INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Calligramme

Linear Logic, Proof Nets and Categorial Grammars

Nancy - Grand Est

THEME SYM

Activity Report

2007

Table of contents

1. Team

Head of project team
Philippe de Groote [DR INRIA]

Vice-head of project team
François Lamarche [DR INRIA]

Administrative assistant
Céline Simon [INRIA]

Permanent researchers
Bruno Guillaume [CR INRIA]
Sylvain Pogodalla [CR INRIA]

Faculty members
Guy Perrier [Professor, University Nancy 2, HdR]

Technical staff
Karën Fort [INRIA]

Ph. D. Students
Mathieu Morey [MESR fellow (since October 1st 2007), defense planned in 2010]
Robert Hein [INRIA CORDI fellow (since September 18th 2006), defense planned in 2009]
Joseph Le Roux [defense on October 17th 2007, Assistant teacher at the "IUT Charlemagne"]
Sarah Maarek [MESR fellow (since October 1st 2006), defense planned in 2009]
Jonathan Marchand [MESR fellow (since October 1st 2006), defense planned in 2009]
Novak Novakovic [INRIA fellow (since December 1st 2007), defense planned in 2010]
Ekaterina Lebedeva [INRIA CORDI fellow (since December 1st 2007), defense planned in 2010]

Internships
Ryo Yoshinaka [INRIA postdoctoral fellow (until September 30th 2007)]
Daniel de Carvalho [INRIA postdoctoral fellow (since October 1st 2007)]

2. Overall Objectives

2.1. Overall Objectives

Keywords: *categorial grammar*, *lambda calculus*, *linear logic*, *proof nets*, *semantics of natural languages*, *sequent calculus*, *syntactic analysis of natural languages*, *type theory*.

Project-team Calligramme's aim is the development of tools and methods that stem from proof theory, and in particular, linear logic, in the area of computational linguistics. Two fields of application are emphasized: the modelling of the syntax and semantics of natural languages.

3. Scientific Foundations

3.1. Introduction

Project-team Calligramme's research is conducted at the juncture of mathematical logic and computer science. The scientific domains that base our investigations are proof theory and the λ-calculus, more specifically linear logic. This latter theory, the brainchild of Jean-Yves Girard [25] results from a finer analysis of the part played by structural rules in Gentzen's sequent calculus [23]. These rules, traditionally considered as secondary, specify that the sequences of formulas that appear in sequents can be treated as (multi) sets. In the case of intuitionistic logic, there are three of them:

$$\frac{\Gamma \vdash C}{\Gamma, A \vdash C}\ \text{(Weakening)} \qquad \frac{\Gamma, A, A \vdash C}{\Gamma, A \vdash C}\ \text{(Contraction)} \qquad \frac{\Gamma, A, B, \Delta \vdash C}{\Gamma, B, A, \Delta \vdash C}\ \text{(Exchange)}$$

These rules have important logical weight: the weakening rule embodies the fact that some hypotheses may be dropped during a derivation; in a similar fashion the contraction rule specifies that any hypothesis can be used an unlimited number of times; as for the exchange rule, it stipulates that no order of priority holds between hypotheses. Thus, the presence of the structural rules in the ordinary sequent calculus strongly conditions the properties of the logic that results. For example, in the Gentzen-style formulations of classical or intuitionistic logic, the contraction rule by itself entails the undecidability of the predicate calculus. In the same manner, the use of the weakening and contraction rules in the right half of the sequent in classical logic is responsible for the latter's non-constructive aspects.

According to this analysis, linear logic can be understood as a system that conciliates the constructivist aspect of intuitionistic logic and the symmetry of classical logic. As in intuitionistic logic the constructive character comes from the banning of the weakening and contraction rules in the right part of the sequent. But simultaneously, in order to preserve symmetry in the system, the same rules are also rejected in the other half.

	Propositional linear logic			
	Rudimentary linear logic			
	Negation	Multiplicatives	Additives	Exponentials
Negation	$A^{\perp}$			
Conjunction		$A \otimes B$	$A \& B$	
Disjunction		$A \parr B$	$A \oplus B$	
Implication		$A \multimap B$		
Constants		$\mathbf{1}, \perp$	$\top, \mathbf{0}$	
Modalities				!A, ?A

The resulting system, called *rudimentary linear logic*, presents many interesting properties. It is endowed with four logical connectors (two conjunctions and two disjunctions) and the four constants that are their corresponding units. It is completely symmetrical, although constructive, and equipped with an involutive negation. As a consequence, rules similar to De Morgan's law hold in it.

In rudimentary linear logic, any hypothesis must be used once and only once during a derivation. This property, that allows linear logic to be considered as a resource calculus, is due, as we have seen, to the rejection of structural rules. But their total absence also implies that rudimentary linear logic is a much weaker system than intuitionistic or classical logic. Therefore, in order to restore its strength, it is necessary to augment the system with operators that recover the logical power of the weakening and contraction rules. This is done via two modalities that give tightly controlled access to the structural rules. Thus, linear logic does not question the usefulness of the structural rules, but instead, emphasizes their logical importance. In fact, it rejects them as epitheoretical rules [21] to incorporate them as logical rules that are embodied in new connectors. This original idea is what gives linear logic all its subtlety and power.

The finer decomposition that linear logic brings to traditional logic has another consequence: the Exchange rule, which so far has been left as is, is now in a quite different position, being the only one of the traditional structural rules that is left. A natural extension of Girard's original program is to investigate its meaning, in other words, to see what happens to the rest of the logic when Exchange is tampered with. Two standard algebraic laws are contained in it: commutativity and associativity. Relaxing these rules entails looking for non-commutative, and non-associative, variants of linear logic; there are now several examples of these. The natural outcome of this proliferation is a questioning of the nature of the structure that binds formulas together in a sequent: what is the natural general replacement of the notion of (multi) set, as applied to logic? Such questions are important for Calligramme and are addressed, for example, in [35].

