauté internationale d'utilisateurs, informaticiens comme linguistes, et a été adopté par de nombreuses équipes pour l'enseignement de l'informatique linguistique. Depuis sa création, il s'est enrichi de nouvelles fonctionnalités développées par des contributeurs intérieurs et extérieurs à notre équipe, mais nous coordonnons son évolution. Un système analogue à Unitex est en construction en partenariat avec un consortium industriel dans le cadre d'une plate-forme du Réseau national des technologies logicielles (RNTL), Outilex.

Parmi ces composants de base réalisant des opérations sur le texte écrit, citons cinq travaux de recherche significatifs.

- La première opération de l'analyse lexicale d'un texte est la *tokenisation*, qui consiste à donner au texte une structure fondée sur les mots simples et autres symboles élémentaires. De la tokenisation dépendent les performances des opérations ultérieures. Une représentation adaptée à l'utilisation de grammaires de grande taille sur des textes de grande taille a été élaborée et intégrée à Unitex [334].
- L'analyse automatique du texte spécialisé met en jeu des techniques spécifiques lorsque l'écriture du texte est très stéréotypée, comme dans le cas des rapports boursiers dans les quotidiens, qui utilisent de façon répétitive certaines expressions. En exploitant ce fait, [328] met en œuvre une reconnaissance quasi-exhaustive des expressions informatives qui apparaissent dans de tels textes, ce qui permet de réaliser un système de recherche d'informations performant en associant directement, a priori, les résultats voulus aux expressions représentées. Le modèle utilisé est celui des graphes d'expressions.
- Une collaboration avec l'Université de Bergen (Norvège) a permis de mettre au point un outil d'analyse automatique de mots polylexicaux en norvégien [288], c'est-à-dire de concaténations de mots existant par ailleurs, comme c'est fréquem-

ment le cas dans la plupart des langues germaniques. Le logiciel développé pour cette application est d'intérêt général. Il a reçu le statut de logiciel libre sous licence LGPL et a été intégré à Unitex.

- Dans le cadre d'une thèse financée par le gouvernement thaïlandais, a été réalisé un système de segmentation du texte thaï par lexique et par règles [367]. Comme le système d'écriture thaïlandais n'utilise pas l'espace entre mots, il est nécessaire de recourir à un système de segmentation pour délimiter les mots. Les règles qui ont été mises en oeuvre tirent parti de contraintes combinatoires sur les consécutives de lettres de l'alphabet thaï. Elles sont exprimées sous la forme de transducteurs finis, c'est-à-dire d'automates finis qui insèrent des délimiteurs dans certains états en fonction du contexte. Ces techniques ont permis d'améliorer sensiblement les performances de la segmentation par rapport aux systèmes existants, fondés sur des approximations. Ici encore, le logiciel et les transducteurs développés sont distribués avec Unitex.
- La levée d'ambiguïtés lexicales est un des intérêts majeurs de l'équipe. Il s'agit d'une opération nécessaire entre l'étiquetage des mots d'un texte par lexique et l'exploitation de cet étiquetage. En effet, l'étiquetage par lexique est celui qui donne les résultats les plus fiables, mais il produit des ambiguïtés lexicales qui doivent être éliminées des résultats pour que ceux-ci puissent à leur tour être traités. Nous utilisons pour cela le système ELAG. Nous avons étudié en détail ses possibilités [279], nous l'avons intégré à Unitex en le rendant indépendant de la langue, et nous participons à la construction manuelle des règles en français et dans d'autres langues, en collaboration avec des linguistes extérieurs [296].

Les membres grecs de l'équipe ont activement travaillé à la réalisation et à l'intégration de ces composants pour le grec moderne [350].

### 4.3.2 Bibliothèques de ressources linguistiques

La communauté internationale est actuellement réceptive aux questions relatives aux ressources linguistiques, y compris aux ressources complexes dont nous sommes spécialistes (mots composés, grammaires lexicalisées). Nous avons entamé une politique de distribution libre de ressources linguistiques en vue d'établir des collaborations avec un nouveau type d'utilisateurs : ceux qui adaptent les méthodes statistiques classiques du domaine à des ressources plus complexes que celles habituellement utilisées. Ces collaborations devraient déboucher sur de nouvelles applications performantes. Soit celles-ci valideront la qualité de nos ressources, soit elles nous permettront d'adapter nos méthodes de formalisation et de construction de ressources linguistiques.

Du point de vue juridique, nous avons créé une licence d'utilisation de ressources linguistiques comparable au statut LGPL pour les logiciels, la licence LGPL-LR. Les ressources linguistiques incluses dans le système Unitex, et donc distribuées gratuitement, sont sous licence LGPL-LR. Les ressources linguistiques libres fournies par l'IGM

dans le cadre de la plate-forme Outilex seront également en LGPL-LR.

Du point de vue de la structuration des données, une partie des ressources ont été traduites dans des formats XML élaborés en collaboration avec le groupe de travail Technolangue sur la normalisation des formats de données pour le traitement des langues naturelles. Ces nouveaux formats sont mieux documentés que les formats antérieurs, prennent mieux en compte des besoins des différents utilisateurs potentiels, et sont mieux à même de servir de formats d'échange entre équipes et d'assurer une interopérabilité entre environnements informatiques. Ce travail s'achemine maintenant vers l'adoption d'une ou plusieurs normes ISO.

Un effort important a été consacré à organiser une partie des ressources linguistiques du laboratoire sous la forme de bibliothèques de données munies d'outils d'acquisition, d'enrichissement, de validation et d'interrogation. Les plus innovantes de ces fonctionnalités sont :

- la lexicalisation des grammaires syntaxiques (cf. ci-dessous),
- de nouveaux algorithmes de compression de textes et de lexiques [369, 335],
- l'interrogation et la visualisation de la base de données sémantico-syntaxique, y compris de sa documentation en ligne [327],
- la recherche dans une bibliothèque de graphes [306, 363],
- la notion de masques lexicaux, filtres permettant la reconnaissance d'un ensemble de mots du vocabulaire à travers des critères explicites [295],
- l'acquisition de ressources linguistiques : acquisition de termes techniques composés [298],
- un ensemble d'outils d'enrichissement et de validation des contenus de lexiques, tels que la génération automatique de formes fléchies [307].

Notre travail sur la lexicalisation des grammaires syntaxiques est une pièce essentielle de cet ensemble d'outils de gestion de bibliothèques de ressources linguistiques. Les grammaires formelles qui représentent la syntaxe des langues naturelles tirent généralement le meilleur parti possible des régularités syntaxiques à l'intérieur d'une langue, mais, en contrepartie, leurs règles doivent être adaptées aux éléments lexicaux particuliers, qui possèdent des propriétés différentes, dites idiosyncrasiques. Ce processus, appelé lexicalisation, intéresse la communauté du traitement des langues naturelles depuis une dizaine d'années, mais nous sommes particulièrement bien placés pour y travailler en raison de la grande taille des lexiques syntaxiques dont nous disposons (environ 80 000 entrées).

La méthode mise en oeuvre à l'IGM [287] consiste à introduire manuellement dans les grammaires syntaxiques des paramètres dont les valeurs dépendent des classes d'éléments lexicaux, ou des éléments lexicaux eux-mêmes, et sont données dans des tables de propriétés syntaxiques. Les grammaires syntaxiques étant représentées sous la forme de RTN, les automates qui composent ces RTN peuvent être lexicalisés séparément. Les valeurs des paramètres sont soit binaires (interruption ou non d'un chemin dans un automate), soit lexicaux (valeur d'un mot), mais jamais numériques.

Une fois les paramètres introduits, la lexicalisation est assurée par un outil logiciel, intégré à Unitex, qui accède aux tables de propriétés et produit des versions lexicalisées des automates. Lorsque les tables de propriétés donnent les valeurs des paramètres pour les classes, on obtient une version lexicalisée par classe [372]; lorsque les valeurs sont données au niveau des éléments lexicaux, on obtient une version lexicalisée pour chaque élément lexical. Les deux niveaux peuvent être appliqués successivement. La méthode peut être adaptée au cas où plusieurs tables de propriétés sont nécessaires pour stocker les valeurs des paramètres [304, 303].

Avec ce formalisme, toutes les ressources linguistiques qui font l'objet d'une maintenance linguistique (les RTN paramétrés et les tables de propriétés) possèdent une forme lisible et compacte dans laquelle une grande quantité d'informations peut être réunie sur un même écran : les RTN sont représentés sous forme de graphes, et les tables de propriétés sous forme de matrices.

Les deux points précédents (applications, gestion de ressources linguistiques) ont en commun l'utilisation fréquente d'automates finis et de transducteurs finis. Un chapitre du prochain livre de la série Lothaire [355] fait le point sur ces utilisations des automates.

### 4.3.3 Extension des ressources

Des extensions des ressources existantes ont été formalisées et mises sous des formes diverses : lexiques [265], tables de propriétés [282], graphes [302]. Les descriptions linguistiques (grammaticales, morphologiques, syntaxiques) étant destinées à être exploitées dans des applications informatiques, les objectifs sont extrêmement concrets : spécifier les formes effectivement observables. En ce qui concerne le français, citons des travaux sur

- les expressions de pourcentage [285],
- les expressions verbales figées dans 4 variétés de français [278],
- les adjectifs non prédicatifs [325],
- les noms propres de lieux [303].

Hormis le français, les principales langues concernées sont

- le coréen [365],
- l'italien [359],
- le grec moderne [276],
- le tchèque [290],
- le roumain.

Dans le domaine de l'informatique linguistique, notre laboratoire est l'un des seuls qui, outre des recherches sur les applications informatiques et la gestion informatisée de ressources linguistiques, contribue également à la construction et à l'extension de telles ressources. Cette spécificité fait partie intégrante de notre stratégie de recherche. Elle sensibilise les informaticiens à l'importance de la lisibilité et de la maintenabilité

des formats dans lesquels sont exprimées les données linguistiques. Inversement, elle sensibilise les linguistes au caractère à la fois formel et concret que doivent présenter leurs résultats pour être exploitables dans les applications.

## 4.4 Activités

### 4.4.1 Contrats

- Outilex, projet financé par le Ministère de l'Industrie, coordonné par Éric Laporte, avec 6 partenaires industriels et 4 partenaires universitaires, sur la construction de logiciel libre d'analyse de textes en langues naturelles par dictionnaires et automates finis (2002–2005).
- Concorlex, projet franco-belge sur la levée d'ambiguïtés lexicales (2002–2004).
- Elag, projet franco-portugais sur la levée d'ambiguïtés lexicales (2002).
- Transweb, projet financé par le Ministère de l'Industrie visant à mettre en place un moteur de mémoire de traduction, avec 3 partenaires industriels et 3 partenaires universitaires (1999–2002).
- Un contrat d'encadrement de bourse CIFRE (Antonio Balvet) avec Thales sur le filtrage d'informations à l'aide de classes d'automates linguistiques (1999–2002).

### 4.4.2 Diffusion

Dépôt auprès de la « Société des Gens de Lettres de France » d'un CD de ressources linguistiques.

Organisation de colloques :

- EACL Workshop on Finite-State Methods in Natural Language Processing, Budapest, avril 2003.
- International Conference on Lexis and Grammar :
  - 23rd, Deauville, novembre 2004.
  - 22nd, Thessalonique, septembre 2003.
  - 21st, Bari (Italie), septembre 2002.
  - 20th, Londres, septembre 2001.
- Lexicon-Grammar Workshop, Pékin, octobre 2004.
- Description linguistique pour le traitement automatique du français, Rimouski (Canada), mai 2003.
- Journées d'hommage à Maurice Gross, Paris, juin 2002.
- Traduction et études coréennes en France, décembre 2004.

### 4.4.3 Collaborations

La construction parallèle de dictionnaires électroniques se traduit par des collaborations régulières avec les laboratoires du réseau RELEX, notamment Annibale Elia (Université de Salerne, Italie), Mirella Conenna (Université de Bari, Italie), Béatrice Lamiroy (Université catholique de Louvain), Cédric Fairon (Université catholique de Louvain-la-Neuve), Elisabete Marques Ranchhod (Université de Lisbonne), Franz Guenther (Université de Munich), Jacques Labelle (Université du Québec à Montréal), Zygmunt Vetulani (Université de Poznan, Pologne) et leurs équipes.

Nous avons également eu des collaborations suivies avec Denis Maurel (Université de Tours), Strahil Ristov (Institut Ruder Boskovic, Croatie), Maria das Graças Nunes (Université de São Paulo), Carmelita Dias (Université pontificale catholique de Rio de Janeiro), Anna Anastasiadis-Symeonidis (Université de Thessalonique, Grèce), le Centre de la langue grecque (Thessalonique), l'Institut pédagogique d'Athènes (Université d'Athènes).

### 4.4.4 Activités doctorales

DEA d'informatique fondamentale : analyse automatique de textes, automates et transducteurs pour le texte écrit et la parole, XML dans le traitement automatique des langues.

### 4.4.5 Thèses et habilitations

#### Habilitations

- Panayota-Tita Kyriacopoulou, sous le titre *Analyse automatique des textes écrits : le cas du grec moderne* à l'Université de Marne-la-Vallée, le 4 juillet 2003.

#### Thèses en cours

- Olivier Blanc, depuis octobre 2002, *Analyse automatique de textes par confrontation d'automates et lexique-grammaire*.
- HUH Hyun-gue, depuis octobre 1999, *Segmentation automatique du texte coréen*.
- JUNG Eun-jin, depuis décembre 1997, *Grammaires locales des adverbes de durée et de date en coréen*.
- Guénaëlle Mercier, depuis mars 2004, *La levée d'ambiguïtés lexicales entre Nom et Adjectif par des grammaires ELAG écrites manuellement*.
- Takuya Nakamura, depuis septembre 2000, *Les compléments nominaux de verbes à complétive*.
- Javier Sastre, depuis novembre 2004, *Optimisation des algorithmes d'application des grammaires locales pour l'analyse syntaxique automatique et exacte des phrases simples du français*.

- Tahirisoa Rakotonanahary, depuis février 2002, *Comparaison trilingue (malgache-espagnol-français) des expressions verbales figées métaphoriques*.
- Rania Voskaki, depuis octobre 2004, *Le lexique-grammaire des verbes du grec moderne : les constructions transitives non locatives*.
- Stavroula Voyatzi, depuis décembre 2002, *Description syntaxico-sémantique des adverbes figés en vue d'un système d'analyse automatique des textes grecs*.
- Anastasia Yannacopoulou, depuis décembre 2002, *Le lexique-grammaire des verbes en grec moderne : les constructions transitives locatives standard*.

### Thèses

- BAE Sun-Mee, *Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies. Méthodes et applications*, 25 septembre 2002.
- Gohsran Chung, *Analyse des constructions à double nominatif/accusatif par l'opération de restructuration en coréen. Classification syntaxique des constructions à adjectif sôngsang*, 19 décembre 2003.
- Matthieu Constant, *Grammaires locales pour l'analyse automatique de textes : Méthodes de construction et outils de gestion*, 8 septembre 2003.
- Catherine Domingues, *Étude d'outils informatiques et linguistiques pour l'aide à la recherche automatique d'information dans un corpus documentaire*, 16 mars 2001.
- EUM Du-eun, *Syntaxe des verbes de communication en coréen*, 6 juillet 2004.
- Krit Kosawat, *Méthodes de segmentation et d'analyse automatique de textes thaï*, 8 septembre 2003.
- Claude Martineau, *Compression de textes en langue naturelle*, 7 décembre 2001.
- Mario Monteleone, *Lexicographie et dictionnaires électroniques. Des usages linguistiques aux bases de données lexicales*, 8 décembre 2003.
- Mavina Pantazara, *Syntaxe dérivationnelle du grec moderne : Les constructions verbales à un complément prépositionnel et les constructions nominales et adjectivales prédicatives associées*, 6 décembre 2003.
- Sébastien Paumier, *De la reconnaissance de formes linguistiques à l'analyse syntaxique*, 4 juillet 2003.

## 4.5 Références bibliographiques

### Articles de revues

- [264] S.-M. Bae. Deux types de représentations des séquences nominales figées en coréen et de leurs formes fléchies avec INTEX. *Linguisticae Investigationes*, XXIV(2) :147–165, 2001.

- [265] **S.-M. Bae.** Construction of an electronic dictionary for compound nouns in Korean. *Sajeon Pyeonchanhak Yeongu (Studies in Lexicography)*, 11(1) :151–164, 2002. Séoul : Centre de recherches sur le coréen de l’Université de Yonsei.
- [266] **G. Chung.** A study of Korean Sôngsang adjective class and its constructions based on the lexicon–grammar. *Linguistics – The linguistic association of Korean*, 11(2) :17–37, 2003. Version coréenne.
- [267] **G. Chung.** A syntactic classification of sungsang adjectival constructions. *Linguistics – The linguistic association of Korean*, 11(4) :163–184, 2003. Version coréenne.
- [268] **G. Chung.** Une classe de verbes transitifs en coréen : verbes de contact-corporel. *Language research*, 39(1) :39–68, 2003.
- [269] **J. Giry-Schneider.** L’adjectif *personnel* anaphore ou prédicat de relation ? *Langue Française*, 136 :20–33, 2002.
- [270] **M. Gross.** Compte rendu de l’ouvrage de Francis Kaplan : Des singes et des hommes. *Pour la science*, 284, 2001.
- [271] **M. Gross.** Les ambiguïtés. *Linguisticæ Investigationes*, XXIV(1) :3–41, 2001.
- [272] **M. Gross.** Consequences of the metalanguage being included in the language. *Current Issues in Linguistic Theory. The Legacy of Zellig Harris Language and information into the 21st century*, Volume 1 : Philosophy of science, syntax and semantics(228) :57–67, 2002.
- [273] **M. Gross.** Les déterminants numériques, un exemple : les dates horaires. *Langages*, 145, 2002.
- [274] **E.-J. Jung.** Verbes spécifiques de durée avec une complétive en *dei* (pour que P) en coréen. *Linguisticæ Investigationes*, XXIV(2) :203–225, 2001.
- [275] **T. Kyriacopoulou**, S. Mrabti, et **A. Yannacopoulou.** Le dictionnaire électronique des noms composés en grec moderne. *Linguisticæ Investigationes*, XXV(1) :7–28, 2002.
- [276] **T. Kyriacopoulou** et V. Sfetsiou. Les constructions nominales à verbe support en grec moderne. *Linguistic Insights : Studies in Language and Communication*, 5, Phrases and Phraseology – Data and Descriptions :163–181, 2003.
- [277] **T. Kyriacopoulou** et **A. Yannacopoulou.** Vers la constitution du lexique-grammaire des verbes en grec moderne. *Linguisticæ Investigationes*, XXVI(2) :217–236, 2004.
- [278] B. Lamiroy, **C. Leclère**, J.-R. Klein, et J. Labelle. Les expressions verbales figées dans quatre variétés de français : Le projet BFQS. *Cahiers de lexicologie*, 83 :153–172, 2003.
- [279] **É. Laporte.** Reduction of lexical ambiguity. *Linguisticæ Investigationes*, XXIV(1) :67–103, 2001.
- [280] **C. Leclère.** Expressions figées dans la francophonie : le projet BFQS. *Lexique, Syntaxe et Sémantique, Mélanges offerts à Gaston Gross*, pages 321–331, 2000. Bulag, n° Hors Série.
- [281] **C. Leclère.** Emplois verbaux, distributions, métaphores. *Langue française*, 134 :78–89, 2002.

- [282] **C. Leclère**. Organization of the lexicon-grammar of French verbs. *Linguisticae Investigationes*, XXV(1) :29–48, 2002.
- [283] **T. Nakamura**. Maurice Gross et le lexique-grammaire, première partie. *Flambeau, Revue Annuelle de la Section Française, Université des Langues Etrangères, Tokyo (TUFSS)*, 28 :79–96, 2002.
- [284] **T. Nakamura**. Maurice Gross et le lexique-grammaire, deuxième partie. *Flambeau, Revue Annuelle de la Section Française, Université des Langues Etrangères, Tokyo (TUFSS)*, 29 :75–92, 2003.
- [285] **T. Nakamura** et **M. Constant**. Les expressions de pourcentage. *Flambeau, Revue Annuelle de la Section Française, Université des Langues Etrangères, Tokyo (TUFSS)*, 27 :27–46, 2001.
- [286] P. Panagiotidis et **T. Kyriacopoulou**. Solutions technologiques pour la formation des traducteurs et interprètes. *Journal de Linguistique Appliquée*, 18, 2004.
- [287] **S. Paumier**. Some remarks on the application of a lexicon-grammar. *Linguisticae Investigationes*, XXIV(2) :245–256, 2001.
- [288] **S. Paumier** et H. Ulland. Analyse automatique de mots polylexicaux en norvégien. *Linguisticae Investigationes*, 2004. À paraître.
- [289] S. Ristov et **É. Laporte**. Ziv Lempel compression of huge natural language data tries using suffix arrays. *J. Discrete Algorithms (Oxf.)*, 1(1) :241–256, 2000.
- [290] **M. Ungermanová**. Préfixation des verbes de déplacement tchèques. *Linguisticae Investigationes*, XXIV(2) :329–351, 2001.

## Actes de conférences

- [291] A. Anastasiadis-Symeonidis, **T. Kyriacopoulou**, **E. Sklavounou**, I. Thilikos, et V. Rania. A system for analysing texts in Modern Greek. Representing and solving ambiguities. In *Proceedings of COMLEX 2000*, pages 113–116. Department of Electrical Engineering and Computer Technology, University of Patras, Greece, 2000.
- [292] **S.-M. Bae**. Vers le traitement des groupes nominaux du coréen avec INTEX. In A. Dieter, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 9–28, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [293] **S.-M. Bae**. Description of frozen nominal sequences in Korean in an electronic dictionary. In *Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues*, pages 77–100, 2004. Actes des quatrièmes et cinquièmes journées INTEX, Bordeaux 2001 et Marseille 2002.
- [294] S. Bakoura et **R. Voskaki**. Désambiguïsation des unités monétaires, des symboles et des langues en grec moderne en vue de leur intégration dans les systèmes de traitement automatique des textes. In *Studies in Greek Linguistics, Proceedings of the 24th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2003*, pages 90–101, 2004.

- 
- [295] **O. Blanc** et A. Dister. Automates lexicaux avec structure de traits. In *RECITAL 2004, VIII<sup>e</sup> Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues*, pages 23–32, 2004. Fès, Maroc.
- [296] **O. Blanc**, K. Ioannidou, et **R. Voskaki**. Automatic elimination of lexical ambiguities in Modern Greek : presentation of the ELAG system. In *Studies in Greek Linguistics, Proceedings of the 25th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2004*, 2004. À paraître.
- [297] **A. Chrobot**. Description des déterminants numériques anglais par automates et transducteurs finis. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 101–118, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [298] **A. Chrobot**. Étude comparative de deux outils d’acquisition de termes complexes. In *Actes des Quatrièmes Rencontres : Terminologie et Intelligence Artificielle*, pages 129–139, 2001. Nancy, 3-4 Mai 2001, INIST-CNRS.
- [299] **G. Chung**. A type of Transitive Inalienable Possession Construction in Korean. In E. Ranchhod et N. Mamede, editors, *Proceedings Advances in Natural Language Processing*, pages 71–74. Springer (LNAI 2389), 2002. Third International Conference, PorTAL 2002, Faro, Portugal.
- [300] **M. Constant**. Description d’expressions numériques en français. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 119–136, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [301] **M. Constant**. Bibliothèques d’automates finis et grammaires indépendantes du contexte : de nouveaux traitements informatiques. In *RECITAL 2001, V<sup>e</sup> Rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues*, volume 1, pages 431–440, 2001. Tours.
- [302] **M. Constant**. Methods for constructing lexicon-grammar resources : the example of measure expressions. In *3rd Language Resources and Evaluation Conference*, pages 1341–1345. Paris : Elda, 2002. Las Palmas.
- [303] **M. Constant**. On the analysis of locative prepositional phrases : the classifier/proper noun pairing. In E. Ranchhod et N. Mamede, editors, *Proceedings Advances in Natural Language Processing*, pages 33–42. Springer (LNAI 2389), 2002. Third International Conference, PorTAL 2002, Faro, Portugal.
- [304] **M. Constant**. Converting linguistic systems of relational matrices into finite-state transducers. In *Proceedings of the EACL Workshop on Finite-State Methods in Natural Language Processing*, pages 75–82, 2003. Budapest.
- [305] **M. Constant**. GRAAL, une bibliothèque de graphes : mode d’emploi. In *Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues*, pages 321–330, 2004. Actes des quatrièmes et cinquièmes journées INTEX, Bordeaux 2001 et Marseille 2002.
- [306] **M. Constant**. Vers la construction d’une bibliothèque en-ligne de grammaires linguistiques. In S. Kasparian et J. De Finney, editors, *Lexicometrica. Numéro spécial*, 2004.

- Actes du colloque L'analyse de données textuelles : De l'enquête aux corpus littéraires, Québec, 2002.
- [307] **M. Constant** et **A. Yannacopoulou**. Le dictionnaire électronique du grec moderne : conception et développement d'outils pour son enrichissement et sa validation. In *Studies in Greek Linguistics, Proceedings of the 23rd annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2002*, volume II, pages 783–791, 2003.
- [308] **C. Domingues**. Étude de classifieurs pour la construction du dictionnaire de mots composés. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 157–180, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [309] **D.-E. Eum**. Les propriétés des complétives en *-ko* en coréen. In *Linguisticae Investigationes*, volume XXVI :(1), pages 71–96. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2003. Actes du Colloque Grammaires et Lexiques Comparés, Bari, 2002.
- [310] S. Grammenidis et **T. Kyriacopoulou**. Formation des traducteurs : du savoir théorique au savoir-faire professionnel. In *Territoires actuels de la traduction. Actes du Colloque International : Traduire l'Europe. Universitatea Babeş-Bolyai, Cluj-Napoca, 2001*, pages 315–326. Cluj-Napoca : Echinox, 2002.
- [311] **E.-J. Jung**. Quelques grammaires locales sur les expressions de durée en coréen. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 201–218, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [312] **E.-J. Jung**. Analyse des adverbes de temps coréens : autour des noms de partie du jour. In *Cahiers de la MSH Ledoux 1, INTEX pour la linguistique et le traitement automatique des langues*, pages 101–120, 2004. Actes des quatrièmes et cinquièmes journées INTEX, Bordeaux 2001 et Marseille 2002.
- [313] D. Kazantzi et **R. Voskaki**. Désambiguïsation des adjectifs à emploi nominal et des adverbes. In *Actes du 6ème Colloque International de Linguistique Grecque, 2003*. Electronic book.
- [314] D. Kazantzi et **R. Voskaki**. Méthodes de traitement automatique des textes en grec moderne. In *Actes du 13ème Colloque International de la Société Grecque de Linguistique Appliquée : Nouvelles Directions dans la Linguistique Appliquée, Faculté des Lettres, Université Aristote de Thessaloniki, 2003*. À paraître.
- [315] **K. Kosawat**. Procédure de reconnaissance des mots et des phrases thaï. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 241–256, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [316] **T. Kyriacopoulou**. Les bases terminologiques multilingues. In *Actes du Colloque International : Traduction humaine, Traduction automatique, Interprétation*, volume Série Linguistique, No. 11, pages 67–75, 2000. Tunis, 2000.
- [317] **T. Kyriacopoulou**. Problèmes de traduction en traitement automatique du langage naturel (TALN). In *Actes du Symposium International : Traduction, Université Nationale et Kapodistrienne d'Athènes*, volume Série Linguistique, No. 11, pages 231–239, 2000. Athènes, 2000.

- [318] **T. Kyriacopoulou.** Une expérience d'enseignement par visio-centre. In *Actes du 12ème Congrès International : L'apport de l'enseignement/apprentissage de langues dans la défense d'une culture de la paix*, volume 8, pages 279–286, 2001. Thessaloniki, 1999.
- [319] **T. Kyriacopoulou.** Formation universitaire et nouvelles technologies. In *Actes du Colloque International, Médiation des savoirs et des cultures : Quelles perspectives pour les Nouvelles Technologies ? Université d'Athènes*, pages 163–173, 2002.
- [320] **T. Kyriacopoulou.** Un système d'analyse de textes en grec moderne : représentation des mots composés. In *Recherches en linguistique grecque, Actes du 5ème colloque international de linguistique grecque*, volume II, pages 51–54. L'Harmattan, Paris, 2002. Sorbone, 2001.
- [321] **T. Kyriacopoulou, I. Michailidis, et O. Tsaknaki.** Une tentative de traduction semi-automatique des proverbes. In *Actes du Colloque International, Traduire au XXIème siècle : Tendances et Perspectives, Université de Thessaloniki, 2002*, pages 329–335, 2003.
- [322] **T. Kyriacopoulou et O. Tsaknaki.** Représentation automatique des proverbes avec des automates finis. In *Studies in Greek Linguistics, Proceedings of the 23rd annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2002*, volume II, pages 860–871, 2003.
- [323] **T. Kyriacopoulou, R. Voskaki, et A. Yannacopoulou.** Le module grec d'INTEX : état de l'art. In *Actes des sixièmes journées INTEX, Sofia 2003*, 2004. À paraître.
- [324] **É. Laporte.** A Lingüística para o processamento das línguas. In A. Silva et M. Lins, editors, *Recortes Lingüísticos, Vitória, Brésil : Saberes*, pages 67–75, 2000. Conférence prononcée à l'Université fédérale de l'Espírito Santo, avril 2000.
- [325] **É. Laporte.** Une classe d'adjectifs de localisation. In *Cahiers de lexicologie, Adjectifs non prédicatifs. Actes du Colloque, Maison des Sciences de l'Homme, novembre 2002*. Paris : Larousse, 2005. À paraître.
- [326] **C. Leclère.** Du verbe simple aux prédicats complexes. In *Actes du colloque Expressions libres, expressions figées - Hommage à Maurice Gross (15 mai 2002), Università Ca' Foscari Di Venezia, Dipartimento di Scienze del Linguaggio*, 2002. À paraître.
- [327] **C. Leclère.** The lexicon-grammar of French verbs : a syntactic database. In Y. Kawaguchi, T. Zaima, Susumu Takagaki, K. Shibano, et M. Usami, editors, *Proceedings of the First International Conference on Linguistic Informatics*, pages 33–46, 2003. UBLLI, Tokyo University of Foreign Studies.
- [328] **T. Nakamura.** Analysing texts in a specific domain with local grammars : The case of stock exchange market reports. In Y. Kawaguchi, T. Zaima, Susumu Takagaki, K. Shibano, et M. Usami, editors, *Proceedings of the First International Conference on Linguistic Informatics*, pages 109–130, 2003. UBLLI, Tokyo University of Foreign Studies.
- [329] **T. Nakamura.** Analyse automatique d'un discours spécialisé au moyen de grammaires locales. In P. G., F. C., et D. A., editors, *Actes des 7èmes Journées internationales d'analyse statistique des données textuelles*, pages 837–847, 2004.

- 
- [330] P. Panagiotidis et **T. Kyriacopoulou**. Development of the technological skills of translators and interpreters. In *Actes du Colloque International, Traduire au XXIème siècle : Tendances et Perspectives, Université de Thessaloniki, 2002*, pages 407–415, 2003.
- [331] **M. Pantazara**. Le lexique-grammaire des verbes : verbes intransitifs à un complément prépositionnel essentiel. Une première approche : les verbes symétriques. In *Studies in Greek Linguistics, Proceedings of the 21st annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2000*, pages 538–549, 2001.
- [332] **M. Pantazara**. Verbes à complément prépositionnel entrant dans de doubles constructions. In *Studies in Greek Linguistics, Proceedings of the 22nd annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2001*, pages 507–518, 2002.
- [333] **S. Paumier**. Nouvelles méthodes pour la recherche d’expressions dans de grands corpus. In A. Dister, editor, *Revue Informatique et Statistique dans les Sciences humaines*, volume 36, pages 289–296, 2000. Actes des troisièmes journées INTEX, Liège, 2000.
- [334] **S. Paumier**. A time-efficient token representation for parsers. In *Proceedings of the EACL Workshop on Finite-State Methods in Natural Language Processing*, pages 83–90, 2003. Budapest.
- [335] S. Ristov et **É. Laporte**. A method for compressing lexicons. In *2002 Data Compression Conference (DCC 2002), 2-4 April, 2002, Snowbird, UT, USA*. IEEE Computer Society, 2002. Poster Session.
- [336] **A. Savary**. Typographical nearest-neighbour search in a finite-state lexicon and its application to spelling correction. In B. Watson et D. Wood, editors, *Proceedings of the 6th Conference on Implementations and Applications of Automata (CIAA)*, number 2494 in Lecture Notes in Computer Science, pages 251–260. Berlin : Springer Verlag, 2001.
- [337] **A. Savary** et C. Jacquemin. Reducing information variation in text. In S. Renals et G. Grefenstette, editors, *Text- and Speech-Triggered Information Access, Proceedings of TESTIA 2000, 8th ELSNET European Summer School on Language and Speech Communication*, number 2705 in Lecture Notes in Artificial Intelligence, pages 145–181. Heidelberg : Springer-Verlag, 2000.
- [338] **E. Sklavounou**. Les adjectifs du grec moderne en *-menos*. In *Studies in Greek Linguistics, Proceedings of the 20th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 1999*, pages 461–474, 2000.
- [339] I. Thilikos et **R. Voskaki**. Représentation des adjectifs à emploi nominal en grec moderne par des automates à état fini. In *Studies in Greek Linguistics, Proceedings of the 23rd annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2002*, volume II, pages 814–825, 2003.
- [340] **S. Voyatzi**. Reconnaissance automatique des locutions temporelles du grec moderne : l’expression de date et d’heure. In *Studies in Greek Linguistics, Proceedings of the*

- 23rd annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2002*, volume II, pages 803–813, 2003.
- [341] **S. Voyatzi**. Une tentative de traduction semi-automatique des adverbes du grec moderne vers le français. In *Actes du Colloque International, Traduire au XXIème siècle : Tendances et Perspectives, Université de Thessaloniki, 2002*, pages 575–583, 2003.
- [342] **S. Voyatzi**. Description par automates des adverbes de date du grec moderne. In *Actes des sixièmes journées INTEX, Sofia 2003*, 2004. À paraître.
- [343] **S. Voyatzi**. Représentation électronique et reconnaissance automatique des adverbes figés grecs introduits par les indices SA(N), OSAN, OPOS. In *Studies in Greek Linguistics, Proceedings of the 25th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2004*, 2004. À paraître.
- [344] **S. Voyatzi**. Représentation électronique et reconnaissance automatique des déterminants numériques du grec moderne. In *Studies in Greek Linguistics, Proceedings of the 24th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2003*, pages 78–89, 2004.
- [345] **A. Yannacopoulou** et E. Fista. Analyse syntactico-sémantique des verbes préfixés en *kse-* du grec moderne et méthodes de leur représentation. In *Studies in Greek Linguistics, Proceedings of the 25th annual meeting of the Department of Linguistics, Faculty of Philosophy, Aristotle University of Thessaloniki, 2004*, 2004. À paraître.

