INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Projet ATOLL

ATelier d'Outils Logiciels pour le Langage naturel

Rocquencourt

THÈME 3A

Table des matières

1 Composition de l'équipe

Responsable scientifique

Bernard Lang [DR]

Responsable permanent

Pierre Boullier [DR]

Assistante de projet

Josy Baron [TR]

Personnel Inria

Philippe Deschamp [CR, depuis mars 1999]

Éric Villemonte de la Clergerie [CR]

Collaborateurs extérieurs

Jean-Marie Larchevêque [MC, IUT de Vélizy jusqu'en septembre 1999, XEROX-XRCE ensuite]

François Barthélemy [MC, CNAM]

Roland Dachelet [MC, CUEJ Université de Strasbourg]

Doctorants

Vitor Rocio [Thèse en co-tutelle avec l'Université Nouvelle de Lisbonne]

François Role [Fonctionnaire au DISTNB-MESR, Université d'Orléans]

Stagiaires

Saad Berrada [février–mai 1999, ENSIAS, Université Mohammed V, Maroc]

Djamé Seddah [février–septembre 1999, DEA de l'Université Paris 7]

Dilek Dustegor [janvier–février 1999, Université de Galatasaray (Turquie)]

Galip Tartanoglu [juillet–août 1999, Université de Galatasaray (Turquie)]

2 Présentation et objectifs généraux

L'équipe Atoll s'est constituée autour d'une compétence dans les techniques d'analyse syntaxique et d'évaluation tabulaire des programmes logiques. Cette compétence, essentiellement acquise dans le cadre de la compilation des langages de programmation, est maintenant appliquée pour le **traitement de la langue naturelle**, dans ses aspects syntaxiques, voire sémantiques. Ce domaine de recherche est en effet riche de problèmes sur le plan scientifique, peut bénéficier d'une approche formelle et algorithmique solide et est prometteur quant aux applications industrielles.

Cependant, notre équipe ne peut couvrir qu'un champ restreint des nombreux problèmes liés au traitement de la langue. Ainsi, mettre en place un système complet de traitement pour l'analyse de documents ou la traduction automatique dépasse nos moyens et compétences actuels.

Nous cherchons donc à développer progressivement des aspects plus appliqués du traitement de la langue en nous appuyant sur nos autres points forts liés à nos compétences informatiques et en nous associant à d'autres acteurs plus directement impliqués dans les problèmes de traitement de documents électroniques et de linguistique appliquée.

L'usage en plein essor des documents électroniques et structurés, dû en grande partie au développement de la «toile» WWW (le «World Wide Web»), nous parait une opportunité à exploiter, notamment en raison de notre expérience concernant les environnements de programmation. En conséquence, nous cherchons à nous diversifier vers des secteurs plus appliqués, à l'occasion de thèses, mémoires et coopérations. Cependant nous souhaitons aussi, au travers de coopérations, établir des liens nous permettant de faire valoir nos résultats algorithmiques et les systèmes qui les implantent.

Le développement de nos activités présente donc actuellement deux aspects, que nous ferons converger à terme :

1. Poursuite de nos travaux sur les techniques fondamentales en analyse syntaxique et évaluation tabulaire de programmes et grammaires logiques, avec développements de prototypes distribuables.

2. Recherche, traitement et gestion des documents électroniques, en particulier dans leur dimension linguistique.

Nos travaux étant nécessairement limités à un champ étroit de la linguistique informatique, il nous faut pouvoir travailler dans le contexte de ressources et d'outils développés par d'autres équipes. Malheureusement, dans ce domaine comme dans d'autres, le libre accès aux ressources scientifiques et techniques se fait de plus en plus difficile et coûteux. Cela nous a amené à nous pencher sur la possibilité du développement de ressources libres. Ce thème est devenu un sujet à part entière, dont l'intérêt scientifique, économique et politique a considérablement crû au cours de cette année.

3 Fondements scientifiques

3.1 Formalismes grammaticaux

Mots clés : analyse syntaxique, linguistique, programmation dynamique, programmation logique.

Participants : Pierre Boullier, Éric Villemonte de la Clergerie, Bernard Lang, Jean-Marie Larchevêque.

Résumé : *Ce thème concerne l'analyse syntaxique de différents formalismes grammaticaux servant au traitement de la langue naturelle. L'ensemble de ces formalismes forme un continuum très large pour lequel des techniques génériques d'analyse sont étudiées qui permettent de traiter au mieux l'ambiguïté inhérente à toute langue.*

Glossaire :
CFG *Context-Free Grammars*
DCG *Definite Clause Grammars*
TAG *Tree Adjoining Grammars*
LIG *Linear Indexed Grammars*
LFG *Lexical Functional Grammars*
HPSG *Head-driven Phrasal Structure Grammars*
RCG *Range Concatenation Grammars*
GIG *Graph Interpolation Grammars*
MCG *Mildly Context-sensitive Grammars*
LPDA *Logical Push-Down Automata*
Programmation Dynamique technique de construction d'algorithmes consistant à diviser un problème en sous-problèmes élémentaires dont les solutions sont tabulées pour pouvoir être réutilisées plusieurs fois si nécessaire.

3.1.1 Des langages de programmation aux grammaires linguistiques

Le passage des grammaires pour les langages de programmation vers des grammaires pour les traitements linguistiques se traduit avant tout par un saut en complexité et l'obligation de gérer les ambiguïtés du langage. Il est bien connu que les problèmes d'ambiguïté en linguistique sont source d'explosions combinatoires mal maîtrisées.

De plus, alors que la syntaxe des langages de programmation se définit souvent par une (sous-classe d'une) grammaire non contextuelle (CFG), aucun formalisme de description de la syntaxe des langues naturelles n'a fait l'unanimité des linguistes. On assiste au contraire à l'éclosion régulière de nouveaux formalismes grammaticaux, avec en particulier les grandes catégories suivantes :

Formalismes dépendant faiblement du contexte Ils regroupent entre autres les grammaires d'arbres adjoints (TAG) et linéaires indexées (LIG) et possèdent une base structurelle qui assure l'existence d'évaluateurs travaillant en temps polynomial.

Grammaires d'unification Elles combinent un squelette non contextuel et une décoration donnée par des attributs logiques. Les représentants les plus connues sont les Grammaires de Clauses Définies (DCG) où l'unification à la PROLOG est utilisée pour calculer et propager ces attributs. Les formalismes plus récents s'appuient sur des structures typées de traits [Car92] ou éventuellement sur des contraintes. Nous avons ainsi les *Lexical Functional Grammars* (LFG) [MK96] et *Head-Driven Phrasal Structure Grammars* (HPSG) [PS94].

Grammaires stochastiques Pratiquement, toute grammaire peut être décorée avec des probabilités ou des pondérations, afin de mieux coïncider avec l'usage de la langue rencontrée sur un corpus de textes. Ces probabilités peuvent être vues comme des décorations prises dans un demi-anneau, dont les propriétés algébriques sont exploitables au cours de l'analyse syntaxique [Ten97].

