

Activités du laboratoire Incognito

Département d'informatique et de recherche opérationnelle

Université de Montréal

Janvier 1993 - Décembre 1994

édité par
Nicolas Anquetil

Janvier 1995

Table des matières

1	Présentation	1
2	Professeurs	2
2.1	Michel BOYER	2
2.2	Guy LAPALME	2
2.3	Jacques MALENFANT	3
2.4	Jian-Yun NIE	4
2.5	Jean VAUCHER	5
3	Etudiants et étudiantes au Ph.D.	6
3.1	Nicolas ANQUETIL (Directeur : Jean Vaucher)	6
3.2	Martin BRISEBOIS (Directeur : Jian-Yun Nie)	6
3.3	Eliana COELHO (Directeur : Guy Lapalme et Vimla L. Patel)	7
3.4	François-Nicola DEMERS (Directeur : Jacques Malenfant)	8
3.5	Massimo FASCIANO (Directeur : Guy Lapalme)	8
3.6	Leila KOSSEIM (Directeurs : Guy Lapalme et Richard Kittredge)	9
3.7	Marco JACQUES (Directeur : Jacques Malenfant)	10
3.8	Gilles PESANT (Directeur : Michel Boyer)	10
3.9	Daniel ROUSSEAU (Directeurs : Guy Lapalme et Bernard Moulin)	11
3.10	Gilles ROY (Directeur : Jean Vaucher)	12
3.11	Akila SARIRETE (Directeur : Jean Vaucher)	12
3.12	Stéphane Sotèg SOMÉ (Directeurs : Rachida Dssouli et Jean Vaucher)	13
3.13	Marcel TURCOTTE (Directeurs : Guy Lapalme et Robert J. Cedergren)	14
4	Étudiante au M.Sc.	15
4.1	Hong ZHAO (Directeur : Jian-Yun Nie)	15
5	Diplômé(e)s	16
5.1	Diplômé(e)s au Ph.D.	16
5.2	Diplômé(e)s au M.Sc.	16
6	Publications	18
7	Projets spéciaux	21
7.1	Réflex	21
7.2	Scriptum	22
8	Les séminaires Incognito	24
9	Incognito par l'Internet	27

1 Présentation

Le laboratoire Incognito (informatique cognitive) a été fondé en 1983 par les professeurs Paul Bratley, Guy Lapalme et Jean Vaucher afin d'étudier la structuration des connaissances, la programmation logique et fonctionnelle, les systèmes experts et d'autres aspects liés à l'intelligence artificielle. C'est un groupe informel sans structure administrative fixe qui réunit présentement les efforts de recherche de Michel Boyer, Guy Lapalme, Jacques Malenfant, Jian-Yun Nie et Jean Vaucher. Les membres d'Incognito ont établi des collaborations suivies avec des professeurs et étudiants des départements de linguistique et philologie ainsi qu'avec des laboratoires étrangers, l'École des Mines de Nantes et l'Institut de Recherche en Informatique de Toulouse.

Incognito, en tant que laboratoire, n'a jamais été subventionné directement; c'est grâce à la bonne volonté et à l'union des modestes moyens des professeurs et chercheurs impliqués qu'il a pu survivre. Au point de vue matériel, nous avons obtenu en 1984 un don de 3 machines Lisp Xerox 1109 (en plus des 3 achetées à un prix préférentiel). En 1988, nous avons obtenu une subvention de l'Université de Montréal qui a payé pour la "moitié" d'une station Sun-3/60 (les professeurs impliqués ont payé l'autre moitié). A l'été 1989, nous avons reçu un don de la compagnie Apollo pour deux stations DN-3500. Ces matériels sont maintenant disparus après avoir rendu de fiers services et permis à plusieurs étudiants de compléter leurs travaux sur des équipements de pointe. Par la suite, deux Mac-CX ont été achetés dans le cadre de subventions pour le projet Alto. En 1992, le projet Macroscope a permis de financer l'achat d'un nouveau serveur Sun (Sparc II), de deux machines Sun (Sparc station ELC), d'un terminal X (NCD) et de différents supports de mémoire de masse (disque de 1,4 Go, lecteur de disquette optique 125 Mo). En 1993-1994, des contributions de professeurs ont permis l'agrandissement de notre parc d'équipement par l'achat de deux nouvelles stations Sparcs LX, d'un terminal X et de mémoire de masse (deux disques de 2 Go). Enfin, divers achats (changement de la carte CPU, et achats de mémoire vive) et le don d'une deuxième carte CPU par le Département d'Informatique et de Recherche Opérationnelle ont permis d'augmenter considérablement la puissance de calcul de notre serveur, le transformant de Sparc II en Sparc 10/512 (2 processeurs, 80 Mo de mémoire vive). Le déménagement, en mai 1994, dans le pavillon André-Aisenstadt nous a permis de rationaliser notre aménagement matériel et de mieux organiser l'espace de travail pour les étudiants.

Ce rapport résume l'ensemble des projets et des activités des deux dernières années (de janvier 1993 à décembre 1994). Il fait suite à des rapports similaires parus durant les années précédentes. D'ailleurs le premier rapport Incognito date de mai 1985 et traite de la période 1984-1985. Celui-ci marque donc le 10^{me} anniversaire d'un des plus anciens laboratoires du DIRO. Il a été rédigé par Nicolas Anquetil à partir des informations disponibles à la fin décembre 1994. Pour obtenir des renseignements supplémentaires, il est préférable de s'adresser directement au professeur responsable du projet.

2 Professeurs

2.1 Michel BOYER

Mon domaine majeur d'intérêt depuis maintenant plusieurs années est la programmation logique et ses extensions. Dans ce cadre s'inscrivent mes travaux sur les grammaires d'unification, sur la méta- interprétation et maintenant sur l'étude des mécanismes de contraintes en Prolog. Cette dernière étude, faite principalement avec un étudiant de doctorat, donnera incessamment lieu à une thèse. On y démontre l'intérêt et la viabilité de l'ajout de contraintes quadratiques dans un Prolog qui possède déjà les contraintes linéaires.

Je m'intéresse aussi à l'application de méthodes mathématiques en informatique. Dans une recherche faite en collaboration avec Neil Stewart sur la modélisation de solides, nous avons utilisé des méthodes de topologie générale et de topologie linéaire par morceaux pour clarifier des définitions suggérées par Requicha en modélisation de solides; nous avons aussi mis à jour une classe naturelle d'exemples satisfaisant nos conditions. D'autre part, dans une coopération avec le professeur Eduard Cerny (VLSI), nous avons pu établir une méthode de vérification formelle (qui s'est avérée être de type "model checking") où l'on mettait à profit les "lookaheads" pour rendre déterministe un système de transitions et réduire le problème à celui de l'inclusion de langages.

2.2 Guy LAPALME

Génération de texte

La génération de texte est un problème complexe et important dans un contexte où les informations traitées par l'ordinateur et devant être communiquées à l'utilisateur se complexifient. Dans ce cadre, nous avons créé, avec Richard Kittredge professeur au département de linguistique, le groupe, nommé SCRIPTUM, qui réunit présentement 3 étudiant(e)s au Ph.D. (Massimo Fasciano, Leila Kosseim et Daniel Rousseau) dont les projets sont décrits dans la section 3. Au cours des dernières années, trois thèses de Ph.D. (Agnès Tutin, Michel Gagnon et Nicole Tourigny) et un mémoire de M.Sc (Benoit Lavoie) ont été complétés dans ce groupe.

Le projet SCRIPTUM s'attaque à des domaines où l'information est stéréotypée ou bien structurée. Tous ces projets comportent une étude linguistique du domaine, la mise au point d'une stratégie d'organisation de texte et l'implantation en Prolog d'un prototype opérationnel de générateur à partir des données "brutes". Cette partie implantation est une caractéristique importante de nos projets qui nous différencie de plusieurs autres où, trop souvent, on se contente d'une étude de faisabilité ou d'une implantation partielle de certains modules.

Programmation par objets et fonctionnelle

Nous travaillons également sur des langages fonctionnels qui permettent la manipulation de fonctions comme des objets à part entière. Cette approche se caractérise par une absence d'effets de bord; en particulier, on n'y trouve ni affectation, ni boucle. Ces choix en font des candidats intéressants pour l'implantation du parallélisme. Nous avons appliqué ce genre de

programmation dans des domaines aussi différents que la résolution de problèmes avec des contraintes ou l'analyse sémantique de langues naturelles.