### Collections, livres et chapitres dans les livres

- [346] A. Anastassiadis-Symeonidis, **E. Sklavounou**, E. Mavrakaki-Polybiou, P. Panayotopoulou-Florou, et A. Simota-Basila. *Dictionnaire inverse du grec moderne*. Thessaloniki : Zitis et Institut d'Études Néohelléniques, Université Aristote de Thessaloniki, 2002. En grec. 745 p.
- [347] I. Antipa, M. Vlavianou, P. Chatzitheoxarous, S. Antonakou, C. Argiropoulou, et **E. Sklavounou**. *Dictionnaire illustré, Mots et Images*. Athènes : Ilektronikes Texnes EPE et Université d'Athènes, 2000. En grec. 277 p.
- [348] **J. Giry-Schneider**. Une construction tronquée du verbe faire : Jean fait le (brave + cachottier + repentant + enfant gâté). In **C. Leclère**, **É. Laporte**, M. Piot, et M. Silberztein, editors, *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*, pages 223–230. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Lingvisticæ Investigationes Supplementa* 24.
- [349] **M. Gross**. Grammaires locales de déterminants nominaux. In X. Blanco, P.-A. Buvet, et Z. Gavriilidou, editors, *Détermination et Formalisation*, pages 177–194. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2001. *Lingvisticæ Investigationes Supplementa* 23.
- [350] **T. Kyriacopoulou**. *Analyse automatique des textes écrits : le cas du grec moderne*. Thessaloniki : University Press, 2004. En français. 153 p.
- [351] **É. Laporte**. Mots et niveau lexical. In J. Pierrel, editor, *Ingénierie des langues*, pages 25–49. Paris : Hermès, 2000. Informatique et systèmes d'information.

- [352] **É. Laporte.** Resolução de ambiguidades. In E. Ranchhod, editor, *Tratamento das Línguas por Computador. Uma introdução à Linguística Computacional e suas aplicações*, pages 44–89. Lisbonne : Caminho, 2001. Une version anglaise est parue dans *Linguisticæ Investigationes*, XXIV(1).
- [353] **É. Laporte.** Préface. In **C. Leclère, É. Laporte**, M. Piot, et M. Silberztein, editors, *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*, pages xi–xxi. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Linguisticæ Investigationes Supplementa* 24.
- [354] **É. Laporte.** Restructuration and the subject of adjectives. In **C. Leclère, É. Laporte**, M. Piot, et M. Silberztein, editors, *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*, pages 373–388. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Linguisticæ Investigationes Supplementa* 24.
- [355] **É. Laporte.** *Symbolic Natural Language Processing*, volume Applied Combinatorics on Words, chapter 3, pages 153–195. Cambridge University Press, 2004. À paraître.
- [356] **É. Laporte**, G. Gross, E. M. Ranchhod, et **C. Leclère**, editors. *Linguisticæ Investigationes*. Amsterdam–Philadelphia : John Benjamins Publishing, 2002-2004. Fondateur Maurice Gross, 350 pages par an, en 2 fascicules.
- [357] **C. Leclère** et J. Brisbois-Leenhardt. Synonymie de mots et synonymie de phrases. In **C. Leclère, É. Laporte**, M. Piot, et M. Silberztein, editors, *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*, pages 389–404. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Linguisticæ Investigationes Supplementa* 24.
- [358] **C. Leclère, É. Laporte**, M. Piot, et M. Silberztein, editors. *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Linguisticæ Investigationes Supplementa* 24.
- [359] **M. Monteleone.** *Lessicografia e dizionari elettronici. Dagli usi linguistici alle basi di dati lessicali*. Fiorentino and New Technology, Napoli, 2002. En italien. 179 p.
- [360] **T. Nakamura.** Publications de Maurice Gross (1963-2002). In **C. Leclère, É. Laporte**, M. Piot, et M. Silberztein, editors, *Lexique, Syntaxe et Lexique-Grammaire. Papers in honor of Maurice Gross*, pages 649–659. Amsterdam–Philadelphia : John Benjamins Publishing Co, 2004. *Linguisticæ Investigationes Supplementa* 24.

## Thèses et habilitations

- [361] **S.-M. Bae.** *Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies - méthodes et applications*. Thèse de doctorat, Université de Marne-la-Vallée, 2002. Jury : **M. Crochemore**, G. Gross, H.K. Kim, **É. Laporte**, D. Maurel.
- [362] **G. Chung.** *Analyse des constructions à double nominatif/accusatif par l'opération de restructuration en coréen : Classsification syntaxique des constructions à adjectifs Sôngsang*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : **D. Arquès**, A. Dugas, **É. Laporte**, D. Le Pesant, M.G. Pak, R. Vivès.

- [363] **M. Constant.** *Grammaires locales pour l'analyse automatique de textes : Méthodes de construction et outils de gestion.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : **M. Crochemore**, **J. Giry-Schneider**, F. Guenthner, **É. Laporte**, D. Maurel.
- [364] **C. Domingues.** *Étude d'outils informatiques et linguistiques pour l'aide à la recherche automatique d'information dans un corpus documentaire.* Thèse de doctorat, Université de Marne-la-Vallée, 2001. Jury : B. Bachimont, M. Bourdeau, P. Lafon, **É. Laporte**, **M. Gross**, M. Silberztein.
- [365] **D.-E. Eum.** *Syntaxe des verbes de communication en coréen.* Thèse de doctorat, Université de Marne-la-Vallée, 2004. Jury : **J. Giry-Schneider**, G. Gross, **T. Kyriacopoulou**, **É. Laporte**, **C. Leclère**, J.S. Nam.
- [366] **S.-H. Han.** *Les prédicats nominaux en coréen. Constructions à verbe support hata.* Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : C. Cortes, **M. Gross**, **J. Giry-Schneider**, **É. Laporte**, I. Tamba.
- [367] **K. Kosawat.** *Méthodes de segmentation et d'analyse automatique de textes thai.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : G. Delouche, **J. Désarménien**, F. Guenthner, **É. Laporte**, D. Maurel.
- [368] **T. Kyriacopoulou.** *Analyse automatique des textes écrits : le cas du grec moderne.* Mémoire d'habilitation à diriger des recherches, Université de Marne-la-Vallée, 2003. Jury : A. Anastassiadis-Symeonidis, F. Guenthner, **É. Laporte**, S. Mejri, **D. Perrin**, E. Ranchhod, T. Symeonidou-Christidou.
- [369] **C. Martineau.** *Compression de textes en langue naturelle.* Thèse de doctorat, Université de Marne-la-Vallée, 2001. Jury : **M. Crochemore**, **É. Laporte**, D. Maurel, G. Plateau, S. Tomi Klein, **M. Zipstein**.
- [370] **M. Monteleone.** *Lexicographie et dictionnaires électroniques. Des usages linguistiques aux bases de données lexicales.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : M. Connena, E. d'Agostino, A. Elia, **É. Laporte**, M. Piot.
- [371] **M. Pantazara.** *Syntaxe dérivationnelle du grec moderne : Les constructions verbales à un complément prépositionnel et les constructions nominales et adjectivales predicatives associées.* Thèse de doctorat, Université de Paris VIII – Saint-Denis, 2003. Jury : A. Anastassiadis-Symeonidis, **C. Leclère**, D. Leeman, M. Piot, R. Vivès.
- [372] **S. Paumier.** *De la reconnaissance de formes linguistiques à l'analyse syntaxique.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : Ch. Choffrut, F. Guenthner, **É. Laporte**, J.S. Nam, **D. Perrin**.
- [373] **A. Savary.** *Recensement et description des mots composés – méthodes et applications.* Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : G. Gross, **M. Gross**, F. Guenthner, J. Humbley, Ch. Jacquemin, **É. Laporte**, M. Silberztein.

## Divers

- [374] **É. Laporte**, **C. Martineau**, et **M. Zipstein.** Compactage des données, 2001. Rapport final, Transweb 2, 12p. + annexes.

- [375] **S. Paumier.** UNITEX - manuel d'utilisation.  
<http://igm.univ-mlv.fr/~unitex/manuelunitex.ps>, 2002.

# 5

## Géométrie discrète et imagerie

### 5.1 Composition

#### Membres

- Gilles Bertrand, professeur ESIEE, responsable
- Lilian Buzer, professeur assistant ESIEE
- Michel Couprie, professeur associé ESIEE
- Yukiko Kenmochi, chargée de recherche CNRS (depuis octobre 2004)
- Laurent Najman, professeur associé ESIEE
- Laurent Perroton, professeur associé ESIEE
- Hugues Talbot, professeur associé ESIEE (depuis novembre 2004)

#### Membre associé

- Éric Incerti, maître de conférences

#### Doctorants

- Cédric Allène, doctorant
- Jean Cousty, doctorant
- Xavier Daragon, doctorant
- André Vital Saude, doctorant

#### Anciens membres

- Francisco Nivando Bezerra, doctorant de 1998 à 2001
- Silvio Jamil Ferzoli Guimarães, doctorant de 2001 à 2003
- Marco Antônio Garcia de Carvalho, doctorant de 2002 à 2003
- Christophe Lohou, doctorant de 1997 à 2001
- Cédric Sibade, doctorant jusqu'en 2003

## 5.2 Thèmes de recherche

Nos activités de recherche/développement durant cette période se regroupent autour de cinq thèmes :

- **Topologie discrète** : étude des transformations homotopiques discrètes et des algorithmes associés, étude des surfaces et variétés discrètes, des transformations qui changent la topologie de façon sélective . . .
- **Topologie discrète et traitement d'images** : transformations topologiques binaires et numériques, visualisation d'images 3D, analyse de séquences d'images, analyse d'images microscopiques, imagerie médicale 3D.
- **Morphologie mathématique et applications** : ligne de partage des eaux et transformations associées, analyse et traitement d'images de documents, indexation et reconnaissance de dessins techniques.
- **Géométrie algorithmique et géométrie discrète** : algorithmes de simplification, reconnaissance de plans discrets, optimisation et géométrie.
- **Parallélisme et imagerie** : segmentation et classification d'images, compression d'images, architectures pour la compression d'images grand format, traitement sur flux compressé.

Dans le proche futur, nous développerons les thèmes « géométrie discrète » et « morphologie mathématique », qui seront renforcés de façon naturelle par l'arrivée de deux nouveaux membres.

- Yukiko Kenmochi, Chargée de Recherche CNRS, sera affectée à notre laboratoire à partir d'octobre 2004. Sa spécialité est la géométrie discrète, en particulier la polyédrisation d'objets 3D, la reconnaissance de plans discrets.
- Hugues Talbot, spécialiste de morphologie mathématique, va rejoindre notre groupe en novembre 2004 en qualité d'enseignant-chercheur ESIEE. Ses centres d'intérêt sont le filtrage et la segmentation d'image, les applications à l'analyse d'images biomédicales et de matériaux.

## 5.3 Résultats

### 5.3.1 Topologie discrète

Une partie de notre travail concerne l'étude des transformations qui préservent la topologie d'un objet (transformations *homotopiques*). À cet effet nous avons proposé des nouvelles caractérisations de la notion de *point simple*, un point simple d'un objet est un point dont la suppression ne modifie pas la topologie de cet objet (1). À partir de cette notion de point simple, nous avons étudié plusieurs algorithmes de *squelettisation* d'un objet. Nous avons également introduit de nouveaux concepts, tels que les *points P-simples*, qui permettent de définir des transformations supprimant en parallèle des points simples d'un objet (en préservant sa topologie). Ceci nous a permis de construire

ce qui est, à notre connaissance, le premier opérateur de *rétraction symétrique* d'un objet.

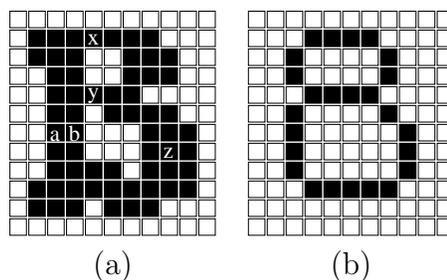


FIG. 1: (a) : Les points (ou pixels)  $x$ ,  $y$  et  $z$  sont non-simples. Les points  $a$  et  $b$  sont tous deux simples : on peut enlever l'un ou l'autre sans modifier la topologie ; cependant on ne peut les enlever tous deux en parallèle sans modifier la topologie. (b) : Si, à partir de l'image (a), on enlève itérativement des points simples, on obtient un *squelette ultime* de (a).

Nous nous sommes intéressés à plusieurs cadres formels permettant de définir un objet discret (*graphes, ordres partiels, complexes simpliciaux*), nous avons étudié un schéma de *discrétisation d'objets*, nous avons proposé plusieurs nouvelles notions de *surface discrète* et mis en évidence des nouvelles propriétés d'une certaine classe de *variétés discrètes* (de dimension quelconque).

Une autre partie de notre travail a consisté à étudier des transformations qui modifient de façon sélective la topologie d'un objet ou d'une image. Un exemple d'une telle transformation est la *fermeture de trous* dans un objet 3D (au sens où un tore plein possède un trou). Nous avons proposé un algorithme qui effectue une telle transformation et qui permet ainsi de réaliser un « filtrage topologique » d'un objet (2). Nous avons également proposé une définition formelle de la notion de *ligne de partage des eaux*. Dans ce contexte, la ligne de partage des eaux d'un objet est une transformation qui préserve une partie des caractéristiques topologiques d'un objet et qui s'autorise à modifier les autres. À notre connaissance il s'agit de la première définition générale de cette notion. Nous avons mis en évidence plusieurs propriétés fondamentales de cette transformation.

### Discrétisation dans les ordres

*Personnes impliquées : Gilles Bertrand, Michel Couprie, Yukiko Kenmochi.*

Parmi les différentes méthodes de discrétisation proposées et étudiées dans la littérature, la *supercouverture* est une des plus naturelles, et possède des propriétés intéressantes. Par contre, une propriété importante lui fait défaut : dans le cadre classique

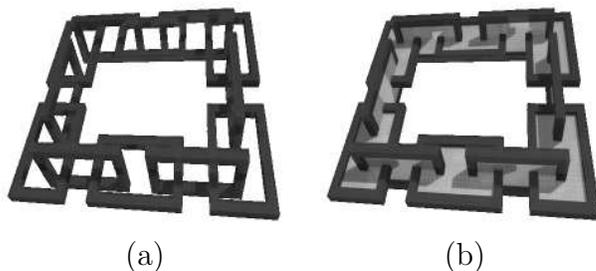


FIG. 2: (a) : Une chaînette (objet 3D) formée de 12 maillons. (b) : Les trous correspondant aux différents maillons ont été fermés automatiquement par notre algorithme.

d'un espace discret constitué de « pixels » (carrés fermés unitaires), la supercouverture d'une droite (resp. d'un plan en 3D) ne constitue pas en général une courbe (resp. une surface) discrète (3).

Nous considérons la généralisation de la supercouverture à un espace discret au sens de P.S. Alexandrov, c'est-à-dire une structure d'ordre. Dans un tel espace, nous avons prouvé que la discrétisation d'un plan de  $\mathbb{R}^3$  est une surface discrète (au sens où le voisinage de chaque point forme une courbe fermée simple, 4), et que la discrétisation de la frontière d'un objet  $X$  de  $\mathbb{R}^3$  suffisamment « régulier » est égale à la frontière de la discrétisation de  $X$ . Cette propriété a des corollaires immédiats pour les demi-espaces et les plans d'une part, et pour les boules et les sphères d'autre part.

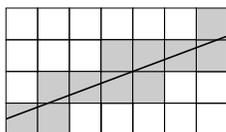


FIG. 3: La supercouverture d'une droite  $D$  est l'ensemble des pixels (carrés fermés unitaires) dont l'intersection avec  $D$  est non vide. À cause des configurations telles que celle ci-dessus, où quatre pixels forment un carré, la supercouverture de certaines droites ne peut être considérée comme une courbe discrète.

*Résultat* : publications [381, 402].

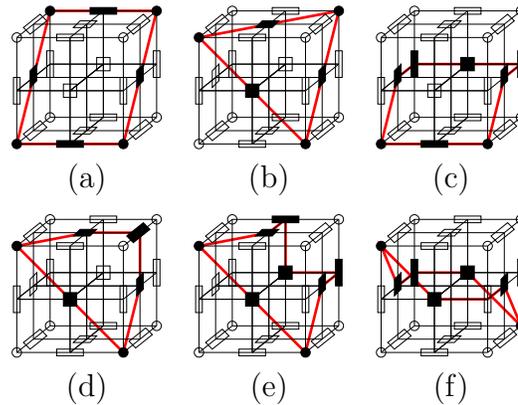


FIG. 4: (a-f) : quelques exemples de configurations du voisinage d'un point dans la discrétisation d'un plan discret. Ce voisinage forme une courbe simple fermée.

### Frontière d'un objet discret

*Personnes impliquées : Gilles Bertrand, Michel Couprie, Xavier Daragon.*

De nombreuses applications, en particulier dans le domaine de l'imagerie médicale, requièrent d'extraire la frontière d'un objet discret. Bien que diverses notions de frontière existent, la plupart fournissent des résultats qui ne sont pas satisfaisants : frontière épaisse, trouée ou ne respectant pas la topologie de l'objet.

Nous proposons une nouvelle définition de la frontière d'un objet discret. Cette définition est valable pour tous les objets discrets pouvant être représentés par un *ordre* (partiel), c'est-à-dire une relation réflexive, transitive et anti-symétrique. La donnée d'un ordre est équivalente à la donnée d'une famille d'ensembles, on voit ainsi que notre définition est très générale puisqu'on considère aussi bien des objets discrets « réguliers » tels que des sous-ensembles d'une maille cubique (ou hyper-cubique), que des objets « irréguliers » tels que des pavages quelconques d'un espace euclidien de dimension arbitraire. Notre cadre est également valable pour les *complexes simpliciaux* (objets triangulés de dimension quelconque) qui peuvent être vus comme un cas particulier d'ordre.

De façon sommaire, si on considère une famille quelconque d'ensembles  $E$  qui correspondent à un espace discret, on définit la *frontière* d'un objet  $X \subset E$  comme étant la famille des ensembles qui touchent à la fois  $X$  et  $E \setminus X$ .

Un des résultats majeurs que nous avons démontré est que la frontière associée à un objet discret quelconque est nécessairement une *variété discrète* (c.a.d. une « surface à  $n$ -dimensions »). Cette propriété indique que la topologie d'un objet et celle de sa frontière - au sens où nous l'avons défini - sont parfaitement cohérentes.

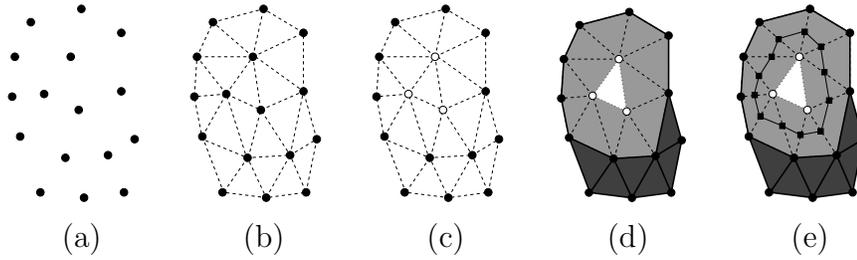


FIG. 5: (a) : Ensemble de points quelconque. (b) : Triangulation de l'espace basée sur ces points, on obtient un complexe simplicial, qui est un ensemble partiellement ordonné. (c) : Un objet est défini comme un sous-ensemble de ces points (points blancs). (d) : Cette partition des points induit une partition du complexe en trois parties. Le sous-complexe objet en blanc, le sous-complexe non objet en noir, et une zone frontalière en gris. (e) : Définition de l'ordre frontière.

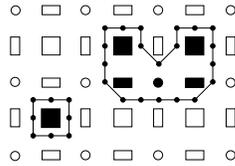


FIG. 6: Exemple d'ordre frontière associé à un objet de  $|H^2|$ , ensemble partiellement ordonné représentant la topologie de  $\mathbb{Z}^2$ .

Résultat : publications [405, 385, 384].

### Points P-simples et algorithmes de squelettisation 3D parallèles

Personnes impliquées : Gilles Bertrand, Christophe Lohou.

Par définition, on peut enlever un point simple sans modifier la topologie d'un objet. Cependant la suppression en parallèle de points simples peut changer la topologie d'un objet (7). En 2 dimensions, certaines approches ont été proposées pour résoudre ce problème : l'une des plus utilisées consiste à classer les points dans quatre catégories, les points de type Nord, Sud, Est, ou Ouest. À chaque itération, seuls les points d'un type donné peuvent être candidats pour la suppression. Cependant cette approche n'est plus valable dans un espace à trois dimensions (7) : en effet, on doit maintenant considérer les six directions Nord, Sud, Est, Ouest, Haut, Bas, et les points  $x$  et  $y$  sont tous deux du type Haut.

À travers la notion de point  $P$ -simple, nous proposons une stratégie générale pour enlever des points en parallèle sans changer la topologie. Cette notion de point  $P$ -simple correspond à une notion d'homotopie forte : un ensemble  $Y$  est fortement homotope à un ensemble  $X$ , si  $Y$  est inclus dans  $X$  et si pour tout  $Z$ , tel que  $Y$  inclus dans  $Z$  et  $Z$  inclus dans  $X$ ,  $Z$  est homotope à  $X$ . Dans ce cas  $P = X \setminus Y$  est constitué de points dits  $P$ -simples (8).

Nous avons proposé une caractérisation des points  $P$ -simples qui peut être effectuée en un temps linéaire. Le problème que nous avons résolu était à priori exponentiel, ce résultat est donc tout à fait inespéré<sup>(1)</sup>.

Récemment, nous avons proposé une nouvelle méthode basée sur les points  $P$ -simples, permettant de construire de nouveaux algorithmes d'amincissement parallèles 3D. À partir d'un algorithme  $A$  existant, nous construisons un nouvel algorithme  $A'$  tel que  $A'$  supprime au moins tous les points supprimés par  $A$ , tout en préservant les mêmes points extrémités.

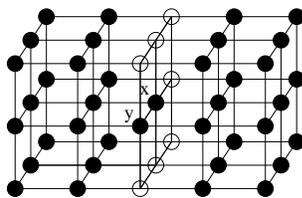


FIG. 7: Un objet composé de deux parallélépipèdes « reliés » par deux points  $x$  et  $y$  : les points  $x$  et  $y$  sont tous deux simples, cependant ils ne peuvent être supprimés en parallèle sans changer la topologie.

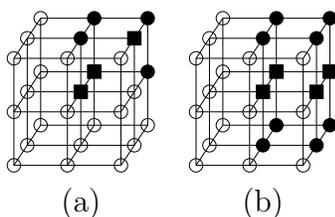


FIG. 8: Les disques noirs représentent les points de  $Y = X \setminus P$ , et les carrés noirs les points de  $P$ . (a) : Le point central est  $P$ -simple, (b) : le point central  $x$  n'est pas  $P$ -simple, car en enlevant certains points de  $P$ , on peut rendre  $x$  non simple.

*Résultat* : publications [426, 397, 427, 390, 391].

<sup>(1)</sup>G.Bertrand : « On  $P$ -simple points », Comptes Rendus Acad. des Sciences, Série Math., Vol. I, No. 321, pp. 1077-1084,1995

### Un algorithme de fermeture de trous 3D

*Personnes impliquées : Zouina Aktouf, Gilles Bertrand, Laurent Perroton.*

Un trou dans un objet en 3D n'est pas une région de l'espace, comme on peut s'en convaincre en considérant un tore plein, qui est un objet comportant un unique trou (voir aussi les exemples de la 9). Intuitivement, la présence d'un trou dans un objet peut être caractérisée par l'existence d'un chemin fermé inclus dans l'objet, et ne pouvant être transformé de façon continue en un point unique, tout en restant dans l'objet.

Fermer des trous dans des objets 3D n'est donc pas un problème trivial, en particulier il n'est pas possible d'utiliser, comme en 2D, des algorithmes de recherche de composantes connexes.

En se basant sur les études que nous avons faites sur la topologie de  $\mathbb{Z}^3$ , nous proposons une méthode pour fermer des trous dans un objet 3D. Cette méthode permet de plus de contrôler par un paramètre la « taille » des trous à fermer (9b). Elle peut être implantée par un algorithme linéaire en temps de calcul. À notre connaissance, il s'agit de la première méthode pour résoudre ce problème.

L'idée de base de la méthode consiste à plonger l'objet  $X$  dont on veut fermer les trous, dans un autre objet  $Y$  connexe et sans trou, tel qu'un cube par exemple. On amincit itérativement  $Y$  en enlevant des points n'appartenant pas à l'objet  $X$ , et en s'assurant que l'enlèvement du point ne crée pas de trou. Ce test est effectué à l'aide des nombres topologiques (voir la section « Points simples et homotopie »). Afin d'obtenir un résultat bien « centré » sur l'objet  $X$ , on guide le processus d'amincissement par une carte de distance à l'objet  $X$ .

*Résultat : publication [376].*

### Mosaïques par connexion

*Personnes impliquées : Gilles Bertrand, Michel Couprie.*

Considérons un ensemble  $R$  de points du plan, composé de  $n$  composantes connexes  $R_1, \dots, R_n$ . À chacune de ces composantes  $R_i$ , on peut associer l'ensemble  $V_i$  des points du plan qui sont strictement plus proches de  $R_i$  que de n'importe quel autre composante de  $R$ . L'ensemble  $V_i$  est nommé *zone d'influence* de  $R_i$ , il s'agit d'une généralisation du concept bien connu de région de Voronoï.

La *transformation par zones d'influence* est l'application qui associe, à tout ensemble  $R$ , l'ensemble  $V$  constitué de l'union des zones d'influences des composantes connexes de  $R$ . Dans le plan discret  $\mathbb{Z}^2$ , muni d'une distance discrète ( $d_4$  ou  $d_8$ ), la transformation par zones d'influence ne préserve aucune propriété topologique, pas même le nombre de composantes connexes.

Nous avons proposé une nouvelle approche basée sur la notion d'ordre. Un ordre est équivalent à un espace topologique discret au sens de P.S. Alexandroff (voir la section « Ordres et topologie numérique »). Dans un tel espace, nous avons défini des transformations qui préservent les composantes connexes : nous les nommons *mosaïques*

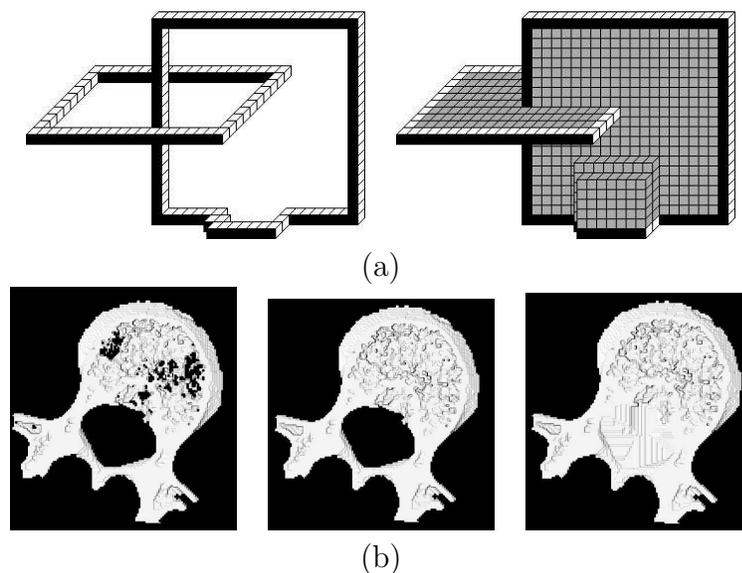


FIG. 9: Illustration d'un algorithme de fermeture de trous dans des objets 3D. (a) : L'utilisation d'une carte de distance permet de « centrer » la position de la surface qui ferme le trou. (b) : Un paramètre permet de régler la « taille » des trous à fermer.

*par connexion.* Nous avons montré que dans un ordre, la transformation par zones d'influences, définie à partir de la notion de plus court chemin et de la distance associée, préserve les composantes connexes de tout ensemble fermé. Nous avons défini, par le biais d'un algorithme parallèle, une mosaïque par connexion particulière qui inclut les zones d'influence et peut donc être qualifiée de « centrée ». De plus, la mosaïque par connexion centrée produit des frontières plus fines que la transformation par zones d'influences. L'extension de cette transformation aux ordres valués peut être vue comme une généralisation de la ligne de partage des eaux.

*Résultat :* publications [379].

### La ligne de partage des eaux

*Personnes impliquées :* Gilles Bertrand.

La ligne de partage des eaux est une notion qui apparaît dès le XIX<sup>ème</sup> siècle avec, notamment, les travaux de J. C. Maxwell. Elle est, depuis une vingtaine d'années, l'objet de nombreuses études motivées par des applications en analyse d'images. De fait, elle constitue l'un des principaux concepts de la morphologie mathématique et, dans ce cadre, elle est considérée comme un des opérateurs les plus puissants permettant de segmenter une image.

Cependant, à notre connaissance, il n'existait pas de cadre formel permettant de

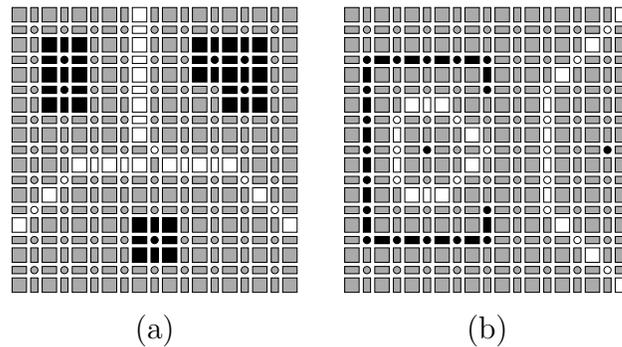


FIG. 10: (a,b) : un ensemble (en noir) et sa mosaïque par connexion centrée (en gris).

disposer d'une définition précise de la ligne de partage des eaux et de mettre en évidence quelques-unes de ses propriétés. Bien au contraire, des propriétés qui nous semblent fondamentales ne sont pas satisfaites par les opérateurs existants. En particulier ces opérateurs ne préservent pas le « contraste » de l'image, les lignes obtenues par ces opérateurs peuvent être mal positionnées...

Le but de ce travail est de montrer qu'un tel cadre existe et que l'approche topologique de la ligne de partage des eaux<sup>(2)</sup> permet non seulement de satisfaire ces propriétés fondamentales mais également d'obtenir plusieurs théorèmes non triviaux. En particulier, nous montrons qu'une fonction  $W$  est une ligne de partage des eaux d'une fonction  $F$  si et seulement si  $W \leq F$  et  $W$  préserve le contraste entre les régions minimales de  $F$ ; le contraste entre deux régions minimales  $M$  et  $M'$  étant défini comme l'altitude minimum à laquelle on est obligé de monter pour aller de  $M$  à  $M'$  (11).

Ce résultat nous paraît d'autant plus intéressant qu'il n'est pas possible d'avoir une telle caractérisation dans le cadre des transformations topologiques « classiques » (transformations homotopiques).

*Résultat* : publications [396, 377], voir aussi [383, 429, 430].

### 5.3.2 Opérateurs topologiques et traitement d'images

L'approche avec laquelle nous abordons l'analyse d'images est une approche typiquement informatique, basée sur des structures discrètes. Plus précisément, nous étudions l'apport des notions de topologie discrète pour concevoir des algorithmes de traitement et d'analyse d'images. Les notions fondamentales de topologie telles que les notions de voisinage, de continuité nous semblent en effet constituer un paradigme qui trouve

<sup>(2)</sup>M. Couprie and G. Bertrand. Topological Grayscale Watershed Transformation. *SPIE Vision Geometry V Proceedings*, Vol. 3168, pp. 136-146, 1997.

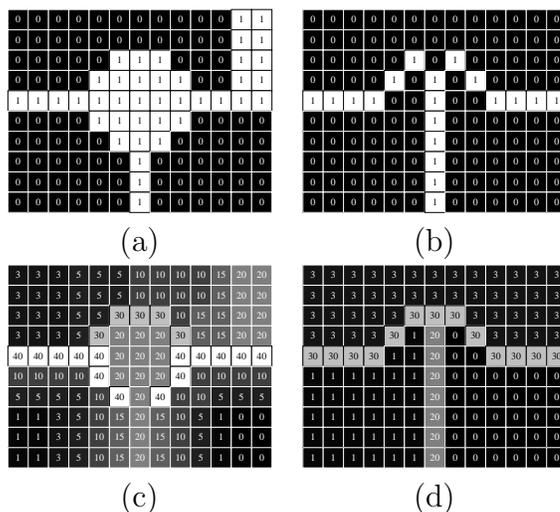


FIG. 11: (a) : un objet binaire  $X$ , (b) : une LPE de  $X$ , (c) : une fonction  $F$  comportant trois régions minimales, (d) : une LPE de  $F$ , le contraste entre les régions minimales de  $F$  est préservé.

toute sa pertinence dans le traitement de données spatiales.

