INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team SIGNES

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

Futurs

THEME SYM

Table of contents

1. Team **1**
2. Overall Objectives **1**
3. Scientific Foundations **2**
3.1. The center: natural language syntax and semantics 2
3.2. Word structure and automata: computational morphology 3
3.3. Sentence structure and formal grammars: syntax 3
3.4. Sentence structure and logic: the syntax/semantics interface 4
3.5. Lexical semantics and derivational morphology 5
3.6. Discourse and dialogue structure: computational semantics and pragmatics 5
3.7. Type systems and functional programming for computational linguistics 5
4. Application Domains **6**
4.1. Sanskrit philology 6
4.2. Towards French Sign Language (LSF) modelling and processing 6
5. Software **7**
5.1. The Zen toolkit 7
5.2. DepLin 7
5.3. Grail 3: natural language analysis with multimodal categorial grammar 8
5.4. Experiments in categorial grammars 8
6. New Results **8**
6.1. Flexional morphology 8
6.2. Robust syntax 9
6.3. Syntax semantics interface for generative grammars 9
6.4. Lexical semantics 9
6.5. Modeling French sign language (LSF) grammar 10
7. Contracts and Grants with Industry **10**
7.1. PicoPeta 10
8. Other Grants and Activities **10**
8.1. Regional research programs 10
8.2. National research programs 10
8.2.1. Groupement de Recherche C.N.R.S. 2521 Sémantique et modélisation 10
8.2.2. Programme Interdiciplinaire du C.N.R.S. Traitement des Connaissances, Apprentissage et Nouvelles Technologies de l'Information et de la Communication 11
8.3. European research programs 11
8.3.1. CoLogNet: European network of Excellence on Computational logic 11
8.3.2. UIL-OTS Utrecht — Signes (Action intégrée van Gogh) 11
9. Dissemination **11**
9.1. Activism within the scientific community 11
9.1.1. Honours 11
9.1.2. Editorial boards 11
9.1.3. Program committees of conferences 12
9.1.4. Academic committees 12
9.1.5. Organization of events 12
9.2. Teaching 13
9.3. Thesis Juries 13
9.4. Academic supervision 14
9.4.1. Student intern supervision – fifth year 14
9.4.2. PhD supervision 14

9.5. Participation to colloquia, seminars, invitations 14
9.5.1. Visiting scientists 14
9.5.2. Seminar Talks, Invitations 15
9.5.3. Participation to conferences and summer school 15
9.5.4. Summer school lectures 16
9.5.5. Colloquium talks 16
10. Bibliography **17**

1. Team

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

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

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

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

Head of team

Christian Retoré [Professor, Université Bordeaux 1]

Administrative assistant

Brigitte Larue-Bourdon [TR, INRIA]

Research scientist (affiliation)

Christian Bassac [Lecturer, Université Bordeaux 3 – since september 2004]
Joan Busquets [Lecturer, Université Bordeaux 3]
Kim Gerdes [Lecturer, Université Bordeaux 3]
Patrick Henry [IR C.N.R.S. LaBRI]
Gérard Huet [DR INRIA, 50% project-team Cristal, 50% Signes]
Renaud Marlet [CR INRIA since december 2004]
Richard Moot [CR C.N.R.S., LaBRI]
Henri Portine [Professor, Université Bordeaux 3]

Research scientist (partner)

Agnès Bracke [Lecturer, Université Bordeaux 3]
Olivier De Langhe [Instructor, Institut national des Jeunes Sourds de Bordeaux Gradignan]
Alain Lecomte [Professor, Université Grenoble 2]
Yannick Le Nir [Lecturer, EISTI Pau]

Ph. D. students

Maxime Amblard [Ministry grant, Université Bordeaux 1]
Roberto Bonato [Italian grant for cotutored PhD, Università di Verona and Université Bordeaux 1]
Pierre Guitteny [Interpret/Official representative of *Direction Régionale des Affaires Sanitaires et Sociales* Aquitaine — Université Bordeaux 3]
Emilie Voisin [Aquitaine Regional grant, Université Bordeaux 3 – since september 2004]

2. Overall Objectives

Joint team with LaBRI (UMR 5800 C.N.R.S.) and the Department of Linguistics of University Michel de Montaigne (Bordeaux 3) — teams ERSS (UMR 5610 C.N.R.S.) and Jeune Equipe JE 2385 TELANCO).

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

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

by means of formal methods such as:

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

Two applications illustrate this approach:

- natural language tools for Sanskrit
- generation in French Sign Language

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

3. Scientific Foundations

3.1. The center: natural language syntax and semantics

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

Since the early days of computer science, natural language is both one of its favorite applicative field and the source of technical inspiration, as exemplified by the relation between formal language theory and linguistics. [36]

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

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

Statistical methods and corpus linguistics [59] have been quite successful for the last years, but there is a renewal of symbolic methods, and especially of logical ones, because of the advances in logic, the improvement of computer abilities for these rather slow algorithms, and overall the need for systems which handle the meaning of phrases, sentences, or discourses. [38]

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

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

- As said above, this enables applications that require access and computation of meaning.
- Up to now semantics only plays a minor role in Natural Language Processing although a linguistic viewpoint, the two sides of the linguistic signs its *signifiant* and *signifié* are a central subject ever since Saussure. The linking of the observable part of the sign or of the sentence and its meaning, is a constant question in linguistics both in Chomsky's Generative Grammar or in the Meaning-Text theory of Mel'cuk. [61][40]

- From a mathematical and algorithmic viewpoint, this interface is the place of some challenges: what is the link between two of the main frameworks, namely generative grammars and categorial grammars? The first ones are exemplified by Tree Adjoining Grammars TAGs [49] or Minimalist Grammars [70]. They enjoy efficient parsing algorithms and a broad covering of syntactic constructs. The second ones (see e.g. [63]) are less efficient but provide more acurate analyses. Indeed these latter systems are used for syntax as well as for logical or predicative semantics like Montague semantics [37][43] and thus allows generation algorithms. Other models, like dependency grammars, [60] provide a different account of the syntax/semantics interface. A comparison between the dependency model and a generative/logical one enables an assessment of the adequation of these families of models, and this is one of the main challenges of contemporary formal linguistics.

