INRIA

Project-Team SIGNES

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

Futurs

THEME SYM

Activity Report

2007

Table of contents

1. Team

INRIA/Signes is a joint team with LaBRI and the Department of Linguistics of Université Bordeaux 3 Michel de Montaigne — in particular with the research ministry EA 4195 and the CNRS UMR 5263 CLLE

LaBRI is a joint CNRS UMR 5800 team, involving Université Bordeaux 1, and the École Nationale Supérieure d'Électronique, d'Informatique, et de Radiocommunications de Bordeaux (ENSEIRB).

CLLE (Cognition, Langues, Langage, Ergonomie) is a joint CNRS team UMR 5263 involving Université Toulouse-Le Mirail and Université Bordeaux 3 Michel de Montaigne in Bordeaux.

TELEM (Textes, littératures : écritures et modèles) is a team of Université Michel de Montaigne, EA 4195 from the Research Ministry.

Head of project-team
Christian Retoré [Professor, Université Bordeaux 1, HdR]

Vice-Head of project-team
Christian Bassac [Lecturer, Université Bordeaux 3, HdR]

Administrative Assistant
Sylvie Embolla [TR]

Research Scientists (affiliation)
Lionel Clément [Lecturer, Université Bordeaux 1]
Irène Durand [Lecturer, Université Bordeaux 1, HdR]
Patrick Henry [IR CNRS LaBRI]
Kupsc Anna [Lecturer, Université Bordeaux 3 – from September 2007]
Renaud Marlet [CR INRIA, HdR]
Richard Moot [CR CNRS, LaBRI]
Henri Portine [Professor, Université Bordeaux 3, HdR]
Benoît Sagot [CR INRIA – Until July 2007]
Sylvain Salvati [CR INRIA – from February 2007]

Research Scientists (partner)
Mauro Gaio [Professor, Université de Pau, HdR]
Kim Gerdes [Lecturer, Université Paris 3]
Pierre Guitteny [Interpret/Official representative of *Direction Régionale des Affaires Sanitaires et Sociales* Aquitaine]
Gérard Huet [DR INRIA, HdR]
Alain Lecomte [Professor, Université Paris 3, HdR]

Research Scientist (visitor)
Michael Moortgat [Invited Porfessor Université Bordeaux 1, October 29th–November 10th and December 10th–22th, HdR]
Reinhard Muskens [Poste Rouge CNRS From Décember 2006 to May 2007, HdR]

Post-doctoral fellows
Marie-Laure Guénot [Temporary lecturer, Université Bordeaux 3 – Until August 2007]
Greg Kobele [INRIA postdoc Until October 2007]
Tristan Vanrullen [INRIA postdoc Until November 2007]

Ph. D. Students
Maxime Amblard [Temporary lecturer, ENSEIRB, Until September 2007 – PostDoctoral Fellow, France Télécom from October 2007]
Houda Anoun [Part Time Lecturer Université Bordeaux 1]
Pierre Bourreau [INRIA Cordi Grant From September 2007]
Bruno Mery [Ministry grant, Université Bordeaux 1]

Émilie Voisin [Aquitaine/INRIA grant Until August 2007 — Temporary Lecturer Université Bordeaux 3 From September 2007]

2. Overall Objectives

2.1. Overall Objectives

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

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

by means of formal methods including:

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

Several applications illustrate this approach among which we mention:

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

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

3. Scientific Foundations

3.1. 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 fields and the source of technical inspiration, as exemplified by the relation between formal language theory and linguistics [62].

Nowadays, the motivation is the need to handle large quantities 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, and computer aided translation.

Statistical methods and corpus linguistics [88] have been quite successful in the past, but there is a renewal of symbolic methods, and especially of logical ones, due to 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 [64].

For all these applications, whether 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 opens the 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 from a linguistic viewpoint, the two sides of the linguistic signs *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'čuk [90], [67].
- From a mathematical and algorithmic viewpoint, this interface gives rise to some challenges: what is the link between two of the main frameworks, namely generative grammars and categorial grammars? The former are exemplified by Tree Adjoining Grammars TAGs [77] or Minimalist Grammars [103]. They enjoy efficient parsing algorithms and broad coverage of syntactic constructs. The latter (see e.g. [92]) 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 [63], [69] and thus allow generation algorithms. Other models, like dependency grammars [89], provide a different account of the syntax/semantics interface. A comparison between dependency models and generative/logical models enables an assessment of the adequacy of these families of models with linguistic descriptions; 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 tranducers, etc. [81], [82].

At the other end, on the semantics 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 [79], 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 [correspondant], Kim Gerdes.

Computational models for phonology and morphology are a traditional application of finite state technology [81], [82], [83], [65]. These models often combine symbolic or logical systems, like rewriting systems, and statistical methods like probabilistic automata which can be learnt from corpora by Hidden Markov Models [88].

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

Existing techniques for compiling such machinery remain 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 a compact representation of such automata. Gérard Huet has implemented a toolkit based on this technique, which allows a very efficient automatic 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], Lionel Clément, Kim Gerdes, Marie-Laure Guénot, Greg Kobele, Anna Kupść, Alain Lecomte, Renaud Marlet, Richard Moot, Christian Retoré [correspondant], Benoît Sagot, Sylvain Salvati.

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.

Rewriting 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 [78]. Among these classes of grammars let us mention Tree Adjoining Grammars [76], [77] and Minimalist Grammars [103], [104], [91]. Dependency Grammars share some properties with them but the general paradigm is quite different [90], [68].

Edward Stabler [103] introduced Minimalist Grammars (MGs) as a formalization of the most recent model of the Chomskian or generative tradition and they are quite appealing to us. They offer a uniform model for the syntax of all human languages.

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

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

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

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

Indeed, a strong reason for using categorial grammars [92], despite their poor computational properties and poor linguistic coverage, is that they provide a correspondence between syntactic analyses and semantic representations. This will be explained in the next section on the syntax/semantics interface.

In order to improve the computational properties of categorial grammars, and to extend their scope, one can try to connect them to more efficient and wider formalisms, like minimalist grammars [85], [84], [100].

A rather new approach to syntax is known as model-theoretic syntax. Its advantages have been underlined by Geoffrey Pullum in [98]. 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. Note that this approach may be considered as another way of describing generative grammars. The advantages of such a description are not in the parsing algorithms (MSO or Constraint Satisfaction are usually of high complexity) but rather in characterising the language class and possibly describing it in a linguistically natural way (as opposed to lexical items of lexicalized grammars). This connection to logic is related to constraint-logic programming or to monadic second order logic.

In the MSO style, the pioneering work of James Rogers on *Government and Binding* and *Tree Adjoining Grammars* must be mentionned [101]. Uwe Mönnich, Jens Michaelis and Frank Morawietz have obtained a two step description of minimalist grammars that we are studying [95], [94].

In the constraint style issued from the Prolog-Definite Clause Grammars, Head Phrase Structure Grammar, Construction Grammars and Property Grammars are defined as sets of constraints. The later ones introduced by Philippe Blache offer a rather natural way to describe grammar rules and are newly studied by our group [70], [66].

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

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

Participants: Maxime Amblard, Pierre Bourreau, Greg Kobele, Alain Lecomte, Renaud Marlet, Michael Moortgat, Richard Moot, Christian Retoré, Sylvain Salvati.

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 [69], [109].

Nevertheless the 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 that are *ad hoc*, and even wrong. For instance, why should the type of a determiner depend of the constituent with which it is involved? 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 [75]. 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 [84], [100].

3.5. Lexical semantics and derivational morphology

Keywords: *computational semantics*, *lexical semantics*.

Participants: Christian Bassac [correspondant], Mauro Gaio, Patrick Henry, Renaud Marlet, Bruno Mery, Reinhard Muskens, Christian Retoré.

The Generative Lexicon [99] 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 regardless of whether they are syntactically mandatory or optional. The event structure follows [79]. It unfolds an event into several ordered sub-events with a mark on the most salient sub-event. Events are typed according to the typology of Vendler: state, process and transition, this latter type including achievement and accomplishment. The qualia structure relates the argument structure and the event structure in roles: formal, constitutive, telic, agentive.

This information and its organization into the generative lexicons allows an explanation of, for instance, polysemy and of compositionality (in particular in compound words). This kind of model relates knowledge representation to linguistic organization and thus 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: Mauro Gaio, Alain Lecomte [correspondant], Reinhard Muskens, Henri Portine.

Montague semantics have some limits. Two of them which, technically speaking, concern the context, can be overcome by using DRT, that is Discourse Representation Theory and its variants. [79], [110] 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. [110].

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

3.7. Type systems and functional programming for computational linguistics

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

Participants: Houda Anoun, Gérard Huet [correspondant], Richard Moot.

The project-team has developed skills in logic, lambda calculus. These models are commonly used in computational linguistics, for example:

- Categorial grammars, with their parsing-as-deduction paradigm, which use proofs in Lambek calculus or linear logic as syntactic trees.

- Montague semantics, which uses the Church description of higher-order logic, implemented in lambda calculus in order to have the compositionality principle of Frege.

- Discourse Representation Theory, which is also 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 [93].

- 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 [97].

- 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 [73].

4. Application Domains

4.1. Sanskrit philology

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

Participant: Gérard Huet [correspondant].

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

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

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

4.2. Towards French Sign Language (LSF) modelling and processing

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

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

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

Given the rarity of linguistic study on the syntax and semantics of sign languages — some exceptions concerning American Sign Language are [96], [86], [87] — before being able to apply our methodology, our first task is to determine the structure of the sentence, using our personal competence as well as our relationship with the deaf community.

We intend to define methods and tools for generation of sign language sentences. It should be noted that there is a continuum of different representations of a sentence in Sign Language, from a grammatical description with agreement features and word/sign order that we are familiar with, to a notation system like Signwriting [105] or to a language for the synthesis of 3D images and movies. Our competences on the interface between syntax and semantics are well designed for work in generation of the grammatical representations.

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

5. Software

5.1. The Zen Toolkit

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

Participant: Gérard Huet [correspondant].

The Zen Toolkit is a library of finite state automata and transducers, called Zen for its simplicity. The algorithmic principles of the Zen library are based on the linear contexts data structure ("zippers") and on the sharing functor (associative memory server) [71]. It has been developed by Gérard Huet and is being used in his Sanskrit modelling platform (see section 5.2). It allows the construction of lexicons, the computation of morphological derivatives and inflected forms, and the segmentation analysis of phonetic streams modulo euphony [71].