The activities of project-team Calligramme are organized around three research actions:

- Proof nets, sequent calculus and typed λ-calculi.
- Grammatical formalisms.
- Implicit complexity of computations.

The first one of these is essentially theoretical, the other two, presenting both a theoretical and an applied character, are our privileged fields of application.

3.2. Proof Nets, Sequent Calculus and Typed Lambda Calculi

Keywords: *Curry-Howard isomorphism*, *denotational semantics*, *lambda calculus*, *proof nets*, *sequent calculus*, *type theory*.

The aim of this action is the development of the theoretical tools that we use in our other research actions. We are interested, in particular, in the notion of formal proof itself, as much from a syntactical point of view (sequential derivations, proof nets, λ-terms), as from a semantical point of view.

Proof nets are graphical representations (in the sense of graph theory) of proofs in linear logic. Their role is very similar to lambda terms for more traditional logics; as a matter of fact there are several back-and-forth translations that relate several classes of lambda terms with classes of proof nets. In addition to their strong geometric character, another difference between proof nets and lambda terms is that the proof net structure of a proof of formula T can be considered as a structure which is *added* to T, as a coupling between the atomic formula nodes of the usual syntactic tree graph of T. Since not all couplings correspond to proofs of T there is a need to distinguish the ones that do actually correspond to proofs; this is called a *correctness criterion*.

The discovery of new correctness criteria remains an important research problem, as much for Girard's original linear logic as for the field of non-commutative logics. Some criteria are better adapted to some applications than others. In particular, in the case of automatic proof search, correctness criteria can be used as invariants during the inductive process of proof construction.

The theory of proof nets also presents a dynamic character: cut elimination. This embodies a notion of normalization (or evaluation) akin to β-reduction in the λ-calculus.

As we said above, until the invention of proof nets, the principal tool for representing proofs in constructive logics was the λ-calculus. This is due to the Curry-Howard isomorphism, which establishes a correspondence between natural deduction systems for intuitionistic logics and typed λ-calculi.

Although the Curry-Howard isomorphism owes its existence to the functional character of intuitionistic logic, it can be extended to fragments of classical logic. It turns out that some constructions that one meets in functional progamming languages, such as control operators, can presently only be explained by the use of deduction rules that are related to proof by contradiction [26].

This extension of the Curry-Howard isomorphism to classical logic and its applications has a perennial place as research field in the project.

3.3. Categorial Grammars

Keywords: *Montague semantics*, *categorial grammar*, *semantics of natural languages*, *syntactic analysis of natural languages*, *syntactic inference*, *tree description*.

Lambek's syntactic calculus, which plays a central part in the theory of categorial grammars, can be seen a posteriori as a fragment of linear logic. As a matter of fact it introduces a mathematical framework that enables extensions of Lambek's original calculus as well as extensions of categorial grammars in general. The aim of this work is the development of a model, in the sense of computational linguistics, which is more flexible and efficient than the presently existing categorial models.

The relevance of linear logic for natural language processing is due to the notion of resource sensitivity. A language (natural or formal) can indeed be interpreted as a system of resources. For example a sentence like *The man that Mary saw Peter slept* is incorrect because it violates an underlying principle of natural languages, according to which verbal valencies must be realized once and only once. Categorial grammars formalize this idea by specifying that a verb such as *saw* is a resource which will give a sentence S in the presence of a nominal subject phrase, NP, and only one direct object NP. This gives rise to the following type assignment:

Mary, Peter:	NP
saw	$(NP \setminus S)/NP$

where the slash ($/$) (resp. the backslash ($\setminus$)) are interpreted as fraction pairings that simplify to the right (resp. to the left). However we notice very soon that this simplification scheme, which is the basis of Bar-Hillel grammars [19], is not sufficient.

Lambek solves this problem by suggesting the interpretation of slashes and backslashes as implicative connectors [28], [29]. Then not only do they obey the *modus ponens* law which turns out to be Bar-Hillel's simplification scheme

$$\frac{\Gamma \vdash A \quad \Delta \vdash A \setminus B}{\Gamma, \Delta \vdash B} \text{ (modus ponens)} \qquad \frac{\Gamma \vdash B/A \quad \Delta \vdash A}{\Gamma, \Delta \vdash B} \text{ (modus ponens)}$$

but also the introduction rules:

$$\frac{A, \Gamma \vdash B}{\Gamma \vdash A \setminus B} \setminus\text{-intro} \qquad \frac{\Gamma, A \vdash B}{\Gamma \vdash B/A} /\text{-intro}$$

The Lambek calculus does have its own limitations. Among other things it cannot treat syntactical phenomena like medial extraction and crossed dependencies. Thus the question arises: how can we extend the Lambek calculus to treat these and related problems? This is where linear logic comes into play, by offering an adequate mathematical framework for attacking this question. In particular proof nets appear as the best adapted approach to syntactical structure in the categorial framework.

Proof nets offer a geometrical interpretation of proof construction. Premises are represented by proof net fragments with inputs and outputs which respectively model needed and offered resources. These fragments must then be combined by pairing inputs and outputs according to their types. This process can also be interpreted in a model-theoretical fashion where fragments are regarded as descriptions for certain classes of models: the intuitionistic multiplicative fragment of linear logic can be interpreted on directed acyclic graphs, while for the implicative fragment, trees suffice [31].

This perspective shift from proof theory to model theory remains founded on the notion of resource sensitivity (e.g., in the form of polarities and their neutralization) but affords us the freedom to interpret these ideas in richer classes of models and leads to the formalism of Interaction Grammars. For example:

- Where previously we only considered simple categories with polarities, we can now consider complex categories with polarized features.
- We can also adopt more expressive tree description languages that allow us to speak about dominance and precedence relations between nodes. In this fashion, we espouse and generalize the monotonic version of Tree Adjoining Grammars (TAG) as proposed by Vijay-Shanker [33].
- Contrary to TAG where tree fragments can only be inserted, Interaction Grammars admit models where the interpretations of description fragments may overlap.

