



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team SIGNES

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

Futurs

THEME SYM

Activity
R *eport*

2006

Table of contents

1. Team	1
2. Overall Objectives	2
2.1. Overall Objectives	2
3. Scientific Foundations	2
3.1. The center: natural language syntax and semantics	2
3.2. Word structure and automata: computational morphology	3
3.3. Sentence structure and formal grammars: syntax	4
3.4. Sentence structure and logic: the syntax/semantics interface	5
3.5. Lexical semantics and derivational morphology	5
3.6. Discourse and dialogue structure: computational semantics and pragmatics	6
3.7. Type systems and functional programming for computational linguistics	6
4. Application Domains	7
4.1. Sanskrit philology	7
4.2. Towards French Sign Language (LSF) modelling and processing	7
5. Software	7
5.1. The Zen toolkit	7
5.2. Sanskrit Site	8
5.3. Grail 3: natural language analysis with multimodal categorial grammar	9
5.4. DepLin	9
5.5. Corpus Arborator	9
5.6. LeFFF	10
5.7. XLFG	10
5.8. French Sign Language HD corpus	10
5.9. Generative Lexicon	10
5.10. Lexed	10
5.11. Efficient and robust LFG parsing: SxPipe and SxLfg	11
5.12. Yab	11
5.13. Tokenizer	11
5.14. Tree-drawing package	11
5.15. Experiments in categorial grammars	11
5.16. Acquisition of linguistic corpora	12
5.17. Farsi corpus	12
5.18. Speech synthesis	12
6. New Results	12
6.1. Flexional morphology and regular relations	12
6.2. Zeugmas in Lexico-Functional Grammar	12
6.3. Efficient and robust LFG parsing: SxPipe and SxLfg	13
6.4. Development of syntactic lexica (the Lefff and its followers)	13
6.5. Categorial grammar construction and parsing	13
6.6. Linguistic modelling within minimalist grammars	14
6.7. Linguistic modelling of French syntax	14
6.8. Generative capacity of minimalist grammars	14
6.9. Pronominal anaphoras and binding theory	15
6.10. Grammatical inference and partly automated grammar construction	15
6.11. Lexical semantics and derivational morphology	15
6.12. Modelling of French Sign Language grammar	16
7. Other Grants and Activities	16
7.1. Regional research programs	16
7.1.1. Traitement Informatique de la langue des Signes Française	16

7.2. National research programs	17
7.2.1. Groupement de Recherche C.N.R.S. 2521 Sémantique et modélisation	17
7.2.2. ARC Mosaïque	17
7.2.3. ANR blanche PRELUDE	17
7.2.4. ILF Project LexSynt	17
7.3. Associate research team	17
8. Dissemination	18
8.1. Activism within the scientific community	18
8.1.1. Honours	18
8.1.2. Editorial boards	18
8.1.3. Program committees of conferences	18
8.1.4. Academic committees	18
8.1.5. Organization of events	19
8.2. Teaching	19
8.3. Thesis Juries	20
8.4. Academic supervision	20
8.4.1. Student internship supervision – fourth and fifth year	20
8.4.2. PhD supervision	20
8.5. Participation to colloquia, seminars, invitations	21
8.5.1. Visiting scientists	21
8.5.2. Seminar Talks, Invitations	21
8.5.3. Participation to conferences and summer schools	22
9. Bibliography	23

1. Team

Joint team with LaBRI and the Department of Linguistics of Université Bordeaux 3 Michel de Montaigne — in particular with the research ministry Jeune Equipe JE 2385 TELANCO and the C.N.R.S. UMR 5610 ERSS.

LaBRI is a joint C.N.R.S. UMR 5800 team, involving Université Bordeaux 1, and the Ecole Nationale Supérieure d'Electronique, d'Informatique, et de Radiocommunications de Bordeaux, ENSEIRB.

ERSS is a joint C.N.R.S. team involving Université Toulouse-Le Mirail and Université Michel de Montaigne in Bordeaux

TELANCO is a team of Université Michel de Montaigne Jeune Equipe 2385 from the Research Ministry

Head of Team

Christian Retoré [Professor, Université Bordeaux 1, HdR]

Vice-Head of Team

Christian Bassac [Lecturer, Université Bordeaux 3 – on INRIA sabbatical, HdR]

Administrative Assistants

Corinne Brisset [TR, INRIA – till August 2006]

Sylvie Embolla [TR, INRIA – from September 2006]

Research Scientists (affiliation)

Lionel Clément [Lecturer, Université Bordeaux 1]

Irène Durand [Lecturer, Université Bordeaux 1, HdR]

Patrick Henry [IR C.N.R.S. LaBRI]

Renaud Marlet [CR INRIA]

Richard Moot [CR C.N.R.S. , LaBRI]

Henri Portine [Professor, Université Bordeaux 3, HdR]

Benoît Sagot [CR INRIA – From July 2006]

Research Scientists (partner)

Kim Gerdes [Lecturer, Université Paris 3 — Affiliated until September 2006 (Lecturer, Université Bordeaux 3)]

Gérard Huet [DR INRIA, HdR]

Alain Lecomte [Professor, Université Grenoble 2, HdR]

Pierre Guitteny [Interpret/Official representative of *Direction Régionale des Affaires Sanitaires et Sociales* Aquitaine — Université Bordeaux 3]

Research Scientist (visitor)

Reinhard Muskens [Poste Rouge C.N.R.S. From Décembre 2006 to May 2007]

Post-doctoral fellows

Roberto Bonato [Temporary Lecturer, Université Bordeaux 1 – Till July 2006]

Marie-Laure Guénot [Temporary lecturer, Université Bordeaux 3 – From September 2006]

Greg Kobele [INRIA Postdoc From December 2006]

Tristan van Rullen [INRIA Postdoc From December 2006]

Ph. D. Students

Maxime Amblard [Temporary lecturer, ENSEIRB]

Houda Anoun [Ministry grant, Université Bordeaux 1]

Bruno Mery [Ministry grant, Université Bordeaux 1 — From October 2006]

Emilie Voisin [Aquitaine/INRIA grant, Université Bordeaux 3]

Internship

Cyril Cohen [ENS Cachan / INRIA Internship — From July to August 2006]

2. Overall Objectives

2.1. Overall Objectives

The Signes team is addressing several domains of computational linguistics such as:

- flexional and derivational morphology
- syntax
- logical (or predicative) semantics
- lexical semantics
- discourse representation

by means of formal methods such as:

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

Two applications illustrate this approach:

- natural language tools for Sanskrit
- modelling of French Sign Language grammar

We also develop the corresponding computational linguistics tools. Ultimately these tools will result in a significant generic NLP platform encompassing analysis, generation and acquisition devices. Some specific languages will deserve particular attention, like Sanskrit, French Sign Language, French.

3. Scientific Foundations

3.1. The center: natural language syntax and semantics

Keywords: *NLP, computational linguistics, formal languages, logic, natural language processing.*

Since the early days of computer science, natural language is both one of its favorite applicative field and the source of technical inspiration, as exemplified by the relation between formal language theory and linguistics. [52]

Nowadays, the motivation is the need to handle lots of digitalized textual and even spoken information, in particular on the Internet, but also interesting mathematical and computational questions raised by computational linguistics, which can lead to other applications.

Most common natural language tools are information retrieval systems, spell checkers, and in a lesser proportion, natural language generation, automatic summary, computer aided translation.

Statistical methods and corpus linguistics [80] have been quite successful for the last years, but there is a renewal of symbolic methods, and especially of logical ones, because of the advances in logic, the improvement of computer abilities for these rather slow algorithms, and overall the need for systems which handle the meaning of phrases, sentences, or discourses. [54]

For all these applications, like queries in natural language, refined information retrieval, natural language generation, or computer aided translation, we need to relate the syntax of an utterance to its meaning. This relation, known as the syntax/semantics interface and its automatization, is the center of this project. This notion is in general used for sentences, but we also work on the extension of this correspondence to discourse and dialogue.

The study of the interface between syntax and semantics makes way for interesting questions of a different nature:

- As said above, this enables applications that require access and computation of meaning.
- Up to now semantics only plays a minor role in Natural Language Processing although a linguistic viewpoint, the two sides of the linguistic signs its *signifiant* and *signifié* are a central subject ever since Saussure. The linking of the observable part of the sign or of the sentence and its meaning, is a constant question in linguistics both in Chomsky's Generative Grammar or in the Meaning-Text theory of Mel'cuk. [82], [59]
- From a mathematical and algorithmic viewpoint, this interface is the place of some challenges: what is the link between two of the main frameworks, namely generative grammars and categorial grammars? The first ones are exemplified by Tree Adjoining Grammars TAGs [70] or Minimalist Grammars [101]. They enjoy efficient parsing algorithms and a broad covering of syntactic constructs. The second ones (see e.g. [84]) are less efficient but provide more accurate analyses. Indeed these latter systems are used for syntax as well as for logical or predicative semantics like Montague semantics [53], [61] and thus allows generation algorithms. Other models, like dependency grammars, [81] provide a different account of the syntax/semantics interface. A comparison between the dependency model and a generative/logical one enables an assessment of the adequation of these families of models, and this is one of the main challenges of contemporary formal linguistics.

At one end of our spectrum stands morphology, and as often in generative grammar, we consider it as part of syntax. It should be nevertheless observed that the computational models involved in the processing of morphology are of different aspects : finite state automata, regular transducers, etc. [73], [74]

At the other end, on the semantical side, we do not consider ontological aspects of semantics, or lexical semantics, but rather extend the logical semantics to discourse and dialog. This is usually done by Discourse Representation Theory [72], which is topdown, incremental and involves state changes.

3.2. Word structure and automata: computational morphology

Keywords: *finite state automata, morphology, transducers.*

Participants: Gérard Huet, Kim Gerdes.

Computational models for phonology and morphology are a traditional application of finite state technology. [73], [74], [75], [55] 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. [80]

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.

3.3. Sentence structure and formal grammars: syntax

Keywords: *categorial grammars, dependency grammars, formal grammars, lexical-functional grammars, minimalist grammars, property grammars, tree adjoining grammars.*

Participants: Maxime Amblard [correspondant], Roberto Bonato, Lionel Clément, Irène Durand, Kim Gerdes, Marie-Laure Guénot, Alain Lecomte, Renaud Marlet, Richard Moot, Christian Retoré, Benoît Sagot.

