INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team signes

Linguistic signs, grammar and meaning: computational logic for natural language

Bordeaux - Sud-Ouest

Theme : Audio, Speech, and Language Processing

Table of contents

1. Team

Research Scientists

Gérard Huet [INRIA, Research Director (DR) located in Rocquencourt, HdR]
Richard Moot [CNRS, Research Associate (CR)]
Sylvain Salvati [INRIA, Research Associate (CR)]

Faculty Members

Lionel Clément [University Bordeaux 1, Assistant Professor (MCF)]
Christian Retoré [University Bordeaux 1, Team Leader, Professor (Pr), HdR]

External Collaborators

Christian Bassac [University Lyon 2, Professor (Pr), HdR]
Mauro Gaio [University of Pau et des Pays de l'Adour, Professor (Pr), HdR]
Alain Lecomte [University Paris 8, Professor (Pr), HdR]

PhD Students

Pierre Bourreau [University Bordeaux 1, since Sep. 2007]
Ivano Ciardelli [INRIA, CORDI/S, since Sep 2010]
Noémie-Fleur Sandillon-Rezer [CNRS, since Nov 2010]
Anaïs Lefeuvre [Bourse Région Aquitaine - INRIA]
Bruno Mery [University Bordeaux 1, since Oct. 2006]
Natalia Vinogradova [INRIA, CORDI/S, since Dec. 2008]

Visiting Scientist

Michele Abrusci [Universita Roma3, Professor (Pr), HdR]

Administrative Assistants

Sylvie Embolla [INRIA, until Oct. 2010]
Anne-Laure Gautier [INRIA, since Oct. 2010]

2. Overall Objectives

2.1. Overall Objectives

The Signes project-team addresses several related domains of computational linguistics: :

- flexional and derivational morphology,
- syntax,
- logical semantics,
- lexical semantics,
- discourse representation.

The interfaces are of special interest to our group: between morphology and syntax, between syntax and semantics, between morphology and semantics, between semantics and discourse structure,...

There are various means to tackle these questions. In Signes, the methological accent is on the formal, symbolic methods issued from logic, which are also studied for themselves, like:

- formal language theory,
- categorial grammars,
- resource logics,
- lambda calculus,
- higher order logic.

We also develop the corresponding computational linguistics tools, which result in natural language processing (NLP) pieces of software for analysis, generation and acquisition devices. Some specific languages deserve particular attention, like Sanskrit, French Sign Language, Dutch and French:

- natural language tools for Sanskrit,
- modelling of French Sign Language grammar,
- large scale grammar for the NWO Dutch Spoken Corpus,
- lexicon and grammar for the analysis of written French.

3. Scientific Foundations

3.1. Word structure and automata for computational morphology

Participant: Gérard Huet [correspondent].

Computational models for phonology and morphology are a traditional application of finite state technology [45], [46], [47], [29]. These models often combine symbolic or logical systems, like rewriting systems, and statistical methods like probabilistic automata which can be learnt from corpus by Hidden Markov Models [51].

Morphology is described by means of regular transducers and regular relations, and lexical data bases, as well as tables of phonological and morphological rules are compiled or interpreted by algebraic operations on automata.

The existing techniques for compiling such machinery are rather confidential, while any naive approach leads to a combinatorial explosion. When transformation rules are local, it is possible to compile them into an invertible transducer directly obtained from the tree which encodes the lexicon.

A generic notion of sharing allows to have compact representation of such automata. Gérard Huet has implemented a toolkit based on this technique, which allows a very efficient automatical segmentation of a continuous phonologic text.

This study of the linear structure of language and of word structures is by itself sufficient for applications like orthographic correctors and text mining. Furthermore, this preprocessing is required for the analysis of other layers of natural language like syntax, semantics, pragmatics, etc.

3.2. Sentence structure and formal grammars: syntax

Participants: Lionel Clément, Alain Lecomte, Richard Moot [correspondent], Christian Retoré, Sylvain Salvati.

Sentence (or phrasal) structure is usually modelled via a tree structure. Different families of syntactic models are studied in Signes: rewriting systems of the Chomsky hierarchy, including tree grammars, deductive systems, i.e. categorial grammars, and constraint-based approaches.

3.2.1. Formal Grammars

Rewriting systems have excellent computational properties and a quite good descriptive adequacy. Relevant classes of grammars for natural language syntax, the so-called mildly context sensitive languages, are just a bit beyond context-free languages, and they are parsable in polynomial time as well well [43]. Among these classes of grammars let us mention Tree Adjoining Grammars [41], [42] and Minimalist Grammars [61], [62], [53]. Dependency Grammars and Lexical Functional Grammars share some properties with them but the general paradigm is quite different [52], [34].

Edward Stabler in [61] introduced Minimalist Grammars (MGs) as a formalization of the most recent model of the Chomskian or generative tradition and they are quite appealing to us. They offer a uniform model for the syntax of all human languages.

- There are two universal, language independent, rules, called merge and move: they respectively manage combination of phrases and movement of phrases (or of smaller units, like heads).
- Next, a language is defined by a (language dependent) lexicon which provides words with features describing their syntactic behavior: some features trigger merge and some others move. Indeed, features have positive and negative variants which must cancel each other during the derivation (this is rather close to resource logics and categorial grammars).

Consequently, MGs are able to describe numerous syntactic constructs, providing the analyzed sentences with a fine grained and complete syntactic structure. The richer the syntactic structure is, the easier it is to compute a semantic representation of the sentence.

MGs also cover phenomena which go beyond syntax, namely morphology via flexional categories, and they also incorporate some semantic phenomena like relations between pronouns and their possible antecedents, quantifiers, etc.

A drawback of rewrite systems, including MGs, is that they do not allow for learning algorithms that can automatically construct or enlarge grammars from structured corpora. But their main drawback comes from the absence of structure on terminals, which gives no hint about the predicative structure of the sentence.

Indeed, a strong reason for Signes using categorial grammars and their extensions [54]. Indeed, despite the inefficiency and the restricted linguistic coverage initial categorial grammars (BA, Lambek) provide a correspondence between syntactic analyses and semantic representations, which we are trying to extend to richer formalisms. This will be explained in the next section on the syntax/semantics interface.

In order to improve the computational properties of categorial grammars, and to extend their scope, we have been working on connecting them to more efficient and wider formalisms, like MGs [49], [48], [60].

A relatively new approach to syntax is known as model-theoretic syntax. Its advantages have been underlined by Geoffrey Pullum in [58]. Instead of viewing the trees or strings as a closure of some base set of expressions, they are viewed as trees or sets satisfying a set of formulae. This approach may be considered as another way of describing generative grammars. The advantages of such a description are not in the parsing algorithms (MSO or Constraint Satisfaction are usually of high complexity) but rather in characterising the language class and possibly describing it in a linguistically natural way (as opposed to lexical items of lexicalized grammars). This connection to logic is related to constraint-logic programming or to monadic second order logic.

In the MSO style, the pioneering work of James Rogers on Government and Binding and Tree Adjoining Grammars must be mentionned in [58]. Uwe Mönnich, Jens Michaelis and Frank Morawietz have obtained a two step description of minimalist grammars that we are studying [56], [55].

In the constraint style issued from the Prolog-Definite Clause Grammars, Head Phrase Structure Grammar, Construction Grammars and Property Grammars are defined as sets of constraints. The later ones introduced by Philippe Blache offer a rather natural way to describe grammar rules and have been studied by Marie-Laure Guénot in our group [35], [30].

3.2.2. High-Level Syntactic Formalisms

Lionel Clément worked on a formal representation of grammatical generalisations implemented for several linguistic formalisms.

This work deals with the problem of same linguistic phenomena expressed in several formalisms, alternative realisations and linguistic generalisations. The project aims at finding a common representation platform for all considered formalisms and factoring out elements shared in different linguistic constructions (i.e. different realizations of a nominal subject). The alternatives describe sets of related grammatical constructions (i.e. diathesis alternations). Finally, the shared part of these descriptions is expressed in a high-level linguistic

formalism closely related to metagrammar representations [31], [32], [33]. For instance, diathesis alternations can be considered an intersection of syntactic realizations of passive, active or causative sentences.

