INRIA

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

SIGNES is an INRIA Project-Team joint with University of Bordeaux (UB1 and UB3 (TELEM, EA 4195)) and CNRS (CLLE, UMR 5263 and LaBRI, UMR 5800)

1. Team

Research Scientist

Gérard Huet [INRIA, Research Director (DR) located in Rocquencourt, HdR]
Renaud Marlet [INRIA, Research Associate (CR), Until November 30, HdR]
Richard Moot [CNRS, Research Associate (CR)]
Sylvain Salvati [INRIA, Research Associate (CR)]

Faculty Member

Lionel Clément [University Bordeaux 1, Assistant Professor (MCF)]
Christian Retoré [University Bordeaux 1, Team Leader, Professor (Pr), HdR]
Émilie Voisin [University Bordeaux 3, Until August 31]

External Collaborator

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 Student

Pierre Bourreau [INRIA, CORDI/S, since Sep. 2007]
Bruno Mery [University Bordeaux 1, since Oct. 2006]
Natalia Vinogradova [INRIA, CORDI/S, since Sept. 2007]

Administrative Assistant

Sylvie Embolla [INRIA]

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 [48], [49], [50], [31]. 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 [55].

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, Renaud Marlet, 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 [45]. Among these classes of grammars let us mention Tree Adjoining Grammars [43], [44] and Minimalist Grammars [66], [67], [57]. Dependency Grammars and Lexical Functional Grammars share some properties with them but the general paradigm is quite different [56], [36].

Edward Stabler in [66] 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 [58]. 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 [52], [51], [65].

A relatively new approach to syntax is known as model-theoretic syntax. Its advantages have been underlined by Geoffrey Pullum in [63]. 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 [63]. Uwe Mönnich, Jens Michaelis and Frank Morawietz have obtained a two step description of minimalist grammars that we are studying [60], [59].

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 [37], [32].

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 [33], [34], [35]. 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 [37], [32]. 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: Renaud Marlet, 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 [62].

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 [42]. 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, Renaud Marlet, 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 [64] 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 [46]. 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, Renaud Marlet, 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. [46], [69] 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. [69].

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. Towards French Sign Language (LSF) modelling and processing

Participants: Lionel Clément, Renaud Marlet, Christian Retoré, Émilie Voisin [correspondent].

After a global prohibition of Sign Languages decided in 1880 (and which lasted until the sixties in the USA and until the eighties in France), deaf people can use sign language and rather recently these languages are the object of new studies and development. A first aspect is the social acknowledgment of sign language and of the deaf community, a second aspect is the linguistic study of this language with a different modality (visual and gestural as opposed to auditive and phonemic), and the third and most recent aspect, which relies on the second, is the need for sign language processing. A first goal is computer aided learning of Sign Language for hearing people and even deaf people without access to sign language. A more challenging objective would be computer aided translation from or to sign language, or direct communication in sign language.

Given the rarity of linguistic study on the syntax and semantics of sign languages — some exceptions concerning American Sign Language are [61], [53], [54] — before being able to apply our methodology, our first task is to determine the structure of the sentence, using our personal competence as well as our relationship with the deaf community.

We intend to define methods and tools for generation of sign language sentences. It should be noted that there is a continuum of different representations of a sentence in Sign Language, from a grammatical description with agreement features and word/sign order that we are familiar with, to a notation system like Signwriting [68] or to a language for the synthesis of 3D images and movies. Our competences on the interface between syntax and semantics are well designed for work in generation of the grammatical representations.

A first application would be a software for teaching Sign Language, like the CD ROM *Les Signes de Mano* by IBM and IVT. Indeed, presently, only dictionaries are available on computers, or examples of sign language videos, but no interactive software. Our generation tools, once developed, could be useful to educative purposes.

4.3. Parsers

Participants: Lionel Clément [correspondent], Renaud Marlet, 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.

4.4. Syntactic and semantic modeling

Participants: Lionel Clément, Renaud Marlet, 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: a wide-coverage categorial grammar of Dutch has been developed while a similar wide-coverage grammar for French is currently is currently in the process of being constructed.

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) [38]. 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 [38].

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 [40], [39]. 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 [41].
- 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 [58]) 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. 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/).

5.7. 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).

5.8. French Sign Language HD Corpus

Participant: Émilie Voisin [correspondent].

Thanks the two HD-video cameras provided by a former regional grant, Émilie Voisin and Patrick Henry recorded two hours of sign language utterances from native deaf speakers. This corpus is made of translations of French texts and short comics strips (with no textual annotation) turned into sign language expressions; it also contains sign language productions by non-native speakers as well as the reactions and comments of native sign language speaker confronted to these productions. Annotation and extraction of sequences in the corpus can be performed thanks to a specific interface.