While linear structure is in general sufficient for morphological structure, trees are needed to depict phrasal structure, and, in particular, sentence structure. Different families of syntactic models are studied in *Signes*: rewriting systems of the Chomsky hierarchy, including tree grammars, and deductive systems, i.e. categorial grammars.

The former grammars, rewrite systems, have excellent computational properties and quite a 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. [71] Among these classes of grammars let us mention Tree Adjoining Grammars, [69], [70], Minimalist Grammars. [101], [102], [83] — Dependency Grammars share some properties with them but the general paradigm is quite different [82], [60].

Edward Stabler 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 they 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.

They also cover phenomena which go beyond syntax, namely they include 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 minimalist grammars, is that they do not allow for learning algorithms which could automatically construct or enlarge grammars from structured corpuses. 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 using categorial grammars, [84] despite their poor computational properties, and poor linguistic coverage, is that they provide a correspondence between syntactic analyses and semantic representations. This is to 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, one can try to connect them to more efficient and wider formalisms, like minimalist grammars. [77], [76], [99]

A rather new approach to syntax is known as model-theoretic syntax opposed to generative-enumerative syntax, whose advantages have been underlined by Geoffrey Pullum in [94]. Instead of viewing the trees or strings as the closure of some base set of expression, they are viewed as the trees or sets satisfying some set of formulae. Beware that this is a different view of the set of parse trees, that can often be also described as a generative process. The advantages of such a description are not on the parsing algorithms (MSO or Constraint Satisfaction are usually of high complexity) but rather on characterising the language class and possibly describing in linguistically natural way (as opposed to lexical items of lexicalised 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* ought to be mentioned [100]. Uwe Mönnich, Jens Michaelis and Frank Morawietz have obtained a two step description of minimalist grammars that we are studying. [90], [88]

In the constraint style issued from the Prolog-Definite Clause Grammars, Head Phrase Structure Grammar, Construction Grammars and Property Grammars defined as sets of constraints. These later ones introduced by Philippe Blache offer a rather natural way to describe grammar rules and are newly studied by our group. [62], [57]

3.4. Sentence structure and logic: the syntax/semantics interface

Keywords: *Montague semantics, categorial grammars, computational semantics.*

Participants: Maxime Amblard, Roberto Bonato, Alain Lecomte, Renaud Marlet, Richard Moot, Christian Retoré.

Why does there exist a simple and computable correspondence between syntax and semantics in categorial grammars? This is mainly due to the internal functional structure of non-terminals in categorial grammars, which yields a correspondence with semantic formulae and functions. This correspondence between syntactic and semantic categories extends to terms, or analyses because the usual logic in use for typed lambda-calculus is an extension of the resource logic used for syntactic deductions or analyses. [61], [104]

Nevertheless this computational correspondence between syntax and semantics provided by categorial grammars is very limited. Firstly, for the correspondence between syntactic and semantic types to hold, we have to provide words with syntactic types which are *ad hoc*, and even wrong. For instance, why should the type of a determiner depend of the constituent it is involved with? Secondly, the truth-conditional aspect of Montague semantics can be discussed both from a theoretical and from a practical viewpoint. According to cognitive sciences, and even to common sense, it is unlikely that human beings develop all possible interpretations when they process and understand a sentence, and in practice such a construction of all models is definitely untractable. [68] Thirdly, a strict compositional principle does not hold, as the famous Geach examples shows.

In this project we address the first issue, which is a real limit, and the third one, in the next section on discourse. The first point is one of the motivations for studying the syntax/semantics interface for minimalist grammars. Indeed, they are rather close to categorial grammars and resource logic, and using this similarity we are able to extend the correspondence to a much richer grammatical formalism, without having strange syntactic types. [76], [99]

3.5. Lexical semantics and derivational morphology

Keywords: *computational semantics, lexical semantics.*

Participants: Christian Bassac, Patrick Henry, Renaud Marlet, Bruno Mery, Christian Retoré.

The generative lexicon [95] is a way to represent the internal structure of the meaning of words and morphemes. Hence it is relevant, not to say mandatory, for computing the semantic counterpart of morphological operations. The information which depicts the sense of a word or morpheme is organized 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 no matter whether they are syntactically mandatory or optional. The event structure follows [72]. 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, transition, this later 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). This kind of model which relates knowledge representation to linguistic organization is especially useful for word sense disambiguation during (automatic) syntactic and semantic analysis.

3.6. Discourse and dialogue structure: computational semantics and pragmatics

Keywords: *DRT, Montague semantics, computational semantics.*

Participants: Alain Lecomte, Henri Portine.

Montague semantics has 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. [72], [105] 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. [105].

3.7. Type systems and functional programming for computational linguistics

Keywords: *functional programming, logic programming, proof assistant, type theory.*

Participants: Houda Anoun, Roberto Bonato, Gérard Huet, Richard Moot.

The team has developed competences in logic, lambda-calculus. These models are commonly used in computational linguistics :

- An example is categorial grammars, with their parsing-as-deduction paradigm, which use proofs in Lambek calculus or linear logic as syntactic trees.
- Another example is Montague semantics which uses the Church description of higher-order logic, implemented in lambda calculus in order to have the compositionality principle of Frege.
- Finally, Discourse Representation Theory also is logic, in a different syntax, and can be combined with Montague semantics to obtain lambda-DRT.

Consequently it is quite natural to develop tools in programming languages relying on logic and type theory:

- The Grail syntactic and semantic parser for Multi Modal Categorial grammars, defined and implemented by Richard Moot, is written in Prolog. This is the most developed and efficient software for categorial grammars, relying on recent development in linear logic, in particular proof nets. [85]
- Under the supervision of Yannick Le Nir and Christian Retoré, a team of students implemented in OCaml the first steps of a platform for parsing and learning categorial grammars and related formalisms. [93]
- Gérard Huet developed a toolkit for morphology, the Zen toolkit, using finite state technology, in OCaml. He obtained excellent performances, thus proving the relevance of *pure* functional programming for computational linguistics. [66]

4. Application Domains

4.1. Sanskrit philology

Keywords: *Indian studies, Internet, Sanskrit, natural language processing.*

Participant: Gérard Huet.

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 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

Keywords: *deaf community, disabled, multimedia communication, sign language.*

Participants: Pierre Guitteny, Renaud Marlet, Henri Portine, Christian Retoré, Emilie Voisin.

After a mundial prohibition decided in 1880 (and which lasted until the sixties in the USA and until the eighties in France) Sign Languages, deaf people can use sign language and rather recently these languages are the object of new studies and development: a first aspect is social acknowledgment of sign language and of the deaf community, a second aspect is 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 objectives 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 [91], [78], [79] — before to be able to apply our methodology, our first task is to determine what the structure of the sentence is, 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 sequence 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 [103] 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 a 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.

5. Software

5.1. The Zen toolkit

Keywords: *computational morphology, finite state technology, functional programming, natural language processing, segmentation.*

Participant: Gérard Huet [correspondant].

This software has been devoped by Gérard Huet for many years, initially in the project-team *Cristal* and it is clearly the most significant software presented in *Signes*.

It is a generic toolkit extracted by Gérard Huet from his Sanskrit modeling platform allowing the construction of lexicons, the computation of morphological derivatives and flexed forms, and the segmentation analysis of phonetic streams modulo euphony. This little library of finite state automata and transducers, called Zen for its simplicity, was implemented in an applicative kernel of Objective Caml, called Pidgin ML. A *literate programming* style of documentation, using the program annotation tool Ocamlweb of Jean-Christophe Filliâtre, is available for Ocaml. The Zen toolkit is distributed as free software (under the LGPL licence) in the Objective Caml Hump site. This development forms a significant symbolic manipulation software package within pure functional programming, which shows the faisability of developing in the Ocaml system symbolic applications having good time and space performance, within a purely applicative methodology.

A number of uses of this platform outside of the Cristal team are under way. For instance, a lexicon of french flexed forms has been implemented by Nicolas Barth and Sylvain Pogodalla, in the Calligramme project-team at Loria. It is also used by Talana (University of Paris 7).

The algorithmic principles of the Zen library, based on the linear contexts datastructure ('zippers') and on the sharing functor (associative memory server), were presented as an invited lecture at the symposium Practical Aspects of Declarative Languages (PADL), New Orleans, Jan. 2003 [64]. An extended version was written as a chapter of the book "Thirty Five Years of Automating Mathematics", edited in honor of N. de Bruijn [63].

5.2. Sanskrit Site

Keywords: *Sanskrit, electronic dictionary, segmentation, tagging.*

Participant: Gérard Huet [correspondant].

Gérard Huet's Sanskrit Site (<http://sanskrit.inria.fr>) provides a unique range of interactive resources concerning Sanskrit philology [67], [65]. These resources are built upon, among other ingredients, the Zen Toolkit (see above). The site registers thousands of visitors monthly.

- 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 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. The process is as follows. The user on his client machine types in a sentence, calls remotely the parser, inspects the small number of surviving taggings, then may inspect each one in order to peruse the semantic analysis, presented as a pseudo-English paraphrase. Some non-determinism may remain — typically, a given segment

may be lemmatized in several ways, either by homonymy, or by morphological ambiguity. Each path in the semantic dependency matrix is shown with its bonus-malus, and the user may select the one he prefers, yielding a completely disambiguated analysis which he may then store on his client machine, 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 LGPL. <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.

Another 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 3: natural language analysis with multimodal categorial grammar

Keywords: *logic programming, parsing, semantic analysis, syntactic analysis.*

Participant: Richard Moot [correspondant].

Within the type-logical grammar paradigm, Multi-Modal Categorial Grammars (MMCG, see e.g. [84]) are one of the richest approach. Richard Moot carefully implemented Grail, an analyzer for MMCG that is the most complete system for natural language analysis based on type logical grammars with lexicon/grammars. Several languages are supported (although with different levels of linguistics coverage): dutch, english, french, italian, hindi. Grail is distributed under Gnu LGPL [86].