At one end of our spectrum stands morphology, and as often in generative grammar, we consider it as part of syntax. It should be nevertheless observed that the computational models involved in the processing of morphology are of different aspects : finite state automata, regular tranducers, etc. [52][53]

At the other end, on the semantical side, we do not consider ontological aspects of semantics, or lexical semantics, but rather extend the logical semantics to discourse and dialog. This is usually done by Discourse Representation Theory [51], which is topdown, incremental and involves state changes.

3.2. Word structure and automata: computational morphology

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

Participants: Gérard Huet, Kim Gerdes.

Computational models for phonology and morphology are a traditional application of finite state technology. [52][53][54][39] These models often combine symbolic or logical systems, like rewriting systems, and statistical methods like probabilistic automata which can be learnt from corpus by Hidden Markov Models. [59]

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

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

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

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

3.3. Sentence structure and formal grammars: syntax

Keywords: *categorial grammars*, *dependency grammars*, *formal grammars*, *tree adjoining grammars*.

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

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

The former grammars, rewrite systems, have excellent computational properties and quite a good descriptive adequacy. Relevant classes of grammars for natural language syntax, the so-called mildly context sensitive

languages, are just a bit beyond context-free languages, and they hare parsable in polynomial time as well. [50] Among theses classes of grammars let us mention Tree Adjoining Grammars, [48][49], Minimalist Grammars. [70][71][62] — Dependency Grammars share some properties with them but the general paradigm is quite different [61][41].

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

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

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

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

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

Indeed, a strong reason for using categorial grammars, [63] despite their poor computational properties, and poor linguistic coverage, is that they provide a correspondence bewteeen syntactic analyses and semantic representations. This is to be explained in the next section on the syntax/semantics interface.

In order to improve the computational properties of categorial grammars, and to extend their scope, one can try to connect them to more efficient and wider formalisms, like minimalist grammars. [56][55][69]

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

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

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

Why does there exists 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 extend 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. [43][73]

Nevertheless this computational correspondence between syntax and semantics provided by categorial grammars is very limited. Firstly, for the correspondence between syntactic and semantic types to hold, we have to provide words with syntactic types which are *ad hoc*, and even wrong. For instance, why should the type of a determiner depend of the constituent it is involved with? Secondly, the truth-conditional aspect of Montague semantics can be discussed both from a theoretical and from a practical viewpoint. According to cognitive sciences, and even to common sense, it is unlikely that human beings develop all possible interpretations when they process and understand a sentence, and in practice such a construction of all models is definitely untractable. [47] 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. [55][69]

3.5. Lexical semantics and derivational morphology

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

Participants: Christian Bassac, Patrick Henry, Renaud Marlet.

The generative lexicon [68] 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 informations which depict the sense of a word or morpheme are organized in three layers: the argument structure (related to logical semantics and syntax), the event structure, and the qualia structure.

The argument structure provides types (in the type-theoretical sense) to the arguments encoded in the qualia structure no matter whether they are syntactically mandatory or optional. The event structure follows [51]. It unfolds an event into several ordered sub-events with a mark on the most salient sub-event. Events are typed according to the typology of Vendler: state, process, transition, this later type including achievement and accomplishment. The qualia structure relates the argument structure and the event structure in rôles: formal, constitutive, telic, agentive.

These informations and their organization into the generative lexicons allows an explanation of, for instance, polysemy and of compositionality (in particular in compound words). This kind of model which relates knowledge representation to linguistic organization is especially useful for word sense disambiguation during (automatic) syntactic and semantic analysis.

3.6. Discourse and dialogue structure: computational semantics and pragmatics

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

Participants: Agnès Bracke, Joan Busquets, Gérard Huet, Alain Lecomte, Henri Portine.

Montague semantics has some limits. Two of them which, technically speaking, concern the context, can be overcome by using DRT, that is Discourse Representation Theory and its variants. [51][74] 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 compositionally of Montage 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. [74].

3.7. Type systems and functional programming for computational linguistics

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

Participants: Roberto Bonato, Gérard Huet, Yannick Le Nir, Richard Moot.

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

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

- Another example is Montague semantics which uses the Church description of higher-order logic, implemented in lambda calculus in order to have the compositionality principle of Frege.
- Finally, Discourse Representation Theory also is logic, in a different syntax, and can be combined with Montague semantics to obtain lambda-DRT.

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

- The Grail syntactic and semantic parser for Multi Modal Categorial grammars, defined and implemented by Richard Moot, is written in Prolog. This is the most developed and efficient software for categorial grammars, relying on recent development in linear logic, in particular proof nets. [64]
- 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. [67]
- Gérard Huet developped 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. [46]

4. Application Domains

4.1. Sanskrit philology

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

Participant: Gérard Huet.

Sanskrit literature is extremely rich, and is part of cultural patrimony. Nowadays, Internet can provide to both specialists and inquiring minds an access to it. For instance such a site exists for ancient Greek and Latin literature http://www.perseus.tufts.edu. This site provides an online access to the texts.