4. Application Domains

4.1. Modelling the Syntax and Semantics of Natural Languages

4.1.1. Abstract Categorial Grammars

Abstract Categorial Grammars (ACGs) are a new categorial formalism based on Girard's linear logic. This formalism, which sticks to the spirit of current type-logical grammars, offers the following features:

- Any ACG generates two languages, an abstract language and an object language. The abstract language may be thought as a set of abstract grammatical structures, and the object language as the sets of concrete forms generated from these abstract structures. Consequently, one has a direct control on the parse structures of the grammar.
- The languages generated by the ACGs are sets of linear λ-terms. This may be seen as a generalization of both string-languages and tree-languages.
- ACGs are based on a small set of mathematical primitives that combine via simple composition rules. Consequently, the ACG framework is rather flexible.

Abstract categorial grammars are not intended as yet another grammatical formalism that would compete with other established formalisms. It should rather be seen as the kernel of a grammatical framework in which other existing grammatical models may be encoded.

4.1.2. Interaction Grammars

Interaction Grammars (IGs) are a linguistic formalism that aims at modelling both the syntax and the semantics of natural languages according to the following principles:

- An IG is a monotonic system of constraints, as opposed to a derivational/transformational system, and this system is multidimensional: at the syntactic level, basic objects are tree descriptions and at the semantic level, basic objects are directed acyclic graph descriptions.
- The synchronization between the syntactic and the semantic levels is realized in a flexible way by a partial function that maps syntactic nodes to semantic nodes.
- Much in the spirit of Categorial Grammars, the resource sensitivity of natural language is built-in in the formalism: syntactic composition is driven by an operation of cancellation between polarized morpho-syntactic features and in parallel, semantic composition is driven by a similar operation of cancellation between polarized semantic features.

The formalism of IG stems from a reformulation of proof nets of Intuitionistic Linear Logic (which have very specific properties) in a model-theoretical framework [31] and it was at first designed for modelling the syntax of natural languages [32].

Hassen Ben Zineb and Bruno Guillaume have adapted Interaction Grammars to allow a syntax/semantics interface with the Montague's semantics. This interface is implementd in the parser Leopar. A small toy grammar (a subset of the whole french grammar) with semantic representation has been written to test this new interface.

Mathieu Morey and Guy Perrier have studied the connections between Dynamic Syntax and Interaction Grammars. A partial adaptation of the first one into the second one was done, exhibiting some fundamental differences between them.

4.1.3. Grammatical and lexical resources for French

The relevance of new linguistic formalisms needs to be proved by experiments on real corpora. Parsing real corpora requires large scale grammars and lexicons. There is a crucial lack of such resources for French and all researchers committed in natural language processing (NLP) projects for French based on different formalisms are confronted with the same problem. Now, building large scale grammars and lexicons for French demands a lot of time and human resources and it is crucial to overcome the multiplicity of existing formalisms by developing common and reusable tools and data. This is the sense of two directions of research:

1. The modular organization of formal grammars in a hierarchy of classes allows the expression of linguistic generalizations and it makes their development and their maintenance on a large scale possible. To be used in NLP applications such modular grammars have to be compiled into operational grammars. By comparison with the area of programming languages, we write source grammars in a language with a high abstraction level and then we compile them automatically to object grammars, directly usable by NLP applications.

   Considering the multiplicity of linguistic formalisms, it would be interesting to express the various source grammars that can be written in different formalisms, in a common abstract language and to compile them with the same tool associated to this language. XMG is a first experiment in this direction: for the moment, it allows the edition and the compilation of source grammars for TAGs and IGs. Moreover, we can hope that the use of a common language of syntactic description with a high level of abstraction makes easier the reusability of some parts of grammars from one formalism to another.
2. With the same preoccupation of reusability, it is important to develop syntactic and semantic lexicons which contain only purely linguistic information and which are independent of the different existing grammatical formalisms. Now, a mechanism must be foreseen to combine these lexicons with the grammars built in the various formalisms. A convenient way of doing this is to design the entries of such lexicons in the form of feature structures and to associate also feature structures with the elementary constructions of the grammars. Then, their anchoring in the lexicons is realized by unification of the two kinds of feature structures. The construction of a syntactic and a semantic lexicon for French can be envisaged either by acquisition from corpora or by re-use of existing lexical information.

5. Software

5.1. Leopar

Keywords: *Interaction Grammar*, *parsing*.

Participants: Bruno Guillaume [correspondant], Guy Perrier, Guillaume Bonfante [CARTE team], Sylvain Pogodalla, Joseph Le Roux, Jonathan Marchand, Hassen Ben Zineb [student].

5.1.1. Software description

LEOPAR is a parser for natural languages which is based on the formalism of Interaction Grammars (IG) [32]. It uses a parsing principle, called “electrostatic parsing” which consists in neutralizing opposite polarities. A positive polarity corresponds to an available linguistic feature and a negative one to an expected feature.

Parsing a sentence with an Interaction Grammar consists in first selecting a lexical entry for each of its words. A lexical entry is an underspecified syntactic tree, a tree description in other words. Then, all selected tree descriptions are combined by partial superposition guided by the aim of neutralizing polarities: two opposite polarities are neutralized by merging their support nodes. Parsing succeeds if the process ends with a minimal and neutral tree. As IGs are based on polarities and under-specified trees, LEOPAR uses some specific and non-trivial data-structures and algorithms.

The electrostatic principle has been intensively considered in LEOPAR. The theoretical problem of parsing IGs is NP-complete; the nondeterminism usually associated to NP-completeness is present at two levels: when a description for each word is selected from the lexicon, and when a choice of which nodes to merge is made. Polarities have shown their efficiency in pruning the search tree for the following two steps:

- In the first step (tagging the words of the sentence with tree descriptions), we forget the structure of descriptions, and only keep the bag of their features. In this case, parsing inside the formalism is greatly simplified because composition rules reduce to the neutralization of a negative feature-value pair $f \longleftarrow v$ by a dual positive feature-value pair $f \longrightarrow v$. As a consequence, parsing reduces to a counting of positive and negative polarities present in the selected tagging for every pair (f, v): every positive occurrence counts for +1 and every negative occurrence for –1, the sum must be 0.
- In the second step (node-merging phase), polarities are used to cut off parsing branches when their trees contain too many non neutral polarities.