The Grail parser/theorem prover for categorial grammars, originally developed at the University of Utrecht, has been rewritten from scratch, taking into account modern insights about proof nets as well as requiring only open-source software to run. This new release also includes computational theoretical improvement in accordance with [85]: parallel use of structural postulates (which introduce flexibility for word order, tree structure etc.) and degree of preference in order to improve the complexity of the analysis due to the exponential number of choices. The parser has also been adapted to allow for a tight integration with the supertagger. Also, several new strategies for reducing the search space have been implemented, significantly improving parsing performance.

5.4. DepLin

Keywords: *natural language syntactic analysis and generation.*

Participant: Kim Gerdes [correspondant].

DepLin takes a syntactic dependency tree as the input. The topological grammar translates such an (unordered) tree to an ordered constituent tree, called topological tree. In the following step, this tree is simplified to a three level prosodic constituent tree (prosodic words, prosodic phrases, prosodic sentences). From this tree, a very simple sound output device can concatenate prerecorded sound files corresponding to the different prosodic words (with their prosodic markup). This allows for auditory tests of the resulting sentences in constructed communicative contexts (question-answer sets). The construction of the prerecorded files is quite time consuming; it has been tested on small vocabulary of Modern Greek.

DepLin was developed by Kim Gerdes. It is distributed as free software (GPL) and, apart from our internal usage at the Signes group (in particular for German and Greek), is mainly used at the University of Paris 7 for the development of different grammars (in particular Arabic and French).

5.5. Corpus Arborator

Keywords: *annotation, corpus, editor, functional dependency.*

Participant: Kim Gerdes [correspondant].

An editor for corpora with functional dependency annotation was developed by Kim Gerdes in collaboration with the ERSS, Toulouse. This “corpus arborator” is distributed under the GPL and used in Bordeaux and ERSS Toulouse.

5.6. LeFFF

Keywords: *French, inflected form, lemma, lexicon, morphological features.*

Participants: Lionel Clément, Benoît Sagot [correspondant].

The *Lefff* (Lexique des Formes Fléchies du Français — Lexicon of French inflected forms) is a large coverage morphological and syntactic lexicon for French. It is freely available under the LGPL for Linguistic Resources. A first version, the *Lefff* 1, was limited to a morphological lexicon of French verbs. It has been developed by Lionel Clément and Benot Sagot (before they joined the Signes team) in collaboration with Bernard Lang, thanks to an original automatic acquisition techniques from raw corpora complemented with manual validation. More recently, the *Lefff* 2 has been released. It includes all categories, as well as fine-grained syntactic information. It is now considered as one of the major lexical resources for French. The *Lefff* 2 is now mostly developed by Benoît Sagot, but previous work by Lionel Clément had been used as starting point. This development is still active, in particular by comparing with other free resources (Dicovaleance, some of the lexicon-grammar tables), by developing new manual or automatic acquisition techniques, and by improving the original lexical model underlying the *Lefff*. The *Lefff* is currently used by many French or international research teams.

5.7. XLFG

Keywords: *LFG, Lexical Functional Grammar, parser.*

Participant: Lionel Clément [correspondant].

XLFG is a parser prototype for research. It implements the Lexical Functional Grammar (LFG) formalism. It used for teaching in various universities. It is distributed as free software. It has been developed by Lionel Clément (before he joined the SIGNES group).

5.8. French Sign Language HD corpus

Participants: Patrick Henry [correspondant], Emilie Voisin.

Patrick Henry and Emilie Voisin built a corpus of LSF which will be used by Signes researchers or others in order to study the grammar of LSF. This original work uses high definition video and uses a specific interface to annotate and extract sequences according to various linguistic criteria.

5.9. Generative Lexicon

Participants: Christian Bassac, Patrick Henry [correspondant].

Patrick Henry and Christian Bassac designed a tool kit for the implementation of a Generative Lexicon which can be shared and used on a network. This tool is in particular designed to filter anaphoric reference in nominal compounds, and other abilities of the generative lexicon will be implemented as well.

5.10. Lexed

Keywords: *dictionary search, lexicaliser.*

Participant: Lionel Clément [correspondant].

Lexed is a lexicaliser. It allows to search a dictionary entry from a string. The finite automata-based algorithm is particularly fast, and offers a good alternative to hashes for large dictionnaires. Lexed is distributed for unix platforms with a GPL Licence. This software has been developed by Lionel Clément (before he joined the SIGNES group).

5.11. Efficient and robust LFG parsing: SxPipe and SxLfg

Participant: Benoît Sagot [correspondant].

Benoît Sagot, before he arrived in the Signes team, has developed with Pierre Boullier (INRIA team Atoll) several software tools which constitute a robust and efficient LFG parsing system for French. Some of these tools rely on Pierre Boullier's system Syntax, which builds very efficient parsers for various formalisms including CFG, TAG, RCG. Some of these tools rely on the Lefff presented above.

This LFG parsing system is divided into three main components. SxPipe is a robust pre-parsing processing chain that transforms any French raw corpus into DAGs of inflected forms known by the Lefff. It performs (among others) named-entities recognition, segmentation, tokenization, spelling-error correction (thanks to the SXSpell component), ambiguous multi-word units identification and ambiguous "light" correction.

The second component is SXLFG, a parser generator for the LFG formalism. Its efficiency allows to parse multi-million-word corpora in a few hours with a large-coverage grammar. Robustness techniques at all levels allow to output relevant and rich information for virtually all sentences, including those which are not strictly recognized by the grammar. Moreover, additional modules (chunker, n-best filtering) have been developed to extend SXLFG's performances and features. The third component is the grammar for French which is given as input to SXLFG so as to generate the parser. All these components are still under strong development, but they already give satisfying results (e.g., state-of-the-art chunking precision).

5.12. Yab

Keywords: *compiler compiler, parsing ambiguities, parsing sharing.*

Participant: Lionel Clément [correspondant].

YAB is a GLR parser generator for S-Attributed grammars. We use this compiler compiler in order to develop a new syntactic parser dealing with homonymies in LFG. We use a specific restriction of the LFG formalism so as to build a polynomial-time syntactic parser.

5.13. Tokenizer

Keywords: *ambiguity, compound words, text segmentation.*

Participant: Lionel Clément [correspondant].

This is a software allowing to segment a text in tokens. Ambiguity between simple and compound words is represented through a direct acyclic graph (DAG). This software has been developed by Lionel Clément (before he joined the SIGNES group) and is part of Lexed (see above).

5.14. Tree-drawing package

Keywords: *Minimalist Grammars, tree drawing.*

Participant: Maxime Amblard [correspondant].

Maxime Amblard developed a tree-drawing package in ML. This package is included as a contribution in the open-source parser for Minimalist Grammars developed and distributed by John Hale (<http://www.linguistics.ucla.edu/people/stabler/hale/index.html>).

5.15. Experiments in categorial grammars

Keywords: *grammatical inference, parsing.*

Participants: Roberto Bonato, Richard Moot [correspondant], Christian Retoré.

This software, *CGTools* is an academic prototype. It is the combination of two *Travaux d'Etude et de Recherche* of 4th year students: Véronique Moriceau et Jérôme Pasquier (Université de Nantes, 2002) which has been reorganized and extended by Thomas Poussevin, Jean-François Deverge, Fahd Haiti, Anthony Herbé (Université Bordeaux 1, 2003). [93] It is written in OCaml, with an interface written in Tcl/Tk and the input and output format are XML files (DAGs for representing analyses, proofs and trees).

Presently, the following algorithms are implemented:

- learning of categorial grammars from structured sentences;
- inter-translation in any possible direction between AB categorial grammars, Lambek grammars, context-free grammars in Greibach normal form, and context-free grammars in Chomsky normal form;
- parsing of categorial grammars by proof search;
- parsing of context-free grammars with the Cocke-Kasami-Younger algorithm.

5.16. Acquisition of linguistic corpora

Participant: Kim Gerdes [correspondant].

Kim Gerdes developed a webrobot specialized in linguistic corpora acquisition with automatic domain and language recognition and recognition of morphological schemas for inflected languages.

5.17. Farsi corpus

Participant: Kim Gerdes [correspondant].

Kim Gerdes with Pollet Samvellian (University Paris 3) developed an automatically annotated large corpus of Farsi.

5.18. Speech synthesis

Participant: Kim Gerdes [correspondant].

Kim Gerdes developed with Cédric Gendrot (University Paris 3) a system converting syntactic dependency into speech synthesis with prosodic structure generation, based on transcribed audiocorpora.

6. New Results

6.1. Flexional morphology and regular relations

In order to validate the transducers constructed by the Zen toolkit, Gérard Huet and Benoît Razet are studying the regular relations which are handled by Zen. Viewing them as a rational part of the product of the input and output alphabet is not satisfactory and one cannot work as for regular languages, since the transitive closure of regular relations is not itself regular. Thus, they are trying to characterize the regular relations handled by Zen, drawing a comparison with regular languages and regular expressions, relying on work by Conway or Eilenberg.

6.2. Zeugmas in Lexico-Functional Grammar

Lionel Clément and Kim Gerdes described a new LFG analysis of coordinations. The goal was to analyse particular cases of coordination known as zeugmas (the use of a word to modify or govern two or more words, usually in such a manner that it applies to each in a different sense, or makes sense with only one): "You held your breath and the door for me." (A. Morissette) or "Le général accuse une défaite et ses amis de le trahir" where the same verb "accuser" or "hold" is used both in a light verb construction and in a full verb construction. Their solution consists in a simple formalisation that allows lexical information to be shared among the conjuncts. They propose the introduction of "lexical capture" in LFG and they add this new operator in the XLFG parser.[37]

6.3. Efficient and robust LFG parsing: SxPipe and SxLfg

During a 3-month stay in the Computer Science Institute of the Polish Academy of Science (Warsaw, Poland), Benoît Sagot extended SxPipe's architecture so as to enable the support of different languages, and developed a Polish version of SxPipe. It has been used to improve the quality of the morphosyntactically annotated corpus developed there. Moreover, Benoît Sagot developed a preliminary LFG grammar for Polish, which is now limited to the level of chunks. Using the Polish SxPipe and the parser generated by SXLFG from this grammar leads to a preliminary parsing system for Polish, which is to be further developed. The Polish team, which has developed an automatic subcategorization frame acquisition technique from chunked corpora, will use the current version of this parsing system to learn a syntactic lexicon for Polish.