A simple click on each word analyses it, and brings back the lexcial item of the dictionnary, possible meanings, statistics on its use etc. The work of Gérard Huet described in the software section enables such computational tools for Sanskrit. It is presently extended to syntax in partnership with Pr Brendan Gillon (Mc Gill University, Montreal) The segmentation tools and the forthcoming inversion of flexional morphology will provide a computer aided for segmentation and tagging of continuous Sanskrit texts.

Looking for a treatment of syntax, a further step which already started is the construction of a tree bank of Sanskrit examples.

When the literature will be annotated this work will utltimatey lead to a Sanskrit analogous of Perseus.

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

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

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

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

Given the rarity of linguistic study on the syntax and semantics of sign languages — some exceptions concerning American Sign Language are [65][57][58] — before to be able to apply our methodology, our first task is to determine what the structure of the sentence is, using our personal competence as well as our relationship with the deaf community.

We intend to define methods and tools for generation of sign language sentences. It should be noted that there is a sequence of different representations of a sentence in Sign Language, from a grammatical description with agreement features and word/sign order that we are familiar with, to a notation system like Signwriting [72] or to a language for the synthesis of 3D images and movies. Our competences on the interface between syntax and semantics are well designed for a work in generation of the grammatical representations.

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

5. Software

5.1. The Zen toolkit

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

Participant: Gérard Huet [correspondant].

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

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

A number of uses of this platform outside of the Cristal team are under way. For instance, a lexicon of french flexed forms has been implemented by Nicolas Barth and Sylvain Pogodalla, in the Calligramme project-team at Loria.

The algorithmic principles of the Zen library, based on the linear contexts datastructure ('zippers') and on the sharing functor (associative memory server), were presented as an invited lecture at the symposium Practical Aspects of Declarative Languages (PADL), New Orleans, Jan. 2003 [45]. An extended version was written as a chapter of the book "Thirty Five Years of Automating Mathematics", edited in honor of N. de Bruijn [44].

5.2. DepLin

Keywords: *natural language syntactic analysis and generation*.

Participant: Kim Gerdes [correspondant].

This software based on dependency grammars was first designed for generation. It is now extended to analysis. The development of grammars for french, german, Modern Greek, Arabic is in different stages of progress.

5.3. Grail 3: natural language analysis with multimodal categorial grammar

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

Participants: Maxime Amblard, Richard Moot [correspondant], Christian Retoré.

Within the type-logical grammar paradigm Multi-Modal Categorial Grammars MMCG see e.g. [63] is one of the most complete system. They have been carefully implemented by Richard Moot into the biggest program for natural language analysis based on type logical grammars with lexicon/grammars for several languages: dutch, english, french, italian, hindi.

Originally written in Utrecht by Richard Moot in Sicstus Prolog with and interface in Tcl/Tk the third release has been completely rewritten by him in SWI Prolog (Gnu Public Licence). It also includes computational theoretical improvement in accordance with [64]: 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.

5.4. Experiments in categorial grammars

Keywords: *grammatical inference*, *parsing*.

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

This software, *CGTools* is an academic prototype. It is the combination of two *Travaux d'Etude et de Recherche* of 4^{th} year students: Véronique Moriceau et Jérôme Pasquier (Université de Nantes, 2002) which has been reorganized and extended by Thomas Poussevin, Jean-François Deverge, Fahd Haiti, Anthony Herbé (Université Bordeaux 1, 2003). [67] It is written in OCaML, with an interface written in Tcl/Tk and the input and output format are XML files (DAGs for representing analyses, proofs and trees).

Presently, the following algorithms are implemented:

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

6. New Results

6.1. Flexional morphology

Gérard Huet continued his work on developing a computational linguistics platform adapted to Sanskrit, based on applicative programming in Ocaml. On the morphology front, he implemented the verbal conjugation paradigms, a rather complex task, since Sanskrit verbal forms admit 3 persons, 3 numbers (singular, dual and plural), 3 voices (active, middle and passive), 3 moods (indicative, imperative, optative) a present system with 10 families for tenses present and imperfect, a future system, a perfect system with a complex reduplicating scheme, an aorist system with 7 paradigm classes. Furthermore many roots admit derived stems for causative, intensive, and desiderative moods. This conjugation machine was instantiated on his digital lexicon, leading to 75000 forms for the 515 roots. These forms can be compressed into a 54KB lexical dag, which yields in its turn a segmenting automaton of 120KB. This supplements the 144000 substantive declension forms (167KB dag, 921KB automaton). The segmenter/ tagger he previously developed for substantive compounds is now able to segment small sentences, with a 3-pass method using 3 automata (substantives, roots, and preverbs).

This work basically concludes the morphological work. The Sanskrit processing site has been enriched with the 3-automaton segmenter/tagger, and a lexicon-directed lemmatizer. Full lists of inflected forms, in XML format, were released as free linguistic resources available for research purposes.

A structural classification of compounds was effected by machine, in order to complete the segmenting automaton with compound forms which cannot be generated by standard euphony (external sandhi) from their components. This is a typical application of using programs to generate a systematic linguistic model from a corpus (here the lexicon). Such analysis was previously unavailable for the Sanskrit language, because of its complex morpho-phonetics structure.

6.2. Robust syntax

Kim Gerdes has designed, within the dependency grammar paradigm, a notion of tree with syntactic places which allows a neat treatment of relatively free word order with a single underspecified analysis for different variants with respect to allowed word permutation. After proving their relevance for German Kim Gerdes applied it to Korean with Hi-Yon Yoo [24] and to French verbal cluster with Sylvain Kahane [23].