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

Keywords: *hyperedge replacement grammar*.

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

Keywords: *parallel multiple context-free grammar.*

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 [47] 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. Relational Machines

Participant: Gérard Huet [correspondent].

Gérard Huet and Benoît Razet (who completed this year his Thèse d'Informatique "Machines d'Eilenberg effectives" under the supervision of the former) developed a computational framework in which actions of a non-deterministic machine are executed on on abstract relational machine called an Eilenberg machine, named after the X-machine model of Samuel Eilenberg (1974). Such machines have two components: a control component — a non-deterministic finite-state automaton, whose transitions are labeled with action generators — and a data component — a relational interpretation of the generators over some data domain. Benoît Razet gave a complete definition of the general model, plus variations such as finite machines, for which a termination criterion ensures completeness of bottom-up search. A general machine operates a sequential interpreter called the reactive engine, parameterized by a strategy guiding the resumption structure. Effective relations are curryfied into stream producers, of which two varieties exist, one for partial computability, the other for total computability. Benoît Razet also gave a complete classification of functional implementations of the various compiling schemes of regular expressions into finite automata transition graphs. His work gives the foundation for an ambitious program of high-level relational programming.

6.1.2. Formal grammars and type theory

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

Christian Retoré and Lutz Stassburger (INRIA EPI Parsifal) answered negatively to both an old (1993) conjecture by Retoré and a more recent (1998) by Gugliemi : there exists a Pomset proof net in the sense of Retoré which does not correspond to a sequential proof in the various formulations of Pomset Logic in sequent calculus, but this proof net does correspond to a proof term of the BV-calculus. Hence Pomset Proof net cannot be sequentialized with any standard sequent calculus, and on the other hand, Pomset Logic is not equivalent to BV-calculus.

Sylvain Pogodalla (INRIA EPI Calligramme) and Christian Retoré studied the non commutative proof nets with cuts, trying to extend the proof nets without links for commutative multiplicative linear logic (2003): up to now, they mainly showed that the unique criterion handling non commutative proof nets with cuts, due to Abrusci and Maringelli, is false.

The work of Christian Retore and Sylvain Salvati on embedding of the non-associative Lambek grammars into the ACG framework has been polished and appeared in the Journal of Logic Language and Information [18]

Sylvain Salvati has proved that the problem of generation in the montagovian framework is decidable. A side effect of this work is that it gave rise to a notion of recognizable sets of simply typed lambda-terms. It has several interesting applications outside computational linguistics, such as in formal language theory, in model checking and software verification. Some work is currently being done in collaboration with other members of Méthodes Formelles (LABRI) in order to promote this concept among other formal methods. [29], [19]

Pierre Bourreau and Sylvain Salvati worked on parsing and generation techniques for abstract categorial grammars, extending Kanazawa's technique to a wider class of λ-terms, where deletion can be performed.

Sylvain Salvati has started to extend formal language theory to handle free-order languages. He proposed an extension of recognizable languages, context free languages and and multiple context free languages (MCFL) so as to close them under permutation. In the two first cases, he gave some new machine characterizations and complexity results. The last case proves to be much more difficult to understand, since the permutation power of MCFLs is mostly unknown. In order to get a better understanding of this, Sylvain Salvati started to study the MIX problem which has been open for nearly 30 years now. It was believed that this language MIX that contains the words having the same numbers of a's, of b's and c's is not an MCFL, but Sylvain Salvati proved that it actually was using techniques inspired from algebraic topology. This results should resume discussions about the notion of mildly context sensitivity that is supposed to capture natural languages. Concerning these permutation closed languages, Sylvain Salvati emphasized a deep connection with Abstract Categorial Grammars which easily encompass them.

6.2. Semantics

6.2.1. Syntax and Semantics of Categorial Minimalist grammars

Participants: Alain Lecomte [correspondent], Christian Retoré.

Maxime Amblard (INRIA EPI Calligramme), Alain Lecomte and Christian Retoré, for Lambek 85 Festschrift to appear in Linguistic Analysis, extended their encoding of minimalist grammars into categorial grammars to *adjunction* and also managed to compute the proper semantical effect of adjunction. [13] On the syntactic side this categorial view of minimalist grammars is closely related to the two step approach of Mönnich and Morawietz, see the note [17] on Morawietz's book [59].

In Toulouse in April in the symposium "Quel sens pour la linguistique" organized in honour of Sir John Lyons who was awarded an Honoris Causa doctorate, Christian Bassac presented a paper intitled "Categorial Grammars" in which he showed that John Lyons had had a good intuition as regards the place and development of categorial grammars in his book "Introduction to theoretical linguistics" published in 1968, and other papers. In his presentation C. Bassac compared the two approaches of Phrase Structure Grammars and Categorial Grammars and showed that Sir John's prediction was partly right (Categorial Grammars took a wider place in linguistics) and partly wrong (as they were never integrated as such in PSG) but gave rise to various formalisms for instance Stabler, Lecomte, Rétoré's Minimalist Categorial Grammars which for instance analyse Merge as the elimination of $/$.