The development of SXLFG itself is still very active, thanks to an extensive collaboration of Benoît Sagot with Pierre Boullier (INRIA team Atoll). Since his arrival in the Signes team, new features have been developed including support for huge grammars (hundreds of thousands of rules and terminals), addition of new operators (optional attributes, several kinds of lexical capture inspired from Lionel Clment's work, disambiguation heuristics generated from a high-level description, and others).

6.4. Development of syntactic lexica (the Lefff and its followers)

Several techniques have been developed and/or used this year to improve the *Lefff*. However, since his arrival in the Signes team, Benoît Sagot performed research in two main directions.

First, thanks to a recent improvement in the lexical formalism used by the *Lefff* (leading to a distinction between syntactic functions and their realizations.), it has been made possible to compare the *Lefff* to the two other available resources for French, namely the lexicon-grammar tables (under their SynLex form) and the Proton (now Dicovalence) lexicon. Furthermore, this allowed the exploitation of linguistic information which is available in other resources, in particular for impersonal verbs and frozen verb phrases (in collaboration with Laurence Danlos).

Second, the development of a morphological lexicon for Polish has been initiated, as a preliminary for a future syntactic lexicon. Benoît Sagot developed a morphological description of Polish and extracted a preliminary morphological lexicon from the morphologically-annotated corpus of the Institute of Computer Science of the Polish Academy of Science. This lexicon has been extended thanks to the last version of the automatic acquisition technique for morphological lexica. This allowed to significantly improve the quality of the corpus. Moreover, this lexicon is used by the preliminary parsing system presented above.

Moreover, the development of a morphological lexicon for Slovak has been continued using the same lexical model and architecture. The development of a lexicon for Spanish is considered in the NLP team of the University of La Cornuña.

6.5. Categorical grammar construction and parsing

Richard Moot continued his work on automated supertag extraction from the Spoken Dutch Corpus, adding information about the adjacency of arguments to their functors in the mode information of the lexicon, which increased the size of the extracted lexicon. Thanks to the addition of several corpus search, transformation and error analysis scripts, the extraction algorithm has been improved so that in spite of this larger and more ambiguous lexicon, the supertagging performance has gone up from 77.13% to 80.10% correctly assigned supertags. In addition, work has begun on extracting grammars from the French TALANA treebank.

Grail has been extended to work more tightly in combination with the supertagger and two new strategies for reducing the axiom links have been implemented: one using first-order approximation of packages of structural rules, implementing the ideas of [87] and a second strategy, which uses a context-free grammar to compute the possible relations between unary modalities, as employed by the grammars described in, for example [56] and [27].

Houda Anoun presented the first steps towards the formalization of Arabic syntax and semantics by means of Multimodal Categorical Grammar. Her study particularly focuses on the description of some linguistic phenomena related to nominal sentences construction (e.g., the word-order between the topic and its comment, annexation phenomenon, the use of adjectives ...). These phenomena are captured in an elegant fashion thanks to the use of constrained structural reasoning which is introduced by some relevant packages of lexically anchored structural rules. [35]

6.6. Linguistic modelling within minimalist grammars

Houda Anoun and Alain Lecomte are presently deepening of the previously elaborated model for a logical formulation of Minimalist Grammars. This model belongs to the same family as de Groot's ACGs and Muskens' Lambda Grammars, in that it articulates a level of abstract grammar and levels of concrete grammar (phonology and semantics). The main difference lies in the fact that the basic operations for generating structures are representations of Merge and Move. They use a notion of a constructed syntactic object which contains lexical entries seen as proper axioms and ordered instances of the logical axiom, in such a way that the insertion of a lexical entry depends on the previous introduction of preliminary hypotheses. Hypotheses are playing the role of traces. The difference between overt and covert movements comes from the definition of the phi-terms (the terms which translate the phonological forms). [36]

A particular question, studied by Maxime Amblard, Houda Anoun and Alain Lecomte in this setting is the representation in that model of some ellipsis and coordination phenomena, by means of the use, either of the exponential (!) of Linear Logic, or of two arrows, one for the resource sensitive consumption and the other for the intuitionistic implication. This work was presented in [33]

6.7. Linguistic modelling of French syntax

In [17], Marie-Laure Guénot proposes a new model for French syntax, based on positions constituting a departure from orthodox theoretical assumptions (non-generativist, non-lexicalist, non-modular and multi-dimensional), and formalized into Property Grammars [62].

A part of the PhD work of Maxime Amblard is to construct a fragment of a French MG grammar. Clitic pronouns are one of the difficult questions. He proposes an extension of Stabler's version of clitics treatment for a wider coverage of French. He presents the lexical entries needed in the lexicon and explains the recognition of complex syntactic phenomena such as (left and right) dislocation, clitic climbing over modal verb and extraction from DP. A further step is the presentation of the syntax-semantic interface for clitics analyses and more specifically clitic climbing over verb (including raising verb). He proposes a formalisation of French clitics and interaction of these in complex positions such as negative form, imperative, raising-verbs, control verb etc... This has been partly published in [32].

6.8. Generative capacity of minimalist grammars

In his master thesis, Bruno Mery with Maxime Amblard, Irène Durand and Christian Retoré explored the translation of a minimalist grammar into a weakly equivalent multiple context-free grammar (which allow polynomial parsing), obtaining precise complexity bounds, depending on several parameters of the initial minimalist grammar. [46]

Cyril Cohen, an internship student from ENS Cachan, and Christian Retoré explored [90], [88] which shows that parse trees of Stabler's minimalist grammars can be described as the image by a regular tree transducer of a regular tree language. This provides a model theoretic description in monadic second order logic of the syntactic structures of natural language sentences. Michaelis, Morawietz and Mönnich focused on string languages, and used a translation of minimalist grammars into multiple context free grammars (which leaves out part of the structure) and then into regular tree grammars encoding both the trees and the history of the MCFG or MG derivation. They are then able to recover the binary structure provided by minimalist parse trees. Cyril Cohen and Christian Retoré extended this result by providing a direct translation into the regular tree language and recovering the full minimalist parse trees, including the head/non-head information on internal nodes.

6.9. Pronominal anaphoras and binding theory

Roberto Bonato completed his co-tutored Ph.D. work in the domain of anaphora resolution. He explored a new computational approach to Binding Theory that stems from an article by Reinhart and Reuland of 1993. By stressing the importance of reflexivity over co-reference, Reinhart and Reuland advocates for a more semantically oriented approach to Binding Theory. Roberto Bonato's work gets inspiration from their approach to integrate it in an original framework insights stemming from approaches to Binding Theory both the classical theory of Chomsky [58] and the bound variable reading of Reinhart and Reuland [96], [97], [98] [15]

6.10. Grammatical inference and partly automated grammar construction

Roberto Bonato also published an internal INRIA report [45] that summarizes his previous work at IRISA (Rennes) in the domain of grammar inference and machine learning. This work describes a formal model for first language acquisition as it was first proposed by E.M. Gold in 1961. The basic idea is to provide a formal framework that allows to explore the issue of learnability for formal languages, that is the possibility to infer the rules of a grammar on the basis of sentences belonging to the language generated by that grammar. Special attention is devoted to the learnability proofs of Makoto Kanazawa, for which he provides a more didactic exposition. Furthermore, he extends Kanazawa's results to prove learnability in Gold's model for rigid Lambek grammars from structured examples.

On the same topic, the participants of the INRIA cooperative research action (ARC) Gracq (Lille, Nantes, Nancy, Rennes — 2001:2002) are publishing a survey of the results that they obtained in the Gold learning paradigm, underlining the interest of such results for natural language processing and for the modelling of human language acquisition. [20].

6.11. Lexical semantics and derivational morphology

Over the past year Christian Bassac carried on his work in the Generative Lexicon theory with a view to providing a formal explanatory account of various linguistic phenomena in the domain of morphology. Some of his results were presented papers that just appeared in [18] and in [19], and are gathered with accepted papers in his recent Habilitation [14].

In the paper that appeared in *Cahiers de grammaire*, Christian Bassac offered an explanatory account for the morphology of appositive coordination in English based on a fine-grained representation of lexical items, which allows formal operations on sets driven by the function of the coordinates. Related empirical phenomena such as the presence of a numeral follow from these and from pragmatic maxims.

In a paper to appear in *Generative approaches to the lexicon* Christian Bassac showed that French and Turkish telic compounds offer interesting generalisations regarding their lexicalisations, various syntactic properties such as anaphoric reference or coercion, and the concurrence between two forms of expression of the telic in Turkish, as well as the impossibility of a form of telic whose modifier encodes a resultative state. These various phenomena are accounted for by the representations allowed by a complex telic role.

In another accepted paper on deverbal nouns Christian Bassac showed that both the semantics of some Turkish deverbal nouns and some properties such as gaps in productivity or selection by an irrealis predicate (a predicate which does not hold, as grammatical moods allows) of the matrix clause readily follow from the semantic information encoded in the qualia structure.

Renaud Marlet worked on the formalisation of meaning inference from text, in the Generative Lexicon framework. He proposed a general approach to inject lexical semantics information into a "classic" compositional analysis based on logic (e.g., in a Montague style). The idea is to synchronize the logical analysis with the key composition mechanisms of the Generative Lexicon. Currently covered mechanisms are type coercion and selective binding.

Renaud Marlet also started a comparative study of the main approaches to semantic lexicons, in particular the Generative Lexicon and the DEC (Dictionnaire Explicatif et Combinatoire). The comparison identifies differences in information organization and in composition models. It studies pros and cons of the two approaches, which leads to propositions to try to get the best of both worlds.

6.12. Modelling of French Sign Language grammar

Pierre Guitteny completed his PhD on the passive diathesis in French sign language. He first drew a distinction between LSF and signed French (a pidgin, that is a kind of word to sign translation from French without respecting the structure of sign language nor its inflections) and between LSF and coverbal gesture. Using the iconicity framework introduced by Christian Cuxac which is, according to him, modified by linguistic processes, into morphologic, syntactic, pragmatic phenomena.