After handling the difficult morphology of Sanskrit Gérard Huet is now addressing its syntax in collaboration with Brendan Gillon, a linguist from Mc Gill University. Brendan Gillon manually constructed over the years a corpus of Sanskrit sentences from a classical treatise on Sanskrit syntax by Apte. From its exemple citations, Pr Gillon built an annotated phrase structure tree bank of over 500 sentences. G. Huet managed to parse the notation and reverse engineer this database into a fully structured tree bank. Work is under progress to adapt G. Huet segmenter so that each phrase structure item may be considered an enrichment of the consistent tagged representation. The next step will be to use this annotated corpus as training data for a parser, whose derivation trees will be consistent with the syntactic annotations. Pr Gillon visited our site of Rocquencourt for a week in June in order to jointly work out a more explicit syntactic notation, where long distance dependencies and anaphora antecedent links will be explicity represented in a notation for a forest of dislocated syntax trees. The various tree operations will correspond to elementary moves of a parser, which will enforce constraints issued from agreement and subcategorization conditions. This constraint machine is still in its design phase.

The design of the lexical database and its Web interface was presented in Geneva in August at the Workshop on Enhancing and Using Electronic Dictionaries, as part of the International Conference on Computational Linguistics (COLING 2004) [25].

6.3. Syntax semantics interface for generative grammars

The coding of Lecomte and Retoré (1999) [56][55] of minimalist grammars into categorial grammars that we call categorial minimalist grammars is used Maxime Amblard, Alain Lecomte and Christian Retoré have improved the computation of semantic representations from these analyses. One way developped by Amblard Lecomte and Retoré is to sue lambda terms with contexts in order to synchronize the syntactic and the semantic part [17] and another way explored by Lecomte is to encode movement within the types themselves [26].

Roberto Bonato has defined an incremental algorithm for computing the binding relationship bewteen words and especially when the bound term is a pronoun: indeed, generative syntax has shown that the binding relation is responsible for the possible or impossible coreference of a pronoun with its antecedent. Nevertheless, up to now there was no incremental computing of this relation, which was defined as a set of constraints on a complete analysis. As this relation filters impossible interpretations, it is much more efficient to compute it while analysing the sentence. [22]

A question in the same style, studied by Joan Busquets, is to study subtle differences between different kind of verbal ellipsis. For instance why can we say *I am happy. She is too.* and not *Je suis content. * Elle est aussi.* ? A distinction between VP ellipsis and IP ellipsis (known as stripping), studied through Catalan which stresses this difference, allows us to determine what can be erased and when, e.g. for generation purposes. This study also shows that depending on the focus two ways to interpret a sentence. [12]

6.4. Lexical semantics

In his contribution to the analysis of English nominal compounds, Christian Bassac challenged the claim (stated for instance in works like the one of Downing [42]) that the meaning of these compounds cannot be

predicted. He shows in [18] how a Generative Lexicon (GL) can be used to characterize the meaning of NN compounds compositionally. The main idea developed here is that the modifier of a NN compound saturates an argument of a predicate encoded in a particular role of the qualia structure of the head, (agentive in *fruit juice*, constitutive in *rubber bullet*, telic in *walking stick*, or as an adjunct of the formal in prepositional compounds such as *road casualties*), thus allowing a compositional treatment .

In the work of Christian Bassac on Turkish morphology, [19] the same theoretical framework of a G.L allows a descriptive account of the concurrence of two deverbal morphemes *Iş* and *mE* and also an explanatory account of various syntactic properties connected to the presence of these morphemes. It is showed here that the former morpheme is ruled out in irrealis contexts because of type selection properties, and that various gaps in the productivity of each morpheme as well as the grammaticality of verbal ellipsis can be accounted for by the quantificational force given to each morpheme by the role it is encoded in.

In [20] Christian Bassac introduces the hypothesis of the presence of a copula variously conditioned and sometimes realised as an empty element. It allows a unified description of nominal and verbal predication, and accounts for phonological empiric phenomena too. The consequence of the hypothesis put forward here on a theory of morphology is analysed and the Pollock's claim [66] that each grammatical morpheme has its own functional projection is challenged.

6.5. Modeling French sign language (LSF) grammar

Olivier De Langhe, Pierre Guitteny, Henri Portine and Christian Retoré have been pursuing their quest of the neutral word order in simple sentences expressed in french sign language (LSF). They observed the surprising fact that *Object Subject Verb* is a rather frequent word order in LSF, especially when the *Objet* is unanimated and the *Subject* is human. This conflicts with existing linguistic theories which claim that *OSV* does not exists, or, rather is the result of a topicalization which emphasizes the *Object*. Nevertheless in their observations, in some sentences there is no other possibility, or at least the *OSV* order is strongly preferred. Their observation and analysis of basic sentences in french sign language has been published in [16]

7. Contracts and Grants with Industry

7.1. PicoPeta

Gérard Huet ported his Sanskrit processing workbench as an application for the Simputer, a hand-held computing device running Linux developed in India. In December he visited the PicoPeta corporation in Bangalore, one of the manufacturers of the Simputer, in order to initiate a possible technology transfer towards a pocket Sanskrit machine.

8. Other Grants and Activities

8.1. Regional research programs

The region Aquitaine is funding (together with INRIA and LABRI-CNRS) a project on sign language processing. Our team will thus possess the video recorder, the software and computers for constituting a very good quality corpus of spontaneous sign language speech. Contact: Christian Retoré Ê

8.2. National research programs

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

Signes is one of the fifteen research team of the Groupe de Recherches 2521 (C.N.R.S.) directed by Francis Corblin (Université Paris IV). This research program is divided into *Opérations*: *Modèles et formats de représentation pour la sémantique*, *Les Modèles à l'épreuve des données*, *Sémantique et corpus*, *Les*

interfaces de la sémantique linguistique, *Sémantique computationnelle*. The *Signes* team is part of the later two operations, which could be translated as *Interfaces of linguistic semantics* and *Computational semantics*.