Christian Bassac presented a paper intitled "Philosophy, linguistics and semantic interpretation" at the international conference on "Philosophy of language and Linguistics" (April 2009) at the University of Łódź in Poland. He analysed to what extent philosophers and linguists disagree on what logical form really is and showed how some formalisms such as linear logic can obviate the step of translation from a syntactic form to a logical form as syntax and semantics are consubstantial here.He then showed how linear logic can be used to account for various linguistic phenomena such as morphological agreement, and coercion phenomena.

6.2.2. Effective Compositional Semantics in Grail

Participants: Richard Moot [correspondent], Natalia Vinogradova.

In addition to the syntactic part of Grail, Richard Moot and Natalia Vinogradova are currently developing a semantic lexicon in parallel with the syntactic lexicon. The goal of this semantic lexicon is to assign lambda terms (the "natural" semantics for categorial grammars) to natural language sentences. Though it currently handles only the most frequently occurring formulas, it already correctly handles a number of interesting constructions (verb clusters, passivization, wh extraction) in addition to the basic structures (nouns, adjectives, prepositions, adverbs).

6.2.3. Computational Semantics with Rich Lexical Semantics

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

Christian Bassac, Bruno Mery and Christian Retore continued the formalization of lexical semantics in an integrated logic framework. Their work on lexical disambiguation in context, in particular for object with facets, with second order lambda calculus has been polished and published in the JoLLI. They went beyond lambda calculus by using the finer grained linear type theory: the multiplicative conjunction of linear logic as a generic type for multi facet objects without systematic projections, yet with a compound type related to its components, and discovered that the exponential can faithfully represent the reversibility or irreversibility of some meaning transformations. [14]

Christian Bassac analyzed the Event Structure of verbal lexical items and showed that an Extended Event Structure such as postulated in the Generative Lexicon theory can offer an explanatory account to such phenomena as passivization, the opposition between resultatives and secondary depictives, and Psych-verbs. This work was presented during the international conference on "Event, Events, Sub-events/ Evénement, événements et sous-événements" organized by the University of Paris 3 in September 2009. [30]

6.2.4. Semantic Zeugma and Coordination

Participant: Lionel Clément [correspondent].

Lionel Clément pursued his work on semantic zeugma, that is a figure where two incompatible meanings of a lexical item get simultaneously instantiated. Le lexical choice is responsible for the opposition between the two parts of the coordination. In his paper he suggest to restore the possibility of a lexical choice for the antecedent to replace the choice previously made. The interplay between the two possible instantiations provides a pragmatical reading to this rhetorical figure. [15]

6.2.5. Pragmatics

Keywords: *argumentation*, *fallacies*, *ludics*, *presupposition*.

Participant: Alain Lecomte [correspondent].

Alain Lecomte has managed the PRELUDE project, a project funded by the ANR until November 2009. In this framework, he worked out a modelling of dialogue and other topics in pragmatics and rhetorics. This work was presented as a joint work with Myriam Quatrini (from IML, Marseille-Luminy) at WoLLIC 09, held in Tokyo in June. Moreover, other points of the project were presented at the Colloquium “Dialogue, Games and Interaction” which was held in Paris in September. Another presentation was made in Florence in early September, in the frame of the LIGC colloquium (“Logique, Interaction et Géométrie de la Cognition”) and in homage to Prof. Michele Abrusci for his 60th Birthday. This modelling of interaction is based on the concepts of Ludics, a new frame invented by J-Y Girard in order to give new foundations to Logic. The main point is that such a view allows us to reconcile two views in semantics and pragmatics: one based on Proof Theory, and the other on Games Semantics. [25]

6.2.6. Spatio-temporal Semantic Analysis

Keywords: *corpus studies*, *geographic information system*, *indexing*, *information retrieval*, *lexical semantics*.

Participants: Mauro Gaio [correspondant], Natalia Vinogradova, Richard Moot.