Détermination de structures de macromolécules

Depuis plusieurs années, nous collaborons avec une équipe au département de biochimie dirigée par Robert Cedergren afin de développer des méthodes informatiques de détermination de macromolécules. A présent, cet axe de recherche a été repris par François Major, un ex-membre du laboratoire Incognito et maintenant chercheur adjoint au DIRO dans le cadre du laboratoire de Biologie informatique et théorique. Marcel Turcotte, étudiant au Ph.D. collabore à ce projet.

2.3 Jacques MALENFANT

L'équipe Réflex s'intéresse à l'étude, la conception et à l'implantation de langages de programmation réflexifs. L'équipe compte actuellement un professeur et trois étudiants gradués. La réflexion est la capacité donnée aux programmes : (1) d'observer et de raisonner sur leur propre état et (2) de modifier leur état d'exécution ou d'altérer leur propre interprétation. Elle requiert un mécanisme de réification pour encoder comme données l'état du programme. La réflexion donne la possibilité de modifier dynamiquement la sémantique du langage en y introduisant de nouvelles possibilités : espionnage pour le déverminage (debugging), nouvelles structures de contrôle, optimisations dynamiques, etc. On distingue deux niveaux de réflexion : la *réflexion de structure* qui vise à réifier le programme et les structures de données du langage, et la *réflexion de comportement* qui s'intéresse au côté dynamique en réifiant l'évaluateur du langage et les représentations internes qu'il utilise pour exécuter le programme.

La réflexion dans les langages de programmation devient actuellement un domaine très important car elle donne au programmeur un modèle flexible lui permettant d'adapter le langage lui-même aux besoins sans cesse changeants de ses applications. La réflexion en programmation (1) donne un fondement global à des idées développées indépendamment dans plusieurs approches (programmation fonctionnelle, logique, par objets, etc.), (2) appréhende ces phénomènes avec un bagage commun et facilite le transfert d'expertise d'un domaine à l'autre et (3) solutionne des problèmes importants autant du point de vue des langages que du point de vue des applications.

L'équipe Réflex poursuit actuellement trois projets principaux :

1. **Réflexion de comportement dans les langages à prototypes.** L'équipe poursuit actuellement un effort de recherche débuté il y a quatre ans et qui vise à mieux comprendre la réflexion de comportement en programmation par objets, à définir un modèle formel de réflexion de comportement dans les langages à prototypes, et à réaliser une implantation efficace de ce type de réflexion. Les deux premiers volets sont actuellement terminés mais le troisième volet, que nous entreprenons maintenant par l'écriture d'un compilateur pour notre langage à prototypes réflexif, est certainement le plus ambitieux.

Ce projet est une collaboration avec les équipes de M. Pierre Cointe (Ecole des Mines de Nantes) et de M. Christophe Dony (LIRMM de l'Université de Montpellier).

2. **Implantation du langage Dylan avec extensions réflexives.** Le langage Dylan a été récemment proposé par Apple comme une alternative avancée aux langages à objets comme C++. Dylan est un langage dynamique dans la lignée de Lisp, Smalltalk et Scheme. L'équipe est actuellement impliquée dans un projet visant à une implantation efficace du langage Dylan dans laquelle elle s'intéresse à la partie objet du système ainsi qu'à des extensions réflexives.

Ce projet est une collaboration avec l'équipe de Marc Feeley du laboratoire de parallélisme de notre département ainsi que les équipes de Pierre Cointe (Ecole des Mines de Nantes), Christian Queinnec (INRIA Rocquencourt) et Charles Consel (IRISA de Rennes).

3. **Méthodologie pour la mise en œuvre de la réflexion de comportement.** Jusqu'à maintenant, la recherche en réflexion a principalement porté sur les modèles et leur implantation. Nous nous intéressons à l'étape subséquente qui est l'utilisation intensive et effective de ces modèles dans de grandes applications. En un mot, nous nous intéressons aux méthodologies de programmation réflexive et aux moyens de leur mise en œuvre. Un volet important de cette recherche concerne l'utilisation de la théorie des monades pour exprimer différentes sémantiques à partir d'un interprète générique. Nous étudions cette approche car le formalisme des monades nous permet d'espérer mettre en œuvre des analyses statiques et des transformations de programme rendant les programmes réflexifs efficaces. D'autres volets de cette recherche concerne la combinaison de méta-objets (objets du méta-niveau exprimant des comportements spécialisés par rapport à la sémantique de base), et à la mise au point d'outils de programmation pour les langages à objets réflexif.

2.4 Jian-Yun NIE

Ma recherche porte sur la recherche d'informations qui vise à extraire, parmi un ensemble volumineux de documents, ceux qui sont pertinents à une requête donnée.

Sur le plan théorique, je m'intéresse particulièrement à la modélisation des systèmes de recherche d'informations. L'objectif est de définir un cadre approprié qui permet de décrire tous les aspects impliqués dans une opération de recherche d'informations, tels que le contenu des documents, le besoin de l'utilisateur et le contexte de l'opération. Les logiques constituent les outils privilégiés dans cette tâche. Mais un des problèmes à résoudre est de traiter de l'incertitude dans un cadre logique, ce qui nécessite l'adjonction de la logique avec, par exemple, la théorie des probabilités. Actuellement, mon accent est mis sur l'utilisation d'un nouveau type de logique - la logique contrefactuelle, identifiée comme la plus adaptée à la recherche d'informations.

Du cote pratique, je me concentre sur l'implantation d'un système de recherche d'informations basée sur la sémantique. Cette méthode de recherche d'informations repose sur la

”compréhension” plus poussée des documents et des requêtes (des textes), d’une part, et d’autre part, sur l’utilisation des liens sémantiques entre les concepts (ou plus généralement, des connaissances). Les méthodes de l’IA sont utilisées dans cette partie de ma recherche.

2.5 Jean VAUCHER

Note biographique

Jean Vaucher détient un baccalauréat en génie électrique de l’Université d’Ottawa. Récipiendaire de la prestigieuse bourse Athlone, il étudia à l’Université de Manchester en Angleterre et obtint un doctorat en 1968. De retour à Montréal, Jean Vaucher a travaillé chez IBM pendant deux ans avant d’entrer à l’Université de Montréal en 1970. Il fut directeur de son Département de 1980 à 1983.

Le professeur Vaucher travaille principalement en intelligence artificielle, génie logiciel et en simulation et il compte plusieurs douzaines d’articles scientifiques et deux livres à son actif. Il a siégé sur plusieurs comités scientifiques importants y compris le comité informatique du CRSNG—dont il a été le président en 1979 et 1980—, le FCAC et le Comité associé du CRSNG concernant l’intelligence artificielle. Il est aussi un des cinq fondateurs du CRIM et il entretient des liens avec l’industrie par le biais de consultation et de projets de recherche conjointes.

Intérêts de recherche

Ma recherche porte sur l’intégration des techniques d’Intelligence Artificielle avec celles du Génie Logiciel, particulièrement au niveau de la structuration des connaissances imbriquées, de la déduction automatique et des interfaces personne-machine. Du côté pratique, je suis intéressé par les systèmes de conception (CASE) et par les systèmes experts. Du côté méthodologique, nous expérimentons avec la programmation par objets (depuis 20 ans) et la programmation logique (depuis 10 ans).

Ma recherche a des implications industrielles évidentes et je suis souvent sollicité par le Centre de Recherche Informatique de Montréal (CRIM) afin de participer à des projets conjoints avec l’entreprise privée. Ce genre de projet donne aux étudiants qui y participent une expérience très enrichissante. Trois projets récents ont ainsi permis de doter de bourses des étudiants au PhD informatique (projets Macroscopie, IGLOO et PurchaseMaster).

3 Etudiants et étudiantes au Ph.D.

3.1 Nicolas ANQUETIL (Directeur : Jean Vaucher)

Dans le cadre de l'ingénierie de connaissance, on cherche à définir des systèmes qui puissent acquérir automatiquement des connaissances d'un haut niveau d'abstraction à partir de données brutes ou aussi peu "préparées" que possible. Le but étant de se passer autant que faire ce peut des compétences d'un expert du domaine considéré. Dans la même optique, un objectif plus modeste (mais plus réaliste) est d'aider l'expert ou une personne qui cherche à le devenir à organiser une grande quantité de connaissances selon un schéma qui lui permettra de valider facilement ses connaissances.

Une méthode classique d'organisation et d'abstraction est la construction automatique de hiérarchies de classes permettant de classer toutes les données que l'on possède. Ceci peut-être réalisé grâce à des algorithmes de clustering hiérarchique (Conceptual Clustering).

