INRIA

Project-Team TALARIS

Traitement Automatique des Langues: Représentations, Inférences et Sémantique

Nancy - Grand Est

THEME SYM

Activity Report

2007

Table of contents

1. Team

TALARIS is an INRIA project-team (UMR 7503) common to INRIA, the CNRS, the University of Nancy 1 (Henri Poincaré), the University of Nancy 2, and the National Polytechnic Institute of Lorraine. For more details, we invite the reader to consult the team web site at http://talaris.loria.fr/.

Head of project-team

Patrick Blackburn [Research Director (DR2), INRIA, HdR]

Vice-head of project team

Claire Gardent [Research Director (DR2), SHS Department CNRS, HdR]

Administrative assistant

Isabelle Blanchard

Research scientist

Carlos Areces [Research Scientist (CR1), INRIA]

Matthieu Quignard [Research Scientist (CR2), SHS Department CNRS]

Jesse Tseng [Research Scientist (CRS), SHS Department CNRS (left TALARIS 2007-09-01)]

Faculty member

Lotfi Bellalem [PRAG, ESIAL, Univerity of Nancy 1 (UHP)]

Nadia Bellalem [Assistant Professor, IUT Nancy-Charlemagne, University of Nancy 2]

Daniel Coulon [Professor, UFR MI, University of Nancy 2]

Samuel Cruz-Lara [Assistant Professor, IUT Nancy-Charlemagne, University of Nancy 2]

Christine Fay-Varnier [Assistant Professor, School of Geology, INPL]

Azim Roussanaly [Assistant Professor, UFR MI, University of Nancy 2]

Fabienne Venant [Assistant Professor, IUT Nancy-Charlemagne, University of Nancy 2]

Post-doctoral fellow

Sylvain Schmitz [2007-08-01 until 2008-08-01]

Ph. D. student

Marilisa Amoia [UHP Nancy 1, Co-tutelle University of the Saarland, from 2006-10-15 until 2008-10-15]

Paul Bedaride [Ministry grant, UHP Nancy 1, from 2006-10-15 until 2009-10-15]

Luciana Benotti [CORDIS grant, UHP Nancy 1, from 2006-06-10 until 2009-12-31]

Alexandre Denis [INRIA grant, UHP Nancy 1, from 2006-10-15 until 2009-10-15]

Daniel Gorin [UHP Nancy 1, Co-tutelle Universidad de Buenos Aires]

Sébastien Hinderer [Ministry grant, UHP Nancy 1, from 2004-10-30 until 2008-10-30]

Eric Kow [UHP Nancy 1, INRIA grant, finished 2007-14-11]

Phuong Le Hong [University of Nancy 2, Co-tutelle University of Hanoi, from 2006-10-01 until 2009-10-01]

Cheng Yu Lu [INRIA (with Taiwan), UHP Nancy 1, from 2005-07-19]

Chia-Hung Lin [INRIA (with Taiwan), UHP Nancy 1, from 2005-07-19]

Sergio Mera [UHP Nancy 1, Co-tutelle Universidad de Buenos Aires]

Yannick Parmentier [UFR STMIA UHP Nancy 1 (INRIA grant - UHP Nancy 1 Monitor), finished 2007-04-06]

Joseph Roumier [ENSEM INPL (CIFRE grant financed par CEGELEC) until 2008-10-01]

Dmitry Sustretov [UHP Nancy 1, INRIA grant from 2005-10-01 until 2008-10-01]

Technical staff

Emilie Colin [until 2007-12-31]

Ingrid Falk [until 2007-12-31]

Karen Fort [until 2008-9-30]

2. Overall Objectives

2.1. Background

TALARIS stands for *Traitement Automatique des Langues: Representation, Inference, et Semantique*. As this name suggests, the aim of the TALARIS team is to investigate semantic phenomena (broadly construed) in natural language from a computational perspective. More concretely, TALARIS's goal is to develop grammars (with a special emphasis on French) with a semantic dimension, to explore the linguistic and computational issues involved in such areas as natural language generation, textual entailment recognition, discourse and dialogue modeling, pragmatics, and multilinguality, and to investigate the interplay between representation and inference in computational semantics for natural language.

2.2. Organization

The work of the TALARIS team can be subdivided into four overlapping and mutually supporting categories.

Computational Semantics. These theme is devoted to the theoretical and computational issues involved in building semantic representations for natural language. Special emphasis is placed on developing large scale semantic coverage for the French Language.

Discourse, Dialogue and Pragmatics. This theme is devoted to developing theoretical and computational models of discourse and dialogue precessing, and investigating the inferential impact of pragmatic factors (that is, the factors affecting how humans being actually use language).

Logics for Natural Language and Knowledge Representation. The theme is devoted to theoretical and computational tools for working with logics suitable for natural language inference and knowledge representation. Special emphasis is place on hybrid logic, higher order logic, and discourse representation theory (DRT).

Multilinguality for Multimedia. This theme is devoted to creating generic ISO-based mechanisms for representing and dealing with multilingual textual information. The center of this activity is the MLIF (Multi Lingual Information Framework) specification platform for elementary multilingual units.

2.3. Overall Objectives

The major long term computational goals of the TALARIS team are:

- The creation of a large scale computational semantics framework for French that supports deep semantic analysis and surface realisation (the production of sentences from meaning representations),
- The creation of dialogue systems (in particular, for French) that support flexible and realistic interaction with the user.
- The creation of efficient inference systems for logics that are capable of representing natural language content and the background knowledge required to support reasoning.

These computational goals will be pursued in the context of theoretical investigations that will rigorously map out the required scientific and mathematical context.

2.4. Highlights

The main highlight of 2007 is simply this: it is the first year of TALARIS's existence! Like many other INRIA project teams, TALARIS was born out of a previously existing project — but the past year has been marked by heavy personnel changes. The result is a smaller team that is tightly focused on the goals just listed.