Defining passive as a demotion of the first argument, the subject, as Claude Muller does [89], Pierre Guitteny retained two criteria. For utterances using personal transfer and featuring a transitive verb, the demotion of the subject is obtained by the patient taking the role in the ongoing personal transfer. For utterances without personal transfer the place of the agent remains unspecified or is specified afterwards with a correlating mark. He thus makes a distinction between passive and imperative constructs and between passive and impersonal constructs. He also made a distinction between passive and inverse, the later one changing the viewpoint without changing the focus. His description can take place in an iconicist grammar and less easily in more standard grammatical framework because of phenomena like transfers, simultaneity or spacial realisation. [16]

Pierre Guitteny also published a study on the difference between sign language and signed French (a kind of word to sign translation from French) that he describes as a pidgin and the social impact of this continuum between these two languages on deaf identity and culture. [24]

Emilie Voisin and Loïc Kervajan proposed a typology for the verbal system in French Signed Language based on the work of Anne-Marie Parisot [92]. Then they examined some verbs called in their typology *flexible verbs*: such a verb can have inflection like proform or locus. A flexible verb can, in a particular context, not realise the inflection and we try to describe this context to explain this particularity. [29]

Emilie Voisin presented a communication is about non finite verb forms in French Sign Language (LSF). She tries to determine whether they follow a system of regular patterns and to locate them. To do so, she first proposes the relevant definitions, both general and specific to LSF, thus providing a theoretical background. Next, she points out in an LSF corpus signed by native deaf speakers, examples which matches her definition of non finite verbs. Finally, she proposes some hypotheses accounting for the observed phenomena and concludes with proposals to handle other similar verbal constructs. [31]

Studies on signed languages are often biased because the observations often modify the signed language which spoken and being studied (e.g. LSF). Emilie Voisin brought a reflection about the concept of grammaticality, in particular for the study of signed languages. How to take into account Signed French? We propose to treat Signed French as an asymmetric mean of communication. [44]

Emilie Voisin extended this earlier work of hers, and showed, thanks to some examples signed by a native deaf speaker, that there are many possible sign/word orders for a given sentence, like “the boy eat an apple” (in LSF) and several different representation of the verb “to eat”, with or without inflection. This lead her to reconsider the status of Signed French (without inflection, see above) and to analyse the border between verbal and nominal categories. [30]

7. Other Grants and Activities

7.1. Regional research programs

7.1.1. *Traitement Informatique de la langue des Signes Française*

Participants: Pierre Guitteny, Renaud Marlet, Richard Moot, Henri Portine, Christian Retoré [correspondant], Emilie Voisin.

The region Aquitaine is funding (together with INRIA and University Bordeaux 3) a PhD grant on the same topic. Given an accurate video recorder and corresponding software and computer, our team should be able to constitute a very good quality corpus of spontaneous sign language speech as well as guided experiments.

7.2. National research programs

7.2.1. *Groupe de Recherche C.N.R.S. 2521 Sémantique et modélisation*

Participants: Maxime Amblard, Houda Anoun, Christian Bassac, Roberto Bonato, Lionel Clément, Marie-Laure Guénot, Patrick Henry, Greg Kobele, Alain Lecomte, Reinhard Muskens, Renaud Marlet, Bruno Mery, Richard Moot, Henri Portine, Christian Retoré, Emilie Voisin.

Signes is one of the fifteen research team of the Groupe de Recherches 2521 (C.N.R.S.) *Sémantique et Modélisation* directed by Francis Corblin (Université Paris IV) 2002-2005, 2005-2008. This research program is divided into *Opérations: 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* team is part of the later two operations, which could be translated as *Interfaces of linguistic semantics* and *Computational semantics*.

7.2.2. *ARC Mosaïque*

Participants: Lionel Clément [correspondant], Kim Gerdes, Marie-Laure Guénot, Renaud Marlet, Richard Moot, Christian Retoré, Benoît Sagot.

Mosaïque directed by Lionel Clément is a two year action (ARC 2006-2007) lead by *Signes* and involving three other INRIA teams (Atoll, Calligramme, Langue et Dialogue) and four CNRS laboratories (LINA, LLF, LPL, Modyco) *Mosaïque* is aimed at designing high-level syntactic formalisms. The idea is to dissociate the description level (with something like Meta-Grammars) from the target operational formalisms (which may be TAG, LFG, HPSG, or something else) in order to reuse different syntactic descriptions and to develop a high-level syntactic description editor for the linguists. The hypothesis is that many existing formalisms share a lot of things, even if they look different in a technical way.

<http://mosaique.labri.fr/>

7.2.3. *ANR blanche PRELUDE*

Participants: Maxime Amblard, Houda Anoun, Alain Lecomte [participants], Reinhard Muskens, Bruno Mery, Richard Moot, Christian Retoré [correspondant].

Signes is part of the national research program *PRELUDE Towards a theoretical pragmatic based on ludics and continuations* (November 2006 - November 2008) directed by Alain Lecomte and its laboratory *Structures Formelles de la Langue*. Other partners are the INRIA team Calligramme and the Institut Mathématique de Luminy.

7.2.4. *ILF Project LexSynt*

Participants: Lionel Clément, Kim Gerdes, Marie-Laure Guénot [correspondant], Renaud Marlet, Benoît Sagot.

LexSynt is a research action (2000-2006) involving 6 ILF (Institut de la Langue Française) teams (ATILF, ERSS, IGM, LPL, Lattice, MoDyCo), 4 INRIA teams (Atoll, Caligramme, Langue et Dialogue, *Signes*), 2 foreign teams (ALA - KU Leuven, OLST - Montréal) and one industrial partner (Lexiques pour le TAL). The aim is to federate the different existing projects of syntactic lexicon development for French, and to help bridging gaps between the different information represented in these lexicons.

<http://lexsynt.inria.fr/>

7.3. Associate research team

Participant: Gérard Huet [correspondant].

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. http://pauillac.inria.fr/~huet/proposition_eq_ass.html

8. Dissemination

8.1. Activism within the scientific community

8.1.1. Honours

- Gérard Huet is member of the *Académie des sciences* since November 2002.
- Gérard Huet is member of the *Academia Europae* since November 2002.

8.1.2. Editorial boards

- Alain Lecomte is on the editorial board of the journal *TAL – Traitement Automatique des Langues*, Editions Hermès, Paris since august 2001.
- Alain Lecomte and Christian Retoré are on the editorial board of the book series *Research in Logic and Formal Linguistics*, Edizione Bulzoni, Roma, since 1999.
- Henri Portine is on the editorial board of the journal *ALSIC – Apprentissage des Langues et Systèmes d'Information et de Communication*
- Christian Retoré is reviewer for *Mathematical Reviews* since october 2003.
- Christian Retoré is editor in chief of the journal *TAL – Traitement Automatique des Langues*, Editions Hermès, Paris since April 2004. (in the editorial board since 2001).

8.1.3. Program committees of conferences

- Christian Bassac is a member of the organisation committee of the *The Second International Conference on the Linguistics of Contemporary English* which will be held in july 2007 in Toulouse.
- Richard Moot was on the reading committee of TALN 2006.
- Christian Retoré is on the program committee of *Traitement Automatique du Langage Naturel 2006* (Leuven)
- Christian Retoré is on the reviewing committee of *Human Language Technology / North American Chapter of the ACL 2006* (New-York)
- Christian Retoré chaired (with David Nicolas) the *Journées de Sémantique et Modélisation 2006* and the *Atelier Logique et Sémantique du Langage Naturel*.

8.1.4. Academic committees

- Christian Bassac is a member of the hiring committee in linguistics of Université Bordeaux 3.
- Gérard Huet is a member of International Advisory Board, NII (National Institute of Informatics), Tokyo, Japan.
- Gérard Huet is a member of the Scientific Committee of the GIS SARIMA.
- Henri Portine is a member of the hiring committees in linguistics of Université Paris 3 and Université Bordeaux 3.
- Henri Portine is an elected member of the board of governors of the Université Bordeaux 3 and of Institut Universitaire de Formation des Maîtres d'Aquitaine.

- Henri Portine is the head of the linguistic and literature faculty of Université Bordeaux 3.
- Henri Portine is the head of the research team *Text, Language, Cognition* JE2385.
- Christian Retoré is a member of the hiring committee in computer-science of Université Bordeaux 1.
- Christian Retoré is a member of the committee of the faculty of mathematics and computer science of the Université Bordeaux 1.
- Christian Retoré is a member of the committee for the SPECIF best PhD award (SPECIF French academic computer-science society).

8.1.5. Organization of events

- Maxime Amblard, Lionel Clément, Marie-Laure Guénot, Patrick Henry, Henri Portine, Christian Retoré Emilie Voisin organised the *Semaine Bordelaise de Sémantique Formelle, SEBOSEFOR* March 27th–31th 2006 including the *Atelier logique et sémantique du langage naturel* and the *Journées de Sémantique et modélisation*.
- Christian Bassac is a member of the organisation committee of the *The Second International Conference on the Linguistics of Contemporary English* which will be held in July 2007 in Toulouse.
- Lionel Clément organized the *Mosaïque* meetings in Bordeaux (March 8th 2006) and Paris (May 9th, July 4th, December 1st 2006).
- Emilie Voisin is organizing the weekly seminar *Linguistique et informatique* Universités Bordeaux 1 et 3.