Mais il existe un autre modèle qui a été jusqu'à présent peu ou pas exploré. Dans le domaine de la classification bibliographique, Ranganathan a introduit en 1967 le schéma de classification à *facettes* qui consiste en une collection de "petites" hiérarchies de classification orthogonales les unes aux autres. L'un de ses avantages, par rapport à une classification hiérarchique classique, est qu'il offre une meilleure compréhension de l'ensemble de données considéré en séparant mieux les diverses relations (par exemple l'héritage, mais il y en a d'autres) qui peuvent unir deux classes.

Notre but est de tenter de formaliser une méthode pour extraire automatiquement un schéma de classification à facettes d'un ensemble de données. Cet objectif est ambitieux et s'attaque à des problèmes complexes comme l'extraction de notions sémantiques à partir d'informations syntaxiques ou le traitement du bruit et des erreurs dans les données. Une restriction plus réaliste à court terme est d'extraire un schéma de classification à facettes à partir d'une hiérarchie déjà définie et donc exempte d'erreurs et possédant en elle même, par les choix qu'ont fait les concepteurs, quelques informations sémantiques qui peuvent être utiles.

3.2 Martin BRISEBOIS (Directeur : Jian-Yun Nie)

Je m'intéresse à la recherche d'informations (RI), plus particulièrement à l'évaluation des requêtes. Les modèles conventionnels de RI (modèle booléen, vectoriel, etc) sont limités par le fait que l'évaluation de la pertinence d'un document est basée sur une comparaison directe entre les termes décrivant le document et ceux de la requête. Comme un même concept peut s'exprimer à l'aide d'une grande variété de termes, un bon nombre de documents pertinents ne sont pas retenus. Pour pallier à ce problème, il est possible d'utiliser un thésaurus afin d'en extraire des termes reliés à ceux de la requête. Par le passé, les thésauri obtenus par des moyens statistiques ont été examinés en profondeur. L'apparition de thésauri de grande taille définis manuellement (WordNet, par exemple) permet actuellement d'explorer de nouvelles avenues.

Avec Jian-Yun Nie, j'ai déjà pu constater que l'utilisation de relations sémantiques pour

étendre les requêtes améliore les performances. Nous avons constaté qu'il est primordial de déterminer le facteur de confiance associé aux relations sémantiques. Nous avons déjà défini une méthodologie qui permet d'estimer le facteur de confiance associé à un type de relations sémantiques (par exemple, hyponymie, méronymie, etc.). À court terme, nous essaierons d'estimer le facteur de confiance associé à chacune des relations sémantiques prises individuellement.

À plus long terme, j'examinerai la sémantique de Montague qui semble capturer les notions à la base de la RI. Son principal attrait est d'établir une correspondance entre les règles de l'analyse syntaxique et celles de l'analyse sémantique. De plus, l'objectif principal de Richard Montague était de déterminer la véracité d'un énoncé pour un état du monde donné. La nature même de l'évaluation de la pertinence d'un document est très semblable: on cherche à déterminer si un document est pertinent pour un usager étant donné une description de ses intérêts, de son intention (simple aperçu de la littérature, recherche exhaustive, etc.), des documents déjà consultés, etc. Sembok et Rijsbergen ont déjà présenté SILOL, un premier système de recherche d'informations exploitant cette idée. SILOL modélise adéquatement l'analyse syntaxique des requêtes et des documents. Toutefois, l'analyse sémantique s'y avère naïve, de l'aveu même des auteurs. J'approfondirai l'aspect sémantique de l'utilisation des grammaires de Montague pour la RI. Il sera ensuite souhaitable de réaliser une implantation du modèle afin de le comparer aux modèles classiques.

3.3 Eliana COELHO (Directeur : Guy Lapalme et Vimla L. Patel)

Notre recherche porte sur la modélisation, l'acquisition et la réutilisation de connaissances. Plus particulièrement, nous nous intéressons à la définition d'un langage pour la spécification de modèles d'inférence. Notre idée est de définir un langage pour la construction de modèles d'inférence qui soit indépendant des connaissances du domaine tout en utilisant un niveau d'abstraction indépendant des formalismes de représentation de connaissances.

Dans la première étape de notre travail, nous étudions le domaine médical afin d'identifier des primitives pour la réalisation des tâches de diagnostic et de planification de thérapie. Pour cette étape, nous travaillons en collaboration avec le Centre de Science Cognitive dans le domaine médical de l'Université McGill (McGill Cognitive Science Centre).

La deuxième étape de notre travail est la définition d'un langage pour la construction de modèles d'inférence. Le but de ce langage formel est d'être un soutien à l'ingénierie de connaissances, en aidant à la construction de modèles d'inférence. À partir de ce langage, l'ingénieur de connaissances aura un bon point de départ pour l'activité de modélisation.

Dans la troisième étape de notre recherche, nous développerons un prototype d'outil d'acquisition de connaissances spécifiques pour la tâche de diagnostic. Cet outil sera développé en utilisant comme spécification un modèle d'inférence décrit selon le langage défini dans deuxième étape. Le développement d'un outil d'acquisition de connaissances permettra l'automatisation de la construction et du raffinement de bases de connaissances de systèmes experts. L'utilisateur de cet outil sera un expert du domaine à qui seront posées des questions en termes du domaine, dans ce cas le diagnostic médical. Au moyen de ce prototype, nous validerons notre travail et nous évaluerons la possibilité de développer d'autres outils à partir

des modèles d'inférence créés en utilisant notre approche.

3.4 François-Nicola DEMERS (Directeur : Jacques Malenfant)

Je m'intéresse à la description formelle de la réflexion de comportement dans les langages de programmation fonctionnelle et par objets en utilisant la théorie des monades.

Les monades ont été proposées pour exprimer la sémantique de différentes constructions d'interprètes en programmation fonctionnelle (par exemple les états globaux, les exceptions, les continuations, les entrées-sorties interactives et le non-déterminisme). Le concept de monades a ses origines de la théorie des catégories. Cette théorie a permis d'unifier et de simplifier plusieurs propriétés de systèmes mathématiques très importants. Des résultats récents montrent que les monades permettent aussi de représenter formellement et de façon générale toute une famille d'interprètes de langages. Et de plus, il serait possible de les composer pour construire de nouveaux interprètes combinant les propriétés des monades de départ. Les monades peuvent aussi servir d'approche modulaire à la sémantique dénotationnelle.

Je m'intéresse à cette approche pour définir les comportements dans un langage fonctionnel et par objets réflexifs. En réflexion de comportement, nous cherchons à représenter dans le langage l'évaluateur du langage lui-même, de façon à introduire facilement des variations sur la sémantique du langage. La façon la plus simple d'exprimer de telles variations est d'écrire des interprètes les implémentant. En passant aux monades, nous pensons obtenir une définition formelle et générale des comportements tout en respectant l'aspect modulaire du langage à objets. Par leur composition, une méthodologie de conception de nouveaux comportements pourraient être couplées avec la réutilisabilité du langage. De plus, il est à espérer que l'utilisation de transformations de programme adéquates rendra ce type d'implémentation de langages réflexifs suffisamment performante. Le but de mon projet est d'explorer ces avenues pour définir et implémenter une telle méthodologie.

3.5 Massimo FASCIANO (Directeur : Guy Lapalme)

Le but du projet est la conception d'un outil permettant la génération automatique de rapports statistiques à partir d'informations contenues dans une base de données. Les rapports en question contiendront des graphiques illustrant l'information contenue dans les données ainsi que des textes descriptifs qui dirigeront l'attention du lecteur sur les éléments importants des graphiques et qui permettront une analyse plus poussée des données. Diverses méthodes d'analyse statistique seront utilisées pour extraire les informations pertinentes des données. Le projet se divise en trois parties :

Le choix d'un graphique

Des recherches sur l'efficacité des différents types de graphiques montrent que les trois facteurs principaux ayant une influence sur le choix d'un graphique sont : tout d'abord, le type des variables que l'on veut décrire (certains graphiques sont très adaptés pour des variables numériques continues alors que d'autres décrivent mieux les variables discrètes), ensuite le message

à transmettre (veut-on montrer une répartition, une évolution, une distribution, etc...), et finalement les problèmes de perception (souvent nos yeux nous jouent des tours... il faut donc éviter de produire des graphiques qui peuvent porter à confusion).

Génération d'un texte descriptif

Un texte descriptif a surtout un rôle d'analyse. En effet, alors qu'un graphique est très utile pour donner une vue d'ensemble des données, un texte permet une analyse plus fine. En d'autres mots, le graphique montre l'information et le texte l'explique. Le rôle du texte est de compléter le graphique qu'il accompagne. Ainsi, si un graphique n'est pas parfait, le texte va aider le lecteur à comprendre celui-ci et à en dégager les points importants.