8.2.2. Programme Interdiciplinaire du C.N.R.S. Traitement des Connaissances, Apprentissage et Nouvelles Technologies de l'Information et de la Communication

Alain Lecomte is supervising a project VALI (*Vers des assistants lecteurs intelligents*) in this setting. it is intended to develop tools to help the new researcher to grasp the contents of research article. To do so, the contents can be organized using linguistic theories like SDRT and logical tools like the proof assistant Coq can be applied to deduce relationship between parts of contents.

8.3. European research programs

8.3.1. CoLogNet: European network of Excellence on Computational logic

The team *Signes* is an active node of this network and, in particular of the section 6 of this network: computational logic for natural language processing, headed by Michael Moortgat. The contact person is Gérard Huet.

8.3.2. UIL-OTS Utrecht — Signes (Action intégrée van Gogh)

A research program entitled *Generative grammar and deductive systems for the processing of natural language syntax and semantics* has been approved for 2004 and renewed for 2005. The other team in this bilateral research program is *Computational linguistics and logic* directed by Michael Moortgat at *Utrecht Institue of Linguistics*. The dutch contact is Willemijn Vermaat, and the french one is Christian Retoré.

9. Dissemination

9.1. Activism within the scientific community

9.1.1. Honours

- Gérard Huet is member of the *Académie des sciences* since November 2002.
- Gérard Huet received in April 2004 an Honorary Doctorate in Technology from Chalmers University in Göteborg.

9.1.2. Editorial boards

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

9.1.3. Program committees of conferences

- Joan Busquets was on the program committee of the 8th annual meeting of the Texas Linguistic Society, Issues at the semantic-pragmatics interface, March 5-7 2004.
- Joan Busquets was on the program committee of the Journées Romanes, ERSS, Toulouse December 2004.
- Gérard Huet was on the program committee of the 11th Workshop on Logic Language Information and Computation 2004 (Paris).
- Kim Gerdes was co-chair of the program committee of the *Journées de la Syntaxe*, held in Bordeaux in November 2004.
- Christian Retoré was on the program committee of the LREC Workshop on the Representation and Processing of Sign Language 2004 (Lisbon)
- Christian Retoré was on the program committee of the Workshop Categorial Grammars 2004 (Montpellier).

9.1.4. Academic committees

- Christian Bassac is a member of the hiring committee in linguistics of Université Bordeaux 3.
- Joan Busquets is a member of the hiring committees in linguistics of Université Toulouse 2 and Université Bordeaux 3.
- Gérard Huet is a nominated scientific personnality of the board of governors of the Université Paris 7.
- Henri Portine is a member of the hiring committees in linguistics of Université Paris 3 and Université Bordeaux 3.
- Henri Portine is an elected member of the board of governors of the Université Bordeaux 3 and of Institut Universitaire de Formation des Maîtres d'Aquitaine.
- Henri Portine is the head of the linguistic department of Université Bordeaux 3.
- Henri Portine is the head of the research team *Text, Language, Cognition* JE2385.
- Christian Retoré is a member of the hiring committee in computer-science of Université Bordeaux 1.
- Christian Retoré is a member of the committee of the faculty of mathematics and computer science of the Université Bordeaux 1.

9.1.5. Organization of events

- Kim Gerdes organized with Claude Muller the *Journées de la syntaxe* in Bordeaux in November 2004.
- Kim Gerdes, Xavier Lavry, Maxime Amblard and Patrick Henry organized the weekly seminar *Linguistique et informatique* Universités Bordeaux 1 et 3.
- Patrick Henry organized a meeting with the Calligramme team.
- Alain Lecomte organized the TCAN-VALI meeting on *La contruction du savoir scientifique dans la langue*, grenoble, October 2004.

9.2. Teaching

Since all its members are university staff, *Signes* is intensively implied in teaching, both in the computer science cursus of the science University (Bordeaux I) and in the linguistic cursus of the letter University (Bordeaux III). Let us cite the lectures whose topic is computational linguistics:

- *Natural language processing*, Bordeaux, PhD students in computer science (Christian Retoré, Kim Gerdes, Joan Busquets)
- *The main theorems of lambda calculus*, Parisian Master of Research in Computer Science 5^{th} year in computer science (Gérard Huet)
- *Deductive systems*, Bordeaux 1, 5^{th} year in computer science (Richard Moot, Christian Retoré)
- *Symbolic natural language processing*, Bordeaux 1, 5^{th} year in computer science (Richard Moot, Christian Retoré)
- *Utterance acts and semantics*, Bordeaux 3, 5^{th} year in linguistics (Henri Portine)
- *The syntax of Wh-clauses and extraction*, Bordeaux 3, 5^{th} year in linguistics (Christian Bassac)
- *Finite state natural language processing*, Bordeaux 1, 4^{th} year in computer science (Christian Retoré)
- *The principle of charity: Quine and Davidson*, Bordeaux 3, 4^{th} year in linguistics (Joan Busquets)
- *Pragmatics*, Bordeaux 3, 4^{th} year in linguistics (Joan Busquets)
- *Word order and its formalization*, Bordeaux 3, 4^{th} year in linguistics (Kim Gerdes)

