

Curriculum Vitae : Sylvain SALVATI

Né le 23 janvier 1979 à Nice, union libre, un enfant (1 an et demi).

<i>Coordonnées professionnelles :</i> National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, JAPAN +81(0)3-4212-2664 salvati@nii.ac.jp http ://research.nii.ac.jp/~salvati/	<i>Coordonnées personnelles :</i> 1-9-11 Kanagawa-ken, Kawasaki-shi, Takatsu-ku, Kitamikata, Tobarihaitsu 1, Appt. 203, 213-0005, JAPAN sylvainsalvati@yahoo.fr
---	--

Cursus académique

2001-2005	Doctorat en informatique de l'INPL au sein de l'équipe Calligramme au Loria et sous la direction de Philippe de Groote (obtenu le 13 juin 2005).
2001-2002	Licence de lettres modernes, Paris III.
1998-2001	Diplôme d'ingénieur civil des mines, École Nationale Supérieure des Mines de Nancy.
2000-2001	Diplôme d'études approfondies en informatique, INPL-Loria. Mémoire réalisé sous la direction de Luigi Liquori et la co-direction de Dominique Colnet.
1999-2000	Maîtrise de mathématiques à l'Université Henri Poincaré - Nancy I.
1996-1998	Classe préparatoire MPSI puis MP*, Lycée Masséna à Nice.
1996	Baccalauréat série S (option math).

Parcours professionnel

11/2005-05/2006	Post-doc au National Institute of Informatics (Tokyo, Japon)
10/2004-09/2005	demi-ATER à l'Ecole Nationale Supérieure de Mines de Nancy
10/2001-10/2004	Allocataire de recherche MENRT à l'INPL et moniteur à l'Université Henri Poincaré (Nancy I)

Recherche

Activités de recherche

Mon travail de thèse et la recherche que j'ai menée ensuite s'intéressent à un formalisme grammatical défini par Philippe de Groote [dG01] : les grammaires catégorielles abstraites (Abstract Categorical Grammars ou ACG). Ce formalisme reprend l'idée sur laquelle se fonde le calcul de Lambek, à savoir, modéliser la grammaticalité d'une phrase par la prouvabilité d'une formule dans une logique donnée. Cependant, alors que le calcul de Lambek code par des formules logiques et les dépendances syntaxiques et l'ordre des mots, les ACG scindent l'analyse en deux niveaux : un niveau abstrait qui rend compte seulement des dépendances syntaxiques et un niveau objet qui décrit l'incidence de ces dépendances sur l'agencement des mots. L'uniformité du formalisme provient de la correspondance démonstration/programme (ou isomorphisme de Curry-Howard appliquée) à la logique linéaire et au λ -calcul linéaire. Ainsi, un type abstrait peut être vu comme une spécification grammaticale et sa réalisation syntaxique au niveau objet comme un programme particulier réalisant cette spécification. Ce couplage démonstration/calcul était déjà mis à l'oeuvre dans la théorie de la sémantique de la langue proposée par Montague [Mon74]. Le formalisme des ACG tend à adapter cette approche à l'analyse grammaticale afin de rendre plus transparente l'interface entre syntaxe et sémantique.

De par sa nature même, ce sujet de thèse est interdisciplinaire et a des prolongements tant en mathématiques, en linguistique qu'en philosophie. Le travail que j'ai mené dans ce cadre a surtout été mathématique. Comme le formalisme grammatical proposé par les ACG repose essentiellement sur la théorie de la démonstration et le λ -calcul, la recherche de démonstration et la résolution d'équations symboliques dans le λ -calcul ont été les sujets privilégiés sur lesquels s'est concentré mon travail. Dans un premier temps, j'ai étudié la difficulté de l'analyse grammaticale des grammaires catégorielles abstraites et les divers processus qu'elle met en oeuvre : démonstration dans la logique linéaire et filtrage dans le λ -calcul linéaire. La plupart de ces problèmes se sont révélés être NP-complets, aussi, par la suite, je me suis attaché à trouver des heuristiques générales pour les résoudre ainsi que des heuristiques dédiées à l'analyse grammaticale de la langue dans le cadre des ACG. Cette méthodologie de travail s'étant avérée fructueuse, j'ai commencé à la remettre en oeuvre pour étendre le formalisme en enrichissant la logique et le calcul qui le sous-tendent.