Les spécificités évoquées précédemment peuvent se combiner, par exemple en ajoutant des contraintes et des attributs logiques sur une grammaire d'arbres adjoints. Ajoutons que nous participons à ce foisonnement de formalismes grammaticaux avec les RCG (Section 5.1) et les GIG (Section 5.3).

Cependant, malgré cette diversité, la plupart des formalismes grammaticaux linguistiques trouvent place dans ce qu'on peut appeler le «**continuum de Horn**», c'est-à-dire un ensemble de formalismes de complexité croissante, allant des clauses de Horn propositionnelles aux clauses de Horn du premier ordre (grosso-modo PROLOG), et même au-delà.

Ce constat motive notre travail de développement de techniques générales d'analyse permettant de couvrir ce continuum, ceci au travers de deux approches complémentaires qui utilisent, toutes les deux, les techniques de la programmation dynamique afin de réduire l'explosion combinatoire due au traitement des ambiguïtés :

Approche multi-passe. Elle consiste, lorsque c'est possible, à découper un traitement en une séquence dont les composants ont une complexité (pratique ou théorique) croissante ;

Approche globale. Elle repose essentiellement sur la description du formalisme grammatical et des stratégies d'analyse à l'aide d'automates à piles.

Ces deux approches ne s'opposent pas. Au contraire, chacune enrichit l'autre. L'examen de particularités mises en évidence par l'approche multi-passe permet des avancées théoriques ; réciproquement, des concepts théoriques bien compris et identifiés se traduisent par un élargissement du champ d'action de l'approche multi-passe.

[Car92] B. Carpenter, *The Logic of Typed Feature Structures with Applications to Unification Grammars, Logic Programs and Constraint Resolution*, ISBN 0-521-41932, Cambridge University Press, 1992.

[MK96] J. T. Maxwell, R. M. Kaplan, «An efficient parser for LFG», *in : Proc. of 1st LFG Conference*, Grenoble, 1996, `http://www.csli.standford.edu/user/mutt/`.

[PS94] C. Pollard, I. A. Sag, *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago, 1994.

[Ten97] F. Tendeau, *Analyse syntaxique et sémantique avec évaluation d'attributs dans un demi-anneau – Application à la linguistique calculatoire*, thèse de doctorat, Université d'Orléans, 1997.

3.1.2 Approche multi-passe

Le traitement des langages de programmation est traditionnellement découpé en phases successives de complexité croissante : analyse lexicale, analyse syntaxique, traitement de la sémantique statique, ... Ce découpage se justifie à la fois par des raisons théoriques et pratiques. Les automates finis qui modélisent l'analyse lexicale n'ont pas la puissance formelle nécessaire pour décrire la partie syntaxique qui nécessite une description par une (sous-classe des) CFG. Les CFG elles-mêmes ne permettent pas de décrire les phénomènes contextuels de la sémantique statique. Outre une efficacité potentielle accrue (chaque phase est traitée avec le bon niveau de formalisme), ce découpage augmente la modularité du processus.

L'approche multi-passe du traitement des langues naturelles résulte d'une vision similaire. On essaie d'isoler dans les formalismes grammaticaux des parties de complexité moindre sur lesquelles le reste du traitement va pouvoir s'appuyer. En fait, on constate que la plupart des formalismes du continuum de Horn sont structurés par une base non-contextuelle forte. Ces grammaires peuvent donc être vues comme une CFG décorée par un système de contraintes. L'approche multi-passe consiste pour tous ces formalismes à utiliser un analyseur non-contextuel général (très performant) sur lequel est greffé le système de contraintes, particulier à chaque formalisme traité. Le traitement du squelette non-contextuel est confié au système SYNTAX.

3.1.3 Approche globale

L'approche multi-passe s'applique moins bien lorsque la structure CF du formalisme est faible (par exemple dans le cas PROLOG) ou lorsque les phases sont interdépendantes (par exemple lorsque le traitement des contraintes conditionne fortement l'analyse CF). Il est alors préférable d'utiliser une approche globale où les contraintes (d'unification ou autres) sont gérées en même temps que l'analyse.

Cette approche, très générale, repose sur des formalismes abstraits d'automates à piles permettant de décrire diverses stratégies d'analyse pour divers formalismes grammaticaux à base logique ou non [4]. Ces automates sont ensuite évalués à l'aide de techniques de programmation dynamique. La notion de pile se prête en effet bien à la division des calculs en sous-calculs élémentaires et réutilisables dans différents contextes : il suffit essentiellement d'oublier provisoirement l'information disponible dans le bas des piles. Ces sous-calculs élémentaires sont représentables sous forme compacte par des *items*. L'utilisation d'automates à 2 piles nous a ainsi récemment permis de traiter les formalismes grammaticaux TAG et LIG [3].

Cette approche trouve ses origines dans les analyseurs à chartes initialement développés par Earley [Ear70]. Elle permet de généraliser différentes méthodes proposées en analyse syntaxique mais aussi en programmation en logique, telles les transformations Magic-Set [Ram88].

Le système DYALOG valide cette approche pour la programmation en logique et pour différents formalismes grammaticaux.

[Ear70] S. EARLEY, « An Efficient Context-Free Parsing Algorithm », *in : Communications ACM 13(2)*, ACM, 1970, p. 94–102.

[Ram88] R. RAMAKRISHNAN, «Magic Templates: A Spellbinding Approach to Logic Programs», *in : Proc. of the 5th Int. Conf. and Symp. on Logic Programming*, p. 140–159, 1988.

3.1.4 Forêts partagées d'analyse

Les deux approches précédentes partagent de nombreuses caractéristiques, par exemple l'utilisation des techniques de programmation dynamique. Nous pouvons également citer la notion de forêt partagée d'analyse ou de dérivation. De telles forêts regroupent sous forme compacte l'ensemble des analyses possibles ou l'ensemble des dérivations possibles pour une phrase et sont en général assimilables à des grammaires ou à des programmes logiques [2]. Ainsi, alors que l'analyse par une CFG peut conduire à un nombre exponentiel (ou même non borné) d'analyses, la forêt d'analyse reste cubique en la longueur de la phrase analysée. Les forêts d'analyse ou de dérivation, qui sont les structures intermédiaires de l'approche multi-passe, constituent de surcroît un point de départ pour des traitements linguistiques ultérieurs (prise en compte de contraintes syntaxiques ou sémantiques complémentaires, traduction, ...).

3.2 Le Poste de Travail Informationnel

Participants : Bernard Lang, François Role.