9.3. Thesis Juries

- Gérard Huet was reviewer of the habilitation of Solange Coupet (*Des preuves et des programmes* Université de Provence, 10-09-04)
- Gérard Huet was reviewer of the habilitation of the habilitation of Claire Gardent (*Inférence et traitement automatique des langues* Université de Nancy 1, 21- 09-04).
- Alain Lecomte was on the jury of the PhD thesis of Francis Brunet-Manquat (*Analyse robuste dans les grammaires de dépendance*, Université Joseph Fourier Grenoble 1 21-12-04).
- Christian Retoré was reviewer of the PhD thesis of Daniela Dudau (*Apprentissage de grammaires catégorielles pour simuler l'acquisition du langage naturel à l'aide d'informations sémantiques*, Université Lille 1, 12-04-04)
- Christian Retoré was reviewer of the PhD thesis of Fanch Lejeune (*Analyse sémantico-cognitive d'énoncés en Langue des Signes Française pour une génération automatique de séquences gestuelle* Université Paris-Sud, 23-11-04)
- Christian Retoré was on the jury of the habilitation of Claire Gardent (*Inférence et traitement automatique des langues* Université de Nancy 1, 21- 09-04).

9.4. Academic supervision

9.4.1. Student intern supervision – fifth year

- Alain Lecomte (with Jean Caelen) supervised the master thesis of Anne Xuereb, *SDRT pour le dialogue guidé par les tâches*.
- Henri Portine supervised the master thesis of Emilie Voisin *Modélisation et Analyse des dialogues de théâtre: la problématique de la révision des croyances*.
- Christian Retoré and Joan Busquets supervised the master thesis of Emmanuel Daviaud *Logique d'ordre supérieur avec égalité dans un cadre compositionnel et intensionnel*.
- Christian retoré supervised the master thesis of Bart Georges *Réseaux de Petri, logique linéaire et grammaires formelles*.

9.4.2. PhD supervision

- Joan Busquets and Andrée Borillo (Université Toulouse 2) surpervised the thesis work of Laurent Prévot *Structures sématiques et pragmatiques pour la modélisation de la cohérence dans des dialogues finalisés.* (Université Toulouse 2) defended in 2004.
- Alain Lecomte is supervising the thesis work of Tran Vu Truc *Logique d'informations partielles pour le traitement des implicites*. (Université Grenoble II)
- Alain Lecomte and Christian Retoré are co-supervising the thesis work of Maxime Amblard, *Calcul de représentations sémantiques dans les grammaires minimalistes*. (Université Bordeaux 1)
- Henri Portine and Renaud Marlet are supervising the thesis work of Emilie Voisin, *Génération automatique d'énoncés en Langue des Signes Française*. (Université Bordeaux 3)
- Henri Portine is supervising the thesis work of Pierre Guitteny, *Le passif en Langue des Signes Française*. (Université Bordeaux 3)
- Christian Retoré and Alexandre Dikovsky (Université de Nantes) are co-supervising the thesis work of Erwan Moreau, *Acquisition de grammaires catégorielles et de grammaires de dépendances*. (Université de Nantes)
- Christian Retoré and Denis Delfitto (Università di Verona) are co-supervising the thesis work of Roberto Bonato, *Algorithmes de calcul de représentations sémantiques à partir d'analyses de type générativiste et algorithmes inverses*. (cotutored PhD Université Bordeaux 1 / Università di Verona)

9.5. Participation to colloquia, seminars, invitations

9.5.1. Visiting scientists

- Denis Delfitto (Università di Verona) visited *Signes* for a week in June 2004 and gave two talks, one on *Bound variable interpretation* and one on *Propositional anaphora*.
- Sacha Dikovsky (LINA, Nantes) gave a talk at the *Signes* seminar on *Categorial dependency grammars*
- Claude Muller (ERSS) gave a talk at the *Signes* seminar on *L'inversion du sujet en français contemporain.*
- Brendan Gillon (Mc Gill University, Montreal) visited Rocquencourt a week in June 2004 (work on Sanskrit with Gérard Huet)
- Willemijn Vermaat (OTS Utrecht) visited *Signes* two weeks in June 2004. PAI van Gogh *Signes*/OTS.

9.5.2. Seminar Talks, Invitations

- Joan Busquets gave a talk *Stripping and Ellipsis: What is deleted and When?* at the Universitat Pompeu Fabra (Barcelona) in May 2004
- Kim Gerdes gave two talks on *Dependency grammars and topological analyses* at the Yonsei University of Seoul in August 2004.
- In February, G. Huet was invited speaker at the Sophia-Antipolis Colloquium. Gérard Huet gave a talk on *La théorie de la fonctionnalité à la croisée des chemins entre Informatique, Logique et Linguistique.*
- In March, Gérard Huet gave a seminar at the LIMSI CNRS lab in Orsay, where he gave a talk *De Zen à Aum.*
- In May, Gérard Huet gave a lecture at Chalmers University, in Göteborg, as part of a workshop in his honor, at the occasion of his reception of a Honoris Causa Doctorate of Technology. He spoke on *Computational Linguistics from Zen to AuM.*
- In August too Gérard Huet gave a lecture on *Functionality theory applied to Informatics, Logic and Linguistics* at the session on *New mathematics for new challenges* at the EuroScience Forum 2004 in Stockholm.
- In October Gérard Huet gave a talk *Transducteurs d'états finis applicatifs et traitement des langues naturelles* at the Colloquium CMAT-CMAP-STIX, Centre de mathématiques, Ecole Polytechnique, Palaiseau.
- In November Gérard Huet gave a talk *Des Lambdas et des Aums*, at the Conférence en l'honneur du Pr Martin-Löf, in Marseille Luminy.
- In June Christian Retoré gave a talk on *Hidden Markov Models* at the biology department in Université Bordeaux 2
- In November Christian Retoré gave a talk *Categorial minimalist grammars and semantics* at the TALANA seminar, Université Paris 7.