J'ai par ailleurs pu isoler un catégorie assez naturelle de grammaires pour lesquelles l'analyse s'effectue en temps polynômiale. Pour ces grammaires j'ai pu élaborer un cadre théorique qui permet de concevoir des algorithmes d'analyse efficaces. En particulier, ce cadre théorique a permis l'élaboration d'un algorithme d'analyse qui reproduit le comportement de l'algorithme de Earley pour les ACG qui représentent des langages hors contexte et qui permet d'analyser les grammaires TAG en $\mathcal{O}(n^6)$. Cet algorithme permet d'étendre pour la première fois les techniques d'analyse à la Earley pour la classe des langages moyennement sensibles au contexte, classe qui est généralement considérée comme un bon candidat pour modéliser les langues naturelles.

Par ailleurs, ce cadre théorique se trouve avoir des applications intéressantes dans la résolution des équations de filtrage, notamment en permettant de définir des algorithmes efficaces et originaux.

L'ensemble du travail réalisé au sein de l'équipe Calligramme autour du formalisme des ACG a éveillé l'intérêt de Makoto Kanazawa, chercheur reconnu pour son travail concernant l'apprentissage automatique de grammaires. De là a découlé une collaboration qui a commencé par une visite de deux semaines que nous a amenés, Philippe de Groote et moi-même, à travailler au NII avec Kanazawa. Cette collaboration se poursuit, à l'heure actuelle par ma visite en tant que chercheur invité, celle à Nancy de Ryo Yoshinaka, puis en février par la venue, pour deux semaines encore, de Philippe de Groote et Sylvain Pogodalla.

Recherche en DEA

Le sujet de mon DEA portait sur le typage des langages orientés objet. En particulier, je me suis intéressé au rapport entre la notion de sous-typage et la notion d'héritage. Ce travail a été effectué dans le cadre du projet FunTalk de l'équipe Mirò qui avait pour but de produire un langage orienté objet certifié possédant entre autre un système de typage à la fois souple et consistant avec la notion de calcul. À cette fin, dans mon rapport de DEA, *FunTalk formalisation et démonstration*, j'ai proposé un système de typage cohérent (*i.e.* qui bénéficie des propriétés de typage le plus général et de la réduction du sujet) et qui allie :

1. l'utilisation du type de l'objet courant (similaire à l'instruction `like current` en Eiffel).
2. une notion de sous-typage découplée de l'héritage qui élargit le polymorphisme à l'ensemble des objets.
3. une relation qui généralise la notion d'héritage et l'étend à l'ensemble des objets afin de pouvoir définir simplement de classes paramétrées.

Publications dans des conférences internationales avec comité de lecture et actes

- Philippe de Groote, Sylvain Salvati, *On the complexity of higher-order matching in the linear λ -calculus*. In *Rewriting Techniques and Applications, 14th International Conference, RTA 2003*, LNCS 2706, pages 234-245 (**prix du meilleur article de la conférence**).
- Philippe de Groote, Bruno Guillaume, Sylvain Salvati, *Vector Addition Tree Automata*. In *Proc. of the 19th Symp. on Logic In Computer Science 2004*, pages 64-73.
- Philippe de Groote, Sylvain Salvati, *Higher-order Matching in the linear λ -calculus with pairing*. In *Computer Science Logic, 13th International Conference, CSL 2004*, LNCS 3210, pages 220-234.

Publications universitaires

- Sylvain Salvati, *Problèmes de filtrage et problèmes d'analyse pour les grammaires catégorielles abstraites*, thèse de doctorat de l'Institut National Polytechnique de Lorraine, soutenue le 13 juin 2005.
- Sylvain Salvati, *Funtalk, formalisation et démonstration*, rapport de DEA de l'Institut National Polytechnique de Lorraine, 2001.

Distinctions

1. Prix du meilleur papier de la conférence RTA 2003 pour l'article *On the complexity of higher-order matching in the linear λ -calculus*.
2. Médaille du Loria 2003.

Références

- [dG01] Philippe de Groote. Towards abstract categorial grammars. In Association for Computational Linguistic, editor, *Proceedings 39th Annual Meeting and 10th Conference of the European Chapter*, pages 148–155. Morgan Kaufmann Publishers, 2001.
- [Mon74] Richard Montague. *Formal Philosophy : Selected Papers of Richard Montague*. Yale University Press, New Haven, CT, 1974.