La recherche de débouchés applicatifs à nos travaux, de pair avec un certain intérêt de l'équipe, nous pousse vers les nouveaux média (principalement cédérom et Internet) dont le rôle économique, social et culturel va croissant. Ceci nous amène naturellement à nous impliquer dans diverses actions dont nous espérons à terme des synergies avec nos compétences en analyse syntaxique et déduction, ainsi qu'avec celles plus anciennes en génie logiciel et traitement de documents structurés.

Plus applicatif, cet axe présente deux volets complémentaires, à savoir d'une part la conception et le développement d'outils pour maîtriser un support matériel des documents qui est en pleine évolution, et d'autre part le développement de techniques d'analyse et de gestion des contenus des documents eux-mêmes. Ces deux aspects sont parfois difficilement dissociables. Par exemple, la réalisation d'un outil de recherche sur le Web requiert à la fois une maîtrise des techniques strictement informatiques de l'accès à l'information, mais aussi des outils sophistiqués d'extraction du contenu des documents (par exemple la lemmatisation des mots pour un indexeur sophistiqué).

Il est également clair que ces problèmes font appel à une grande variété de techniques liées au traitement des documents, à l'analyse de la langue naturelle et à la recherche documentaire. Bien entendu, il ne saurait être question d'acquérir une expertise universelle avec les moyens dont nous disposons, et nous cherchons au maximum à réutiliser des outils existants pour nos travaux, tout en nous efforçant d'identifier et d'explorer des problèmes originaux.

Le thème unificateur que nous fixons à ces activités est le développement d'un *Poste de Travail Informationnel*, permettant à un travailleur intellectuel de gérer facilement son capital d'informations et de documents, tant en ce qui concerne la recherche de nouveaux documents, qu'en ce qui concerne leur mémorisation et leur organisation (indexation) pour une réutilisation ultérieure.

4 Logiciels

4.1 Logiciel SYNTAX

Participants : Pierre Boullier, Philippe Deschamp.

La version 3.8h de SYNTAX est actuellement disponible sous FTP à `http://www-rocq.inria.fr/oscar/FNC-2/getfnc2.html` pour les plateformes Linux/Pentium, SunOs/Sparc et Solaris/Sparc.

La version 3.9 de SYNTAX est actuellement en cours de réalisation. Elle comprendra

- un traitement amélioré des caractères 8 bits ;
- un constructeur de dictionnaires utilisant des techniques de représentation de matrices creuses [1] ;
- un prototype pour le traitement des RCG (voir la section 5.1).

Cette version 3.9 est en cours de portage sur les plateformes Linux, SunOs, Solaris et, nouveauté, sur Windows-NT.

4.2 Logiciel DYALOG

Participants : Éric Villemonte de la Clergerie, Djamé Seddah.

Le logiciel DYALOG est un compilateur de grammaires et de programmes logiques produisant des exécutables tabulaires. Il est plus spécifiquement dédié à la construction d'analyseurs syntaxiques pour le traitement de la langue naturelle mais est aussi utile pour remplacer des systèmes PROLOG traditionnels dans le cadre d'applications très ambiguës avec potentiellement du partage de calculs.

Les sources ou une distribution RPM de la version courante de DYALOG (1.1) sont disponibles pour les plateformes Linux (Pentium) sous FTP à `http://atoll.inria.fr/~clerger`.

Cette version courante permet le traitement des DCG (*Definite Clause Grammars*) et des TAG (*Tree Adjoining Grammars*). Elle offre la possibilité d'utiliser des structures typées de traits pour des écritures plus compactes des grammaires.

5 Résultats nouveaux

5.1 Analyse contextuelle

Participant : Pierre Boullier.

Mots clés : formalismes grammaticaux contextuels, forêts partagées, temps d'analyse polynomial, modularité grammaticale.

1. Ces techniques sont appliquées pour représenter sous forme compacte les automates à états finis utilisés. On peut noter que cette optimisation en place ne s'effectue pas au détriment de l'efficacité : on assure qu'un mot de n caractères est reconnu (ou rejeté) en au plus n comparaisons. Ce constructeur sera utilisé dans le module de LECL qui traite les mots-clés.

Glossaire :
MCS *Mildly Context-sensitive Grammars*
RCG *Range Concatenation Grammars*

Résumé : *Nos recherches sur les grammaires à concaténation d'intervalles se sont poursuivies selon trois axes : une utilisation directe pour décrire des phénomènes linguistiques difficiles, une utilisation indirecte comme formalisme intermédiaire pour implanter d'autres types de grammaires et enfin l'étude de leurs propriétés formelles.*

Nous avons introduit en 1998 un nouveau formalisme syntaxique, la grammaire à concaténation d'intervalles (RCG) qui définit une classe de langages appelée RCL. Les RCG sont puissantes : elles englobent les grammaires non-contextuelles (CFG) et même les formalismes faiblement dépendant du contexte (*mildly context-sensitive*—MCS) tout en conservant un temps d'analyse polynomial. Ce formalisme grammatical possède en outre un certain nombre de propriétés théoriques (citons par exemple sa clôture par intersection et complémentation) qui lui permettent de briguer la place occupée actuellement par les CFG au cœur des systèmes définissant les langues naturelles. Nos recherches se sont poursuivies selon trois axes : une utilisation directe des RCG pour décrire des phénomènes linguistiques difficiles, une utilisation indirecte comme formalisme intermédiaire pour implanter d'autres types de grammaires et enfin l'étude de leurs propriétés formelles.

5.1.1 Utilisation directe des RCG

Il existe certains phénomènes linguistiques que les méthodes syntaxiques usuelles, et en particulier les formalismes MCS, ne peuvent pas décrire. Citons par exemple les nombres en chinois et les constructions permutantes (*scrambling*) de l'allemand.

Ainsi, en dialecte mandarin, le nom pour 10^{12} est *zhao* et 5 se dit *wu*. La séquence *wu zhao zhao wu zhao* est un nombre chinois bien formé (c'est $510^{24} + 510^{12}$) alors que *wu zhao wu zhao zhao* ne l'est pas : le nombre de *zhao* consécutifs doit décroître strictement de gauche à droite. Le langage de l'ensemble des nombres chinois n'est pas MCS car il ne possède pas la propriété de croissance constante[2]. Radzinski a montré que les nombres chinois appartiennent à l'ensemble des langages indexés[3]. Au cours de son étude des nombres chinois, il a même noté que sa recherche pour trouver un formalisme qui décrirait les nombres chinois, sans définir toute la classe des langages indexés, a échoué. Nous avons montré que les nombres chinois peuvent se décrire par une RCG et être analysés en temps linéaire.

Nous avons également montré que les constructions permutantes de l'allemand peuvent se décrire par des RCG et s'analyser dans le pire cas en temps quadratique. Ces travaux sont décrits en [14] et en [22]. Au cours de cette étude, nous avons noté que les constructions permutantes sont apparentées au célèbre langage *MIX* d'Emmon Bach[4]. Jusqu'à présent, on

2. Si on classe les phrases d'un langage par taille croissante, la différence de taille de deux phrases successives ne peut pas être bornée par une constante.