As exposed on the ARC Mosaïque web site http://mosaique.labri.fr/, the new idea introduced in the metagrammar paradigm is the fact that metagrammars handle two kinds of factorized informations: structural (and formalism dependent: tree structures, graphs, dependancies), and linguistic. The latter presupposes introducing of a way to represent non generative data and linguistic knowledge, without redundancy.

3.2.3. Linguistic Formalisations

In addition to studying formal properties of the models mentioned above, Signes use them to describe linguistic phenomena in various languages. Dependency Grammars have been applied to a detailed analysis of word order in German, whereas various French phenomena have been formalised and implemented as computational grammars adopting Property Grammar frameworks [35], [30]. Lexical Functional Grammar: the XLFG parser implemented by Lionel Clément. Finally, a morphosyntacitc analysis of Polish past tense and conditional verb forms has been modelled in HPSG. This formalism has also been used by members of the group to account for French inflection.

3.3. Sentence structure and logic: the interaction between syntax and semantics

Participants: Richard Moot, Christian Retoré, Sylvain Salvati [correspondent].

The principle of compositionnality, enounciated by Frege and formalized by Montague, makes syntax be the main window through which the meaning of natural language can be studied. This is the reason why mathematical linguistics is not only interested in the kinds of languages formalisms may describe, but also to the structures those formalisms assign to valid sentences. In his work Montague uses the lambda calculus so as to build the semantics of fragments of English, and the connection with the lambda calculus that the Curry-Howard correspondence naturally establishes with many kinds of intuitionnistic logics has made categorial formalisms quite popular for developing Montague semantics. But Montague's work may also be adapted to other syntactic formalisms and it gives some rather elegant formalization of semantics [57].

In Montague's approach the syntactic structure of a sentence is also the structure of its meaning and thus a from a grammar *à la* Montague, one can extract a grammar of a language and a grammar of meaning. This leads to the study of languages of structures denoting meanings. In particular, as in Montague tradition, when meanings are represented as logical formulae, themselves represented as lambda terms, this leads to a study of languages of lambda terms.

Montague's approach is only a first step toward modeling semantics and it has several shortcomings. For example, the particular kind of transformation it uses to map syntactic structures to meanings sometimes require to model syntax in a way that is counter-intuitive. The truth-conditional and intentional models of meaning, even though they can explain some phenomena of entailment, are not relevant from a cognitive point of view [40]. It is also the case that the strict compositional principle does not hold in general, as the famous Geach examples show.

The interests of the project at the interface between syntax and semantics are oriented in three directions. First, it seems that the minimalist approach of syntax advocated by Chomsky provides a description of language that is very close to it semantics. Thus we try to provide techniques for giving semantics to minimalist grammars. Second the various representations of discourse semantics are a way to extend Montague semantics and to make it more sensitive to the context. Finally, we try to understand the expressive power of Montague semantics so as to assess its possibilities.

3.4. Lexical semantics

Participants: Christian Bassac, Mauro Gaio, Bruno Mery, Christian Retoré [correspondent].

One of the most exciting challenges in computational linguistics is the question of lexical semantics, that is a proper treatment of word meanings and the way they relate one to another and finally how to handle the minimal interaction with knowledge representation. This part of semantics is relevant, not to say mandatory, for computing the semantic counterpart of composition be it lexical or syntactic.

The Generative Lexicon [59] is one of the most common frameworks for representing the internal structure of the meaning of words and morphemes, an alternative being the lexical functions of Mel'c̆uk and Polguère. The former is better suited for the logical apparatus developed by Signes, since Pustejovsky's set up can be viewed as an extension of Montague semantics, with which it shares the compositionality and the type theoretical formulation.

The information which depicts the sense of a word or morpheme is organised in three layers: the argument structure (related to logical semantics and syntax), the event structure, and the qualia structure. The argument structure provides types (in the type-theoretical sense) to the arguments encoded in the qualia structure regardless of whether they are syntactically mandatory or optional. The event structure follows [44]. It unfolds an event into several ordered sub-events with a mark on the most salient sub-event. Events are typed according to the typology of Vendler: state, process and transition, this latter type including achievement and accomplishment. The qualia structure relates the argument structure and the event structure in roles: formal, constitutive, telic, agentive.

This information and its organization into the generative lexicons allows an explanation of, for instance, polysemy and of compositionality (in particular in compound words or in simple phrase structure). This kind of model relates knowledge representation to linguistic organization and thus is especially useful for word sense disambiguation during (automated) syntactic and for computing the semantics of a compound, a phrase, a sentences and a discourse.

Signes is for instance interested in the so-called logical polysemy, that is how some occurrences may refer to one or another aspect (corresponding to a semantic type) of a given word. In order to get a better interface with syntax, our research rather try to extend logical and compositional sentence-semantics like Montague semantics and lambda-DRT, than to encode the structure that one finds in dictionaries and lexical studies.

3.5. Discourse and dialogue structure: computational semantics and pragmatics

Participants: Mauro Gaio, Alain Lecomte, Richard Moot, Christian Retoré [correspondent].

Montague semantics have some limits. Two of them which, technically speaking, concern the context, can be overcome by using DRT, that is Discourse Representation Theory and its variants. [44], [63] Firstly, if one wants to construct the semantics of a piece of text, one has to take into account sequences of sentences, either discourse or dialogue, and to handle the context which is incrementally defined by the text. Secondly, some constructs do not obey the strict compositionality of Montague semantics, since pronouns can refer to bound variables. For instance a pronoun of the main clause can be bound in a conditional sub-clause.

For these reasons, Discourse Representation Theory was introduced. This model defines an incremental view of the construction of discourse semantics. As opposed to Montague semantics, this construction is top-down, and proceeds more like state change than like functional application — although lambda-DRT present DRT in a Montague style, see e.g. [63].

These approaches may be used for constructing semantic representations of fragments of natural language. Such representations are relevant for applications like information extraction and retrieval, question answering system, and human-computer interaction, among others.

4. Application Domains

4.1. Sanskrit philology

Participant: Gérard Huet [correspondent].

Sanskrit literature is extremely rich, and is part of the world cultural patrimony. Nowadays, Internet can provide to both specialists and inquiring minds an access to it.

This kind of resource already exists for ancient Greek and Latin literature. For instance, Perseus (http://www.perseus.tufts.edu) provides an online access to texts. A simple click on each word analyses it, and brings back the lexical item of the dictionary, possible meanings, statistics on its use, etc.

The work described in the following sections enables such computational tools for Sanskrit, some of which are already developed and made available on a web site (http://sanskrit.inria.fr). These tools efficiently and accurately assist the annotation of Sanskrit texts. Besides, a tree bank of Sanskrit examples also is under construction. Such a corpus annotation tool is a prerequisite to the implementation of a Perseus-like facility for Sanskrit.

4.2. Parsers

Participants: Lionel Clément [correspondent], Richard Moot, Sylvain Salvati.

In the implementation of a robust parser, one of the major issue arises from homonymous words and phrases. Natural language is highly ambiguous and each sentence, taken without any pragmatic or semantic context, has a huge number of possible meanings. In written languages this combinatorial problem necessitates the use of subtle techniques; but in spoken languages, where normative rules have less influence, those techniques do not seem to be able to cope with ambiguity. The recent developments of natural language processing concerning the problem of ambiguity is based on stochastic and low-level methods. Those techniques try only to represent surface dependencies and forget about the various structures of phrases and about their meanings. They are quite efficient for applications such as information retrieval and lack accuracy in others like automatic translation.

We would like to develop new techniques so as to allow robust parsing of spoken language, but also so as to deal with the computation of meaning regardless the ambiguity of sentences. Usually the various possible analyses of a sentence are represented in a structure called "shared forest". Such a structure can be seen as a tree automaton. This remark gives us several directions of research. A first one would be to adapt various techniques coming from automata theory especially concerning automaton transformations and transductions. A second one consists in using the connection between tree automaton theory and the weak MSO theory of trees so as to perform selections of certain sets of analyses.