8.2. Teaching

Since most 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 school for PhD students and colleagues. Let us cite the lectures whose topic is computational linguistics:

- *Méthode logiques pour la syntaxe et la sémantique du langage naturel* at the *Ecole Jeunes Chercheurs en Algorithmique et Calcul Formel*, Bordeaux, May 2006. (Christian Retoré)
- *From Syntactic Structures to Logical Semantics* European Summer School on Logic Language and Information, ESSLLI 2006, Malagà, August 2006. (Christian Retoré and Alexandre Dikovsky).
- *Sémantique de Montague* and *Quantificateurs Généralisés* at the *Atelier logique et sémantique du langage naturel* during the *Semaine Bordelaise de Sémantique Formelle, SEBOSEFOR* (Alain Lecomte and Christian Retoré)
- *Structures Informatiques et Logiques pour la Modélisation Linguistique*, Parisian Master of Research in Informatics (MPRI). (Gérard Huet, Philippe de Groote)
- *Symbolic natural language processing*, Bordeaux 1, 5th year in computer science (Christian Retoré)
- *Utterance acts and semantics*, Bordeaux 3, 5th year in linguistics (Henri Portine)
- *The syntax of Wh-clauses and extraction*, Bordeaux 3, 5th year in linguistics (Christian Bassac)
- *Finite state natural language processing*, Bordeaux 1, 4th year in computer science (Roberto Bonato, Lionel Clément, Christian Retoré)
- *Word order and its formalization*, Bordeaux 3, 4th year in linguistics (Kim Gerdes)
- *Linguistic formalisms*, Bordeaux 3, 4th year in linguistics (Lionel Clément, Kim Gerdes, Renaud Marlet)

8.3. Thesis Juries

- Gérard Huet was on the jury of the PhD of Patrick Thévenin entitled *Vers un assistant la preuve en langue naturelle* defended on 5 December 06 at the Université de Savoie
- Gérard Huet was on the jury (reviewer) of the PhD of Benoît Sagot *Analyse automatique du français : lexicques, formalismes, analyseurs.* defended on 7 April 2006 at the Université Paris 7
- Alain Lecomte was on the jury of the PhD of Sylvain Degeilh entitled *Applications des prégroupes de Lambek au traitement automatique des langues naturelles* defended on 22 November 2006 at Université Montpellier 2.
- Alain Lecomte was on the jury of the PhD of Willemijn Vemaat entitled *The Logic of Variation, a cross-linguistic account of Wh-question formation* defended on 13 January 2006 at Universiteit Utrecht
- Henri Portine was on the jury (supervisor) of the PhD of Pierre Guitteny entitled *Le passif en langue des signes* defended on 13 December 2006 at the Université Michel de Montaigne Bordeaux 3.
- Christian Retoré was on the jury (supervisor) of the PhD of Roberto Bonato entitled *An integrated computational approach to binding theory* defended on 2 May 2006 at the Università degli studi di Verona.
- Christian Retoré was on the jury (supervisor) of the PhD of Erwan Moreau entitled *Acquisition de grammaires lexicalisées pour les langues naturelles* defended on 18 October 2006 at the Université de Nantes
- Christian Retoré was on the jury (reviewer) of the PhD of Sylvain Degeilh entitled *Applications des prégroupes de Lambek au traitement automatique des langues naturelles* defended on 22 November 2006 at Université Montpellier 2.
- Christian Retoré was on the jury of the PhD of Marie-Laure Guénot entitled *Éléments de grammaire du français pour une théorie descriptive et formelle de la langue* defended on 7 December 2006 at the Université de Provence.

8.4. Academic supervision

8.4.1. Student internship supervision – fourth and fifth year

- Christian Retoré supervised the internship of Cyril Cohen on *Minimalist derivations as a regular tree-language*
- Maxime Amblard, Irène Durand and Christian Retoré supervised the master thesis of Bruno Mery *Grammaires légèrement contextuelles pour l'analyse syntaxique du langage naturel*

8.4.2. PhD supervision

- Alain Lecomte is supervising with Pierre Castéran the thesis work of Houda Anoun on *Proof theoretic methods in computational linguistics.* (Université Bordeaux 1)
- Christian Bassac and Christian Retoré are co-supervising the thesis work of Bruno Mery on *Type theory for lexical semantics.* (Université Bordeaux 1)
- Alain Lecomte is supervising the thesis work of Tran Vu Truc *Logique d'informations partielles pour le traitement des implicites.* (Université Grenoble II)
- Alain Lecomte and Christian Retoré are co-supervising the thesis work of Maxime Amblard, *Calcul de représentations sémantiques dans les grammaires minimalistes.* (Université Bordeaux 1)
- Henri Portine and Renaud Marlet are co-supervising the thesis work of Emilie Voisin, *Génération automatique d'énoncés en Langue des Signes Française.* (Université Bordeaux 3)

- Henri Portine supervised the thesis work of Pierre Guitteny, *Le passif en Langue des Signes Française*. (Université Bordeaux 3)
- Christian Retoré and Alexandre Dikovskiy (Université de Nantes) co-supervised the thesis work of Erwan Moreau, *Acquisition de grammaires catégorielles et de grammaires de dépendances*. (Université de Nantes)
- Christian Retoré and Denis Delfitto (Università di Verona) co-supervised the thesis work of Roberto Bonato, *Algorithmes de calcul de représentations sémantiques à partir d'analyses de type générative et algorithmes inverses*. (cotutored PhD Université Bordeaux 1 / Università di Verona)

8.5. Participation to colloquia, seminars, invitations

8.5.1. Visiting scientists

- 27 November 2006 : Philip Miller (Université de Lille 3) "Les verbes de perception, causatifs, de permission et de coercition et l'opposition entre *Montée* et *Contrôle*: syntaxe ou sémantique ?"
- 13 November 2006 Patrice Dalle (Institut de Recherche en Informatique de Toulouse) et Loïc Kervajan (France Télécom Recherche et Développement & Université de Provence): Journée sur la formalisation et l'informatisation des langues signées
- 6 November 2006 : Bruno Courcelle (LABRI, Bordeaux) "Une présentation des grammaires non contextuelles centrée sur les arbres de dérivation"
- During March 2006, Gaëtane Dostie (University of Sherbrooke) visited our team and gave talks on Discourse Markers.
- 06 March 2006 : Thierry Poibeau (Laboratoire d'Informatique de Paris-Nord), "Traitement automatique des emplois métaphoriques des entités nommées".
- 06 February 2006 : Benoît Sagot (INRIA - projet Atoll), "Modélisation et analyse automatique des langues".
- 09 January 2006: Leo Wanner (Universitat Pompeu Fabra, Barcelone), "Extraction and semantic classification of lexical co-occurrences from corpora".

8.5.2. Seminar Talks, Invitations

- Lionel Clément, Kim Gerdes, Marie-Laure Guénot, Renaud Marlet, Richard Moot, Benoît Sagot attended most of Mosaïque meetings in Bordeaux (March 8th 2006) and Paris (May 9th, July 4th, December 1st 2006).
- Lionel Clément, Kim Gerdes, Marie-Laure Guénot, Renaud Marlet, Benoît Sagot attended most of LexSynt meetings in Nancy (January 19th 2006), Bordeaux (March 9th, 2006) and Paris (April 28th, September 28th, July 4th, December 4th 2006).
- Kim Gerdes visited the NLP group at the Technology Department of the Pompeu Fabra University, Barcelona, <http://www.recerca.upf.edu/taln/> in May and June 2007.
- Kim Gerdes gave a talk in the joint MPRI and Talana seminar, Paris: "Topological Syntax, Theory and Applications", in March 2006
- Gérard Huet presented his work on Mathematics, Linguistics, and Types. at the Journée IHPST: Théories modernes des types, Paris in March 2006.
- Gérard Huet presented his work on Parsing Sanskrit by Computer. at the 13th World Sanskrit Conference, Edinburgh in July 2006.
- Gérard Huet presented his work on Vérité Mathématique, cohérence logique et vérification informatique. at the Conférence M.U.R.S., cycle "Qu'est-ce que la vérité scientifique ?", Palais de la découverte, Paris in April 2006.

- Gérard Huet presented his work on Analyse syntaxique superficielle guidée par des contraintes d'accord et de gouvernement. Application au sanskrit. au Séminaire TAL de Nancy in April 2006.
- Gérard Huet presented his work on Shallow syntax analysis in Sanskrit guided by semantic nets constraints. International Workshop On Research Issues in Digital Libraries at IWRIDL December 2006, Calcutta, India.
- Alain Lecomte gave an invited talk on *Labelled Linear Grammars* at Utrecht University, January 2006.
- Alain Lecomte gave a talk on *Logics in Contemporary Sciences* at the seminar *Mathématisation* in the framework of the PRCT "Mathématisation" funded by Région Rhône-Alpes,
- Alain Lecomte gave a talk on *La mathématisation dans les sciences du langage* at the seminar *Mathématisation* in the framework of the PRCT "Mathématisation" funded by Région Rhône-Alpes
- Alain Lecomte gave a talk on *L'esquisse d'une grammaire pure et la linguistique contemporaine* at the *Hommage to Jean-Louis Gardies* organized at Maison des Sciences de l'Homme de Grenoble,
- Pierre Guitteny gave a talk on *Langue des signes, communauté sourde et CMT* at the *Colloque CMT (communication médiatisée par les technologies de l'information et de la communication)*, Université Bordeaux III-Telanco, 2006
- Pierre Guitteny gave a talk on *La loi du 11 février 2005* at the *Université d'été 2LPE (Deux langues pour une éducation)*, 2006
- Pierre Guitteny Loi 2005-102 et interprétariat, Université d'automne de IÕAFILS (Association française des interprètes en langue des signes), 2006
- Christian Retoré was invited two weeks at the Università degli studi di Verona where he gave post-graduate lectures on the syntax and semantics of categorial grammars.
- Christian Retoré gave a talk on *Une présentation rationnelle des grammaires minimalistes* at the Nancy TAL seminar in December 2006.
- Christian Retoré présentés his cursus and the birth of the project team Calligramme for the 20th anniversary of INRIA-Lorraine in December 2006.
- Benoît Sagot gave a talk at the Computer Science Institute of the Polish Academy of Science (Warsaw, Polan) on his ongoing work on Building a morphosyntactic lexicon and a pre-syntactic processing chain for Polish, in August 2006. This talk was part of his 3-month stay in this Institute.
- Emilie Voisin was invited to present her work in France Telecom, Lannion in July 2006.
- Emilie Voisin was invited to work with the team based in Montreal, UQàM, about sign languages (comparison between LSF and LSQ) from October 8th to October 18th.

8.5.3. Participation to conferences and summer schools

- Maxime Amblard, Greg Kobele and Christian Retoré presented their poster on La linguistique computationnelle entre traitement automatique des langues et sciences cognitives at the colloquium of the Association pour la Recherche en Science Cognitive.
- Maxime Amblard, Houda Anoun, Alain Lecomte presented their work on Ellipsis and Coordination in a Type Logical Grammar at the *Journées de Sémantique et Modélisation, JSM'06*
- Maxime Amblard, Houda Anoun, Alain Lecomte, Christian Retoré and Emilie Voisin attended ESSLLI 2006 in Malaga in August 2006.
- Maxime Amblard presented his works on clitics at the Formal Grammar conference in Malaga in July 2006.
- Houda Anoun and Alain Lecomte presented their works on Logical Grammars and Emptiness at the Formal Grammar conference in Malaga in July 2006.