For the present Mauro Gaio with the colleagues involved in the GEONTO project http://geonto.lri.fr/ focuses his research on interoperability of diverse data related to geographic information. In this project we particularly intend to facilitate applications dealing with such a diversity, from the integration of database schemas to the intuitive query of textual documents, being either technical or common documents. In this context, Mauro Gaio has proposed several approaches which depend on natural language processing techniques. The goal of these approaches is to analyse existing geographical texts of the public library *la médiathèque de Pau* looking for geographical expressions linked with geographical names and typical linguistic situations surrounding them. He also has contributed to propose a set of tools to extract and formalize these expressions in order to build a geographic ontology reflecting document content points of view. The integration of this linguistically anchored geographical ontology in the semantical part of Grail is being discussed with Richard Moot and Natalia Vinogradova. All the works of this project will be developed and tested on actual data. Real data in the geographic field, more particularly in the topographic domain, will be provided by the French national mapping agency and a public library in the town of Pau for experiments. In France, there are numerous and huge databases in this domain (about one hundred of giga-bytes). [21], [24], [16], [26], [27], [28]

6.3. Morphology and Syntax

6.3.1. Automated grammar checking

Participant: Lionel Clément.

Lionel Clément, Renaud Marlet with Kim Gerdes (U. Paris 3) defined the central algorithm of an open system for grammar checking, based on deep parsing. The grammatical specification is a context-free grammar with flat feature structures (Datalog). After a shared-forest analysis where feature agreement constraints are relaxed, error detection globally minimizes the number of corrections and alternative correct sentences are automatically proposed. [23], [22]

6.3.2. The Grail parser/theorem prover

Participants: Richard Moot [correspondent], Natalia Vinogradova.

Richard Moot has improved the Grail parser/theorem prover in several respects. First is the addition of an online tutorial (http://www.labri.fr/perso/moot/tutorial/), giving an introduction on how to use Grail, detailing how to write grammars, use the parser interactively and explaining the more advanced features, which help the user make their grammar as efficient as possible. Secondly, the semantic component has seen several additions, such as a type checker for the lambda terms in the lexicon and an interface to database queries. Finally, Grail has been updated to integrate more tightly with the supertagger and a user interface has been added to facilitate the use of Grail in combination with a part-of-speech tagger and supertagger.

6.3.3. Towards a wide coverage categorial grammar for French

Participant: Richard Moot [correspondent].

Richard Moot has started the development of a wide-coverage categorial grammar for the french language, which is automatically extracted from the Paris VII treebank, then corrected manually. A very early version of this categorial grammar, though still containing noise and artifacts, already gives promising first results. Because of the large lexicon, which assigns several hundreds of different lexical formulas to many frequent words (forms of the auxiliary verbs "être" and "avoir" and conjunctions like "et" and "ou", for example have between 300 and 500 entries in the current lexicon), the lexical ambiguity is prohibitive to parsing with the extracted grammar. A solution with a supertagger, which assigns only the most probable formula to a word based on its immediate context (preceding and succeeding words, formulas assigned to previous words) give promising first results: trained on over 250.000 words of the treebank it assigns between 80,81% and 95,67% of the correct formulas to words in unseen sentences (the lower precision corresponds to exactly one formula per words, whereas the higher precision assigns, on average, 6,98 formulas). Work to clean up the extracted grammar continues, with each correction both reducing the number of different formulas extracted (currently 2.531, though only 1.052 occur more than once) and improving the performance of the supertagger.

6.3.4. Sanskrit Processing

Participant: Gérard Huet [correspondent].

The cooperation with the Sanskrit Studies Department of Hyderabad University continued successfully. In January 2009 the 3rd Sanskrit Computational Linguistics Symposium was organized in Hyderabad by Amba Kulkarni. Amba Kulkarni and Gérard Huet are joint editors of the Proceedings, edited in the Springer Verlag Lecture Notes series. In November 2009 Amba Kulkarni spent a month at INRIA Paris-Rocquencourt, plus visits at various European sites in view of forming a Euro-India Consortium on the topic of Sanskrit Computational Linguistics Symposium. A Memorandum of Understanding between INRIA and Hyderabad University for the continuation of the joint team effort in a bilateral framework has been drafted, and should be signed in 2010.

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, Renaud Marlet, 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 research programs

7.2.1. ARC CAuld Automated construction of logical representations of discourse

Participants: Christian Bassac, Pierre Bourreau, Alain Lecomte, Renaud Marlet, Bruno Mery, Richard Moot, Christian Retoré, Sylvain Salvati [correspondent], Natalia Vinogradova.

ARC CAULD gathers two INRIA EPI (Calligramme, Signes) a team from Toulouse (LALIC, IRIT) and one from Paris (LLF). Its aim is to give an operational contents to Discourse Reprentation Theory, to overpass its limits. In particular ompositional treatment of DRT is advocated for, as well as a proper treatment of context, references, anaphora and dynamic meaning construction. Typed lambda calculus is the main framework to achieve this.

7.2.2. Groupement de Recherche C.N.R.S. 2521 Sémantique et modélisation

Participants: Christian Bassac, Pierre Bourreau, Alain Lecomte, Renaud Marlet, Bruno Mery, Richard Moot, Christian Retoré [correspondent], Sylvain Salvatti, Émilie Voisin.