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

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

5.2. Sanskrit Site

Keywords: *Sanskrit*, *electronic dictionary*, *parsing*, *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 [74], [72]. These resources are built upon, among other ingredients, the Zen Toolkit (see section 5.1). The site registers thousands of visitors daily.

- The *declension engine* gives the declension tables for Sanskrit substantives.
- The *conjugation engine* conjugates verbs for the various tenses and modes.
- The *lemmatizer* tags inflected words.
- A *dictionary* lists inflected forms of Sanskrit words. Full lists of inflected forms, in XML format (given with a specific DTD), are released as free linguistic resources available for research purposes. This database, developed in collaboration with Pr. Peter Scharf, from the Classics Department at Brown University, has been used for research experiments by the team of Pr. Stuart Shieber, at Harvard University.

- The *Sanskrit Reader* segments simple sentences, where the (optional) finite verb form occurs in final position. This reader enhances the hand-tagged Sanskrit reader developed by Peter Scharf, that allows students to read simple texts differently: firstly in davanagari writing, then word-to-word, then in a word-to-word translation, then in a sentence-to-sentence translation.

- The *Sanskrit Parser* eliminates many irrelevant pseudo-solutions (segmentations) listed by the Sanskrit reader.

- The *Sanskrit Semantic Analyzer*, based on the notion of *kāraka* of Pāṇini, controls overgeneration using a pertinence principle [42].

- The *Sanskrit Tagger* is an assistant for the tagging of a Sanskrit corpus. Given a sentence, the user chooses among different possible interpretations listed by the morpho-syntactic tools and may save the corresponding unambiguously tagged sentence on disk as an hypertext document indexing in the Sanskrit Heritage Dictionary (our structured lexical database). This service has no equivalent worldwide.

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

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

5.3. Grail: Natural Language Analysis with Multimodal Categorial Grammar

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. [92]) are one of the richest approaches. 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 (http://www.labri.fr/perso/moot/grail.html).

The Grail parser/theorem prover for categorial grammars, originally developed at the University of Utrecht, has been rewritten from scratch in Grail 3, taking into account modern insights about proof nets as well as requiring only open-source software to run. This release (http://www.labri.fr/perso/moot/grail3.html) also includes computational theoretical improvement in accordance with [93]: 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 [50]. Additionally, several strategies for reducing the search space have been implemented, significantly improving parsing performance.

Grail has received an extensive update in 2007: the graph library has been improved by the inclusion of a faster and more compact implementation of the Floyd-Warshall algorithm for computing the transitive closure. Improvements also include an implementation of Regin's algorithm for enforcing all-different constraints and of Gabow's algorithm for computing strong components.

5.4. Suite of Corpus Tools for Type-Logical Grammars

Keywords: *corpus*, *type-logical grammar*.

Participant: Richard Moot [correspondant].

A suite of tools for grammar extraction, corpus transformation, supertagger training and evaluation, and error analysis has been revised and updated. It includes a tighter interaction with Grail (see section 5.3) as well as new search, extraction and evaluation options. A first version has been publicly released and can be downloaded from http://www.labri.fr/perso/moot/Corpus/.

5.5. Experimental LFG Parsing: 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 is distributed as free software (http://lionel-clement.dyndns.org/xlfg/). It was developed by Lionel Clément (before he joined the Signes project-team).

XLFG5 is a major reimplementation of XLFG. The parsing produces a shared forest (of c-structures) in order to speed-up ambiguous sentences analyzes. The sharing of functional dependency structures (f-structures) is under development. XLFG5 is a parser prototype for research. It has been developed by Lionel Clément and is distributed as free software (http://lionel-clement.dyndns.org/xlfg5/).

Both XLFG and XLFG5 are used for teaching in various universities.

5.6. Efficient and Robust LFG Parsing: SxPipe and SxLfg

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

Participant: Benoît Sagot [correspondant].

Benoît Sagot, before he arrived in the Signes project-team, developed with Pierre Boullier (INRIA project-team Atoll, now Alpage) 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 Le*fff* (see section 5.11).

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 Le*fff*. 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 parsing multi-million-word corpora in a few hours with a large-coverage grammar. Robustness techniques at all levels allow it 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 active developement, but they already give satisfying results (e.g., state-of-the-art chunking precision).

5.7. Yab

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

Participant: Lionel Clément [correspondant].

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

5.8. Lexed

Keywords: *dictionary search*, *lexicalizer*.

Participant: Lionel Clément [correspondant].

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

5.9. Tokenizer

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

Participant: Lionel Clément [correspondant].

Tokenizer is a program that segments a text into tokens. Ambiguity between simple and compound words is represented through a directed acyclic graph (DAG). This software has been developed by Lionel Clément (before he joined the Signes project-team) and is part of Lexed (see section 5.8).

5.10. MERLin

Participants: Tristan Vanrullen [correspondant], Lionel Clément.

MERLin (Models for Expression and Representation of Linguistic information) is a program that supports a formal language that allows the expression of GPi grammars, their computation, their representation and their transformation into grammars for several NLP formalisms. The formal language for GPi Grammars is based on some ideas that are common in the metagrammar paradigm, as well as on some new features. Its specifications can be accessed from the ARC Mosaïque web site (http://mosaique.labri.fr/). See also section 6.3.4.

This software project, which is currently being designed and developed by Tristan Vanrullen, is still work in progress. It can be accessed on the Inria GForge Platform by authorized members at https://gforge.inria.fr/projects/merlin/. The access to this project should become public by the end of the year.

5.11. LeFFF: A Lexicon of French Inflected Forms

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

Participants: Lionel Clément [correspondant], Benoît Sagot.

The Le*fff* (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 (http://www.labri.fr/perso/clement/lefff/). A first version, the Le*fff* 1, was limited to a morphological lexicon of French verbs. It has been developed by Lionel Clément and Benoît Sagot (before they joined the Signes project-team) in collaboration with Bernard Lang, thanks to an original automatic acquisition techniques from raw corpora complemented with manual validation. More recently, the Le*fff* 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 Le*fff* 2 is now mostly developed by Benoît Sagot, but previous work by Lionel Clément had been used as a starting point. This development is still active (see section 6.5.3), in particular in comparison with other free resources (Dicovalence, some of the lexicon-grammar tables), by developing new manual or automatic acquisition techniques, and by improving the original lexical model underlying the Le*fff*. The Le*fff* is currently used by many French or international research teams.

5.12. TreeLex

Keywords: *adjective*, *lexicon*, *treebank*, *valence*, *verb*.

Participant: Anna Kupść [correspondant].

TreeLex is a subcategorisation lexicon automatically extracted from a syntactically annotated corpus (treebank). It contains about 2000 contemporary French verbs (types) with their valency frames. 160 different verb frames have been identified, with 1.91 frame per verb on average. Recently, the work has been extended to the extraction of adjective frames.

The lexicon is still under development. It is being created and maintained by Anna Kupść (the work started before she joined Signes). The lexicon will be freely available at http://erssab.u-bordeaux3.fr/.

5.13. 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.

5.14. Text Generation from Syntactic Dependencies: DepLin

Keywords: *natural language generation*, *syntactic analysis*, *syntactic dependency*, *topological grammar*.

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 a 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 under GPL and, apart from its use in Signes (in particular for German and Greek), it is mainly used at the University of Paris 7 for the development of different grammars (in particular Arabic and French).

5.15. Speech Synthesis with Prosodic Structure Generation

Keywords: *audiocorpora*, *prosody*, *speech synthesis*.

Participant: Kim Gerdes [correspondant].

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

5.16. Bilingual Text Alignment: Alignator

Keywords: *bilingual text alignment*, *textometric analysis*.

Participant: Kim Gerdes [correspondant].

A preliminary version of a bilingual text alignment tool, called Alignator, has been developed made available (http://elizia.net/alignator). It is written Python, C, and Javascript, It runs on a web server for high accessibility. It aims at being a useful tool for researchers working on bilingual text who need fast paragraph alignment for textometric analyses (cf. section 6.5.1).

5.17. Acquisition of Linguistic Corpora on the Web

Keywords: *corpora*, *web-robot*.

Participant: Kim Gerdes [correspondant].

Kim Gerdes developed a web-robot specialized in linguistic corpora acquisition with automatic domain and language recognition and recognition of morphological schemas for inflected languages. (See also section 6.5.2.)

5.18. Farsi Corpus

Keywords: *Farsi*, *Persian*, *annotation*, *corpus*.

Participant: Kim Gerdes [correspondant].

In collaboration with Pollet Samvellian (University Paris 3), Kim Gerdes developed an automatically annotated large corpus of Farsi (see section 6.5.2).

5.19. French Sign Language HD Corpus

Keywords: *French Sign Language*, *LSF*, *coprus*, *high definition*.

Participants: Patrick Henry [correspondant], Émilie Voisin.

Thanks the two HD-video cameras provided by a former regional grant, Émilie Voisin and Patrick Henry recorded two hours of sign language utterances from native deaf speakers. This corpus is made of translations of French texts and short comics strips (with no textual annotation) turned into sign language expressions; it also contains sign language productions by non-native speakers as well as the reactions and comments of native sign language speaker confronted to these productions. Annotation and extraction of sequences in the corpus can be performed thanks to a specific interface. In 2008 it is planned to extend the corpus and start its diffusion within the scientific community.

5.20. Generative Lexicon Toolkit

Keywords: *Generative Lexicon*, *anaphoric reference*, *network*, *nominal compound*.

Participants: Christian Bassac, Patrick Henry [correspondant].

Patrick Henry and Christian Bassac designed a toolkit 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 (see also section 6.2.1). The implementation of other features of the generative lexicon are planned.

5.21. Turkish Conjugation

Keywords: *Turkish conjugation*.

Participants: Christian Bassac, Patrick Henry [correspondant].

Christian Bassac and Patrick Henry started working on the development of a platform designed to implement the generation of verb forms of Turkish conjugation paradigms. Some work remains to get done mainly regarding the details of the features of number and person, and on the labelling of forms.

6. New Results

6.1. Properties of the formalisms

6.1.1. *Formalisation of Minimalist Grammars*

Keywords: *Categorial Grammar*, *Government and Binding (GB)*, *Lambek calculus*, *Linear Logic*, *Minimalist Grammar*, *syntax-semantic interface*.

Participants: Maxime Amblard, Houda Anoun, Greg Kobele, Alain Lecomte, Christian Retoré [correspondant].