Une partie de notre travail concerne l'étude de transformations d'images binaires qui préservent la topologie d'une image, basées sur la notion de point simple : un point simple d'un objet est un point dont la suppression ne modifie pas la topologie de cet objet (voir la 5.3.1 « Topologie discrète »). Deux images binaires sont dites homotopes si l'on peut passer de l'une à l'autre en enlevant ou en ajoutant séquentiellement des points simples. En particulier, nous proposons une nouvelle méthode pour lisser des objets dans des images binaires à 2 et à 3 dimensions, tout en garantissant que le résultat est homotope à l'image originale.

Nous avons également travaillé à l'introduction de notions topologiques pour l'analyse des images numériques (en niveaux de gris). Notre approche consiste à définir l'équivalence topologique entre deux images numériques à partir de l'équivalence topologique entre les images binaires que constituent les coupes (ou seuils) de l'image numérique aux différents niveaux. Dans ce cadre, deux images en niveaux de gris  $F$  et  $G$  sont dites homotopes si chaque coupe de l'image  $F$  (l'ensemble des points de niveau supérieur à un niveau donné) est homotope, au sens binaire, à la coupe correspondante de l'image  $G$  (12).

Dans des travaux antérieurs, G. Bertrand a introduit la notion de *point destructible* qui généralise la notion de point simple : un point destructible peut être abaissé d'une unité sans changer la topologie d'aucune des coupes. Cette notion permet de concevoir des opérateurs de traitement d'images numériques qui préservent la topologie. Cependant, la préservation stricte de la topologie des coupes peut s'avérer trop contraignante

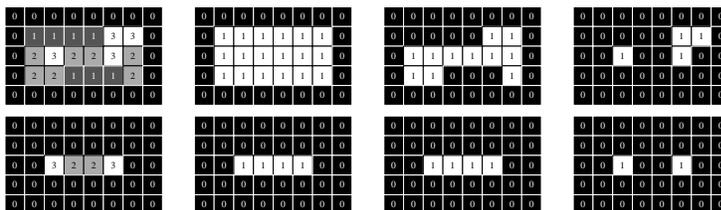


FIG. 12: Première rangée : une image  $F$  et ses coupes aux niveaux 1, 2, 3. Seconde rangée : une image  $G$  et ses coupes aux niveaux 1, 2, 3. Les images  $F$  et  $G$  sont homotopes au sens de la topologie des coupes.

pour certaines applications. C'est pourquoi nous avons introduit des opérateurs permettant de modifier sélectivement la topologie, en se basant sur des critères de contraste local, de taille, ou encore des critères purement topologiques.

Le cadre de la topologie des coupes nous permet de classer les points d'une image selon leurs caractéristiques topologiques locales (pic, col, crête...). Ceci nous a permis de proposer un nouvel opérateur permettant la détection de lignes de crêtes, et le réhaussement de cols étroits dans ces crêtes, avec des applications à l'amélioration de méthodes de segmentations par contours ou par régions.

Nous avons appliqué ces différentes méthodes à l'analyse d'images de cellules biologiques, ainsi qu'à des problèmes d'imagerie médicale 3D. Nous avons également travaillé sur la détection de transitions dans des séquences d'images, ainsi que sur la visualisation d'images 3D.

## Un opérateur de lissage de forme préservant la topologie

*Personnes impliquées : Gilles Bertrand, Michel Couprie.*

L'opération de lissage joue un rôle important pour le traitement d'images et la reconnaissance de formes. Dans le cas des images binaires, on cherche à lisser les contours des objets. Dans tous les travaux précédents, il était supposé que la forme à lisser se composait d'un unique objet, autrement dit, que son contour était une courbe fermée simple (en 2D) ou une surface fermée simple (en 3D). Mais que se passe-t-il si nous voulons appliquer le lissage à une scène composée de plusieurs objets ? En appliquant l'une quelconque des méthodes existantes à chaque objet séparément, puis en fusionnant les résultats, nous n'avons aucune garantie que l'image de deux objets disjoints sera disjointe. Plus généralement, peu d'attention a été accordée aux propriétés topologiques des méthodes de lissage.

Nous proposons une nouvelle méthode pour lisser des objets dans des images binaires à 2 et à 3 dimensions, tout en préservant la topologie. Les objets sont définis par des ensembles de points dans  $\mathbb{Z}^2$  ou  $\mathbb{Z}^3$ , et la préservation de la topologie est assurée par

l'emploi exclusif de transformations homotopiques définies dans le cadre de la topologie digitale. L'action de lissage est obtenue par l'utilisation d'ouvertures et de fermetures morphologiques par des disques ou des boules de rayon croissant, à la manière des filtres séquentiels alternés. Ces filtres morphologiques ne préservent pas la topologie, c'est pourquoi nous avons introduit de nouveaux opérateurs : la coupure homotopique et le remplissage homotopique, qui combinent un effet de filtrage avec la garantie de préserver la topologie ; et le filtre séquentiel alterné homotopique, qui est une composition des deux opérateurs précédents, en prenant comme paramètres des boules de rayon croissant. Ce dernier opérateur prend en entrée une image binaire  $X$  et une image de contrôle  $C$ , et lisse  $X$  autant que possible tout en préservant la topologie de  $X$  et en respectant des contraintes géométriques représentées implicitement par  $C$ . En nous basant sur ces nouveaux opérateurs, nous proposons une méthode générale pour lisser une image binaire, dont l'effet de lissage est contrôlé par un paramètre unique. De plus, de petites variations du paramètre n'entraînent que de petites variations du résultat de la méthode. Nous proposons également une méthode sans paramètre pour lisser des images binaires 2D ou 3D agrandies, tout en préservant la topologie (13).

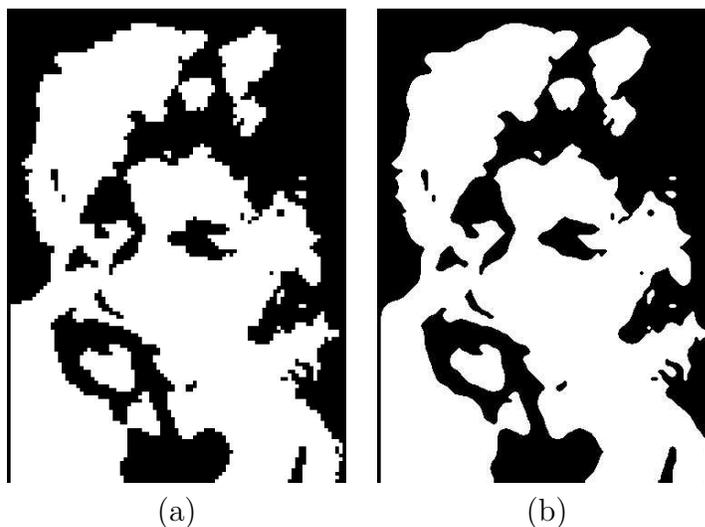


FIG. 13: (a) : image obtenue en agrandissant 4 fois une image plus petite ; (b) : lissage avec préservation de la topologie de (a).

*Résultat* : Un article de revue [380].

### Topologie des images en niveaux de gris

*Personnes impliquées* : Gilles Bertrand, Francisco Nivando Bezerra, Michel Couprie.

Lors de travaux antérieurs, G. Bertrand a étendu les notions de topologie des images binaires 2D à des images en niveaux de gris 2D, c'est-à-dire à des fonctions dont le domaine est  $\mathbb{Z}^2$  (ce travail peut se généraliser aisément aux fonctions dont le domaine est  $\mathbb{Z}^3$ ). Pour cela nous considérons les différentes « coupes » d'une fonction : soit  $f$  une image 2D en niveaux de gris, i.e. une fonction de  $\mathbb{Z}^2$  dans  $\mathbb{Z}$ , nous appelons *coupe de  $f$  au niveau  $k$*  le sous-ensemble de  $\mathbb{Z}^2$  :  $f_k = \{x \in \mathbb{Z}^2, f(x) \geq k\}$ . Nous dirons qu'une transformation « préserve la topologie » de  $f$  si cette transformation « préserve la topologie » (au sens binaire) de toutes les coupes  $f_k, k \in \mathbb{Z}$ . Les notions de *point destructible* et de *point constructible* généralisent, dans ce cadre, la notion usuelle de point simple.

On obtient un *squelette ultime numérique* d'une image, en sélectionnant un point destructible, en abaissant la valeur de ce point jusqu'à une valeur où il n'est plus destructible, et en répétant ces opérations jusqu'à stabilité. Nous avons proposé plusieurs notions de squelettes numériques, et défini des opérateurs de filtrage sur de tels squelettes.

Cependant, la préservation stricte de la topologie des coupes peut s'avérer trop contraignante pour certaines applications. C'est pourquoi nous avons introduit la notion de point  $\lambda$ -destructible, qui nous permet de modifier sélectivement la topologie d'une fonction, sous le contrôle d'un paramètre de contraste local  $\lambda$  (14).

Dans ce même cadre, nous avons également proposé une méthode originale pour filtrer du bruit impulsional, qui donne de meilleurs résultats que les filtres par rang et que des méthodes basées sur les ouvertures morphologiques.

L'anisotropie est un problème lié à l'utilisation de squelettes d'images numériques : le squelette obtenu à partir d'une rotation de l'image originale  $F$  peut être très différent du squelette de  $F$ , soumis à la même rotation. En binaire, l'anisotropie est habituellement minimisée par l'utilisation de la transformation de distance euclidienne. Nous avons proposé une méthode pour minimiser l'anisotropie des squelettes numériques, basée sur une généralisation, au cas des images numériques, de la transformation de distance euclidienne.

*Résultat* : Une thèse [442] et une communication [399].

## Amincissement et restauration de crêtes dans les images en niveaux de gris

*Personnes impliquées* : Gilles Bertrand, Francisco Nivando Bezerra, Michel Couprie.

L'amincissement et la squelettisation sont des applications importantes, en traitement d'image, des notions topologiques. De très nombreux algorithmes d'amincissement pour les images binaires ont été proposés. Cependant, l'utilisation de tels algorithmes suppose une segmentation préalable de l'image, ce qui implique une perte d'information.

Nous proposons un algorithme de squelettisation agissant directement sur une image en niveaux de gris. Cet algorithme garantit la préservation de la topologie des coupes

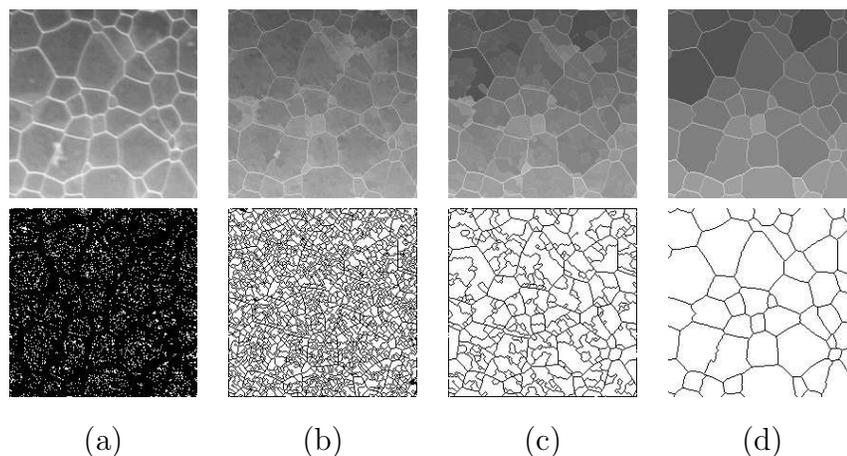


FIG. 14: Première rangée : images, seconde rangée : minima régionaux (en blanc) des images correspondantes. (a) : Image originale. (b) : Squelette ultime numérique de (a). (c) : Squelette filtré de (a), paramètre = 6. (d) : Squelette filtré de (a), paramètre = 23.

de l'image. Nous proposons également une méthode pour « filtrer » de tels squelettes, c'est-à-dire pour simplifier sélectivement la topologie en se basant sur un critère de contraste local (voir également la section 5.3.2).

D'autre part, lorsque l'on rencontre dans des images des objets allongés et minces, il est fréquent que de tels objets soient « coupés » par l'effet du bruit, et l'on souhaiterait pouvoir reconnecter ces objets. Si l'on considère une image comme un relief topographique, un objet « filiforme » peut être vu comme une ligne de crête, et la reconnection de cet objet se ramène à la détection et au réhaussement de cols étroits dans la ligne de crête (15). Le cadre de la topologie des coupes nous permet de classer les points d'une image selon leurs caractéristiques topologiques locales (pic, col, crête ...). Ceci nous a permis de concevoir une méthode efficace et rigoureusement définie pour la restauration de crêtes. Cette méthode s'applique en particulier à des images de gradient, qui par nature sont constituées de lignes minces correspondant aux contours des objets de l'image originale (16).

*Résultat* : un article de revue [382].

### Algorithme de type « Marching Cubes » dans la grille de Khalimsky

*Personnes impliquées* : Gilles Bertrand, Michel Couprie, Xavier Daragon

L'algorithme des « Marching Cubes » (cubes balladeurs) permet de générer, à partir d'une image 3D binaire ou en niveaux de gris définie sur la grille cubique, une surface (sous la forme d'une triangulation) qui correspond à la frontière d'objets présents dans cette image. Son application essentielle est la visualisation de données spatiales. Il doit

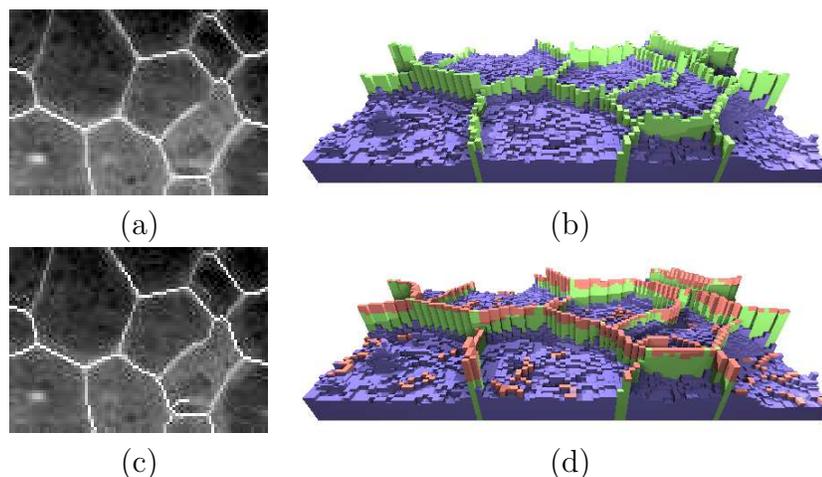


FIG. 15: (a) : Une image en niveaux de gris amincie. (b) : La même image, vue comme un relief topographique. (c,d) : Après restauration de crêtes.

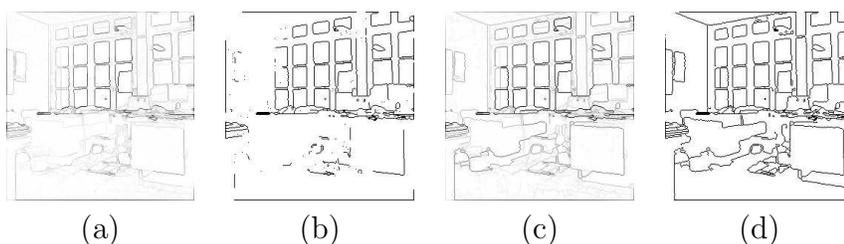


FIG. 16: (a) : Une image de gradient, inversée et amincie. (b) : Image (a) seuillée au niveau 58. (c) : Après la restauration des crêtes appliquée à (a). (d) : Image (c) seuillée au niveau 58.

sa popularité à la grande simplicité de son principe, qui consiste à préparer et à utiliser une table indiquant le morceau de surface devant être associé à un cube unitaire en fonction de l'appartenance de ses sommets à l'objet ou à son complémentaire. Cependant, la version originale de l'algorithme des Marching Cubes avait le défaut de pouvoir générer des surfaces « trouées ». Différents travaux, dont celui présenté ici, ont visé à garantir la topologie des frontières générées par des algorithmes de ce type. L'espace de Khalimsky  $H^n$  est un espace topologique discret construit sur  $\mathbb{Z}^n$ , il permet d'exprimer dans un cadre formel des opérateurs pour le traitement des images digitales et d'en analyser les propriétés topologiques. Il s'agit d'une topologie au sens de P.S. Alexandroff, ou, de manière équivalente, d'un ordre partiel. Cet espace possède de plus la propriété d'être une  $n$ -surface discrète quelle que soit la dimension  $n$ .

Nous avons introduit une notion d'ordre *frontière* associé à un sous-ensemble quel-

conque de  $H^n$  (un « objet »), et démontré que cet ordre frontière possède toujours la propriété d'être une union de  $(n - 1)$ -surfaces disjointes. Ceci nous permet de concevoir un algorithme de type Marching Cubes pour la grille de Khalimsky en dimension quelconque, et de garantir les bonnes propriétés topologiques du résultat. Dans le cas de la dimension 3, il est de plus possible de simplifier la surface obtenue par le biais d'opérateurs garantissant la conservation de ses propriétés topologiques. Cette simplification peut par ailleurs être effectuée sur la table elle-même et n'alourdit donc pas le processus d'extraction de la surface. Les résultats sont alors, du point de vue du nombre de faces, comparables avec ceux d'un algorithme de Marching Cubes standard (17).

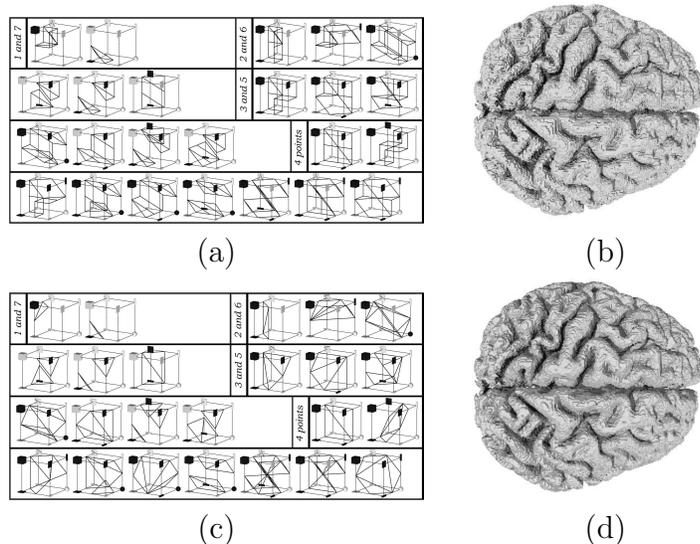


FIG. 17: (a) : Table donnant le morceau de surface devant être généré pour un cube unitaire de  $H^3$ , en fonction de l'appartenance ou non de chaque sommet à l'objet. (b) ; Surface associée à une segmentation du cortex dans  $H^3$ , d'après la table (a). (c) : Table (a) après simplification. (d) : Surface obtenue d'après la table (b).

Résultat : une communication [404].

### Aide au diagnostic de certaines dyskinésies ciliaires

*Personnes impliquées : Michel Couprie, Estelle Escudier (Hôpitaux de Paris, INSERM).*

Cette étude nous a été proposée par le Dr. Estelle Escudier, qui travaille au service d'histologie de l'hôpital de la Pitié-Salpêtrière (Paris) et à l'INSERM; elle se poursuit actuellement dans le cadre d'un projet INSERM/AFM/MR de recherche sur les maladies rares. Plus précisément, nous sommes associés à un réseau pluridisciplinaire sur « Les dyskinesies ciliaires primitives : prises en charge cliniques, diagnostic morphologique et génétique moléculaire », auquel participent des équipes de l'INSERM (U468) et les hôpitaux Henri Mondor (Créteil), Pitié Salpêtrière et Armand Trousseau (Paris). *Résultat* : une publication dans une revue médicale [387], participation à un réseau financé INSERM/AFM/MR de recherche sur les maladies rares, un logiciel de traitement d'images a été fourni, il est actuellement utilisé par des médecins à H. Mondor.

### Extraction des sillons du cortex cérébral à partir de données IRM

*Personnes impliquées : Michel Couprie, Xavier Daragon, Petr Dokládál.*

L'imagerie par résonance magnétique (IRM) permet d'obtenir des images tri-dimensionnelles du cerveau, que l'on peut ensuite segmenter en ses différentes parties (cortex, cervelet, bulbe rachidien ...). Le cortex cérébral possède une structure topologique particulièrement simple, par contre sa géométrie est complexe : on peut le comparer à une feuille de papier (épais) froissée, repliée sur elle-même.

Dans plusieurs applications médicales, il est nécessaire d'extraire avec exactitude les sillons corticaux. Par exemple dans le traitement de formes aiguës d'épilepsie, nécessitant une intervention chirurgicale, l'analyse de ces sillons permet au chirurgien de préparer son opération.

Nous proposons une procédure d'extraction automatique des sillons corticaux à partir de données IRM. Nous privilégions une approche basée sur les propriétés topologiques du cortex : celles-ci étant connues et simples, elles peuvent fournir des indications pour la « reconstruction » des sillons. Notre méthode comporte deux étapes : une croissance homotopique guidée par les niveaux de gris de l'IRM, à partir d'un point, permet de segmenter la matière blanche; ensuite ce processus est poursuivi jusqu'aux limites du cortex, en ajoutant une contrainte liée à l'épaisseur (considérée localement comme constante) de celui-ci.

À l'occasion d'une collaboration avec l'équipe d'I. Bloch de l'ENST, nous avons également contribué à l'amélioration d'une méthode automatique de segmentation du cerveau dans des IRM.

*Résultat* : un article de revue [386], une communication [403].

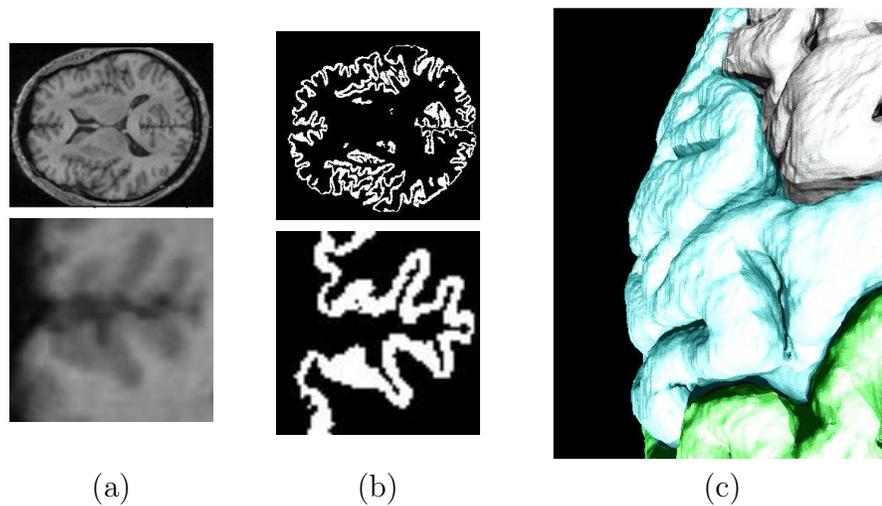


FIG. 18: (a) : Une coupe extraite d'une IRM 3D de la tête (au dessus, un détail). (b) : La coupe correspondante de la zone correspondant au cortex, extraite par notre méthode. (c) : Visualisation par synthèse d'image d'une partie de la surface du cortex.

### 5.3.3 Morphologie mathématique et applications du traitement d'images

L'interprétation d'une image et la reconnaissance de certains objets qui s'y trouvent requièrent généralement deux étapes :

- la première consiste à repérer les structures intéressantes et à les isoler, c'est ce qu'on appelle la segmentation ;
- la deuxième consiste à quantifier ces objets en leur associant des valeurs (nombres ou symboles), en vue de leur classification ou d'une utilisation ultérieure.

Bien qu'elle ne soit pas limitée au traitement des images, la *morphologie mathématique* y trouve un immense champ d'applications. Cette technique repose sur le principe consistant à comparer des structures inconnues (les images que l'on étudie) à un ensemble de formes, les *éléments structurants*, dont on maîtrise les caractéristiques. La comparaison est effectuée au travers de relations booléennes telles l'intersection ou l'inclusion. Nous participons au développement de la morphologie mathématique, tant du côté théorique que du côté algorithmique, ainsi qu'au travers d'applications.

Les ensembles de niveaux d'une image sont les ensembles de pixels au dessus d'un seuil. Les composantes connexes des ensembles de niveaux peuvent être organisées en structure d'arbre au travers de la relation d'inclusion. Un tel arbre est appelé *arbre des composantes*. L'arbre des composantes capture des caractéristiques essentielles de l'image, et permet de caractériser certaines entités topologiques. Cet arbre a été uti-

lisé dans de nombreuses applications, parmi lesquelles nous pouvons citer : filtrage d'images, segmentation d'images fixes et animées, recalage d'images et compression d'images. Nous avons proposé un algorithme quasi-linéaire pour le calcul de l'arbre des composantes.

La ligne de partage des eaux est l'outil fondamental de la morphologie pour la segmentation d'images. Une approche originale, la *ligne de partage des eaux topologique* a été proposée il y a quelques années par M. Couprie et G. Bertrand. Poursuivant cette approche, et afin de permettre son utilisation effective dans la pratique, nous proposons des algorithmes quasi-linéaires pour la ligne de partage des eaux topologique, algorithmes reposant en partie sur l'arbre des composantes. De tels algorithmes sont possibles à construire et à prouver en particulier grâce au formalisme récemment introduit par G. Bertrand. Par ailleurs, nous montrons que la plupart des algorithmes de ligne de partage des eaux existant aujourd'hui ne préservent pas certaines caractéristiques essentielles des images. Seule l'approche topologique de la ligne de partage des eaux garantit des contours satisfaisants.

Combinée aux techniques plus classiques du traitement d'images, la morphologie mathématique permet d'obtenir d'excellents résultats dans les applications. Nous développons plusieurs applications dans le domaine du traitement d'images de documents (estimation morphologique de l'angle d'inclinaison d'un document, indexation et reconnaissance de dessins techniques, utilisation d'une algèbre de composants pour la création automatique de documents intelligents, évaluation des systèmes de reconnaissance de caractères).

Il est à noter que certains des travaux décrits ici ont été en partie réalisés alors que L. Najman travaillait pour Océ Print Logic Technologies. L. Najman, qui a rejoint l'équipe GDI comme professeur associé en novembre 2002, était chercheur associé à ce laboratoire depuis l'année 2000.

## Arbre des composantes

*Personnes impliquées : L. Najman et M. Couprie*

Les ensembles de niveaux d'une image sont les ensembles de pixels au dessus d'un seuil. Les composantes connexes des ensembles de niveaux peuvent être organisées en structure d'arbre au travers de la relation d'inclusion. Un tel arbre est appelé *arbre des composantes*. L'arbre des composantes capture des caractéristiques essentielles de l'image, et permet de caractériser des entités topologiques. Cet arbre a été utilisé dans de nombreuses applications, parmi lesquelles nous pouvons citer : filtrage d'images, segmentation d'images fixes et animées, recalage d'images et compression d'images. Notons aussi que cet arbre est essentiel pour le calcul efficace de la ligne de partage des eaux topologiques.

Plusieurs algorithmes ont été proposés dans la littérature pour calculer l'arbre des composantes. Les plus rapides ont une complexité de calcul en  $O(n \ln(n))$ . Nous avons

proposé un algorithme quasi-linéaire pour calculer cet arbre, basé sur le principe Union-Find développé par Tarjan. De plus, cet algorithme est simple à implémenter.

Décrivons informellement comment nous construisons l'arbre des composantes, en utilisant des références topographiques. Nous pouvons voir l'image comme la surface d'un relief dont l'altitude correspondrait au niveau de gris. Imaginons la surface complètement couverte d'eau, et le niveau d'eau décroissant lentement. Des îles (les maxima) apparaissent. Ces îles forment les feuilles de l'arbre. Au fur et à mesure que le niveau d'eau descend, les îles grandissent, construisant les branches de l'arbre. Quelquefois, à un niveau donné, plusieurs îles fusionnent en un seul morceau connexe. Ces morceaux sont les nœuds de l'arbre. Nous arrêtons quand toute l'eau a disparu. Nous remarquons que nous devons garder trace à la fois des composantes connexes d'un niveau donné et de l'union de ces composantes lors de la variation d'altitude. Aussi, le cœur de notre algorithme consiste à utiliser deux implémentations de l'Union-Find de Tarjan, une pour construire les composantes connexes d'un niveau donné, et le deuxième pour construire leur union suivant l'altitude. Notre algorithme parcourt tous les pixels de l'image une seule fois, en commençant par les plus hauts, les fusionnant avec leurs voisins de même altitude, et construisant l'arbre à partir des parties construites aux niveaux précédents.

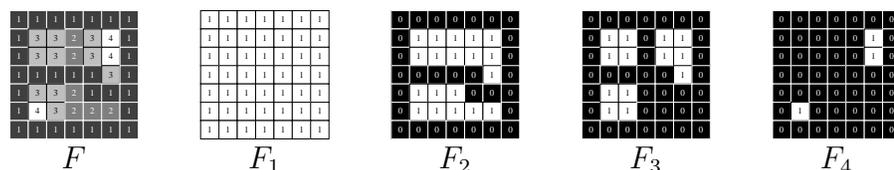


FIG. 19: Une image  $F$  et ses ensembles de niveaux aux niveaux 1,2,3,4

*Résultat* : une communication [430].

### Algorithmes quasi-linéaires pour la ligne de partage des eaux topologique

*Personnes impliquées* : G. Bertrand, M. Couprie, L. Najman.

La ligne de partage des eaux a été introduite en tant qu'outil pour segmenter des images en niveaux de gris par S. Beucher et C. Lantuéjoul vers la fin des années 1970. Désormais cette opérateur intervient comme une étape fondamentale dans de nombreuses méthodes de segmentation.

La transformation nommée *ligne de partage des eaux topologique*<sup>(3)</sup> modifie une fonction (par exemple, une image en niveau de gris) tout en préservant certaines

<sup>(3)</sup>M. Couprie and G. Bertrand. Topological Grayscale Watershed Transformation. *SPIE Vision Geometry V Proceedings*, Vol. 3168, pp. 136-146, 1997.

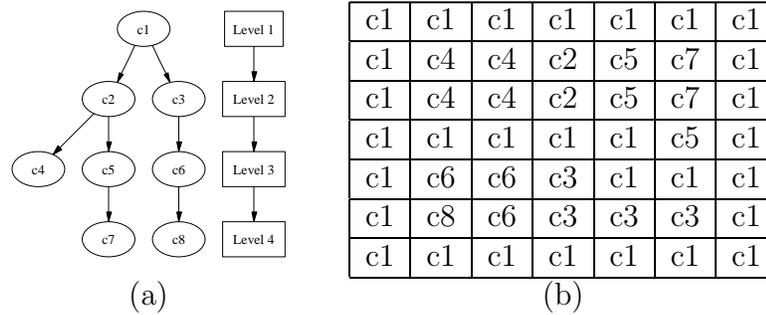


FIG. 20: L'arbre des composantes (a) de l'image  $F$  de la 19 et la carte de correspondance des composantes (b)

caractéristiques topologiques, plus précisément, les composantes connexes de chaque coupe inférieure. Cette transformation a pour effet d'étendre les minima régionaux d'une image, et d'amincir les « crêtes » qui séparent ces minima, qui peuvent être considérées comme les lignes de partage des eaux de l'image originale (21). Le cadre nouvellement développé par G. Bertrand [396, 377] montre que la ligne de partage des eaux topologique possède des propriétés remarquables. Par ailleurs, en comparant cette transformation avec des algorithmes plus classiques, nous avons montré que ces autres algorithmes ne possèdent pas de bonnes propriétés topologiques (voir la section 5.3.3).

Nous avons étudié l'algorithmique de la ligne de partage des eaux et ses liens étroits avec une structure de données fondamentale en analyse d'images nommée arbre des composantes. Un algorithme naïf pour calculer la ligne de partage des eaux topologique pourrait être le suivant : pour tout point  $p$  ( $n$  points), calculer le nombre de composantes connexes de la coupe inférieure au niveau de  $p$  qui sont adjacentes à  $p$  (coût :  $O(n)$  pour chaque point), et abaisser la valeur de  $p$  d'une unité si ce nombre est égal à un. Ceci doit être répété jusqu'à stabilité. Au pire, la boucle externe peut s'exécuter  $\max(n, g)$  fois, où  $g$  représente l'étendue des niveaux de gris, la complexité de cet algorithme est donc en  $O(n^2 \times \max(n, g))$ . Une telle complexité est incompatible avec des applications réelles en traitement d'images.

Nous avons proposé et démontré une caractérisation des points destructibles qui peut être vérifiée localement et en temps constant. Nous obtenons ce résultat en précalculant une structure de données appelée arbre des composantes. Nous avons montré que, grâce à un algorithme original (voir la section 5.3.3), l'arbre des composantes peut être construit en temps quasi-linéaire, au sens de l'*Union-Find* de Tarjan. Nous avons proposé un algorithme quasi-linéaire pour calculer une ligne de partage des eaux topologique. Une variante plus simple a également été introduite lorsque le résultat attendu est une image binaire (l'ensemble des points de la ligne de partage des eaux) et non une fonction.