Intégration des deux formes d'expression

Cette partie du projet vise à coordonner deux formes d'expression très différentes l'une de l'autre : le graphique et le texte. Certains messages sont mieux transmis à travers un graphique alors que d'autres ressortent mieux dans un texte. Il faut donc être capables d'identifier ceux-ci et de les exprimer de la façon la plus efficace. On peut imaginer une interface entre les deux générateurs qui serait capable d'identifier l'efficacité d'un graphique et d'ajuster la quantité d'information contenue dans le texte en conséquence (si le graphique dit tout, inutile de le répéter dans le texte. par contre, si on n'a pas pu trouver de graphique parfaitement adapté aux données, on peut compenser par un texte plus détaillé).

3.6 Leila KOSSEIM (Directeurs : Guy Lapalme et Richard Kittredge)

Mes travaux de doctorat portent sur la planification textuelle de textes d'instructions dans cadre de la génération automatique de texte.

La génération automatique d'instructions est un axe de recherche assez prometteur. En effet, à partir d'une représentation conceptuelle de la tâche à accomplir, on peut automatiser la rédaction initiale ainsi que la mise à jour de textes d'instructions ou tout simplement assister les rédacteurs techniques humains. Mon projet de doctorat s'est penché sur les aspects sémantiques et rhétoriques de la planification de textes d'instructions français.

Lors de la rédaction d'un texte, deux questions sont primordiales: comment sélectionner le contenu informationnel et comment le présenter le façon cohérente. Ces deux étapes n'étant pas reliées de façon directe, pour les déterminer de façon automatique, nous avons d'abord effectué l'analyse d'un corpus *naturel*. Cette étape nous a permis d'identifier un ensemble d'*éléments sémantiques* qui définissent le contenu informationnel du texte et un ensemble de relations rhétoriques (basées sur la RST) utilisées pour présenter ces éléments de sens. Finalement, l'analyse de corpus a permis de développer des heuristiques pour sélectionner les éléments sémantiques à partir de la représentation procédurale de la tâche; et des heuristiques pour les présenter par les relations rhétoriques les plus appropriées.

Pour évaluer les heuristiques de sélection et de présentation, nous avons implanté le prototype informatique SPIN. Les résultats de ce système démontrent la validité des heuristiques,

et leurs avantages par rapport aux méthodes antérieures.

Ce projet de doctorat permettra de mieux comprendre les processus de rédaction humaine, et de s'approcher de plus en plus d'une rédaction complètement automatique de manuels d'instructions.

3.7 Marco JACQUES (Directeur : Jacques Malenfant)

Mon projet consiste en l'implantation d'un compilateur pour un langage à objets avec réflexion de comportement.

La réflexion est un paradigme de programmation permettant à un programme d'examiner et de modifier ses propres structures au cours de son exécution. Nous nous intéressons tout particulièrement à la réflexion de comportement, qui permet la modification de la sémantique du langage. Les langages à objets sont de bon candidats pour l'implantation de la réflexion. En effet, les principes de modularité des objets assurent la localisation des effets de la réflexion.

Cependant, l'intégration de la réflexion de comportement dans un langage à classes soulève quelques problèmes conceptuels. Également, à cause de sa puissance intrinsèque, l'implantation de ce type de réflexion se fait présentement avec un grand sacrifice au niveau de l'efficacité.

Le but de notre projet est de réaliser une implantation efficace d'un langage réflexif. Malenfant, Cointe et Dony ont proposé un modèle conceptuel de réflexion de comportement pour un langage à prototypes (i.e. un langage à objets sans classes). Dans le cadre de mon projet de maîtrise, j'ai réalisé un interprète pour ce langage (Proto-Reflex).

Mon projet de doctorat consiste en l'implantation d'un compilateur pour ce langage. À l'aide de différents langages et techniques de compilation, nous verrons comment effectuer l'intégration de nouvelles définitions réflexives au compilateur.

3.8 Gilles PESANT (Directeur : Michel Boyer)

Une contrainte représente quelque chose de fondamental: elle est l'abstraction d'un lien entre des entités. Qu'elle porte le nom de relation ou de prédicat, on la retrouve un peu partout en Intelligence Artificielle et généralement au cœur de la spécification d'un problème à résoudre. L'intégration des contraintes parmi les primitives d'un langage de programmation a récemment reçu une attention particulière. Plusieurs paradigmes ont été proposés comme candidats.

La programmation en logique est un paradigme de programmation de très haut niveau puisque de nature déclarative. Une contrainte, puisqu'elle spécifie une relation entre des variables, est aussi essentiellement déclarative. La programmation en logique avec contraintes (CLP) accroît la puissance descriptive de la programmation logique en élevant les contraintes au niveau de primitives du langage et en généralisant son mécanisme d'exécution, l'unification, à la résolution de contraintes. L'implantation de solveurs pour des domaines particuliers de contraintes peut donner naissance à autant de langages issus de ce paradigme.

La majorité des langages de programmation en logique avec contraintes incorporent déjà un traitement des contraintes arithmétiques linéaires en empruntant principalement des tech-

niques d’algèbre et de programmation linéaires. Bien plus rares sont ceux qui vont permettre une réelle manipulation des contraintes non linéaires. Ces dernières posent des problèmes difficiles dont les solutions connues dans leur pleine généralité ne sont pas applicables parce que trop coûteuses dans ce contexte. Mon apport s’est situé au niveau du compromis: se restreindre principalement aux contraintes quadratiques, un sous-problème riche offrant un traitement efficace par le biais d’une classification géométrique et d’approximations linéaires. Son aboutissement a pris la forme du langage expérimental QUAD-CLP(\Re).

Mes travaux ne sont pas limités au paradigme de la programmation logique. Ils peuvent également s’appliquer au contexte plus général des systèmes de raisonnement sur contraintes ou des bases de données déductives.

3.9 Daniel ROUSSEAU (Directeurs : Guy Lapalme et Bernard Moulin)

Les conversations sont des entités linguistiques fascinantes qui se déploient habituellement de manière cohérente grâce à la contribution des agents participants, même si ceux-ci ont des connaissances et des buts différents. Nous proposons une approche pour modéliser et simuler des conversations entre des planificateurs intelligents capables d’interpréter ce qui leur est communiqué, de raisonner sur leurs propres états mentaux et ceux d’autres agents et de réagir de manière appropriée en s’adaptant au contexte. Nous supposons que les agents locuteurs s’échangent des objets conversationnels reflétant des états mentaux, tels les croyances et les buts, par le biais d’actes de communication composés d’énonciations et de gestes, tout en interagissant à d’autres niveaux: gestion du sujet de conversation, maintenance de la communication et gestion de l’initiative (tour de parole).

Nous énonçons certaines catégories d’états mentaux et de relations conceptuelles entre ces états que l’on peut retrouver dans le modèle mental d’un agent locuteur et dans le réseau d’objets conversationnels reflétant les états mentaux échangés au cours d’une conversation. Nous ajoutons au réseau d’objets conversationnels deux autres structures de connaissances, la structure de conversation et l’agenda d’initiative, pour permettre au modèle de la conversation de représenter divers niveaux d’interaction. Nous suggérons certains protocoles sophistiqués de communication pour des agents locuteurs artificiels qui permettent d’exprimer les actes de communication à un niveau conceptuel en prenant en compte les divers niveaux d’interaction présents dans une conversation et les diverses transitions permises. Nous proposons une approche utilisant un moteur d’inférence et des règles de production pour interpréter des actes de communication en fonction du modèle explicite de la conversation et des connaissances du locuteur. Cette approche est appliquée dans un agent spécial, l’agent de conversation, responsable de la gestion du modèle de la conversation. Finalement, nous élaborons une architecture pour un agent locuteur capable de percevoir et d’exécuter des actes de communication et des actions non-linguistiques, de raisonner sur le contenu de son modèle mental et sur celui du modèle d’une conversation et de planifier son comportement.

Un prototype simulant des conversations entre des agents locuteurs est en cours d’implantation. L’approche proposée pourrait être exploitée dans des modules de communication personne-machine ou machine-machine.

3.10 Gilles ROY (Directeur : Jean Vaucher)

Les systèmes de programmation par contraintes (SPC) doivent faire appel à une panoplie de solveurs de façon à supporter un large spectre de types de contraintes et couvrir ainsi des domaines d'application variés.