4.3. Syntactic and semantic modeling

Participants: Lionel Clément, Richard Moot [correspondent], Sylvain Salvati.

The SIGNES team develops Grail, for multimodal categorial grammars, and XLFG, for Lexical-Functional Grammars as well as several different grammars with different levels of grammatical coverage for these formalisms, ranging from the specific - giving descriptions of linguistic phenomena such as French clitics and extraposition - to wide-coverage: wide-coverage categorial grammars for both Dutch and French have been developed and grammar models.

Categorial grammars have a transparent syntax-semantics interface by means of the Curry-Howard homomorphism, where a proof of the grammaticality of a sequence of typed words immediately gives us the way of combining the corresponding semantic expressions. It is our goal to develop this correspondence to allow for wide-coverage semantic analysis, using Pustejovsky's generative lexicon to help with semantic disambiguation.

A specific application we envisage, collaborating with researchers from Pau and Toulouse, is to perform syntactic, semantic and discourse analysis of a corpus reciting voyages through the region of the Pyrénées. Naturally, this chain of analysis will be specialized towards the application domain, benefiting from specific knowledge concerning the region, means of transport and the conventions of this style of literature.

5. Software

5.1. The Zen Toolkit

Participant: Gérard Huet [correspondent].

The Zen Toolkit is a library of finite state automata and transducers, called Zen for its simplicity. The algorithmic principles of the Zen library are based on the linear contexts data structure ("zippers") and on the sharing functor (associative memory server) [36]. It has been developed by Gérard Huet and is being used in his Sanskrit modelling platform (see section 5.2). It allows the construction of lexicons, the computation of morphological derivatives and inflected forms, and the segmentation analysis of phonetic streams modulo euphony [36].

The Zen Toolkit is implemented in an applicative kernel of Objective Caml, called Pidgin ML. It follows a *literate programming* style of documentation, using the program annotation tool Ocamlweb of Jean-Christophe Filliâtre, available for Ocaml. The Zen toolkit is distributed as free software (under the LGPL licence) in the Objective Caml Hump site as well as at URL http://sanskrit.inria.fr/ZEN/. This development forms a significant symbolic manipulation software package within pure functional programming, which shows the feasibility of developing in the Ocaml system symbolic applications having good time and space performance, within a purely applicative methodology.

The Zen Toolkit has been used, e.g., to implement a lexicon of french flexed forms (Nicolas Barth and Sylvain Pogodalla, Calligramme project-team at Loria). It is also used by Aarne Ranta (Chalmers University) as a morphological engine of the Grammatical Frameworks software.

5.2. Sanskrit Site

Participant: Gérard Huet [correspondent].

Gérard Huet's Sanskrit Site (http://sanskrit.inria.fr) provides a unique range of interactive resources concerning Sanskrit philology [38], [37]. These resources are built upon, among other ingredients, the Zen Toolkit (see section 5.1). The site registers thousands of visitors daily.

- The *declension engine* gives the declension tables for Sanskrit substantives.
- The *conjugation engine* conjugates verbs for the various tenses and modes.
- The *lemmatizer* tags inflected words.
- A *dictionary* lists inflected forms of Sanskrit words. Full lists of inflected forms, in XML format (given with a specific DTD), are released as free linguistic resources available for research purposes. This database, developed in collaboration with Pr. Peter Scharf, from the Classics Department at Brown University, has been used for research experiments by the team of Pr. Stuart Shieber, at Harvard University.
- The *Sanskrit Reader* segments simple sentences, where the (optional) finite verb form occurs in final position. This reader enhances the hand-tagged Sanskrit reader developed by Peter Scharf, that allows students to read simple texts differently: firstly in davanagari writing, then word-to-word, then in a word-to-word translation, then in a sentence-to-sentence translation.
- The *Sanskrit Parser* eliminates many irrelevant pseudo-solutions (segmentations) listed by the Sanskrit reader.
- The *Sanskrit Semantic Analyzer*, based on the notion of *kāraka* of Pāṇini, controls overgeneration using a pertinence principle [39].
- The *Sanskrit Tagger* is an assistant for the tagging of a Sanskrit corpus. Given a sentence, the user chooses among different possible interpretations listed by the morpho-syntactic tools and may save the corresponding unambiguously tagged sentence on disk as an hypertext document indexing in the Sanskrit Heritage Dictionary (our structured lexical database). This service has no equivalent worldwide.

- The *morphological data* for Sanskrit have been released by Gérard Huet under LGPLLR (http://sanskrit.inria.fr/DATA/XML/). The precise lexer used by the shallow parser is specified as a *modular transducer* whose top-level states are the lexical categories corresponding to the flexed forms banks, and whose arcs correspond to (the inversion of) euphony (*sandhi*) rules.

An on-going project is the construction of a tree bank of Sanskrit examples, in collaboration with Pr. Brendan Gillon, from McGill University in Montreal.

5.3. Grail: Natural Language Analysis with Multimodal Categorial Grammar

Participants: Richard Moot [correspondent], Natalia Vinogradova.

Grail is a modern, flexible and robust parser/automated theorem prover for multimodal categorial grammars (MMCG [54]) developed by Richard Moot. It is designed to allow students and researchers to design and experiment with their grammars while at the same time offer the advanced users many optional optimisation strategies.

Grail can be run either as a command line script or as an application with a graphical interface, with the possibility to follow Grail's partial parses/proof attempts interactively. It is freely available from http://www.labri.fr/perso/moot/grail3.html.

5.4. Suite of Corpus Tools for Type-Logical Grammars

Participants: Richard Moot [correspondent], Natalia Vinogradova.

A suite of corpus tools has been developed by Richard Moot. It contains tools for the display, search, transformation and extraction of grammars on the base of an annotated corpus. In addition, there is are tools for the generation of training and test data for maximum entropy models, a supertagger and scripts for error analysis included in the tools. Grail and the supertagger are designed to work in tight integration. This suite of corpus tools is available from http://www.labri.fr/~moot/Corpus/.

5.5. XLFG5: Experimental LFG Parsing

Participant: Lionel Clément [correspondent].

XLFG5 is a parser prototype for research. It implements a variant of the Lexical Functional Grammar (LFG) formalism. The parsing produces a shared forest (of c-structures) in order to speed-up ambiguous sentences analyzes. Sharing functional dependency structures (f-structures) is under development.

XLFG5 is used for teaching in various universities, amongst which Université Bordeaux 3, Université Paris IV, and others in Spain or Algeria. Languages known to be used with XLFG5 are french, arabic, mandarin, spanish, english, german and thai.

XLFG5 has been developed by Lionel Clément and is available as a parsing server (http://www.xlfg.org/). The new version has been made widely available.

5.6. Enriching a geographic ontology

Participant: Mauro Gaio [correspondent].

Within the ANR GEONTO (http://geonto.lri.fr), a software (without any official name yet) has been released that extracts terms/concepts candidates from raw texts. This extraction aims at enriching a geographic ontology.

5.7. Yab: a GLR Parser Generator for S-Attributed Grammars

Participant: Lionel Clément [correspondent].

YAB is a GLR parser generator for S-Attributed grammars. This compiler has been used to develop a syntactic parser dealing with homonymies in LFG. It relies on a specific restriction of the LFG formalism to build a polynomial-time syntactic parser. This software has been developed by Lionel Clément (before he joined the Signes project-team). It is publicly available (http://www.labri.fr/perso/clement/yab/).

5.8. Lexed: a Dictionary Lexicalizer

Participant: Lionel Clément [correspondent].

Lexed is a lexicalizer. It allows one to search a dictionary entry for a string. The finite automata-based algorithm is particularly fast, and offers a good alternative to hashes for large dictionnaries. Lexed is a C++ library distributed with a GPL Licence (http://www.labri.fr/perso/clement/lexed/). This software has been developed by Lionel Clément (before he joined the Signes project-team).