Maxime Amblard proposed a new version of Minimalist Grammars based on an algebraic definition of trees. This new version provides a framework where hypothesized properties of these grammars can be checked [16]. Maxime Amblard and Christian Retoré have provided a natural deduction system for Partially Commutative Intuitionistic Multiplicative Linear Logic (PCIMLL) and established its normalisation and subformula properties [31]. Such a system involves both commutative and non commutative connectives and deals with contexts that are series-parallel multisets of formulae. This calculus is the extension of the one introduced by de Groote presented by the second author for modelling Petri net execution, with a full entropy which allows order to be relaxed into any suborder — as opposed to the Non Commutative Logic of Abrusci and Ruet. Their result also includes, as a special case, the normalisation of natural deduction within the Lambek calculus with product, which is unsurprising but heretofore unproven. Up to now PCIMLL with full entropy had no natural deduction system. In particular for linguistic applications, such a syntax seems particularly well-suited for constructing semantic representations from syntactic analyses.

Maxime Amblard proposed a new formalisation of minimalist grammars based on Partially Commutative Intuitionistic Multiplicative Linear Logic (PCIMLL) that implements the main propositions of Chomsky in the Minimalist Program [16]. This formalism, called Categorial Minimalist Grammars (CMG), provides a framework that uses assumptions for modelling of reasoning in derivations. The non commutative properties are used for implicative elimination and the commutative one for the product elimination. Maxime Amblard proved the inclusion of generated languages into minimalist languages.

Maxime Amblard also proposed a syntax-semantic interface for Categorial Minimalist Grammars (CMG) [16]. For this, the semantic calculus is synchronised on the syntactic one (proof of CMG). The results of these calculi are formulae of first order logic. This interface accounts for uniform theta assignment hypothesis (UTAH) which associate thematic roles to variables of the formulae. To represent the quantifier scope ambiguities, Maxime Amblard includes properties of $\lambda\mu$-calculus which implement the continuations.

Houda Anoun and Alain Lecomte proposed another framework for implementing ideas coming from the minimalist program into a logical setting [32]. The logical rules include elimination and introduction rules but the use of hypotheses is constrained. Based on this work, Houda Anoun and Alain Lecomte produced a quasi exhaustive study of binding phenomena in a type-logical framework [33].

6.1.2. *Type-Theoretic Grammars and the Distinction Between Tectogrammars and Phenogrammars*

Keywords: *Abstract Categorial Grammar*, *Lambek calculi*, *Linear Logic*, *Minimalist Grammars*, *membership problem*, *mildly context sensitive formalisms*, *parsing*, *phenogrammar*, *tectogrammar*.

Participants: Pierre Bourreau, Greg Kobele, Richard Moot, Christian Retoré, Sylvain Salvati [correspondant].

Sylvain Salvati investigated in [57] the respective complexities of the membership and the universal membership problems for Abstract Categorial Grammars (ACGs) [108]. Although this problem has already been addressed in [107], Sylvain Salvati presents some more precise results as well as some new ones.

In [58], Sylvain Salvati has showed that the parsing problem is decidable for context-free languages of simply typed λ-terms. Generating sentences from their meaning representation, when the relation between syntax and semantics is described in a Montagovian fashion, is a particular instance of this problem and is therefore decidable. This result relies on the use of intersection types which seem to be adequate for studying syntactic problems, such as higher order matching, in the simply typed λ-calculus.

On the syntactic side, several formalisms have been investigated through the tectogrammar/phenogrammar distinction using classical tools from formal language theory:

- the non-associative Lambek calculus can be polynomially embedded into a second order ACG [53] (this shows that the parsing problem for non-associative Lambek grammar is polynomial);

- the derivation trees of minimalist grammars with the shortest move constraint are regular and the derived sentence can be obtained from them with a finite copying tranducer [46];

- the derivations of Lambek grammars can be described in a free algebra; this yields several results such as:

    – derivations in Lambek grammars, seen as normal natural deduction trees, are not context free sets of trees;

    – Lambek grammars can be embedded in second order ACGs;

    – cut-free proof-nets representing derivations in Lambek grammars are sets of hypergraphs that can be described as the languages of hyperedge replacement grammars;

- the derivations of minimalist grammars are faithfully described by proof in Multiplicative Exponential Linear Logic.

Makoto Kanazawa and Sylvain Salvati [43] related k-control languages defined by Weir [106] and mildly context sensitive formalisms by showing the relation of this hierarchy with the hierarchy of l-multiple context free languages defined in [102]. The proofs use some non-trivial properties of ACGs and show the interest of this formalism as a formal tool.

Besides, Richard Moot introduced a proof-net calculus for the Lambek-Grishin calculus (LG), an extension of the Lambek calculus proposed by Grishin in 1983. The calculus and its soundness and correctness results extend the proof nets for the multimodal Lambek calculus in several ways. With the help of these proof-nets an embedding result of Lexicalized Tree Adjoining Grammars into LG is shown, giving a first result with respect to the descriptive complexity of LG.

6.1.3. *Logics, Linguistics, and Other Sciences*

Keywords: *Ludics*, *computational biology*, *epistemology*, *linguistics*, *logic*, *pragmatics*.

Participant: Alain Lecomte [correspondant].

In collaboration with Marie-Rénée Fleury and Myriam Quatrini, Alain Lecomte produced a synthesis of the ideas of Ludics [61], a theory originally developed by Jean-Yves Girard. This synthesis presents in an intuitive way the basic concepts of Ludics and continuations, as well as some applications to pragmatics.

In [23], Alain Lecomte showed the consequences of the recent developments in logic (and its applications to calculus) concerning sciences like biology or linguistics which deal with informational processes. Alain Lecomte also re-stated and analyzed the problem of the frontier between language and logics [24].

6.2. Semantics

6.2.1. *Computational Semantics with Rich Lexical Semantics*

Keywords: *Generative Lexicon (GL)*, *Montague*, *computational semantics*, *lexical semantics*, *syntax-semantic interface*.

Participants: Christian Bassac [correspondant], Patrick Henry, Renaud Marlet, Bruno Mery, Henri Protine, Christian Retoré.

Computational semantics has long relied upon the Montague correspondance between syntax and semantics, which is not by itself well suited for the computation of some phenomena, such as logical polysemy, that have been addressed by recent advances in lexical semantics. In fact, generative lexical semantics aims, among other things, to model co-compositional meanings that classical Montague semantics is unable to express. This discrepancy has been studied in various directions in the project-team.

Christian Bassac, Bruno Mery and Christian Retoré pursued their formalisation of computational lexical semantics within type theory, starting from Pustejovsky's Generative Lexicon (GL). Instead of providing constructs for each of the operations in GL, they considered another way to use a lexicon à la Montague. An object is provided with *several* lambda terms: the main, usual one, ought to be used, while the others, said to be optional, are only used whenever a type clash needs to be overcome as in the case of dot-objects. These lambda terms, involving constants, can turn the object into its relevant aspect, but, as opposed to what happens with canonical type morphisms, such transformations cannot apply when no term from the lexicon enables them. Two versions of this principle are being studied: one [49] considers only the various aspects of the type, without a general type that specialises into them, but with transformations between aspects, while the other approach [48] actually provides a general type corresponding to the dot-type, and terms whose type is the one of the projections from a dot-type to its component. More recently, the authors have explored the possibility to use the multiplicative conjunction of linear logic as a generic type for dot-objects without systematic projections, yet with a compound type related to its components.

Renaud Marlet also observed that computational semantics and lexical semantics have so far mostly been studied separately, and that, as a result, computational semantics often constructs meanings with poor or little lexical sense, while lexical semantics generally only apply to simple and small phrases (as opposed to complete sentences). He proposed a general framework to incorporate lexical semantics information originating from a generative lexicon into a standard (Montagovian) analysis of computational semantics [47]. His framework is able to deal with phenomena such as type coercion and selective binding. In this framework, the coupling between the analysis and the lexicon is kept as low as possible to facilitate separate evolutions and to adapt to the partial availability of lexical semantics information.

Besides, Christian Bassac and Patrick Henry designed and implemented a software toolkit for the construction, maintenance and collaborative use of a Generative Lexicon (cf. section 5.20). They showed that this toolkit can be used among other things to filter the adequate form of anaphoric reference to the modifier in French endocentric compounds [41].

In collaboration with Agnès Bracke, Henri Protine also studied scientific lexemes [37].

6.2.2. *Representation of Semantics*

Keywords: *event*, *semantics*.
Participant: Christian Bassac [correspondant].

Christian Bassac studied the motivations and consequences of the introduction of an event variable in the arguments of a predicate and showed that this approach proves superior to concurrent approaches such as those in which events are considered as elements to which a sentence refers [34]. He argued that the best-suited tool to construct the object the event variable refers to is Hilbert's ε function.

6.2.3. *Virtual Itineraries in Pyrenees*

Keywords: *corpus studies*, *geographic information system*, *indexing*, *information retrieval*, *lexical semantics*.
Participant: Mauro Gaio [correspondant].

Local cultural heritage document collections are characterized by their content, which is strongly attached to a territory and its land history (i.e., geographical references). The aim of the PIV project (Virtual Itineraries in Pyrenees) is to add value to a legacy localised corpus in a geographic (spatial) oriented digital library.

In [20], Mauro Gaio, Christian Sallaberry, Patrick Etcheverry, Christophe Marquesuzaà and Julien Lesbeguéries present the PIV project. Spatial and temporal core models are proposed to give a formal representation of geographical information. Semantic processes have been built to automatically manage the spatial and temporal information from non-structured data. A "back office" prototype, which adds these processes to classic information extraction approaches, while associating a geographical information retrieval (GIR) service is proposed. This service searches for any links between formal representations of geographic information in document collections, and similar representations in a user's information query. This paper presents the design work, giving the details of the principles of result visualization and navigation, while proposing a "front office" first implementation of the system.

In collaboration with Christian Sallaberry, Julien Lesbegueries and Pierre Loustau, Mauro Gaio worked at making the content retrieval process more efficient whenever a query includes geographic criteria [27]. The authors proposed a core model for a formal representation of spatial information. The model takes into account characteristics of different modes of expression, such as written language, captures of drawings, maps, photographs, etc. The authors have developed a prototype that fully implements spatial information extraction (IE) and spatial information retrieval (IR) processes on text. All prototype processing resources are designed as Web Services. The authors proposed a geographic IE process based on semantic treatment as a supplement to classical IE approaches.

In collaboration with Christian Sallaberry, Mustapha Baziz and Julien Lesbegueries, Mauro Gaio proposed an approach (implemented within PIV prototype) based on a semantic analysis of digital corpora and free text queries [56]. They defined requirements and a methodology of semantic annotation for automatic indexing and geo-referencing of text documents. They also reported on a case study where the spatial-based IR process is evaluated and compared to classical (statistical-based) IR approaches using first, pure spatial queries and then, more general ones containing both spatial and thematic scopes. The main result in these first experiments shows that combining a spatial approach with a classical (statistical-based) IR one improves in a significant way retrieval accuracy, namely in the case of general queries.