Année	Volume (h. eq. TD)	Public	Intitulé	Responsable
2004-2005 ($\frac{1}{2}$ -ATER)	63	Troisième année d'école d'ingénieur	Statistique et Algorithmique pour la bioinformatique	Karl Tombre
	23	Première année d'école d'ingénieur	Initiation à la programmation Java	Guillaume Bonfante
	10	Troisième année d'école d'ingénieur	Suivi de Stage ingénieur	Karl Tombre
2003-2004 (moniteur)	34	DEUG MIAS 2 ^{ème} année	TD de mathématiques discrètes	Adam Cichon
	28	DEUG SM 1 ^{ère} année	TD/TP de programmation CAML	Adam Cichon
2002-2003 (moniteur)	36	DEUG MIAS 2 ^{ème} année	TD de mathématiques discrètes	Pierre Bouchet
	35	DESS CCI	Cours/TD HTML et XML	Jean-Marie Pierrel
2001-2002 (moniteur)	34	DEUG MIAS 2 ^{ème} année	TD de mathématiques discrètes	Pierre Bouchet
	36	DEUG MIAS 1 ^{ère} année	TD/TP de programmation CAML	Jean-Pierre Jacquot

FIG. 1 – Durée, et nature des enseignements

Enseignement

D'octobre 2001 à octobre 2004, j'ai effectué un total de 203 heures équivalent TD en tant que moniteur à l'Université Henri Poincaré-Nancy 1, encadré par Jean-Pierre Jacquot (Jean-Pierre.Jacquot@loria.fr), mon tuteur. Au cours de ces trois années, j'ai eu à intervenir dans de nombreux cours de DEUG où j'ai assuré les travaux dirigés. Ces cours portaient essentiellement sur les bases de l'informatique : l'initiation au langage CAML et les mathématiques discrètes. J'ai eu également à préparer un cours de DESS double compétence. Ce cours portait sur la structuration de document, d'abord avec HTML, puis avec XML.

D'octobre 2004 à septembre 2005, j'ai effectué un demi-ATER à l'Ecole Nationale Supérieure des Mines de Nancy. L'axe principal de mon travail pédagogique était la mise en place de la formation en bioinformatique à l'Ecole des Mines qui débutait à la rentrée 2004. Suivant un ensemble d'objectifs prédéfinis, j'ai en collaboration avec Arnaud Courtois, également demi-ATER cette année-là, défini le programme de ce nouvel enseignement ainsi que préparé une partie de ces cours.

Vous trouverez le détail de ces heures dans le tableau récapitulatif 1.

Détail des enseignements

Prog. Java (23h.) Ce cours s'adresse à des élèves ingénieur de première année. Il s'agit de les initier aux problématiques de la programmation à travers le

langage Java. Cet enseignement se déroule tout au long de l'année. Je n'y suis intervenu que pendant le second semestre où ont été abordés : la notion de variables statiques, l'héritage, les classes abstraites et les interfaces, le traitement des exceptions et les bases de la création d'interfaces graphiques.

Bioinformatique (63h.) L'ouverture de la filière commune à l'ENSIC, l'ENSAIA, et l'ENSMN de la spécialité bioinformatique a été pour moi l'occasion de concevoir un cours dans son intégralité. N'avoir aucune expertise en bioinformatique a présenté pour moi de nombreux intérêts. Tout d'abord, cela m'a apporté l'occasion d'acquérir une vue d'ensemble sur un domaine de l'informatique à la fois intéressant et en plein essor. Ensuite cela m'a permis d'exercer ce à quoi j'ai pris goût en doctorat à savoir créativité, autoformation et esprit critique. Finalement, mon domaine de recherche, tout comme la bioinformatique, est pluri-disciplinaire. J'ai été curieux d'observer comment la pluri-disciplinarité de la bioinformatique et celle du traitement automatique des langues recouvraient des réalités différentes et contrastées.