5.9. Other On-going Software Developments

In the following, we list various on-going software developments. Most of these software and resources are in an early stage, and not yet packaged or available. Please get in touch with their correspondent to know the current status.

5.9.1. Hyperion: a parser for HR grammars

Participant: Richard Moot [correspondent].

Hyperion is a hypergraph parser developed by Richard Moot. It analyses graphs using hyperedge replacement grammars. The core parser is a very small and very general implementation of Clemens Lautemann's dynamic programming algorithm for parsing graphs using hyperedge replacement grammars. A rudimentary Dot/Graphviz interface for portraying the hypergraphs and grammars is included. Hyperion is written in XSB Prolog.

Hyperion is a first implementation of many of the author's ideas which were presented at TAG+9 as well as in talks in Bordeaux and Chieti. An early version of the system can be dowloaded from http://www.labri.fr/perso/moot/hyperion/.

5.9.2. Datalog parser for PMCFG

Participant: Sylvain Salvati [correspondent].

Sylvain Salvati has developed a recognizer for PMCFG (Parallel Multiple Context Free Grammars) which has the correct prefix property. This recognizer is based on a compiler that transforms a grammar (a PMCFG) into a datalog program which in turn is optimized in several steps. Recognition is reduced to solving a query with this datalog program, the analyzed sentence being encoded as an extensional database. A datalog query solver is also provided with this prototype. The optimization of the datalog program is based on a modular approach which composes several atomic transformations. The composition of these transformations is enforcing properties of the recognition algorithm. Therefore composing different transformations yields to different recognition algorithms. This program is based on an extension of Kanazawa [50] prefix correct algorithm for MCFG (Multiple Context Free Grammars) proposed by Sylvain Salvati. This program is already used by Kanazawa and by the Calligramme INRIA project-team in Nancy. In the future it should be improved in several ways; first it should become a parser; second it should be interfaced with certain classes of Abstract Categorial Grammars, so as to perform text generation; finally, so as to face the scaling problem of handling big grammars, compositions of transformations should be compiled into one efficient transformation.

6. New Results

6.1. Properties of the Formalisms

6.1.1. Formal grammars and type theory

Participants: Pierre Bourreau, Sylvain Salvati [correspondent].

Pierre Bourreau and Sylvain Salvati have put in correspondence a syntactic criterion (negatively non-duplicating) on types and the λ-terms (almost-affine) that inhabit them. This result has been proved using a game theoretic approach. Furthermore, the interest in this correspondence lies in the fact that negatively non-duplicating types have at most one inhabitant.

Pierre Bourreau and Sylvain Salvati completed the extension of Kanazawa's technique for parsing almost linear ACGs to parsing almost affine ACGs.

Sylvain Salvati studied a newly defined class of automata, *higher-order pushdown automata with collapse*, and proved they where computing the (possibly infinite) tree generated by a higher-order programming scheme using Krivine machines. This result main interest consists in showing that Krivine machine may well be a good way of studying the properties of higher-order programming schemes.

Sylvain Salvati worked further on the notion of recognizability in the simply typed λ-calculus. He has given a definition in terms of congruences of finite index in order to study a possible extension of Eilenberg variety Theorem to recognizable sets of λ-terms. Surprisingly a difficulty of this line of research is to prove that the congruential definition of recognizability is equivalent to the ones that use standard models or intersection types.

6.1.2. Properties of mildly context sensitive formalisms

Participant: Sylvain Salvati [correspondent].

Sylvain Salvati obtained, in a collaboration with Makoto Kanazawa, a precise account of the copying power of well-nested Multiple Context-Free Language (MCFL). This result is, to the best of our knowledge, providing the simplest way of separating well-nested MCFL from MCFL.

Sylvain Salvati has pursued his research on MCFL and tried to understand the iteration (or pumping) properties of this class of languages. It turns out that he now conjectures that there are languages in the class of MCFL that are not iterable. He gave a candidate of a language that he believes not to be iterable.

Still on the difference between MCFL and well-nested MCFL, Sylvain Salvati worked on the problem of whether MIX is a well-nested MCFL or not. As a starting point, he tried to prove that it was not a well-nested MCFL of rank 2 (*i.e.* a Tree Adjoining Language). He was able to prove that MIX is not in a certain subclass well-nested MCFL of rank 2 and he is now working on extending this result so as to prove that MIX is actually not a Tree Adjoining Language.

6.2. Semantics

6.2.1. Wide-coverage semantics lexicons

Participant: Richard Moot [correspondent].

On the level of semantics, Richard Moot has begun the development of two wide-coverage semantic lexicons, one producing DRT semantics (Kamp & Reyle 2003) and one producing a continuation-based semantics in the style of Philipp de Groote (2006). The system has been presented at the TALN 2010 conference with 2 papers and a demonstration and at the workshop on Logical Methods for Discourse with a presentation and a demonstration.

6.2.2. Text generation with XLFG

Participant: Lionel Clément [correspondent].

XLFG is still used for teaching and research in formal syntax. Now Lionel Clément is extending with some automatic text generation capabilities. This extension involves some new algorithmic and software developments.

6.2.3. Lexical semantics

Participants: Christian Bassac, Christian Retoré [correpondent], Bruno Mery.

Bruno Mery Christian Bassac, Christian Retoré pursued the development of their type theoretical model of lexical semantics.

Bruno Mery gave an elegant account of the quantification over the different aspects of a lexeme by using the proper notion of projections. In some sentences, some objects might be counted according to one aspect and to another aspect when a pronoun refers to them. In the following example: books are first counted as physical object, and thereafter as contents, since Boas did not read each copy of the same book: *All the books in the library burnt but Boas already read them*. This appears in the manuscript of Bruno Meryâs PhD.

Christian Retoré provided a further refinement by using linear types. That way, the control of transformations from an aspect of a word to another is blocked or licensed by the type themselves and not by an external rule. This gives a purely type theoretical account of the proper co-predication over different meanings (This book is interesting but heavy.) and rejects the infelicitous ones (∗Washington borders the Potomac and attacked Iraq). He gave talks in Paris, Montpellier and Verona on such issues.

6.2.4. Logic and linguistics

Participants: Michele Abrusci, Ivano Ciardelli, Christian Retoré [Correspondent].

Ivano Ciardelli, Jean Gillibert, Christian Retoré are producing a direct proof of the completeness of sheaf of classical models with respect to intuitionistic predicate calculus.

Michele Abrusci and Christian Retoré studied the logical formalisation of ordinary language quantification (âsomeâ, âfor allâ, but also âthe majority ofâ, âmost ofâ) and they found evidences that set theoretic models do not provide a relevant account. They started a proof theoretical view of these phenomena, relying on proofs and refutations. A paper is submitted to the congress CLMPS.

Sylvain Pogodalla and Christian Retoré found a counter example to the criterion of Michele Abrusci for viewing Lambek proofs with cuts as graphs satisfying a certain criterion. Some graphs not satisfying the criterion to correspond to proof with cuts.

6.3. Syntax

6.3.1. POS-tagging and Suppertagging for French

Participant: Richard Moot [correspondent].

Richard Moot has continued his work on the extraction of categorial grammars from the French Treebank (annotated by Anne AbeillÃ© and her team at Paris VII). Much work has been spent on the development of POS-tagger and Supertagger models (two essential components for the performance of the Grail parser when using the extracted grammar: the POS-tagger assigns syntactic categories like "name", "present tense verb" etc., whereas the Supertagger assigns categorial grammar formulas to words). The POS-tagger currently performs at the same level as the best POS-taggers available for French, whereas the Supertagger performs at the same level as the best categorial Supertaggers for English (no comparable software exists for French for a direct comparison).

7. Other Grants and Activities

7.1. Regional research programs

7.1.1. ITIPY Automated extraction of itineraries from historical-geographical corpora

Participants: Mauro Gaio, Anaïs Lefeuvre, Bruno Mery, Richard Moot [correspondent], Christian Retoré, Natalia Vinogradova.