9.5.3. Participation to conferences and summer school

- Maxime Amblard, Patrick Henry, Henri Portine, attended the workshop TCAN-VALI organized by Alain Lecomte, Grenoble, October 2004.
- Maxime Amblard, Roberto Bonato, Gérard Huet, Alain Lecomte, Renaud Marlet attended the *European Summer School in Logic, Language and Information*, Nancy, August 2004.
- Agnès Bracke attended the *Ecole d'été de linguistique de Corpus : Constitution, Archivage, Evaluation*, Caen, June 2004.
- Pierre Guitteny attended the workshop *Representation and Processing of Sign Languages*, LREC-workshop, Lisboa, May 2004.
- Gérard Huet attended the ACL 2004 conference in Barcelona.
- Renaud Marlet attended the EALing autumn school in linguistics, ENS, Paris.
- Emilie Voisin attended the conference *Verbal and Signed Languages*, Università di Roma tre, September 2004.

9.5.4. Summer school lectures

- Richard Moot with Pierre Castéran (INRIA, LORIA, Nancy) gave a fifteen hour lecture at the *European Summer School in Logic, Language and Information*, Nancy, August 2004.
- In August 2004 Gérard Huet gave one of the Evening Lecture at the ESSLLI 2004 summer school in Nancy, on the topic *Splitting the Entropy Gordian Knot.*

9.5.5. Colloquium talks

- Maxime Amblard gave a talk at the workshop *Journées Sémantique et Modéélisation* on *Interface syntaxe/sémantique pour les grammaires minimalistes catégorielles*, ENS-Lyon, March 2004. (joint work with A. Lecomte and C. Retoré) [17]
- Roberto Bonato presented his work on *Towards inductive semantics for coreference* at the NASSLLI Student Session, Los-Angeles, June 2004.
- Roberto Bonato presented his work with Francesco Bellomi on *Lexical Authorities in an Encyclopedic Corpus: a case study on Wikipedia* at the *Word Structure and Lexical Systems: models and applications* conference in Paiva, December 2004.
- Kim Gerdes presented with Hi-Yon Yoo their work on *A dependency approach to Korean word order*, at the 2004 Linguistic Society of Korea International Conference, Seoul August 2004.
- Gérard Huet gave a talk on *Design of a Lexical Database for Sanskrit* at the Workshop on Enhancing and Using Electronic Dictionaries, of the Coling 2004 conference in Geneva.[25]
- Alain Lecomte gave an invited lecture on *Minimalist derivations as proofs: the control of hypotheses in Lambek-style deductions* at CG 2004, Montpellier. [26]
- Yannick Le Nir gave a talk on *From AB grammars to NL* at CG 2004, Montpellier. [27]
- Richard Moot gave a talk on *Graph Algorithms for Improving Type-Logical Proof Search* at CG 2004, Montpellier. [28]
- Richard Moot and Christian Retoré gave a talk on *L'ordre des mots dans les grammaires catégorielles* aux *Journées de la syntaxe* Bordeaux, November 2004. [29]
- Christian Retoré gave an invited lecture (with Sylvain Pogodalla) on *Handsome Non-Commutative Proof-Nets: perfect matchings, series-parallel orders and Hamiltonian circuits* at CG 2004, Montpellier. [34]

10. Bibliography

Major publications by the team in recent years

[1] N. ASHER, D. HARDT, J. BUSQUETS. *Discourse Parallelism, Ellipsis, and Ambiguity*, in "Journal of Semantics", vol. 18, n° 1, 2001, p. 1–25.

[2] A. FORET, Y. LE NIR. *Lambek rigid grammars are not learnable from strings*, in "COLING'2002, 19th International Conference on Computational Linguistics, Taipei, Taiwan", vol. 1, August 2002, p. 274–279.

[3] K. GERDES. *Topologie et grammaires formelles de l'allemand*, Thèse de Doctorat, Université Paris 7, 2002.

[4] 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.

[5] A. LECOMTE. *Rebuilding the Minimalist Program on a logical ground*, in "Journal of Research on Language and Computation", n° 1, 2004, p. 27–55.

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

[7] 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.

[8] H. PORTINE, A. ROUSSEAU. *Tesnière et la syntaxe structurale*, in "Modèles linguistiques", vol. 23, n° 2, 2002, p. 99-121.

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

Books and Monographs

[10] C. RETORÉ, E. STABLER (editors). *Special Issue on Resource Logics and Minimalist Grammars*, Journal of Research on Language and Computation, vol. 2(1), Kluwer, 2004.

[11] C. BASSAC. *Principes de morphologie anglaise*, Linguistica, Presses Universitaires de Bordeaux, 2004.

Articles in referred journals and book chapters

[12] J. BUSQUETS. *La position du focus et la distinction stripping/ellipse du groupe verbal en catalan*, in "Cahiers de grammaire", vol. 29, 2004.

[13] P. FLAJOLET, G. HUET. *Mathématiques et Informatique*, in "Les mathématiques dans le monde contemporain", J.-C. YOCCOZ (editor)., chap. 2, Lavoisier, 2004.

[14] H. PORTINE. *Les modalités: entre logique et langue*, in "Revue belge de philologie et d'histoire", vol. 81, n° 3, 2003.

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

[16] O. DE LANGHE, P. GUITTENY, H. PORTINE, C. RETORÉ. *A propos des structures OSV en Langue des Signes Française*, in "Silexicales", vol. 4, 2004, p. 115–130.

Publications in Conferences and Workshops

[17] M. AMBLARD, A. LECOMTE, C. RETORÉ. *L'interface entre syntaxe et sémantique pour les grammaires minimalistes catégorielles*, in "Journées Scientifiques Sémantique et Modélisation, ENS-LSH Lyon", 2004.