6.3. Morphology and Syntax

6.3.1. *Linguistic Modelling of the Syntax of French Sign Language*

Keywords: *French Sign Language (LSF)*, *Generative Grammar*, *language typology*, *morphology*, *syntax*.

Participants: Maxime Amblard, Pierre Guitteny, Patrick Henry, Renaud Marlet, Henri Portine, Christian Retoré, Émilie Voisin [correspondant].

The studies of signed languages are often confronted with a certain number of biases that alter in a important way language production. Émilie Voisin studied cases of diglossia in French Sign Language (LSF) and their impact on syntax [30]. In collaboration with Loïc Kervajan, she also proposed a formalization of the modifier mechanisms (such as "péniblement") [44]. This builds on previous work regarding the flexional typology of verbs in LSF and its relationship with syntax [29], [28].

Maxime Amblard and Émilie Voisin studied the interplay between verbal inflection and sign order (e.g., SOV or OSV) in French Sign Language [59]. Based on an analysis of verbal inflection as noun incorporation, they proposed a model for sign order. For this, they rely on Minimalist Grammars (MGs) to provide a flexible framework where the incorporation phenomena are represented by a new concept, called "signeme". Signemes in a sentence are build with traces in the MGs.

On a more sociological issue, Pierre Guitteny studied how new technologies for visual communication allow distant sign speakers to communicate [22]. These technologies not only impact deaf people individually, regarding access to knowledge and autonomy, they also impact them collectively, regarding community structuring as well as social and professional integration. Moreover, these technologies also influence the development of sign langage itself.

6.3.2. *Linguistic Modelling of French Syntax*

Keywords: *French grammar*, *Pronouns*, *Property Grammar*.

Participants: Maxime Amblard, Marie-Laure Guénot, Henri Portine, Lionel Clément [correspondant].

Marie-Laure Guénot proposed a grammar for French that is the implementation of a linguistic model based on corpus descriptions (notably coming from the Approche Pronominale of Claire-Blanche Benveniste) and represented in the Property Grammars formalism [40]. It accounts for a new proposition among formal grammars, taking part into the works that aim to promote convergence between the various researchs of descriptive linguistics and the diversity of formal representation possibilities. It is freely available on the Spoken Data Resource Center (CRDO), as a representation and analysis resource (http://crdo.fr).

In his PhD, Maxime Amblard also constructed a fragment of a French grammar in the Minimalist Grammars formalism [16]. He proposed an extension of Stabler's version of clitics treatment for a wider coverage of French. He presented the lexical entries needed in the lexicon and explained the recognition of complex syntactic phenomena such as left and right dislocation, clitic climbing over modal verb and extraction from a 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). The proposed formalisation of French clitics includes the treatment of clitics in complex positions such as negative form, imperative, raising-verbs, control verb, etc.

Henri Portine proposed a new analysis for the *avec* preposition [52]. He also studied the modal value of the construction "se devoir (à soi-même) de Vinf" [25].

6.3.3. *Formal account for 'A' and 'A'-bar movement*

Keywords: *Generative Grammar*, *Government and Binding (GB)*, *Minimalist Grammars*.
Participant: Greg Kobele [correspondant].

Greg Kobele proposed a formal foundation for 'A' and 'A-bar' movement [45]. 'A' dependencies (typified by raising, passivization, etc) are those that allow for re-binding, disallow reconstruction, and do not license parasitic gaps. 'A-bar' dependencies (typified by wh-movement, relativization, etc) disallow re-binding, allow reconstruction, and do license parasitic gaps. Furthermore, when an expression enters in to both A and A-bar dependencies, all of its A dependencies must "precede" its A-bar dependencies. This relational property of dependencies is known as the ban on improper movement. In the government and binding (GB) tradition, all of the properties of these two dependency types must be independently stipulated-none follow from any of the others. Greg Kobele presented a simple and constrained formal system in the GB tradition with two kinds of long distance dependency forming operations, the interaction of which gives rise to the major syntactic property of A and A-bar movement: the ban on improper movement. Additionally, the resulting system seems to provide the right kinds of structures over which to naturally enforce the other characteristic semantic properties (reconstruction and re-binding) of each kind of movement. The difference between A and A-bar dependencies, as regards the licensing of parasitic gaps, cannot be made to follow from the formal architecture of the system; nor do parasitic gaps. However, natural grammars can be written for fragments of English in which the difference in licensing of parasitic gaps obtains without stipulation.

6.3.4. *High-Level Syntactic Formalisms*

Keywords: *factorization*, *linguistics*, *metagrammar*.
Participants: Lionel Clément [correspondant], Tristan Vanrullen.

Tristan Vanrullen worked on the representation of linguistic information for several NLP formalisms. This work deals with the problem of factorization for linguistic descriptions. As exposed on the ARC Mosaïque web site (http://mosaique.labri.fr/), the new idea introduced in the metagrammar paradigm is the fact that meta-grammars handle two kinds of factorized informations: the first one is structural (and formalism dependant: tree structures, graphs, dependancies), and the second one is linguistic. The latter supposes the introduction of a way to represent non generative data and linguistic knowledge, without redundancy. (See also section 5.10.)

6.3.5. *Topological Grammars*

Keywords: *German*, *Meaning-Text Theory*, *dependency*, *topological grammar*, *word order*.
Participant: Kim Gerdes [correspondant].

In collaboration with Sylvain Kahane, Kim Gerdes investigated the notion of phrase in non phrase structure grammars [21]. Following Tesnière and Mel'čuk, Gerdes and Kahane defend the idea that word order must be separated from the syntactic representation proper, and that phrases only intervene when word order is at play. Gerdes and Kahane try to characterize a new notion they call topological phrase (partially inherited from the classical topological model for German) and distinguish it from the classical notion of phrase in X-bar Syntax. Their proposition is illustrated by the puzzling case of German word order for which Gerdes and Kahane propose a simple and powerful grammar giving all the possible word orders and topological phrase structures of verbal syntax. This discussion of the notion of phrase opens a new perspective for the comparison of the entire architectures of Chomskyan and Mel'čukian linguistic models.

6.3.6. Lexical Morphology

Keywords: *junggrammatiker*, *linguistics*, *method of lexical exceptions*.

Participant: Christian Bassac [correspondant].

In the context of the method of lexical exceptions, Christian Bassac showed how the notion of exception to a rule was understood by the linguistic school of the *junggrammatiker* and discussed the relevance of the creed of this linguistic school in today's linguistics [35].

6.4. Parsers

6.4.1. Parser Evaluation

Keywords: *Spoken Dutch Corpus*, *proof net*.

Participant: Richard Moot [correspondant].

Richard Moot evaluated the different heuristics used by the Grail parser (cf. 5.3) to filter out different axiom links which cannot be part of a total linking for a proof net, including the new implementations of Regin's algorithm and the use of first-order to encode word order constraints [51]. He evaluated Grail against a series of 478 randomly generated sequents (containing 4.012 atomic formulas and 2.330 connectives: Grail performs 2.825 axioms out of 15.946 possibilities) and against 5454 sequents of the grammar extracted from the Spoken Dutch Corpus (containing 49.156 atomic formulas and 23.094 connectives: Grail performs 70.953 axioms out of over 1,1 billion possiblities).

6.4.2. Parser for Sanskrit

Keywords: *dependencies*, *morphology*, *syntax*.

Participant: Gérard Huet [correspondant].

Gérard Huet designed and implemented a syntax analyser for classical Sanskrit, guided by semantic net constraints expressing dependencies between the word forms [42]. Finite verb forms demand semantic roles, according to valency patterns depending on the voice (active, passive) of the form and the governance (transitive, etc) of the root. Conversely, noun/adjective forms provide actors which may fill those roles, provided agreement constraints are satisfied. Tool words are mapped to transducers operating on tagged streams, allowing the modelling of linguistic phenomena such as coordination by abstract interpretation of actor streams. The parser ranks the various interpretations (matching actors with roles) with penalties, and returns to the user the minimum penalty analyses, for final validation of ambiguities. This parser is part of a whole computational platform for the analysis of classical Sanskrit (cf. section 5.2).

6.4.3. Parsers for Large Grammars

Keywords: *Earley parser*, *context-free grammar*, *scalability*.

Participant: Benoît Sagot [correspondant].

In collaboration with Pierre Boullier, Benoît Sagot proposed a method which, in practice, allows to use parsers for languages defined by very large context-free grammars (over a million symbol occurrences) [36]. The idea is to split the parsing process in two passes. A first pass computes a sub-grammar which is a specialized part of the large grammar selected by the input text and various filtering strategies. The second pass is a traditional parser which works with the sub-grammar and the input text. This approach has been validated by practical experiments performed on an Earley-like parser running on a test set with two large context-free grammars.

6.4.4. Parser for Polish

Keywords: *Polish lexicon*, *Polish parser*, *corpus*, *named entity recognition*, *segmentation*, *spelling correction*, *tokenization*.

Participant: Benoît Sagot [correspondant].

During a 3-month stay at IPI PAN (Warsaw, Poland), Benoît Sagot developed a new set of tools and resources for Polish which covers all the steps required to transform a raw unrestricted text into a reasonable input for a parser [55]. This includes (1) a large-coverage morphological lexicon, developed thanks to the IPI PAN corpus as well as a lexical acquisition technique, and (2) multiple tools for spelling correction, segmentation, tokenization and named entity recognition. This processing chain is also able to deal with the XCES format both as input and output, hence allowing to improve XCES corpora such as the IPI PAN corpus itself. This allows a brief qualitative evaluation of the lexicon and of the processing chain.

6.5. Corpus

6.5.1. *Bilingual Text Alignment*

Keywords: *Chinese*, *French*, *bilingual text alignment*, *corpus*.

Participant: Kim Gerdes [correspondant].

Bilingual corpora are essential for the construction of bilingual resources just as for any other work in translation studies, but the alignment itself needs bilingual resources or important interventions of bilingual speakers. Kim Gerdes has work in progress on bilingual text alignment with a dynamic time warping algorithm. All other algorithms work with bilingual resources and on the assumption of similarities between source and target languages (lexical or punctuation cognates): the signal alone of the corpus to be aligned is compared with the signals of the words in the target text. Gerdes's approach gives reasonably good results even for language couples like French (rich inflection) and Chinese (isolating language and lingua continua). A preliminary version of the bilingual text alignment tool has been made available (cf. section 5.16).