Il est remarquable que l'arbre des composantes, qui est au cœur de l'algorithmique de la ligne de partage des eaux topologique, soit aussi un outil essentiel pour analyser et filtrer une image : il permet d'implémenter simplement et efficacement des opérations comme la reconstruction géodésique, les  $h$ -minima, les ouvertures par aire, le filtrage par dynamique, les filtres connexes ... qui sont très souvent associées à la ligne de partage des eaux dans les méthodes morphologiques de segmentation d'images.

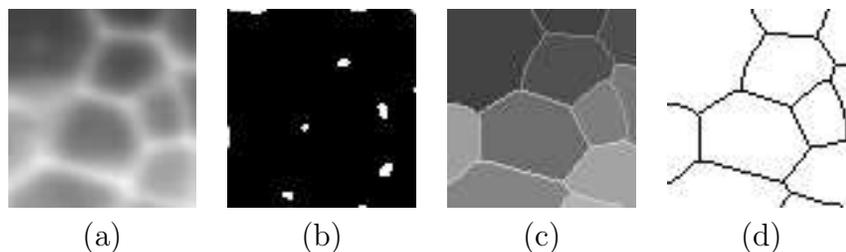


FIG. 21: (a) : image originale, (b) : en blanc : minima régionaux de (a), (c) : ligne de partage des eaux topologique de (a), (d) : en blanc : minima régionaux de (c).

*Résultat* : un article de revue [383].

### Lignes de partage des eaux

*Personnes impliquées* : L. Najman, M. Couprie et G. Bertrand.

Nous avons étudié le comportement des algorithmes de lignes de partage des eaux les plus utilisés dans la pratique : approches basées sur un principe d'inondation (algorithme de Vincent-Soille et algorithme de Meyer) et ligne de partage des eaux topologique. Grâce à l'introduction d'un concept de hauteur de col, nous montrons que la plupart des algorithmes existants ne permettent pas de retrouver des caractéristiques importantes de l'image (en particulier, les points selles ne sont pas correctement calculés). Une conséquence importante de ce résultat est qu'il n'est pas possible de calculer des mesures correctes du volume, de l'aire ou de la hauteur des bassins en utilisant la plupart des algorithmes de ligne de partage des eaux. Un seul principe de ligne de partage des eaux, la *ligne de partage des eaux topologique* garantit des contours corrects.

Renversant le principe d'inondation, nous avons proposé le concept d'*émergence*. Un processus d'émergence est une transformation basée sur un critère topologique dans laquelle les points sont traités par altitude décroissante tout en préservant la connectivité des sections inférieures. Notre résultat principal affirme que toute ligne de partage des eaux par émergence est une ligne de partage des eaux topologique, et plus remarquablement, que toute ligne de partage des eaux topologique d'une image peut être obtenue par un processus d'émergence.

*Résultat* : une communication [429] et un article de revue [392].

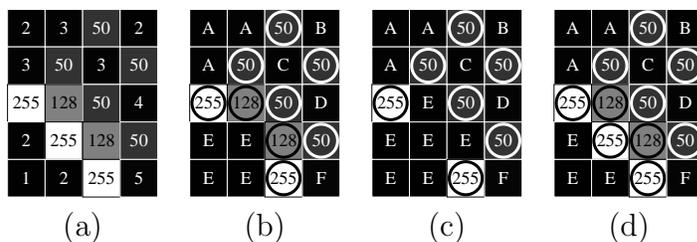


FIG. 22: Contre-exemple à la préservation des hauteurs de cols. Une image en niveau de gris (a) et quelques contours résultats d'algorithmes de ligne de partage des eaux : (b) contours produits par l'algorithme de Vincent-Soille (c) contours produits par l'algorithme de Meyer et (d) contours produits par la lignes de partage des eaux topologique. On peut voir que la hauteur du col entre E et n'importe quel autre bassin est de 50 pour (c) et de 128 pour (b). La ligne de partage des eaux topologique préserve la hauteur de col de 255.

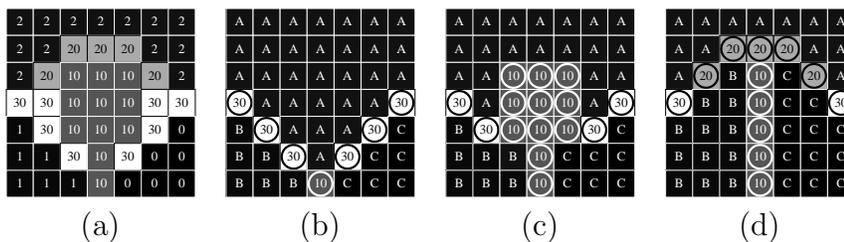


FIG. 23: Un autre contre-exemple à la préservation des hauteurs de col. (a) image originale de type « boutonnière », (b) les contours de la ligne de partage des eaux de Meyer, (c) ceux de l'algorithme de Vincent-Soille, et (d) les contours de la ligne de partage des eaux topologique. On peut noter que le contour d'altitude 20 n'est pas gardé ni par l'algorithme de Vicent-Soille, ni par celui de Meyer. On peut aussi noter que la ligne de partage des eaux topologique préserve les hauteurs de cols de la boutonnière (a).

### 5.3.4 Géométrie algorithmique et géométrie discrète

La géométrie algorithmique est apparue dans les années 70 dans la mouvance de l'étude théorique des programmes informatiques. Cette discipline est aujourd'hui un domaine reconnu principalement grâce à sa très large communauté de chercheurs animant plusieurs revues et de multiples conférences internationales ainsi que de nombreux ouvrages de référence. Son succès s'explique d'une part par la beauté des problèmes et des solutions rencontrées et d'autre part par la richesse et la variété des domaines d'application : informatique graphique, systèmes d'information géographique, robotique, conformation des protéines et tous les autres domaines où la géométrie joue un rôle fondamental. Les premières approches de résolution furent généralement lentes et difficiles à implémenter. Aujourd'hui, ces années de recherche successives ont permis de simplifier et d'améliorer la plupart des approches précédentes.

Les années 60 ont vu l'émergence de la géométrie discrète sous l'impulsion des travaux d'A. Rosenfeld. Ce domaine regroupe des problématiques géométriques se trouvant à la frontière des mathématiques et de l'algorithmique. Il se base sur l'étude d'ensembles de points positionnés sur un maillage régulier. L'imagerie numérique étant basée sur des données à caractère discret, la géométrie discrète a immédiatement trouvé de nombreuses applications en reconnaissance de formes, en vision artificielle et en imagerie médicale.

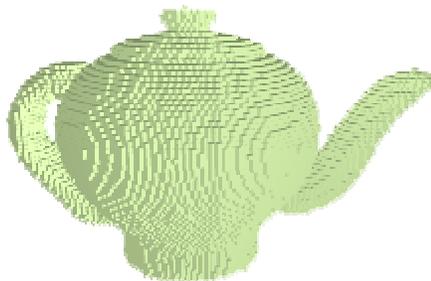


FIG. 24: *Présentation d'un objet discret dans l'espace.*

Ces deux domaines étant tous deux reliés à des notions géométriques, il apparaît entre eux des connexions autant sur le plan de la construction des solutions algorithmiques que du côté des problématiques soulevées. En étudiant chacun des thèmes principaux, il est alors possible de mettre en place des ponts entre ces deux matières. Ainsi certains résultats forts dans un domaine peuvent amener à déverrouiller certains problèmes difficiles de l'autre domaine.

## Reconnaissance des objets discrets

*Personne impliquée : Lilian Buzer.*

De plus en plus la place de l'information digitale s'accroît dans le monde moderne. Les nouvelles technologies d'acquisition : capteurs CCD, Imagerie par Résonance Magnétique... apportent ce type de données. En traitant ces images, nous sommes parfois amenés à vouloir reconnaître des entités géométriques basiques telles que les droites et les plans. Ainsi apparaît le problème de la reconnaissance des droites et des plans discrets : nous cherchons à déterminer si un ensemble de points donnés est une partie d'une droite discrète (ou d'un plan discret).

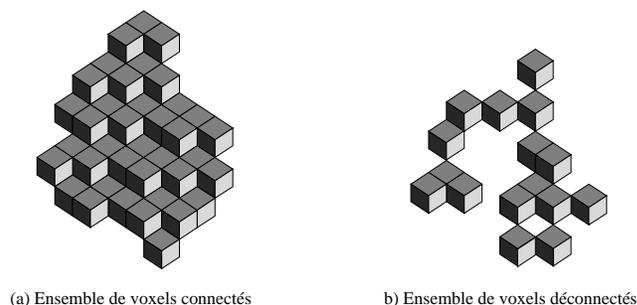


FIG. 25: Reconnaissance de morceaux de plans discrets suivant différents types d'ensemble de points.

Notre algorithme incrémental et optimal de reconnaissance a été construit et démontré dans la thèse de L. Buzer<sup>(4)</sup>. Nous nous sommes ensuite attachés à exposer une liste très exhaustive des méthodes algorithmiques pouvant être utilisées pour la reconnaissance : calcul d'enveloppe convexe, détection d'intersection, calcul d'épaisseur, méthode des cordes, approches algébriquo-géométriques, techniques combinatoires, programmation linéaire et méthodes de décimation. Divers aspects supplémentaires ont été abordés et en particulier le caractère dynamique, incrémental et probabiliste de ces méthodes. L'étude de la reconnaissance des plans discrets est fortement liée à la résolution d'un problème de programmation linéaire. Nous nous attachons maintenant à construire des algorithmes de programmation linéaire spécifiquement adaptés à la résolution de systèmes à faible nombre d'inconnues. Le but visé est d'apporter des techniques efficaces pour la facettisation d'objets voxélisés suivant des familles géométriques plus complexes, par exemple les ellipsoïdes discrets.

*Résultat* : une communication [400], un article de revue [378].

<sup>(4)</sup> *Reconnaissance des plans discrets & Simplification polygonale*, Clermont-Ferrand 2002.

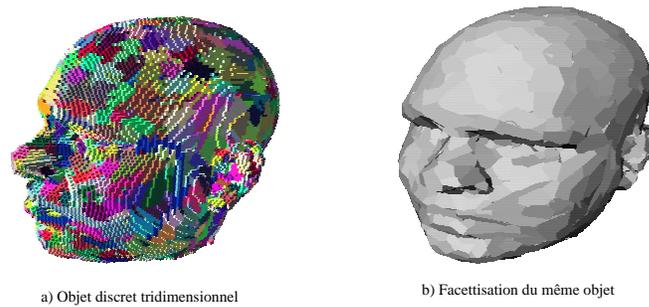


FIG. 26: Exemple de facettisation d'un objet discret à partir de morceaux de plans discrets.

### Simplification polygonale

Personne impliquée : Lilian BUZER.

Les courbes polygonales sont souvent utilisées pour représenter les frontières des entités en cartographie, en graphisme ou en reconnaissance de formes. Simplifier ce type de courbe par une version plus grossière est un problème fondamental étudié dans plusieurs disciplines comme les Systèmes d'Informations Géographiques (S.I.G.), l'analyse des images digitales ou encore la géométrie algorithmique.

Le calcul d'une approximation est toujours lié à des besoins d'efficacité et à une recherche de la préservation de certaines informations topologiques. Nous étudions le problème min-# correspondant à la minimisation du nombre de segments en sortie relativement à un seuil d'erreur fixé. Les algorithmes habituels utilisent des familles géométriques simples (des rectangles par exemple) pour recouvrir la courbe polygonale à simplifier et ainsi obtenir les sommets de la nouvelle courbe.

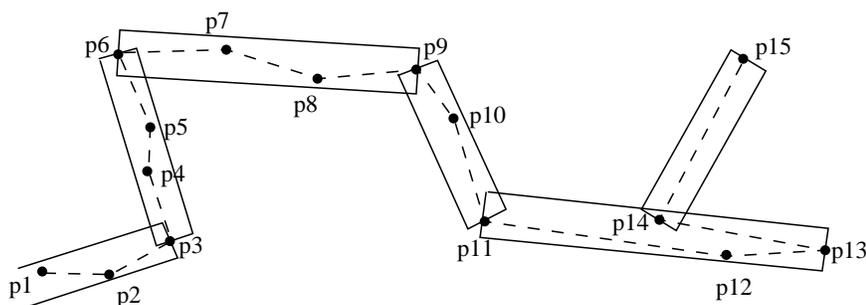


FIG. 27: Exemple de simplification en utilisant une famille de rectangles.

En important une nouvelle famille d'objets géométriques issue du monde du discret, nous sommes arrivés à créer le premier algorithme sous-quadratique de simplification

résolvant le problème min-# de manière optimale. De plus cette méthode garantit que la courbe polygonale obtenue aura son rendu graphique au plus éloigné d'un 1/2 pixel de la courbe polygonale euclidienne d'origine. L'ensemble de ces qualités permet donc de disposer d'un algorithme original et concurrent de l'ensemble des méthodes existantes.

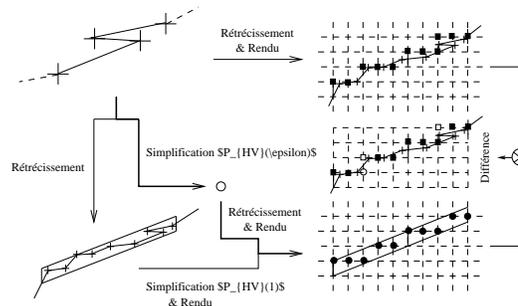


FIG. 28: Critères de qualité pour le rendu de la nouvelle méthode d'approximation.

*Résultat* : article en préparation : A new competitive method for the planar min-# problem.

### 5.3.5 Compression d'image

L'activité la plus importante sur cette période du rapport d'activité est la thèse de C. Sibade sur la compression d'images appliquée aux images grand format et au domaine de l'impression. Cette thèse CIFRE à été effectuée en collaboration avec l'industriel Océ, spécialiste mondial dans le domaine de l'impression grand format. L'objet principal de cette thèse était l'étude de nouveaux algorithmes de compression d'image que l'adéquation entre les algorithmes existants et la problématique spécifique de l'impression grand format. Ce travail a abordé plusieurs points : un état de l'art des algorithmes de compression d'image dans le cadre de l'application de l'impression grand format, la modélisation même des systèmes d'impression afin d'évaluer l'apport de la compression au sein de ces systèmes, et finalement le développement d'algorithmes originaux de traitement de données directement sur flux compressé, notamment sur des données spécifiques au monde de l'impression grand format.

#### Compression d'image appliquée aux documents grand format

*Personnes impliquées* : C. Sibade, L. Perroton, M. Akil & S. Barizien (Océ PLT)

Les systèmes de traitement de documents proposent de multiples fonctionnalités, telles que l'impression, la copie ou la numérisation, auxquelles il est possible d'appliquer

différentes opérations de traitement d'image. Dans le cadre des documents grand format et pour des utilisations productives, ces architectures de traitement de données sont en présence de flux massifs d'image. L'étude de l'évolution de tels systèmes permet d'envisager une augmentation des tailles et des débits de ces flux.

Une étude algorithmique du monde de la compression a été réalisée. Elle présente une large revue des techniques de compression. L'accent est mis sur l'évaluation du comportement de la compression par rapport au grand format et aux types de données spécifiques de l'impression. En effet, différentes profondeurs de couleur et différents contenus sémantiques, depuis les dessins filaires jusqu'aux images photographiques peuvent se trouver et cohabiter sur des images rasters traitées.

Une phase de mesure des performances et du coût fonctionnel pour le système complète cette étude algorithmique. L'évaluation porte sur les paramètres propres aux algorithmes de compression (taux de compression et erreur de reconstruction dans le cas d'une compression avec pertes), sur le temps ou le débit et sur une évaluation de la quantité de mémoire utilisée pour la compression et la décompression.

JPEG2000 est aujourd'hui l'algorithme de compression qui présente le plus de potentialités et qui permet de traiter la plus large gamme d'images. Il possède de plus de nombreuses autres fonctionnalités de parcours (par résolution, par qualité, par position spatiale et par composante couleur), des modes de codage avec ou sans pertes, des codages par régions d'intérêt ou la possibilité d'effectuer des transformations sur le flux compressé. Malgré son lourd coût calculatoire et, plus généralement, son utilisation massive des ressources de la plate-forme qui le supporte, il vise une utilisation très large de la compression.

*Résultat* : une communication [435].

### **Intégration de la compression au sein des architectures de traitement documentaire grand format : modélisation et simulation**

*Personnes impliquées* : C. Sibade, L. Perroton, M. Akil & S. Barizien (Océ PLT)

Parallèlement à l'étude de la compression, le système sur lequel elle doit s'appliquer a été analysé. L'utilisation d'une ou de plusieurs formes de compression d'image raster au sein de la chaîne de traitement pour les systèmes grand format implique un changement de représentation de données et l'ajout d'un ou plusieurs modules de compression et de décompression. Les stratégies d'intégration de ces formats compressés peuvent être *locales* afin de viser une diminution d'un espace de stockage ou de permettre une meilleure utilisation de la bande passante des liens de communication. Nous avons aussi introduit une stratégie d'intégration *globale* dans laquelle les données traversent plusieurs étapes de la chaîne à l'état compressé.

La phase de description des flux de données et des éléments qui les composent a mis à jour le besoin d'une évaluation des performances du système. Une approche spécifique de modélisation comportementale du système logiciel supportant de telles architectures

de traitement de données a été développée. Celle-ci se focalise sur la synchronisation et l'échange des données-images, ainsi que sur le partage des ressources de calcul et de bande passante, pour le transfert et le stockage. Un outil de simulation a été spécifié et implémenté à partir de ce modèle comportemental, il permet d'effectuer des tests grandeur nature de système, en modifiant la représentation des flux de données sous un format compressé.

La modélisation se focalise sur la synchronisation de l'échange de données-images et sur le partage des ressources. Une chaîne de traitement est ainsi décrite sous forme de composants opératifs, communiquant de façon asynchrone par l'intermédiaire de paquets de données (soit une image entière ou une séquence de parties d'image).

*Résultat* : une communication [436]

### Traitement sur flux compressé

*Personnes impliquées* : C. Sibade, L. Perroton, M. Akil & S. Barizien (Océ PLT)

La compression est un changement de représentation de la donnée initiale. Les opérations de traitement d'image sont des transformations appliquées sur une représentation brute de l'image. Pour un chemin de données qui manipule des images compressées, les algorithmes de compression et de décompression servent de « transcodage » entre un format compressé et un format adéquat pour les transformations. Si le nombre d'utilisations de la compression se multiplie, de nombreuses utilisations de ces modules de compression et de décompression peuvent avoir un impact important sur les performances globales du système.

La mise au point de « Traitements d'image sur Flux Compressé » (TFC) va permettre de diminuer le coût de ces transcodages. Il s'agit d'algorithmes de traitement d'image qui consomment et/ou produisent des données compressées. Le but est d'adapter l'opération initiale sur une représentation non-compressée.

Ces TFC ont été initialement développés pour les applications vidéo : la contrainte de productivité qui était prépondérante pour le codage vidéo M-JPEG ou MPEG, qui permettait d'éviter une décompression-transformation-recompression à chaque trame de l'image, est adaptée à notre contrainte de productivité pour les documents grand format. Nous avons décrit un ensemble de méthodes qui permettent de réaliser de nombreuses transformations géométriques et colorimétriques sur des images compressées par des algorithmes de compression de type RLE (Standard télécopie CCITT et HP-RTL), DCT (JPEG) et à base d'ondelettes (JPEG2000).

Dans le contexte particulier du traitement d'image pour l'impression, un TFC original de *halftoning* appliqué aux images JPEG en tons continus a été créé. Cet algorithme, aussi appelé algorithme de production de demi-tons, permet de créer l'illusion visuelle d'une image en teintes continues par un judicieux arrangement des pixels binaires (d'après Ulichney) à partir des données JPEG.

*Résultat* : deux communications [439, 434] et un brevet [458].

## 5.4 Activités

### 5.4.1 Formation doctorale

#### Thèses soutenues

- Petr Dokládál, « Grey-scale image segmentation : a topological approach », thèse en co-tutelle avec l'université de Brno (République Tchèque), soutenue le 31 janvier 2000. Jury : G. Bertrand, Zdeněk Smékal (co-directeurs de thèse), I. Bloch, J.M. Chassery (rapporteurs), D. Arquès, J. Jiří (examineurs).
- Christophe Lohou, « Contribution à l'analyse topologique des images : étude d'algorithmes de squelettisation pour images 2D et 3D, selon une approche topologie digitale ou topologie discrète », thèse soutenue le 20/12/2001. Jury : G. Bertrand (directeur de thèse), R. Malgouyres et C. Ronse (rapporteurs), D. Arquès et A. Manzanera (examineurs).
- Francisco Nivando Bezerra, « Opérateurs topologiques pour le traitement d'images en niveaux de gris », thèse soutenue le 29/11/2001. Jury : M. Couprie (directeur de thèse), I. Bloch et F. Prêteux (rapporteurs), D. Arquès, F. Meyer et S. Philipp (examineurs).
- Silvio Jamil Ferzoli Guimarães, « Segmentation et indexation sur des séquences d'images ». Thèse en co-tutelle, soutenue en mars 2003. Jury : M. Couprie (co-directeur de thèse), A. de Albuquerque Araújo (co-directeur de thèse), S. Philipp-Foliguet et R. Lotufo (rapporteurs), N.J. Leite (examineur).
- Cédric Sibade, « Compression d'images grands formats en vue d'améliorer la productivité des plates-formes d'impression », thèse en convention CIFRE avec la société Océ Print Logic Technologies soutenue le 16/12/2003. Jury : M. Akil (directeur de thèse), L. Perroton (co-directeur), P. Garda, M. Kunt (rapporteurs), M. Crochemore, S. Barizien (examineurs)

#### Thèses en cours

- Xavier Daragon, en thèse depuis octobre 2000, « Algorithmique dans les ordres, application à l'analyse topologique des images ». Directeur de thèse : M. Couprie
- Marco Antônio Garcia de Carvalho, en thèse depuis 1999, « Indexation et mise en correspondance d'images basées sur l'arbre des composantes ». Directeur de thèse : Roberto de Alencar Lotufo, Université de Campinas, Brésil. Co-directeur de thèse : M. Couprie

## Diplôme d'Études Approfondies

Nous assurons le cours *traitement topologique des images* (G. Bertrand et M. Couprie) du DEA *Informatique Fondamentale et Applications* (co-habilitation ENPC, ESIEE, UMLV, Paris VII).

## Jurys de thèse

Nous avons participé à de nombreux jurys de thèse, à l'ENST, l'ENSMP, et dans les universités de Paris VI, Paris VII, Grenoble, Caen, Clermont-Ferrand, Strasbourg, Lyon, Marne-la-Vallée.

## 5.4.2 Participation à la vie scientifique

### Appartenance à des comités scientifiques de conférences

- SPIE Vision Geometry'2001, San Diego, California, USA (G. Bertrand).
- International Workshop on Combinatorial Image Analysis-IWCIA'2001, Philadelphia, USA (G. Bertrand).
- Discrete Geometry for Computer Imagery-DGCI'2002, Bordeaux (G. Bertrand).
- SPIE Vision Geometry'2002, Seattle, California, USA (G. Bertrand).
- Discrete Geometry for Computer Imagery-DGCI'2003, Naples, Italie (G. Bertrand, M. Couprie).
- SIBGRAP'01, Florianópolis, Brésil (M. Couprie).
- Image and Vision Computing New Zealand (IVCNZ'01), Dunedin, New Zealand (M. Couprie).
- SIBGRAP'02, Fortaleza, Brésil (M. Couprie).
- Image and Vision Computing New Zealand (IVCNZ'02), Auckland, New Zealand (M. Couprie).
- SIBGRAP'03, São Carlos, Brésil (M. Couprie).
- Image and Vision Computing New Zealand (IVCNZ'03), Palmerston North, New Zealand (M. Couprie).
- Discrete Geometry for Computer Imagery (DGCI'2003), Naples (M. Couprie).
- Première Conférence Fédérative sur le Document en Langue Française CFD'02, Hammamet-Tunisie, (L. Najman).
- Congrès International Francophone sur l'Écrit et le Document (CIFED 2002) (L. Najman).
- Conférence Internationale sur l'Analyse Multivoque, le Contrôle, la Viabilité et les Applications, en l'honneur de Jean-Pierre Aubin (21-25 juin 2004, Roscoff, France) (L. Najman).

### Journaux scientifiques

Nous participons de façon régulière à l'évaluation d'articles; parmi les journaux, citons :

- Pattern Recognition Letters (G. Bertrand, M. Couprie);
- Theoretical Computer Science (G. Bertrand, M. Couprie);
- Discrete Applied Mathematics (G. Bertrand, M. Couprie);
- Image and Vision Computing (M. Couprie);
- Computer Graphics, Vision and Image Processing (G. Bertrand, L. Perroton);
- Computer Vision and Image Understanding (L. Najman);
- Journal of Mathematical Imaging and Vision (G. Bertrand, L. Perroton);
- IEEE Pattern Analysis and Machine Intelligence (G. Bertrand, M. Couprie, L. Najman);
- Graphical Models (G. Bertrand);
- Optical Engineering (L. Najman).

### Conférences invitées

- G. Bertrand, « On P-simple points », *Workshop on topology*, New-York, USA, 2002.
- G. Bertrand, « Three-dimensional parallel thinning algorithms and P-simple points », *Conf. Denis Richard*, Clermont-Ferrand, France, 2002.
- G. Bertrand, « Topologie discrète et imagerie bio-médicale », *Journées Biosciences*, Créteil, France, 2002.
- M. Couprie, « Digital topology and cross-section topology for grayscale image processing », tutorial session, *SIBGRAP'01*, Florianópolis, Brésil, 2001.

### Organisation de conférences

- Co-organisation, du 26 juin au 3 juillet 2001, de la première École d'Été Franco-Nordique de Mathématiques (G. Bertrand), Erken, Suède.
- Co-organisation (avec Ch. Ronse et E. Decencière) du prochain ISMM 2005 (International Symposium on Mathematical Morphology) (L. Najman).

### Animation scientifique

- Participation à l'AS CNRS-STIC « Géométrie discrète pour l'analyse spatio-temporelle d'images » (A2SI-ESIEE, CMM-ENSMP, ERIC-Univ. Lyon 2, IRCOM SIC-Univ. Poitiers, LABRI-Univ. Bordeaux, LERI-Univ. Reims, LIS-Univ. Grenoble) (G. Bertrand, M. Couprie).
- Bonus Qualité Recherche (BQR) de l'université de Créteil « Calcul scientifique pour la segmentation et la fusion d'images biologiques et médicales » (M. Cou-

prie).

- Membre du GRCE (Groupe de Recherche en Communication Ecrite) (L. Najman).
- Mise en place d'une politique de rénovation des ouvrages informatiques de la bibliothèque du groupe ESIEE (L. Buzer).
- Présentation de thèmes de recherche spécifiques pour l'AS-CNRS géométrie algorithmique et géométrie discrète de janvier 2004 (L. Buzer).
- Organisation du second séminaire de l'AS-CNRS géométrie algorithmique et géométrie discrète de septembre 2004 (L. Buzer, M. Couprie).

### 5.4.3 Coopérations

- Université de Reims, Laboratoire d'Informatique : Laurent Lucas ; représentation d'objets 3D.
- École Nationale des Ponts et Chaussées – CERGRENE : Jean-Marie Mouchel ; reconnaissance de bactéries pour l'analyse de la qualité des eaux fluviales.
- École Nationale Supérieure des Télécommunications (ENST) : Isabelle Bloch ; opérateurs morphologiques et topologiques pour la segmentation d'images.
- Hôpital de la Pitié-Salpêtrière (Paris) : Estelle Escudier ; analyse d'images médicales, aide au diagnostic.
- Université de Campinas (Brésil) : Neucimar J. Leite ; analyse d'images.
- Université de Campinas (Brésil) : Roberto de Alencar Lotufo ; analyse d'images.
- Université Fédérale du Minas Gerais (Belo Horizonte, Brésil) : Arnaldo de Albuquerque Araújo ; analyse d'images de matériaux.
- Océ Print Logic Technologies (2000-2002) L. Najman a été détaché de la société Océ comme chercheur associé dans l'équipe GDI une journée par semaine.
- Laboratoire Lorrain en Informatique et ses applications, (LORIA, UMR 7503) : Karl Tombre et Bart Lamiroy (équipe QGAR) ; indexation automatique de documents.

### 5.4.4 Contrat

- INSERM (2002-2003) : participation à un réseau pour l'étude des maladies rares. Logiciel d'aide au diagnostic de certaines dyskinésies ciliaires (M. Couprie).

## 5.5 Références bibliographiques

### Articles de revues

- [376] Z. Aktouf, **G. Bertrand**, et **L. Perroton**. A three-dimensional holes closing algorithm. *Pattern Recognition Letters*, 23 :523–531, 2002.

- 
- [377] **G. Bertrand.** On topological watersheds. *Journal of Mathematical Imaging and Vision*, 2004. À paraître (voir aussi : IGM2004-10).
- [378] **L. Buzer.** A linear incremental algorithm for naive and standard digital lines and planes recognition. *Graphical Models*, 65(1-3) :61–76, 2003.
- [379] **M. Couprie et G. Bertrand.** Tesselations by connection. *Pattern Recognition Letters*, 23 :637–647, 2002.
- [380] **M. Couprie et G. Bertrand.** Topology preserving alternating sequential filter for smoothing 2d and 3d objects. *Journal of Electronic Imaging*, 2004. À paraître.
- [381] **M. Couprie, G. Bertrand, et Y. Kenmochi.** Discretization in 2d and 3d orders. *Graphical Models*, 65(1-3) :77–91, 2003.
- [382] **M. Couprie, F. N. Bezerra, et G. Bertrand.** Topological operators for grayscale image processing. *Journal of Electronic Imaging*, 10(4) :1003–1015, 2001.
- [383] **M. Couprie, L. Najman, et G. Bertrand.** Quasi-linear algorithms for the topological watershed. *Journal of Mathematical Imaging and Vision*, 2004. À paraître (voir aussi : IGM2004-11).
- [384] **X. Daragon, M. Couprie, et G. Bertrand.** Derived neighborhoods and frontier orders. *Discrete and Applied Mathematics*, 2004. À paraître (voir aussi : IGM2004-8).
- [385] **X. Daragon, M. Couprie, et G. Bertrand.** Discrete surfaces and frontier orders. *Journal of Mathematical Imaging and Vision*, 2004. À paraître (voir aussi : IGM2004-5).
- [386] **P. Dokládal, I. Bloch, M. Couprie, D. Ruijters, R. Urtasun, et L. Garnero.** Segmentation of 3d head MR images using morphological reconstruction under constraints and automatic selection of markers. *Pattern Recognition*, 36 :2463–2478, 2003.
- [387] E. Escudier, **M. Couprie**, B. Duriez, F. Roudot-Thoraval, M. Millepied, V. Prulière-Escabasse, L. Labatte, et A. Coste. Computer-assisted analysis helps to detect inner dynein arm abnormalities. *American Journal of Respiratory and Critical Care Medicine*, 166(9) :1257–1262, 2002.
- [388] **S. Guimarães, M. Couprie, A. de Albuquerque Araújo, et N. Leite.** Video segmentation based on 2d image analysis. *Pattern Recognition Letters*, 24(7) :947–957, 2002.
- [389] **S. Guimarães, N. Leite, M. Couprie, et A. de Albuquerque Araújo.** Flat zone analysis and a sharpening operation for gradual transition detection on video images. *EURASIP Journal on Applied Signal Processing*, 2004. À paraître.
- [390] **C. Lohou et G. Bertrand.** A 3d 12-subiteration thinning algorithm based on p-simple points. *Discrete and Applied Mathematics*, 139 :171–195, 2004.
- [391] **C. Lohou et G. Bertrand.** A 3d 6-subiteration curve thinning algorithm based on p-simple points. *Discrete and Applied Mathematics*, 2004. À paraître.
- [392] **L. Najman, M. Couprie, et G. Bertrand.** Watersheds, mosaics and the emergence paradigm. *Discrete and Applied Mathematics*, 2004. À paraître (voir aussi : IGM2004-4).

## Actes de conférences

- [393] M. Akil, **L. Perroton**, S. Gailhard, J. Denoulet, et F. Bartier. Architecture for hardware compression / decompression of large images. In *SPIE Electronic Imaging*, 2001.
- [394] S. Alayrangues, **X. Daragon**, J.-O. Lachaud, et P. Lienhardt. équivalence des n-g-cartes fermées sans multi-incidence et des n-surfaces. In *actes du Groupe de Travail en Modélisation Géométrique (GDR ALP)*, pages 35–46, mars 2004.
- [395] A. Araújo, **L. Perroton**, R. Oliveira, L. Claudino, S. Guimarães, et E. Bastos. Non-linear features extraction applied to pollen grain images. In E. Dougherty et J. Astola, editors, *IS&T/SPIE Electronic Imaging*, volume 4304, pages 35–45, San Jose-CA USA, January 2001. SPIE, SPIE.
- [396] **G. Bertrand**. Some properties of topological greyscale watersheds. In *procs. SPIE Vision Geometry XII*, volume 5300, pages 182–191, 2004.
- [397] **G. Bertrand** et **C. Lohou**. Three dimensional parallel thinning algorithms based on p-simple points. In *Denis Richard's birthday conference, Clermont-Ferrand*, 2002.
- [398] **F. N. Bezerra** et **M. Couprie**. Reducing anisotropy of topological operators for grayscale images. In *SPIE Vision Geometry IX*, volume 4117, pages 46–57, 2000.
- [399] **F. N. Bezerra** et **M. Couprie**. Réduction d'anisotropie des squelettes en niveaux de gris. In *procs. congrès RFIA*, volume 3, pages 819–828, 2002.
- [400] **L. Buzer**. An incremental linear time algorithm for digital line and plane recognition using a linear incremental feasibility problem. In *Discrete Geometry for Computer Imagery : 10th International Conference, DGCI 2002, Proceedings*, volume 2301 of *Lecture Notes in Computer Science*, pages 372–381. Springer, 2002.
- [401] **M. Couprie** et **G. Bertrand**. Tesselations by connection in orders. In *Discrete geometry for computer imagery*, volume 1953 of *Lecture Notes in Computer Science*, pages 15–26. Springer, 2000.
- [402] **M. Couprie**, **G. Bertrand**, et **Y. Kenmochi**. Discretization in 2d and 3d orders. In *Digital and Image Geometry, LNCS, Springer Verlag*, volume 2301, pages 301–312, 2002.
- [403] **X. Daragon** et **M. Couprie**. Segmentation topologique du neo-cortex cérébral depuis des données IRM. In *procs. congrès RFIA*, volume 3, pages 809–818, 2002.
- [404] **X. Daragon**, **M. Couprie**, et **G. Bertrand**. Marching chains algorithm for Alexandroff-Khalimsky spaces. In *SPIE Vision Geometry XI*, volume 4794, pages 51–62, 2002.
- [405] **X. Daragon**, **M. Couprie**, et **G. Bertrand**. Discrete frontiers. In *DGCI, LNCS, Springer Verlag*, volume 2886, pages 236–245, 2003.
- [406] A. de Albuquerque Araújo, **L. Perroton**, R. Oliveira, L. Claudino, **S. Guimarães**, et E. Bastos. Non-linear features extraction applied to pollen grain images. In *SPIE Electronic Imaging*, volume 4304, pages 35–45, 2001.
- [407] **M. G. de Carvalho**, R. Lotufo, et **M. Couprie**. Spatiotemporal segmentation of MR image sequence based on hierarchical analysis. In *procs. ISSPA*, 2003. À paraître.