But our birthday year was marked by another highlight — the successful inauguration of a brand new *Erasmus Mundus* Masters degree in *Language and Communication Technology* (for detailed information, see http://lct-master.org). This degree, which is taught as one of the tracks of the Masters in Cognitive Science at the University of Nancy 2, is designed to attract to Europe the very best international students in computational linguistics. Students must spend one year at each of two partner universities in two different European countries (in addition to the University of Nancy 2, they can choose between the University of Bolzano in Italy, Charles University in the Czech Republic, the University of Groningen in the Netherlands, the University of the Saarland in Germany, and the University of Malta in Malta).

Needless to say, this Masters degree is not just relevant to TALARIS — indeed it is interesting precisely because it is relevant to the entire Nancy computational linguistics and language engineering community (which includes such LORIA based teams as CALLIGRAMME, ORPAILLEUR, PAROLE and READ and the Lexique team at ATILF). Nonetheless, TALARIS members (notably Carlos Areces, Patrick Blackburn and Claire Gardent) were heavily involved in setting up the program — and so its successful start in 2007 is special cause for celebration in the team. Seven *Erasmus Mundus* Masters students (3 first year, 4 second year) chose to come to Nancy 2 this year; we hope this success will be repeated in the remaining four years of the program.

2.5. Computational Linguistics and Computational Logic

Keywords: *computational logic*, *computational semantics*, *empirical studies*, *inference*, *knowledge representation*, *linguistic resources*, *logic engineering*.

We said above that the central research theme of TALARIS was computational semantics (where "semantics" is broadly construed to cover various pragmatic and discourse level phenomena) and that TALARIS is particularly focused on investigating the interplay between representation and inference. Another way of putting this would be to say that the scientific foundations of TALARIS's work boil down to the motto: *computational linguistics* meets *computational logic* and *knowledge representation*.

From computational linguistics we take the large linguistic and lexical semantics resources, the parsing and generation algorithms, and the insight that (whenever possible) statistical information should be employed to cope with ambiguity. From computational logic and knowledge representation we take the various languages and methodologies that have been developed for handling different forms of information (such as temporal information), the computational tools (such as theorem provers, model builders, model checkers, sat-solvers and planners) that have been devised for working with them, together with the insight that, whenever possible, it is better to work with inference tools that have been tuned for particular problems, and moreover that, whenever possible, it is best to devote as little computational energy to inference as possible.

This picture is somewhat idealized. For example, for many languages (and French is one of them) the large scale linguistic resources (lexicons, grammars, WordNet, FrameNet, PropBank, etc.) that exist for English are not yet available. In addition, the syntax/semantics interface often cannot be taken for granted, and existing inference tools often need to be adapted to cope with the logics that arise in natural language applications (for example, existing provers for Description Logic, though excellent, do not cope with temporal reasoning). Thus we are not simply talking about bringing together known tools, and investigating how they work once they are combined — often a great deal of research, background work and development is needed. Nonetheless, the ideal of bringing together the best tools and ideas from computational linguistics, knowledge representation and computational logic and putting them to work in coordination is the guiding line.

Another simplification involved in the "computational linguistics meet computational logic and knowledge representation" motto is that often the goal is to find out when the use of computational logic can be *avoided* or *minimized*. Logical inference can be computationally expensive, and if simpler statistical methods can be used, or if only computationally tractable inference methods (such as model checking) are required, then it is highly desirable to turn to them. Empirically inspired heuristics are needed so that the tools of computational logic are only applied when truly needed, and only to the smallest problems possible.

To ensure that theoretically plausible ideas really are applicable, and to gain insight as to when empirically oriented methods can be usefully employed, TALARIS focuses on concrete semantic phenomena (for example, tense and aspect, presupposition and anaphora resolution, dialogue structure, etc.). By carefully examining the empirical data, we aim to determine which phenomena require inference and which not; which can be dealt with using weak logics and which not; which can be handled statistically and which not; what scales up successfully and what does not...

2.6. Semantics and Inference

Over the next decade, progress in natural language semantics will likely depend on obtaining a deeper understanding of the role played by inference. One of the simplest levels at which inference enters natural language is as a disambiguation mechanism. Utterances in natural language are typically highly ambiguous: inference allows human beings to (seemingly effortlessly) eliminate the irrelevant possibilities and isolate the intended meaning. But inference can be used in many other processes, for example, in the integration of new information into a known context. This is important when generating natural language utterances. For this task we need to be sure that the utterance we generate is suitable for the person being addressed. That is, we need to be sure that the generated representations fit in well with the recipient's knowledge and expectations of the world, and it is inference which guides us in achieving this.

Much recent semantic research actively addresses such problems by systematically integrating inference as a key element. This is an interesting development, as such work redefines the boundary between semantics and pragmatics. For example, van der Sandt's algorithm for presupposition resolution (a classic problem of pragmatics) uses inference to guarantee that new information is integrated in a coherent way with the old information.

The TALARIS team investigates such semantic/pragmatic problems from various angles (for example, from generation and discourse analysis perspectives) and tries to combine the insights offered by different approaches. For example, for some applications (e.g., the textual entailment task) shallow syntactic parsing combined with fast inference in description logic may be the most suitable approach. In other cases, deep analysis of utterances or sentences and the use of a first-order inference engine may be better. Our aim is to explore these approaches and their limitations.

2.7. Linguistic Resources

In an ideal world, computational semanticists would not have to worry overly much about linguistic resources. Large scale lexica, treebanks, and wide coverage grammars (supported by fast parsers and offering a flexible syntax semantics interface) would be freely available and easy to combine and use. The semanticist could then focus on modeling semantic phenomena and their interactions.

Needless to say, in reality matters are not nearly so straightforward. For a start, for many languages (including French) there are no large-scale resources of the sort that exist for English. Furthermore even in the case of English, the idealized situation just sketched does not obtain. For example, the syntax/semantics interface cannot be regarded as a solved problem: phenomena such as gapping and VP-ellipsis (where a verb, or verb phrase, in a coordinated sentence is missing and has to be somehow "reconstructed" from the previous context) still offer challenging problems for semantic construction.

Thus a team like TALARIS simply cannot focus exclusively on semantic issues: it must also have competence in developing and maintaining a number of different lexical resources (and in particular, resources for French).

TALARIS is involved in such aspects in a number of ways. For example, it participates in the development of an open source syntactic an synonymic lexicon for French, in an attempt to lay the ground for a French version of FrameNet; and it also works on developing a large scale, reversible (i.e., usable both for parsing and for generation) Tree Adjoining Grammar for French.