Le modèle général établi qui consiste à passer la main à un deuxième solveur si un premier échoue tend à disparaître au profit de deux tendances :

- exploitation dans le SPC d'un seul solveur très efficace mais à portée limitée ;
- adoption d'un paradigme unificateur extensible permettant d'une part le choix du solveur directement selon le type de contrainte et d'autre part l'addition modulaire de nouveaux solveurs au SPC. La réécriture de termes augmentée (RTA) est un paradigme relevant de cette approche.

Nous nous inscrivons dans cette deuxième tendance en proposant un nouveau modèle basé sur la satisfaction incrémentale de contraintes par la réécriture de termes, de façon à pallier aux faiblesses du paradigme : son inefficacité inhérente et une approche non-interactive à la satisfaction de contraintes.

Notre modèle, la réécriture de termes augmentée incrémentale (D-RTA), tire profit des développements les plus récents en matière d'implantation des langages de programmation fonctionnelle, notamment ceux s'appuyant sur la logique équationnelle.

3.11 Akila SARIRETE (Directeur : Jean Vaucher)

On assiste, ces dernières années, à une croissance considérable dans le développement des systèmes à base de connaissances (SABC). Depuis les premiers systèmes se basant sur de simples architectures (base de connaissances, moteur d'inférence, interface usager), on est arrivé à définir des architectures plus complexes intégrant même différentes techniques de l'intelligence artificielle.

Un problème qui peut être soulevé avec l'explosion des réalisations, dans ce domaine, est comment en bénéficier. Ce que nous proposons c'est de réutiliser de telles réalisations et de les intégrer dans un même système. Ceci sera bénéfique puisque la réalisation de systèmes intelligents est une tâche complexe et nécessite beaucoup de travail. Dans cette optique, nous proposons une méthodologie pour la conception et l'implantation d'un SABC qui intègre plusieurs sources de connaissances et rend ces connaissances accessibles aux usagers.

Le système se composera alors de plusieurs modules supportant chacun une certaine expertise. L'exploitation de cette expertise se fera via une interface générale. Ceci implique que pour arriver à une bonne communication entre les sources de connaissances, il faut établir une interface complexe et puissante qui permet de les intégrer et d'avoir une sorte d'"interlingua".

La validation de notre système se fera dans le domaine de l'aide à la décision et la considération d'un système d'achat.

Nous dirigeons notre recherche sur deux aspects:

1. l'aspect interface et coopération intelligente entre les différentes sources de connaissances;
2. l'aspect méthodologique dans lequel nous établirons une méthodologie d'acquisition de connaissances où l'accent sera mis sur l'aspect utilisateur.

Dans le premier aspect, nous considérons que les différentes sources de connaissances sont indépendantes et donc, nous définirons des primitives sémantiques qui réaliseront la coopération entre les différentes sources. Quant au second aspect, nous considérons l'utilisateur du système comme une source de connaissances dont nous modéliserons les connaissances et établirons une stratégie d'interaction.

3.12 Stéphane Sotèg SOMÉ (Directeurs : Rachida Dssouli et Jean Vaucher)

Mon travail est un volet du projet IGLOO sur les méthodologies de spécification et de conception par objets. Je travaille avec Rachida Dssouli du groupe de recherche sur les protocoles de communication et Jean Vaucher du laboratoire Incognito.

Le développement de systèmes informatiques s'effectue selon un processus débutant par la définition des exigences des usagers et se terminant à la fin de la vie du système.

Il existe plusieurs paradigmes de développement de systèmes informatiques dont certains bénéficient d'un support automatique. Toutefois, l'étape de la définition des besoins des usagers est difficilement automatisable en raison du caractère informel des exigences qui sont généralement exprimés dans un langage naturel. Cette non-automatisation de la définition des exigences constitue une des grandes causes d'erreurs de conception et des coûts qu'ils engendrent.

Les exigences des usagers concernant certains types de systèmes informatiques sont énoncées sous forme de scénarios décrivant des séquences d'interactions entre le système et son environnement. Notre objectif est de développer une approche efficace et automatique pour l'acquisition des besoins des usagers, en utilisant des scénarios définis par eux, et de dériver des spécifications complètes et valides.

Nous accordons une attention particulière aux faits suivants :

- Les besoins des usagers doivent être exprimés dans une notation compréhensible à la fois par les utilisateurs et les experts du domaine. Cette notation doit de plus être supportée par un formalisme pour permettre la vérification systématique des spécifications. Elle doit également permettre l'inclusion de contraintes temporelles, utiles pour la description du comportement de systèmes temps réels.
- Généralement, les scénarios ne définissent pas le comportement d'un système conformément à tous les cas possibles. Cela suppose le développement de méthodes de complétions de spécifications.
- Le comportement défini par les utilisateurs doit pouvoir être intégré à la spécification et rester compatible avec elle.

3.13 Marcel TURCOTTE (Directeurs : Guy Lapalme et Robert J. Cedergren)

Mon projet a trait à l'étude du repliement des macromolécules. La taille et la complexité de ces molécules font en sorte que la détermination de leur structure tridimensionnelle par des moyens physiques est un processus long et onéreux. Cette connaissance est néanmoins nécessaire. En effet, la détermination d'une nouvelle structure fait progresser les études fondamentales de la régulation, de l'évolution moléculaire et plus généralement de la compréhension des mécanismes cellulaires.

J'ai participé au développement d'un système, MC-SYM, permettant la génération d'un ensemble de conformations satisfaisant à un ensemble de contraintes. Afin d'automatiser la détermination des contraintes, j'étudie différentes mesures d'associations qui sont appliquées aux positions d'un alignement de séquences de même fonction. Les résultats pour les molécules d'ARN sont excellents. Une autre partie du projet consiste à étudier différentes alternatives à l'algorithme de recherche de MC-SYM. Le but global est de produire un système utilisant en entrée un ensemble des séquences homologues et produisant en ensembles de conformations possibles pour la molécule à l'étude.

4 Étudiante au M.Sc.

4.1 Hong ZHAO (Directeur : Jian-Yun Nie)

Je m'intéresse à la recherche d'informations sur les documents structurés. Le but de ce travail est de rendre le système capable de retrouver les portions de documents à travers la structure décrite en SGML. Pour ce faire, il faut d'abord définir une méthode d'indexation en considérant la structure logique des documents. À la fin de cette indexation, chaque élément logique d'un document sera associé à un ensemble de mots clés. L'opération de recherche se base sur cette indexation structurée. Le langage d'interrogation doit aussi être enrichi afin de permettre aux utilisateurs d'indiquer le niveau des réponses qu'ils souhaitent obtenir (les documents complets, les chapitres, les sections, etc.).

5 Diplômé(e)s

5.1 Diplômé(e)s au Ph.D.

Directeurs : Guy Lapalme

- Michel Gagnon “Expression de la localisation temporelle dans un générateur de texte”, (co-direction avec Richard Kittredge), 1993; maintenant en stage post-doctoral à l’Institut de Recherche en Informatique de Toulouse (IRIT).
- Sylvain Giroux “Agents et systèmes, une nécessaire unité”, (co-direction avec Alain Senteni), 1993; maintenant chercheur associé à la Télé-Université.
- Nicole Tourigny “La génération automatique de descriptions de systèmes dynamiques”, 1994; maintenant professeure au département d’informatique de l’université Laval.

5.2 Diplômé(e)s au M.Sc.

Directeur : Michel Boyer

- Dominique Boucher “La théorie des catégories en informatique: notions de base et application”, 1993; maintenant étudiant au PhD informatique au DIRO.

Directeur : Guy Lapalme

- Massimo Fasciano “Génération intégrée de texte et de graphiques”, Maîtrise accordée par équivalence pour études accomplies au niveau du doctorat, 1993; maintenant étudiant au PhD, DIRO.
- Benoit Lavoie “Étude sur la planification de texte: application au sous-domaine de la population active”, (co-direction avec R. Kittredge), 1994; maintenant analyste à CoGenTex, Ithaca, USA.
- Annabel Sadan “Classification de conformations macromoléculaires”, (co-direction avec Robert Cedergren), 1993; maintenant analyste chez Vidéotron.
- Marcel Turcotte “L’étude de la dynamique de conformation des protéines et des acides nucléiques” (co-direction avec Robert Cedergren), Maîtrise accordée par équivalence pour études accomplies au niveau du doctorat, 1993; maintenant étudiant au PhD, DIRO.

Directeur : Jacques Malenfant

- Francois-Nicola Demers “Réflexion de comportement et évaluation partielle en prolog”, 1994; maintenant étudiant au PhD informatique au DIRO.