La filière étant en création, cette année seuls cinq étudiants en ont suivi les cours. Quatre d'entre eux étaient des ingénieurs ergonomes avec de faibles connaissances en informatique. J'ai eu la charge du cours intitulé "algorithmiques et de probabilités pour la bioinformatique". Les objectifs pédagogiques que je m'étais fixés étaient d'apprendre à utiliser les logiciels couramment employés par les biologistes, de comprendre leur fonctionnement et de savoir en créer qui puissent répondre à des problèmes simples. J'ai essayé de présenter mon cours en partant d'une tâche que les biologistes souhaiteraient voir automatiser, en élaborant ensuite avec les étudiants une proposition de modèle mathématique, en donnant les outils conceptuels nécessaires à la résolution du problème, puis en concevant les algorithmes de résolution proprement dits. J'ai constamment cherché à montrer que les algorithmes et les modèles proposés étaient le résultat d'un compromis entre le réalisme biologique d'une part et l'efficacité algorithmique d'une autre. A mesure de la progression dans le cours, j'ai préparé des travaux pratiques pour apprendre l'utilisation des outils implémentant les divers algorithmes vus en cours. Par ailleurs, j'ai accompagné le cours d'une initiation à la programmation avec le langage Perl, initiation qui s'est achevée avec la découverte et l'utilisation des bibliothèques de BioPerl. Etant face à un public ayant déjà de bonnes bases de probabilité, je ne les ai utilisées que comme outil pour expliquer la significations des résultats des algorithmes, et la conceptions d'algorithme basées sur des probabilités. L'ensemble de mon cours était structuré autour de l'analyse de séquences et j'ai présenté :

1. les problèmes de la recherche de motifs,
2. de l'alignement de deux séquences,
3. celui de l'alignement de plusieurs séquences,
4. l'alignement de séquences à l'aide des chaînes de Markov cachées ainsi que les techniques d'apprentissage associées
5. diverses techniques de construction d'arbres philogénétiques
6. la prédiction des structures secondaires des chaînes protéiques ou ARN

J'ai essayé lorsque cela était possible de faire ressortir les notions informatiques importantes comme les langages réguliers, les automates, les algorithmes tabulaires, la notion de NP-complétude et celle de polynômialité, la correction et la complétude des algorithmes, les langages algébriques. . .

Pour moi, la difficulté essentielle de ce cours a été de trouver des exemples biologiques pertinents. Je me suis pour l'essentiel inspiré de ce que j'ai pu trouver dans les livres et les publications dont je disposais. Mais par moment, une vraie connaissance des pratiques en biologie aurait été nécessaire.

Math. discrètes (104 h.) J'ai assuré les travaux dirigés de mathématiques discrètes durant trois années. Au cours de ce module sont abordés :

1. la logique formelle (algèbre de Boole, calcul des propositions et calcul des prédicats) où l'on développe la notion de démonstration, puis des techniques de recherche automatique de démonstrations comme les tableaux sémantiques ou la résolution de Robinson.
2. la théorie des graphes où l'on aborde les notions de graphe planaire et de graphe eulérien.
3. la théorie des langages qui pour l'essentiel traite des langages réguliers et des automates d'états finis. À la fin du cours sont cependant introduites les grammaires hors contexte.

Prog. CAML (64 h.) J'ai assuré les travaux dirigés et les travaux pratiques pour deux cours d'introduction à la programmation en DEUG première année. Ces enseignements se basent sur la proximité du langage CAML avec les mathématiques pour initier les étudiants aux problématiques de la programmation. On y aborde notamment la notion de typage, de structure conditionnelle et de récursivité tant sur des entiers que sur des structures comme des listes. Au profit d'un nombre d'heures plus conséquent, la notion de type énuméré a pu être abordée en DEUG MIAS.

HTML XML (35 h.) Dans le cadre du DESS CCI proposé par l'Université Henri Poincaré-Nancy 1, j'ai eu l'occasion de me voir remettre la responsabilité d'un cours sur la structuration des documents à l'aide de XML. Ce cours commençait par une présentation de la notion de langages de balises avec une initiation à HTML. Par la suite, j'ai montré comment, à l'aide des feuilles de style, on pouvait se servir de la structuration logique des documents pour obtenir une présentation claire et homogène. Finalement, les limites de cette approche m'ont permis d'introduire le langage XML ainsi que les outils et concepts qui l'entourent comme la validité des documents contre une DTD ou un schéma XML, la transformation de documents avec XSLT, ou comment affiner ces notions à l'aide de Xpath et définir des liens multiples avec Xlink.

Ayant à assumer l'élaboration du cours dans son ensemble, j'ai dû composer avec le fait que les étudiants n'étaient pas des spécialistes en informatique. J'ai pu me rendre compte que le recours à des notions mathématiques abstraites étaient la plupart du temps éclairant pour des étudiants souvent perdus dans la profusion des notions et la syntaxe de XML.