A project including travel money and a PhD grant funded by Région Aquitaine and INRIA involves Signes, DESI LIUPPA (Pau) and LALIC IRIT (Toulouse). The purpose of this project is to automatically extract spatial and temporal information from reports of travel through the Pyrénées, in order to reconstruct itineraries. It should provide an easy access, in particular for tourists, to the wide literary corpora gathered and digitalized by the Médiathèque de Pau. At the same time it will put forward the regional patrimony. This task involves information retrieval, deep syntactic and semantic analysis of relevant parts.

7.2. National Initiatives

7.2.1. ANR project FREC Frontiers of recognizability

Participant: Sylvain Salvati [correspondent].

FREC is a four-years project (2010-2013) funded by the French National Research Agency (ANR). Its goal is to push further the frontiers of recent developments in the theory of automata and recognizability to more complex and realistic structures: trees, λ-terms and models of quantitative automata. The involvement of Sylvain Salvati in the project is mainly related to the notion of recognizability in the λ-calculus.

7.2.2. ANR GEONTO Programme Masse de Données et Connaissances

Participant: Mauro Gaio [correspondent].

The GEONTO project is a 3-year project (2008-2010) funded by the French National Research Agency (ANR) involving 4 partners: LRI, LIUPPA, COGIT, IRIT. It focuses on interoperability of diverse data related to geographic information. The first part of the project consists in building several geographic ontologies reflecting several (*different* might be better) points of view. In order to complete these objectives, various approaches relying on techniques taken from natural language processing will be used. The second part of the project will study the alignment of ontologies built in the previous part. See http://geonto.lri.fr/ for details.

7.2.3. ANR project LOCI

Participants: Christian Bassac, Christian Retoré [correspondent], Richard Moot, Sylvain Salvati.

Signes is one of the five sites of the ANR Blanc research program LOCI *Locativity and Interaction in Logic, Linguistics and Computer Science* with three labs SFL (Paris), LIPN (Paris), IML (Marseille), and the association Résurgences. With linear logic and ludics as a common underlying technique, this program is exploring several related topics from the humanities (speech acts, dialogue and lexicon; social interaction and learning; communication and web ontologies).

7.2.4. ANR project Rhapsodie

Participant: Lionel Clément [correspondent].

Rhapsodie is a 4-year project (2008-2011) funded by the French National Research Agency (ANR). It aims at building an annotated corpus of spoken French that will be freely distributed within the research community. This corpus will contain several kinds of spoken discourses, and its annotations will focus both on prosody and syntax, including topology and syntactic dependency. It will be a precious resource to understand the status of prosody in spoken French and its relation with syntax and informational structure. It will also be a suitable resource to train probabilistic parsers targetted at spoken French. In particular it has been shown how to annotate a (French) spoken corpus. The project includes partipants from several research groups in France, as well as Belgium and Switzerland. The involvement of Signes members in Rhapsodie mainly concerns syntax. See http://rhapsodie.risc.cnrs.fr for details.

7.2.5. Groupement de Recherche C.N.R.S.: MAGIS

Participant: Mauro Gaio [correspondent].

MAGIS "Methods and Applications for Geographic Information Science" is a CNRS "Collaborative Research Network". At present the network is composed of 35 French research centers and laboratories (more than 200 researchers). MAGIS is organized in four GIS domains whose objective is to develop fundamental methods and favor synergy across different scientific projects: (1)data integration, (2)spatio-temporal modeling and representation, (3)spatio-temporal methods and techniques, (4)GI organizational and societal contexts. Major events supported by MAGIS :

1. MAGIS Summer school - SAGEO : conference on Spatial Analysis and Geomatics,
2. CQFD-Géo : Quebec&French conference for the development of the Geomatic
3. European Journal of GIS and Spatial Analysis (Hermès edition)

7.2.6. TUL 2010-2014 project: Approches typologiques des constructions elliptiques

Participant: Lionel Clément [correspondent].

We may define elliptic constructions as syntactic structures which are missing at least the head or a complement whose interpretation and, in a certain way, whose syntax are fixed by the immediate context. This property distinguishes them from non-verbal sentences such as *A nous la liberté!* which are complete. The terminology establishes the following cases:

1. head ellipsis: Gapping, Comparative deletion, Sluicing, Stripping, Conjunction reduction / Argument Cluster coordination, N'-deletion,
2. complement ellipsis: Verb Phrase Ellipsis, Pseudo-gapping, Antecedent-contained deletion, Null-complement anaphora, Right Node Raising

The aim of the project is to compare syntax, semantics and prosody of those constructs in various languages. The project is more particularly concerned with the following problems:

1. what kind of ellipsis exist in each studied languages?
2. what are the constraints and the properties of those ellipsis?
3. which of those properties can be related to other properties of the considered languages?

The project shall:

1. create a multilingual corpus of elliptical constructions and make it available on the Internet.
2. organize two international conferences (in 2011 and in 2013) in Paris.
3. publish a book in the collection Studies in Languages, Complementary series of John Benjamins, and edited by Werner Abraham and Michael Noonan.

8. Dissemination

8.1. Animation of the scientific community

8.1.1. Organization of seminars

- Pierre Bourreau has been organizing the seminar of the team until September 2010.
- Anaïs Lefeuvre and Nathalia Vinogradova are in charge of the team seminar for the academic year 2010-2011.
- Jean Gillibert, Christian RetorÃ© as well as Michele Abrusci and Boas Erez gave many lectures at their weekly reading group on Sheaves in logic and geometry (2009/2010) and Logic, categories geometry (2010/2011) .
- Sylvain Salvati is now organizing the seminar *Graphes et logique* of the LaBRI research group *Méthodes formelles*.

8.1.2. Editorial boards

- Mauro Gaio has been the head of the committee of a special issue of the *Revue Internationale en Géomatique*", published by Hermès and that will bee released during the first semester of 2011.
- Christian Retoré has been a reviewer for *Mathematical Reviews* since October 2003.
- Christian Retoré has been the editor in charge of the relations between mathematics, logic and computer science in La Gazette des Mathématiciens (quarterly, Société Mathématique de France) since July 2008.

8.1.3. Program committees of conferences and schools

- Christian Bassac was a member of the program committee of the *The Seventh International Conference on Morphology* which was held in december 2010 in Toulouse.
- Mauro Gaio has been co-president of the committee of the conference SAGEO 2010.
- Mauro Gaio is a member of the program committee of the conference SAGEO (2010, 2011).
- Mauro Gaio is member of the French program committee of ICC 2011 (http://www.icc2011.fr/).
- Mauro Gaio is a member of the program committee of *Conference on Geospatial Semantics* http://geosco.org/geos2011/committee.html.
- Christian Retoré has been member of the program committee of the 48th Association for Computational Linguistics 2010 (Uppsala), section mathematics of language.
- Christian Retoré has been a member of the program committee of the Formal semantics week in Nancy march 22â26 2010.
- Richard Moot was a member of the reading committee of TALN 2010 in Montreal.
- Richard Moot is a member of the reading committee of TALN 2011 in Montpellier.
- Richard Moot is a member of the program committee of LACL 2011 in Montpellier.
- Natalia Vinogradova has been co-chair of the student session of ESSLLI 2010.

8.1.4. Academic committees

- Christian Retoré is a member Hiring Comittee of UniversitÃ© Toulouse III Lecturer in computer science.
- Since January 2010, Sylvain Salvati has been a member of the teacher-and-researcher committee of INRIA Bordeaux - Sud-Ouest. Tasks include the evaluation of proposals concerning invitations of professors and temporary assignments at INRIA.

8.1.5. Organization of events

- Richard Moot organized two ITIPY meetings in Bordeaux (2-3 June 2010, 9-10 December 2010).
- Christian Retoré, together with Jean Gillibert, organized the international workshop on Logic, categories, semantics with Jean Gillibert. (Cartier, Girard, Asher, Vickers, Lamarche, Streicher, Pollard, Preller, Gehrke, Melliès, Moortgat, Abrusci).
- Sylvain Salvati has been the president of the program committee of the Rencontres INRIA Industries *les industries du numérique pour la santé* held in Bordeaux April 15.