- 
- [408] C. Doublier, **M. Couprie**, J. Garot, et Y. Hamam. Computer assisted segmentation, quantification and visualization of an infarcted myocardium from MRI images. In *procs. BiomedSim'03*, pages 151–156, 2003.
- [409] D. Dulac, S. Guezguez, et **G. Bertrand**. Parallel segmentation based on topology with the associative net model. In *Computer Architecture for Machine Perception*, pages 95–104, 2000.
- [410] C. Fares, Y. Hamam, **M. Couprie**, R. El-Abyad, et A. Abche. Virtual arthroscopic surgery trainer: a virtual reality based training system for arthroscopic surgery. In *procs. BiomedSim'03*, pages 146–151, 2003.
- [411] **S. Guimarães**, **M. Couprie**, N. Leite, et A. de Albuquerque Araújo. A new method for cut detection based on visual rhythm. In *procs. SIBGRAP'01*, pages 297–304, 2001.
- [412] **S. Guimarães**, **M. Couprie**, N. Leite, et A. de Albuquerque Araújo. Video fade detection by discrete line identification. In *procs. ICPR*, volume 2, pages 1013–1016, 2002.
- [413] **S. Guimarães**, A. de Albuquerque Araújo, **M. Couprie**, et N. Leite. An approach to detect video transitions based on mathematical morphology. In *procs. ICIP*, 2003. electronic edition.
- [414] **S. Guimarães**, N. Leite, **M. Couprie**, N. Leite, et A. de Albuquerque Araújo. A directional and parametrized algorithm to gradual transition detection. In *procs. SIBGRAP'02*, pages 261–268, 2002.
- [415] **Y. Kenmochi** et A. Imiya. Naive planes as discrete combinatorial surfaces. In *Discrete geometry for computer imagery*, volume 1953 of *Lecture Notes in Computer Science*, pages 249–261. Springer, 2000.
- [416] **Y. Kenmochi** et A. Imiya. Discrete polyhedrization of a lattice point set. In *Digital and Image Geometry, LNCS, Springer Verlag*, volume 2243, pages 148–160, 2001.
- [417] **Y. Kenmochi** et A. Imiya. Polyhedral set operations for 3d discrete object deformation. In *International Conference on Scale-Space, LNCS, Springer Verlag*, volume 2106, pages 417–425, 2001.
- [418] **Y. Kenmochi**, A. Imiya, T. Nomura, et K. Kotani. Extraction of topological features from sequential volume data. In *International Workshop on Visual Form, LNCS, Springer Verlag*, volume 2059, pages 333–345, 2001.
- [419] **Y. Kenmochi**, C. Li, et K. Kotani. Flatness analysis of three-dimensional images for global polyhedrization. In *Computer Analysis of Images and Patterns, LNCS, Springer Verlag*, volume 2124, pages 482–492, 2001.
- [420] H. Köhler, **M. Couprie**, S. Bouattour, et D. Paulus. Extraction and analysis of coronary tree from single X-ray angiographies. In *procs. SPIE Medical Imaging*, 2004. to appear.
- [421] B. Lamiroy et **L. Najman**. Scan-to-XML : Using software component algebra for intelligent document generation. In D. Blostein et Y.-B. Kwon, editors, *GREC, LNCS, Springer Verlag*, volume 2390, pages 211–221, 2002.

- 
- [422] B. Lamiroy, **L. Najman**, R. Ehrhard, C. Louis, F. Quelain, N. Rouyer, et N. Zegache. Scan-to-XML for vector graphics : an experimental setup for intelligent browsable document generation. In *Fourth IAPR International Workshop on Graphics Recognition (GREC)*, pages 312–321, 2001.
- [423] J. Lecoq, **L. Najman**, O. Gibot, et E. Trupin. Benchmarking commercial OCR engines. In *Sixth International Conference on Document Analysis and Recognition (ICDAR)*, pages 138–142, 2001.
- [424] **C. Lohou** et **G. Bertrand**. New parallel thinning algorithms for 2d grayscale images. In *SPIE Vision Geometry IX*, volume 4117, pages 58–69, 2000.
- [425] **C. Lohou** et **G. Bertrand**. Nouvel algorithme de squelettisation parallèle d’images binaires 2d. In *12e congrès RFIA*, volume 1, pages 493–504, 2000.
- [426] **C. Lohou** et **G. Bertrand**. A new 3d 12-subiteration thinning algorithm based on p-simple points. In *procs. IWCIA’01, Electronic Notes in Theoretical Computer Science*, volume 46, 2001.
- [427] **C. Lohou** et **G. Bertrand**. A new 3d 6-subiteration thinning algorithm based on p-simple points. In *Digital and Image Geometry, LNCS, Springer Verlag*, volume 2301, pages 102–113, 2002.
- [428] **L. Najman**. Using mathematical morphology for document skew estimation. In *procs. SPIE Document Recognition and Retrieval XI*, volume 5296, pages 182–191, 2004.
- [429] **L. Najman** et **M. Couprie**. Watershed algorithms and contrast preservation. In *DGCI, LNCS, Springer Verlag*, volume 2886, pages 62–71, 2003.
- [430] **L. Najman** et **M. Couprie**. Quasi-linear algorithm for the component tree. In *procs. SPIE Vision Geometry XII*, volume 5300, pages 98–107, 2004.
- [431] **L. Najman**, O. Gibot, et M. Barbey. Automatic title block location in technical drawings. In *Fourth IAPR International Workshop on Graphics Recognition (GREC)*, 2001.
- [432] **L. Najman**, O. Gibot, et S. Berche. Indexing technical drawings using title block structure recognition. In *Sixth International Conference on Document Analysis and Recognition (ICDAR)*, pages 587–591, 2001.
- [433] S. Prévost, L. Lucas, **G. Bertrand**, et **M. Couprie**. A multiresolution model of implicit skeleton shapes. In *procs. ICCVG*, volume 2, pages 602–609, 2002.
- [434] **C. Sibade**, M. Akil, **L. Perroton**, et S. Barizien. Stratégie d’application de traitement d’image sur des flux compressés. In *Actes de la conférence CORESA 2003 (COmpression et REprésentation des Signaux Audiovisuels)*, pages 215–218, 2003.
- [435] **C. Sibade**, S. Barizien, M. Akil, et **L. Perroton**. Wide format image manipulation and compression in a printing environment. In *Proceedings of SPIE ITCOM 2002 conference on Multimedia Systems and Applications V*, 2002.
- [436] **C. Sibade**, S. Barizien, M. Akil, et **L. Perroton**. Wide format raster compression applied to a printing environment. In *Proceedings of IS&T NIP 18 conference (Digital Printing Technologies conference)*, 2002.

- [437] **C. Sibade**, S. Barizien, M. Akil, et **L. Perroton**. Evaluation of raster image compression in the context of large format document processing. In *Procs. of IS&T/SPIE Electronic Imaging, session Color Imaging IX : Processing, Hardcopy, and Applications IX*, 2004.
- [438] **C. Sibade**, S. Barizien, M. Akil, et **L. Perroton**. Halftoning processing on a JPEG-compressed image. In *Procs. of IS&T/SPIE Electronic Imaging, session Color Imaging IX : Processing, Hardcopy, and Applications IX*, 2004.
- [439] **C. Sibade**, **L. Perroton**, M. Akil, et S. Barizien. Compressed-domain processing for wide format printing application. In *Proceedings of ICSHPP conference (International Congress on High Speed Photography and Photonics)*, pages 230–235, 2002.

## Collections, livres et chapitres dans les livres

- [440] **G. Bertrand**, A. Imiya, et R. Klette, editors. *Digital and Image Geometry*, volume 2243. Springer Verlag, 2001.
- [441] **G. Bertrand** et R. Malgouyres, editors. *Discrete geometry for computer imagery*, volume 283-1. Elsevier, 2002.

## Thèses et habilitations

- [442] **F. N. Bezerra**. *Opérateurs topologiques pour le traitement d'images en niveaux de gris*. Thèse de doctorat, Université de Marne-la-Vallée (France), 2001.
- [443] **M. Couprie**. *Notions et opérateurs topologiques pour le traitement d'images*. Mémoire d'habilitation à diriger des recherches, Université de Marne-la-Vallée, 2004. Jury : E. Andres, **G. Bertrand**, J.-M. Chassery, **M. Crochemore**, R. Malgouyres, **D. Perrin**, C. Ronse.
- [444] **M. G. de Carvalho**. *Hierarchical Image Analysis through the Tree of Critical Lakes*. Thèse de doctorat, Universidade Estadual de Campinas (Brésil), 2004.
- [445] **P. Dokládal**. *Grey-scale image segmentation : a topological approach*. Thèse de doctorat, Université de Marne-la-Vallée (France) and University of technology of Brno (Czech Republic), 2000.
- [446] **S. Guimarães**. *Video transition identification based on 2D image analysis*. Thèse de doctorat, Université Fédérale du Minas Gerais (Brésil), Université de Marne-la-Vallée (France), 2003.
- [447] **C. Lohou**. *Contribution à l'analyse topologique des images : étude d'algorithmes de squelettisation pour images 2D et 3D, selon une approche topologie digitale ou topologie discrète*. Thèse de doctorat, Université de Marne-la-Vallée (France), 2001.
- [448] **C. Sibade**. *Compression de données pour les systèmes de traitement de document grand format*. Thèse de doctorat, Université de Marne-la-Vallée, dec 2003.

**Divers**

- [449] S. Berche et **L. Najman**. A method for recognizing and indexing documents. European Patent Application EP 1 136 938 A1, Sept. 2001. Application number 01200872.8.
- [450] **G. Bertrand**. On topological watersheds. Technical Report IGM2004-10, Université de Marne-la-Vallée, 2004.
- [451] **M. Couprie, L. Najman, et G. Bertrand**. Quasi-linear algorithms for the topological watershed. Technical Report IGM2004-11, Université de Marne-la-Vallée, 2004.
- [452] **M. Couprie** et R. Zrour. Discrete bisector function and euclidean skeleton in 2d and 3d. Technical Report IGM2004-12, Université de Marne-la-Vallée, 2004.
- [453] **X. Daragon, M. Couprie, et G. Bertrand**. Derived neighborhoods and frontier orders. Technical Report IGM2004-08, Université de Marne-la-Vallée, 2004.
- [454] **X. Daragon, M. Couprie, et G. Bertrand**. Discrete surfaces and frontier orders. Technical Report IGM2004-05, Université de Marne-la-Vallée, 2004.
- [455] **L. Najman**. Skew detection. European Patent, 2002. Filled at August 27, 2002 as a European filing the French Patent Office.
- [456] **L. Najman** et M. Barbey. Automatic table location in documents. European Patent Application EP 1 237 115 A2, Sept. 2002. Application number 02075780.3.
- [457] **L. Najman, M. Couprie, et G. Bertrand**. Watersheds, extension maps and the emergence paradigm. Technical Report IGM2004-04, Université de Marne-la-Vallée, 2004.
- [458] **C. Sibade** et S. Berche. Method and apparatus for generating a halftoned image from a compressed image, 2003.

# 6

## Signal et communications

### 6.1 Composition

#### Membres

- Antoine Chevreuil, maître de conférences
- Philippe Loubaton, professeur
- Jean-Christophe Pesquet, professeur, responsable
- Christophe Vignat, maître de conférences

#### Doctorants

- Marc Castella, doctorant, allocataire normalien
- Caroline Chaux, doctorante, allocataire de recherche
- Julien Dumont, doctorant CIFRE
- Jérôme Gauthier, doctorant CIFRE
- Pierre Jallon, doctorant, allocataire de recherche
- Wided Miled, doctorante, boursière INRIA
- Belkacem Mouhouche, doctorant CIFRE
- Teodora Petrisor, doctorante, boursière GET

#### Anciens membres

- Samson Lasaulce, doctorant jusqu'en 2001
- Mérouane Debbah, doctorant jusqu'en 2002
- Jean-Marie Chaufray, doctorant jusqu'en 2002
- Sébastien Houcke, doctorant jusqu'en 2002
- Pascal Bianchi, doctorant jusqu'en 2003

## 6.2 Thèmes de recherche

Les thématiques de l'équipe concernent les communications numériques, la séparation de sources, la théorie de l'information, et l'analyse en ondelettes 2D.

### Communications numériques

Le premier volet de nos activités a concerné l'étude des récepteurs pour les systèmes CDMA et MC-CDMA (estimation de canal, récepteurs linéaires), ainsi que l'évaluation de leurs performances. Nous nous sommes en particulier intéressés au cas de systèmes de grandes tailles à codes aléatoires. Dans ce contexte, nous avons utilisé diverses techniques provenant de la théorie des matrices aléatoires de grande taille, et obtenu des résultats significatifs. Nous avons par ailleurs approfondi notre expertise des problèmes liés à la surveillance du spectre radio-électrique. Cette thématique, assez peu explorée dans la communauté, a permis la mise en évidence de problèmes peu conventionnels de traitement statistique du signal. Nos travaux ont bénéficié d'un soutien conséquent de la DGA et de Thalès Communication via trois allocations DGA-CNRS et deux contrats.

### Séparation de sources

Nos travaux ont principalement concerné les techniques de séparation de mélanges convolutifs de sources par maximisation de fonctions de contrastes. Alors que la très grande majorité des travaux existants s'intéressent à des signaux sources obtenus par filtrage de suites indépendantes et identiquement distribuées, nous avons considéré le cas de signaux stationnaires non linéaires, mais aussi cyclostationnaires du fait de leur pertinence dans le contexte de l'écoute passive.

### Théorie de l'information

Nos travaux ont porté sur les inégalités entropiques et sur une famille d'entropies généralisées. Nous avons notamment obtenu une extension de l'inégalité de la puissance entropique à une famille de lois discrètes ; une généralisation des inégalités sur les informations de Fisher dans les systèmes linéaires a été étudiée. Nous avons entrepris l'étude des entropies de Rényi et Tsallis en caractérisant leurs lois maximisantes et leur applicabilité au problème de l'égalisation aveugle.

### Analyse en ondelettes 2D

Des recherches sont menées sur divers développements récents de la théorie des ondelettes en vue de la restauration et de la compression d'images. Les champs d'applications envisagés sont variés : imagerie multispectrale, analyse d'images sismiques (en liaison avec l'Institut Français du Pétrole), compression vidéo,... Nos contributions portent aussi bien sur la conception de trames d'ondelettes appropriées aux traitements considérés que sur la définition de nouveaux estimateurs non linéaires tirant profit des spécificités des décompositions espace-échelles. Parmi les estimateurs étudiés, mention-

nons des approches d'optimisation convexe sous contraintes, lesquelles sont calculées de manière adaptative, en fonction des données.

En termes de perspectives, chacune des thématiques présentées plus haut est suffisamment riche pour pouvoir être poursuivie dans le cadre du prochain plan quadriennal. Nous comptons toutefois investir davantage le domaine des grandes matrices aléatoires en abordant des problèmes de traitement statistique des signaux multivariés liés à nos problématiques dans le cas, en pratique très fréquent, où la dimension des vecteurs observés est du même ordre de grandeur que le nombre d'observations. Dans ce contexte, certaines techniques d'estimation paramétrique usuelles fournissent des estimateurs non consistants, et doivent donc être repensées. Bien que certains résultats soient déjà disponibles<sup>(1)</sup>, il reste un travail très important à accomplir afin de disposer d'outils statistiques bien adaptés à nos contextes applicatifs.

## 6.3 Résultats

### 6.3.1 Communications numériques

**Membres permanents :** A. Chevreuil, Ph. Loubaton.

**Doctorants participant ou ayant participé :** P. Bianchi, J.M. Chaufray, Ph. Ciblat, M. Debbah, J. Dumont, W. Hachem, S. Houcke, P. Jallon, S. Lasaulce, B. Mouchouche.

Du fait de l'explosion du secteur des télécommunications, les problèmes liés à la conception de la couche physique des systèmes de communication numériques ont connu depuis une dizaine d'années un grand regain d'intérêt. Nous développons donc depuis cette date des méthodologies avancées de traitement statistique du signal appliquées au domaine des communications numériques. Après avoir effectué divers travaux relatifs à l'estimation et à l'égalisation aveugles de canaux de transmission (voir par exemple [473, 472, 496, 471] pour les articles les plus récents), nos travaux dans la période 2001-2004 ont concerné d'une part des problèmes d'estimation aveugle peu conventionnels rencontrés dans le contexte de la surveillance du spectre radio-électrique, et d'autre part la conception et l'étude des performances de systèmes à accès multiples à répartition par les codes (CDMA) dont l'importance est liée au développement des systèmes de communication mobiles de troisième génération (CDMA 2000, UMTS) et de quatrième génération.

---

<sup>(1)</sup>V.L. Girko, "An introduction to Statistical Analysis of Random Arrays", VSP, The Netherlands, 1998.

**Estimation aveugle pour la surveillance du spectre radio-électrique.**

Les recherches développées dans le cadre de ce thème sont motivées par des problèmes applicatifs rencontrés dans la surveillance du spectre radio-électrique. Dans ce contexte, il convient de détecter un ou plusieurs émetteurs actifs dans une bande de fréquence donnée, et d'obtenir le plus d'informations possible sur leurs caractéristiques techniques. Nous nous intéressons en particulier à l'estimation de paramètres techniques tels que le rythme des symboles, la valeur de la fréquence porteuse, l'indice de modulation dans le cas de modulations de fréquence à phase continue, ou la nature de la constellation utilisée par les symboles. Cet axe de recherche est soutenu par la DGA (3 thèses financées par des allocations DGA-CNRS [563, 566, 560] ont été soutenues entre 2000 et 2003, et une quatrième thèse de ce type est en cours) et par Thalès-Communications (2 contrats entre 2001 et 2003 liés à la thèse de P. Bianchi).

Lorsque le signal reçu est modulé linéairement par un train de symboles, le rythme symbole est le plus souvent estimé en remarquant qu'il coïncide avec la plus petite fréquence cyclique du signal reçu. Il est donc possible de l'estimer en maximisant dans le domaine cyclique la norme d'un vecteur construit à partir de coefficients de corrélations cycliques estimés. Nous avons étudié en détail les propriétés asymptotiques des estimateurs de ce type (consistance, normalité asymptotique) en remarquant qu'ils sont définis en maximisant un périodogramme, et en adaptant les résultats connus en la matière à notre contexte un peu particulier. Nous avons également évalué les variances asymptotiques des estimateurs de façon à mettre en évidence l'impact de paramètres tels que le nombre de corrélations cycliques prises en compte sur leurs performances [475]. Ces techniques ont été adaptées au contexte de l'estimation de la porteuse du signal reçu dans [474].

Les approches cycliques sont réputées pour être peu performantes dans le cas, en pratique très fréquent, où le signal reçu a un très faible excès de bande. Nous avons donc développé (thèse de S. Houcke [566]) une approche complètement différente consistant à chercher à estimer conjointement le débit symbole et extraire les symboles transmis. Pour ceci, nous échantillons le signal reçu à un rythme variable, et adaptons sur les diverses versions échantillonnées un égaliseur aveugle maximisant une fonction de contraste. Nous avons alors établi que la période symbole coïncide avec la période d'échantillonnage pour laquelle le critère de contraste mesuré en sortie de l'égaliseur est maximum [486], et proposé un algorithme très performant, mais relativement complexe, permettant de mettre en œuvre cette idée. Cette technique a également été étudiée dans un cadre multi-sources [487].

Enfin, nous avons étudié des problèmes du même type dans le cadre spécifique des modulations de fréquences à phase continues (modulations CPM, thèse de P. Bianchi,

[560]), peu abordé dans la littérature. Dans ce contexte, le problème le plus délicat consiste à estimer un paramètre appelé indice de modulation. Pour ceci, nous avons proposé une approche basée sur le fait que l'inverse de l'indice coïncide avec la plus petite puissance à laquelle il faut élever le signal reçu afin d'y faire apparaître une composante sinusoidale. Afin de se faire une idée claire des performances de cette approche, nous avons entrepris une étude des propriétés asymptotiques de l'estimateur qui fait apparaître un comportement très inhabituel (loi limite non gaussienne, vitesse de convergence en l'inverse de la fenêtre d'observation, [466]). Cette idée a ensuite été utilisée pour mettre en évidence un estimateur conjoint de l'indice, du résidu de porteuse, et de la période symbole [513]. Les paramètres de la modulation étant estimés, nous nous sommes intéressés à la restitution des symboles transmis par l'émetteur. Lorsque la réception du signal est perturbée par un canal de propagation sélectif en fréquence, il convient de compenser l'effet du canal par un égaliseur aveugle. Nous avons donc considéré le problème de l'égalisation aveugle de signaux CPM, qui en dépit de son importance dans le contexte de l'écoute passive, n'avait pas été beaucoup étudié par le passé. Les signaux CPM étant de module 1, nous avons tout naturellement étudié le comportement de l'algorithme du module constant, et montré que contrairement au contexte des modulations linéaires, son utilisation ne permettait pas nécessairement de compenser le canal [512].

A l'exception du travail [487], nous nous sommes intéressés au cas où la bande de fréquence analysée ne contient qu'un seul émetteur. Dans le cas où plusieurs émetteurs différents interfèrent dans la bande, les techniques que nous avons développées ne fonctionnent pas. Lorsque le récepteur dispose de plusieurs capteurs, il est potentiellement possible de séparer spatialement les signaux qu'ils transmettent afin de se ramener au cas mono-émetteur. Ce point fait l'objet de la thèse de P. Jallon, et sera développé dans le paragraphe consacré à nos travaux en séparation de sources.

### **Conception de récepteurs et évaluation de performances pour les systèmes CDMA**

L'accès multiple à répartition par les codes (CDMA) est un mode d'accès multiple dans lequel on fait transiter à tout instant tous les utilisateurs du système (en fait d'une même cellule) dans la même bande de fréquence. Les signaux convoyés par les différents émetteurs sont multipliés par des fonctions orthogonales entre elles, et peuvent être séparés au niveau du récepteur en effectuant des produits scalaires pertinents. Cette vision d'un récepteur CDMA est cependant quelque peu idyllique car l'orthogonalité entre les signaux correspondants à chaque utilisateur n'est en pratique jamais assurée, d'une part en raison de l'asynchronisme inévitable existant entre ces signaux dans le cas des communications montantes (communications des mobiles à destination d'un récepteur situé au niveau d'une station de base), et d'autre part du

fait que même dans le cas synchrone, la présence d'un canal de transmission dispersif entre émetteur(s) et récepteur(s) détruit l'orthogonalité. Il convient donc d'étudier des récepteurs plus sophistiqués, nécessitant l'estimation et la compensation des canaux de transmission entre les divers émetteurs et le récepteur, et dans chaque type d'application, de mettre en évidence le bon compromis performances / complexité. Ces problèmes sont évidemment similaires à ceux que l'on rencontre dans les systèmes mono-utilisateurs, mais ils sont considérablement compliqués par le caractère multi-utilisateurs du système. L'importance de cette thématique est renforcée actuellement par le développement du système de troisième génération UMTS qui est basé sur le CDMA. Nos travaux ont en particulier fait l'objet de collaborations et contrats industriels : les thèses de S. Lasaulce [567] et M. Debbah [564] ont été financées ou co-financées par le centre de recherche de Motorola de Saint-Aubin, la thèse de J.M. Chaufray [562] a été financée par Thalès-Communications, et celles de B. Mouhouche et J. Dumont font l'objet d'une convention Cifre avec la société Wavecom et FranceTelecom Recherche et Développement, respectivement. Nous avons par ailleurs participé au projet RNRT DOLIE piloté par la Sagem (2000-2002) destiné à étudier la faisabilité du CDMA pour les transmissions haut débit sur le réseau électrique, ainsi qu'au projet IST du 5ème PCRD ANTIUM piloté par Thalès-Communications (2001-2003) au sein duquel nous avons développé des algorithmes permettant d'étudier finement les scénarios de brouillage dans les futurs réseaux UMTS.

En termes de résultats académiques, nos travaux ont plus particulièrement été consacrés au problème de l'estimation de canal, et à celui de l'application de techniques issues du domaine des grandes matrices aléatoires à l'évaluation des performances des différents récepteurs.

**Estimation de canal dans les systèmes CDMA** L'une des causes limitant le plus les performances d'un système CDMA est la présence d'un canal de propagation à trajets multiples entre l'émetteur et le récepteur. Afin de pallier la perte d'orthogonalité entre codes que le canal produit, il convient de l'estimer et de compenser son effet. Pour ceci, l'émetteur transmet des symboles pilotes grâce auxquels l'équivalent temps discret du canal peut être estimé de façon basique. En pratique, les performances de cet estimateur conventionnel sont insuffisantes dès que la charge du système est quelque peu conséquente. Nous avons donc tout d'abord cherché à mettre en évidence des estimateurs plus performants utilisant conjointement les symboles pilotes et le signal reçu correspondant à la transmission de données inconnues [488]. L'utilisation de ce type d'approche, dite semi-aveugle, ne peut toutefois être envisagée que dans le cas d'un système CDMA dont les codes sont invariants dans le temps. Puisque cette condition n'est pas respectée dans le contexte de systèmes importants tels que IS-95 ou l'UMTS, nous avons considéré une toute autre approche basée sur l'observation bien connue que

la connaissance des propriétés statistiques du second ordre du canal permet, grâce à un filtrage de Wiener ou une procédure de réduction de rang, d'améliorer les performances de l'estimateur conventionnel. Nous nous sommes donc intéressés au problème de l'estimation consistante des statistiques du canal. La plupart des travaux existants proposent pour cela d'estimer la matrice de covariance du canal par la matrice de covariance empirique de l'estimateur conventionnel. Cependant, cet estimateur n'est pas consistant, et nous avons montré comment la présence d'un code de scrambling pouvait être utilisée pour mettre en évidence un estimateur consistant [470].

**Matrices aléatoires et évaluation des performances de récepteurs** Le second volet de nos recherches actuelles concerne l'évaluation des performances des grands systèmes CDMA. Notre préoccupation rejoint des travaux récents menés à Berkeley (D. Tse<sup>(2)</sup>) et Princeton (S. Verdu<sup>(3)</sup>) visant à analyser les performances de certains récepteurs linéaires. Dans ce contexte, l'indicateur de performance le plus naturel est le rapport signal à interférence plus bruit (SINR) en sortie du récepteur. Son évaluation analytique est tout à fait triviale, mais la formule le définissant, qui dépend de façon complexe des codes d'étalement alloués aux divers utilisateurs, n'est quasiment pas exploitable. On ne peut en particulier pas l'utiliser pour analyser, autrement que par de lourdes simulations, l'impact sur les performances de paramètres tels que la charge du système, la politique de contrôle de puissance, l'existence de récepteurs multi-capteurs, le compromis entre codage et étalement.... Afin de résoudre ce problème, Tsé et Verdu ont proposé de modéliser la matrice formée des codes d'étalement par une matrice aléatoire à coefficients indépendants et identiquement distribués (i.i.d.), et de s'intéresser au comportement du SINR quand le nombre d'utilisateurs et le facteur d'étalement convergent vers l'infini à un rythme comparable. Il a été établi que ceci peut se ramener à étudier la distribution des valeurs propres de grandes matrices aléatoires formées à partir d'entrées i.i.d. Grâce à divers résultats, on peut alors établir que le SINR converge presque sûrement vers une quantité déterministe, ne dépendant évidemment pas de la réalisation particulière de la matrice des codes. De plus, le SINR est donné par une formule explicite qui permet de bien comprendre l'influence des paramètres les plus importants sur les performances du récepteur. Si le modèle de matrice de code i.i.d. permet de bien représenter la réalité dans le cas d'une liaison montante asynchrone, il constitue un modèle tout à fait inapproprié pour décrire les matrices utilisées dans les liaisons descendantes, qui sont quasiment toujours orthogonales. Nous nous sommes donc intéressés aux mêmes questions, mais quand on remplace les codes i.i.d. par les matrices aléatoires orthogonales les plus courantes, i.e. celles dont la loi de probabilité est uniforme sur le groupe des matrices unitaires (distribution de Haar). Dans ce

<sup>(2)</sup>D.Tse, S. Hanly, IEEE Trans. on Information Theory, vol. 45, no. 2, pp. 641-657, March 1999.

<sup>(3)</sup>S. Verdu, S. Shamai, IEEE Trans. on Information Theory, vol. 45, no. 2, pp. 622-640, March 1999.

contexte, les outils statistiques sont d'une nature tout à fait différente, et nous avons utilisé des résultats issus de la théorie des probabilités libres afin de parvenir à analyser les SINR de différents types de récepteurs et à établir qu'ils convergent également vers une quantité déterministe. Ce travail a été mené dans le cadre des systèmes CDMA [469] et des systèmes MC-CDMA [478, 479]; voir également l'article de synthèse issu de la conférence invitée [543]. Nous avons également abordé de la même manière le problème de l'évaluation des performances de récepteurs de complexité réduite basé sur la technique du filtrage de Wiener à rang réduit, qui consistent à estimer le symbole courant en le projetant sur un sous-espace de petite dimension, appelée rang du récepteur, formé à partir du signal observé. En utilisant des résultats relatifs au comportement asymptotique de certains polynômes orthogonaux, nous avons mis en évidence l'impact du rang sur les performances du récepteurs [544, 546, 548]). D'une façon générale, l'utilisation des grandes matrices aléatoires dans le domaine des communications numériques et du traitement statistique du signal nous semble être une direction de recherche très prometteuse. En effet, la plupart des résultats qui ont été obtenus jusqu'à maintenant ne concernaient que des modèles simples ne représentant pas toujours fidèlement la réalité. Afin de traiter des scénarios plus réalistes, il convient d'adapter les résultats mathématiques existants, voire même de développer de nouvelles approches. Compte tenu de l'ampleur de la tâche, nous collaborons avec d'autres équipes du domaine, notamment dans le cadre du réseau d'excellence NEWCOM, et avec des mathématiciens spécialistes des matrices aléatoires. Nous avons en particulier mis en place, en collaboration avec J. Najim (LTCI), l'ACI Nouvelles Interfaces des Mathématiques MALCOM (Application des Matrices Aléatoires à l'Evaluation des Performances des Systèmes de Communication) associant des jeunes chercheurs que nous avons formé (Ph. Ciblat, M. Debbah, W. Hachem, S. Lasaulce) et des mathématiciens de très haut niveau (O. Khorunzhy de l'université de Versailles, L. Pastur de l'Institut de Physique des Basses Températures à Kharkov). Par ailleurs, nous avons invité un mois à l'UMLV V. Girko qui est sans doute l'un des tous premiers spécialistes de ces questions.

### 6.3.2 Séparation de sources

**Membres permanents :** A. Chevreuil, Ph. Loubaton, J.-C. Pesquet.

**Doctorants participant ou ayant participé :** M. Castella, S. Houcke, P. Jallon.

#### Présentation de la problématique

Dans de nombreuses applications, on observe un signal vectoriel  $\mathbf{y}(n)$  de dimension  $N$  qui peut être modélisé comme la sortie d'un système linéaire  $K$ -entrées /  $N$ -sorties inconnu excité par un signal vectoriel  $\mathbf{s}(n)$  non observable, de dimension  $K$ . Autrement

dit,

$$\mathbf{y}(n) = \sum_k \mathbf{H}_k \mathbf{s}(n - k)$$

où  $\sum_k \mathbf{H}_k z^{-k}$  est la fonction de transfert du système. Les composantes de l'entrée  $\mathbf{s}(n)$  représentent des signaux dus à des sources indépendantes qui se propagent dans un milieu, tandis que le système linéaire inconnu approxime l'effet de la propagation dans le milieu. Dans un contexte de ce type, le problème de la séparation de sources consiste à tenter de reconstituer les signaux sources, c'est-à-dire les composantes de  $\mathbf{s}(n)$ , à partir de la seule connaissance du signal observé. L'une des applications possibles de cette problématique est celle de l'écoute passive déjà évoquée plus haut dans le cas où plusieurs émetteurs interfèrent dans la bande passante analysée. Il existe cependant bien d'autres contextes où la séparation de sources est un problème pertinent, par exemple la prise de son multi-locuteurs, le contrôle non destructif, l'analyse de certains types d'images,...