2.8. Logic Engineering

Once again, in the ideal world, not only would computational semanticists not have to worry about the linguistic resources at their disposal, but they would not have to worry about the inference tools available either. These could be taken for granted, applied as needed, and the semanticist could concentrate on developing linguistically inspired inference architectures. But in spite of the spectacular progress made in automated theorem proving (both for very expressive logics like predicate logics, and for weak logics like description logics) over the last decade, we are not yet in the ideal world. The tools currently offered by the automated reasoning community still have a number of drawbacks when it comes to natural language applications.

For a start, most of the efforts of the first-order automated reasoning community have been devoted to theorem proving; model building, which is also a required technology for natural language processing, is nowhere nearly as well developed, and far fewer systems are available. Secondly, the first-order reasoning community has adopted a resolutely 'classical' approach to inference problems: their provers focus exclusively on the satisfiability problem. The description logic community has been much more flexible, offering architectures and optimisations which allow a greater range of problems to be handled more directly. One reason for this has been that historically, not all description logics offered full Boolean expressivity. So there is a long tradition in description logic of treating a variety of inference problems directly, rather than via reduction to satisfiability. Thirdly, many of the logics for which optimised provers exists do not directly offer the kinds of expressivity required for natural language applications. For example, it is hard to encode temporal inference problems in implemented versions of description logics. Fourth, for very strong logics (notably higher-order logics) few implementations exists and their performance is currently inadequate.

These problems are not insurmountable, and TALARIS members are actively investigating ways of overcoming them. For a start, logics such as higher-order logic, description logic and hybrid logic are nowadays thought of as various fragments of (or theories expressed in) first-order logic. That is, first-order logic provides a unifying framework that often allows transfer of tools or testing methodologies to a wide range of logics. For example, the hybrid logics used in TALARIS (which can be thought of as more expressive versions of description logics) make heavy use of optimization techniques from first-order theorem proving.

Moreover — and from a logical perspective, this is the most interesting point — the interaction between natural language and computational logic is not a one way street. The problems that arise in natural language may well be significant for developments in computational logic. As an example of this, early versions of the CURT software (an educational system for computational semantics developed by Patrick Blackburn and Johan Bos) made use of a standard first-order model builder called MACE. The inference problems that the system generated were then used as tests when the PARADOX model builder was developed, leading to considerable performance improvements. Similarly, natural language applications have also inspired significant performance enhancements to the RACER description logic prover. Feedback from natural language to logic is likely to be an important theme in future developments.

2.9. Empirical Studies

The role of empirical methods (model learning, data extraction from corpora, evaluation) has greatly increased in importance in both linguistics and computer science over the last fifteen years. TALARIS members have been working for many years on the creation, management and dissemination of linguistic resources reusable by the scientific community, both in the context of implementation of data servers, and in the definition of standardized representation formats like TAG-ML. In addition, they have also worked on the applications of linguistic ideas in multimodal settings and multimedia.

The work in this area is in concordance with our scientific projects. As we said above, one of the most important points that needs to be understood about logical inference is how its use can be minimized and intelligently guided. Ultimately, such minimization and guidance must be based on empirical observations concerning the kinds of problems that arise repeatedly in natural language applications.

Finally, is should be remarked that the emphasis on empirical studies lends another dimension to what is meant by inference. While much of TALARIS's focus is on symbolic approaches to inference, statistical and probabilistic methods, either on their own or blended with symbolic approaches, are likely to play an increasingly important role in the future. TALARIS researchers are well aware of the importance of such approaches and are interested in exploring their strengths and weaknesses, and where relevant, intend to integrate them into their work.

3. Application Domains

3.1. Modular Grammar Building.

The development of large scale grammars is a complex task which usually involves factorising information as much as possible. While good grammar writing and factorisation environments exist for "non tree grammars" (e.g., HPSG, LFG), this is not the case for "tree based grammars" such as TAG, Interaction Grammars or Tree Description Grammars. The Extended Metagrammar Compiler (XMG) developed at TALARIS remedies this shortcoming while additionally providing a clean and modular way to describe several linguistic dimensions thereby supporting the production of tree grammars with semantic information[1].

3.2. Referential Expressions

TALARIS has a longstanding interest in the semantics and the processing of referential expressions. In recent years, an extensive corpus annotation has been carried out on 5.000 definite descriptions[2]; an algorithm for generating bridging definite descriptions has been specified and implemented which illustrates the interaction of realisation and inference[3]; a constraint based algorithm for definite description has been proposed which differs from the standard one in that it uses constraints to produce a minimal description[4]; and a shallow anaphora resolver for French has been developed and evaluated within the national evaluation campaign MeDIA.

3.3. Surface Realization

The tree adjoining grammar for French developed by TALARIS associates with each NL expression not only a syntactic tree but also a semantic representation. Interestingly, the semantic calculus used is reversible in that the association between strings and semantic representations is non-directional (declarative). We put this feature to work and have been working over the years towards developing a surface realiser for French called GenI[5]. At present GenI is the only surface realiser available for French. Current work concentrates on improving both coverage and efficiency.

3.4. Textual Entailment.

In essence, the textual entailment recognition task is an inference task, namely deciding whether the information contained in a given text T_1 can be inferred from the information provided by another text T_2.

It is crucial to be able to answer this question. One important characteristic of natural language is the large number of ways in which it can express the same information. Many natural language processing applications like question answering, information retrieval, generation, and anaphora resolution need to deal with this diversity efficiently and accurately, and recognising textual entailments is a key step towards this.

---

[1][65], [67], [66], [68], [64], [69], [70], [94], [91], [92], [93], [81], [90]
[2][85], [89], [86], [88], [87], [84], [78], [79]
[3][95], [80], [82], [57]
[4][71], [57]
[5][83], [77]

Textual entailment recognition is a difficult task. The approach we are experimenting with is to encode lexical information as a description logic ontology (or a hybrid logic theory) and then to use logical inference to compute the result.