8.2. Thesis committee

- Christian Retoré was in the HDR committee of Dominique Archambault who defended a habilitation at Université Paris 6 on *Interaction et usages des modalitÃ©s non visuelles, accessibilitÃ© des contenus complexes* (December 2010).

- Christian Retoré was on the PhD committee of Paul BÃ©daride who defended a thesis at Université Nancy 1 on *Implication Textuelle et RÃ©Ã©criture* (October 2010).

8.3. Academic supervision

8.3.1. Student internship supervision – fourth and fifth year

- Lionel Clément is supervising the Master thesis of Jean-SÃ©bastien Brilleau on *Les anaphores associatives en franÃ§ais parlé spontané* (the defense should take place in May 2011).
- Lionel Clément supervised a programming project on automatic text generation.
- Sylvain Salvati supervised Jérôme Kirman in initiation to research.

8.3.2. PhD supervision

- Christian Bassac and Christian Retoré are co-supervising the PhD thesis work of Bruno Mery (Université Bordeaux 1, ministry grant): *type theory for lexical semantics*.
- Mauro Gaio is supervising with Christian Sallaberry the PhD thesis work of Nguyen Van Tien (Université de Pau et des Pays de l'Adour): *improving a geographical ontology: a method based on a semantic corpus analysis*.
- Alain Lecomte is supervising the PhD thesis of Christophe Onambele (Université Paris 8): *Minimalist Grammars and application to linguistic descriptions*.
- Alain Lecomte is co-supervising the PhD thesis of Mawusse Kpakpo Akue Adotevi (Université de Lomé, Togo): *Dialogical Games and Language Games*.
- Richard Moot and Christian Retoré are co-supervising the PhD thesis of Natalia Vinogradova (Université Bordeaux 1, INRIA CORDI grant): *analyse sèmantique à large échelle du français avec une application aux expressions géographiques dans un corpus historique.*
- Christian Retoré is co-supervising the PhD thesis of Ivano Ciardelli on *Du sens des mots au sens d'une phrase : modèles en logique linéaire du premier ordre* with Simone Martini (Bologna) et Jean Gillibert (IMB) Bourse CordiS INRIA - cotutelle France-Italie
- Christian Retoré and Richard Moot are co-supervising the PhD thesis of Noémie-Fleur Sandillon-Rezer on *Acquisition de grammaires catégorielles à partir de corpus annotés co-encadrement* (CNRS fellowship)
- Christian Retoré and Mauro Gaio are co-supervising Anaïs Lefeuvre on *Extraction automatique d'itinéraires dans des récits de voyages* (Bourse RÃ©gion Aquitaine - INRIA)
- Christian Retoré and Sylvain Salvati are co-supervising the PhD thesis of Pierre Bourreau (Université Bordeaux 1, INRIA CORDI grant): *the treatment and use of non-linearity in computationnal linguistics*.

8.4. Participation to colloquia, seminars, invitations

8.4.1. Talks at conferences, seminar talks and invitations

- Pierre Bourreau presented his PhD work during the LaBRI PhD students seminar.
- Pierre Bourreau gave a presentation to high school students during the FÃªte de la Science in October.
- Pierre Bourreau gave a talk to present the parsing of almost affine second order ACG at the CAuLD meeting in December 2010.
- Anaïs Lefeuvre presented a talk entitled *Bientôt: portée et repérage* at the CAuLD meeting in December 2010.

- Anaïs Lefeuvre and Natalia Vinogradova gava a talk at FÃªte de la science, octobre 2010 : *un ordinateur peut-il comprendre notre langue ?*
- Anaïs Lefeuvre presented the article *Temporal and Nominal Anaphora* (Barbara Partee 1984) during the working group of the team.
- Christian Retoré gave a talk at FÃªte de la science, octobre 2010 : *un ordinateur peut-il comprendre notre langue ?*
- Richard Moot presented his work on the extraction of categorial grammars for French at TALN 2010 conference, on 22 July 2010 and at the Nancy CauLD meeting, on 14 December 2010.
- Richard Moot gave two presentations on the topic *Can a computer understand our language* to high school students during the FÃªte de la Science, 21 october 2010.
- Christian Retoré gave a talk entitled *Un peu de nuance dans une logique de brutes: la question du sens en linguistique informatique*. Séminaire Unithé ou café, INRIA Bordeaux Sud-Ouest, 17 décembre 2010.
- Christian Retoré gave a talk entitled *Parler et comprendre : processus cognitifs et modèles informatiques de la faculté langage*. Conférence proposé par l'Asco et club cognitique, Bordeaux, jeudi 2 décembre 2010.
- Christian Retoré gave a talk entitled *Which type theory for lexical semantics*, 16e rencontres Logique, Algèbre, Calcul, GDR Informatique Mathématique, 16-17 novembre 2010, PPS, U. Paris 7.
- Christian Retoré gave a talk entitled *Inserting lexical semantics into montagovian compositional semantics*, séminaire TAL du LIRMM, Montpellier, 5 juillet 2010.
- Christian Retoré and Jean Gillibert gave a talk entitled *Faisceaux en logique et en géométrie*, journées maths-info de l'UFR éponyme - 2 juillet 2010.
- Christian Retoré gave a talk entitled *Logic in linguistics: some issues in the philosophy of language* UniversitÃ degli studi di Verona FacoltÃ di Lettere e Filosofia. Avril-Mai 2010. 16h
- Christian Retoré gave a talk entitled *Sémantique et modélisation dans le cadre de la logique catégorique*. Séminaire ANR GENIUS IJN/IHESS/ENS 15 avril 2010.
- Christian Retoré gave a talk entitled *Une intégration de la sémantique lexicale Ã la sémantique compositionnelle montagovienne*, LIPN U. Paris 13, 25 jan. 2010.
- Sylvain Salvati presented his work on recognizability in the simply typed λ-calculus on several occasions this year:
  - at the ANR Modulo meeting in January (Bordeaux),
  - at the workshop on Automata Concurrency and Times Systems in in February (Chennai),
  - at the workshop on Higher-Order Recursion Schemes and Pushdown Automata in March (Paris),
  - at the seminar on Semantics at PPS in June (Paris)
  - at the seminar of the Plume team in November (Lyon)
- Sylvain Salvati presented his proof that MIX is a 2-MCFL on several occasions this year:
  - at the seminar Graphes et Logiques in February (Bordeaux)
  - at the workshop Multiple Context-Free Grammars and Related Formalisms in September (Tokyo),
  - at the LINA in October (Nantes)
- Sylvain Salvati gave a presentation on the topic *Can a computer understand our language?* to high school students during the FÃªte de la Science in October.

- Natalia Vinogradova presented a talk entitled *Bientôt: portée et repérage* at the ITIPY meeting in December 2010.
- Natalia Vinogradova presented the articles *Categorizing binary topological relations between regions, lines and points in geographical databases* (by M. Egenhofer et J. Herring) and *The 9+-Intersection: A Universal Framework for Modeling Topological Relations* (Y. Kurata) during the working group of the team.

8.4.2. Participation to conferences and summer schools

- Pierre Bourreau, Anaïs Lefeuvre, Natalia Vinogradova and Sylvain Salvati attended the *Journées Sémantique et modélisation* in Nancy.
- Anaïs Lefeuvre and Natalia Vinogradova attended the 22nd European Summer School in Logic, Language and Information.
- Richard Moot visited TALN 2010, 19-23 july 2010, Montreal, Canada.
- Richard Moot and Sylvain Salvati visited Logic, categories, semantics, 12-13 november 2010, Bordeaux.
- Pierre Bourreau, Anaïs Lefeuvre, Richard Moot, Sylvain Salvati and Noémie-Fleur Sandillon-Rezer visited Logical Methods for Discourse, 13-14 december 2010, Nancy.
- Sylvain Salvati attended LATA 2010, in May (Trier).