3. Les langages indexés forment une classe de langages pour laquelle aucun algorithme d'analyse polynomial n'est connu.

4. Les phrases du langage *MIX* sont des suites de a, b et c dans un ordre quelconque, mais en nombre identique.

ne savait même pas si MIX était un langage indexé, nous avons montré que c'est un RCL dont les phrases peuvent être analysées en temps linéaire.

5.1.2 Les RCG comme formalisme d'implantation

Le formalisme MCS le plus populaire en linguistique est très certainement celui des grammaires d'arbres adjoints (TAG). L'an dernier, nous avions déjà montré qu'on pouvait traduire une TAG en une RCG équivalente. Cependant, cette transformation n'était proposée que pour une classe restreinte de contraintes d'adjonctions et surtout, le temps d'analyse en $\mathcal{O}(n^6)$ n'était obtenu que pour une certaine forme normale de la TAG initiale. Nous avons généralisé cet algorithme pour lui faire accepter en entrée des TAG et des contraintes d'adjonction quelconques tout en assurant que les phrases de la RCG produite s'analysent au pire en temps $\mathcal{O}(n^6)$. De plus, des formes restreintes de TAG peuvent produire des RCG dont les phrases s'analysent avec de meilleures complexités dans le pire des cas. Ainsi les grammaires à insertion d'arbre (TIG) de Schabes et Waters ou la forme proposée par Satta et Schuler donnent respectivement des complexités en temps cubique ou en $\mathcal{O}(n^5)$. Ces travaux sont décrits en [24] et en [17]. Ce dernier article a d'ailleurs été distingué comme l'un des trois meilleurs de la $6^{\text{ème}}$ conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'99).

Le formalisme MCS le plus puissant est actuellement celui des TAG ensemblistes où l'adjonction simultanée d'arbres est localisée aux ensembles d'arbres auxiliaires (MC-TAG). Les MC-TAG partagent cette première place avec deux autres formalismes qui lui sont équivalents, les systèmes de réécriture non-contextuels linéaires (LCFRS) et les grammaires non-contextuelles multiples (M-CFG). En fait une TAG est une MC-TAG dont les ensembles d'arbres élémentaires sont tous des singletons. Nous savions déjà que les LCFRS et les M-CFG pouvaient se traduire directement en des RCG équivalentes. Cette année, nous avons montré que c'était également le cas pour les MC-TAG. L'algorithme qui produit une RCG équivalente a été présenté en [16]. C'est une généralisation de celui proposé pour les TAG. Cependant, la phase d'optimisation, qui, pour une TAG, permet d'obtenir un temps d'analyse indépendant de la taille et de la forme des arbres élémentaires, ne se généralise pas aux MC-TAG. Cette dernière propriété explique que, dans le cas général, le degré du polynôme qui exprime l'ordre de grandeur du temps d'analyse dépende à la fois du nombre de nœuds par arbre et du nombre d'arbres par ensemble.

5.1.3 Propriétés formelles des RCG

Dans ce dernier axe, nous avons étudié les propriétés formelles de la sous-classe des RCG qui ne contient que des prédicats unaires, les 1-RCG qui définissent des langages appelés 1-RCL. Cette sous-classe, présentée en [15] et en [23], possède à la fois des propriétés formelles et des possibilités de description linguistique intéressantes. Nous avons tout d'abord montré que l'analyse de textes en 1-RCG s'effectue au pire en temps cubique. Comme les RCL, les 1-RCL sont clos à la fois par union, concaténation, fermeture de Kleene, intersection et complémentation. Toutes ces propriétés de clôture, et en particulier les deux dernières, sont obtenues sans modifier les grammaires composantes. Par exemple, considérons deux 1-RCG G_1 et G_2 décrivant les 1-RCL L_1 et L_2, la 1-RCG G décrivant le 1-RCL $L = L_1 \cap L_2$ s'obtient sans toucher

ni à G_1 ni à G_2 (on se contente d'ajouter une clause qui spécifie cette intersection). Puisque G_1 et G_2 peuvent être réutilisées telles quelles, elles peuvent constituer des bibliothèques de composants grammaticaux réutilisables. La théorie des langages formels nous a appris que l'intersection et la complémentation de langages non-contextuels étaient contextuels. Nous avons montré que l'intersection de deux CFL est un 1-RCL et que le complémentaire d'un CFL est un 1-RCL, langages dont les phrases peuvent bien entendu s'analyser au pire en temps cubique. Nous avons également montré que les 1-RCG sont presque MCS ; en fait ils sont un peu trop puissants, au sens où ils peuvent définir des langages qui n'ont pas la propriété de croissance constante. Cependant, cette propriété, nécessaire pour être un formalisme MCS, ne fait pas l'unanimité parmi les spécialistes et certains proposent de l'amender. Dans ce cas, les 1-RCG deviendraient le formalisme MCS le plus efficace.

5.2 DyALog: Automates à piles et Programmation dynamique

Participants : Éric Villemonte de la Clergerie, Djame Seddah.

Mots clés : tabulation, linguistique, programmation en logique, programmation dynamique, automate à pile.

Glossaire :
TAG *Tree Adjoining Grammars*
LIG *Linear Indexed Grammars*
TFS *Typed Feature Structures*
2SA *2-stack automata*
LIA *Linear Indexed Automata*
EPDA *Embedded Push-Down Automtata*

Résumé : *Le développement du système* DyALog *se poursuit et valide l'approche globale par automates (section 3.1.3). Outre un nouveau compilateur et une meilleure gestion des structures typées de traits, ce système autorise depuis peu le traitement des grammaires d'arbres adjoints (TAG). Ce résultat s'appuie sur des travaux théoriques en cours.*

Automates et Programmation Dynamique En 1998, en collaboration avec M. Alonso Pardo, nous avons introduit un formalisme d'automates à 2 piles (2SA) [3] permettant de décrire les pas de calculs de diverses stratégies d'analyse pour les grammaires d'arbres adjoints (TAG) et les grammaires linéaires indexées (LIG), deux formalismes grammaticaux faiblement dépendants du contexte. Une interprétation par Programmation Dynamique de ces automates était également spécifiée permettant une évaluation tabulaire de ces automates avec des complexités dans le pire des cas en $O(n^6)$ pour le temps et en $O(n^5)$ pour la place, où n dénote la longueur de la chaîne analysée.

Cette année, ces résultats ont été complétés par différents articles [20, 19, 27] où nous explorons différentes variantes d'automates, d'interprétations par programmation dynamique et de stratégies d'analyse. En particulier, nous avons étudié les «linear indexed automata» (LIA), forme particulière d'automates à piles enchâssés (EPDA). Ces automates sont une alternative

aux 2SA mais des équivalences sont en fait possibles entre les différents formalismes. Nous avons aussi examiné les spécialisations possibles des interprétations par programmation dynamique pour des stratégies d'analyse purement descendantes ou ascendantes.