3.5. Computational Logics and Computational Semantics.

Members of TALARIS are among the main figures proposing the idea of using inference (and in particular, using computational tools like model builders and theorem provers) as an integral part of different tasks in computational semantics, mainly during semantic construction[6]. The book "Representation and Inference for Natural Language: A First Course in Computational Semantics"[7] by Blackburn and Bos is nowadays an important reference in this area.

3.6. Hybrid Automated Deduction

TALARIS main contribution in this topic has been the design of resolution calculi for hybrid logics, that were then implemented into the HyLoRes theorem prover. In particular, TALARIS members have proved that such calculi can be enhanced with optimisations of order and selection functions without losing completeness. Moreover, the first 'effective' (i.e., directly implementable) termination proof for $\mathcal{H}(@)$ has been recently established and the technique is being extended to more expressive languages[8]. Current work includes optimising performance by use of ordered resolution, adding a temporal reasoning component to HYLORES, extending the architecture to allow querying against a background theory without having to re-saturate the theory with each new query, and testing HYLORES performance against dedicated first-order provers using a structure-preserving translation method.

During this year, and making use of modularized code from HYLORES, we have implemented a tableaux based prover for hybrid logics called HTAB [44]. HTAB is an optimised tableaux prover for hybrid logics, using algorithms that ensure termination [15]. It ultimately aims to cover a number of frame conditions (i.e., reflexivity, symmetry, antisymmetry, etc.), as far as it is possible to ensure termination [28]. Moreover, we are interested in providing a range of inference services beyond satisfiability checking. For example, the current version of HTAB includes model generation (i.e., HTAB can generate a model from a saturated open branch in the tableau).

We have also started to explore other decision methods (e.g., game based decision methods) which are useful for non-standard semantics like topological semantics. The prover HYLOBAN is an example of this work [51].

3.7. Multimedia

MLIF (Multi Lingual Information Framework) is being designed as a generic ISO-based mechanism for representing and dealing with multilingual textual information. A preliminary version of MLIF has been associated to digital media within the ISO/IEC MPEG context and dealing with subtitling of video content, dialogue prompts, menus in interactive TV, and descriptive information for multimedia scenes. MLIF comprises a flexible specification platform for elementary multilingual units that may be either embedded in other types of multimedia content or used autonomously to localise existing content.

4. Software

4.1. The eXtended Meta-Grammar (XMG) Compiler and Tools

A metagrammar compiler generates automatically a grammar from a reduced description called a MetaGrammar. This description captures the linguistic properties underlying the syntactical rules of a grammar. Various past and present TALARIS members have been working on metagrammar compilation since 2001 and several tools have been developed within this framework starting with the MGC system of Bertrnad Gaiffe (now of ATILF, *Analyse et Traitment Informatique de la Langue Francaise* a Nancy-based CNRS unit) to the newly developed XMG system of Crabbé et al.

---

[6][58], [62]
[7][63]
[8][61], [60]

The XMG system is a 2nd generation compiler that proposes (a) a representation language allowing the user to describe in a factorised and flexible way the linguistic information contained in the grammar, and (b) a compiler for this language (using a Warren Abstract Machine-like architecture). An innovative feature of this compiler is the fact that it makes it possible to describe several linguistic dimensions, and in particular it is possible to define a natural Syntax/Semantics interface within the Metagrammar.

The compiler actually supports two syntactic formalisms (Tree Adjoining Grammars and Interaction Grammars) and the description both of the syntactic and of the semantic dimension of natural language. The generated grammars are in XML format, which makes them easy to reuse. Plug-ins have been realised with the LLP2 parser, with Eric de la Clergerie's DyALog parser and with the GENI generator. Future work concerns the modularisation and the extension of XMG to define a library of languages describing linguistic data allowing the user to describe his own target formalism.

Developed under the supervision of Denys Duchier, the XMG compiler is the result of an intensive collaboration with CALLIGRAMME. It has been implemented in Oz/Mozart and runs under the Linux, Mac, and Windows platforms. It is available with tools easing its use with parsers and generators (tree viewer, duplicate remover, anchoring module, metagrammar browser).

The system is currently being used and tested by Owen Rambow (University of Columbia, USA) and Laura Kallmeyer (University of Tuebingen, Germany).

Version: 1.1.4
License: CeCILL
Last update: 27/09/2005
Web site: http://sourcesup.cru.fr/xmg/
Documentation: http://sourcesup.cru.fr/xmg/#Documentation
Authors: Benoit Crabbé, Denys Duchier, Joseph Le Roux, Yannick Parmentier
Contact: Benoit Crabbé, Yannick Parmentier

4.2. Frolog

Frolog is a dialogue system based on current technology from computational linguistics, artificial intelligence planning and theorem proving. It implements a text adventure game engine that uses natural language processing techniques to analyse the player's input and generate the system's output.

The Frolog core is implemented in Prolog, but it uses external tools for the most heavy-loaded tasks. It performs syntactic analysis of the input based on an English grammar developed using XMG and computes a flat semantic representation using the SEMCONST semantic construction tool. It then uses the constructed semantic representation and an off-the-shelf planner to interpret the player's intention and change the world model accordingly. The world is modelled as a knowledge base in description logics, and accessed using the Description Logic theorem prover *Racer*. Finally, the results of the action, or descriptions of objects, are generated automatically, using the GENI generator.

Frolog's main utility is to serve as a laboratory in order to test pragmatic theories about presupposition accommodation. However, it will also result in the first integrated system to use SEMTAG (the LORIA toolbox for TAG-based Parsing and Generation).

Version: 0.9
License: GPL
Last update: 2007-11-26
Web site: http://trac.loria.fr/projects/frolog/wiki
Documentation: http://trac.loria.fr/projects/frolog/wiki/Documentation
Authors: Luciana Benotti, Alejandra Lorenzo, Laura Perez
Contact: Luciana Benotti

4.3. GenI Generator

The GENI generator is a successor of the InDiGen generator. Also based on a chart algorithm, it is implemented in Haskell (one of the leading functional programming languages available nowadays) and aims for modularity, re-usability and extensibility. The system is "stand-alone" as we use the Glasgow Haskell compiler to obtain executable code for Windows, Solaris, Linux and Mac OS X.