- Houda Anoun presented her works on the logical analysis of nominal sentences in Standard Arabic at the ESSLLI student session in Malaga in July 2006.
- Lionel Clément presented his work with Kim Gerdes on Zeugmas at the LFG conference in Konstanz in July 2006.
- Kim Gerdes presented his work with Sylvain Kahane on A polynomial parsing algorithm for the topological model. Synchronizing Constituent and Dependency Grammars, Illustrated by German Word Order Phenomena at the ACL meeting in Sydney in August 2006.
- Alain Lecomte gave a three hour lecture on *Quantificateurs Généralisés* at the *Atelier logique et sémantique du langage naturel* during the *Semaine Bordelaise de Sémantique Formelle, SEBOSEFOR*.
- Alain Lecomte gave two talks at the the workshop "Logics and Linguistics" in the framework of GEOCAL'06 (Geometry of Calculus) organized at Marseille-Luminy, February, introductory talk: "may we speak of a logic of language?", research talk : "Labelled Linear Grammars - 2" in February 2006.
- Christian Retoré gave a three hour lecture on *Sémantique de Montague* at the *Atelier logique et sémantique du langage naturel* during the *Semaine Bordelaise de Sémantique Formelle, SEBOSEFOR*.
- Christian Retoré with Alexandre Dikovskiy gave a 8-hours lecture on The Interface between Syntax and Semantics at ESSLLI 2006, in Malaga.
- Christian Retoré gave three talks at the the workshop "Logics and Linguistics" in the framework of GEOCAL'06 (Geometry of Calculus) organized at Marseille-Luminy, February, introductory talk: "Logic, Linguistics and Informatics", research talks: "Categorial Minimalist Grammars" and "Handsome non commutative proof-nets". in February 2006.
- Benoît Sagot gave a joint talk with Laurence Danlos on "Les expressions verbales figées en français : Des tables du lexique-grammaire aux applications du TAL" at the Lexicon and Grammar Conference in Palermo, Italy, in September 2006.
- Emilie Voisin presented her work on Langue des Signes Française et Français Signé : frontire et grammaticalité at the Grenoble colloque international des étudiants chercheurs en didactique des langues et en linguistique in Grenoble in July 2006.
- Emilie Voisin presented her work on Les formes verbales non finies en Langue des Signes Française at the Cerlico Symposium on Les formes non finies du verbe in Bordeaux in June 2006.
- Emilie Voisin presented her work on Manifestations syntaxiques de la diglossie : le cas du Français Signé at the Montreal Symposium of PhD student in linguistics in October 2006.
- Emilie Voisin presented her work with Loïc Kervajan on Typologie des verbes et forme verbale non marquée en Langue des Signes Française : incidences sur l'organisation syntaxique at the international colloquium on Syntax, Interpretation and Lexicon in Sign Languages in Lille in June 2006.

9. Bibliography

Major publications by the team in recent years

- [1] C. BASSAC. *Principes de morphologie anglaise*, Linguistica, Presses Universitaires de Bordeaux, 2004.
- [2] C. BASSAC. *Morphologie et Information Lexicale*, Mémoire d'Habilitation à Diriger des Recherche, Université Michel de Montaigne Bordeaux 3, Décembre 2006.

- [3] L. CLÉMENT, A. KINYON. *Generating parallel multilingual LFG-TAG grammars from a MetaGrammar*, in "Proc. of ACL'03", 2003.
- [4] K. GERDES. *Topologie et grammaires formelles de l'allemand*, Thèse de Doctorat, Université Paris 7, 2002.
- [5] G. HUET. *Transducers as lexicon morphisms, phonemic segmentation by euphony analysis, application to a sanskrit tagger*, in "Journal of Functional Programming", 2005, <http://pauillac.inria.fr/~huet/PUBLIC/tagger.ps>.
- [6] A. LECOMTE. *Categorical Grammar for Minimalism*, in "Logic and Grammar", C. CASADIO, P. SCOTT, R. SEELY (editors)., CSLI, 2005.
- [7] R. MOOT. *Proof nets for linguistic analysis*, Ph. D. Thesis, 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] H. PORTINE. *La syntaxe de Damourette et Pichon comme outil de représentation du sens*, in "Modèles linguistiques", vol. 23, n^o 2, 2002, p. 21–46.
- [10] C. RETORÉ. *Logique linéaire et syntaxe des langues*, Mémoire d'habilitation à diriger des recherches, Université de Nantes, Janvier 2002.
- [11] 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.
- [12] B. SAGOT. *Analyse automatique du français: lexiques, formalismes, analyseurs*, Thèse de Doctorat, Université Paris 7, avril 2006.

Year Publications

Books and Monographs

- [13] K. GERDES, C. MULLER (editors). *Ordre des mots et topologie de la phrase française*, vol. 29, 2006, p. 1-10.

Doctoral dissertations and Habilitation theses

- [14] C. BASSAC. *Morphologie et Information Lexicale*, Mémoire d'Habilitation à Diriger des Recherche, Université Michel de Montaigne Bordeaux 3, Décembre 2006.
- [15] R. BONATO. *An Integrated Computational Approach to Binding Theory*, Tesi di Dottorato di Ricerca and Thèse de Doctorat, Università degli Studi di Verona and Université Bordeaux I, May 2006.
- [16] P. GUITTENY. *Le passif en langue des signes*, Thèse de doctorat, Université Michel de Montaigne Bordeaux 3, Décembre 2006.
- [17] M.-L. GUÉNOT. *Eléments de grammaire du français pour une théorie descriptive et formelle de la langue*, Thèse de doctorat, Université de Provence, Aix en Provence, Décembre 2006.

Articles in refereed journals and book chapters

- [18] C. BASSAC. *A compositional treatment for English compounds*, in "Research in Language", september 2006.
- [19] C. BASSAC. *Un problème de morphologie flexionnelle: l'apposition coordinative en anglais*, in "Cahiers de grammaire", september 2006.
- [20] D. BECHET, R. BONATO, A. DIKOVSKY, A. FORET, Y. LE NIR, E. MOREAU, I. TELLIER. *Modèles algorithmiques de l'acquisition de la syntaxe : concepts et méthodes, résultats et problèmes*, in "Recherches Linguistiques de Vincennes", 2007.
- [21] K. GERDES. *Sur la non-équivalence des représentations syntaxiques : comment la représentation en X-barre nous amène au concept du mouvement*, in "Cahiers de grammaire", vol. 30, 2006.
- [22] K. GERDES, S. KAHANE. *L'amas verbal au coeur d'une modélisation topologique du français*, in "Linguisticae Investigationes", vol. 29, 2006, p. 75-89.
- [23] K. GERDES, C. MULLER. *Présentation*, in "Linguisticae Investigationes", vol. 29, 2006, p. 1-10.
- [24] P. GUITTENY. *Langue, pidgin et identité*, in "Glottopol", vol. 7, 2006, p. 128-143, <http://www.univ-rouen.fr/dyalang/glottopol>.
- [25] G. HUET. *Lexicon-directed Segmentation and Tagging of Sanskrit*, in "Themes and Tasks in Old and Middle Indo-Aryan Linguistics, Delhi", B. TIKKANEN, H. HETRICH (editors). , Motilal Banarsidass, 2006, p. 307-325.
- [26] G. HUET. *Vérité Mathématique, cohérence logique et vérification informatique*, in "Cahiers du Mouvement Universel de la Responsabilité Scientifique", vol. 49, n° 3, 2006, p. 42-61.
- [27] R. MOOT, C. RETORÉ. *Les indices pronominaux du français dans les grammaires catégorielles*, in "Linguisticae Investigationes", vol. 29, 2006, p. 137-146.
- [28] C. RETORÉ. *Les mathématiques de la linguistique computationnelle – premier volet: la théorie des langages*, in "La Gazette des Mathématiciens", 2007.
- [29] É. VOISIN, L. KERVAJAN. *Typologie des verbes et forme verbale non marquée en Langue des Signes Française : incidences sur l'organisation syntaxique*, in "Silexicales", 2007.
- [30] É. VOISIN. *Manifestations syntaxiques de la diglossie : le cas du Français Signé*, in "RELQ / QSJL Revue des étudiants en linguistique du Québec / Quebec Student Journal of Linguistics", 2006.
- [31] É. VOISIN. *Les formes verbales non finies en Langue des Signes Française*, in "Les formes non finies du verbe", F. LAMBERT (editor). , Presses Universitaires de Rennes, 2007.

Publications in Conferences and Workshops

- [32] M. AMBLARD. *Treating clitics with minimalist grammars*, in "Formal Grammar, Malaga", 2006.

- [33] M. AMBLARD, H. ANOUN, A. LECOMTE. *Ellipse et coordination dans les grammaires de type logique*, in "Journée Sémantique et Modélisation", 2006.
- [34] M. AMBLARD, G. KOBELE, C. RETORÉ. *La linguistique computationnelle entre traitement automatique des langues et sciences cognitives*, in "Colloque de l'Association pour la Recherche Cognitive", Poster, 2006.
- [35] H. ANOUN. *Towards a logical analysis of nominal sentences in Standard Arabic*, in "European Summer School in Logic Language and Information ESSLLI, student session", 2006.
- [36] H. ANOUN, A. LECOMTE. *Logical Grammars with Emptiness*, in "Formal Grammar", 2006, p. 15–29.
- [37] L. CLÉMENT, K. GERDES. *Analyzing Zeugmas In XLFG*, in "LFG conference", 2006.
- [38] L. DANLOS, B. SAGOT, S.-A. SUSANNE. *French frozen verbal expressions: from lexicon-grammar tables to NLP applications*, in "Actes du Colloque Lexique et Grammaire 2006, Palermo, Italy", 2006.
- [39] K. GERDES, S. KAHANE. *A polynomial parsing algorithm for the topological model. Synchronizing Constituent and Dependency Grammars, Illustrated by German Word Order Phenomena*, in "44th meeting of the Association for Computational Linguistics 2006", 2006.
- [40] G. HUET, B. RAZET. *The Reactive Engine for Modular Transducers*, in "Algebra, Meaning and Computation, Essays Dedicated to Joseph A. Goguen on the Occasion of His 65th Birthday", K. FUTATSUGI, J.-P. JOUANNAUD, J. MESEGUER (editors)., LNCS, vol. 4060, Springer, 2006, p. 355–374.
- [41] 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.
- [42] B. SAGOT, É. DE LA CLERGERIE. *Error mining in parsing results*, in "Proceedings of ACL-CoLing 06, Sydney, Australia", 2006, p. 329–336.
- [43] D. SEDDAH, B. SAGOT. *Modeling and Analysis of Elliptic Coordination by Dynamic Exploitation of Derivation Forests in LTAG parsing*, in "Proceedings of TAG+8, Sydney, Australia", 2006, p. 147-152.
- [44] É. VOISIN. *Langue des Signes Française et Français Signé : frontière et grammaticalité*, in "Actes du colloque international des étudiants chercheurs en didactique des langues et en linguistique, Grenoble", 2006.