Signes is one of the fifteen research teams involved in the Groupe de Recherches CNRS 2521, GDR *Sémantique et Modélisation*, directed by Francis Corblin (Université Paris IV) 2002-2005, 2005-2008. This research program is divided into five “operations”: modèles et formats de représentation pour la sémantique; les modèles à l’épreuve des données; sémantique et corpus; Les interfaces de la sémantique linguistique; sémantique computationnelle. The Signes project-team is part of the latter two operations (interfaces of linguistic semantics and computational semantics). See http://semantique.free.fr/ for details.

7.2.3. 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.

7.2.4. ANR blanche PRELUDE

Participants: Alain Lecomte, Bruno Mery, Richard Moot, Sylvain Salvati, Christian Retoré [correspondent].

Signes is part of the national research program PRELUDE *Towards a theoretical pragmatics based on ludics and continuations* is a 3-year project (2008-2011) funded by the French National Research Agency (ANR) (November 2006 – November 2009) directed by Alain Lecomte (team *Structures Formelles de la Langue*, Univ. Paris 8). Other partners are the LaBRI (most of the Signes members from the LaBRI), the INRIA project-team Calligramme (LORIA, Nancy) and the Institut Mathématique de Luminy (Marseille). See http://anr-prelude.fr/ for details.

7.2.5. ANR project Rhapsodie

Participants: Lionel Clément, Renaud Marlet [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. [20] 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.

7.3. Associate research team

Participant: Gérard Huet [correspondent].

Signes and an Indian group in computational linguistics lead by Amba Kulkarni (University of Hyderabad) and Puspak Battacharyya (IIT Mumbai) has been approved as an INRIA Franco-Indian Research Network in Computational Linguistics. This network is aimed at enforcing the relationship and exchanges in computational linguistics between France and India. In particular, a Sanskrit WordNet should be designed. See http://yquem.inria.fr/~huet/EA/Sanskrit08.html for details.

8. Dissemination

8.1. Involvement within the scientific community

8.1.1. Honours

- Gérard Huet obtained the EATCS 2009 award.

8.1.2. Editorial boards

- Alain Lecomte has been on the editorial board of the international journal *Traitement Automatique des Langues (TAL)* since August 2001.
- Christian Retoré has been a reviewer for *Mathematical Reviews* since October 2003.

- Christian Retoré has been editor in chief of the international journal *Traitement Automatique des Langues (TAL)* since April 2004 (and in the editorial board since 2001) until Summer 2009.
- 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

- Mauro Gaio was co-chair of the Colloque Sageo 09 http://sageo09.univ-pau.fr/.
- Gérard Huet is co-editing with Amba Kulkarni and Peter Scharf the proceedings of the 1st and 2nd International Sanskrit Computational Linguistics Symposiums, held respectively in INRIA Rocquencourt (France) in October 2007 and Brown University (Providence, Rhode Island, USA) in May 2008.
- Gérard Huet and Amba Kulkarni will also co-edit the proceedings of the 3rd Symposium, that will take place in University of Hyderabad (India) in January 2009.
- Alain Lecomte and Christian Retoré were on the program committee of workshop Games, Dialogs and Interactions in 2009 at Paris.
- Richard Moot and Christian Retore were in the Reading Committee of TALN 2009 (Senlis).
- Richard Moot was co-chair of the program committee of ESSLLI 2009.
- Richard Moot was part of the program committee for the special issue "Machine Learning for Natural Language Processing" of the journal Traitement Automatique des Langues.
- Richard Moot was part of the program committee of Formal Grammar 2009 in Bordeaux.
- Richard Moot was part of the program committee of the ESSLLI 2009 Workshop "Parsing with Categorial Grammars".
- Christian Retoré was on the committee of the *Journées de sémantique et modélisation / Conference on Semantics and Formal Modelling* held in Paris in March 2009.

8.1.4. Academic committees

- Renaud Marlet was a member of the hiring committee for associate scientists (chargés de recherche) at INRIA Bordeaux - Sud-Ouest, 2009.
- Since November 2007, Renaud Marlet has been a member of the Incentive Action Working Group (GTAI) of the Scientific and Technological Orientation Council (COST) of INRIA. Tasks include selecting and evaluating Collaborative Research Initiatives (ARC), Exploratory Actions (AE), Software Development Operations (ODL) and Technological Development Actions (ADT).
- Since November 2007, Renaud Marlet has been a member of the International Affairs Working Group (GTRI) of the Scientific and Technological Orientation Council (COST) of INRIA. Tasks include selecting and evaluating Associated Teams (EA) as well as proposals made in internatinal programs such as CONICyT/INRIA (with Chile), EuroMéditerrané 3+3 (with Algeria, Spain, Italy, Morocco, Tunisia), SECyT/INRIA-CNRS (with Argentina), STIC-AmSud (with South America), STIC-Tunisie (with Tunisia), and ERCIM post-doctorates.
- Since January 2008, Renaud Marlet 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.
- Christian Retoré was on a hiring committee for a professorship in Université Paris XIII.
- Christian Retoré was the head of the Master on Algorithms and Formal Methods of Université Bordeaux 1 until June 2009.
- Christian Retoré is a member of the scientific committee (conseil scientifique) of the LaBRI.