The GENI generator uses efficient datatypes and intelligent rule application to minimise the generation of redundant structures. It also uses a notion of polarities as a means first, of coping with lexical ambiguity and second, of selecting variants obeying given syntactic constraints.

The grammar used by the GENI generator is produced using the MetaGrammar Compiler and covers the basic syntactic structures of French as described in Anne Abeillé's book "An electronic grammar for French".

The system can process the output of the XMG Metagrammar compiler mentioned above.

Version: 0.8
License: GPL
Last update: 2005-10-17
Web site: http://wiki.loria.fr/wiki/GenI
Documentation: http://wiki.loria.fr/wiki/GenI/Manual
Project(s): GENI
Authors: Carlos Areces, Claire Gardent, Eric Kow
Contact: Claire Gardent

4.4. HyLoRes, a Resolution Based Theorem Prover for Hybrid Logics

HYLORES is a resolution based theorem prover for hybrid logics (it is complete for the hybrid language H(@,↓), a very expressive but undecidable language, and it implements a decision method for the sublanguage H(@)). It implements a version of the "given clause" algorithm which is the underlying framework of many current state of the art resolution-based theorem provers for first-order logic; and uses heuristics of order and selection function to prune the search space on the space of possible generated clauses.

HYLORES is implemented in Haskell, and compiled with the Glasgow Haskell compiler (thus, users need no additional software to use the prover). We have also developed a graphical interface.

The interest of HYLORES is twofold: on one hand it is the first mature theorem prover for hybrid languages, and on the other, it is the first modern resolution based prover for modal-like languages implementing optimisations and heuristics like order resolution with selection functions.

Version: 2.4
License: GPL
Last update: 2007-12-01
Web site: http://trac.loria.fr/projects/hylores
Documentation: http://trac.loria.fr/projects/hylores
Authors: Carlos Areces, Daniel Gorín and Juan Heguiabehere
Contact: Carlos Areces

4.5. HTab, a Tableau Based Theorem prover for Hybrid Logics

The main goal behind HTAB is to make available an optimised tableaux prover for hybrid logics, using algorithms that ensure termination. We ultimately aim to cover a number of frame conditions (i.e., reflexivity, symmetry, antisymmetry, etc.), as far as we can ensure termination. Moreover, we are interested in providing a range of inference services beyond satisfiability checking. For example, the current version of HTAB includes model generation (i.e., HTAB can generate a model from a saturated open branch in the tableau).

HTAB and HYLORES are actually being developed in coordination, and a generic inference system involving both provers is being designed. The aim is to take advantage of the dual behaviour existing between the resolution and tableaux algorithms: while resolution is usually most eficient for unsatisfiable formulas (i.e., a contradiction can be reported as soon as the empty clause is derived), tableaux methods are better suited to handle satisfiable formulas (i.e., a saturated open branch in the tableaux represents a model for the input formula).

Version: 1.2.1
License: GPL
Last update: 2007-12-01
Web site: http://trac.loria.fr/projects/htab
Documentation: http://trac.loria.fr/projects/htab
Authors: Carlos Areces, Guillaume Hoffmann
Contact: Guillaume Hoffmann

4.6. HyLoBan, a Game Based Theorem Prover for Topological Hybrid Logics

HYLOBAN is a game-based prover, resulting from a direct implementation of Sustretov's game-based proofs of the PSPACE-completeness of the hybrid logics of T0 and T1 topological spaces. The interest of this approach is that termination is guaranteed and in addition the underlying game-based architecture is of independent interest; its disadvantage is that (at present) it is still extremely ineficient.

Version: 0.2
License: GPL
Last update: 2007-12-01
Web site: http://trac.loria.fr/projects/hyloban
Documentation: http://trac.loria.fr/projects/hyloban
Authors: Carlos Areces, Guillaume Hoffmann, Dmitry Sustretov
Contact: Guillaume Hoffmann

4.7. hGEN, a Random Formula Generator

hGen is a random CNF (conjunctive normal form) generator of formulas for sublanguages of H(@, ↓, A, P). It is an extension of the latest proposal of Patel-Schneider and Sebastiane, nowadays considered the standard testing environment for classical modal logics. The random generator is used for assessing the performance of different provers.

Version: 1.1
License: GPL
Last update: 2007-12-01
Web site: http://trac.loria.fr/projects/hgen
Documentation: http://trac.loria.fr/projects/hgen
Authors: Carlos Areces, Daniel Gorín, Juan Heguiabehere and Guillaume Hoffmann
Contact: Carlos Areces

4.8. SynLex: Extracting a Syntactical Lexicon from the LADL Tables

Maurice Gross' grammar lexicon contains extremely rich and exhaustive information about the morphosyntactic and semantic properties of French syntactic functors (verbs, adjectives, nouns). Yet its use within natural language processing systems is still restricted.

The aim of our work is to translate this information into a format which is more suitable for use by NLP systems and also compatible with the state of the art practice in lexical data representation.

The lexicon should assign to each verb a set of subcategorisation frames. Frames are defined by a list of atoms (e.g., A0 V A1) representing the verb and its arguments, and by a list of atoms/feature structure pairs specifying the feature values associated with each of these atoms.

Two sets of subcategorisation lexicons (called LADL-SynLex and NLP-SynLex) were extracted from the LADL tables. The current SynLex contains the LADL- and NLP-SynLex lexicons for the LADL-tables 1, 2, 4, 5, 7, 8, 10, 11, 13, 14 and 16 which amounts to roughly 2.000 verb usages. Work is underway to process the remaining available tables which should yield a description of roughly 6.500 verbs.

SynLex is the result of joint work between TALARIS, ATILF and CALLIGRAMME[9]. It is currently being evaluated by the members of the ILF (Institut de la Langue Française) funded LexSynt project, and is partially available at http://www.loria.fr/~gardent/ladl.