6.5.2. *Ressource-Free Corpus Construction and Partial Annotation of Light Verb Constructions for Farsi*

Keywords: *Farsi*, *Persian*, *annotation*, *corpus*, *learning algorithm*.

Participant: Kim Gerdes [correspondant].

In collaboration with Pollet Samvelian, Kim Gerdes have worked on the construction and partial annotation of a corpus of Persian [39]. The authors' goal in this work is the extraction and quantification of light verb constructions. The long term lexicological goal is to obtain reasonably complete lists of verbal lexical functions for a large set of nouns. The corpus is spidered from the web (cf. section 5.17), cleaned, and morphological learning algorithms are paired with heuristics (cf. section 5.18). The first results gave simple quantified lists of light verb constructions in Persian.

6.5.3. *A French Lexicon for NLP Applications*

Keywords: *French lexicon*, *Lefff*, *Lexicon-Grammar tables*.

Participant: Benoît Sagot [correspondant].

Benoît Sagot continued his work aiming at developing a large-coverage morphological and syntactic lexicon for French (Le*fff*, cf. section 5.11). In collaboration with Laurence Danlos, he devised a way to incorporate in the Le*fff* informations orginating from the Lexicon-Grammar tables, whose development has been initiated by Maurice Gross. Sagot and Danlos illustrated and implemented this approach on one kind of non-standard verbal and adjectival entries: impersonal structures [26]. In collaboration with Karën Fort, Benoît Sagot implemented another Le*fff* enrichment with information originating the Lexicon-Grammar tables concerning adverbs ending with *-ment* [54].

In collaboration with Laurence Danlos, Benoît Sagot compared Dicovalence and the Lexicon-Grammar of full verbs, two syntactic lexical resources for French that have been developed by linguists for numerous years. They focused on differences and overlaps between both underlying lexical models. They showed that the Le*fff* model is able to integrate lexical information present in the Lexicon-Grammar and in Dicovalence [38]. The long term goal of this study is the consitution of a high-quality syntactic lexicon for NLP.

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

Participants: Maxime Amblard, Houda Anoun, Christian Bassac, 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 teams 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 latter 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, Tristan Vanrullen.

Mosaïque is directed by Lionel Clément and 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 aims 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, *etc...*) in order to reuse different syntactic descriptions and to develop a high-level editor dedicated to syntax for linguists. The hypothesis is that many existing formalisms share a lot of things, even if they look different in a technical way.

The ARC Mosaïque has finished in December 2007. It lead to several events: ten scientific meetings and a Workshop organized as part of the conference TALN07. The project also settled the bases for an ongoing implementation of a generic tool to design grammars.

http://mosaique.labri.fr/

7.2.3. *ANR GEONTO Programme Masse de Données et COnnaissances*

Participant: Mauro Gaio.

The GEONTO project is a research action involving 5 teams (LRI, LIUPPA, COGIT, IRIT). It focuses on interoperability of diverse data related to geographic information. The first part of the project consists in building several geographic ontologies reflecting several (*different* might be better) points of view. In order to complete these objectives, various approaches relying on techniques taken from natural language processing will be used. The second part of the project will study the alignment of ontologies built in the previous part.

7.2.4. ANR blanche PRELUDE

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

Signes is part of the national research program PRELUDE *Towards theoretical pragmatics 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.

7.2.5. ANR RHAPSODIE

Participants: Lionel Clément, Kim Gerdes.

The ANR RHAPSODIE aims at building an annotated corpus of spoken French. This corpus will contain several kinds of discourses and its annotations will focus both on prosody and syntax so as to understand the status of prosody in spoken French and its relation with syntax and the informationnal structure. Lionel Clément will be more specifically concerned with the syntactic annotation of a textual transcription of the audio corpus and the format of intonational and syntactic annotations. The corpus will be freely distributed within the French community.

7.2.6. ILF Project LexSynt

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

LexSynt is a research action (2000-2008) 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.

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://yquem.inria.fr/~huet/EA/Rapport2007.html

8. Dissemination

8.1. Involvement 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

- Kim Gerdes is co-editor of the proceedings of the third International Conference on Meaning Text Theory (MTT) (May 21 - 24, 2007, Klagenfurt, Austria).
- Marie-Laure Guénot is secretary of the *Association pour le Traitement Automatique des Langues* (ATALA).

- 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 has been a member of the program committee of the *Fourth International Workshop on Generative Approaches to the Lexicon* held in Paris in May 2007, and of the International conference on adjectives held in Lille in october 2007.

- Lionel Clément, Renaud Marlet, Christian Retoré and Benoît Sagot were members of the program committee of the *Atelier sur les formalismes syntaxiques de haut niveau* (high-level syntactic formalisms), which was a satellite workshop of the TALN 2007 conference (Toulouse, France).

- Marie-Laure Guénot and Richard Moot were on the reading committee of Traitement Automatique des Langages Naturelles 2007 (Toulouse).

- Alain Lecomte and Christian Retoré were members of the Program Committee of the workshop *New directions in Type-Theoretical Grammars*, held during ESSLLI 07 (Dublin).

- Alain Lecomte and Christian Retoré were members member of the Program Committee of *Journées de Sémantique et Modélisation*, Paris.

- Alain Lecomte was a member of the Program Committee of WOLLIC 07 (Rio de Janeiro).

- Alain Lecomte was a member of the Program Committee of FG 07 (Formal Grammar), held during ESSLLI 07 (Dublin)

- Alain Lecomte was a member of the Program Committee of the First Workshop on Computational Sanskrit (Rocquencourt, October, 29-30)

- Henri Portine is a member of the Program Committee of the journal *Information grammaticale*.

- Christian Retoré was on the committee of Mathematics of Language held in Los- Angeles in July 2007.

- Christian Retoré was on the committee of the workshop on Semantics Beyond Set Theory held in Paris, October 18th-19th 2007.

8.1.4. Academic committees

- Renaud Marlet was a member of the associate scientist (i.e., chargé de recherche) hiring committees for INRIA Bordeaux - Sud-Ouest and for INRIA Lille - Nord Europe.

- Christian Bassac is a member of the hiring committee in linguistics of Université Bordeaux 3.

- Kim Guerdes is a member of the hiring commitee of Université Sorbonne nouvelle.

- 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 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.

- Henri Portine is the head of the *Association des Linguistes de l'Enseignement supérieur (ALES)*.

- Henri Portine has been expert member for the *Mission scientifique, technique et pédagogique du Ministère de la recherche* until September 15th 2007.

- Henri Portine is expert member for the *Direction générale de la recherche et de l'innovation* concerning human sciences (DGRI-A6).

- 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 head of the master on Algorithms and Formal Methods of Université Bordeaux 1.

- Christian Retoré was on the evaluating committee of the Laboratoire d'Informatique Fondamentale (CNRS et Université d'Aix-Marseille) which met in January 2007.

- Christian Retoré is a member of the committee for the SPECIF best PhD award (SPECIF French academic computer-science society).

- Christian Retoré is a member of the Standing Committee of the European Summer School in Logic and Language and Information since August 2007.

8.1.5. Organization of events

- Christian Bassac was a member of the organization committee of the *International Conference on the Linguistics of Contemporary English* held in Toulouse, on the 2nd-4th July 2007.

- Alain Lecomte co-organized (with Laurent Roussarie and Brenda Laca) the "Journées de Sémantique et Modélisation 07" at Paris - Saint-Denis.

- April Alain Lecomte co-organized (with M. Quatrini, M-R. Fleury and S. Tronçon) the "Workshop Prelude" at Carry-le-Rouet, June 13-15.

- Bruno Mery is organizing the weekly seminar Linguistique et informatique Universités Bordeaux 1 et 3.

- Lionel Clément, Marie-Laure Guénot, Renaud Marlet, Christian Retoré, Benoît Sagot and Tristan Vanrullen co-organized the *Atelier sur les formalismes syntaxiques de haut niveau* (high-level syntactic formalisms), which was a satellite workshop of the TALN 2007 conference (Toulouse, France).

- Reinhart Muskens has organized the Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007) as part of the European Summer School on Logic, Language and Information 2007, (ESSLLI 2007), which was be held in Dublin, 6 - 17 August, 2007. The workshop dates were August 6 - 10, 2007.

- Christian Retoré together with Sylvain Pogodalla (INRIA Calligramme) and Myriam Quatrini (IML, MArseille) organised a workshop in honor of Alain Lecomte in Pauillac (Novembre 2-3 2007).

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

- Christian Bassac taught 12 hours in a master 2 seminar on logic and linguistics. He also taught 15 hours on English syntax for students taking the agregation.

- *Structures Informatiques et Logiques pour la Modélisation Linguistique*, Parisian Master of Research in Informatics (MPRI). (Gérard Huet, Philippe de Groote)

- Anna Kupsc gave a Licence 3 semester course at University of Bordeaux 3, on *Linguistics and computer science*.

- Alain Lecomte gave a Master 2 seminar at Ecole Normale Supérieure (master of "Linguistique Formelle et Théorique", co-organized by University Paris 8 and Department of Cognitive Studies at ENS) on "Formal and Logical Grammars".

- Henri Portine gives a course entitled *Utterance acts, semantics and history of linguistics (from 18th to 20th centuries)*, at Université Bordeaux 3, 5th year in linguistics.

- Logique et Languages Bordeaux 1, 5 t h year in computer science (Bruno Courcelle, Christian Retoré)

- Traitement Automatique des Langages Naturelles, Bordeaux 1, 4 t h year in computer science (Lionel Clément, Christian Retoré)

- Christian Retoré was invited for two weeks at the Università degli Studi di Verona where he gave a 12 hours master course on *Categorial grammar and the logical structure of sentences*.

- On November 29th, Christian Retoré gave a talk at the *Université ouverte* in Cergy entitled *Modéliser le langage: la linguistique computationnelle au carrefour de l'informatique, de la linguistique, des sciences cognitives et des outils pour l'internet*.

8.3. Defended Theses

- Maxime Amblard defended his PhD thesis on *Calculs de représentations sémantique et syntaxe générative : les grammaires minimalistes catégorielles* at the Université de Bordeaux (September 24th 2007).

- Houda Anoun defended her PhD thesis on *Approche logique des grammaires pour les langues naturelles* at the Université de Bordeaux (October 24th 2007).

- Renaud Marlet defended his Habilitation à Diriger les Recherches (HDR) on *Spécialiser les programmes, spécialiser les langages* at the Université Bordeaux 1 (November 23rd, 2007) [18].