La majorité des travaux qui ont été consacrés à ce problème ont considéré le cas où les composantes de  $\mathbf{s}(n)$  sont des suites indépendantes et identiquement distribuées (i.i.d. en abrégé) non gaussiennes. Dans ce contexte, on peut généraliser simplement les approches basées sur la maximisation de fonctions de contrastes introduites dans le contexte de la déconvolution aveugle. Deux approches de la séparation de sources i.i.d par fonction de contraste peuvent être distinguées :

- Les approches de séparation par bloc consistent à chercher un filtre  $N$ -entrées /  $K$ -sorties de fonction de transfert  $\mathbf{G}(z)$  maximisant une fonction bien choisie  $J(\mathbf{G})$  s'exprimant en fonction des statistiques du signal  $\mathbf{r}(n) = [\mathbf{G}(z)]\mathbf{y}(n)$ . Des fonctions de coût dont le maximum est atteint si et seulement si chaque composante de  $\mathbf{r}(n)$  est une version retardée et/ou pondérée de chaque composante de  $\mathbf{s}(n)$  ont ainsi été mises en évidence [492]. Cependant, la plupart de ces fonctions doivent être maximisées sous une contrainte portant sur  $\mathbf{G}(z)$  permettant d'assurer que la même source n'est pas extraite plusieurs fois. La contrainte la plus fréquemment rencontrée est celle de para-unitarité, i.e.  $\mathbf{G}(e^{2i\pi f})\mathbf{G}(e^{2i\pi f})^* = \mathbf{I}_K$ . Bien entendu, l'ensemble des filtres para-unitaires est d'une nature complexe, et y maximiser une fonction n'est en pratique pas facile.
- Les approches par déflation (aussi qualifiées d'approches séquentielles), proposées à l'origine par Delfosse et Loubaton<sup>(4)</sup> consistent à extraire la première source, à identifier et soustraire sa contribution au signal  $\mathbf{y}(n)$  afin de former un nouveau mélange convolutif de  $K - 1$  sources. L'étape initiale peut alors être répétée  $K - 1$  fois pour séparer toutes les sources. Afin d'extraire la première source, on cherche une fonction de transfert  $\mathbf{g}(z)$   $N$ -entrées / 1-sortie maximisant une fonction bien choisie  $J(\mathbf{g})$  s'exprimant en fonction des statistiques du signal sca-

<sup>(4)</sup>N. Delfosse, Ph. Loubaton, Signal Processing, 45, 1995, pp. 59-83.

laire  $r(n) = [\mathbf{g}(z)]\mathbf{y}(n)$ . Comme dans les approches par bloc, on peut mettre en évidence facilement des fonctions  $J$  dont le maximum est atteint si et seulement si  $r(n)$  coïncide avec l'une des composantes de  $\mathbf{s}(n)$  à un retard près. L'exemple le plus simple est la valeur absolue du kurtosis de  $r(n)$  définie par

$$J(r) = \left| \frac{c_4(r(n))}{(\mathbb{E}(|r(n)|^2))^2} \right| \quad (6.1)$$

où  $c_4(r(n))$  représente le cumulants d'ordre 4 de la variable aléatoire  $r(n)$ .

Ces approches s'étendent immédiatement aux signaux sources qui sont des processus linéaires, c'est-à-dire des signaux obtenus par filtrage linéaire de suites i.i.d. non gaussiennes. En effet, il suffit de remplacer dans tout ce qui précède les signaux sources  $s_k$  par les suites i.i.d.  $\nu_k$  qui les engendrent pour se ramener au cas i.i.d. Les algorithmes de séparation décrits plus haut reconstituent alors les  $\nu_k$  (c'est-à-dire des versions filtrées particulières des  $s_k$ ) à un retard près.

## Résultats obtenus

L'hypothèse suivant laquelle les signaux sources sont des suites i.i.d. ou des processus linéaires est assez restrictive en pratique. Nous nous sommes donc intéressés à la séparation des mélanges convolutifs de signaux stationnaires (classe beaucoup plus large que celle des processus linéaires) par des techniques de fonctions de contraste.

Dans le cadre des approches par bloc, une partie non négligeable des contrastes considérés dans le cas i.i.d. ne fonctionne plus dans le cas stationnaire. Nous avons cependant montré qu'il était possible d'aborder le problème dans le domaine spectral, par une approche originale. De nouveaux contrastes fréquentiels s'exprimant à l'aide des polyspectres des observations ont ainsi été proposés dans [468]. L'équivalence de ces critères avec des formes temporelles a également été montrée, dans certains cas, ce qui permet la construction d'une vaste classe de contrastes pour des sources non i.i.d. L'un des inconvénients de cette approche est néanmoins de requérir un pré-blanchiment des données. Une autre difficulté est qu'il est nécessaire de mettre en œuvre des algorithmes d'optimisation [517] permettant d'éviter les maxima locaux parasites que présentent généralement ces contrastes par bloc.

Une façon de pallier ces problèmes est de recourir à des approches séquentielles où les sources sont estimées les unes après les autres. Nous avons en particulier établi dans [498] que la plupart des contrastes utilisés dans le cas i.i.d. par les approches par déflation continuaient à être valides, la différence étant que l'algorithme d'optimisation fournit des versions filtrées des signaux sources. Des alternatives à ces méthodes de déflation ont également été envisagées afin de limiter les phénomènes d'accumulation d'erreurs au fil des itérations et d'accélérer la convergence des algorithmes d'optimisation mis en œuvre. Ces travaux réalisés pendant la thèse de M. Castilla ont montré qu'il

pouvait être avantageux de remplacer la méthode de moindres carrés utilisée dans la déflation pour soustraire les sources déjà extraites, par une procédure de décorrélation [519]. Il a également été souligné l'intérêt de techniques de post-optimisation où les contraintes sont relaxées au voisinage d'une solution afin de réduire l'erreur d'estimation. Dans le cadre d'une collaboration avec E. Moreau (ISITV-Toulon), une autre voie prometteuse a été prospectée qui consiste à maximiser des « contrastes avec référence » [516] employant des cumulants croisés du type  $c_4(r(n), r(n), z(n), z(n))$  entre une estimation  $r(n)$  d'une source et un signal de référence  $z(n)$ . Ce dernier peut être choisi comme une version filtrée quasiment quelconque des sources. L'intérêt de cette approche est de conduire à des critères quadratiques dont la maximisation est aisée.

Les signaux générés par des systèmes de communication numérique ne sont pas stationnaires, mais cyclostationnaires. Afin de pouvoir utiliser des techniques de séparation de sources dans le contexte de l'écoute passive, nous étudions le problème de la séparation de sources cyclostationnaires (thèse de P. Jallon). Par rapport au cas stationnaire, deux difficultés nouvelles apparaissent. Il faut tout d'abord revoir les fonctions de contraste du cas stationnaire puisque, souvent basées sur les statistiques du signal de sortie du séparateur, leur expression fait apparaître une dépendance temporelle. A titre d'exemple, le kurtosis défini par (6.1) dépend de l'instant  $n$ . Comme les filtres séparateurs sont invariants au cours du temps, il est évidemment hors de question de maximiser ce contraste à chaque instant car les filtres obtenus dépendraient alors du temps. Il faut donc reconsidérer les fonctions de contraste du cas stationnaire de façon à les rendre invariantes au cours du temps. Dans le contexte du kurtosis, nous avons par exemple établi dans [540] que la maximisation de

$$J'(r) = \left| \frac{\langle c_4(r(n)) \rangle}{(\langle \mathbb{E}(|r(n)|^2) \rangle)^2} \right| \quad (6.2)$$

permet d'extraire un signal source. Le symbole  $\langle \rangle$  désigne l'opérateur de moyenne temporelle défini par  $\langle u(n) \rangle = \lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{n=1}^N u(n)$ .

Le deuxième problème posé par la présence de signaux cyclostationnaires concerne l'estimation des fonctions de contraste. En effet, les estimateurs des divers moments et statistiques utilisés dans le cas stationnaire pour évaluer les fonctions de contraste les plus usuelles ne convergent pas vers les valeurs souhaitées. Dans ces conditions, on ne maximise pas les bonnes fonctions, et les performances des méthodes de séparation peuvent en être gravement affectées. A titre d'exemple, estimer de façon consistante  $\langle c_4(r(n)) \rangle$  nécessite l'estimation des fréquences cycliques du signal observé, problème statistique qui est parfois difficile à résoudre.

Afin de contourner cette difficulté, nous avons montré que l'approche d'estimation/égalisation conjointe mise en évidence dans [486] pouvait être généralisée dans le cas multi-sources [487] quand les signaux sources sont des modulations linéaires.

L'algorithme résultant de [487] étant très complexe à mettre en œuvre, l'approche que nous étudions actuellement consiste à mettre en évidence des fonctions de contrastes dont l'estimation consistante ne pose aucun problème, et qui, au moins dans certains cas, permettent de séparer des sources cyclostationnaires. C'est par exemple le cas de la fonction définie par

$$\left| \frac{\langle \mathbb{E}(|r(n)|^4) \rangle}{(\langle \mathbb{E}(|r(n)|^2) \rangle)^2} - 2 \right|$$

qui s'estime simplement, et se comporte comme une fonction de contraste dans le cas où les signaux à séparer ont des fréquences cycliques différentes [540], ou encore de la fonction

$$\frac{\langle \mathbb{E}(|r(n)|^4) \rangle}{(\langle \mathbb{E}(|r(n)|^2) \rangle)^2}$$

qui fonctionne si les signaux sources sont obtenus en modulant linéairement des symboles de module constant. Ces exemples montrent que l'on peut être en mesure de séparer certains mélanges de sources cyclostationnaires par le biais d'approches dont la complexité et les performances sont du même ordre de grandeur que dans le cas stationnaire. Il reste cependant à accomplir un travail important afin d'identifier clairement les situations permettant de mettre en œuvre des solutions de ce type.

### 6.3.3 Théorie de l'information

**Membre permanent :** C. Vignat.

Notre activité de recherche en théorie de l'information concerne l'étude des notions d'entropie et d'information de Fisher dans leurs applications à certains problèmes de communications.

Si  $X$  est une variable aléatoire de densité de probabilité  $f_X(x)$ , nous avons abordé l'étude des entropies de Rényi  $H_X^{(\alpha)}$  définies de la façon suivante pour tout  $\alpha \geq 0$  :

$$H_X^{(\alpha)} = \frac{1}{1-\alpha} \ln \int f_X^\alpha dx$$

Remarquons que cette famille inclut l'entropie de Shannon  $H_X = - \int f_X(x) \ln(f_X(x)) dx$  comme cas particulier lorsque  $\alpha \rightarrow 1$ .

Nous avons évalué la possibilité d'utiliser ces entropies de Rényi dans le cadre de la déconvolution aveugle [554]. Cette étude a mis en évidence le rôle fondamental de l'inégalité de puissance entropique - dont aucune extension aux entropies de Rényi n'est actuellement connue - pour la construction d'une fonction de contraste réaliste : l'égalisation d'un canal basée sur les entropies de Rényi avec  $\alpha \neq 1$  implique la maîtrise de la norme infinie du canal équivalent ; dans le cas de l'entropie de Shannon, l'application de l'inégalité de la puissance entropique permet l'égalisation à condition que la

norme 2 du canal équivalent soit contrainte, ce qui correspond à un schéma beaucoup plus réaliste.

Soit  $J_X$  l'information de Fisher associée à la variable aléatoire  $X$  définie par

$$J_X = \int \left( \frac{f'_X}{f_X} \right)^2 f_X dx$$

L'inégalité de la puissance entropique est une conséquence de la propriété de sous-additivité de l'information de Fisher, qui s'écrit

$$J_{X+Y}^{-1} \geq J_X^{-1} + J_Y^{-1}$$

où  $X$  et  $Y$  sont deux variables aléatoires indépendantes. R. Zamir<sup>(5)</sup> a proposé une extension au cas multi-dimensionnel de ce résultat, de la forme

$$J_{AX}^{-1} \geq AJ_X^{-1}A^T$$

- où  $A$  est une matrice  $m \times n$  de rang plein avec  $m \leq n$  - et a caractérisé les cas d'égalité. Avec J.-F. Bercher [501, 553], nous avons apporté une nouvelle preuve simplifiée de ces résultats, basée sur certaines propriétés de la fonction score ; cette approche met de plus en évidence l'importante notion de composante identifiable d'un système linéaire non-inversible.

Dans le cadre d'une collaboration avec P. Harremoës, nous avons étudié la possibilité d'étendre au cas des probabilités discrètes des inégalités établies uniquement dans le cas continu. Nous avons d'abord proposé dans [485] une extension au cas de certaines lois discrètes d'un résultat dû à Cover<sup>(6)</sup> dans le cas Gaussien : étant donné un canal additif instantané, il s'agit de caractériser la loi optimale des données que l'émetteur doit adopter afin de maximiser l'information transmise en présence d'un brouilleur de loi fixée. Nous avons ensuite proposé une extension [484] de l'inégalité de la puissance entropique au cas de certaines lois discrètes, à savoir au cas de lois binômiales de paramètre 1/2.

Une collaboration avec A. Hero, du département E.E.C.S. de l'université du Michigan, a permis l'étude des lois multidimensionnelles maximisantes, sous contrainte de covariance, des entropies de Rényi. Ces lois avaient été identifiées par Kapur, en 1988, comme des lois de Cauchy généralisées, mais dans le cas d'une matrice de covariance unité uniquement. Nous avons notamment étudié les représentations stochastiques des variables aléatoires associées [530] et caractérisé leurs transformations par convolutions

<sup>(5)</sup>R. Zamir, "A Proof of the Fisher Information Matrix via a Data Processing Argument", IEEE trans. on Information Theory, IT 44, 3, pp. 1246-1250, 1998.

<sup>(6)</sup>T. Cover and J. A. Thomas, Elements of Information Theory. Wiley, 1991, Ex.1 p.263.

[555, 502].

Les entropies de Rényi sont des fonctions monotones des entropies de Tsallis, introduites en 1988 par C. Tsallis<sup>(7)</sup> dans le domaine de la thermostatistique, afin de modéliser certains systèmes complexes. Nous avons entrepris des collaborations avec deux chercheurs de cette communauté (A. Plastino, Université de La Plata, J. Naudts, Université d'Anvers) afin d'évaluer les apports que peut susciter une approche physique de ces problèmes d'optimisation.

### 6.3.4 Analyse en ondelettes 2D

**Membre permanent :** J.-C. Pesquet.

**Doctorants participant ou ayant participé :** C. Chaux, J. Gauthier, W. Miled, T. Petrisor.

L'analyse en ondelettes et ses applications constituent l'un des principaux thèmes de recherche de la communauté du Traitement du Signal et des Images, depuis une quinzaine d'années. Au cours de cette période, les transformations en ondelettes ont été généralisées sous diverses formes, notamment dans le but de mieux décrire les structures géométriques présentes dans les images. Dans le même temps, les besoins applicatifs se sont multipliés, particulièrement en restauration et en compression d'images, faisant ainsi apparaître de nouvelles problématiques.

Une part importante de notre travail concerne l'imagerie multi-composantes et est menée en collaboration avec A. Benazza-Benyahia (Sup'Com Tunis). Dans ce contexte, on dispose de plusieurs images de la même scène recueillies par des capteurs de caractéristiques (longueur d'onde, résolution,...) différentes. Ce type de situation se produit typiquement dans les systèmes d'imagerie satellitaires multi- ou hyperspectraux. Il existe alors des redondances/dépendances fortes entre les images correspondant aux différentes bandes spectrales. L'exploitation de ces relations inter-bandes permet d'espérer des performances accrues par rapport aux traitements actuels qui sont le plus souvent effectués sur chaque image prise isolément.

Dans le cadre de la compression de données satellitaires, nous avons souligné l'intérêt de l'utilisation d'analyses en ondelettes vectorielles, permettant de mieux réduire les redondances inter-bandes, tout en fournissant une approche de codage progressif. Pour la construction de ces nouvelles représentations des images satellitaires, nous avons montré qu'il était possible d'exploiter des bancs de filtres vectoriels mis en œuvre sous forme de *lifting* [463]. Ces structures permettent une conception très souple des opérateurs de décomposition et intègrent notamment une prédiction vectorielle des signaux, combinant les différentes bandes. Des simulations ont montré que ces méthodes conduisent à un gain de performances significatif, aussi bien pour un codage avec

---

<sup>(7)</sup>C. Tsallis, J. Stat. Phys. 52 (1988) 479.

perte que sans perte. Dans ce dernier cas, il est nécessaire d'inclure dans le lifting des opérateurs de troncature (« integer to integer ») qui n'altèrent cependant pas les propriétés de reconstruction parfaite [462]. D'autres non-linéarités peuvent également être introduites, par exemple de façon à préserver les contours des objets au cours d'une analyse multirésolution. Des liens peuvent être établis entre ces approches et une discrétisation de méthodes espace-échelle reposant sur la résolution d'équations aux dérivées partielles [464].

Dans les nouveaux systèmes de satellite tels que SPOT 5, un échantillonnage en quinconce des données est réalisé pour certaines bandes. Par ailleurs, ce type de sous-échantillonnage se prête à la conception d'analyses adaptatives où les filtres sont optimisés sans contrainte de séparabilité. Il semble donc intéressant d'étudier de manière plus précise ce type de schémas d'analyse en quinconce [509].

Un autre point fort de notre activité concerne le débruitage d'images. Le principe des méthodes de régression par ondelettes est de concentrer l'information utile sur un nombre réduit de coefficients qu'il est possible de discriminer du bruit par des estimateurs non linéaires. Suivant les transformations en ondelettes appliquées et les techniques d'estimation employées, diverses méthodes sont obtenues. La façon la plus simple de procéder a été proposée par D. Donoho<sup>(8)</sup> dans les années 90 et consiste à seuiller les coefficients issus de la décomposition du signal bruité, sur une base orthonormale d'ondelettes. Cette approche initialement justifiée pour un bruit gaussien peut être étendue à des bruits non gaussiens. Nous avons ainsi montré dans [461], en utilisant une approche bayésienne, que des seuillages spécifiques sont obtenus pour des distributions de bruit gaussiennes généralisées (GG) et de Cauchy, quand les coefficients d'ondelettes du signal utile sont supposés suivre une loi GG. Des résultats similaires ont été établis lorsqu'on fait l'hypothèse que le signal appartient à un certain espace de Besov [489].

L'inconvénient de ces méthodes bayésiennes est de supposer la connaissance parfaite de la loi *a priori* des données à estimer. Lorsque cette hypothèse est mise en défaut, des erreurs de modélisation apparaissent et peuvent être amplifiées par les méthodes numériques auxquelles on doit souvent recourir, ne serait-ce que pour estimer les hyperparamètres de ces lois. Une approche permettant de contourner ces problèmes consiste à se donner une structure d'estimateur non linéaire qui peut être obtenue à partir de considérations bayésiennes (par exemple, une modélisation Bernoulli-gaussienne des coefficients d'ondelettes) et d'optimiser les paramètres de l'estimateur de façon à directement minimiser le risque. Cette technique se révèle plus robuste car elle permet de réduire l'impact des imprécisions pesant sur le modèle. Elle n'est cependant applicable que si le risque, qui dépend naturellement des données originales (donc inconnues), peut lui-même être estimé de manière consistante. Or, dans le cas d'un coût

---

<sup>(8)</sup>D. L. Donoho and I.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika*, no. 81, pp. 425-455, 1994.

quadratique et d'un bruit additif gaussien, le principe de Stein<sup>(9)</sup> permet d'effectuer une estimation non biaisée du risque. Cette stratégie a été employée avec succès dans des problèmes de débruitage d'images satellitaires multispectrales [505]. Un estimateur multivarié est alors appliqué aux coefficients d'ondelettes de façon à prendre en compte les dépendances statistiques existant entre les bandes spectrales. Des simulations exhaustives nous ont permis de comparer notre méthode aux approches concurrentes actuelles et ont démontré ses excellentes performances. Ces travaux devraient se poursuivre dans le cadre de l'ACI MULTIM (2004-2007) où l'objectif est de résoudre des problèmes plus complexes de restauration/déconvolution d'images multispectrales pouvant provenir non seulement de sources satellitaires mais aussi astronomiques. Une autre voie de recherche consiste également à généraliser les estimateurs que nous avons proposés de façon à tirer parti des dépendances inter-échelles [506].

Une façon différente d'appréhender les problèmes de restauration consiste à rechercher l'image comme la solution d'un problème d'optimisation sous contrainte. Si l'on observe

$$y = Hx + u$$

où  $x$  est l'image originale,  $u$  désigne le bruit et  $H$  l'opérateur de dégradation (flou par exemple). On détermine l'image restaurée  $\hat{x}$  de façon à minimiser une fonctionnelle  $J$  sous la contrainte que  $\hat{x}$  appartienne à certains ensembles de contraintes  $(S_i)_{i \in I \subset \mathbb{N}}$  traduisant l'information *a priori* dont on dispose sur la solution (par exemple, bornes inférieure et supérieure sur les composantes de  $x$ ). On a souvent intérêt à choisir  $J$  et les  $(S_i)_{i \in I}$  convexes pour pouvoir disposer d'algorithmes d'optimisation efficaces. Ainsi, si l'on se place dans un espace de Hilbert muni d'une norme  $\|\cdot\|$ ,  $J : x \mapsto \|y - Hx\|^2$  constitue un critère usuel. Dans ce cadre, notre contribution a porté sur la proposition d'ensembles de contraintes servant à conférer une certaine régularité à l'image restaurée. Nous avons notamment considéré l'appartenance à des boules BV (Bounded Variation). La méthode que nous avons développée présente l'avantage par rapport à l'approche historique de Rudin et Osher<sup>(10)</sup>, de pouvoir gérer simultanément un nombre arbitraire de contraintes convexes supplémentaires [476]. L'algorithme proposé en collaboration avec P. Combettes (Univ. Paris 6) présente une convergence très rapide et peut être parallélisé. D'autres ensembles de contraintes tels que des boules de Besov ont également été étudiés. Rappelons que ces boules s'expriment facilement dans le domaine ondelettes. Une limitation des contraintes envisagées est cependant de nécessiter la détermination des rayons des boules, ce qui conduit à réaliser une étude statistique préalable portant sur la classe d'images traitées. Pour éviter cette démarche, nous cherchons à proposer d'autres formes de contraintes pouvant être définies à partir de coefficients d'ondelettes ou de gradients directionnels, et qui sont calculées automatiquement

<sup>(9)</sup>C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, no 9, pp. 1135-1151, 1981.

<sup>(10)</sup>L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D*, vol. 60, pp. 259-268, 1992.

à partir de l'image observée. Nous avons, en particulier, considéré des contraintes de type hyperplan affine dont les caractéristiques peuvent être estimées, dans un contexte de bruit gaussien, à l'aide du principe de Stein déjà évoqué. Une analyse statistique asymptotique a été effectuée pour fournir des intervalles de confiance sur la solution obtenue [477]. Comme perspective à ce travail, il pourrait être intéressant de s'attaquer à des problèmes de restauration d'archives cinématographiques, dans le prolongement de la spécialité de Master ouverte en 2004 au pôle de Val d'Europe de l'Université de Marne la Vallée.

Une autre voie d'amélioration des méthodes de compression/restauration à l'aide d'ondelettes consiste à sortir du cadre assez restrictif des décompositions sur des *bases* d'ondelettes. Ces dernières années ont ainsi vu émerger d'autres types de transformations (curvelets, bandelettes,...). Notre équipe a fait porter ses efforts sur la construction de trames (« frames ») de fonctions appropriées à l'analyse d'images. Dans le cadre de la thèse de C. Chaux, nous avons présenté une extension au cas  $M$ -bandes des travaux concernant la construction de décompositions en ondelettes formant des paires de Hilbert [524]. Ces représentations présentent de nombreux avantages notamment en terme d'analyse invariante par translation et de directionnalité  $2D$ . Nous avons établi les conditions que doivent satisfaire les bancs de filtres en arbre dual servant à l'analyse/la synthèse des signaux traités. Nous avons également fourni une justification théorique des pré-traitements qu'il est nécessaire d'appliquer à des données discrètes. Ces décompositions introduisant typiquement une redondance d'un facteur 2, elles constituent des trames à partir desquelles on peut aisément calculer une reconstruction optimale. Des applications de ces outils à des données sismiques sont en cours, en collaboration avec l'Institut Français du Pétrole. La thèse de T. Petrisor (menée en liaison avec le LTCI-ENST Paris) considère d'autres types de trames d'ondelettes, dans le but de développer des algorithmes de codage à descriptions multiples de séquences vidéo. Ces techniques sont utiles pour assurer un codage robuste dans des réseaux où peuvent se produire des pertes de paquets.

En amont de ces thèmes, on peut mentionner quelques études théoriques visant à déterminer les caractéristiques statistiques (cumulants) des coefficients d'ondelettes de certains processus aléatoires non stationnaires [499].

## 6.4 Activités

### 6.4.1 Contrats

#### Contrats industriels

- Contrat (en collaboration avec E. Moulines, ENST Paris) de 120 KF avec Motorola (Novembre 98/ Octobre 2001) correspondant à l'encadrement de la thèse de S. Lasaulce [567].

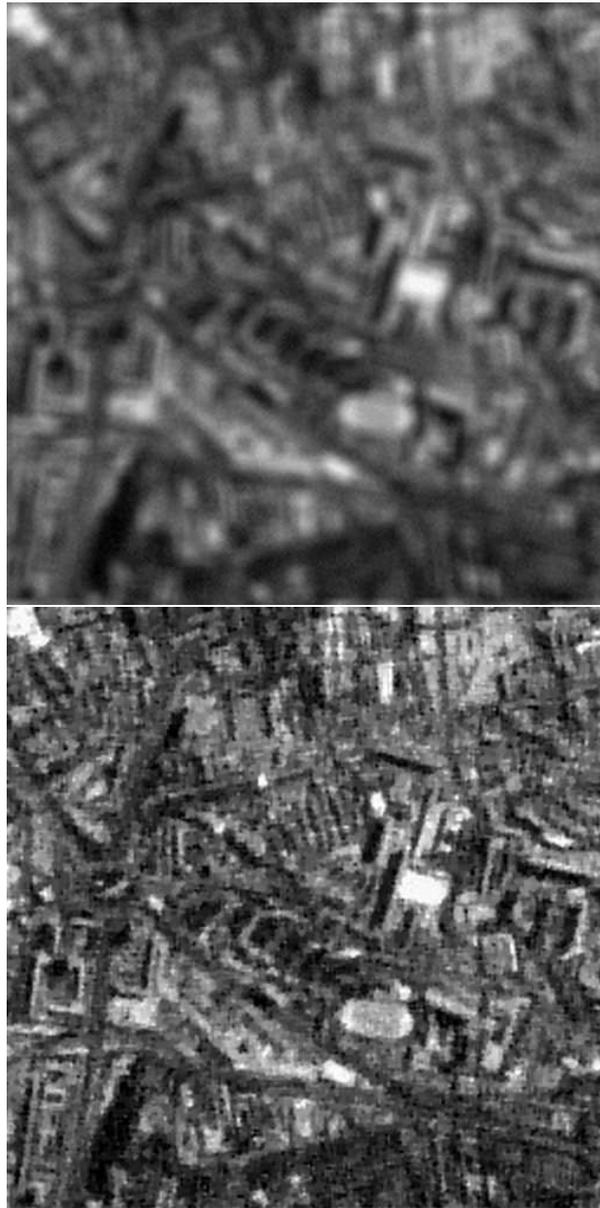


FIG. 1: *Images satellitaires dégradée (à gauche) et restaurée (à droite) à l'aide d'une contrainte d'appartenance à une boule BV.*

- Participation au projet RNRT DOLIE (150 KF pour l'UMLV, Janvier 2000/ Juin 2002). Objet du projet : évaluer la faisabilité du CDMA pour la transmission haut débit sur le réseau électrique.
- Participation au projet européen IST ANTIUM (65 KE pour l'UMLV, Janvier

- 2001/Décembre 2003). Objet du projet : développer un équipement susceptible d'analyser les scénarios de brouillage dans les réseaux UMTS/FDD.
- Participation au projet LOLITA2 financé par la DGA (25 KE pour l'UMLV, Janvier 2001/Juin 2002). Objet des travaux de l'UMLV : mettre en évidence des algorithmes de démodulation aveugle de modulations CPM.
  - Contrat avec Thalès Communication correspondant à l'encadrement de la thèse de J.M. Chaufray [562] (150 KF, Octobre 1999, Septembre 2002).
  - Contrat avec Thalès Communication correspondant à l'encadrement de la thèse de P. Bianchi [560] (25 KE, Octobre 2000, Septembre 2003).
  - Contrat avec Thalès Communication correspondant à la réalisation d'une tâche dans un contrat confidentiel (45KE, Janvier 2003, Décembre 2003).
  - Contrat avec France Telecom Recherche et Développement correspondant à l'encadrement de la thèse de J. Dumont (27 KE, Novembre 2003, Octobre 2006).
  - Contrat avec l'Institut Français du Pétrole correspondant à l'encadrement de la thèse de J. Gauthier (18 KE, Octobre 2004, Septembre 2007).

### **Financements institutionnels (ACI, Réseaux d'excellence,...)**

- Participation à un projet MathSTIC (Algorithmes parallèles de décomposition pour la restauration d'images satellitaires) en 2002.
- Contrat ONR (Office of Naval Research, USA) avec Drexel University (Philadelphie) sur la séparation de sources, 2000-2004.
- Partenaires d'un contrat de coopération franco-tunisien (CMCU) portant sur le thème « Auto-similarités : théorie et applications », impliquant 7 laboratoires tunisiens et 13 français, 2002-2004.
- Participation et co-responsabilité de l'ACI Nouvelles Interfaces des Mathématiques MALCOM (Application des Matrices Aléatoires à l'Evaluation de Performances des Systèmes de Communication Numériques), 2004-2006.
- Participation à l'ACI Nouvelles Interfaces des Mathématiques MULTIM (Nouvelles méthodes mathématiques pour la restauration d'images multi-canaux), 2004-2006.
- Participation (A. Chevreuil, Ph. Loubaton) au Réseau d'Excellence NEWCOM (Network of Excellence in Wireless Communications) du 6ème PCRD (à partir de 2004).

## **6.4.2 Diffusion**

### **Livres**

Nous avons participé à 2 ouvrages du traité IC2 :

- « Signal et Télécommunications », Hermès Science, 2004, coordonné par Ph. Loubaton.

- « Le Traitement d’Images », Hermès Science, 2003, Chapitre « Ondelettes et Traitement d’Images » écrit par B. Pesquet-Popescu et J.C. Pesquet.

### **Organisation de sessions spéciales dans des conférences**

Ph. Loubaton a organisé les sessions spéciales suivantes :

- « Large random matrices in digital communications and signal processing » à la conférence EUSIPCO-2004, Vienne, 2004,
- « Large random matrices and performance evaluation of large digital communication systems » qui se déroulera à la conférence ICASSP-2005, Philadelphie, 2005.

### **6.4.3 Collaborations**

#### **Collaborations nationales**

- Service Radio-Electricité de Supelec (W. Hachem).
- Laboratoire Jacques-Louis Lions, Université Paris 6 (P. Combettes).
- Institut des Sciences de l’Ingénieur de Toulon et du Var (E. Moreau).
- Laboratoire de Modélisation et Calcul, Institut d’Informatique et Mathématiques Appliquées de Grenoble (A. Antoniadis).
- Laboratoire Traitement et Communication de l’Information, CNRS-ENST Paris (Ph. Ciblat, E. Moulines, J. Najim, B. Pesquet-Popescu).
- Laboratoire des Signaux et Systèmes, CNRS-SUPELEC (S. Lasaulce).
- Département Communications Mobiles, Institut Eurecom (M. Debbah).

#### **Collaborations internationales**

- Département Mathématiques Appliquées, Signal et Communications, Ecole Supérieure des Communications de Tunis (A. Benazza-Benyahia).
- Département d’Electrical and Computer Engineering de l’Université de Caroline du Nord (H. Krim).
- Département d’Electrical and Computer Engineering de Drexel University, Philadelphie (A. Petropulu).
- Département d’Electrical Engineering and Computer Science de l’Université du Michigan (A. Hero).
- Département d’Electrical Engineering and Computer Science de l’Université de Minneapolis (G. Giannakis).
- Département de Mathématiques de l’Université de Copenhague (P. Harremoës).
- Département de Physique de l’Université d’Anvers (J. Naudts).
- Institut de Mathématique de l’Université de Kiev (V. Girko).

### 6.4.4 Activités doctorales

Ph. Loubaton et J.-C. Pesquet sont responsables d'un cours intitulé « Bancs de filtres et applications » dans le cadre du Master Recherche Automatique et Traitement du Signal de l'Ecole Doctorale STITS (Univ. Paris 11). Par ailleurs, Ph. Loubaton est responsable d'un cours intitulé « Traitement Statistique du Signal pour les Communications » dans le cadre du Master Recherche Telecom et Réseaux de l'Ecole Doctorale STITS (Univ. Paris 11).

### 6.4.5 Thèses et habilitations

#### Thèses

- Samson Lasaulce, *Estimation de canal et détection multi-utilisateurs pour les systèmes UMTS-TDD* 2001.
- Mérouane Debbah, *Précodeurs linéaires pour les transmissions OFDM sans fils* 2002.
- Sébastien Houcke, *Séparation autodidacte d'un mélange de sources émettant à des débits inconnus et éventuellement différents* 2002.
- Jean-Marie Chaufray, *Détection et démodulation de stations de bases dans un réseau UMTS* 2002.
- Pascal Bianchi, *Démodulation aveugle de modulations non linéaires à phases continues* 2003.