Last update: 2005-10-14
Web site: http://www.loria.fr/~gardent/ladl/
Documentation: http://www.loria.fr/~gardent/ladl/
Project(s): SynLex
Authors: Claire Gardent, Guy Perrier, Bruno Guillaume, Ingrid Falk
Contact: Claire Gardent

4.9. MEDIA

In the framework of the MEDIA project, software has been developed to process transcriptions of a spoken dialogue corpus and to provide a semantic representation of their task-related content. This software contains a tokeniser, a TAG parser (LLP2), a TAG grammar, an OWL ontology and a set of rules in description logic, and works together with a reasoner such as RACER. The modularity of its architecture and the use of the Java programming language enable this software to be run on multiple platforms and to be easily adapted to other transactional contexts besides hotel reservation (it original application domain). This software aims to be further improved to implement reference resolution and dialogic contextual understanding (during the second stage of the MEDIA project) and eventually to be embedded in dialogue systems.

Version: 0.3
License: GPL
Last update: 08/11/2005
Project(s): MEDIA
Authors and Contact: Alexandre Denis

4.10. CURT (Clever Use of Reasoning Tools)

The CURT (Clever Use of Reasoning Tools) family is a series of simple dialogue systems which illustrate how tools for building semantic representations can be combined with inference tools.

The behaviour of the different CURT programs is as follows: the user extends CURT's knowledge by entering English sentences, and can query it about its acquired knowledge.

The CURT family is composed of Baby CURT (the backbone of the CURT system using no inference services), Rugrat CURT (including either a simple free variable tableau prover or resolution prover to check the consistency of the current dialog), Clever CURT (which performs consistency checking by running a sophisticated first-order theorem prover and model checker in parallel), Sensitive CURT (which checks in addition for informativeness of the discourse), Scrupulous CURT (which eliminates equivalent interpretations), Knowledgeable CURT (which adds lexical and world knowledge) and Helpful CURT (which is able to handle simple natural language questions from the user).

A multilingual version of CURT is being developed (covering French, Romanian and Spanish).

Version: 1.0

---

[9][74], [72]

License: GPL
Last update: 2005-10-11
Web site: http://www.comsem.org
Documentation: http://www.comsem.org
Authors: Carlos Areces, Patrick Blackburn, Johan Bos, Sébastien Hinderer, Daniela Solomon
Contact: Carlos Areces, Patrick Blackburn, Sébastien Hinderer.

4.11. Nessie

Nessie is a library providing facilities for semantic construction. It is written in OCaml and uses typed lambda-calculus and first-order logic as underlying formalisms. It allows the user to flexibly build terms and term trees; and once a lambda-term tree is built, its semantic representation can be efficiently computed.

For test purposes, this library has been interfaced with CURT's parser.

Future developments of Nessie will include an extension to other logics (modal logic, hybrid logic...), interfaces with French grammars (e.g., those generated by the XMG system), interface with inference tools, etc.

Last update: 2005-10-11
Authors: Sébastien Hinderer
Contact: Sébastien Hinderer

4.12. DeDe Corpus

DeDe is a corpus of roughly 50.000 words where around 5.000 definite descriptions have been annotated as coreferential, contextually dependent, non referential or autonomous. The corpus consists of articles from the newspaper *Le Monde* and is annotated with Multext-based morphosyntactic information[10].

Authors: Claire Gardent, Hélène Manuelian Contact: Claire Gardent

4.13. SemFRaG

A TAG grammar developed with the XMG metagrammar compiler and which describes both the syntax and the semantics of natural language expressions. Syntactically, the grammars covers the TSNLP testsuite and work is in progress to acquire an equivalent semantic coverage. Used both for parsing and for generation.

Authors: Claire Gardent, Benoit Crabbé
Contact: Claire Gardent

4.14. FRoG — French Resource Grammar

An HPSG grammar of French developed with the LKB platform. The grammar incorporates a treatment of interface phenomena (syntax-semantics, phonology-syntax, morphology-syntax) in a constraint-based framework designed for bidirectionality (parsing and generation).

Version: 0.1
License: LGPL-LR
Last update: 01/11/2005
Project(s): Delph-In
Authors and Contact: Jesse Tseng

---

[10][78]

4.15. LLP2

LLP2 is an LTAG Parser based on the bottom up algorithm described in Patrice Lopez's thesis. The present version is restricted to TIG. The parser is compliant with the TAGML2 resources format and is capable of processing a graph of words as input. Furthermore, an external utterance segmenter can be plugged in. The distribution comes with graphical exploration and debugging tools.

Version: 1.0
Last update: 31/05/2005
Web site: http://www.loria.fr/~azim/LLP2/help/fr/
Documentation: http://www.loria.fr/~azim/LLP2/help/fr/
Project(s): Passage
Authors and Contact: Azim Roussanaly

5. New Results

5.1. Introduction

To structure our discussion of of the new results for TALARIS in 2007, we shall discuss the four main themes in turn:

- Computational Semantics
- Discourse, Dialogue and Pragmatics
- Logics for Natural Language and Knowledge Representation,
- Multilinguality for Multimedia

5.2. Computational Semantics

The Computational Semantics group in TALARIS focuses on two main points:

- The development of a computational infrastructure for the semantic processing of French
- The interfacing of natural language processing (NLP) systems with knowledge based inference

The bulk of the work in these area is led by Claire Gardent, who guides the work of Marilisa Amoia, Paul Bedaride, Ingrid Falk, Eric Kow, Yannick Parmentier, Sylvain Schmitz, and Fabienne Venant. In addition, Patrick Blackburn and Sébastien Hinderer work on computational semantics (though applications in French are not the focus of their work). Many of the tools produced by Computational Semantics group are used by other TALARIS researchers, notably from the Discourse, Dialogue and Pragmatics group.

5.2.1. *Computational Semantics for French*

In 2007, work concentrated mainly on finalising a basic parsing and generating architecture for the semantic processing of French. This involves developing, integrating and evaluating several modules including: a syntactic lexicon (SYNLEX) and a lexicon for verb synonyms (SYN2), a grammar (SEMFRAG), a module for semantic construction (SEMCONST) and a module for surface realisation (GENI). The computational semantic group also initiated and organised in collaboration with CALLIGRAMME a workshop on XMG, a grammar writing environmment developed in Nancy.