8.4. Thesis Juries

- Gregory Kobele, Alain Lecomte and Christian Retoré were on the jury of the PhD of Maxime Amblard entitled *Calculs de représentations sémantiques et syntaxe générative : les grammaires minimalistes catégorielles*, defended on September 24th 2007 at université de Bordeaux.

- Alain Lecomte was on the jury of the PhD of Joseph Le Roux entitled *La coordination dans les grammaires d'interaction* on October 17th 2007 at Institut National Polytechnique de Lorraine.

- Alain Lecomte and Sylvain Salvati were on the jury of the PhD of Houda Anoun entitled *Approche logique des grammaires pour les langues naturelles* on October 24th 2007 at université de Bordeaux.

- Alain Lecomte was on the jury of the Habilitation à diriger les recherches of Jean-Baptiste Joinet on *Logique et interaction* on december 12th 2007 at université Paris 1.

- Alain Lecomte was on the jury of the PhD of Laurent Keiff entitled *Le pluralisme logique. Vers une dynamique. Eléments pour une étude des interactions dynamiques entre la sémantique dialogique et certains contextes de l'activité rationnelle. La cladistique. Les taxonomies " primitives"* on December 8th 2007 at Université de Lille 3.

- Richard Moot was on the jury of the PhD of Matteo Capelletti entitled *Parsing with structure-preserving categorial grammars* defended on the July 9th 2007 at the University of Utrecht.

- Christian Retoré was on the PhD committee of Véronique Moriceau *Intégration de données dans un système question-réponse sur le Web* (Université Paul Sabatier, Toulouse, February 15th 2007).

8.5. Academic supervision

- Christian Retoré and Greg Kobele supervised the master thesis of Benoît Wagler entitled: *Analyse syntaxique dans les grammaires minimalistes* defended in june 2007.
- Richard Moot and Reinhart Muskens supervised the master thesis of Pierre Schweitzer entitled: *Introduction à l'Interface Syntaxe-Sémantique* defended in september 2007.

8.5.1. *Student internship supervision – fourth and fifth year*

- Within ARC Mosaïque and in collaboration with Université de Provence (Aix-Marseille) and the Laboratoire Parole et Langage (LPL, Aix en Provence), Lionel Clément and Tristan Vanrullen are co-supervising the internship of Xavier Cornillier on *Validation d'un système formel pour la représentation de phénomènes linguistiques non syntaxiques connectés à la syntaxe*.

8.5.2. *PhD supervision*

- Christian Bassac supervises the thesis work of Olivier Taïs on *The Dynamics of the Lexicon*.
- Christian Bassac and Christian Retoré are co-supervising the thesis work of Bruno Mery on *Type theory for lexical semantics* (Université Bordeaux 1).
- Mauro Gaio is supervising with Thierry Nodenot the thesis work of Pierre Loustau *From displacements to itinerary, from clauses to discourse.* (Université de Pau et des Pays de l'Adour).
- Alain Lecomte and Christian Retoré were co-supervising the thesis work of Maxime Amblard, *Calcul de représentations sémantiques dans les grammaires minimalistes* (defended in September 2007 at Université Bordeaux 1).
- Alain Lecomte is supervising with Pierre Castéran the thesis work of Houda Anoun on Proof theoretic methods in computational linguistics (defended in October 2007 at Université Bordeaux 1).
- Alain Lecomte is supervising the thesis work of Christophe Onambele, *Grammaires Minimalistes et Applications à la Description Linguistique*, (Université Paris 8).
- Henri Portine and Renaud Marlet are co-supervising the thesis work of Emilie Voisin (Université Bordeaux 3) on the analysis and modelling of French Sign Language (LSF).
- Christian Retoré and Sylvain Salvati are co-supervising the thesis of Pierre Bourreau Université Bordeaux 1 (supported by CORDI funds) on *the treatment and the use of non-linearity in computationnal linguistics*.

8.6. Participation to colloquia, seminars, invitations

8.6.1. *Visiting scientists*

- February 26th 2007, Jacques Lamarche (département d'études françaises, Université Western Ontario) *Vers une définition combinatoire des catégories grammaticales*.
- March 26th 2007, Nicholas Asher (CNRS - IRIT Toulouse) *sémantique lexicale formelle, coercion, coprédication et une théorie de prédication typée*
- October 15th 2007, Pascal Amsili (Université Paris 7) *Les présuppositions obligatoires : contraintes discursives et pragmatiques*.
- November 12th 2007, Sylviane Schwer (Université Paris 13), *Smots et Slangages*.

- Ryo Yoshinaka (INRIA Lorraine) spent one week in Bordeaux to work with Sylvain Salvati on Kanazawa's interpolation theorem [80]. This theorem is the basis of several learning algorithms which infer the sentence meaning relation from a finite number of examples.

- Michael Moortgat invited professor (Université de Bordeaux) ooctober 29th -10th and december 10th-22nd.

8.6.2. Seminar Talks, Invitations

- Maxime Amblard and Christian Retoré presented their joint work on *normalisation of Partialy Commutative Linear Calculi* at Computation in Europe, CiE 2007.

- Maxime Amblard presented his works on syntax semantic interface at France Telecom R&D.

- Maxime Amblard presented his works on Syntax-Semantic interface at the colloquium in honor of Alain Lecomte.

- Marie-Laure Guénot gave an invited talk on *Hiérarchie de constructions verbales dans une grammaire du français* , during the *Journées d'étude sur les Grammaires de Construction*, in Université Paris X Nanterre, 27-28 septembre 2007.

- Marie-Laure Guénot gave an invited talk on *Des descriptions de corpus au traitement automatique des langues : Une grammaire formelle du français parlé* during *Séminaire ERSSàB*, Université du Mirail, Toulouse 2, 22 mars 2007.

- Marie-Laure Guénot gave an invited talk on *La formalisation des entassements paradigmatiques dans une grammaire du français* during *Séminaire du Lattice*, Université Denis Diderot, Paris 7, 19 mars 2007.

- Marie-Laure Guénot presented her work on *Une grammaire du français pour une théorie descriptive et formelle de la langue* at TALN conference in Toulouse June 6-8th 2007.

- Kim Gerdes and Pollet Samvelian presented their work on *Ressource-Free Corpus Construction and Partial Annotation of light verb constructions: Preliminary results on Persian journalistic corpora* at the Workshop on Cognitive and Computational Aspects of Extended Lexical Units (ELU2007), Leiden, December 2007.

- Kim Gerdes gave a talk during the Master seminar "Génération automatique de textes" at Université Paris 3.

- Kim Gerdes gave an invited talk at Rencontres Jeunes Chercheurs en Parole 2007 (RJCP 2007).

- Anna Kupsc gave a talk entitled *A comparative study of past tense and conditional constructions across Slavic languages* at the ERSSàB/SIGNES seminar, on November 19th 2007.

- Anna Kupść gave a talk at the Master TAL Seminar, ERRS, on 16 November 2007 entitled *Chronological order of events in encyclopedic texts*.

- Anna Kupść is programe committee member of the 15the International Conference on Head-driven Phrase Structure Grammar, Kyoto, July 2008.

- Alain Lecomte gave an invited talk at the seminar of UMR 7023 ("Structures Formelles de la Langue"), November, 9, 2006, Paris, Université Paris 8.

- Alain Lecomte gave several talks related to ANR-Prelude:
    - "les Journées Prelude" February, Ecole Normale Supérieure,
    - Workshop Prelude, Carry-le-Rouet, June, 13-15,
    - "Journée Prelude", October, 31, Pauillac.

- Bruno Mery presented his work on *A Montagovian Generative Lexicon* at FG'07.

- Bruno Mery presented his work on A Montague-based model of Generative Lexical Semantics at the workshop on *New Direction in Type Theoretic Grammars*, organized as part of ESSLLI'07.

- Renaud Marlet presented a poster on *When the Generative Lexicon meets Computational Semantics* at the 4th International Workshop on Generative Approaches to the Lexicon (GL 2007, Paris, France).

- Richard Moot presented his work on filtering axiom links for proof nets at Formal Grammar, 4-5 August 2007.

- Richard Moot gave an invited talk on proof nets for display logic at Utrecht University, 9 july 2007 .

- Richard Moot gave an invited talk on proof nets and minimalist grammars at Pauillac, 2 November 2006.

- Henri Portine presented his work on *La préposition* avec *et l'effet de cadrage* at the conference *autour de la préposition* (Université de Caen, September 20-22nd 2007).

- Henri Portine presented his work (joint with Agnès Bracke) on *Décrire les zones frontières des lexèmes scientifiques: Quand une approche prototypique croise une approche sémique* at *Journées d'études* Sémantique et lexicologie des langues d'Europe: Des aspects théoriques aux applications, *Réseau Euralangues*, Université de Lille 3, October 22-23rd 2007.

- Lionel Clément and Tristan Vanrullen presented the ARC Mosaïque project at the journées des ARCs (Octobre 1-2nd 2007, Rennes, France).

- Christian Retoré presented his work on *Perfect matchings and directed cographs: proofnets for pomset logic and the origins of deep inference* at the workshop on Deep Inference at the Ecole Polytechnique, Palaiseau on June 21st 2007.

- Christian Retoré gave an invited lecture *On the trees underlying natural language sentences* at the Colloquium in Honor of Gérard Huet, held in ENS Paris, June 22nd-23rd 2007.

- Christian Retoré gave a talk *What do we know about tree languages used in syntax?* at the workshop on Lambek's formalisms held in Toulouse, June 9th.

- Christian Retoré was invited for two weeks at the Università degli Studi di Verona in MAy 2007 where he gave a 12 hours master course on *Categorial grammar and the logical structure of sentences*.

- Christian Retoré gave a talk at the *Université ouverte* in Cergy on November 29th entitled *Modéliser le langage: la linguistique computationnelle au carrefour de l'informatique, de la linguistique, des sciences cognitives et des outils pour l'internet*.

- Cap Science invited Émilie Voisin to debate around a movie *l'esquive* by Abdelatif Kechiche (March 2007).

- Émilie Voisin presented her work in Toulouse (ERSS team) (March 2007).

- Denis Delfitto invited Émilie Voisin to present her work at the university of Verona (April 2007).

- Émilie Voisin went to Lille for the Aflico colloquium and, with Loïc Kervajan, presented a formalization to treat the mechanisms of modifiers in LSF (May 2007).

- The university of Aix en Provence invited Émilie Voisin (LPL team) to show her work aroud the theme of nominal incorporation (December 2007).

- Sylvain Salvati was invited to present his work on *Lambek Derivations* at the fourth workshop on Lambda Calculus and Formal Grammars in Nancy (18-19 September 2007)

- Sylvain Salvati presented his work on *the Complexity of Abstract Categorial Grammars* at the tenth workshop on Mathematics of Language in Los Angeles (28-30 July 2007).