Internal Reports

- [45] R. BONATO. *A Study on Learnability for Rigid Lambek Grammars*, Research Report, n^o 5964, INRIA, 08 2006, <https://hal.inria.fr/inria-00088818>.
- [46] B. MERY, M. AMBLARD, I. DURAND, C. RETORÉ. *A Case Study of the Convergence of Mildly Context-Sensitive Formalisms for Natural Language Syntax: from Minimalist Grammars to Multiple Context-Free Grammars*, Technical report, n^o 6042, INRIA, 2006, <https://hal.inria.fr/inria-00117306>.

References in notes

- [47] G. BOUMA, E. HINRICH, G.-J. M. KRUIJFF, R. OEHRLE (editors). *Constraints and Resources in Natural Language Syntax and Semantics*, distributed by Cambridge University Press, CSLI, 1999.
- [48] P. DE GROOTE, G. MORRILL, C. RETORÉ (editors). *Logical Aspects of Computational Linguistics, LACL'2001*, LNCS/LNAI, n° 2099, Springer-Verlag, 2001.
- [49] M. MOORTGAT (editor). *Logical Aspects of Computational Linguistics, LACL'98, selected papers*, LNCS/LNAI, n° 1014, Springer-Verlag, 2001.
- [50] C. RETORÉ (editor). *Logical Aspects of Computational Linguistics, LACL'96*, LNCS/LNAI, vol. 1328, Springer-Verlag, 1997.
- [51] C. RETORÉ, E. STABLER (editors). *Special Issue on Resource Logics and Minimalist Grammars*, vol. 2(1), Kluwer, 2004.
- [52] G. ROZENBERG, A. SALOMAA (editors). *Handbook of Formal Languages*, Springer Verlag, Berlin, 1997.
- [53] R. THOMASON (editor). *The collected papers of Richard Montague*, Yale University Press, 1974.
- [54] J. VAN BENTHEM, A. TER MEULEN (editors). *Handbook of Logic and Language*, North-Holland Elsevier, Amsterdam, 1997.
- [55] K. R. BEESLEY, L. KARTTUNEN. *Finite-State Morphology: Xerox Tools and Techniques*, Cambridge University Press, 2002.
- [56] R. BERNARDI, R. MOOT. *Generalized Quantifiers in Declarative and Interrogative Sentences.*, in "Logic Journal of the IGPL", vol. 11, n° 4, 2003, p. 419-434.
- [57] P. BLACHE. *Property Grammars: A Fully Constraint-based Theory*, in "Constraint Satisfaction and Language Processing", H. CHRISTIANSEN, P. SKADHAUGE, J. VILLADSEN (editors). , Springer, 2005.
- [58] N. CHOMSKY. *Some concepts and consequences of the theory of Government and Binding*, MIT Press, Cambridge, MA, 1982.
- [59] N. CHOMSKY. *The minimalist program*, MIT Press, Cambridge, MA, 1995.
- [60] A. DIKOVSKY, L. MODINA. *Dependencies on the other side of the Curtain*, in "Traitement Automatique des Langues", vol. 41, n° 1, 2000, p. 67-95.
- [61] L. T. F. GAMUT. *Logic, Language and Meaning – Volume 2: Intensional logic and logical grammar*, The University of Chicago Press, 1991.
- [62] 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.
- [63] G. HUET. *Linear Contexts and the Sharing Functor: Techniques for Symbolic Computation.*, in "Thirty Five Years of Automating Mathematics", F. KAMAREDDINE (editor). , Kluwer, 2003, <http://pauillac.inria.fr/~huet/PUBLIC/DB.pdf>.

- [64] G. HUET. *Zen and the Art of Symbolic Computing: Light and Fast Applicative Algorithms for Computational Linguistics*, in "Practical Aspects of Declarative Languages (PADL) symposium, New Orleans", Invited lecture, 2003, <http://pauillac.inria.fr/~huet/PUBLIC/padl.pdf>.
- [65] G. HUET. *Design of a Computational Linguistics Platform for Sanskrit*, in "Logical Aspects of Computational Linguistics, LACL 05", Invited Lecture, 2005.
- [66] G. HUET. *Transducers as lexicon morphisms, phonemic segmentation by euphony analysis, application to a sanskrit tagger*, in "Journal of Functional Programming", 2005, <http://pauillac.inria.fr/~huet/PUBLIC/tagger.ps>.
- [67] G. HUET. *Un système de traitement informatique du sanskrit*, in "Journé e ATALA: Traitement Automatique des Langues Anciennes", Invited Lecture, 2005.
- [68] R. JACKENDOFF. *The Architecture of the Language Faculty*, Linguistic Inquiry Monographs, n^o 28, M.I.T. Press, Cambridge, Massachusetts, 1995.
- [69] A. JOSHI, L. LEVY, M. TAKAHASHI. *Tree Adjunct Grammar*, in "Journal of Computer and System Sciences", vol. 10, 1975, p. 136–163.
- [70] 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.
- [71] 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.
- [72] H. KAMP, U. REYLE. *From Discourse to Logic*, D. Reidel, Dordrecht, 1993.
- [73] R. M. KAPLAN, M. KAY. *Regular Models of Phonological Rule Systems*, in "Computational Linguistics", vol. 20,3, 1994, p. 331–378.
- [74] L. KARTTUNEN. *Applications of Finite-State Transducers in Natural Language Processing*, in "Proceedings, CIAA-2000", 2000.
- [75] K. KOSKENNIEMI. *A general computational model for word-form recognition and production*, in "10th International Conference on Computational Linguistics", 1984.
- [76] 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.
- [77] A. LECOMTE, C. RETORÉ. *Towards a Minimal Logic for Minimalist Grammars: a Transformational Use of Lambek Calculus*, in "Formal Grammar, FG'99", FoLLI, 1999, p. 83–92.
- [78] 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.

- [79] D. LILLO-MARTIN. *Universal Grammar and American Sign Language: Setting the Null Argument Parameters*, Kluwer, 1991.
- [80] C. MANNING, H. SCHUTZE. *Foundations of statistical natural language processing*, MIT Press, 1999.
- [81] I. MELCUK. *Communicative Organization in Natural Language: The Semantic-communicative Structure of Sentences*, John Benjamins, 2001.
- [82] I. MELCUK. *Dependency syntax – theory and practice*, Linguistics, State University of New York Press, 1988.
- [83] J. MICHAELIS. *Derivational minimalism is mildly context sensitive*, in "Logical Aspects of Computational Linguistics, LACL'98, selected papers", M. MOORTGAT (editor). , LNCS/LNAI, n^o 2014, Springer-Verlag, 2001, p. 179–198.
- [84] 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.
- [85] R. MOOT. *Proof nets for linguistic analysis*, Ph. D. Thesis, UIL-OTS, Universiteit Utrecht, 2002.
- [86] R. MOOT. *Grail*, Technical report, 2005, <http://www.labri.fr/perso/moot/grail3.html>.
- [87] R. MOOT, M. PIAZZA. *Linguistic Applications of First Order Multiplicative Linear Logic*, in "Journal of Logic, Language and Information", vol. 10, n^o 2, 2001, p. 211–232.
- [88] F. MORAWIETZ. *Two-Step Approaches of Natural Language Formalisms*, Studies in Generative Grammar, Mouton de Gruyter, Berlin · New York, 2003.
- [89] C. MULLER. *Les bases de la syntaxe – syntaxe contrastive français / langues voisines*, Linguistica, Presses Universitaires de Bordeaux, 2002.
- [90] 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.
- [91] 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.
- [92] A.-M. PARISOT. *Accord et cliticisation: l'accord des verbes à forme rigide en Langue des Signes Québécoise*, Ph. D. Thesis, Université du Québec à Montréal, 2003.
- [93] T. POUSSEVIN, J.-F. DEVERGE, F. HAITI, A. HERBÉ. *Traitement Automatique des Langues: analyse syntaxique dans les grammaires catégorielles*, Mémoire de Maîtrise – TER, Université Bordeaux 1, May 2003.
- [94] G. K. PULLUM, B. C. SCHOLZ. *On the distinction between model-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.

-
- [95] J. PUSTEJOVSKY. *The Generative Lexicon*, MIT Press, 1995.
- [96] T. REINHART. *Anaphora and semantic interpretation*, University of Chicago Press, 1983.
- [97] T. REINHART. *Coreference and bound anaphora: a restatement of the anaphora questions*, in "Linguistics and Philosophy", vol. 6, n^o 1, 1983, p. 47–88.
- [98] T. REINHART, E. REULAND. *Reflexivity*, in "Linguistic Inquiry", vol. 24, 4 1993, p. 657–720.
- [99] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", vol. 2(1), 2004, p. 3–25.
- [100] J. ROGERS. *A Descriptive Approach to Language-Theoretic Complexity*, CSLI Publications, Stanford, California, 1998.
- [101] E. STABLER. *Derivational Minimalism*, in "Logical Aspects of Computational Linguistics, LACL'96", C. RETORÉ (editor). , LNCS/LNAI, vol. 1328, Springer-Verlag, 1997, p. 68–95.
- [102] 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). , distributed by Cambridge University Press, CSLI, 1999, p. 299–326.
- [103] V. SUTTON. *Lessons in SignWriting*, 2002, <http://www.signwriting.org>.
- [104] P. DE GROOTE, C. RETORÉ. *Semantic readings of proof nets*, in "Formal Grammar, Prague", G.-J. KRUIJFF, G. MORRILL, D. OEHRLE (editors). , FoLLI, 1996, p. 57–70.
- [105] 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.