5.1.2. Current state of the implementation

The current implementation started in 2004 (by Guillaume Bonfante, Bruno Guillaume, Guy Perrier and Sylvain Pogodalla).

This implementation (http://www.loria.fr/equipes/calligramme/leopar/) is a public project on the InriaGforge platform (http://gforge.inria.fr/projects/leopar/). It is freely available under the CECILL License (http://www.cecill.info). A release for a larger audience is planned for the beginning of 2008.

The main features of the current implementation are:

- automatic parsing of a sentence or a set of sentences,
- manual parsing (the user chooses the couple of nodes to merge),
- visualization of grammars produced by XMG or of set of description trees associated to some French word,
- a graphical interface (using GTK) which is useful for debugging grammars.

The main changes this year:

- new implementation of the automata library,
- integration of a syntax/semantic interface based on Montague's semantics,
- grammar of french coordination inspired from [7].

The current implementation comes with a middle-size coverage grammar for French (2200 tree descriptions in the grammar produced with XMG). It also includes hand-made morphological and syntactical lexicons that cover the French examples of the TSNLP (Test Suite for Natural Language Processing) [30].

5.2. XMG

Keywords: *metagrammar*.

Participants: Joseph Le Roux [correspondant], Yannick Parmentier [TALARIS team].

The eXtensible MetaGrammar (XMG) is a tool for generating large coverage grammars from concise descriptions of linguistic phenomenena (the so-called metagrammar). This software is a Calligramme and TALARIS joint work and was formerly known as The Metagrammar Workbench.

This software is based on two important concepts from logic programming, namely the Warren's Abstract Machine and constraints on finite sets. It has been developed by Benoît Crabbé, Yannick Parmentier, Denys Duchier and Joseph Le Roux. It is available at http://sourcesup.cru.fr/xmg. It is now maintained by Ph.D students Yannick Parmentier and Joseph Le Roux.

At current stage of implementation, XMG generates Tree Adjoining Grammars, Multi-Component Tree Adjoining Grammars and Interaction Grammars but the underlying formalism is generic so it could be extended to others grammars like dependency grammars or lexical functional grammars, depending on users' requests.

XMG is used in the research field (by Guy Perrier, Claire Gardent, Laura Kallmayer and Owen Rambow) to design lexicalized grammars for NLP parsers and in computational linguistics teaching.

In 2007, Calligramme and Talaris organized the XMG workshop (Nancy, June 21st–22nd). This workshop included both theoretical presentations and experiences with XMG. The overall aim was to define the next features to add, from the user point of view. The workshop also included a more general discussion about high level linguistic formalisms.

5.3. ACG support system

Keywords: *Abstract Categorial Grammars*.

Participants: Sylvain Pogodalla [correspondant], Philippe de Groote, Sarah Maarek.

The current ACG development toolkit is being rewritten. It aims at providing support for the planned extension of the ACG type system and to offer a more modular architecture and integrate the new proposed algorithms for parsing (in case of second order [27] and in the general case).

6. New Results

6.1. Proof Nets, Sequent Calculus and Typed Lambda Calculi

Keywords: *linear logic*, *proof nets*, *sequent calculus*.

Participant: François Lamarche.

6.1.1. Denotational Semantics of Classical Logic

François Lamarche begins the paper [6] by giving a full categorical axiomatization of the Medial rule, which is an essential component for the presentation of classical logic by the means of deep inference. It takes the form of additional structure on a *-autonomous category (so the symmetry of classical logic is kept), which can be expressed in the language of monoidal functors. He then shows abstract conditions that are necessary for a class of "bimonoids" in such a category to become a model of classical logic. The reason a model of this kind does not fall prey to "Joyal's paradox" (i.e., collapse to a poset) is that the full bimonoid structure is not preserved by every linear map between bimonoids.

Then he shows that such things actually exist in nature, by first exhibiting an example of Medial structure in a modified version of the category of coherence spaces and linear maps, made famous by J.-Y. Girard. The final step is the extraction of classes of bimonoids in that category that obey the necessary "intrinsicness" condition to become models of classical logic. Several such classes are exhibited. One of them is very natural, but it does not have the ability to count how many times a given axiom link is reused by means of contraction. The final model has this ability (up to a finite, but arbitrary number of reuses), but its construction is much more involved. One interesting aspect of these semantics is that they contain additive counterparts to conjunction and disjunction, which are equivalent to the traditional connectives from the point of view of provability, but not from the point of view of naming proofs.

6.2. Categorial Grammars

Keywords: *Abstract Categorial Grammars*, *Earley algorithm*, *Interaction Grammars*, *discourse dynamics*, *scope ambiguity*.

Participants: Philippe de Groote, Guy Perrier, Sylvain Pogodalla, Bruno Guillaume, Joseph Le Roux, Jonathan Marchand.

6.2.1. Abstract Categorial Grammars

2007 was the first year of the associate team LAMBDA & GRAMMARS between Calligramme and Makoto Kanazawa (NII, Tokyo). The new results about ACG are threefolds: about the study and the characterisation of ACG as a formal language, about the extension of the underlying type system, and the development of case studies in the modelling of linguistic phenomena.

6.2.1.1. Theoretical properties of the formalism

From a method developed by Makoto Kanazawa to compile second order ACG as Datalog queries [27], Philippe de Groote proposed a way to reduce the membership problem for any ACG (without order restriction) to a proof search problem in the multiplicative exponential fragment of linear logic (invited talk at the Colloquium in honour of Gérard Huet).