- Sylvain Salvati presented his work on *the Membership Problem for Non-linear Abstract Categorial Grammars* at the Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007) organized as part of the European Summer School on Logic, Language and Information 2007, (ESSLLI 2007), which was be held in Dublin, 6 - 17 August, 2007.

- Sylvain Salvati presented his work on *the Non-Associative Categorial Grammars and Abstract Categorial Grammars* Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007) organized as part of the European Summer School on Logic, Language and Information 2007, (ESSLLI 2007), which was be held in Dublin, 6 - 17 August, 2007.

- Gregory Kobele presented his work on a *Formal Foundation for A and A-bar Movement* at the tenth workshop on Mathematics of Language in Los Angeles (28-30 July 2007).

- Gregory Kobele presented his work on *An Automata-Theoretic Approach to Minimalism* at the workshop on Model Theoretic Syntax at 10 organized as part of the European Summer School on Logic, Language and Information 2007, (ESSLLI 2007), which was be held in Dublin, 6 - 17 August, 2007.

8.6.3. Participation to conferences and summer schools

- Christian Bassac attended the Summer School on the Method of Lexical Exceptions in the University of St Andrews, 2nd Đ 8th September 2007.

- Maxime Amblard has participated to the Prelude session of the mars 16th 2007.

- Gregory Kobele, Renaud Marlet, Bruno Mery, Christian Retoré, Sylvain Salvati, Emilie Voisin attended the *Journées de Sémantique et Modélisation, JSM'07*

- Houda Anoun (first day only), Maxime Amblard, Christian Bassac, Pierre Bourreau, Patrick Henry (first day only), Anna Kupsc, Alain Lecomte, Renaud Marlet, Bruno Méry (first day only), Richard Moot, Christian Retoré, Sylvain Salvati, Emilie Voisin attended the *workshop in honor of Alain Lecomte* on November 2-3 2007 at Pauillac.

- Christian Bassac, Bruno Mery, Renaud Marlet attended the *4th International Workshop on Generative Approaches to the Lexicon, GL'07*

- Bruno Mery, Sylvain Salvati attended the *12th Conference on Formal Grammar, FG'07*

- Kim Gerdes attended the *third International Conference on Meaning Text Theory (MTT)* (May 21 - 24, 2007, Klagenfurt, Austria).

- Alain Lecomte attended the second week of the *European School on Language, Logic and Information, ESSLLI'07,*

- Gregory Kobele, Renaud Marlet, Bruno Mery, Sylvain Salvati attended the *European School on Language, Logic and Information, ESSLLI'07.*

- Richard Moot attended the following Prélude Meetings: 29 November 2006 Paris, 6 February 2007 Paris, 31 october, 1 november 2007 Pauillac.

- Gregory Kobele, Renaud Marlet, Bruno Mery and Sylvain Salvati attended the 19th European Summer School in Logic, Language and Information (ESSLLI) in Dublin, Ireland (6-17 August, 2007).

- Lionel Clément, Marie-Laure Guénot, Anna Kupsc, Renaud Marlet, Benoît Sagot and Tristan Vanrullen attended the ARC Mosaïque meetings (March 2nd, May 14th, September 24th, December 6-7th, 2007) as well as the *Atelier sur les formalismes syntaxiques de haut niveau* (high-level syntactic formalisms), which was a satellite workshop of the TALN 2007 conference (June 8th, 2007, Toulouse, France).

- Christian Retoré attended the TALN conference in Toulouse June 6-8th the Mosaïque workshop, the Workshop on Sign Languages Processing (9th), and the workshop on Lambek Formalisms (10th).

- Christian Retoré attended the workshop on Deep Inference (June 21st) the workshop in honor of Gérard Huet (June 22nd-23rd).

- Sylvain Salvati attended all the Prelude meetings from February 2007 (*i.e.* Paris (6 February 2007, 16 March 2007), Carry-le-Rouet (15 June 2007), Pauillac (31 October 2007))

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 recherches, Université Michel de Montaigne Bordeaux 3, December 2006.

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

[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", vol. 15, n° 4, July 2005, p. 573–614.

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

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

[8] R. MOOT. *Proof nets for linguistic analysis*, Ph. D. Thesis, UIL-OTS, Universiteit Utrecht, 2002.

[9] 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.

[10] E. NYBERG, T. MITAMURA, J. P. CALLAN, J. G. CARBONELL, R. E. FREDERKING, K. COLLINS-THOMPSON, L. HIYAKUMOTO, Y. HUANG, C. HUTTENHOWER, S. JUDY, J. KO, A. KUPSC, L. V. LITA, V. PEDRO, D. SVOBODA, B. V. DURME. *The JAVELIN Question-Answering System at TREC 2003: A Multi-Strategh Approach with Dynamic Planning*, in "TREC", 2003.

[11] H. PORTINE. *La syntaxe de Damourette et Pichon comme outil de représentation du sens*, in "Modèles linguistiques", vol. 23, n° 2, 2002, p. 21–46.

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

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

[14] S. SALVATI. *Problèmes de filtrage et problèmes d'analyse pour les grammaires catégorielles abstraites*, Ph. D. Thesis, Institut National Polytechnique de Lorraine, 2005.

Year Publications

Books and Monographs

[15] K. GERDES, T. REUTHER, L. WANNER (editors). *Proceedings of the 3rd international conference on Meaning-Text Theory (MTT 2007), Klagenfurt, Austria (May 20-24, 2007)*, Wiener Slawistischer Almanach, Linguistische Reihe, Special Issue (Sonderband), n° 69, Kubon & Sagner, Vienna, July 2007.

Doctoral dissertations and Habilitation theses

[16] M. AMBLARD. *Calcul de représentations sémantiques et syntaxe générative : les Grammaires Minimalistes Catégorielles*, Ph. D. Thesis, Université de Bordeaux 1, September 2007.

[17] H. ANOUN. *Approche logique des grammaires pour les langues naturelles*, Ph. D. Thesis, Université de Bordeaux 1, 2007.

[18] R. MARLET. *Spécialiser les programmes, spécialiser les langages*, Habilitation à diriger les recherches, Université de Bordeaux 1 / INRIA, November 2007.

Articles in refereed journals and book chapters

[19] D. BECHET, R. BONATO, A. DIKOVSKY, A. FORET, Y. LE NIR, E. MOREAU, C. RETORÉ, I. TELLIER. *Modèles algorithmiques de l'acquisition de la syntaxe : concepts et méthodes, résultats et problèmes*, in "Recherches Linguistiques de Vincennes", vol. 36, 2007, p. 123–152, http://rlv.revues.org/.

[20] M. GAIO, C. SALLABERRY, P. ETCHEVERRY, C. MARQUESUZAÀ, J. LESBEGUERIES. *A Global Process to Access Documents' Contents from a Geographical Point of View*, in "Journal of Visual Languages and Computing", Special Issue on Spatial and Image-Based Information Systems, vol. 18, n° 6, December 2007.

[21] K. GERDES, S. KAHANE. *Phrasing it differently*, in "Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'čuk.", L. WANNER (editor), Benjamins, Amsterdam, Philadelphie, 2007, p. 293–332.

[22] P. GUITTENY. *La communication médiatisée par les technologies de l'information et la communauté des sourds*, in "La langue du cyberespace : de la diversité aux normes", J. GERBAULT (editor), To appear, L'Harmattan, 2007.

[23] A. LECOMTE. *Conversion en logique et dans quelques domaines appliqués : mots, gènes et programmes*, in "Le mot : conversion", J. BAILLÉ (editor), Presses Universitaires de Grenoble, 2007, p. 113–138.

[24] A. LECOMTE. *La "grammaire pure" : une grammaire des normes*, in "Volume d'hommage à J.-L. Gardies", J.-M. LARDIC (editor), To appear, Vrin, Paris, 2007.

[25] H. PORTINE. *La construction "se devoir (à soi-même) de Vinf" et sa valeur modale*, in "Problèmes de sémantique et de syntaxe : Hommage à André Rousseau", L. BEGIONI, C. MULLER (editors), Coll. UL3, Université Charles-de-Gaulle - Lille 3, 2007, p. 273–286.

[26] B. SAGOT, L. DANLOS. *Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : constructions impersonnelles*, in "Cahiers du Cental", To appear, 2007.

[27] C. SALLABERRY, M. GAIO, J. LESBEGUERIES, P. LOUSTAU. *A semantic approach for geospatial information extraction from unstructured documents*, in "The Geospatial Web", A. SCHARL, K. TOCHTERMANN (editors), Advanced Information and Knowledge Processing Series, ISBN 1-84628-826-6, Springer, 2007, p. 3–105.

[28] É. VOISIN, L. KERVAJAN. *Typologie des verbes et forme verbale non marquée en Langue des Signes Française : incidences sur l'organisation syntaxique*, in "Silexicales", n° 5, 2007.

[29] É. VOISIN. *Les formes verbales non finies en Langue des Signes Française*, in "Les formes non finies du verbe 2", F. LAMBERT, C. MOREAU, J. ALBRESPIT (editors), Proceedings of the 20th Colloque du CerLiCO, Bordeaux, France (June 2006), n° 20, Presses Universitaires de Rennes, 2007.

[30] É. VOISIN. *Manifestations syntaxiques de la diglossie : la cas du français signé*, in "Revue des étudiants en linguistique du Québec", To appear, 2007.

Publications in Conferences and Workshops

[31] M. AMBLARD, C. RETORÉ. *Natural Deduction and Normalisation For Partially Commutative Linear Logic and Lambek Calculus with product*, in "Computation and Logic in the Real World (Computing in Europe 2007)", S. B. COOPER, T. F. KENT, B. LÖWE, A. SORBI (editors), Quaderni del Dipartimento di Scienze Matematiche e Informatiche "Roberto Magari", vol. ID487, Università degli Studi di Siena, september 2007, p. 28–35, http://www.mat.unisi.it/~sorbi/Proceedings.pdf.

[32] H. ANOUN, A. LECOMTE. *Linear grammars with labels*, in "Proceedings of the 11th conference on Formal Grammar (FG 2006), Malaga, Spain (June 2006)", ISSN 1935-1569, CSLI Publications, 2007, p. 21–34, http://cslipublications.stanford.edu/FG/.

[33] H. ANOUN, A. LECOMTE. *Towards a Logical Account of Binding Theory*, in "7th International Tbilisi Symposium on Language, Logic and Computation (Tbilisi 2007)", October 2007.