8.1.5. Organization of events

- Alain Lecomte organized an international workshop Games, Dialogues and Interactions in Paris in September 2009 to conclude the Prélude national research program.
- ESSLLI 2009 (520 participants, 48 lectures and workshops, 2 weeks) organizing committee included Christian Retoré (chair), Pierre Bourreau, Lionel Clément, Alain Lecomte, Bruno Mery, Richard Moot, Sylvain Salvati, Natalia Vinogradova, Emilie Voisin: Signes project-team was the main organizer of the European Summer School in Logic and Language and Information (ESSLLI), Bordeaux, France, July 2009..

8.2. 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 computational linguistics:

- Lionel Cément taught linguistic formalization to linguistics students at the University Bordeaux 3. The aim was to present the Lexical Functional Grammar (LFG) theory as well as an implementation of a few linguistic phenomena with the XLFG5 parser (see section 5.5).
- Lionel Clément and Christian Retoré are teaching a lecture on Natural Language Processing (60h, 4th year), Université Bordeaux 1.
- Gérard Huet gave a course on algebraic and functional calculi at the Master Parisien de Recherche en Informatique (MPRI, Master 2, 12h).
- Christian Retoré is teaching (with Jean Gillibert) a course on *Sheaves in Geometry and Logic*, one hour per week, (9 hours in 2009), école doctorale Math-Info Université Bordeaux 1.
- Sylvain Salvati taught a course at ESSLLI 2009 entitled *An introduction to Abstract Categorial Grammars* (with Ph. de Groote).

8.3. Academic supervision

8.3.1. Student internship supervision – fourth and fifth year

- Renaud Marlet supervised the internship of Diego LLarull (Universidad Nacional de Rosario, Argentine) on *Semantic analysis and inference for plural*.
- Lionel Clément and Christian Retoré supervised the internship of Paul Brunet (ENS Cachan) on *LFG and Glue Semantics*.

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 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

- Christian Bassac gave talk on *Philosophy, linguistics, and semantic interpretation* at the International Conference on Philosophy of Language and Linguistics in May 2009.
- Christian Bassac gave a talk entitled *Les grammaires catégorielles* during the Symposium organized for the Doctorat Honoris Causa of Sir John LYONS in April 2009.
- Christian Bassac gave a talk on *Représentation linguistique des événements* at the colloque international "Evénement, événements et sous-événements" in September 2009.
- Pierre Bourreau presented his work on *Characterizing unique inhabited types* at the Workshop on Lambda Calculus and Formal Grammars in March 2009.
- Lionel Clément gave a talk on *The treatment of collocations in XLFG* during the PERGRAM ANR-meeting in March 2009.
- Lionel Clément gave a talk on *Grammaire d'erreur - correction grammaticale avec analyse profonde et proposition de corrections minimales* at TALN in June 2009.
- Lionel Clément gave a talk on *Zeugmes sémantiques* in a workshop held at Université Bordeaux 3 in October 2009.
- The relational programming methodology was presented by G. Huet at the Journées Francophones des Langages Applicatifs, Saint-Quentin sur Isère (February 2009), at the Journées au vert du projet Signes, Saintes (May 2009), at a seminar at the Department of Computer Science, University of Texas at Austin (October 2009) and at the Department of Computer Science, Jawaharlal Nehru University, New Delhi (December 2009).
- Alain Lecomte gave talk on *Ludics and Its Applications to Natural Language Semantics* at WoLLIC 2009.
- Alain Lecomte gave talk on *Ludics and Its Applications to Natural Language Semantics* at CoRR 2009.
- Alain Lecomte gave a talk during the Colloquium "Logique, Interaction et Géométrie de la Cognition" LIGC 2009, held in honour of the 60th birthday of Michele Abrusci in Florence.
- Renaud Marlet gave a talk on *A Grammar Correction Algorithm - Deep Parsing and Minimal Corrections for a Grammar Checker* at FG 2009.
- Christian Retoré presented his work entitled *Vers une modélisation logique de certains aspects de la sémantique lexicale* during the first CAuLD meeting in June 2009.
- Christian Retoré presented his work on *Une intégration de la sémantique lexicale à la sémantique compositionnelle montagovienne* at the LPL Aix-en-Provence seminar in November 2009.
- Christian Retoré presented his work entitled *Towards a Type-Theoretical Account of Lexical Semantics* in December 2009 at the Cauld Workshop held in Nancy.
- Sylvain Salvati presented his work on *Free Word Orders and Formal Grammars* at the Workshop on Lambda Calculus and Formal Grammars in March 2009.
- Sylvain Salvati has been invited to give a talk at WoLLIC 2009 on his work on recognizable sets of lambda-terms.
- Sylvain Salvati has given an introductory course at ESSLLI 2009 on Abstract Categorial Grammars.