6.2.1.2. Formalism extension

Philippe de Groote and Sarah Maarek proposed an extension to the formalism based on the extension of the underlying type system [16]. This extension adds feature structure constructs as in unification grammars. The resulting formalism can encode any recursively enumerable language, as shown by Philippe de Groote, Sarah Maarek and Ryo Yoshinaka [17].

6.2.1.3. Modelling of linguistic phenomena

Sylvain Pogodalla showed how to model quantifier scope ambiguity using the composition properties of the ACG [14]. He used this construct to propose a way of expressing scope ambiguity in Tree Adjoining Grammars without using any underspecified semantic representation formalism [13].

6.2.2. Interaction Grammars

We developed original techniques for lexical selection –a major issue with strongly lexicalized formalisms– based on the notion of polarity which is the heart of Interaction Grammars. This appoach is twofold:

1. at the sentence level, we check the global neutrality of a selection, which is needed for correct parsing;
2. at the constituent level, we take advantage of the syntactical modelisation of some phenomena (e.g., coordination) to refine the global neutrality constraint by checking polarities before and after distinguished tree descriptions (e.g., tree description corresponding to *and*).

This method is based on multiple automata intersection. Although the problem of guessing the best order to perform this intersection has been proved to be NP-Complete [8], we worked on the implementation to design good heuristics.

6.3. Development of linguistic resources

Keywords: *French formal grammar*, *Gross' grammar lexicon*, *lexicon*, *subcategorisation*.

Participants: Guy Perrier, Bruno Guillaume, Kären Fort, Mathieu Morey.

6.3.1. Development of a syntactical lexicon for french preposition

PrepLex is a lexicon of French prepositions which provides all the syntactic information needed for parsing. It was built by comparing and merging several authoritative lexical sources. This lexicon also includes information about the prepositions or classes of prepositions that appear in French verb subcategorization frames. This resource has been developed as a first step in making current French preposition lexicons available for effective natural language processing. This work is published in [9], [10]

6.3.2. Development of a syntactical lexicon for french adverbs

Lefff (Lexique des Formes Fléchies du Français) is a large coverage lexicon for French that contains morphosyntactic and syntactic information. Lefff is designed for NLP (Natural Language Processing) and in particular for parsing. Most of the recent work on Lefff focused on verbs. However, we show here how we enriched Lefff using the freely available Lexique-Grammaire (lexicon grammar) tables for "-ment" adverbs, published by Molinier (Molinier & Levrier 2000). This implied both a linguistic and a modelling work in order to exploit their content in a NLP lexicon such as Lefff. We also made a brief evaluation of this lexicon. This work is published in [15].

6.3.3. Development of tools for human validation of syntactic lexicons

A precise syntactic lexicon should contain detailed information for predicative words. In order to ensure the high quality of such a resource, human validation is unavoidable. For this purpose, we propose a freely available Web-based framework, named Sylva. The main point of our framework is that it handles multiple validations and keeps track of the resource's history. The expert linguist task is made easier: (s)he has only to consider data for which there is a disagreement between validators.

6.3.4. Development of an Interaction Grammar for French

Guy Perrier has developed an interaction grammar for French using XMG [11]. The methodology is inspired by Benoit Crabbé, who has developed a large French TAG [20].

The source grammar is composed of 449 classes organized in an inheritance hierarchy with two operators of conjunction and disjunction. The leaves of the hierarchy describe elementary phenomena of the grammar. Conjunctions and disjunctions express two ways of representing complex phenomena: for instance, a particular diathesis for a verb can result from the conjunction of classes representing specific realizations of its arguments and the realization of a particular predicate argument structure can be expressed by the disjunction of the classes representing the different diatheses.

The compiled grammar is composed of 830 tree descriptions mainly covering the following phenomena of the French syntax:

- most subcategorisation frames for verbs, predicative adjectives and nouns,
- active, passive, middle and reflexive diatheses combined with personal and impersonal subject constructions,
- grammatical words and related syntactic constructions (clitics, personal, relative and interrogative pronouns, complementizers, prepositions, negations, auxiliary verbs, ...),
- some hard to model phenomena such as: pied-piping in relative and interrogative clauses, islands for wh-extraction, long distance dependencies related to negative expressions ("ne...aucun", "ne...personne"), past participle agreement in presence of the auxiliary "avoir", control of the subject for the infinitives...

The grammar was evaluated on the TSNLP [30] and Eurotra [22] test suites by Jennifer Planul. Previously, she extended the grammar to the causative constructions. She has also partially represented the difficult phenomenon of comparatives.

7. Other Grants and Activities

7.1. National Actions

7.1.1. Agence Nationale de la Recherche (ANR) Inval

Headed by Eric Goubault at the CEA, this three-year "programme blanc" action (started in November 2005) aims at the study and development of algebraic invariants of computation, inspired by traditional homology and homotopy in algebraic topology. Two meetings have been held in 2007, one in Paris on April 6th and one in Nancy on September 7th. The co-ordinator for the Loria site is François Lamarche and Yves Guiraud (Protheo) is also one of the original members.

Webpage: http://www.pps.jussieu.fr/~inval/

7.1.2. Agence Nationale de la Recherche (ANR) Infer

This three-year "programme blanc" project on the theoretical and applicative development of deep inference began in December 2006. Two Inria-Lorraine, Calligramme and Protheo, are involved in it, along with teams at INRIA-Futurs and the PPS lab (Université Paris VII). The head of the project is Lutz Straßburger (Parsifal, INRIA-Futurs), and the local co-ordinator is François Lamarche. The first meeting was held at École Polytechnique on June 21.

Webpage: http://www.lix.polytechnique.fr/~lutz/orgs/infer.html

7.1.3. LexSynt project

Calligramme is involved in the LexSynt project. Thirteen French-speaking research teams work on this project. It aims at developing a syntactic lexicon with large coverage for French. In order to be usable in various NLP applications, this lexicon is independent of any grammatical formalism.

The web page of the project is http://lexsynt.inria.fr.