[34] C. BASSAC. *Events and reference*, in "First Res Per Nomen colloquium on Naming, Referring and Usage in Linguistics, Reims, France", P. FRATH (editor), May 2007.

[35] C. BASSAC. *Rules and exceptions: the Junggrammatiker and Verner's law*, in "International Conference on the Method of Lexical Exceptions, St Andrews, Scotland", C. BEEDHAM (editor), September 2007.

[36] P. BOULLIER, B. SAGOT. *Are very large context-free grammars tractable?*, in "Proceedings of 10th International Conference on Parsing Technologies (IWPT 2007), Prague, Czech Republic", June 2007.

[37] A. BRACKE, H. PORTINE. *Décrire les zones frontières des lexèmes scientifiques : Quand une approche prototypique croise une approche sémique*, in "Journées d'études "Sémantique et lexicologie des langues d'Europe : des aspects théoriques aux applications" (réseau thématique Euralangues), Lille, France", Université Charles-de-Gaulle - Lille 3, October 2007.

[38] L. DANLOS, B. SAGOT. *Comparaison du Lexique-Grammaire des verbes pleins et de* DICOVALENCE *: vers une intégration dans le Lefff*, in "Actes de la 14e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2007), Toulouse, France", IRIT Press, June 2007, p. 229–238.

[39] K. GERDES, P. SAMVELIAN. *Ressource-free corpus construction and partial annotation of light verb constructions: preliminary results on Persian journalistic corpora*, in "Proceedings of the Workshop on Cognitive and Computational Aspects of Extended Lexical Units (ELU 2007), Leiden, The Netherlands", December 2007.

[40] M.-L. GUÉNOT. *Une grammaire du français pour une théorie descriptive et formelle de la langue*, in "Actes de la 14e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2007), Toulouse, France", IRIT Press, June 2007, p. 155–164.

[41] P. HENRY, C. BASSAC. *A toolkit for a Generative Lexicon*, in "4th International Workshop on Generative Approaches to the Lexicon (GL 2007), Paris, France", P. BOUILLON, K. KANZAKI (editors), Poster, May 2007.

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

[43] M. KANAZAWA, S. SALVATI. *Generating control languages with Abstract Categorial Grammars*, in "Proceedings of the 12th conference on Formal Grammar (FG 2007), Dublin, Ireland", L. KALLMEYER, P. MONACHESI, G. PENN, G. SATTA (editors), To appear, ISSN 1935-1569, CSLI Publications, August 2007, http://cslipublications.stanford.edu/FG/.

[44] L. KERVAJAN, É. VOISIN. *Analyse et traitement des modificateurs en Langue des Signes Française*, in "Deuxième Colloque International de l'Association Française de Linguistique Cognitive (AFLICO 2007)", To appear, 2007.

[45] G. KOBELE. *A formal foundation for A and A-bar movement*, in "10th conference on Mathematics of Language (MOL 10), Los Angeles, CA", M. KRACHT, G. PENN, E. STABLER (editors), July 2007.

[46] G. KOBELE, C. RETORÉ, S. SALVATI. *An automata-theoretic approach to Minimalism*, in "Workshop on Model Theoretic Syntax (MTS@10), Dublin, Ireland", J. ROGERS, S. KEPSER (editors), MOL, August 2007, http://cs.earlham.edu/esslli07mts/.

[47] R. MARLET. *When the Generative Lexicon meets computational semantics*, in "4th International Workshop on Generative Approaches to the Lexicon (GL 2007), Paris, France", P. BOUILLON, K. KANZAKI (editors), Poster, May 2007.

[48] B. MERY, C. BASSAC, C. RETORÉ. *A Montagovian Generative Lexicon*, in "Proceedings of the 12th conference on Formal Grammar (FG 2007), Dublin, Ireland", L. KALLMEYER, P. MONACHESI, G. PENN, G. SATTA (editors), To appear, ISSN 1935-1569, CSLI Publications, August 2007, http://cslipublications.stanford.edu/FG/.

[49] B. MERY, C. BASSAC, C. RETORÉ. *A Montague-based model of Generative Lexical Semantics*, in "Proceedings of the Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007), Dublin, Ireland", R. MUSKENS (editor), Foundation of Logic, Language and Information (FoLLI), August 2007, p. 90–97.

[50] 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.

[51] R. MOOT. *Filtering axiom links for proof nets*, in "Proceedings of the 12th conference on Formal Grammar (FG 2007), Dublin, Ireland", L. KALLMEYER, P. MONACHESI, G. PENN, G. SATTA (editors), To appear, ISSN 1935-1569, CSLI Publications, August 2007, http://cslipublications.stanford.edu/FG/.

[52] H. PORTINE. *La préposition avec et l'effet de cadrage*, in "International colloquium "Autour de la préposition"", Caen, France", Université de Caen, September 2007.

[53] C. RETORÉ, S. SALVATI. *Non-associative Categorial Grammars and Abstract Categorial Grammars*, in "Proceedings of the Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007), Dublin, Ireland", R. MUSKENS (editor), Foundation of Logic, Language and Information (FoLLI), August 2007, p. 51–58.

[54] B. SAGOT, K. FORT. *Améliorer un lexique syntaxique à l'aide des tables du Lexique-Grammaire : adverbes en -ment*, in "Proceedings of the 26th Colloque Lexique et Grammaire, Bonifacio, France", October 2007.

[55] B. SAGOT. *Building a morphosyntactic lexicon and a pre-syntactic processing chain for Polish*, in "3rd international Language and Technology Conference (LTC 2007), Poznań, Poland", Z. VETULANI (editor), Wydawnictwo Poznańskie Sp. z o. o., October 2007.

[56] C. SALLABERRY, M. BAZIZ, J. LESBEGUERRIES, M. GAIO. *Towards a geographical IE and IR system dealing with spatial information scope in textual digital libraries*, in "9th International Conference on Enterprise Information Systems (ICEIS 2007), Funchal, Madeira, Portugal", ACM Press, June 2007.

[57] S. SALVATI. *On the complexity of Abstract Categorial Grammars*, in "10th conference on Mathematics of Language (MOL 10), Los Angeles, CA", M. KRACHT, G. PENN, E. STABLER (editors), July 2007.

[58] S. SALVATI. *On the membership problem for Non-linear Abstract Categorial Grammars*, in "Proceedings of the Workshop on New Directions in Type-theoretic Grammars (NDTTG 2007), Dublin, Ireland", R. MUSKENS (editor), Foundation of Logic, Language and Information (FoLLI), August 2007, p. 43–50.

Internal Reports

[59] M. AMBLARD, É. VOISIN. *Vers une analyse automatique de la LSF : une modélisation de la Langue des Signes Française par les Grammaires Minimalistes*, Technical report, n⁰ HAL-00186965, 2007, http://hal.archives-ouvertes.fr/hal-00186965/.

[60] R. MOOT. *Proof nets for display logic*, Technical report, INRIA, 2007.

Miscellaneous

[61] A. LECOMTE, M.-R. FLEURY, M. QUATRINI. *Introduction à la ludique et applications à la pragmatique*, 2007, http://anr-prelude.fr/IMG/pdf/LudAppPrag.pdf.

References in notes

[62] G. ROZENBERG, A. SALOMAA (editors). *Handbook of Formal Languages*, Springer Verlag, Berlin, 1997.

[63] R. THOMASON (editor). *The collected papers of Richard Montague*, Yale University Press, 1974.

[64] J. VAN BENTHEM, A. TER MEULEN (editors). *Handbook of Logic and Language*, North-Holland Elsevier, Amsterdam, 1997.

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

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

[67] N. CHOMSKY. *The minimalist program*, MIT Press, Cambridge, MA, 1995.

[68] A. DIKOVSKY, L. MODINA. *Dependencies on the other side of the Curtain*, in "Traitement Automatique des Langues", vol. 41, n⁰ 1, 2000, p. 67-95.

[69] L. T. F. GAMUT. *Logic, Language and Meaning – Volume 2: Intensional logic and logical grammar*, The University of Chicago Press, 1991.

[70] 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.

[71] 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.

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

[73] 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.

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

[75] R. JACKENDOFF. *The Architecture of the Language Faculty*, Linguistic Inquiry Monographs, n⁰ 28, M.I.T. Press, Cambridge, Massachusetts, 1995.

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

[77] 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.

[78] 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.

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

[80] M. KANAZAWA. *Computing interpolants in implicational logics*, in "Annals of Pure and Applied Logic", vol. 142, 2006, p. 125–201.

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

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

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

[84] 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.

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

[86] 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.

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

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

[89] I. MELCUK. *Communicative Organization in Natural Language: The Semantic-communicative Structure of Sentences*, John Benjamins, 2001.

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

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

[92] 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.

[93] R. MOOT. *Proof nets for linguistic analysis*, Ph. D. Thesis, UIL-OTS, Universiteit Utrecht, 2002.

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

[95] 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.

[96] 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.

[97] 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.

[98] G. K. PULLUM, B. C. SCHOLZ. *On the distinction bewteen moedl-theoretic and generative-enumerative syntax*, in "Logical Aspects of Computational Linguistics (LACL 2001)", P. DE GROOTE, G. MORRILL, C. RETORÉ (editors), LNCS/LNAI, n⁰ 2099, Springer-Verlag, 2001, p. 17–43.

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

[100] C. RETORÉ, E. STABLER. *Generative Grammar in Resource Logics*, in "Journal of Research on Language and Computation", vol. 2(1), 2004, p. 3–25.

[101] J. ROGERS. *A Descriptive Approach to Language-Theoretic Complexity*, CSLI Publications, Stanford, California, 1998.

[102] H. SEKI, T. MATSUMURA, M. FUJII, T. KASAMI. *On multiple context free grammars*, in "Theoretical Computer Science", vol. 88, n⁰ 2, 1991, p. 191–229.

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

[104] 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.

[105] V. SUTTON. *Lessons in SignWriting*, 2002, http://www.signwriting.org.

[106] D. J. WEIR. *A geometric hierarchy beyond context-free languages*, in "Theoretical Computer Science", vol. 104, n⁰ 2, 1992, p. 235–261.

[107] R. YOSHINAKA, M. KANAZAWA. *The complexity and generative capacity of lexicalized Abstract Categorial Grammars*, in "5th international conference on Logical Aspects of Computational Linguistics (LACL 2005)", April 2005, p. 330–346.

[108] P. DE GROOTE. *Towards Abstract Categorial Grammars*, in "Proceedings of the 39th Annual Meeting on Association for Computational Linguistics (ACL 2001), Morristown, NJ, USA", Association for Computational Linguistics, July 2001, p. 252–259.

[109] 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.

[110] 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.