- Sylvain Salvati has been invited at the National Institute of Informatics in Tokyo to present his work on the MIX languages in September 2009.
- Sylvain Salvati has been invited to present his work on recognizable sets of lambda-terms at the seminar of the Computer Science Department of the University of Manchester in October 2009.
- Sylvain Salvati gave a talk on *Modèles finis et résolution d'anaphores* at the second CAuLD meeting in October 2009.

8.4.2. Participation to conferences and summer schools

- Alain Lecomte and Sylvain Salvati participated to WoLLIC 2009
- Pierre Bourreau, Alain Lecomte, Bruno Méry, Richard Moot, Christian Retoré, Sylvain Salvati, Natalia Vinogradova and Emilie Voisin participated to ESSLLI 2009
- Alain Lecomte, Renaud Marlet and Sylvain Salvati participated to Formal Grammars 2009.
- Pierre Bourreau, Alain Lecomte, Bruno Méry, Richard Moot, Christian Retoré, Sylvain Salvati, Natalia Vinogradova participated to the 7th Workshop on Lambda Calculus and Formal Grammars - March 2009
- Pierre Bourreau, Anaïs Lefeuvre, Bruno Méry, Richard Moot, Christian Retoré, Sylvain Salvati and Natalia Vinogradova participated to the Workshop on Logical Methods for Discourse - Décembre 2009

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, http://hal.archives-ouvertes.fr/hal-00413217/PDF/zeugmes_semantiques-2.pdf, à paraître.

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

[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", vol. 2, n^{o} 1, 2004, 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, 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/.

Year Publications

Articles in International Peer-Reviewed Journal

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

[14] C. BASSAC, B. MERY, C. RETORÉ. *Towards a Type-Theoretical Account of Lexical Semantics*, in "Journal of Logic Language and Information", 2009, http://hal.inria.fr/inria-00408308/en/.

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

[16] P. LOUSTAU, T. NODENOT, M. GAIO. *Design principles and first educational experiments of PIIR, a platform to infer geo-referenced itineraries from travel stories*, in "Interactive Technology and Smart Education", vol. 6, n^{o} 1, 10 2009, p. 23 - 39, http://hal.archives-ouvertes.fr/hal-00452668/en/.

[17] C. RETORÉ. *Note de lecture sur Two-Step Approaches to Natural Language Formalisms de Frank Morawietz*, in "Traitement Automatique des Langues", vol. 48, n^{o} 3, 2009, p. 227–229, http://hal.archives-ouvertes.fr/hal-00306394/en/.

[18] C. RETORÉ, S. SALVATI. *A Faithful Representation of Non-Associative Lambek Grammars in Abstract Categorial Grammars*, in "Journal of Logic Language and Information", 2009, p. 45–54, http://hal.inria.fr/inria-00409557/en/.

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

International Peer-Reviewed Conference/Proceedings

[20] C. BENZITOUN, A. DISTER, K. GERDES, S. KAHANE, R. MARLET. *annoter du des textes tu te demandes si c'est syntaxique tu vois*, in "28th International Conference on Lexis and Grammar (LGC 2009), Bergen Norvège", Arena Romanistica, vol. 4, Presses de l'Université de Bergen, 2009, p. 16–27, http://hal.inria.fr/inria-00403626/en/.

[21] M.-N. BESSAGNET, E. KERGOSIEN, M. GAIO. *Extraction de termes, reconnaissance et labellisation de relations dans un thésaurus*, in "Patrimoine 3.0 CIDE'12: 12e Colloque International sur le Document Electronique, Montréal Canada", Europia, 10 2009, p. 275-286, http://hal.archives-ouvertes.fr/hal-00451960/PDF/CIDE_MNBEKMG.pdf.

[22] L. CLÉMENT, K. GERDES, R. MARLET. *A Grammar Correction Algorithm – Deep Parsing and Minimal Corrections for a Grammar Checker*, in "14th conference on Formal Grammar (FG 2009), Bordeaux France", 2009, http://hal.inria.fr/inria-00396240/en/.