#### Thèses en cours

- Marc Castella, *Séparation de sources non linéaires dans le cas de mélanges convolutifs* (depuis septembre 2001).
- Caroline Chau, *Analyse de signaux par bancs de filtres M-bandes ; applications au traitement de signaux sismiques* (depuis octobre 2003).
- Julien Dumont, *Utilisation d'informations partielles dans les systèmes de communication multi-entrées / multi-sorties* (depuis novembre 2003).
- Jérôme Gauthier, *Analyse, détection et filtrage de signaux et d'images par bancs de filtres ; applications aux géosciences* (depuis octobre 2004).
- Pierre Jallon, *Séparation aveugle de signaux cyclostationnaires* (depuis septembre 2003).
- Wided Miled, *Analyse d'images de scènes routières par approches multi-résolution pour la détection des obstacles routiers* (depuis février 2004).
- Belkacem Mouhouche, *Récepteurs avancés pour la liaison descendante de l'UMTS* (depuis septembre 2002).
- Teodora Petrisor, *Décompositions en ondelettes redondantes pour le codage vidéo par descriptions multiples* (depuis novembre 2003).

## 6.4.6 Rayonnement

### Au niveau national

- J.-C. Pesquet a été membre de la commission thématique Signal et Circuits Intégrés associés du RNRT de 1999 à 2002, et Ph. Loubaton est membre de cette même commission depuis 2003.
- Ph. Loubaton a été membre élu (2000-2003), puis membre nommé (à partir de 2004) de la section 61 du Conseil National des Universités.

### Au niveau international

- Ph. Loubaton a été Editeur Associé à IEEE Transactions on Signal Processing de 1998 à 2001, à IEEE Communication Letters de 2000 à 2002, et est Editeur Associé à IEEE Transactions on Signal Processing depuis Novembre 2004. J.-C. Pesquet est Editeur Associé à IEEE Signal Processing Letters depuis Janvier 2004.
- Ph. Loubaton a été membre du comité technique « Signal Processing for Communications » de la IEEE Signal Processing Society entre 1998 et 2004, et J.-C. Pesquet est membre du comité technique « Signal Processing Theory and Methods » de la IEEE Signal Processing Society depuis 2002.
- J.-C. Pesquet a été co-Technical Chairman de la conférence ICASSP-2005.
- A. Chevreuil et Ph. Loubaton sont membres du réseau d'excellence NEWCOM (Network of Excellence in Wireless Communications, 6ème PCRD).

## 6.5 Références bibliographiques

### Articles de revues

- [459] D. Alpay, **A. Chevreuil**, et **P. Loubaton**. An extension problem for discrete time periodically correlated stochastic processes. *J. of Time Series Analysis*, 22(1) :1–11, Jan. 2001.
- [460] D. Alpay, B. Freydin, et **P. Loubaton**. An extension problem for discrete-time almost periodically correlated stochastic processes. *Linear Algebra and its applications*, 308(1–3) :163–181, 2000.
- [461] A. Antoniadis, D. Leporini, et **J.-C. Pesquet**. Wavelet thresholding for some classes of non-Gaussian noise. *Statistica Neerlandica*, 56(4) :434–453, Dec. 2002.
- [462] A. Benazza-Benyahia et **J.-C. Pesquet**. A unifying framework for lossless and progressive image coding. *Pattern Recognition*, 35 :627–638, 2002.
- [463] A. Benazza-Benyahia, **J.-C. Pesquet**, et M. Hamdi. Vector lifting schemes for lossless coding and progressive archival of multispectral image. *IEEE Trans. on Geoscience and Remote Sensing*, 40 :2011–2024, Sept. 2002.
- [464] A. Benazza-Benyahia, **J.-C. Pesquet**, et H. Krim. A nonlinear diffusion-based 3-band filter bank. *IEEE Signal Processing Letters*, 10 :360–363, Dec. 2003.

- 
- [465] J.-F. Bercher et **C. Vignat**. Estimating the entropy of a signal with applications. *IEEE Trans. on Signal Processing*, 48(6) :1687–1694, 2000.
- [466] **P. Bianchi**, **P. Loubaton**, et F. Sirven. Non data aided estimation of the modulation index of continuous phase modulations. *IEEE Trans. on Signal Processing*, 52(10) :2847–2861, Oct. 2004.
- [467] O. Cappé, E. Moulines, **J.-C. Pesquet**, A. Petropulu, et X. Yang. Long-range dependence and heavy-tail modeling for teletraffic data. *IEEE Signal Processing Magazine*, pages 14–27, May 2002.
- [468] **M. Castella**, **J.-C. Pesquet**, et A. P. Petropulu. A family of frequency- and time-domain contrasts for blind separation of convolutive mixtures of temporally dependent signals. *IEEE Trans. on Signal Processing*, 2004. À paraître.
- [469] **J.-M. Chaufray**, **W. Hachem**, et **P. Loubaton**. Asymptotic analysis of optimum and sub-optimum CDMA downlink MMSE receivers. *IEEE Transactions on Information Theory*, 2004. À paraître en novembre.
- [470] **J.-M. Chaufray**, **P. Loubaton**, et P. Chevalier. Consistent estimation of Rayleigh fading channel second order statistics in the context of the wideband CDMA mode of the UMTS. *IEEE Trans. on Signal Processing*, 49(12) :3055–3064, Dec. 2001.
- [471] **A. Chevreuil**, **P. Loubaton**, et L. Vandendorpe. Transmitter induced cyclostationarity : analysis based on a MMSE-DF equalizer. *IEEE Trans. on Signal Processing*, 48(11) :3072–3086, Nov. 2000.
- [472] **A. Chevreuil**, E. Serpedin, **P. Loubaton**, et G. Giannakis. Blind channel identification and equalization using non-redundant periodic modulation precoders : performance analysis. *IEEE Trans. on Signal Processing*, 48(6) :1570–1586, June 2000.
- [473] **P. Ciblat**, **A. Chevreuil**, et **P. Loubaton**. Alpha repetition-modulation and blind second order identification. *IEEE Trans. on Signal Processing*, 48(11) :3153–3161, Nov. 2000.
- [474] **P. Ciblat**, **P. Loubaton**, E. Serpedin, et G. Giannakis. Asymptotic analysis of blind cyclic correlation based symbol rate estimation. *IEEE Trans. on Information Theory*, 48(7) :1922–1934, July 2002.
- [475] **P. Ciblat**, **P. Loubaton**, E. Serpedin, et G. Giannakis. Performance analysis of blind carrier offset estimation for non-circular transmissions through frequency-selective channels. *IEEE Trans. on Signal Processing*, 50(1) :130–140, Jan. 2002.
- [476] P. L. Combettes et **J.-C. Pesquet**. Image restoration subject to a total variation constraint. *IEEE Trans. on Image Processing*, 13(9) :1213–1222, Sept. 2004.
- [477] P. L. Combettes et **J.-C. Pesquet**. Wavelet-constrained image restoration. *International Journal on Wavelets, Multiresolution and Information Processing*, 2004. À paraître.
- [478] **M. Debbah**, **W. Hachem**, **P. Loubaton**, et M. de Courville. MMSE analysis of certain large isometric random precoded systems. *IEEE Trans. on Information Theory*, 49(5) :1293–1311, May 2003.

- 
- [479] **M. Debbah, P. Loubaton**, et M. de Courville. Asymptotic performance of successive interference cancellation in the context of linear precoded ofdm. *IEEE Transactions on Communications*, 52(9) :1144–1448, Sept. 2004.
- [480] N. Delfosse et **P. Loubaton**. Adaptive blind separation of independent sources : a second order stable algorithm. *IEEE Trans. on Circuit and Systems*, 47(7) :1056–1071, July 2000.
- [481] **W. Hachem**, F. Desbouvries, et **P. Loubaton**. Identification of certain Noisy MA Models : New results. *System and Control Letters*, 39 :237–43, 2000.
- [482] **W. Hachem**, F. Desbouvries, et **P. Loubaton**. A MIMO channel blind identification algorithm in the presence of spatially correlated noise. *IEEE Trans. on Signal Processing*, 50(3) :651–661, Mar. 2002.
- [483] F. J. Hampson et **J.-C. Pesquet**. Motion estimation in the presence of illumination variations. *Signal Processing : Image Communication*, 16(4) :373–381, 2000.
- [484] P. Harremoës et **C. Vignat**. An entropy power inequality for the binomial family. *Journal of Inequalities in Pure and Applied Mathematics*, 4(5, Article 93), 2003.
- [485] P. Harremoës et **C. Vignat**. A Nash Equilibrium related to the Poisson Channel. *Communications in Information and Systems*, 3–3 :183–190, 2004.
- [486] **S. Houcke, A. Chevreuil**, et **P. Loubaton**. Blind equalization : case of an unknown symbol period. *IEEE Trans. on Signal Processing*, 51(3) :781–793, Mar. 2003.
- [487] **S. Houcke, A. Chevreuil**, et **P. Loubaton**. Blind source separation of a mixture of communication sources emitting at various baud-rates. *Transactions of IEICE*, E86–A(3) :564–572, Mar. 2003. Invited paper.
- [488] **S. Lasaulce, P. Loubaton**, et E. Moulines. A semi-blind estimation technique based on second-order blind method for CDMA systems. *IEEE Trans. on Signal Processing*, vol.51(7) :1894–1904, July 2003.
- [489] D. Loporini et **J.-C. Pesquet**. Bayesian wavelet denoising : Besov priors and non-gaussian noises. *Signal Processing*, 81 :55–67, 2001.
- [490] **P. Loubaton** et E. Moulines. On blind multiuser forward link channel estimation by the subspace method : identifiability results. *IEEE Trans. on Signal Processing*, 48(8) :2366–2376, Aug. 2000.
- [491] A. Mansour, C. Jutten, et **P. Loubaton**. Adaptive subspace algorithm for blind separation of independent sources in convolutive mixtures. *IEEE Trans. on Signal Processing*, 48(2) :583–586, Feb. 2000.
- [492] **J.-C. Pesquet** et E. Moreau. Cumulant based independence measures for linear mixtures. *IEEE Trans. on Information Theory*, 47 :1947–1956, July 2001.
- [493] B. Pesquet-Popescu et **J.-C. Pesquet**. Synthesis of bidimensional alpha-stable models with long-range dependence. *Signal Processing*, 82 :1927–1940, 2002.
- [494] A. P. Petropulu, **J.-C. Pesquet**, X. Yang, et J. Yin. Power-law shot noise and its relationships with long-memory alpha-stable processes. *IEEE Trans. on Signal Processing*, 48(7) :1883–1892, July 2000.

- [495] F. Sagnard, F. Bentabet, et **C. Vignat**. Theoretical study of methods based on ellipsometry for measurement of complex permittivity of materials. *Electronic Letters*, 22 :1843–1845, Oct. 2000.
- [496] E. Serpedin, **A. Chevreuril**, G. Giannakis, et **P. Loubaton**. Blind joint estimation of carrier frequency offset and channel using non-redundant periodic modulation precoders. *IEEE Trans. on Signal Processing*, 48(8) :2389–2405, Aug. 2000.
- [497] E. Serpedin, **P. Ciblat**, G. Giannakis, et **P. Loubaton**. Performance analysis of blind carrier phase estimators for general QAM constellations. *IEEE Trans. on Signal Processing*, 49(8) :1816–1823, Aug. 2001.
- [498] C. Simon, **P. Loubaton**, et C. Jutten. Separation of a class of convolutive mixtures : a contrast function approach. *Signal Processing*, 81 :883–887, 2001.
- [499] **S. Touati** et **J.-C. Pesquet**. Some results on the wavelet packet decomposition of nonstationary processes. *Journal of Applied Signal Processing*, 2002(11) :1289–1295, Nov. 2002.
- [500] **C. Vignat** et J.-F. Bercher. Analysis of signals in the Fisher-Shannon information plane. *Physics Letters A*, 312(3) :27–33, 2003.
- [501] **C. Vignat** et J.-F. Bercher. On fisher information inequalities and score functions in non-invertible linear systems. *JIPAM*, 4(4, Article 17), 2003.
- [502] **C. Vignat**, A. Hero, et J. Costa. About closedness by convolution of the Tsallis maximizers. *Physica A*, 340(1-3) :147–152, Sept. 2004.
- [503] Y. Wang, E. Serpedin, **P. Ciblat**, et **P. Loubaton**. Performance analysis of a class of non-data aided carrier frequency offset and symbol timing delay estimators for flat-fading channels. *IEEE Trans. on Signal Processing*, 50(9) :2295–2305, Sept. 2002.

## Actes de conférences

- [504] A. Benazza-Benyahia et **J.-C. Pesquet**. Wavelet-based multispectral image denoising with Bernouilli-Gaussian models. In *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, June.
- [505] A. Benazza-Benyahia et **J.-C. Pesquet**. An extended sure approach for multicomponent image denoising. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2004*, pages 945–948, Montréal, Canada, May 2004.
- [506] A. Benazza-Benyahia et **J.-C. Pesquet**. An interscale multivariate map estimation of multispectral images. In *Proc. European Signal and Image Processing Conference, EUSIPCO 04*, Vienna, Austria, Sept. 2004.
- [507] A. Benazza-Benyahia, **J.-C. Pesquet**, et M. Gharbia. Adapted vector-lifting schemes for compression of polarimetric and multifrequency SAR images. In *IEEE 2003 International Geoscience and Remote Sensing Symposium, Toulouse*, July 2003.
- [508] A. Benazza-Benyahia, **J.-C. Pesquet**, et H. Masmoudi. Block-based adaptive lifting schemes for lossless and progressive image coding. In *Third International workshop on Spectral Methods and Multirate Signal, SMMSP'03*, pages 207–211, Sept. 2003.

- 
- [509] A. Benazza-Benyahia, **J.-C. Pesquet**, et H. Masmoudi. Block-based adaptive lifting schemes for multiband image compression. In *Wavelet Applications in Industrial Processing, Symposium Photonics East 2003, SPIE Conference*, volume 5266, pages 118–128, Oct. 2003.
- [510] **P. Bianchi, P. Loubaton**, et F. Sirven. Performances of a non data-aided estimator of the modulation index of continuous-phase modulations. In *Proc. ICASSP-2002*, pages 2377–2380, May 2002.
- [511] **P. Bianchi, P. Loubaton**, et F. Sirven. Estimation aveugle du débit symbole de modulations cpm. In *Actes du Colloque GretsI 2003*, pages 379–382, Sept. 2003.
- [512] **P. Bianchi, P. Loubaton**, et F. Sirven. On the blind equalization of continuous phase modulated signals using a constant modulus algorithm. In *Proc. SPAWC 2003, Rome.*, pages 442–446, June 2003.
- [513] **P. Bianchi, P. Loubaton**, et F. Sirven. Performances of a non data aided joint estimator of the technical parameters of continuous phase modulated signals. In *Proc. Globecom Communication Theory Symposium, San Francisco*, pages 1731–1735, Dec. 2003.
- [514] **M. Castella, P. Bianchi, A. Chevreuril**, et **J.-C. Pesquet**. Blind mimo detection of convolutively mixed cpm sources. In *Proc. European Signal and Image Processing Conference, EUSIPCO 04*, Vienna, Austria, 2004.
- [515] **M. Castella, A. Chevreuril**, et **J.-C. Pesquet**. Séparation aveugle d’un mélange convolutif de sources non linéaires par une approche hiérarchique. In *GRETSI*, Paris, France, 2003.
- [516] **M. Castella**, E. Moreau, et **J.-C. Pesquet**. A quadratic miso contrast function for blind equalization. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing ICASSP 2004*, pages 681–684, Montréal, Canada, 2004.
- [517] **M. Castella** et **J.-C. Pesquet**. Source separation of a class of non linear time series. In *IEEE-Eurasip Workshop on Non Linear Signal and Image Processing (NSIP)*, Grado, Italy, 2003.
- [518] **M. Castella** et **J.-C. Pesquet**. An iterative blind source separation method for convolutive mixtures of images. In *Proc. International Conference on Independent Component Analysis, ICA 04*, Granada, Spain, 2004.
- [519] **M. Castella, J.-C. Pesquet**, et A. P. Petropulu. New contrasts for blind separation of non iid sources in the convolutive case. In *Proc. European Signal and Image Processing Conference, EUSIPCO 02*, pages 107–110, Toulouse, France, 2002.
- [520] **J.-M. Chaufray, W. Hachem**, et **P. Loubaton**. Asymptotic analysis of optimum and sub-optimum CDMA downlink MMSE receivers. In *Proc. ISIT-2002*, page 189, June 2002.
- [521] **J.-M. Chaufray, P. Loubaton**, et P. Chevalier. Consistent estimation of Rayleigh fading channel second order statistics in the context of the wideband CDMA mode of the UMTS. In *Proc. ICASSP-2001, Salt Lake City*, pages 2189–2192, May 2001.

- 
- [522] **J.-M. Chaufray, P. Loubaton, et W. Hachem.** Asymptotic analysis of optimum and suboptimum CDMA downlink MMSE receivers. In *Proceedings Seventh ISSPA Conf.*, pages 493–496, July 2003.
- [523] **J.-M. Chaufray, P. Loubaton, F. Pipon, et D. Depierre.** Performances asymptotiques de récepteurs sur le lien descendant de l'UMTS-FDD. In *Actes du Colloque Gretsi 2003*, pages 243–246, Sept. 2003.
- [524] **C. Chaux, L. Duval, et J.-C. Pesquet.** Hilbert pairs of  $m$ -band orthonormal wavelet bases. In *Proc. European Signal and Image Processing Conference, EUSIPCO 04*, pages 1187–1190, Vienna, Austria, Sept. 2004.
- [525] **P. Ciblat, P. Loubaton, E. Serpedin, et G. Giannakis.** Performance of non-data aided carrier offset estimation for non-circular transmissions through frequency-selective channels. In *Proc. ICASSP-2000, Istanbul*, pages 2525–2528, June 2000.
- [526] P. L. Combettes et **J.-C. Pesquet.** Image deconvolution with total variation bounds. In *Proc. of the Seventh International Symposium on Signal Processing and Its Applications, Paris*, pages 441–444, July 2003.
- [527] P. L. Combettes et **J.-C. Pesquet.** Incorporating total variation information in image recovery. In *Proc. of the IEEE International Conference on Image Processing, Barcelona*, pages 373–376, Sept. 2003.
- [528] P.-L. Combettes et **J.-C. Pesquet.** Constraint construction in convex set theoretic signal recovery via stein's principle. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2004*, pages 945–948, Montréal, Canada, 2004.
- [529] P.-L. Combettes et **J.-C. Pesquet.** Estimating first-order finite-difference information in image restoration problems. In *IEEE International Conference on Image Processing, ICIP 2004*, Singapore, Oct. 2004.
- [530] J.-A. Costa, A. Hero, et **C. Vignat.** A characterization of the multivariate distributions maximizing Renyi entropy. In *IEEE International Symposium on Information Theory, Lausanne*, page 263, July 2002.
- [531] **M. Debbah, W. Hachem, P. Loubaton, et M. de Courville.** Asymptotic performance analysis for redundant block precoded OFDM systems. In *Proceedings of the 11th IEEE Statistical Signal Processing Workshop, Singapore*, pages 389–392, Aug. 2001.
- [532] **M. Debbah, W. Hachem, P. Loubaton, et M. de Courville.** MMSE analysis of certain large isometric random precoded systems. In *2001 Information Theory Workshop, Cairns, Australia*, pages 158–160, Sept. 2001.
- [533] **M. Debbah, P. Loubaton, et M. de Courville.** Spread OFDM performance with MMSE equalization. In *Proc. ICASSP-2001*, pages 2385–2388, May 2001.
- [534] **M. Debbah, P. Loubaton, et M. de Courville.** Linear precoders for OFDM wireless communications with MMSE equalization : facts and result. In *Proc. Eusipco 2002, Toulouse*, Sept. 2002.
- [535] E. Grosicki, K. Abed-Meraim, **P. Loubaton, et J.-M. Chaufray.** Comparison of downlink mobile positioning methods for the UMTS FDD mode without using IPDL periods. In *Proceedings Seventh ISSPA Conf., Paris*, pages 347–350, July 2003.

- 
- [536] **W. Hachem**, F. Desbouvries, et **P. Loubaton**. Blind channel estimation for CDMA systems : an induced cyclostationarity approach. In *Proc. ICASSP-2000, Istanbul*, pages 2477–2480, June 2000.
- [537] J. Hattay, A. Benazza-Benyahia, et **J.-C. Pesquet**. Adaptive lifting schemes using variable-size block segmentation. In *Advanced Concepts for Intelligent Vision Systems, ACIVS 2004*, Brussels, Belgium, Aug. 2004.
- [538] **S. Houcke**, **A. Chevreuil**, et **P. Loubaton**. Joint blind equalization and estimation of the symbol period : a contrast function approach. In *Proc. ICASSP-2001, Salt Lake City*, pages 2545–2548, May 2001.
- [539] **S. Houcke** et **P. Loubaton**. A weighted linear prediction approach for the blind CDMA forward link channel estimation. In *Proc. ICASSP-2000, Istanbul*, pages 2937–2940, June 2000.
- [540] **P. Jallon**, **A. Chevreuil**, **P. Loubaton**, et P. Chevalier. Separation of convolutive mixtures of cyclostationary sources : a contrast function based approach. In *Proc. ICA '04*, Grenade, Spain, Sept. 2004.
- [541] **S. Lasaulce**, **P. Loubaton**, et E. Moulines. Performance of a subspace based semi-blind technique in the UMTS TDD mode context. In *Proc. ICASSP-2000, Istanbul*, pages 2481–2484, June 2000.
- [542] **S. Lasaulce**, **P. Loubaton**, E. Moulines, et S. Buljore. Training-based channel estimation and de-noising in the UMTS TDD mode. In *Proc. VTC Fall, Atlantic-City*, pages 1908–1911, Oct. 2001.
- [543] **P. Loubaton**. Asymptotic distribution of large random matrices and performance analysis of large CDMA systems. In *Proceedings of Seventh ISSPA, Paris*, pages 205–214, July 2003. Invited Conference.
- [544] **P. Loubaton** et **W. Hachem**. Asymptotic analysis of reduced rank wiener filters. In *Proc. Information Theory Workshop 2003, Paris*, pages 328–331, Apr. 2003.
- [545] P. Maillé, **M. Debbah**, **P. Loubaton**, et M. de Courville. Asymptotic analysis of successive versus parallel interference cancellation schemes for block precoded OFDM. In *Proceedings of the 2001 Allerton Conference*, 2001.
- [546] **B. Mouhouché**, **P. Loubaton**, et **W. Hachem**. Asymptotic analysis of reduced rank chip level MMSE equalizers in the downlink of CDMA systems. In *Proc. of IEEE Int. Conf. SPAWC 2004*, Lisbonne, Portugal, July 2004.
- [547] **B. Mouhouché**, **P. Loubaton**, **W. Hachem**, K. Abed-Meraim, et N. Ibrahim. Analyse asymptotique de certains filtres de Wiener à rang réduit. In *Actes du Colloque GretsI 2003*, pages 163–166, Sept. 2003.
- [548] **B. Mouhouché**, **P. Loubaton**, **W. Hachem**, et N. Ibrahim. Asymptotic analysis fo reduced rank downlink CDMA wiener receivers. In *Proc. of Conf. EUSIPCO-2004*, pages 201–204, Sept. 2004.
- [549] **B. Mouhouché**, K. Meraim, N. Ibrahim, et **P. Loubaton**. Reduced-rank adaptive chip-level MMSE equalization for the forward link of long-code DS-CDMA systems. In *Proceedings. Seventh ISSPA Conf*, pages 497–500, July 2003.

- [550] F. Sagnard, D. Seetharamdoo, et **C. Vignat**. Reflection ellipsometry for in-situ measurements of complex permittivity and thickness of a single-layer material at microwave frequencies : Theory and experiments. In *32nd EUMC, Milan*, Sept. 2002.
- [551] F. Sagnard, **C. Vignat**, V. Moncourtois, et E. Rolland. Détermination de la permittivité complexe de matériaux de construction dans le domaine micro-onde : Comparaison des méthodes de fresnel et d'ellipsométrie micro-onde par réflexion. In *Journées Nationales Microondes, (JNM), Poitiers Futuroscope*, May 2001.
- [552] **S. Touati** et **J.-C. Pesquet**. Wavelet estimation of cyclopectra. In *Proc. International Conference on Acoustics, Speech and Signal Processing, Hong Kong*, pages 289–292, Apr. 2003.
- [553] **C. Vignat** et J.-F. Bercher. Matrix Fisher inequalities for non-invertible linear systems. In *IEEE International Symposium on Information Theory, ISIT 2002, Lausanne*, page 237, July 2002.
- [554] **C. Vignat** et J.-F. Bercher. A Renyi entropy convolution inequality with application. In *EUSIPCO 2002, Toulouse*, 2002.
- [555] **C. Vignat**, J. Costa, et A. Hero. On solutions to multivariate maximum alpha-entropy problems. In *Lecture Notes in Computer Science, Springer-Verlag*, volume 2683, pages 211–228, July 2003.
- [556] Y. Wang, E. Serpedin, **P. Ciblat**, et **P. Loubaton**. Performance analysis of blind carrier frequency offset and symbol timing delay estimators in flat-fading channels. In *Proc. ICASSP-2001, Salt Lake City*, pages 2321–2324, May 2001.

## Collections, livres et chapitres dans les livres

- [557] **P. Loubaton**, E. Moulines, et P. Regalia. *Subspace Methods for Blind Identification and Deconvolution*, volume I of *Signal Processing Advances in Wireless Communications*, chapter 3. Prentice-Hall, 2000.
- [558] B. Pesquet-Popescu et **J.-C. Pesquet**. *Techniques de l'Ingénieur, traité Télécoms, TE 5 215-1*, chapter Ondelettes et applications. Aug. 2001.
- [559] B. Pesquet-Popescu et **J.-C. Pesquet**. *Ondelettes et traitement d'images*, chapter Le traitement des images, pages 193–213. Paris : Hermès Science, 2003.

## Thèses et habilitations

- [560] **P. Bianchi**. *Démodulation aveugle de modulations non linéaires à phases continues*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : P. Comon, P. Duhamel, C. Le Martret, **P. Loubaton**, P. Regalia, F. Sirven, G. Tantot.
- [561] **M. Castella**. *Séparation de sources non linéaires dans le cas des mélanges convolutifs*. Thèse de doctorat, Université de Marne-la-Vallée, 2004. Jury : P. Comon, C. Jutten, É. Moreau, P. Regalia, **A. Chevreuil**, **J.-C. Pesquet**.

- 
- [562] **J.-M. Chaufray.** *Détection et démodulation de stations de bases dans un réseau UMTS.* Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : P. Chevalier, P. Forster, J.-F. Hélar, C. Jutten, **P. Loubaton**, F. Pipon, D. Slock.
- [563] **P. Ciblat.** *Quelques problèmes d'estimation relatifs aux télécommunications non coopératives.* Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : O. Besson, P. Duhamel, M. Granger, **P. Loubaton**, E. Moulines, P. Regalia, G. Vezzosi.
- [564] **M. Debbah.** *Précodeurs linéaires pour les transmissions OFDM sans fils.* Thèse de doctorat, ENS de Cachan, 2002. Jury : J.-C. Belfiore, E. Biglieri, G. Caire, M. de Courville, P. Duhamel, P. Larzabal, **P. Loubaton**.
- [565] **W. Hachem.** *Déconvolution autodidacte au second ordre de signaux vectoriels en présence de bruit additif.* Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : K. Abed-Meraim, F. Desbouvries, P. Duhamel, **P. Loubaton**, P. Regalia, A. Richard, D. Slock.
- [566] **S. Houcke.** *Séparation autodidacte d'un mélange de sources émettant à des débits inconnus et éventuellement différents.* Thèse de doctorat, Université de Marne-la-Vallée, 2002. Jury : **A. Chevreuril**, P. Comon, C. Jutten, C. Le Martret, **P. Loubaton**, E. Moreau, G. Tantot.
- [567] **S. Lasaulce.** *Channel estimation and multiuser detection for TD-CDMA systems.* PhD thesis, ENST, 2001. Jury : J.-C. Belfiore, S. Buljore, P. Duhamel, **P. Loubaton**, E. Moulines, D. Slock.

# Annexe

## Données complémentaires

Les annexes contiennent :

- une liste des thèses et habilitations soutenues dans les équipes pendant la période 2001–2004 ;
- une liste de tous les rapports internes édités par le laboratoire pendant la période 2001–2004 ;
- une liste des invités au séminaire général pendant la période 2001–2004 ;
- une liste d’invités pendant la période 2001–2004 ;
- une description des moyens et de l’environnement technique et administratif du laboratoire.

## Thèses et habilitations

### Habilitations à diriger des recherches

- M.-P. Béal. *Codage symbolique*. Habilitation à diriger des recherches, Université de Marne la Vallée, Jan. 2001. Jury : J.-P. Allouche, G. Cohen, R. Cori, M. Crochemore, C. Frougny, B. Marcus, F. Morain et D. Perrin.
- O. Carton. *Automates et mots infinis*. Habilitation à diriger des recherches, Université de Marne la Vallée, Dec. 2001. Jury : J. Berstel, C. Choffrut, D. Perrin, J.-É. Pin, P. Schupp, G. Sénizergues et W. Thomas.
- F. Hivert. *Combinatoire et calcul symbolique dans les algèbres de Hopf*. Habilitation à diriger des recherches, Université de Marne la Vallée, 2004. Jury : N. Bergeron, C. Frougny, J.-L. Loday, M. Morvan, P. Paule, M. Petkovsek, J.-Y. Thibon.
- T. Kyriacopoulou. *Analyse automatique des textes écrits : le cas du grec moderne*. Mémoire d’habilitation à diriger des recherches, Université de Marne-la-Vallée, 2003. Jury : A. Anastassiadis-Symeonidis, F. Guenthner, É. Laporte, S. Mejri, D. Perrin, E. Ranchhod, T. Symeonidou-Christidou.

- S. Michelin. *Modélisation, simulation et approche temps-réel pour des environnements virtuels complexes*. Habilitation à diriger des recherches, Université de Marne-la-Vallée, 2004. Jury : D. Arquès, R. Caubet, M. Crochemore, J.-C. Grossetie, D. Marini, B. Peroche.
- J.-C. Novelli. *Applications de la combinatoire bijective et algébrique*. Habilitation à diriger des recherches, Université de Marne la Vallée, 2001. Jury : F. Bergeron, M. Bousquet-Melou, M. Crochemore, M. Delest, P. Gastin, M. Habib, D. Krob, J.-Y. Thibon.
- G. Roussel. *Grammaires et automates comme outils pour le développement logiciel*. Habilitation à diriger des recherches, Université de Marne-la-Vallée, Dec. 2003. Jury : J. Berstel, I. Attali, S. Crespi-Reghizzi, B. Lorho et P. Minet.

### Thèses de doctorat

- C. Allauzen. *Combinatoires sur les mots et recherche de motifs*. Thèse de doctorat, Université de Marne la Vallée, Jan. 2001. Jury : J. Berstel, C. Choffrut, M. Crochemore, G. Kutcherov, F. Mignosi et J.-C. Spehner.
- S.-M. Bae. *Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies - méthodes et applications*. Thèse de doctorat, Université de Marne-la-Vallée, 2002. Jury : M. Crochemore, G. Gross, H.K. Kim, É. Laporte, D. Maurel.
- S. Bentolila. *La logique et le vivant ; les formalismes de représentation des connaissances en biologie*. Thèse de doctorat, Université de Marne la Vallée, 2002. Jury : M. Crochemore, M.-C. Maurel, C. Rayssiguier, M.-F. Sagot et D. Thieffry.
- F. Bezerra. *Opérateurs topologiques pour le traitement d'images en niveaux de gris*. Thèse de doctorat, Université de Marne-la-Vallée (France), Nov. 2001. Jury : M. Couprie (directeur de thèse), I. Bloch et F. Prêteux (rapporteurs), D. Arquès, F. Meyer et S. Philipp (examineurs).
- V. Biri. *Techniques d'animation dans les méthodes globales d'illumination*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : D. Arquès, J.-M. Dischler, J.-P. Jessel, S. Michelin, B. Péroche, Y. Rémyon.
- P. Blayo. *Une approche comparative combinatoire pour la prédiction de gènes chez les eucaryotes*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : M. Crochemore, C. Rayssiguier, T. Schiex, M. Gouy, M.-F. Sagot et P. Rouzé.
- G. Chung. *Analyse des constructions à double nominatif/accusatif par l'opération de restructuration en coréen : Classification syntaxique des constructions à adjectifs Sôngsang*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : D. Arquès, A. Dugas, É. Laporte, D. Le Pesant, M.G. Pak, R. Vivès.
- M. Constant. *Grammaires locales pour l'analyse automatique de textes : Méthodes de construction et outils de gestion*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : M. Crochemore, J. Giry-Schneider, F. Guenthner, É. Laporte, D. Maurel.