5.2.1.1. SynLeX.

A verbal syntactic lexicon lists for each verb the type and the number of its arguments. Such a lexicon is required for any NLP application involving either parsing or generation. However until recently, no such lexicon was available for French. Over the last three years, TALARIS has worked on extracting such a lexicon from a large scale linguistic resource manually developed under Maurice Gross' guidance namely, the LADL tables. Although these tables are extremely rich, they cannot directly be used in NLP for two reasons. First, their format does not fit the requirements of NLP systems and second, much of its information (in particular inter-dependencies) is only implicitly represented. To remedy these shortcomings, we devised a method to extract the relevant information from these tables and translate it into a format amenable to NLP[11]. However it was immediately clear that the lexicon extracted using these methods was imperfect. We therefore carried out an evaluation of this lexicon using corpus based methods and computing precision (percent of correct entries wrt a gold standard) and recall (percent of gold standard entries present in the extracted lexicon). The results [35] show that SYNLEX has a low precision but reasonably good recall thus suggesting that manual validation is a good way to clean up the lexicon and thereby produce a reasonably extensive and correct resource. To this end, we currently cooperate with CALLIGRAMME within the MISN CPER operation BDSyn, where the objective is to develop a database to normalise, store and manipulate the lexicon content; and to then interface this database with a web service supporting validation by multiple annotators.

5.2.1.2. Syn2.

To reason about the meaning of text, lexical semantic knowledge is necessary and in particular, knowledge about synonyms (to detect that two sentences carry the same meaning, it is necessary to be able to detect synonymous words). Together with ATILF, we initiated the MISN CPER operation Syn2 whose aim is to merge 5 synonym dictionaries on the basis of the TLFi definitions. In a first step, we defined a method based on similarity measures between definitions which permits regrouping synonyms by senses. It remains to be seen how this method performs on the data. Next we intend to evaluate and compare several variants of the proposed methods wrt a gold standard. Once the best method is identified, a unique synonym lexicon will be created which assigns a word and its possible meanings, the sets of synonyms corresponding to each given meaning.

5.2.1.3. SemFrag.

To parse and generate sentences a grammar is required. CALLIGRAMME and TALARIS have been developing over the last 5 years medium size grammars for French using the XMG grammar writing environment developed by Denys Duchier, Yannick Parmentier and Joseph LeRoux. The TALARIS grammar SEMFRAG is a Tree Adjoining Grammar augmented with a unification based compositional semantics. A distinguishing feature of SEMFRAG is its reversibility: the grammar can be used with a parser to derive the semantic representation of a sentence or conversely, with a realiser to produce the sentences associated by the grammar with a given meaning. In [38], we show that the type of semantics used in SEMFRAG obey some general principles which are common to other types of unification based semantics such as the glue semantics used in Lexical Functional Grammar and make it easier to integrate the required semantic information in a large scale grammar. In future, we plan to use these principles to reformulate the semantic dimension of SEMFRAG in a more general and compact way.

5.2.1.4. SemConst.

To derive semantic representations from sentences, we use SEMFRAG together with Eric de la Clergerie's TAG parser and a semantic construction module implemented by Claire Gardent. The module is available on the web and was presented both at TALN (the French NLP conference) and at ACL (Association for Computational Linguistics) [43], [42]. A more detailed description of the system and of its theoretical underpinnings is given in Yannick Parmentier's PhD thesis [14].

---

[11][73], [75], [76]

5.2.1.5. GenI.

The surface realiser GENI takes as input SEMFRAG and a semantic representation and produces as output the set of strings associated by the grammar with the input semantic representation. In [41], [39], we show how the realiser's input can be parameterised to enforce symbolic based determinism (the realiser produces a single output obeying the symbolic constraints encoded in the enriched input). In [40], we show further how GENI can be used to detect over-generation in a grammar – we show in particular that the proposed methods permits reducing over-generation by 70% in roughly 13 hours linguist work. A detailed description of GENI and of its theoretical underpinnings is given in Eric Kow's PhD thesis [13].

5.2.1.6. XMG workshop.

The XMG grammar writing environment was developed by Denys Duchier, Yannick Parmentier and Joseph LeRoux in 2006 and is now used by a medium size international (France, Germany, Israel, USA) community. To assess the needs of the users and better plan further extensions, TALARIS and CALLIGRAMME organised a 2 day workshop in Nancy[12] which gathered 31 participants from 4 countries and 12 distinct institutions. The workshop resulted in a number of changes to XMG including the addition of a graphical interface for visualising and manipulating an XMG hierarchy and the adaptation of the framework to Multi Components Tree Adjoining Grammars [47].

5.2.2. NLP and Inference

Work on inference has concentrated on textual entailment recognition for English. [24] shows how description logic can be used to model lexical semantic based reasoning when assessing whether one sentence entails another while [22] presents a first order semantics for adjectives which permits capturing the complex interplay of compositional, lexical and morphoderivational semantics determining whether or not entailment holds. In additon, Patrick Balckburn and Sébastien Hinderer have worked on the automatic generation of models for Polish temporal expressions.

5.3. Discourse, Dialogue and Pragmatics

The Discourse, Dialogue and Pragmatics group in TALARIS focuses on two main themes:

- The study of grounding, mutual understanding, and collaboration.
- The study of presupposition and information accommodation for a planning-based perspective.

The bulk of the work in this area is conducted by Matthieu Quignard, Patrick Blackburn, Alexandre Denis, Luciana Benotti, Daniel Coulon and Carlos Areces. The group is making increasingly heavy use of the tools provided by the Computational Semantics group (notably the GENI generator, XMG based grammars, and the SEMCONST grammar constructor), a trend that is likely to continue. In addition, the group is making increasing use of a inference tools, which leads to links with the themes explored by the the Logics for Natural Language and Knowledge Representation group. Indeed, this is as it should. Linguistics have long viewed pragmatics, the study of how natural languages are actually used, as providing an insight into a level of meaning over an above (though mediated by) the meanings provided by pragmatics. And, crucially, this level of meaning crucially involves inference. Thus, in a sense this group provides a bridge between the semantically and logically oriented work of the other themes.

5.3.1. Grounding and Mutual Understanding

One of the most difficult problems in discourse is guaranteeing mutual understanding. Early AI approaches to dialog (that is, work from the 1970s and 1980s) tended to ignore the problem: such models assumed perfect understanding or had very crude models of repair.