[23] L. CLÉMENT, K. GERDES, R. MARLET. *Grammaires d'erreur – correction grammaticale avec analyse profonde et proposition de corrections minimales*, in "16e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2009), Senlis France", 06 2009, electronique, http://hal.inria.fr/inria-00396229/en/.

[24] E. KERGOSIEN, M. KAMEL, C. SALLABERRY, M.-N. BESSAGNET, N. AUSSENAC- GILLES, M. GAIO. *Construction et enrichissement automatique d'ontologie à partir de ressources externes*, in "JFO'09 JFO'09: 3es Journées Francophones sur les Ontologies, Poitiers France", 12 2009, p. 1-10, http://hal.archives-ouvertes.fr/hal-00451980/PDF/paper21_jfo2009.pdf.

[25] A. LECOMTE, M. QUATRINI. *Ludics and its Applications to natural Language Semantics*, in "16th Workshop on Logic, Language, Information and Computation", H. ONO, M. KANAZAWA, R. J. G. B. DE QUEIROZ (editors), Lecture Notes in Artificial Intelligence, vol. 5514, Springer, 2009, p. pp 242–255, http://hal.archives-ouvertes.fr/hal-00422680/en/.

[26] T. NGUYEN VAN, M. GAIO, C. SALLABERRY. *Recherche de relations spatio-temporelles : une méthode basée sur l'analyse de corpus textuels*, in "TIA 2009 : 8th International Conference on Terminology and Artificial Intelligence TIA'09WS: Acquisition et modélisation de relations sémantiques, Toulouse France", 11 2009, p. 1-6, http://hal.archives-ouvertes.fr/hal-00452005/PDF/tia_2009_preversion.pdf.

[27] D. PALACIO, C. SALLABERRY, M. GAIO. *Normalizing Spatial Information to Better Combine Criteria in Geographical Information Retrieval*, in "ECIR GIIW: 31st European Conference on Information Retrieval, Geographic Information on the Internet Workshop (GIIW) ECIR, 31st European Conference on Information Retrieval, Toulouse France", 04 2009, p. 37-49, http://hal.archives-ouvertes.fr/hal-00451924/PDF/article_v13_eng_wgii.pdf.

[28] C. SALLABERRY, A. ROYER, P. LOUSTAU, M. GAIO, T. JOLIVEAU. *GeoStream: Spatial Information Indexing Within Textual Documents Supported by a Dynamically Parameterized Web Service*, in "OGRS 2009: International Opensource Geospatial Research Symposium OGRS 2009: International Opensource Geospatial Research Symposium, Nantes France", Springer, 07 2009, 14, http://hal.archives-ouvertes.fr/hal-00451949/PDF/ogrs.pdf, GEOTOPIA CNRS.

[29] S. SALVATI. *Recognizability in the Simply Typed Lambda-Calculus*, in "16th Workshop on Logic, Language, Information and Computation", H. ONO, M. KANAZAWA, R. J. G. B. DE QUEIROZ (editors), Lecture Notes in Artificial Intelligence, vol. 5514, Springer, 2009, p. 48–60, http://hal.inria.fr/inria-00412654/en/.

Scientific Books (or Scientific Book chapters)

[30] C. BASSAC. *Evénements et référence*, in "Res per nomen", J. P. PIERRE FRATH (editor), Epure, 03 2009, p. 177–194, http://hal.archives-ouvertes.fr/hal-00418581/en/.

References in notes

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

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

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

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

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

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

[37] 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.

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

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

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

[41] 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.

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

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

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

[45] 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.

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

[47] M. KANAZAWA. *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.

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

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

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

[51] 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.

[52] 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.

[53] D. LILLO-MARTIN, E. S. KLIMA. *Pointing out differences: ASL pronouns in syntactic theory*, in "Theoretical issues in sign language research – Vol 1 Linguistics", S. D. FISHER, P. SIPLE (editors), University of Chicago Press, 1990, p. 191–210.

[54] D. LILLO-MARTIN. *Universal Grammar and American Sign Language: Setting the Null Argument Parameters*, Kluwer, 1991.

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

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

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

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

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

[60] 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.

[61] C. NEIDLE, J. KEGL, D. MACLAUGHLIN, B. BAHAN, R. G. LEE. *The Syntax of American Sign Language – Functional Categories and Hierarchical Structure*, MIT Press, 2000.

[62] 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.

[63] 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, n^{o} 2099, Springer-Verlag, 2001, p. 17–43.

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

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

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

[67] 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.

[68] V. SUTTON. *Lessons in SignWriting*, Deaf Action Committee for SignWriting, La Jollla, CA, 2002, http://www.signwriting.org.

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