- Marco Jacques “Implantation d’un langage à prototypes avec réflexion de comportement”, 1994; maintenant étudiant au PhD informatique au DIRO.
- François Rousseau “Visualisation d’objets en milieu réflexifs: le cas du browser à la smalltalk-80”, 1994; maintenant employé chez Vidéotron.

Directeur : Jian-Yun Nie

- Martin Brisebois “Utilisation des relations sémantiques dans la recherche d’informations”, Maîtrise accordée par équivalence pour études accomplies au niveau du doctorat, 1994; maintenant étudiant au PhD, DIRO.
- Sergio Carmona “Indexation syntaxique dans la recherche d’informations”, 1994; maintenant employé dans un ministère du gouvernement Mexicain.
- François Paradis “Indexation sémantique de documents: application à un corpus technique”, 1993; maintenant en doctorat à l’université Joseph Fourier (Grenoble).
- Éric Prévot “Une nouvelle approche connexionniste pour la recherche de documents”, (co-direction avec Jan Gecsei), 1993.

Directeur : Jacques Savoy

- Martin Choquette, “L’application des systèmes hypertextes dans un cadre juridique”, (co-direction avec Paul Bratley), 1993; maintenant analyste informatique au Centre de Recherche en Droit Public de la Faculté de Droit de l’Université de Montréal.
- Atoulé Agbeko Kokoroko “Mise à jour automatique des textes de loi”, (co-direction avec Paul Bratley), 1993; maintenant analyste informatique.
- Anani Kuassivi “Système d’aide multicritère à la décision et hypertexte: une alliance intéressante”, 1994; maintenant analyste informatique.

6 Publications

Revues avec comité de lecture

- M. Feeley, M. Turcotte et G. Lapalme, Using Multilisp for Solving Constraint Satisfaction Problems: an Application to Nucleic Acid 3D Structure Determination, *Journal of Lisp and Symbolic Computation*, 7(2-3):229-244, 1994.
- S. Giroux, A. Senteni et G. Lapalme, Les problèmes d'adaptation dans les systèmes ouverts. *Analyse de Systèmes*, 19(1-2):121-144, juin 1993.
- G. Lapalme et F. Lavier, Using a functional language for parsing and semantic processing. *Computational Intelligence*, 9(2):111-131, 1993.
- J. Malenfant, C. Dony, et P. Cointe, A Semantics of Introspection in a Reflective Prototype-Based Language, à paraître dans *Lisp and Symbolic and Computation*, juin 1994.
- G. Pesant. et M. Boyer, Approximations of Quadratic constraints, *soumis à SIAM Computing*, 30 pages.
- J.-Y. Potvin, G. Lapalme et J.-M. Rousseau, A microcomputer assistant for the development of vehicle routing and scheduling heuristics. *Decision Support Systems*, 12(1):41-56, 1994.

Conférences avec comité de lecture

- N. Anquetil et J. Vaucher, Extracting Hierarchical Graphs of Concepts from an Object Set : Comparison of two Methods, *actes du "Workshop on Knowledge Acquisition using Conceptual Graph Theory", ICCS'94*, pages 26-45, College Park, Maryland, aout 1994.
- M. Fasciano et G. Lapalme, A Prolog implementation of the functional unification grammar formalism. Dans *Proceedings of the Fourth International Workshop on Natural Language Understanding and Logic Programming (NLULP4)*, pages 37-45, Nara, Japan, 29 sept. - 1 oct. 1993. version révisée soumise à *New Generation Computing*.
- S. Giroux, A. Senteni et G. Lapalme, Reactalk, du monde réel à des systèmes informatiques ouverts et adaptables. Dans *RFIA-94*, Paris, France, janvier 1994.
- L. Kosseim et G. Lapalme, Content and rhetorical status selection in instructional texts. Dans *7th International Generation Workshop*, pages 53-60, Kennebunkport, Maine, juin 1994.
- G. Lapalme, D. Rousseau, B. Moulin, Une approche multi-agents pour modéliser les conversations. Dans *Deuxièmes Journées Francophones d'intelligence Artificielle Distribuée et Systèmes Multi-Agents*, pages 3-14, Voiron, France, mai 1994.

- G. Lapalme, B. Moulin, D. Rousseau, A multi-agent approach for modelling conversations. Dans *Fourteenth International Avignon Conference AI 94, Natural Language Processing sub-conference*, pages 35–50, Paris, juin 1994.
- F. Major, M. Turcotte et G. Lapalme, Constraint satisfaction in functional programming. Dans *First Workshop on Principles and Practice of Constraint Programming*, pages 184–188, Rhode-Island, pages 28–30 avril 1993.
- J. Malenfant et F.-N. Demers, Évaluation partielle pour l’implantation efficace de la réflexion de comportement, *Actes des Journées Francophones sur la Programmation Logique, JFPL’93*, mai 1993.
- J.Y. Nie, M. Brisebois, Using a General Thesaurus to Set Fuzzy Relevance of Terms in Information Retrieval, *Expersys’94*, Houston, Texas, pages 493–498, décembre 1994.
- J.Y. Nie, J. Vaucher, N. Anquetil, A flexible architecture for collaborative information retrieval, *RIAO’94*, New-York, pages 600–611, octobre 1994.
- J.Y. Nie, W. Jin, M.-L. Hannan, A hybrid approach to unknown word detection and segmentation of Chinese, *International Conference on Chinese Computing*, Singapore, pages 326–335, juin 1994.
- J.Y. Nie, F. Paradis,, J. Vaucher, Adjusting the performance of an information retrieval system, *ACM 2nd Conference on Information and Knowledge Management*, Washington (DC), pages 726–728, novembre 1993.
- J.Y. Nie, N. Anquetil, J. Vaucher, A multi-agent distributed retrieval system, *Annual conference of the Canadian Association for Information Systems*, pages 150–162, juillet 1993.
- J.Y. Nie, F. Paradis, J. Vaucher, Using information retrieval for software reuse, *International conference on computing and information*, Sudbury, pages 448–452, juin 1993.
- J.Y. Nie, F. Paradis, J. Vaucher, Un système de recherche d’information pour la réutilisation, *ICO’93*, Montréal, pages 163–172, mai 1993.
- G. Pesant et M. Boyer, QUAD-CLP(R): Adding the Power of Quadratic Constraints, *Proceedings of the Second Workshop on Principles and Practice of Constraint Programming (PPCP’94)*, Springer Verlag, Lectures Notes in Computer Science (874) .
- G. Pesant and M. Boyer, A Geometric Approach to Quadratic Constraints in Logic Programming, *proceedings of the Third Workshop on Constraint Logic Programming (WCLP’93)*, Marseilles, mars 1993.
- D. Rousseau, B. Moulin et G. Lapalme, Modélisation des conversations basée sur une perspective de systèmes multiagents. Dans *Colloque Informatique et langue naturelle*, pages 129–150. Institut de Recherche en Informatique de Nantes, Université de Nantes – Ecole Centrale, décembre 1993.

F. Rousseau and J. Malenfant, Browsing in Explicit Metaclass Languages : an Essay in Reflective Programming Environments, *Actes du "Third Workshop on Reflection and Metalevel Architectures in Object-Oriented Programming"*, OOPSLA '93, Washington (DC), octobre 1993.

Revue sans comité de lecture

G. Lapalme, R. Kittredge, M. Fasciano, L. Kosseim, M. Gagnon, B. Lavoie, D. Rousseau, N. Tourigny et A. Tutin, Natural language generation at the Université de Montréal: Research summary of the SCRIPTUM group. *Canadian Artificial Intelligence*, pages 13–19, Spring/Summer 1994.

Conférences sans comité de lecture

M. Fasciano et G. Lapalme, Génération intégrée de textes et de graphiques statistiques. Dans *Traitement automatique du français écrit: développements théoriques et applications*, 62e congrès de l'ACFAS, mai 1994.

D. Rousseau, B. Moulin et G. Lapalme, Believable agents in conversations. Dans *AAAI Spring Symposium on Believable Agents*, Stanford University, California, mars 1994.

D. Rousseau, B. Moulin et G. Lapalme, Representation and simulation of conversations. Dans *AAAI Spring Symposium on Natural Language Understanding in Integrated Systems*, Stanford University, California, mars 1994.

D. Rousseau, G. Lapalme et B. Moulin, A model of speech act planner adapted to multiagent universes. Dans O. Rambow, editor, *Proceedings of the Workshop on Intentionality and Structure in Discourse Relations*, pages 110–113. Special Interest Group on Generation, Association for Computational Linguistics, 21 juin 1993.