D'autre part, une synthèse de différents travaux sur la tabulation (dont les nôtres) a été réalisée par É. de la Clergerie à l'occasion du tutoriel «Tabulation et traitement de la langue» [5] présenté lors de la 6ème conférence sur le *Traitement Automatique des Langues Naturelles* (TALN'99).

Le système DyALog En parallèle aux travaux théoriques, le développement du système DyALog s'est poursuivi. En premier lieu, un nouveau compilateur pour DyALog a été écrit en DyALog et «bootstrapé». Il remplace un ancien compilateur écrit en Scheme et produit maintenant du code en pseudo assembleur (plutôt que du code C comme précédemment). Ce code est ensuite traduit en assembleur et est lié à la librairie implantant la machine abstraite de DyALog. Ce travail prouve d'une part une certaine maturité de DyALog qui lui permet de se compiler lui-même en un temps raisonnable. D'autre part, le nouveau compilateur offre de bien meilleures possibilités d'extension que le compilateur précédent.

En particulier, lors de son stage de DEA, D. Seddah a étendu le compilateur DyALog pour traiter les grammaires d'arbres adjoints (avec attributs logiques) [30, 28]. Il n'a pas été nécessaire de modifier la machine abstraite pour ce travail. Les résultats préliminaires montrent un comportement correct vis à vis des complexités attendues.

Les réalisations du nouveau compilateur et de son extension pour les TAG ont permis de mieux dégager l'architecture d'un système comme DyALog et les points d'évolution possibles. Le modèle actuel contient 3 niveaux principaux:

1. Une phase de compilation allant des grammaires vers les automates permettant de se focaliser sur différentes stratégies d'analyse.

2. Une phase de compilation allant des automates vers le code et s'appuyant sur la déclaration d'une interprétation par programmation dynamique.

3. Une machine abstraite gérant les aspects bas niveau liés à la tabulation (gestion de la table et de l'agenda, partage de structure, indexation de la table, ...)

Chacun de ces niveaux est largement indépendant et évolue à son rythme. L'utilisation d'un formalisme logique pour les phases de compilation permet une approche déclarative facilitant ces évolutions.

En marge de ces développements majeurs, nous avons aussi réécrit l'outil de compilation des hiérarchies de structures typées de traits (TFS) [Car92,San98] pour une meilleure intégration avec le nouveau compilateur DyALog. Ces structures typées de traits sont employées dans de nombreux formalismes grammaticaux.

[Car92] B. Carpenter, *The Logic of Typed Feature Structures with Applications to Unification Grammars, Logic Programs and Constraint Resolution*, ISBN 0-521-41932, Cambridge University Press, 1992.

[San98] F. Sanches, «Étude et implantation modulaire d'algorithmes d'analyse syntaxique pour des grammaires utilisées en langue naturelle (grammaires d'arbres adjoints ou grammaires lexicales fonctionnelles», Mémoire d'Ingénieur CNAM, 1998.

Enfin, nous avons aussi cherché à faciliter la distribution et l'installation de DyALog, en particulier par la création d'un «distribution RPM» Linux (Section 4.2).

5.3 Grammaire à Interpolation de Graphe

Participant : Jean-Marie Larchevêque.

Glossaire :
GIL *Graph Interpolation Language*
TAL *Tree Adjoining Language*

La Grammaire à Interpolation de Graphes est un formalisme grammatical doté d'une sémantique opérationnelle qui permet d'analyser les phrases mot par mot, c'est-à-dire en construisant incrémentalement une représentation syntaxique à mesure que chaque mot est lu.

Les travaux en cours, présentés lors d'un exposé au XRCE à Grenoble, prouvent que la capacité générative faible des Grammaires à Interpolation de Graphes est au moins égale à celle des Grammaires d'Adjonction d'Arbres.

5.4 Thesaurus

Participants : Éric Villemonte de la Clergerie, Roland Dachelet.

Dans le cadre de l'étude de faisabilité européenne TermIT (Section 6.1), É. de la Clergerie a poursuivi son tour d'horizon des problèmes d'acquisition de connaissances à partir d'entrées de thesaurus dans le but d'identifier (semi-automatiquement) les entrées similaires dans différents thesaurus. Ce travail a donné lieu à un rapport [29] qui s'appuie en partie sur des indications expérimentales découlant de l'examen des 3 thesauri (AAT, RCHME et MERIMEE) à notre disposition. Cet examen a été rendu possible par la réalisation préalable d'une interface de navigation et de recherche dans ces thesauri.

5.5 Bibliothèques électroniques

Participant : François Role.

Mots clés : métadonnée, bibliothèque électronique.

Glossaire :
TEI *Text Encoding Initiative*
DOM *Document Object Model*

Poursuivant son travail de thèse, F. Role essaie de mettre en évidence l'apport des *métadonnées* et de la *documentation structurée* dans la conception d'un environnement de travail pour l'étude des documents présentant un intérêt tant du point de vue de leur contenu que de leur aspect graphique [AFR95].

Concernant la notion de métadonnée, F. Role a fait une synthèse des recherches en cours dans ce domaine, travail qui a donné lieu à la publication d'un rapport de recherche INRIA

[AFR95] J. André, J.-D. Fekete, H. Richy, « Traitement mixte image/texte de documents anciens », *Cahiers GUTenberg 21*, 1995, p. 75–85.

[25]. Il a également publié dans la revue *Document numérique* un article sur la représentation formelle de métadonnées complexes [13].

Concernant la documentation structurée, F. Role a évalué certaines techniques proposées pour manipuler des documents structurés, notamment le *Document Object Model* (DOM) et a publié un article comparant les différentes implantations du DOM à l'occasion du congrès GUT'99 [21]. Dans le même domaine, il a aussi publié un article sur la *Text Encoding Initiative* (TEI) et ses perspectives d'évolution [12].

5.6 Logiciels libres

Participant : Bernard Lang.

Mots clés : Logiciel libre, Linux.

L'évolution du marché et de la disponibilité des ressources logicielles et linguistiques (dictionnaires, grammaires, corpus) nous a amené à nous intéresser au développement des ressources libres.[5] Ce nouveau modèle de production et de distribution des biens immatériels a émergé cette année comme une composante majeure de l'évolution économique et politique, autant que technique, des technologies de l'information, ce qui justifie le travail que nous lui avons consacré depuis un peu plus de deux ans.

Les logiciels libres, et plus généralement les ressources libres, sont des ressources immatérielles qui, ayant été produites, sont mises à la disposition du public avec tous les moyens techniques et autorisations juridiques de les utiliser, de les faire évoluer, et de les rediffuser. Il s'agit donc d'un modèle de création très similaire à celui de la recherche scientifique traditionnelle, mais s'appliquant aussi à des ressources pouvant être directement utilisables, par des spécialistes, par des entreprises, ou par le grand public.