7.1.4. Action de Recherche Concertée (ARC) Mosaïque

Calligramme is involved in the Mosaïque INRIA ARC. Nine French-speaking research teams work on this project. It aims at developing a high level description language for the syntax of natural languages and a software environment for building large scale grammars, especially French grammars.

The web page of the project is http://mosaique.labri.fr.

7.1.5. Agence Nationale de la Recherche (ANR) Prelude

Calligramme is involved in the ANR-Blanc action PRELUDEhttp://prevert.upmf-grenoble.fr/~alecomte/PRELUDE.htm. This action is starting and aims at giving a theory of pragmatics based on ludics [24] and continuations [34]. The partner teams are: Structures Formelles de la Langue[1](coordinator), Institut Mathématique de Luminy[2], the Signes INRIA project[3] and Calligramme.

7.2. International Actions

7.2.1. Associate Team

In 2007 started an associate team program with Calligramme and Makoto Kanazawa2007 (NII, Japan). This program is supported by the Associate Team Program of INRIA [4]. The first annual report is published at http://www.loria.fr/equipes/calligramme/FormulaireRenouv08.html

7.2.2. NWO

The Calligramme project belongs to the network entitled "A Global Network for Lambda Grammars and Abstract Categorial Grammars"[5]. The other partners of this network are: the NII, Japan, the Technion, Israel, and Tilburg University, Netherlands (coordinator of this network). It started in 2007 for a three years period. It is supported at 80% by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO, national agency for science of the Netherlands).

7.2.3. Van Gogh Action

The Calligramme team is part of a Van Gogh action (from the Hubert Curien program) with the Utrecht Institues of Linguistics OTS (Utrecht University) about "Dynamique du discours et continuations".

[1] UMR 7023, Paris 8, http://recherche.univ-paris8.fr/red_fich_equ.php?OrgaNum=48
[2] CNRS, http://iml.univ-mrs.fr/
[3] http://signes.labri.fr/
[4] http://www-direction.inria.fr/international/EQUIPES_ASSOCIEES/index.eng.html
[5] http://www.nwo.nl/projecten.nsf/pages/2300136194?opendocument

7.3. Visits and invitation of researchers

- Lutz Straßburger (Parsifal, Inria-Futurs) and François Lamarche met through many of the workshops mentioned elsewhere, and also conducted several personal visits.
- François Lamarche was invited to give a talk at the Plume seminar at ENS Lyon on March 15, on the denotational semantics of classical logic.
- Steve Awodey (Carnegie-Mellon) and his student Henrik Forsell visited François Lamarche in September 7th to September 10th.
- Philippe de Groote and Sylvain Pogodalla visited Makoto Kanazawa (NII, Tokyo) from January 21st to February 2nd.
- Philippe de Groote, Sylvain Pogodalla and Ryo Yoshinaka visited Makoto Kanazawa (NII, Tokyo) from July 7th to July 20th.
- Makoto Tatsuta (Nii, Japan) visited the Calligramme team from June 13th to June 24th.
- Makoto Kanazawa (NII, Japan) visited the Calligramme team from September 16th to September 22nd.
- Reinhard Muskens (Tilburg University, Netherlands)) visited the Calligramme team from September 16th to September 22nd.

8. Dissemination

8.1. Activism within the scientific community

- François Lamarche organized a one-day meeting for the ANR Inval project on September 7th.
- François Lamarche was the organizer of the 86th edition of the Peripatetic Seminar on Sheaves and Logic (PSSL86), which was held at Institut Élie Cartan on the weekend of September 8–9. This venerable institution, which began in the early Seventies, is the number one European colloquium in the field of category theory. Participants came from Britain, Germany, the Netherlands, Norway, Italy and France and 12 talks were given. Webpage: http://www.loria.fr/~lamarche/psslHomeEN.html
- Philippe de Groote is President of the INRIA-Lorraine Projects Committee, and a member of INRIA's evaluation board.
- Philippe de Groote is a member of the LORIA management board, and of the LORIA laboratory council.
- Philippe de Groote is an associate editor of the journal Higher-Order and Symbolic Computation. He belongs to the editorial board of the series Papers in Formal Linguistics and Logic (Bulzoni, Roma), and Cahiers du Centre de Logique (Academia-Bruylant, Louvain-la-Neuve).
- Sylvain Pogodalla was member of the program committee of the 10th conference on Mathematics of Language (MOL-10).
- Sylvain Pogodalla and Philippe de Groote organized the 4th Workshop on Lambda Calculus and Formal Grammars[6], September 18-19.
- Sylvain Pogodalla was member of the program and organising committee of the Colloquium on Honour of Alain Lecomte[7], November 2-3.
- Guy Perrier and Joseph Le Roux were members of the organising committee of the XMG Workshop (June 21st–22nd).
- Guy Perrier is a member of the editorial board of the journal "Traitement Automatique des Langues".

[6] http://www.loria.fr/equipes/calligramme/acg/workshops/lcfg-04
[7] http://www.loria.fr/~pogodall/AL/

- Guy Perrier is a member of the program committee of the 6th International Conference on Natural Language Processing GOTAL 2008, which will be held in Gothenburg (Sweden) August 25- 27th 2008.
- Guy Perrier was a member of the program committee of the workshop "formalismes syntaxiques de haut niveau", which was held in Toulouse (France) on June 8th 2007, conjointly with the conference TALN 2007. Guy Perrier is a member of the board of the "Département de Formation Doctorale de l'école Doctorale IAEM-Lorraine".

8.2. Teaching

- François Lamarche gave an informal course on category theory at the Loria during the spring, which totalled about 20 hours.
- Bruno Guillaume is teaching the course ”Grammatical formalisms” of the Nancy computer science master, specialization ”Traitement Automatique des Langues”.
- Philippe de Groote is teaching the course “Sémantique computationnelle” of the Nancy computer science master specialization “Traitement Automatique des Langues”.
- Philippe de Groote are teaching the course “Structures Informatiques et Logiques pour la Modélisation Linguistique” of the “Master Parisien de Recherche en Informatique” with Gérard Huet.
- Sylvain Pogodalla is teaching the course “Corpus linguistics and linguistics resources” of the Nancy computer science master, specialization “Traitement Automatique des Langues”.
- Guy Perrier is the head of the common specialisation "Traitement Automatique des Langues" of the computer science and cognitive science masters from Nancy.
- Guy Perrier is teaching the courses "tools and algorithms for NLP", "initiation to NLP" and "programming for NLP" in the common specialisation "Traitement Automatique des Langues" of the computer science and cognitive science masters from Nancy.