- 
- M. G. de Carvalho. *Hierarchical Image Analysis through the Tree of Critical Lakes*. Thèse de doctorat, Universidade Estadual de Campinas (Brésil), 2004.
  - P. Dokládal. *Grey-scale image segmentation : a topological approach*. Thèse de doctorat, Université de Marne-la-Vallée (France) and University of technology of Brno (Czech Republic), Dec. 2000. Jury : G. Bertrand, Zdeněk Smékal (co-directeurs de thèse), I. Bloch, J.M. Chassery (rapporteurs), D. Arquès, J. Jiří (examineurs).
  - C. Domingues. *Étude d'outils informatiques et linguistiques pour l'aide à la recherche automatique d'information dans un corpus documentaire*. Thèse de doctorat, Université de Marne-la-Vallée, 2001. Jury : B. Bachimont, M. Bourdeau, P. Lafon, É. Laporte, M. Gross, M. Silberztein.
  - P. Dumont-Bècle. *Simulation de l'aspect des revêtements peinture sur véhicules virtuels*. Thèse de doctorat, Université de Marne-la-Vallée, 2002. Jury : D. Arquès, R. Caubet, A. Kemeny, S. Michelin, B. Péroche, F. Viénot.
  - D.-E. Eum. *Syntaxe des verbes de communication en coréen*. Thèse de doctorat, Université de Marne-la-Vallée, 2004. Jury : J. Giry-Schneider, G. Gross, T. Kyriacopoulou, É. Laporte, C. Leclère, J.S. Nam.
  - R. Forax. *Les multi-méthodes en Java*. Thèse de doctorat, Université de Marne la Vallée, Dec. 2001. Jury : J. Berstel, G. Hedin, C. Queinnec, J. Malenfant, G. Roussel et É. Simon.
  - J. Grazzini. *Analyse multiéchelle d'images météorologiques : application à la détection des zones précipitantes*. Thèse de doctorat, Université de Marne la Vallée, Dec. 2003. Jury : J. Berstel, I. Herlin, F.-X. Le Dimet, E. Memin, J.-P. Nadal, A. Szantai, A. Turiel et H. Yahia.
  - S. Guimarães. *Video transition identification based on 2D image analysis*. Thèse de doctorat, Université Fédérale du Minas Gerais (Brésil), Université de Marne-la-Vallée (France), Mar. 2003. Jury : M. Couprie (co-directeur de thèse), A. de Albuquerque Araújo (co-directeur de thèse), S. Philipp-Foliguet et R. Lotufo (rapporteurs), N.J. Leite (examineur).
  - S.-H. Han. *Les prédicats nominaux en coréen. Constructions à verbe support hata*. Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : C. Cortes, M. Gross, J. Giry-Schneider, É. Laporte, I. Tamba.
  - I. Icart. *Modèles d'illumination pour les couches et multicouches prenant en compte les phénomènes interférentiels*. Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : M. Adler, D. Arquès, R. Caubet, B. Péroche, D. Perrin, F. Sillion.
  - K. Kosawat. *Méthodes de segmentation et d'analyse automatique de textes thaï*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : G. Delouche, J. Désarménien, F. Guenthner, É. Laporte, D. Maurel.
  - V. Le Maout. *Expérience de programmation générique sur des structures non-séquentielles : les automates*. Thèse de doctorat, Université de Marne la Vallée,

- July 2003. Jury : M. Crochemore, T. Lecroq, D. Perrin, D. Revuz et J.-M. Rifflet.
- P. Lecocq. *Simulation d'éclairage temps-réel par des sources lumineuses mobiles et statiques : outils pour la simulation de conduite*. Thèse de doctorat, Université de Marne-la-Vallée, 2001. Jury : D. Arquès, R. Caubet, P. Gauriat, A. Kemeny, S. Michelin, B. Péroche.
  - C. Lohou. *Contribution à l'analyse topologique des images : étude d'algorithmes de squelettisation pour images 2D et 3D, selon une approche topologie digitale ou topologie discrète*. Thèse de doctorat, Université de Marne-la-Vallée (France), Dec. 2001. Jury : G. Bertrand (directeur de thèse), R. Malgouyres et C. Ronse (rapporteurs), D. Arquès et A. Manzanera (examineurs).
  - J. Marchadier. *Modélisation fonctionnelle et topologique pour la vision par ordinateur : application au relèvement de clichés urbains*. Thèse de doctorat, Université de Marne-la-Vallée, 2002. Jury : D. Arquès, G. Bertrand, J.-M. Chassery, Y. Egels, H. Maître, S. Michelin.
  - L. Marsan. *Inférence de motifs structurés : algorithmes et outils appliqués à la détection de sites de fixation dans des séquences génomiques*. Thèse de doctorat, Université de Marne la Vallée, 2002. Jury : S. Dulucq, A. Viari, R. Grossi, J. Berstel, M.-F. Sagot, M. Crochemore et A. Vanet.
  - C. Martineau. *Compression de textes en langue naturelle*. Thèse de doctorat, Université de Marne la Vallée, Dec. 2001. Jury : M. Crochemore, É. Laporte, D. Maurel, G. Plateau, S. Tomi Klein et M. Zipstein.
  - A. Micheli. *Combinatoire des cartes de genre quelconque et arborescences multicoloreuses*. Thèse de doctorat, Université de Marne-la-Vallée, 2001. Jury : D. Arquès, R. Cori, P. Flajolet, H. De Fraysseix, J.-Y. Thibon.
  - M. Monteleone. *Lexicographie et dictionnaires électroniques. Des usages linguistiques aux bases de données lexicales*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : M. Connena, E. d'Agostino, A. Elia, É. Laporte, M. Piot.
  - M. Pantazara. *Syntaxe dérivationnelle du grec moderne : Les constructions verbales à un complément prépositionnel et les constructions nominales et adjectivales predicatives associées*. Thèse de doctorat, Université de Paris VIII – Saint-Denis, 2003. Jury : A. Anastassiadis-Symeonidis, C. Leclère, D. Leeman, M. Piot, R. Vivès.
  - S. Paumier. *De la reconnaissance de formes linguistiques à l'analyse syntaxique*. Thèse de doctorat, Université de Marne-la-Vallée, 2003. Jury : Ch. Choffrut, F. Guenthner, É. Laporte, J.S. Nam, D. Perrin.
  - C. Rispal. *Automates sur les ordres linéaires : Complémentation*. Thèse de doctorat, Université de Marne la Vallée, Dec. 2004. Jury : J. Berstel, V. Bruyère, D. Caucal, O. Carton, D. Perrin, J.-É. Pin.
  - A. Savary. *Recensement et description des mots composés – méthodes et applications*. Thèse de doctorat, Université de Marne-la-Vallée, 2000. Jury : G. Gross, M. Gross, F. Guenthner, J. Humbley, Ch. Jacquemin, É. Laporte, M. Silberztein.

- C. Sibade. *Compression de données pour les systèmes de traitement de document grand format*. Thèse de doctorat, Université de Marne-la-Vallée, Dec. 2003. Jury : M. Akil (directeur de thèse), L. Perroton (co-directeur), P. Garda, M. Kunt (rapporteurs), M. Crochemore, S. Barizien (examineurs).

## Rapports internes

Cette section contient la liste des rapports internes produits entre 2001 et 2004. Ces rapports constituent des prépublications qui permettent aux chercheurs de diffuser les résultats de certaines de leurs recherches rapidement, avant leur publication dans les revues et colloques adaptés.

### Année 2004

- IGM 2004-12, M. Couprie, R. Zour, Discrete bisector function and Euclidean skeleton in 2D and 3D, Institut Gaspard Monge, 2004
- IGM 2004-11, M. Couprie, L. Najman, G. Bertrand, Quasi-linear algorithms for the topological watershed, Institut Gaspard Monge, 2004
- IGM 2004-10, G. Bertrand, On topological watersheds, Institut Gaspard Monge, 2004
- IGM 2004-09, F. Hivert, N. M. Thiéry, MuPAD-Combinat, an Open-Source Package for Research in Algebraic Combinatorics, Institut Gaspard Monge, 2004
- IGM 2004-08, X. Daragon, M. Couprie and G. Bertrand, Derived neighborhoods and frontier orders, Institut Gaspard Monge, 2004
- IGM 2004-07, Gohsran Chung, Analyse des constructions à double nominatif/accusatif par l'opération de restructuration en coréen. Classification syntaxique des constructions à adjectif sôngsang – Annexes, « Thèse » Institut Gaspard Monge, 2004
- IGM 2004-06, M.P. Béal, F. Fiorenzi, D. Perrin, A hierarchy of irreducible sofic shifts, Institut Gaspard Monge, 2004
- IGM 2004-05, X. Daragon, M. Couprie and G. Bertrand, Discrete surfaces and frontier orders, Institut Gaspard Monge, 2004
- IGM 2004-04, L. Najman, M. Couprie and G. Bertrand, Watersheds, extension maps, and the emergence paradigm, Institut Gaspard Monge, 2004
- IGM 2004-03, J. Allali, M.-F. Sagot, The at-most K-deep factor tree, Institut Gaspard Monge, 2004
- IGM 2004-02, Ch. Deleray, Bedon, G. Roussel, E. Duris, Corosol : a component-based and reflexive JVM dynamically customizable, Institut Gaspard Monge, 2004
- IGM 2004-01, M. Crochemore, R. Giancarlo, M.-F. Sagot, Longest Motifs with a Functionally Equivalent Central Block , Institut Gaspard Monge, 2004

## Année 2003

- IGM 2003-14, M.P. Béal, M. Crochemore, G. Fici, Presentations of constrained systems with unconstrained positions, Institut Gaspard Monge, 2003
- IGM 2003-13, M.P. Béal, F. Fiorenzi, D. Perrin The syntactic graph of a sofic shift, Institut Gaspard Monge, 2003
- IGM 2003-12, M.P. Béal, A. Bergeron, S. Corteel, M. Raffinot, An algorithmic view of gene teams, Institut Gaspard Monge, 2003
- IGM 2003-11, K. Kosawat, Méthodes de segmentation et d'analyse automatique de textes thaï- Annexes, « Thèse » Institut Gaspard Monge, 2003
- IGM 2003-10, S. Paumier, De la reconnaissance de formes linguistiques à l'analyse syntaxique, « Thèse » Institut Gaspard Monge, 2003
- IGM 2003-09, D. Perrin, G. Rindone, On syntactic groups, Institut Gaspard Monge, 2003
- IGM 2003-08, D. Perrin, Automata and formal languages, Institut Gaspard Monge, 2003
- IGM 2003-07, Sun-Mee Bae, Le dictionnaire électronique des séquences nominales figées en coréen et de leurs formes fléchies. Méthodes et applications, « Thèse » Institut Gaspard Monge, 2003
- IGM 2003-06, P. Blayo, Une approche comparative combinatoire pour la prédiction de gènes chez les eucaryotes, « Thèse » Institut Gaspard Monge, 2003
- IGM 2003-05, M.P. Béal, A note on Cerny's Conjecture an rational series, Institut Gaspard Monge, 2003
- IGM 2003-04, M.P. Béal, M. Crochemore, F. Mignosi, A. Restivo, M. Sciortino, Forbidden words of regular languages, Institut Gaspard Monge, 2003
- IGM 2003-03, M.P. Béal, D. Perrin, On the generating sequences of regular languages on k symbols , Institut Gaspard Monge, 2003
- IGM 2003-02, M.P. Béal, F. Fiorenzi, F. Mignosi, Minimal forbidden patterns of multi-dimensional shifts , Institut Gaspard Monge, 2003
- IGM 2003-01, S. Bentolila, La logique et le vivant les formalismes de représentation des connaissances en biologie « Thèse », Institut Gaspard Monge, 2003

## Année 2002

- IGM 2002-10, N. Pisanti, M. Crochemore, R. Grossi, M.F. Sagot, A Basis for Repeated Motifs in pattern Discovery and Text Mining, Institut Gaspard Monge, 2002
- IGM 2002-09, J. Marchandier, Modélisation fonctionnelle et topologique pour la vision par ordinateur : Application au relèvement de clichés « Thèse », Institut Gaspard Monge, 2002

- IGM 2002-08, P. Dumont-Bècle, Simulation de l'aspect des revêtements peinture sur véhicules virtuels « Thèse », Institut Gaspard Monge, 2002
- IGM 2002-07, L. Marsan, Inférence de motifs structurés : algorithmes et outils appliqués à la détection de sites de fixation dans des séquences génomiques « Thèse », Institut Gaspard Monge, 2002
- IGM 2002-06, J.M. Champarnaud, G. Hansel, D. Perrin, Unavoidable sets of constant length, Institut Gaspard Monge, 2002
- IGM 2002-05, P. Lecocq, Simulation d'éclairage temps réel par des sources lumineuses mobiles et statiques : outils pour la simulation de conduite « Thèse », Institut Gaspard Monge, 2002
- IGM 2002-04, C. Martineau, Compression de textes en langue naturelle « Thèse », Institut Gaspard Monge, 2002
- IGM 2002-03, A. Micheli, Combinatoire des cartes de genre quelconque et arborescences multicolours « Thèse », Institut Gaspard Monge, 2002
- IGM 2002-02, M. Crochemore, G.M. Landau, M. Ziv-Ukelson, A Sub-quadratic Sequence Alignment Algorithm for Unrestricted Cost Matrices « revision », Institut Gaspard Monge, 2002
- IGM 2002-01, O. Carton, Automates et mots infinis « Mémoire d'habilitation à diriger des recherches », Institut Gaspard Monge, 2002

## Année 2001

- IGM 2001-08, M. Crochemore, G.M. Landau, M. Ziv-Ukelson A Sub-quadratic Sequence Alignment Algorithm for Unrestricted Cost Matrices, Institut Gaspard Monge, 2001
- IGM 2001-07, O. Carton, M.P. Béal, Determinization of transducers over infinite words II, Institut Gaspard Monge, 2001
- IGM 2001-06, O. Carton, M. Michel, Unambiguous Büchi automata, Institut Gaspard Monge, 2001
- IGM 2001-05, O. Carton, R-Trivial languages of words on countable ordinals, Institut Gaspard Monge, 2001
- IGM 2001-04, C. Domingues, Etude d'outils informatiques et linguistiques pour l'aide à la recherche automatique d'information dans un corpus documentaire « Thèse », Institut Gaspard Monge, 2001
- IGM 2001-03, J. Berstel, S. Crespi Reghizzi, G. Roussel, P. San Pietro A Scalable Format Method for Design and Automatic Checking of User Interfaces, Institut Gaspard Monge, 2001
- IGM 2001-02, R. Forax, E. Duris, G. Roussel, A Simple Dispatch Technique for Pure Java Multi-Methods, Institut Gaspard Monge, 2001
- IGM 2001-01, G. Duchamp, E. Laugerotte, J-G. Luque, Extending the scalars of minimization, Institut Gaspard Monge, 2001

## Séminaires

Cette section contient la liste des invités au séminaire général pour les années allant de 2001 à 2004.

### Année universitaire 2004-2005 (début)

- Gad M. Landau (Université de Haifa) « *Sparse Normalized Local Alignment* »
- James Currie (LRI, Orsay) « *Abelian Pattern Avoidance* »
- Jean-Pierre Borel (Université de Limoges) « *Facteurs palindromes dans les mots du billard* »
- Isabelle Fagnot (IGM) « *Splicing Systems and Chomsky Hierarchy* »
- Arnaud Carayol (IRISA) « *La hiérarchie des graphes des automates à pile de piles* »
- Gabriele Fici (IGM) « *Le problème de l'assemblage de fragments à travers les mots interdits minimaux* »

### Année universitaire 2003-2004

- Jens Stoye (Univ. Bielefeld) « *Finding Repetitive Structures in Large Sequences* »
- Alfredo Viola (Monte Video, Uruguay) « *Adaptive Sampling for Quickselect* »
- Gilles Bertrand (ESIEE, IGM) « *Sur la ligne de partage des eaux* »
- Bodo Lass (Institut Girard Desargues, Université Claude Bernard, Lyon 1) « *Dénombrement des colorations par listes* »
- Nicolas Gürel (LIX, IGM) « *Introduction à l'algorithmique des courbes superelliptiques, applications à la cryptographie* »
- Rajeev Raman (Computer Science Department, University of Leicester) « *Succinct tree representations for XML documents* »
- Amihoud Amir (Bar-Ilan University, Israel), Gad M. Landau (University of Haifa, Israel) « *Efficient Rotated Matching* »
- Juha Kärkkäinen (Max-Planck-Institut für Informatik, Saarbrücken, Germany) « *Sorting suffixes with difference cover samples* »
- Fabien Coulon (LIFAR, Université de Rouen) « *Algorithmes rapides et compacts pour la recherche d'expressions rationnelles* »
- Marc Chemillier (Université de Caen), Gérard Assayag (Ircam) « *Système d'improvisation par ordinateur utilisant l'oracle des facteurs* »
- Jean-Éric Pin (LIAFA, Paris) « *Opérations qui préservent les langages reconnaissables* »
- Martin Matamala (Santiago, Chile) « *Upper bounds for domination numbers* »
- Bertil Folliot (LIP6, Paris) « *Systèmes dynamiquement reconfigurables : la machine virtuelle* »

- Éric Tannier (Université de Lyon 1 et INRIA) « *Une méthode rapide pour le tri par inversions* »
- Alain Lascoux (IGM, Marne-la-Vallée) « *Monge et l'équation des coniques planes* »
- Violetta Lonati (Università degli Studi di Milano) « *Local limit distributions in pattern statistics : beyond the Markovian models* »
- Nicolas Thiéry (IGM et Lyon I) « *MuPAD-Combinat, a package for Computer exploration in algebraic combinatorics* »
- Jérôme Petazzoni (IGM, Marne la Vallée) « *Deplika : un protocole de transfert de fichier Multicast fiable sur des réseaux Ethernet* »
- Olivier Serre (LIAFA, Paris 7) « *Jeux sur des graphes d'automates à pile* »
- Irina Rusu (IRIN, Université de Nantes) « *Sur la complexité algorithmique de quelques problèmes posés par la biologie* »
- Marni Mishna (LaBRI) « *On the benefits of being D-finite : Combinatorial classification in theory and in practice* »
- Jérôme Durand-Lose (LIP, projet MC2 et I3S, Université de Nice-Sophia Antipolis) « *Calculer géométriquement sur le plan* »
- Alexis Nasr (Lattice, Paris 7) « *Utilisation d'automates finis pondérés pour l'étiquetage morpho-syntaxique et l'analyse syntaxique partielle* »
- Jacques Malenfant (LIP6, UMR 7606 UPMC/CNRS) « *Programmation systématique = Réflexion asynchrone + Contrôle* »
- Jean-Christophe Novelli (IGM, Marne-la-Vallée) « *Algèbres de Hopf combinatoires : l'exemple des fonctions de parking* »
- Chloé Rispal (IGM) « *Mots indexés par des ordres linéaires : Complémentation* »
- Jacques Mazoyer (LIP Lyon) « *automates cellulaires(Sujet en rapport avec le workshop)* »
- Christian Mauduit (IML, Marseille) « *Suites finies pseudo-aléatoires* »
- Christiane Frougny (LIAFA, Paris 8) « *Ensembles de Meyer arithmétiques* »
- Pablo Arrighi « *Calcul quantique à l'aveugle* »

### Année universitaire 2002-2003

- Dominique Perrin (IGM, Marne-la-Vallée) « *Sur les relations entre un nombre fini de mots* »
- Renaud Pawlak (CNAM, Paris) « *JAC : un environnement de développement orienté aspect* »
- Christophe Morvan (IGM, Marne-la-Vallée) « *Les graphes rationnels sur les traces des langages contextuels* »
- Mathieu Raffinot « *The Algorithmic of Gene Teams* »
- Jean-Gabriel Luque (IGM) « *Hankel hyperdeterminants and Selberg integrals* »
- Srečko Brlek (UQAM) « *Combinatorial properties of smooth infinite words* »
- Juhani Karhumäki (Dept. Math and TUCS, University of Turku, Finland) « *Chal-*

- lenges on finite sets of words* »
- Sophie Schbath (INRA) « *Statistiques des comptages de mots dans les séquences* »
  - Edmund Harriss (LRI) « *Non-periodic tilings, substitutions and sturmian words* »
  - Julien Cervelle (IGM) « *Dynamique des automates cellulaires et complexité algorithmique. Automates de sable* »
  - Franck Pommereau (Paris 12) « *Temps causal dans les M-nets* »
  - Massimiliano Goldwurm (Dip. Scienze dell’Informazione, Università degli Studi di Milano) « *On the number of occurrences of a symbol in words of regular languages* »
  - William Y. C. Chen (Center for Combinatorics Nankai University, China) « *Symmetries on plane trees* »
  - Emmanuel Briand (IGM) « *Polynômes multisymétriques* »
  - Rémi Forax (IGM) « *Objets persistants et héritage* »
  - Tanguy Urvoy (IRISA) « *Familles abstraites de graphes* »
  - François Morain (LIX) « *La primalité est dans P* »
  - Luc Boasson (LIAFA) « *Mélange de langages et langages mélangés* »
  - Jean-Loup Guillaume (LIAFA) « *Modélisation de graphes petit-monde* »
  - Claude Del Vigna (CAMS) « *Ambiguïtés Irréductibles dans les Monoïdes de Mots* »
  - Pascal Weil (LaBRI, CNRS et Université Bordeaux-1) « *Sur le rang de l’intersection de sous-groupes finiment engendrés du groupe libre et la conjecture de Hanna Neumann* »
  - Dominique Poulalhon (LIX) « *Construction bijective de triangulations* »
  - Stéphane Vialette (LGM-ENS) « *Identification et classification de promoteurs* »
  - Clémence Magnien (LIX) « *Étude du modèle du tas de sable : points de vue algorithmique et algébrique* »
  - Benjamin Audit (EMBL-European Bioinformatics Institute, Cambridge) « *Modeling the percolation of annotation errors in a database of protein sequences* »
  - Rémy Malgouyres (LLAIC, Clermont) « *Complexité de la décision d’existence d’une déformation continue entre deux objets topologiques discrets* »
  - Thierry Cachat (RWTH Aachen) « *Jeux sur des graphes infinis* »
  - Nathalie Friburger (Université de Tours) « *Reconnaissance automatique des noms propres et application à la classification automatique de textes journalistiques* »
  - Alessandra Carbone (Institut des Hautes Etudes Scientifiques) « *Codon Adaptation Index as a measure of dominating codon bias* »
  - Rémy Malgouyres (LLAIC, Clermont) « *Complexité de la décision d’existence d’une déformation continue entre deux objets topologiques discrets* »
  - Renaud Keriven (ENPC, CERMICS / DI-ENS) « *Vision et réalité augmentée* »
  - Michel Nguyen The (IGM, LIX) « *Distribution limite de la taille d’arbres réduits* »
  - Michel Couprie (A2SI, ESIEE) « *Topologie des images en niveaux de gris* »

**Année universitaire 2001-2002**

- Philippe Flajolet (INRIA) « *Langages, dénombrements et asymptotique algébriques* »
- Bernadette Charron-Bost (LIX) « *Problèmes d'accord dans les systèmes distribués tolérant les pannes* »
- Gilles Schaeffer (LORIA) « *Sur le nombre de noeuds et d'entrelacs premiers alternants* »
- Jean-Christophe Novelli (IGM) « *Combinatoire des pseudo-permutations* »
- Pierre-Cyrille Héam (LIAFA) « *Idéaux de shuffle* »
- Maurice Nivat (LIAFA) « *Suites doubles F-homogènes* »
- Marie-France Sagot (Rhône-Alpes) « *Some approximation results for the Maximum Agreement Forest (MAF) problem* »
- Anna Gambin (LORIA) « *Contextual alignment of biological sequences* »
- Georges Hansel « *Ensembles inévitables et classes de conjugaison* »
- Thierry Lecroq (LIFAR) « *La recherche delta et gamma approchée dans les séquences musicales* »
- Venceslas Biri (IGM) « *Animation dans les méthodes globales d'illumination* »
- Gilles Roussel (IGM) « *Ordonnement des déplacements dans un réseau de robots* »
- Boris Adamczewski (IML) « *Propriétés d'équilibre pour les points fixes de substitutions primitives* »
- Jean-Christophe Dubacq (LRI) « *Signaux pour automates cellulaires en dimension 2 ou plus* »
- Jean-Christophe Aval (Université de Bordeaux) « *Fonctions quasi-symétriques et chemins de Catalan* »
- Raffaele Giancarlo (Universita' Di Palermo) « *Improving Table Compression with Combinatorial Optimization* »
- Carole Delporte (LIAFA), Hugues Fauconnier (LIAFA) « *Diffusion générique et diffusion atomique averse* »
- Marianne Durand (INRIA) « *Tout ce que vous avez toujours voulu savoir sur Quicksort...* »
- Volker Strehl (Computer Science Institute, University of Erlangen-Nürnberg, Germany) « *Pavages, Chemins, Résultants, Tableaux* »
- Francesca Fiorenzi (LIX) « *Cellular Automata and Finitely Generated Groups* »
- Shmuel Safra « *The Importance of Being Biased* »
- Anne Bergeron (LACIM, UQAM) « *Combinatoire et algorithmique du tri des permutations signées* »
- Antonio Restivo (Université de Palerme) « *Decipherability conditions and Kraft inequality* »
- Arnaud Lefebvre (LIFAR, Rouen) « *Mots sur un alphabet ordonné et permutation* »

- standard des suffixes* »
- Anahí Gajardo Schulz (Universidad de Concepcion, CHILI) « *Représentation symbolique de la dynamique d'un agent : la fourmi de Langton* »
  - Arash Habibi (Université Louis Pasteur, Strasbourg) « *De la fumée aux bulles de savon, Animation et visualisation des fluides en synthèse d'images* »
  - Julien Cassaigne (IML Marseille) « *The Kolakoski sequence and its conjectured subword complexity* »
  - Gabriela Andrejková (Department of Computer Science, Faculty of Science P. J. Safarik University Kosice, Slovakia) « *Computational power of neural networks* »
  - Mireille Bousquet-Mélou (laBRI, Bordeaux) « *Combinatoire énumérative : l'exemple des animaux* »
  - Christophe Reutenauer (LaCIM, UQAM) « *La théorie des caractères non commutatifs du groupe symétrique de Armin Jöllenbeck* »
  - Jérémie Bourdon (GREYC, université de Caen) « *Sources dynamiques et recherche de motifs* »
  - Filippo Mignosi (Palerme) « *Compact representation of local automata* »
  - Ralf Klasing (King's College London) « *Communication in Interconnection Networks in the Disjoint-Paths Modes* »
  - Paul Schupp (University of Illinois) « *Generic-case complexity and decision problems in group theory* »
  - Giulio Pavesi (Università di Milano-Bicocca) « *Pattern discovery in biological sequences* »
  - Giancarlo Mauri (Università di Milano-Bicocca) « *Circular splicing systems* »
  - Christine Heitsch (University of British Columbia) « *The Unavoidability of Generalized Zimin Word Constructions* »

## Année universitaire 2000-2001

- Pierre Boullier (INRIA) « *Les grammaires à concaténation d'intervalles* »
- Maxime Crochemore (IGM) « *Indexation de textes* »
- Philippe Narbel (LaBRI) « *Ensembles de courbes, mots de complexité linéaire et substitutions* »
- Olivier Carton (IGM) « *Prédicats morphiques et extensions* »
- Mesaac Makpangou (INRIA) « *Réplication des données sur l'Internet : concilier cohérence et performance* »
- Roberto Di Cosmo (Université Paris VII) « *Sécurité, informatique et vie privée : le problème propriétaire, le danger monopoliste, la solution libre* »
- Journée de présentation du laboratoire d'informatique
- Sylvain Lombardy (ENST) « *Dérivation d'expressions avec multiplicité* »
- Nadia Pisanti (Université de Pise) « *On syntenic distance between genomes* »
- Sylvain Michelin (IGM) « *Les milieux participants en synthèse d'images (brouillard,*

- fumée...)* »
- Didier Caucal (IRISA, Rennes) « *Une hiérarchie de Chomsky de familles de graphes* »
  - Alain Lascoux (IGM) « *Calcul algébrique en plusieurs variables* »
  - Jean-Gabriel Luque (IGM) « *Minimisation d'automates à multiplicité dans des anneaux principaux* »
  - Anne Siegel (Institut de Mathématiques de Luminy) « *Propriétés géométriques des fractales de Rauzy et automates* »
  - Julien Cassaigne (Institut de Mathématiques de Luminy) « *Complexité palindromique* »
  - Julien Clément (GREYC, Université de Caen) « *Arbres digitaux et sources dynamiques* »
  - Sam Toueg (Département d'Informatique) « *On the Quality of Service of Failure Detectors* »
  - Cédric Chauve (LaBRI) « *Recherche de motifs dans un arbre* »
  - Fabrice Guillemin (France Télécom R & D) « *Modélisation des réseaux large bande et problèmes combinatoires* »
  - Michel Habib (LIRM) « *Algorithmes de graphes et affinage de partition* »
  - Jocelyn Marchadier (IGM) « *Extraction de primitives* »
  - Nicolas Thiéry (LAPCS, Lyon I) « *Invariants algébriques de graphes et reconstruction* »
  - Nozha Boujemaa (INRIA) « *Indexation et recherche d'images par le contenu* »
  - Véronique Bruyère (Université de Mons-Hainaut) « *Autour du théorème de Kleene* »
  - Richard Cole (New York University/King's College London) « *Cache oblivious search trees* »
  - Laurent Marsan (IGM) « *Extraction de modèles structurés communs à plusieurs textes - Application à la détection de mots conservés dans des séquences d'ADN* »
  - Annie Luciani (INPG) « *Synthèse d'images animées et simulation multisensorielle à retour d'effort par modèles physiques particuliers temps réel* »
  - Christiane Frougny (LIAFA) « *Addition en ligne en base réelle* »
  - Juraj Hromkovic (RWTH, Aachen) « *Stability of Approximation algorithms for hard optimization problems* »
  - Alain Pajor (Université de Marne-la-Vallée) « *Algorithmes géométriques en temps polynomial* »
  - Gohsran Chung (IGM) « *Classification syntaxique des constructions adjectivales à substantifs appropriés en coréen* »
  - Laurent Viennot (INRIA) « *Surcharge de trafic dans les réseaux radio ad-hoc* »
  - Wojciech Szpankowski (Purdue University) « *Pattern matching image and video compression : theory algorithms and experiments* »
  - Roberto Grossi (Universita di Pisa) « *Full text indexing in small space* »

## Moyens et environnement

### Dotations budgétaires

Il s'agit des dotations hors salaires et infrastructures.

	2001	2002	2003	2004
Ministère				
Fonctionnement	30 k€	30 k€	123 k€	100 k€
Équipement	85 k€	65 k€	70 k€	80 k€
ESIEE				
Fonctionnement	5 k€	5 k€	5 k€	5 k€
Équipement	10 k€	10 k€	10 k€	10 k€
CNRS				
Fonctionnement	24 k€	26 k€	36 k€	37 k€
Équipement	12 k€	12 k€		
AS	30 k€	41 k€	23 k€	
DRI	3 k€			1 k€
Contrats	50 k€	68 k€	168 k€	370k€

### Locaux

L'université de Marne-la-Vallée et l'ESIEE accueillent les équipes du laboratoire dans leurs locaux. À l'université, le laboratoire occupe une partie du quatrième étage du bâtiment Copernic et à l'ESIEE au deuxième étage de l'épi 5 du bâtiment principal. Les équipes sont réparties essentiellement de façon thématique.

### Personnel administratif et technique

L'administration du laboratoire est assurée par Andrée Impérial et Nelly Muntz-Berger dont les bureaux se trouvent au quatrième étage du bâtiment Copernic de l'université.

L'administration système et réseau des équipements du laboratoire est assurée par Patrice Hérault à l'université et par Eric Llorens à l'ESIEE.

Outre ses activités de développement scientifique, Teresa Gomez-Diaz assure la maintenance des sites internet et intranet du laboratoire. Depuis 2004, elle est également correspondante formation du laboratoire pour le CNRS.

Le laboratoire accueille aussi l'administration de l'Institut Gaspard-Monge, en particulier, son secrétariat assuré par Line Fonfrède.



FIG. 2: *Point de rencontre autour d'un café*

## **Bibliothèque**

La bibliothèque de recherche, commune aux laboratoires de mathématiques et d'informatique de l'université est située dans le bâtiment Copernic, à proximité des locaux du laboratoire. Elle constitue une composante de la bibliothèque universitaire et bénéficie ainsi de ses services, dont l'accès en ligne au catalogue et l'accès aux périodiques des partenaires du Polytechnicum de la cité Descartes de Marne-la-Vallée. Elle a pour vocation d'accueillir les chercheurs et les étudiants de troisième cycle en informatique et mathématiques. Elle regroupe déjà plusieurs fonds :

- un fond important de livres (environ 5000 volumes) de Linguistique et d'Informatique Linguistique contenant des livres anciens, certains datant du 16ème siècle, des dictionnaires et des encyclopédies ;
- un fond de livres d'informatique dont la plupart a été acquise dans les dernières années ; des thèses et compte-rendus de conférences récents ; plusieurs collections de revues.

Ces livres sont entièrement catalogués et le service de prêt fonctionne depuis le premier semestre 1995.

## **Équipement informatique**

Les chercheurs du laboratoire ont accès à différents types de matériels informatiques qui comprennent principalement :

- à l'université :
  - 103 postes fixes (PC) avec systèmes Linux et Windows ;

- 
- 13 portables dont 2 sont réservés pour les déplacements ;
  - 2 serveurs de fichiers (Bi-P4 2.8Ghz 1Go/300Go) sous NFS (Unix) et SAMBA (Windows) qui stockent de façon centralisée les fichiers des membres du laboratoire et assure leur sauvegarde ;
  - 1 serveur WEB/Messagerie (P4 2.8Ghz 512Mo/80 Go) ;
  - 3 serveurs de calcul (Bi-P3 1Ghz 4Go/55Go, Bi-P4 3Ghz 6Go/180Go, Bi-Opteron-64bit 2.2 GHz 4Go/80Go) utilisés pour les applications gourmandes en mémoire et en temps de calcul telles que les applications de calcul symbolique ou de bio-informatique ;
  - 1 serveur CVS (Bi-P3 1Ghz 512Mo/144Go) pour assurer la pérennité des logiciels développés dans le laboratoire comme Unitex, JMMF, Corosol, SMILE ou les bibliothèques MuPAD ;
  - 2 salles équipées de vidéo-projecteur pour le séminaire et les cours de DÉA ;
  - 2 vidéo-projecteurs portables ;
  - 3 imprimantes laser ;
  - 2 imprimantes couleur ;
  - à l'ESIEE :
    - 15 postes fixes (PC) ;
    - 3 portables ;
    - 1 imprimante laser ;
    - 1 imprimante couleur ;
    - 1 vidéo projecteur.