Nos travaux dans ce domaine ne portent pas spécifiquement sur les aspects linguistiques, mais plus généralement sur une analyse de l'intérêt et de l'impact des logiciels libres sur l'économie, sur le fonctionnement de la recherche, et sur la stratégie des entreprises. Il s'agit d'un travail de défrichage d'un modèle nouveau de production et qui comporte notamment des volets économiques et juridiques [11, 10, 6]. Notre travail a demandé une forte composante d'activité de terrain, permettant notamment une bonne communication avec les entreprises.

Notre réflexion [11] sur l'intérêt des logiciels libres pour la maîtrise des standards (et donc de certains marchés), ainsi que sur leur intérêt compétitif dans le cas des systèmes embarqués est en voie de confirmation par les décisions stratégiques de plusieurs grandes entreprises françaises.

5.7 Maintenance de traduction

Participants : Saad Berrada, Bernard Lang.

Mots clés : traduction, maintenance.

5. Notre attention fut initialement attirée sur ce sujet par la mise en oeuvre du système d'exploitation libre Linux. Des discussions avec plusieurs collègues nous ont amené à voir ce problème sous l'angle de la disponibilité des ressources scientifiques.

Notre implication dans l'analyse du phénomène des logiciels libres et de leur mode de production nous a amené à nous poser le problème de la maintenance de la documentation de ces logiciels. L'intérêt de ce problème vient de ce que cette documentation est maintenue simultanément dans des langues multiples, sans qu'il y ait nécessairement une langue particulière qui serve de référence. En outre le travail de (maintenance de) traduction est généralement fait par des bénévoles, compétents sur les techniques concernées, mais peu préparés à utiliser des outils d'assistance demandant une forte technicité linguistique, et peu à même d'en payer le prix commercial.

Dans le cadre d'un stage, nous avons debroussaillé des solutions possibles et construit un prototype d'outil [26]. La méthode envisagée s'appuie sur les techniques de mémoire de traduction, couramment utilisées dans les produits professionnels. Cette méthode s'impose dans notre cas, dans la mesure où il s'agit surtout d'assurer une maintenance, et donc de revenir souvent sur un travail déjà fait. Pour préserver une simplicité technique, nous avons limité la recherche des alignements de corpus bilingues à maintenir en n'utilisant que des indices simples liées à la mise en page et au découpage en phrases. Cette hypothèse est raisonnable dans la mesure où nous avons affaire à des documents en SGML dont la structure est explicite, et que l'utilisation en maintenance implique un fort taux de réutilisation de larges fragments déjà traduits antérieurement.

6 Contrats industriels (nationaux, européens et internationaux)

6.1 Projet TermIT

Participants : Éric Villemonte de la Clergerie, Roland Dachelet.

TermIT `http://www.mda.org.uk/term-it/` (LE4-8356) est une étude de faisabilité financée par la Communauté Européenne concernant le «développement de méthodes et d'outils pour la production, dissémination et exploitation de ressources terminologiques multilingues dans le domaine culturel»

Outre l'INRIA représenté par le projet Atoll, les partenaires étaient MDA, Forth (*Foundation for Research and Technology*), ILSP (*Institute for Language and Speech Processing*), le ministère de la culture et de la communication et SSL (*System Simulation Ltd*).

Cette étude s'est terminée cette année et a donné lieu, de notre côté, à la rédaction d'un rapport [29] concernant la tâche 3.3.2 «Multilingual Terminology Production Through an Intermediate Knowledge Level: Knowledge Acquisition Methods and Techniques». Ce rapport doit prendre place dans un document plus large. Concrètement, l'objectif de cette tâche était d'étudier les possibilités d'appariements d'entrées entre thesauri monolingues en se fondant sur la similarité de ces entrées au niveau conceptuel.

6.2 Action Xerox

Participants : Pierre Boullier, Éric Villemonte de la Clergerie.

Cette action est actuellement en suspens suite à des changements de thèmes de recherche de la part de notre correspondant Marc Dymetman.

7 Actions régionales, nationales et internationales

7.1 Actions nationales

Ph. Deschamp est membre de la Commission spécialisée de terminologie de l'informatique et des composants électroniques, et diffuse sur la toile le glossaire résultant de ses travaux.

Ph. Deschamp a également fait partie de la Commission de normalisation de l'AFNOR CGTI/CN 1 «Vocabulaire» jusqu'à sa clôture le 27 avril 1999 pour des raisons financières.

B. Lang est secrétaire de l'AFUL (`http://www.aful.org`), Association Francophone des Utilisateurs de Linux et des Logiciels Libre, et membre du conseil d'administration de l'ISoc-France (`http://www.isoc.asso.fr`), chapitre français de l'Internet Society.

7.1.1 Logiciels Libres

B. Lang a présenté les logiciels libres dans des séminaires, tables-rondes et conférences organisés par plusieurs entreprises, collectivités locales et administrations, dont : la région Poitou-Charentes, Aérospatiale CIMPA, la technopole Rennes Atalante, Exoffice Technologies, le Centre National de Documentation Pédagogique, Sycomore, l'Université de la Communication Hourtin, les rencontres de l'Orme (CRDP, Marseille), Thomson-CSF/LCR et First Internationale CET Forum de Thomson-CSF, Matra Datavision, Centre de Formation à l'Informatique pour les Personnels de L'Education Nationale (CFIPEN - Créteil), Saint-Gobain Recherche, France-Télécom (infocom 99, Clermont-Ferrand).

Il est également intervenu en tant que conseil sur le développement des logiciels libres, notamment pour la région Poitou-Charentes et les sociétés Matra Datavision et X*** (confidentiel).

7.2 Réseaux et groupes de travail internationaux

En raison de son changement d'activité, Ph. Deschamp a quitté fin 1998 l'Academic Advisory Council de la société Sun Corp.

B. Lang est membre du groupe d'experts sur le logiciel libre réuni par la DG 13 de la Commission Européenne. (`http://eu.conecta.it/`).

Il a également participé à la réalisation du CDROM « Internet au sud» produit par l'UNITAR (Institut des Nations Unies pour la Formation et la Recherche) et l'IRD (Institut de Recherche pour le Développement).

7.2.1 Réseau franco-portugais de formation par la recherche

V. Rocio de l'Université Nouvelle de Lisbonne et É. de la Clergerie ont poursuivi leur collaboration portant sur l'emploi de DyALog pour la réalisation d'un analyseur syntaxique robuste pour le Portugais. Cet analyseur comprend plusieurs couches, dont deux exploitent DyALog, et s'appuie en particulier sur le formalisme des *grammaires à mouvements restreints* (BMG). Un article «Tabulation for multi-purpose partial parsing» est en cours de rédaction relatant cette expérience.