[18] C. BASSAC. *Composionality and Nominal Compounds*, in "Journées de linguistique anglaise à Toulouse", 2004.

[19] C. BASSAC, M. CIÇEK. *Les noms déverbaux dérivés par affixation de -(y)Iş et mE*, in "Noms déverbaux, Lille", 2004.

[20] C. BASSAC, M. CIÇEK. *Morphologie de la prédication verbale et non verbale en turc*, in "Colloque sur la prédication", 2004.

[21] F. BELLOMI, R. BONATO. *Lexical authorities in an encyclopedic corpus: a case study on wikipedia*, in "Word structure and lexical systems: models and applications, Pavia", 2004.

[22] R. BONATO. *Towards inductive semantics for coreference*, in "NASSLLI Student Session", 2004.

[23] K. GERDES, S. KAHANE. *L'amas verbal au coeur d'une modélisation topologique du français*, in "Pré-Actes des Journées de la syntaxe: ordre des mots dans la phrase française, positions et topologie", K. GERDES, C. MULLER (editors)., E.R.S.S.-C.N.R.S., 2004.

[24] K. GERDES, H.-Y. YOO. *A dependency approach to Korean word order*, in "2004 Linguistic Society of Korea International Conference", 2004.

[25] G. HUET. *Design of a Lexical Database for Sanskrit*, in "Workshop on Enhancing and Using Electronic Dictionaries, COLING 2004", International Conference on Computational Linguistics, 2004, http://pauillac.inria.fr/~huet/PUBLIC/coling.pdf.

[26] A. LECOMTE. *Minimalist derivations as proofs: the control of hypotheses in Lambek-style deductions*, in "Categorial grammars – an efficient tool for natural language processing, Montpellier", C.N.R.S., June 2004, p. 288-289.

[27] R. MOOT. *From NL to AB grammars*, in "Categorial grammars – an efficient tool for natural language processing, Montpellier", C.N.R.S., June 2004, p. 290–304.

[28] R. MOOT. *Graph Algorithms for Improving Type-Logical Proof Search*, in "Categorial grammars – an efficient tool for natural language processing, Montpellier", C.N.R.S., June 2004, p. 13–28.

[29] R. MOOT, C. RETORÉ. *L'ordre des mots dans les grammaires catégorielles*, in "Pré-Actes des Journées de la syntaxe: ordre des mots dans la phrase française, positions et topologie", K. GERDES, C. MULLER (editors)., E.R.S.S.-C.N.R.S., 2004, p. 56–59.

[30] C. RETORÉ. *A description of the non-sequential execution of Petri nets in partially commutative linear logic*, in "Logic Colloquium 99", J. VAN EIJCK, V. VAN OOSTROM, A. VISSER (editors)., Lecture Notes in Logic, ASL and A. K. Peters, 2004, p. 152–181.

Internal Reports

[31] M. AMBLARD. *Représentation sémantique pour les grammaires minimalistes*, Technical report, n° 5360, INRIA, 2004, http://www.inria.fr/rrrt/rr-5360.html.

[32] A. BRACKE. *Construction des concepts scientifiques par des enfants de 9 à 14 ans: structuration cognitive et langagière*, Technical report, Université Bordeaux 3, 2004.

[33] J. BUSQUETS. *A propos de fer-ho ("le faire") anaphorique en catalan*, Technical report, ERSS Université Toulouse Le Mirail, 2004.

[34] S. POGODALLA, C. RETORÉ. *Handsome Non-Commutative Proof-Nets: perfect matchings, series-parallel orders and Hamiltonian circuits*, Technical report, n° 5409, INRIA, 2004, http://www.inria.fr/rrrt/rr-5409.html.

[35] C. RETORÉ. *Syntaxe et Traitement Automatique des Langues*, Technical report, n° 5459, INRIA, 2004, http://www.inria.fr/rrrt/rr-5459.html.

Bibliography in notes

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

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

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

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

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

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

[42] P. DOWNING. *On the Creation and Use of English Compound Nouns*, in "Language", vol. 53, n° 4, 1977, p. 810–842.

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

[44] 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.

[45] G. HUET. *Zen and the Art of Symbolic Computing: Light and Fast Applicative Algorithms for Computational Linguistics*, in "Practical Aspects of Declarative Languages (PADL) symposium, New Orleans", Invited lecture, 2003, http://pauillac.inria.fr/~huet/PUBLIC/padl.pdf.

[46] 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.

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

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

[49] 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.

[50] 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.

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

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

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

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

[55] 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, http://www.labri.fr/Recherche/LLA/signes/.

[56] A. LECOMTE, C. RETORÉ. *Towards a Minimal Logic for Minimalist Grammars: a Transformational Use of Lambek Calculus*, in "Formal Grammar, FG'99", FoLLI, 1999, p. 83–92.

[57] 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.

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

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

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

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

[62] J. MICHAELIS. *Derivational minimalism is mildly context sensitive*, in "Logical Aspects of Computational Linguistics, LACL'98, selected papers", M. MOORTGAT (editor)., LNCS/LNAI, n° 2014, Springer-Verlag, 2001, p. 179–198.

[63] 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.

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

[65] 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.

[66] J.-Y. POLLOCK. *Verb movement, Universal grammar and the structure of IP*, in "Linguistic Inquiry", 1989.

[67] 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.

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

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

[70] E. STABLER. *Derivational Minimalism*, in "Logical Aspects of Computational Linguistics, LACL'96", C. RETORÉ (editor)., LNCS/LNAI, vol. 1328, Springer-Verlag, 1997, p. 68–95.

[71] 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.

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

[73] 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.

[74] 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.