8.5. Teaching

As half of its members are university staff, Signes is intensively implied in teaching, both in the computer science cursus (University Bordeaux 1) and in the linguistic cursus (University of Bordeaux 3). Signes is also teaching in summer schools for PhD students and colleagues. What follows only covers lectures whose topic is related to computational linguistics:

- Mauro Gaio created (2010-2011) a new lecture in first year of master entitled *Structures et modèles de contenus pour le Web*.
- Mauro Gaio created a new lecture in second year of master entitled: *Concepts et Méthodes Informatiques pour la Recherche d'Informations par le Contenu*.
- Anaïs Lefeuvre is giving 60 hours teaching in second year of *Licence de sciences du langage*.
- Pierre Bourreau gave a course in IUT second year on formal language theory.
- Pierre Bourreau gave a course in IUT *licence professionnelle* on applied formal methods.
- Lionel Clément gave a course on *automatic translation* in the second year of Master on *Traitement Automatique des Langues*.
- Lionel Clément gave courses in the first year of Master on *Traitement Automatique des Langues*.
- Lionel Clément gave a course on Linguistics and Computer Science in third year of Licence.
- Lionel Clément gave a course on Linguistics and Formalisation in first year of Master.
- Christian Retoré gave a Seminar/graduate lecture between April 24 and May 16 2010 : *Logic in linguistics: some issues in the philosophy of language* (16hours)
- Sylvain Salvati gave an introductory lecture on Mathematics of Languages at the Ecole d'Automne en Linguistique (ENS) in September (3 hours).
- Natalia Vinogradova gave 50 hours of courses at Université Bordeaux 3 (30h in the first year of License in Science du langage (syntax, morphology, phonetic,...) and 20h in third year of Licence (computational linguistics, introduction to Python).

9. Bibliography

Major publications by the team in recent years

[1] C. BASSAC. *Morphologie et Information Lexicale*, Université Michel de Montaigne Bordeaux 3, December 2006, Mémoire d'habilitation à diriger des recherches.

[2] L. CLÉMENT, A. KINYON. *Generating parallel multilingual LFG-TAG grammars from a MetaGrammar*, in "Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Sapporo, Japan", July 2003.

[3] L. CLÉMENT. *Zeugme sémantique*, in "Revue de Sémantique et de Pragmatique", 2010, p. 1-17, à paraître, http://hal.archives-ouvertes.fr/hal-00413217/PDF/zeugmes_semantiques-2.pdf.

[4] G. HUET. *Transducers as lexicon morphisms, phonemic segmentation by euphony analysis, application to a sanskrit tagger*, in "Journal of Functional Programming", July 2005, vol. 15, n^{o} 4, p. 573–614.

[5] A. LECOMTE. *Categorial Grammar for Minimalism*, in "Logic and Grammar", C. CASADIO, P. SCOTT, R. SEELY (editors), CSLI Publications, 2005.

[6] J. LESBEGUERIES, M. GAIO, P. LOUSTAU. *Geographical information access for non-structured data*, in "Proceedings of the 2006 ACM Symposium on Applied Computing (SAC), Dijon, France, April 23-27, 2006", H. HADDAD (editor), ACM, 2006, p. 83–89, http://doi.acm.org/10.1145/1141277.1141296.

[7] R. MOOT. *Proof nets for linguistic analysis*, UIL-OTS, Universiteit Utrecht, 2002.

[8] R. MOOT. *Automated Extraction of Type-Logical Supertags from the Spoken Dutch Corpus*, in "The Complexity of Lexical Descriptions and its Relevance to Natural Language Processing: A Supertagging Approach", S. BANGALORE, A. JOSHI (editors), MIT Press, 2007.

[9] C. RETORÉ. *Logique linéaire et syntaxe des langues*, Université de Nantes, January 2002, Mémoire d'habilitation à diriger des recherches.

[10] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", 2004, vol. 2, n^{o} 1, p. 3–25.

[11] S. SALVATI. *On the Membership Problem for Non-Linear Abstract Categorial Grammars*, in "Journal of Logic, Language and Information", 2009, p. 0925-8531 [*DOI :* 10.1007/s10849-009-9110-0], http://hal.inria.fr/inria-00446785/PDF/generation.pdf.

[12] S. SALVATI. *Recognizability in the Simply Typed Lambda-Calculus*, in "16th Workshop on Logic, Language, Information and Computation", Tokyo Japon, 2009, http://hal.inria.fr/inria-00412654/en/.

Publications of the year

Articles in International Peer-Reviewed Journal

[13] M. AMBLARD, A. LECOMTE, C. RETORÉ. *Categorial Minimalist Grammar: From Generative Syntax To Logical Form*, in "Linguistic Analysis", 12 2010, vol. 36, n^{o} 1–4, p. 273–306, http://hal.archives-ouvertes.fr/hal-00545748/PDF/ALR_festschrift.pdf.

[14] M. AMBLARD, A. LECOMTE, C. RETORÉ. *From Lambek style minimalist analyses to logical forms*, in "Linguistic Analysis", 04 2010, vol. 35, n^{o} 1-2, http://hal.archives-ouvertes.fr/hal-00414306/en/.

[15] C. BASSAC, B. MERY, C. RETORÉ. *Towards a Type-Theoretical Account of Lexical Semantics*, in "Journal of Logic Language and Information", 2010, p. 229-245, http://hal.archives-ouvertes.fr/inria-00408308/PDF/bassac-mery-retore-revised.pdf.

[16] C. RETORÉ, S. SALVATI. *A Faithful Representation of Non-Associative Lambek Grammars in Abstract Categorial Grammars*, in "Journal of Logic Language and Information", 04 2010, vol. 19, p. 185–200, http://hal.archives-ouvertes.fr/inria-00409557/PDF/nl_acg_retore_salvati_jolli.pdf.

[17] S. SALVATI. *On the Membership Problem for Non-Linear Abstract Categorial Grammars*, in "Journal of Logic, Language and Information", Apr 2010, vol. 19, n^{o} 2, p. 163-183 [*DOI :* 10.1007/s10849-009-9110-0], http://www.springerlink.com/content/f3g38386p568n7j4/, http://hal.inria.fr/inria-00446785/en.

Articles in National Peer-Reviewed Journal

[18] C. BASSAC. *Sémantique de la coordination: une brève introduction*, in "Revue de sémantique et pragmatique", september 2010.

[19] C. BASSAC, M. ÇIÇEK. *Factorisation des affixes de pluriel et de possessif en turc*, in "Revue de sémantique et pragmatique", september 2010.

[20] L. CLÉMENT. *Zeugme sémantique*, in "Revue de Sémantique et de Pragmatique", 2010, p. 1-17, http://hal.inria.fr/hal-00413217/en.

[21] L. CLÉMENT. *Zeugme sémantique*, in "Revue de Sémantique et de Pragmatique", 2010, p. 1-17, à paraître, http://hal.archives-ouvertes.fr/hal-00413217/PDF/zeugmes_semantiques-2.pdf.

International Peer-Reviewed Conference/Proceedings

[22] M.-N. BESSAGNET, M. GAIO, E. KERGOSIEN, C. SALLABERRY. *Extraction automatique d'un lexique Ã connotation géographique Ã des fins ontologiques dans un corpus de récits de voyage*, in "TALN 2010 - 17e Conférence sur le Traitement Automatique des Langues Naturelles", Canada Montréal, Jul 2010, 10 pages, http://hal.inria.fr/hal-00536083/en.

[23] K. MAKOTO, S. SALVATI. *The Copying Power of Well-Nested Multiple Context-Free Grammars*, in "Language and Automata Theory and Applications", Allemagne Trier, A.-H. DEDIU, H. FERNAU, C. MARTÃN-VIDE (editors), LNCS, Springer, 2010, vol. 6031, p. 344-355 [*DOI :* 10.1007/978-3-642-13089-2_29], http://www.springerlink.com/content/b8w7r53761614lk7/, http://hal.inria.fr/inria-00525077/en.

[24] R. MOOT. *Semi-automated Extraction of a Wide-Coverage Type-Logical Grammar for French*, in "TALN 2010", Canada Montréal, Jun 2010, http://hal.inria.fr/inria-00494062/en.