8.3. Academic Supervision

- François Lamarche is supervising the thesis work of Robert Hein (since October 2006) and Novak Novakovic (since December 2007).
- François Lamarche is supervising the postdoctoral research of Daniel de Carvalho, which began in October.
- Guy Perrier and Bruno Guillaume are co-supervising the thesis work of Jonathan Marchand.
- Guy Perrier supervised the thesis work of Joseph Le Roux.
- Bruno Guillaume is supervising the master thesis work of Hassen Ben Zineb.
- Philippe de Groote is supervising the thesis work of Sarah Maarek (started from September 2006).
- Philippe de Groote is co-advisor with Michael Moortgat (Utrecht University, Netherlands) of Anna Chernilovskaya (started from September 2007)
- Philippe de Groote is supervising the thesis work of Ekaterina Lebedeva (started from December 2007).
- Guy Perrier was supervised the master thesis of Mathieu Morey and now he is supervising his thesis work.
- Guy Perrier supervised a 2,5 month internship of Jennifer Planul in the framework of the Mosaïque ARC.

8.4. Thesis defenses

- Joseph Le Roux defended his thesis on October 17th 2007 (jury: F. Simonot-Lion, G. Perrier, A. Lecomte, A. Ranta, O. Rambow, D. Duchier)

8.5. Thesis juries

- François Lamarche was member of the jury for Daniel de Carvalho's thesis defense (IML, Marseille), on September 4.
- Guy was a member of the jury for Joseph Le Roux thesis defense on October 17th.

8.6. Participation to colloquia, seminars, invitations

- François Lamarche attended the 85th edition of the Peripatetic Seminar on Sheaves and Logic (PSSL85) in Nice on March 25th–26th, and gave a talk.
- François Lamarche attended the "Journées Girard", given in honor of Jean-Yves Girard, on September 11th–12th at Institut Henri Poincaré, Paris.
- Philippe de Groote, François Lamarche, Guy Perrier and Sylvain Pogodalla attended the "Colloque en l'honneur d'Alain Lecomte", given in Pauillac on November 2th and 3th. Philippe de Groote gave a talk.
- François Lamarche and Robert Hein visited Kai Brünnler and Richard McKinley at Bern on November 19th–22th.
- Philippe de Groote, François Lamarche and Sylvain Pogodalla visited Michael Moortgat at Utrecht on December 5th–7th.
- Kären Fort, Robert Hein, Sarah Maarek Jonathan Marchand and Ryo Yoshinaka attended ESSLLI summer school in Dublin, Ireland (August 6th–17th).
- Sarah Maarek attended the Types Summer School in Bertinoro, Italy (August 19th–31th).
- Robert Hein attended the HyLo worskhop on Hybrid Logics, held during the ESSLI summer school (Dublin, Ireland, August 6th–10th) .
- Kären Fort, Bruno Guillaume and Joseph Le Roux attended the ACL annual meeting in Prague, Czech Republic (June 23rd-30th).
- Kären Fort, Bruno Guillaume, Joseph Le Roux, Guy Perrier and Sylvain Pogodalla attended TALN 07 in Toulouse, France (June 5th-8th).
- Guy Perrier attended the International Conference RANLP 2007 in Borovets (Bulgaria) on September 27th-29th.

9. Bibliography

Major publications by the team in recent years

[1] G. PERRIER. *Interaction Grammars*, in "CoLing 2000, Sarrebrücken, Germany", International Committee on Computational Linguistics, August 2000, p. 600–606.

[2] G. PERRIER. *La sémantique dans les grammaires d'interaction*, in "Traitement Automatique des Langues", vol. 45, n^{o} 3, 2004, p. 123–144.

[3] P. DE GROOTE. *Towards abstract categorial grammars*, in "Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Toulouse, France", July 2001, p. 148–155.

[4] P. DE GROOTE, F. LAMARCHE. *Classical Non Associative Lambek Calculus*, in "Studia Logica", vol. 71, n^o 3, August 2002, p. 355–388.

Year Publications

Doctoral dissertations and Habilitation theses

[5] J. LE ROUX. *La coordination dans les grammaires d'interaction*, Ph. D. Thesis, Ecole Doctorale IAEM Lorraine; Institut National Polytechnique de Lorraine - INPL, 10 2007, http://tel.archives-ouvertes.fr/tel-00185248/en/.

Articles in refereed journals and book chapters

[6] F. LAMARCHE. *Exploring the Gap between Linear and Classical Logic*, in "Theory and Applications of Categories", vol. 18, 2007, p. 473–535 (63 pages), http://hal.inria.fr/inria-00113785/en/.

[7] J. LE ROUX, G. PERRIER. *Modélisation de la coordination dans les grammaires d'interaction*, in "Traitement Automatique des Langues (TAL)", 2007, http://hal.inria.fr/inria-00185237/en/.

Publications in Conferences and Workshops

[8] G. BONFANTE, J. LE ROUX. *Intersection Optimization is NP-Complete*, in "Sixth International Workshop on Finite-State Methods and Natural Language Processing - FSMNLP 2007, Postdam Allemagne", 2007, http://hal.inria.fr/inria-00185232/en/.

[9] K. FORT, B. GUILLAUME. *PrepLex : un lexique des prépositions du français pour l'analyse syntaxique*, in "Traitement Automatique des Langues Naturelles - TALN 2007, Toulouse France", 2007, http://hal.inria.fr/inria-00186761/en/.