7.3 Visites, et invitations de chercheurs

A la demande de F. Barthélemy, nous avons accueilli Mlle Dilek Dustegor et Mr Galip Tartanoglu, deux étudiants de l'Université francophone de Galatasaray (Turquie). L'objectif de ces stages était de familiariser ces étudiants avec le domaine de la Linguistique Informatique en général et de nos travaux en particulier. Galip Tartanoglu a ainsi réalisé un générateur jouet en Prolog pour les formes fléchies des verbes français du premier groupe.

8 Diffusion de résultats

8.1 Animation de la communauté scientifique

8.2 Encadrement

É. de la Clergerie a encadré le stage de DEA de D. Seddah, portant sur l'intégration dans le système DyALog d'un compilateur d'analyseurs pour les grammaires d'arbres adjoints [TAG].

É. de la Clergerie suit le travail de thèse de M. Alonso Pardo de l'Université de la Corogne (Espagne).

B. Lang suit le travail de thèse de F. Role et a encadré le stage d'ingénieur ENSIAS de S. Berrada [26].

8.3 Jury

É. de la Clergerie a participé à la Commission Mixte d'Évaluation en Informatique de l'Université d'Orléans.

B. Lang est membre de la commission de spécialiste du CNAM pour les enseignements d'informatique.

B. Lang était membre du jury de doctorat de Patrice Lopez, pour sa thèse intitulée «Analyse d'énoncés oraux pour le dialogue homme-machine à l'aide de grammaires lexicalisées d'arbres», en octobre à l'Université H. Poincaré (Nancy 1).

8.4 Enseignement

Enseignement universitaire. É. de la Clergerie est intervenu dans l'option «Langage Naturel» du DEA d'Informatique de l'Université d'Orléans.

B. Lang a participé en juillet à l'université d'été « La contribution des logiciels et ressources libres à l'amélioration de l'environnement de travail des enseignants et des élèves sur les réseaux », formation pour les enseignants du secondaire organisée à l'Université Pierre Mendès France de Grenoble, et à l'université d'été francophone « Développement durable et systèmes d'information » à l'École Nationale Supérieure des Mines de Saint-Etienne. Il a également contribué à une formation des moniteurs du CIES Versailles en avril.

8.5 Comités de programme

B. Lang est membre du comité de programme de IWPT 2000, Sixth International Workshop on Parsing Technologies, février 2000, Trento, Italie.

Il est également membre des comités de programme de plusieurs manifestations professionnelles : les salons Interop 1999 et Interop 2000, ainsi que le salon Linux-Expo 2000.

8.6 Participation à des colloques, séminaires, invitations

É. de la Clergerie et R. Dachelet ont participé aux différentes réunions du projet TermIT.

É. de la Clergerie a présenté ses travaux à l'Université de Namur. Il a également animé un tutoriel de 6 heures sur le thème «*Tabulation et traitement de la langue*» lors de la 6ème conférence sur le *traitement Automatique des Langues Naturelles* (TALN'99) à Cargèse. Les notes et support de cours de ce tutoriel sont disponibles en ligne [5].

F. Role a animé un séminaire de recherche sur les documents structurés à l'Université de Paris 8. Il a également effectué une intervention sur les techniques XML lors du 2ème forum XML qui s'est tenu à Paris en novembre. F. Role participe aux réunions du projet Philectre, projet financé par le GIS "cognition" du CNRS et auquel participent des équipes de l'IRISA, de l'ENST et de l'École des Mines de Nantes.

B. Lang a contribué à de nombreux colloques ou salons portant sur l'utilisation des logiciels libres et leur rôle économique :

- Table ronde « Internet et logiciels libres Quel sens technique et économique? », Les troisièmes rencontres Isoc-France d'Autrans (janvier);
- « L'intérêt du libre pour l'entreprise », Journées-rencontres Autour du Libre, ENST Bretagne (Brest, janvier);
- Table ronde « La Fronde du Logiciel Libre », École de Paris (février);
- « Logiciels Libres et Strategie d'entreprise», séminaire à l'invitation du Prof. Pierre-Yves Schobbens, Institut d'Informatique de l'Université de Namur (mai);
- « Transition », Keynote présentée à Linux-Expo (Paris, juin);
- Séminaire sur la Société de l'Information au Sud, Université de la Communication d'Hourtin (août);
- Table-ronde « Linux » au salon Interop 99 (Paris, septembre);
- Table-ronde « Linux, Apache Samba, Quelle place pour le LL dans le système d'information de l'entreprise », Club de la presse Informatique et des Télécom (Paris, octobre)
- « Open Source Software », ERCIM 10th Anniversary (Amsterdam, novembre);
- Table-ronde « Free Software - Behind the scene», Information Society Technology Conference 1999 (IST 99), à l'invitation de la Commission Européenne (Helsinki, novembre);

- Animateur de la table ronde « Les nouveaux business models du libre et l'émergence de la net economy », Colloque Logiciels et Contenus Libres : Un défi pour l'Europe, Institut National des Télécommunications (Evry, décembre);
- et divers exposés à l'invitation d'associations.

B. Lang est intervenu dans plusieurs manifestations concernant la propriété intellectuelle, notamment en ce qui concerne le développement des logiciels ou l'édition scientifique :

- « Les auteurs: leur place dans le développement de l'art et des industries numériques », table ronde organisée à la Cité des Sciences et de l'Industrie (Paris) sous l'égide du Prix Möbius International (mars);
- URANUS : Université d'été pour la recherche documentaire appliquée aux sciences humaines et sociales, Université Pierre Mendès France de Grenoble et Institut d'Etudes Politiques de Grenoble (juillet);
- « The influence of intellectual property on world economic development », conférence internationale organisée à Monte-Carlo (septembre);
- « Logiciels: Les Apports de la Propriété Industrielle », forum de l'Agence pour la Protection des Programmes, Université Pierre Mendès France de Grenoble (novembre);
- Réunion sur la brevetabilité des logiciels, Secrétariat d'état à l'Industrie (octobre).

É. de la Clergerie, P. Boullier et Ph. Deschamp ont participé aux réunions du groupe de travail **TAG/XML** `http://www.loria.fr/~lopez/TAG_XML/` concernant l'utilisation des Grammaires d'Arbres Adjoints en général et l'emploi de représentations XML de ces grammaires en particulier.

É. de la Clergerie a participé aux réunions du groupe de travail **A3CTE** `http://www-lipn.univ-paris13.fr/groupes-de-travail/A3CTE/` qui cherche à faire émerger des applications combinant Apprentissage et Traitement Automatique des Langues Naturelles.

9 Bibliographie

Ouvrages et articles de référence de l'équipe

[1] B. Lang, « Complete Evaluation of Horn Clauses: an Automata Theoretic Approach », *rapport de recherche n° 913*, INRIA, Rocquencourt, France, novembre 1988.

[2] B. Lang, « Towards a Uniform Formal Framework for Parsing », *in : Current issues in Parsing Technology*, M. Tomita (éditeur), Kluwer Academic Publishers, 1991, ch. 11, also appear in the Proc. of Int. Workshop on Parsing Technologies – IWPT89.