[25] R. MOOT. *Wide-Coverage French Syntax and Semantics using Grail*, in "TALN 2010", Canada Montréal, Jun 2010, Publication to accompany a TALN 2010 system demo., http://hal.inria.fr/inria-00494064/en.

[26] D. PALACIO, C. SALLABERRY, M. GAIO. *Normalizing Spatial Information to Improve Geographical Information Indexing and Retrieval in Digital Libraries*, in "14th International Symposium on Spatial Data Handling", Hong-Kong Hong-Kong, 2010, p. 229-234, http://hal.inria.fr/hal-00458059/en.

Workshops without Proceedings

[27] C. BASSAC. *On the semantics of blends*, in "International conference on lexical blending", 2010.

Scientific Books (or Scientific Book chapters)

[28] R. MOOT. *Extraction of Type-Logical Supertags from the Spoken Dutch Corpus*, in "Supertagging: Using Complex Lexical Descriptions in Natural Language Processing", A. JOSHI, S. BANGALORE (editors), MIT Press, Jun 2010, http://hal.inria.fr/inria-00494072/en.

References in notes

[29] K. R. BEESLEY, L. KARTTUNEN. *Finite-State Morphology: Xerox Tools and Techniques*, Cambridge University Press, 2002.

[30] P. BLACHE. *Property Grammars: A Fully Constraint-based Theory*, in "Constraint Satisfaction and Language Processing", H. CHRISTIANSEN, P. SKADHAUGE, J. VILLADSEN (editors), Springer, 2005.

[31] M.-H. CANDITO. *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées, Application au français et à l'italien.*, Université de Paris 7, 1999.

[32] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *A new metagrammar compiler*, in "TAG+6", 2002.

[33] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Représentation et gestion du lexique d'une grammaire d'arbres adjoints*, in "Traitement automatique des langues", 2004, vol. 43, n^{o} 3.

[34] A. DIKOVSKY, L. MODINA. *Dependencies on the other side of the Curtain*, in "Traitement Automatique des Langues", 2000, vol. 41, n^{o} 1, p. 67-95.

[35] M.-L. GUÉNOT, P. BLACHE. *A descriptive and formal perspective for grammar development*, in "Workshop on Foundations of Natural-Language Grammars", Edimbourg, Royaume-Uni, 2005.

[36] G. HUET. *Linear Contexts and the Sharing Functor: Techniques for Symbolic Computation.*, in "Thirty Five Years of Automating Mathematics", F. KAMAREDDINE (editor), Kluwer, 2003.

[37] G. HUET. *Design of a Computational Linguistics Platform for Sanskrit*, in "Logical Aspects of Computational Linguistics, LACL 05", 2005, Invited Lecture.

[38] G. HUET. *Un système de traitement informatique du sanskrit*, in "Journée ATALA : Traitement Automatique des Langues Anciennes", 2005, Invited Lecture.

[39] G. HUET. *Shallow syntax analysis in Sanskrit guided by semantic nets constraints*, in "International Workshop on Research Issues in Digital Libraries (IWRIDL 2006), Kolkata, India", December 2007, To appear in ACM Digital Library.

[40] R. JACKENDOFF. *Languages of the mind*, MIT Press, 1995.

[41] A. JOSHI, L. LEVY, M. TAKAHASHI. *Tree Adjunct Grammar*, in "Journal of Computer and System Sciences", 1975, vol. 10, p. 136–163.

[42] A. JOSHI, Y. SCHABES. *Tree Adjoining Grammars*, in "Handbook of Formal Languages", Berlin, G. ROZENBERG, A. SALOMAA (editors), Springer Verlag, Berlin, 1996, vol. 3, chap. 2.

[43] A. JOSHI, K. VIJAY-SHANKER, D. WEIR. *The convergence of mildly context-sensitive grammar formalisms*, in "Fundational issues in natural language processing", P. SELLS, S. SCHIEBER, T. WASOW (editors), MIT Press, 1991.

[44] H. KAMP, U. REYLE. *From Discourse to Logic*, D. Reidel, Dordrecht, 1993.

[45] R. M. KAPLAN, M. KAY. *Regular Models of Phonological Rule Systems*, in "Computational Linguistics", 1994, vol. 20,3, p. 331–378.

[46] L. KARTTUNEN. *Applications of Finite-State Transducers in Natural Language Processing*, in "Proceedings, CIAA-2000", 2000.

[47] K. KOSKENNIEMI. *A general computational model for word-form recognition and production*, in "10th International Conference on Computational Linguistics", 1984.

[48] A. LECOMTE, C. RETORÉ. *Extending Lambek grammars: a logical account of minimalist grammars*, in "Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, ACL 2001", Toulouse, ACL, July 2001, p. 354–361.

[49] A. LECOMTE, C. RETORÉ. *Towards a Minimal Logic for Minimalist Grammars: a Transformational Use of Lambek Calculus*, in "Formal Grammar (FG 1999)", FoLLI, 1999, p. 83–92.

[50] K. MAKOTO. *A prefix-correct Earley recognizer for multiple context-free grammars*, in "9th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+9)", Tübingen, Germany, C. GARDENT, A. SARKAR (editors), June 2008, p. 49–56.

[51] C. MANNING, H. SCHUTZE. *Foundations of statistical natural language processing*, MIT Press, 1999.

[52] I. MELCUK. *Dependency syntax – theory and practice*, Linguistics, State University of New York Press, 1988.

[53] J. MICHAELIS. *Derivational minimalism is mildly context sensitive*, in "Logical Aspects of Computational Linguistics (LACL 1998), selected papers", M. MOORTGAT (editor), LNCS/LNAI, Springer-Verlag, 2001, n^{O} 2014, p. 179–198.

[54] M. MOORTGAT. *Categorial Type Logic*, in "Handbook of Logic and Language", Amsterdam, J. VAN BENTHEM, A. TER MEULEN (editors), North-Holland Elsevier, Amsterdam, 1996, chap. 2, p. 93–177.

[55] F. MORAWIETZ. *Two-Step Approaches ot Natural Language Formalisms*, Studies in Generative Grammar, Mouton de Gruyter, Berlin · New York, 2003.

[56] U. MÖNNICH, J. MICHAELIS, F. MORAWIETZ. *On Minimalist Attribute Grammars and Macro Tree Transducers*, in "Linguistic Form and its Computation", C. ROHRER, A. ROSSDEUTSCHER, H. KAMP (editors), CSLI Publications, 2004.

[57] S. POGODALLA. *Computing Semantic Representation: Towards ACG Abstract Terms as Derivation Trees*, in "Proceedings of the Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+7)", May 2004, p. 64–71.

[58] G. K. PULLUM, B. C. SCHOLZ. *On the distinction bewteen moedl-theoretic and generative-enumerative syntax*, in "Logical Aspects of Computational Linguistics (LACL 2001)", P. DE GROOTE, G. MORRILL, C. RETORÉ (editors), LNCS/LNAI, Springer-Verlag, 2001, n^{o} 2099, p. 17–43.

[59] J. PUSTEJOVSKY. *The Generative Lexicon*, MIT Press, 1995.

[60] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", 2004, vol. 2, n^{o} 1, p. 3–25.

[61] E. STABLER. *Derivational Minimalism*, in "Logical Aspects of Computational Linguistics (LACL 1996)", C. RETORÉ (editor), LNCS/LNAI, Springer-Verlag, 1997, vol. 1328, p. 68–95.

[62] E. STABLER. *Remnant movement and structural complexity*, in "Constraints and Resources in Natural Language Syntax and Semantics", G. BOUMA, E. HINRICHS, G.-J. M. KRUIJFF, R. OEHRLE (editors), CSLI, 1999, p. 299–326, distributed by Cambridge University Press.

[63] J. VAN EIJCK, H. KAMP. *Representing Discourse in Context*, in "Handbook of Logic and Language", Amsterdam, J. VAN BENTHEM, A. TER MEULEN (editors), North-Holland Elsevier, Amsterdam, 1996, chap. 3, p. 179–237.