Publications du Département d'informatique et de recherche opérationnelle, Université de Montréal

886 J. Malenfant (1993) Équipe Réflex, rapport d'activités 1992/93

7 Projets spéciaux

Pour donner un cadre à une collaboration suivie entre eux et des chercheurs d'autres facultés ou universités, deux professeurs du laboratoire Incognito (G.Lapalme et J.Malenfant) ont senti la nécessité de former des sous-groupes de recherche à l'intérieur du laboratoire. Il s'agit des groupes SCRIPTUM et Réflex.

7.1 Réflex

Pour obtenir plus de renseignements sur le groupe Réflex, veuillez contacter le professeur Jacques Malenfant (malenfan@IRO.UMontreal.CA, v. adresses électroniques §9).

Présentation générale de l'équipe

L'équipe Réflex du laboratoire INCOGNITO s'intéresse à la conception et à l'implantation de langages de programmation réflexifs. L'équipe compte actuellement un professeur et trois étudiants gradués.

La réflexion en programmation consiste à permettre à un programme d'inspecter et de modifier l'état de sa propre exécution. Le but est de permettre à ce programme d'adapter son propre comportement aux conditions dynamiques de son exécution, ceci pour optimiser son code, expliquer son comportement (trace, "debugging"), ajouter des fonctionnalités au langage de base (gestion d'exception par exemple), etc.

Un langage de programmation réflexif est un langage qui supporte la réflexion en représentant ses programmes ainsi que leur état d'exécution à l'aide de ses propres procédures et structures de données. Ces procédures et structures de données sont rendues accessibles au programme de l'utilisateur à l'exécution. Actuellement, plusieurs langages de programmation offrent, au moins en partie, des capacités réflexives : les langages à objets Smalltalk-80 et CLOS, le langage de programmation logique Prolog, le langage fonctionnel Scheme, etc.

Nos projets se développent selon trois axes de recherche principaux :

1. Langage de programmation par objets réflexif. Cet axe de recherche s'attarde principalement aux problèmes d'implantation efficace des langages réflexifs, mais aussi à la conception des modèles, méthodes et techniques de programmation réflexive dans les langages à objets qui favorisent la facilité d'utilisation et la réutilisation du code réflexif.
2. Application des techniques de transformation de programmes basées sur la sémantique (évaluation partielle, interprétation abstraite) pour résoudre les problèmes d'efficacité des langages réflexifs.
3. Étude des langages réflexifs par des approches théoriques (sémantique dénotationnelle, systèmes de réécriture, monades, etc.) à la fois pour en définir la sémantique et pour obtenir des méthodes d'implantation efficace formellement démontrables.

Les projets de recherche en cours portent sur l'implantation d'un langage à prototypes (objets) réflexif réifiant les continuations comme objet de plein droit, sur l'utilisation des monades pour représenter la partie comportementale du langage, l'évaluation partielle pour une

implantation efficace de la réflexion en en Prolog, et sur le développement d'outils de programmation, fouineur (*browser*), dévermineur (*debugger*) et autres, pour langages à métaclasses explicites en Smalltalk-80. Un projet d'implantation efficace du langage Dylan avec des extensions réflexives est en cours de démarrage en collaboration avec l'équipe de programmation fonctionnelle et parallèle de Marc Feeley.

Collaborateurs

- Pierre Cointe, professeur, et Philippe Mulet, thésard, École des Mines de Nantes.
- Marc Feeley, professeur, Département d'informatique et r.o., Université de Montréal.
- Christian Queinnec, professeur a l'École Polytechnique (Palaiseau) et chargé de recherche à l'INRIA-Rocquencourt.
- Charles Consel, professeur à l'Université de Rennes.
- Christophe Dony, maître de conférence, LIRMM, Université de Montpellier.

7.2 Scriptum

Pour obtenir plus de renseignements sur le groupe SCRIPTUM, vous pouvez vous adresser à l'adresse électronique suivante : `scriptum@IRO.UMontreal.CA`

Note : Scriptum signifie "écrit" en latin ; les amateurs d'acrostiches pourront retrouver ce sigle dans "Système Capable de Reproduire Intelligemment des Phrases et des Textes à l'Université de Montréal" ou bien "Système de CRéation Informatique de Petits Textes à l'Université de Montréal".

Introduction

Un aspect très important des interfaces personne-machine est la possibilité d'obtenir une génération automatique de textes en langue naturelle à partir des structures d'information conservées dans l'ordinateur. En effet, la complexité de celles-ci rend pratiquement inutilisable l'approche consistant à prévoir à l'avance toutes les possibilités d'interactions au moyen de textes préparés à l'avance. Il y a bien certaines possibilités de paramétrisation au moyen de "formats variables" mais elles restent très limitées et très en deça des interactions qu'on serait en droit d'attendre de la part d'un système qu'on voudrait qualifier d'intelligent.

La génération de texte couvre un très large éventail de problèmes intéressants: la représentation des connaissances en fonction des domaines et des applications, les problèmes conceptuels comme le choix des informations à exprimer, les problèmes linguistiques comme le découpage en paragraphes ou en phrases, la sélection des constructions et des mots, la syntaxe des phrases et la morphologie. Or ces aspects du traitement de la langue naturelle ont été peu étudiés en comparaison avec les problèmes d'analyse qui eux ont fait l'objet de multiples théories et méthodes. Nous nous concentrons sur les aspects linguistiques et informatiques en ne considérant pas les problèmes de représentation de connaissances dans leur ensemble.

Nous visons surtout deux classes de domaines où l'information est stéréotypée ou bien structurée: dans la première classe, nous nous intéresserons à des situations stables ou évolutives décrites par des informations numériques souvent conservées sous forme de tables semblables à celles rencontrées dans les banques de données relationnelles. Les "procédures" forment la deuxième classe de domaines qui nous intéressent. On y trouve des instructions d'assemblage, de réparation ou d'entretien qui spécifient des suites d'actions à effectuer dans un monde de référence concret pour obtenir un résultat voulu. Ce qui caractérise ces types de textes est la présence d'informations "objectives" sous forme de nombres ou de relations entre les faits.

Plusieurs étudiants ont effectué récemment ou effectuent encore leur recherche au sein du groupe SCRIPTUM, en voici la liste :

- Massimo Fasciano (Ph.D. Informatique) travaille sur la "Génération intégrée de textes et de graphiques statistiques" (voir aussi p. 8).
- Michel Gagnon (Ph.D. Informatique obtenu en 1993) a conçu "PréTexte: un générateur pour l'expression de concepts temporels".
- Leila Kosseim (Ph.D. Informatique) travaille sur la planification de textes d'instructions, la sélection du contenu et de la structure rhétorique (voir aussi p. 9).
- Daniel Rousseau (Ph.D. Informatique) travaille à la modélisation de la simulation de conversations dans un univers multi-agents (voir aussi p. 11).
- Nicole Tourigny (Ph.D. Informatique obtenu en 1994), a étudié la "Génération de description de systèmes dynamiques";
- Agnès Tutin (Ph.D. Linguistique obtenu en 1992) pour la recherche et Leila Kosseim (M.Sc. Informatique) pour l'implantation, ont travaillé sur la "Génération automatique de procédés anaphoriques dans les instructions".

8 Les séminaires Incognito

L'une des traditions du laboratoire est le séminaire hebdomadaire donné par les étudiants.

Le séminaire prend la forme d'une présentation de vingt à trente minutes, généralement en français, suivie d'une discussion informelle. Les présentations sont faites par les étudiants du laboratoire sur un sujet qui les intéresse particulièrement, typiquement il s'agit des sujets de maîtrise ou doctorat actuel ou passé.

Exceptionnellement, ces séminaires sont donnés par des chercheurs externes de passage à Montréal et invité par l'un des membres du laboratoire.

Le but de ces séminaires est triple :

- faciliter la communication entre les étudiants de façon à ce que chacun sache ce que les autres font,
- pouvoir discuter avec d'autres de son sujet de recherche et ainsi affiner ses idées,
- donner aux étudiants l'habitude de présenter et défendre leurs idées et leur travail devant une assistance qui leur est déjà acquise.