[3] Éric Villemonte de la Clergerie, M. A. Alonso Pardo, « A tabular interpretation of a class of 2-Stack Automata », *in : Proc. of ACL/COLING'98*, août 1998, `ftp://ftp.inria.fr/INRIA/Projects/Atoll/Eric.Clergerie/SD2SA.ps.gz`.

[4] Éric Villemonte de la Clergerie, *Automates à Piles et Programmation Dynamique. DyALog : Une application à la programmation en Logique*, thèse de doctorat, Université Paris 7, 1993.

Livres et monographies

[5] Éric Villemonte de la Clergerie, « Tabulation et traitement de la langue », ATALA, Cargèse, Corse, France, juillet 1999, Tutoriel présenté à la 6ème conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'99), `http://pauillac.inria.fr/~clerger/TALN99.html`.

Articles et chapitres de livre

[6] B. Lang, J.-P. Smets, « Un brevet pour tuer », *Libération*, 15 janvier 1999, `http://www.liberation.com/multi/cahier/articles/sem99.03/cah990115k.html`.

[7] B. Lang, « Préface », *in : Logiciels Libres - Liberté, Egalite, Business*, J.-P. Smets-Solanes et B. Faucon (éditeurs), Éditions Edispher, Paris, avril 1999, ISBN 2-911-968-7, `http://www.freepatents.org/liberty/`.

[8] B. Lang, « Ressources libres et indépendance technologique dans les secteurs de l'information», *Techniques et science informatique 18*, 8, octobre 1999, p. 901–914, `http://pauillac.inria.fr/~lang/ecrits/hanoi/`.

[9] B. Lang, «Internet libère les logiciels», *La Recherche*, février 2000, à paraître, `http://pauillac.inria.fr/~lang/ecrits/larecherche`.

[10] B. Lang, «Le nouveau protectionnisme est intellectuel», *in : Libres enfants du savoir numérique*, O. Blondeau et F. Latrive (éditeurs), Éditions de l'éclat, Perreux, mars 2000, A paraître, ISBN 2-84162-043-3, `http://pauillac.inria.fr/~lang/ecrits/latrive`.

[11] B. Lang, «Logiciels libres et entreprises», *Terminal*, 80/81, 2000, à paraître, `http://pauillac.inria.fr/~lang/ecrits/monaco`.

[12] F. Role, «La DTD TEI», *in : Techniques de l'ingénieur - traité informatique, H7 158*, TO BE FILLED, septembre 1999.

[13] F. Role, « Représentation et exploitation de métadonnées complexes », *Document numérique. Editions Hermes*, 1999, p. 75–85, A paraître en novembre 1999.

Communications à des congrès, colloques, etc.

[14] P. Boullier, «Chinese Numbers, MIX, Scrambling, and Range Concatenation Grammars», *in : Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL'99)*, p. 53–60, Bergen, Norway, juin 1999.

[15] P. Boullier, « A Cubic Time Extension of Context-Free Grammars », *in : Sixth Meeting on Mathematics of Language (MOL6)*, p. 37–50, University of Central Florida, Orlando, Florida, USA, juillet 1999.

[16] P. Boullier, «On Multicomponent TAG Parsing», *in : 6ème conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'99)*, p. 321–326, Cargèse, Corse, France, juillet 1999.

[17] P. Boullier, «On TAG Parsing», *in : 6ème conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'99)*, p. 75–84, Cargèse, Corse, France, juillet 1999.

[18] I. Debourges, G. Hains, S. Guilloré, . de la Clergerie, «Vers un analyseur syntaxique parallèle», *in : Proc. of the Fourth International Symposium on Economic Informatics*, p. 71–79, Bucharest, Romania, mai 1999, `ftp://ftp.inria.fr/INRIA/Projects/Atoll/Eric.Clergerie/mol.ps.gz`.

[19] M. V. Miguel A. Alonso Pardo, David Cabrero Souto, Éric Villemonte de la Clergerie, « Tabular Algorithms for TAG Parsing », *in : Proc. of EACL'99*, 1999, `ftp://ftp.inria.fr/INRIA/Projects/Atoll/Eric.Clergerie/eacl.ps.gz`.

[20] M. A. A. Pardo, D. C. Souto, Éric Villemonte de la Clergerie, « Tabulation of Automata for Tree Adjoining Languages», *in : Sixth Meeting on Mathematics of Language (MOL6)*, 1999, `ftp://ftp.inria.fr/INRIA/Projects/Atoll/Eric.Clergerie/mol.ps.gz`.

[21] F. Role, P. Verdret, «Le Document Object Model (DOM)», *in : Actes du congrès GUT'99*, M. G. ed. (éditeur), Lyon - INPL, mai 1999. à paraître en novembre 1999.

Rapports de recherche et publications internes

[22] P. Boullier, « Chinese Numbers, MIX, Scrambling, and Range Concatenation Grammars », *Research Report n° RR-3614*, INRIA, Rocquencourt, France, janvier 1999, 14 pages, `http://www.inria.fr/RRRT/RR-3614.html`.

[23] P. Boullier, « A Cubic Time Extension of Context-Free Grammars», *Research Report n° RR-3611*, INRIA, Rocquencourt, France, janvier 1999, 28 pages, `http://www.inria.fr/RRRT/RR-3611.html`.

[24] P. Boullier, « On TAG and Multicomponent TAG Parsing », *Research Report n° RR-3668*, INRIA, Rocquencourt, France, avril 1999, 19 pages, `http://www.inria.fr/RRRT/RR-3668.html`.

[25] F. Role, « Panorama des travaux en cours dans le domaine des métadonnées », *rapport de recherche n° 3628*, INRIA, INRIA Rocquencourt, février 1999, `http://www.inria.fr/RRRT/RR-3628.html`.

Divers

[26] S. Berrada, « Un outil d'aide à la maintenance de traductions », Mémoire d'ingénieur de l'ENSIAS, Université Mohammed V, Maroc, 1999.

[27] M. A. Pardo, J. G. na, M. Vilares, Éric Villemonte de la Clergerie, «New Tabular Algorithms for LIG Parsing», Soumis à IWPT00, novembre 1999.

[28] Éric Villemonte de la Clergerie, M. A. Pardo, D. Seddah, « Pratical Aspects in implementing a FTAG parser», novembre 1999.

[29] Éric Villemonte de la Clergerie, « Multilingual Terminology Production Through an Intermediate Knowledge Level: Knowledge Acquisition Methods and Techniques », Tâche 3.3.2 du Projet LE4-8356 Term–IT, devant être inclus dans le document D3.1, juin 1999.

[30] D. Seddah, *Intégration d'un compilateur de grammaire TAG au sein du système tabulaire DyALog*, Mémoire, Université Paris 7, septembre 1999.