[10] K. FORT, B. GUILLAUME. *PrepLex: a lexicon of French prepositions for parsing*, in "4th ACL-SIGSEM Workshop on Prepositions, Prague Tchèque, République", 2007, http://hal.inria.fr/inria-00186777/en/.

[11] G. PERRIER. *A French Interaction Grammar*, in "International Conference on Recent Advances in Natural Language Processing - RANLP 2007, Borovets Bulgarie", G. ANGELOVA, K. BONTCHEVA, R. MITKOV, N. NICOLOV, K. SIMOV (editors), INCOMA Ltd, Shoumen, Bulgaria, IPP & BAS & ACL-Bulgaria, 2007, p. 463-467, http://hal.inria.fr/inria-00184108/en/.

[12] G. PERRIER. *Une grammaire d'interaction du français*, in "Traitement Automatique des Langues Naturelles - TALN 2007, Toulouse France", N. HATHOUT, P. MÜLLER (editors), IRIT Press, CLLE-ERSS & IRIT, 2007, p. 453-462, http://hal.inria.fr/inria-00184102/en/.

[13] S. POGODALLA. *Ambiguïté de portée et approche fonctionnelle des TAG*, in "Traitement Automatique des Langues Naturelles - TALN 2007, Toulouse/France", 06 2007, http://hal.inria.fr/inria-00141913/en/.

[14] S. POGODALLA. *Generalizing a Proof-Theoretic Account of Scope Ambiguity*, in "7th International Workshop on Computational Semantics - IWCS-7, Tilburg/The Netherlands", 01 2007, http://hal.inria.fr/inria-00112898/en/.

[15] B. SAGOT, K. FORT. *Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : Adverbes en -ment*, in "26e Colloque International sur le Lexique et grammaire 2007, Bonifacio France", 2007, http://hal.inria.fr/inria-00186779/en/.

[16] P. DE GROOTE, S. MAAREK. *Type-theoretic Extensions of Abstract Categorial Grammars*, in "Proceedings of Workshop on New Directions in Type-theoretic Grammars", R. MUSKENS (editor), 2007, p. 19–30.

[17] P. DE GROOTE, R. YOSHINAKA, S. MAAREK. *On two Extensions of Abstract Categorial Grammars*, in "Proceedings of International Conference on Logic for Programming, Artificial Intelligence and Reasoning (LPAR)", 2007.

Miscellaneous

[18] G. BONFANTE, J. LE ROUX. *Intersection Optimization is NP Complete*, 05 2007, http://hal.archives-ouvertes.fr/hal-00146671/en/.

References in notes

[19] Y. BAR-HILLEL. *A quasi-arithmetical notation for syntactic description*, in "Language", vol. 29, 1950, p. 47-58.

[20] B. CRABBÉ. *Computational representation of strongly lexicalised syntactic formalisms: application to Tree Adjoining Grammars*, thèse de doctorat, University Nancy 2, 2005.

[21] H. CURRY. *Foundations of mathematical logic*, Dover Publications, 1977.

[22] L. DANLOS, O. LAURENS. *Présentation du projet EUROTRA et des grammaires d'Eurotra-France*, Technical report, CNRS, Université Paris 7, 1991.

[23] G. GENTZEN. *Recherches sur la déduction logique (Untersuchungen über das logische schließen)*, Traduction et commentaire par R. Feys et J. Ladrière, Presses Universitaires de France, 1955.

[24] J.-Y. GIRARD. *Locus Solum*, in "Mathematical Structures in Computer Science", vol. 11, 2001, p. 301–506.

[25] J.-Y. GIRARD. *Linear Logic*, in "Theoretical Computer Science", vol. 50, 1987, p. 1-102.

[26] T. G. GRIFFIN. *A formulae-as-types notion of control*, in "Conference record of the seventeenth annual ACM symposium on Principles of Programming Languages", 1990, p. 47-58.

[27] M. KANAZAWA. *Parsing and generation as Datalog queries*, in "Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics", 2007.

[28] J. LAMBEK. *The mathematics of sentence structure*, in "Amer. Math. Monthly", vol. 65, 1958, p. 154-170.

[29] J. LAMBEK. *On the calculus of syntactic types*, in "Studies of Language and its Mathematical Aspects, Providence", Proc. of the 12th Symp. Appl. Math., 1961, p. 166-178.

[30] S. LEHMANN, S. OEPEN, S. REGNIER-PROST, K. NETTER, V. LUX, J. KLEIN, K. FALKEDAL, F. FOUVRY, D. ESTIVAL, E. DAUPHIN, H. COMPAGNION, J. BAUR, L. BALKAN, D. ARNOLD. TSNLP — *Test Suites for Natural Language Processing*, in "Proceedings of COLING 1996, Kopenhagen", 1996.

[31] G. PERRIER. *Intuitionistic Multiplicative Proof Nets as Models of Directed Acyclic Graph Descriptions*, in "8th International Conference on Logic for Programming, Artificial Intelligence and Reasoning - LPAR 2001, Havana, Cuba", R. NIEUWENHUIS, A. VORONKOV (editors), Lecture Notes in Artificial Intelligence, vol. 2250, Springer, Dec 2001, p. 233-248.

[32] G. PERRIER. *Descriptions d'arbres avec polarités : les Grammaires d'Interaction*, in "9ième Conférence annuelle sur le Traitement Automatique des Langues Naturelles - TALN'02, Nancy, France", Jun 2002, http://www.loria.fr/publications/2002/A02-R-123/A02-R-123.ps.

[33] K. VIJAY-SHANKER. *Using Description of Trees in a Tree Adjoining Grammar*, in "Computational Linguistics", vol. 18, n° 4, 1992, p. 481-517.

[34] P. DE GROOTE. *Towards a Motagovian account of dynamics*, in "Proceedings of Semantics and Linguistic Theory XVI", CLC Publications, 2006.

[35] P. DE GROOTE, F. LAMARCHE. *Classical Non Associative Lambek Calculus*, in "Studia Logica", 2000.