La liste ci-dessous (séminaires Incognito de 1993 et 1994) donne une idée de la diversité des sujets abordés lors de ces séminaires :

Session Hiver 1993

- | | |
|---------|---|
| 11 jan. | PIERRE SAINT-VINCENT (doctorat), " <i>La création et l'utilisation d'outils informatiques pour aider à l'interprétation des lois</i> " |
| 18 jan. | TANGUY KERVAHUT (analyste informatique), " <i>Recherche tabou pour le problème de tournées de véhicules avec fenêtre de temps</i> " |
| 25 jan. | ANNABEL SADAN (maîtrise), " <i>Classification de conformation macromoléculaire</i> " |
| 1 fév. | FRANÇOIS ROUSSEAU (maîtrise), " <i>Adaptation dynamique des outils de programmation à l'aide de protocoles d'accès aux structures des classes</i> " |
| 8 fév. | ÉLIANA COELHO (doctorat), " <i>Modèles d'explications pour systèmes basés en logique</i> " |
| 8 mars | FRANÇOIS-NICOLA DEMERS (maîtrise), " <i>L'application de l'évaluation partielle à l'optimisation des tours réflexives</i> " |
| 15 mars | ATOULÉ KOKOROKO (maîtrise), " <i>Mise à jour automatique de texte de loi</i> " |
| 22 mars | JEAN VAUCHER (professeur), " <i>Compte rendu du projet Macroscope</i> " |
| 29 mars | ANANI KUASSIVI (maîtrise), " <i>L'aide multicritère à la décision</i> " |
| 5 avr. | GILLES ROY (doctorat), " <i>Réécriture incrémentale de termes augmentés et hiérarchie de contraintes</i> " |

Session Automne 1993

- 13 sep. MASSIMO FASCIANO (doctorat), “*A Prolog Implementation of the Functional Unification Grammar Formalism*”
- 27 sep. NICOLAS ANQUETIL (doctorat), “*Sélection de nœuds dans un treillis de concepts*”
- 4 oct. GILLES PESANT (doctorat), “*Les contraintes, quossé ça ?*”
- 18 oct. MARCO JACQUES (maîtrise), “*Implantation d’un langage à prototypes réflexif avec continuations*”
- 25 oct. MARTIN BRISEBOIS (maîtrise), “*Intégration d’une inférence sémantique à un modèle classique de RI*”
- 1 nov. FRANÇOIS ROUSSEAU (maîtrise), “*Un browser pour langage à méta-classes explicites*”
- 8 nov. SAMY BENGIO (stage post-doctoral), “*Utilisation des algorithmes génétiques pour le problème des tournées de véhicules avec fenêtres de temps*”
- 15 nov. DANIEL ROUSSEAU (doctorat), “*Modélisation des conversations basées sur une perspective de systèmes multi-agents*”
- 22 nov. ILHAM BENYAHIA (stage post-doctoral), “*Une architecture pour l’automatisation de la circulation des informations en temps réel*”
- 29 nov. CHRISTIAN FOISY (doctorat), “*Parallélisme de donnée + ML = DPML*”

Session Hiver 1994

- 17 jan. FRANÇOIS-NICOLA DEMERS (maîtrise), “*L’application de l’évaluation partielle à la réflexion de comportement en Prolog*”
- 24 jan. LEILA KOSSEIM (doctorat), “*Planification de textes d’instructions pour différents buts communicatifs*”
- 31 jan. GILLES PESANT (doctorat), “*Les contraintes, quossé ça? II: Le retour...*”
- 7 fév. STÉPHANE SOMÉ (doctorat), “*Définition des exigences des usagers par des scénarios et génération de spécifications complètes*”
- 4 fév. AKILA SARIRETE (doctorat), “*Intégration intelligente de sources multiples de connaissances pour l’aide à la décision*”
- 1 fév. AMINA ARFI (maîtrise), “*Hiérarchie d’interfaces des collections de classes de Smalltalk-80*”
- 7 mars ISABELLE THORÉ (maîtrise), “*Intelligence Artificielle dans le domaine légal*”
- 14 mars MASSIMO FASCIANO (doctorat), “*Génération intégrée de textes et de graphiques statistiques*”
- 21 mars MARCEL TURCOTTE (doctorat), “*Exercice de style: implémentations Miranda, C, Scheme et Multilisp du système MC-SYM*”
- 28 mars MICHEL GAGNON (stage post-doctoral), “*Vers une représentation granulaire du temps pour le traitement de la langue naturelle*”
- 11 avr. BENOIT LAVOIE (maîtrise), “*Étude sur la modélisation de la planification de textes dans le domaine des rapports statistiques*”
- 18 avr. PAUL-ANDRÉ MONNEY (chercheur invité), “*La théorie de Dempster-Shafer appliquée au diagnostique logique*”

Session Automne 1994

- 26 sep. NICOLAS ANQUETIL (doctorat), "*Classification, ontologies, graphes de concepts et le modèle à facettes*"
- 3 oct. JIAN-YUN NIE (professeur), "*La ségmentation du chinois*"
- 17 oct. CHARLES DE LEAN (baccalauréat), "*Un browser graphique dans SmallTalk*"
- 24 oct. HONG ZHAO (maîtrise), "*Système d'aide à la généralisation dans les bases de données géographiques*"
- 31 oct. PHILLIPE MULET (doctorat à l'École des Mines de Nantes), "*La composition de méta-objets*"
- 7 nov. JACQUES SAVOY (chercheur associé), "*La conférence TREC (Text REtrieval Conference) et les liens interdocuments en recherche d'informations*"
- 14 nov. MARCO JACQUES (maîtrise), "*De la parenté entre les environnements de MIT Scheme et les prototypes de Self*"
- 21 nov. MARTIN BRISEBOIS (doctorat), "*Using General Thesaurus to Set Fuzzy Relevance of Terms in Information Retrieval*"
- 28 nov. FRANÇOIS-NICOLA DEMERS (doctorat), "*Comparaison d'interprètes écrits dans un langage fonctionnel pur*"
- 5 déc. FRÉDÉRIC RIVARD (doctorat à l'École des Mines de Nantes), "*Extension du compilateur SmallTalk: Application à la paramétrisation de l'envoi de message*"

9 Incognito par l'Internet

Cette section regroupe les différentes adresses électroniques pour rejoindre les membres du laboratoire ou aller chercher de l'information.

Le laboratoire Incognito en lui même possède une adresse de courrier électronique : `cogito@IRO.UMontreal.CA` à laquelle peuvent être adressées toutes les demandes qui ne relèvent pas d'un membre particulier.

Membres du laboratoire

Nous listons ici les adresses électronique (e-mail) de tous les membres du laboratoires mentionnés dans ce document :

Professeurs	Michel Boyer	<code>boyer@IRO.UMontreal.CA</code>
	Guy Lapalme	<code>lapalme@IRO.UMontreal.CA</code>
	Jacques Malenfant	<code>malenfant@IRO.UMontreal.CA</code>
	Jian-Yun Nie	<code>nie@IRO.UMontreal.CA</code>
	Jean Vaucher	<code>vaucher@IRO.UMontreal.CA</code>
Étudiants	Nicolas Anquetil	<code>anquetil@IRO.UMontreal.CA</code>
	Martin Brisebois	<code>briseboi@IRO.UMontreal.CA</code>
	Éliana Coelho	<code>coelho@IRO.UMontreal.CA</code>
	François-Nicola Demers	<code>demers@IRO.UMontreal.CA</code>
	Massimo Fasciano	<code>fasciano@IRO.UMontreal.CA</code>
	Leila Kosseim	<code>kosseim@IRO.UMontreal.CA</code>
	Marco Jacques	<code>marcoj@IRO.UMontreal.CA</code>
	Gilles Pesant	<code>pesant@IRO.UMontreal.CA</code>
	Daniel Rousseau	<code>roussead@IRO.UMontreal.CA</code>
	Gilles Roy	<code>royg@IRO.UMontreal.CA</code>
	Akila Sarirete	<code>sarirete@IRO.UMontreal.CA</code>
	Stéphane Somé	<code>some@IRO.UMontreal.CA</code>
	Marcel Turcotte	<code>turcotte@IRO.UMontreal.CA</code>
Hong Zhao	<code>zhao@IRO.UMontreal.CA</code>	

Autres adresses électroniques

Le Département d'Informatique et de Recherche Opérationnelle de l'Université de Montréal dispose d'un serveur WWW (World Wide Web) (<http://www.IRO.UMontreal.CA>) et FTP (<ftp.IRO.UMontreal.CA>) sur lequel le laboratoire Incognito a deux adresses :

- une page WWW en français qui donne accès à une version électronique du présent document (<http://www.IRO.UMontreal.CA/labs/incognito>)
- un serveur ftp autorisant les connexions anonymes (anonymous) qui permet d'accéder à certaines thèses récentes d'étudiants du laboratoire (en format PostScript) ou à des articles (<ftp.IRO.UMontreal.CA:/pub/incognito>).