From a theoretical perspective such approaches are very bad. Dialog is essentially one long mutual effort at negotiating and checking understanding. Moreover, from a practical perspective such a model leads to inflexible dialogue systems.

---

[12]http://www.loria.fr/~parmenti/WSP-XMG/

One of the successes of the group this year was to propose a detailed computational model of the process of *grounding*, that is, the exchange of signals that takes place during the process of accepting/rejecting dialog information [34]. A notable feature of this model is that it offered an approach to deal with situations where the evidence of understanding has itself been misunderstood.

This model has now been implemented, and forms part of the ongoing work by Alexander Denis and Matthieu Quignard on obtaining a detailed computational model of dialog suitable for implementing robust dialogue systems — one of the long term goals of the group.

5.3.2. *Presupposition and Information Accommodation*

A pervasive feature of the way we use natural language is the heavy use made of inference to smooth the process on communication. We don't have to spell everything out: we rely on the fact that the people we talk with have lots of knowledge and experience that lets them find their way to the correct interpretation.

For example, when giving people instructions, we typically don't give all the details: if we ask someone to make a salad, we typically don't tell them that they should wash the lettuce as a part of this process. We rely on the fact that people can successfully "fill in" such tacit actions. The study of such linguistic inferences belongs to the area known as pragmatics, and in particular, the study of presupposition and accommodation.

An explicit computational model of part of this process was given by Luciana Benotti [27]. Taking as her starting point a text adventure game called FrOz that made use of Description Logic inference tools, she added planning capability to it (yielding FrOzA, or FrOz Advanced system). The use of planning techniques enables the game dialogue system to "fill in" tacit action required by the players instructions, which results in far more natural, and linguistically plausible interactions.

This work is currently being extended by Luciana Benotti and Patrick Blackburn. A reimplementation, called Frolog, of this system is nearing completion. This reimplementation allows all language processing and inference and planning tasks to be handed over to external modules so that the ideas can be applied in more sophisticated settings.

5.4. Logics for Natural Language and Knowledge Representation

The Logics for Natural Language and Knowledge representation focuses on two main points:

- The theoretical study of hybrid logic (propositional, first-order, and higher-order) and the implementation of efficient proof methods for them.

- Investigating other logics of relevance to natural language and knowledge representation, notably description logic, dedicated planning methods, and discourse representation theory (DRT).

The bulk of the work in this area is conducted by Carlos Areces, Patrick Blackburn, Dmitry Sustretov, Guillaume Hoffmann, Daniel Gorín and Sergio Mera. The inference methods studied by this group are relevant to the work of the Computational Semantics group and the Discourse, Dialogue and Pragmatics group, in particular the work of Luciana Benotti on presupposition and information accomodation (which makes heavy used of description logic and planning) and Paul Bedaride's work on textual entailment (which explores the use of description and hybrid logics).

5.4.1. *Automated Theorem Proving for Hybrid Logics*

During 2007, this topic received an important impulse. On the one hand, HYLORES, the resolution based theorem prover for hybrid logics, has finally arrived to a very stable stage of developement. Important work has been done during this year, on modularization, testing and performance improving. A graphical interface has been developed, and model generation capabilities included. More importantly, we have started working in the two topics which we want to address during the next year: paralelization and moving into a client-server architecture. We will discuss these two topics in more detail below.

Taking advantage of the modularized code of HYLORES, the second important result of 2007 was the development of HTAB, a tableaux based prover for hybrid logics [44]. HTAB is an optimised tableaux prover for hybrid logics, using algorithms that ensure termination [15]. It ultimately aims to cover a number of frame conditions (i.e., reflexivity, symmetry, antisymmetry, etc.), as far as it is possible to ensure termination [28]. Moreover, we are interested in providing a range of inference services beyond satisfiability checking. For example, the current version of HTAB includes model generation (i.e., HTAB can generate a model from a saturated open branch in the tableau). Even though other provers for languages similar to $\mathcal{H}l(@)$ exists, HTAB has a number of particularities that make it a potentially useful tool. It has already outperformed HyLoTab [13], a previously existing tableaux based prover for hybrid logics, and in many inputs it also outperforms HYLORES.

After our experience in designing and developing HYLORES and HTAB, we are currently in the process of drawing the main lines of a new system that we call INTOHYLO.

INTOHYLO is actually an integrated collection of tools that work in collaboration to offer a varied spectrum of inference services for different hybrid logics. The system will be developed under the GPL licence and the source code will be available on-line. In addition to the core inference tools, we will also make available the testing environment used to test INTOHYLO optimizations and heuristics, in order to encourage independent development.

The main inference task addressed by INTOHYLO will be satisfiability checking, but the system will also be able to offer more varied and complex services, like model generation, model checking and instance retrieval. Initially, INTOHYLO will be created from the integration of HYLORES and HTAB, and this is what we are going to discuss in detail below. But in the future we will consider the addition of other tools (like the HYLOBAN [51] for topological semantics).

The core idea behind INTOHYLO is to take advantage of the inherent dual behavior existing between the resolution based and the tableaux based calculi: while the resolution method performs better on unsatisfiable formulas (during resolution we only need to derive the empty clause to detect that a formula is unsatisfiable), the tableau method performs better on satisfiable formulas (while constructing a tableau we only need to find a saturated open branch to detect that a formula is satisfiable).[14]

Our first step will be to transform HYLORES and HTAB into server applications, while HYLORUN will act as a client application which will connect to the provers submitting queries and displaying the results. HYLORUN will detect whether HYLORES and/or HTAB are running as servers and connect to them using either HTTP or TCP services. This architecture is the one used by the description logic prover *RACER* and we believe that it has some important benefits:

1. To start with, the different components of INTOHYLO (currently, the two provers and the front-end) can evolve independently without interfering with applications making use of them, as long as the communication protocols are maintained. In addition, new inference tools can be added as additional servers and only the front-end will need to be modified to offer these additional services.

2. Secondly, and as we will further explain later, we want to investigate ways in which the two provers can collaborate while working in a given problem. With this idea in mind, we want the provers to be able to exchange information (i.e., partial results) in a manner that is transparent to the user.

3. But the most important reason for choosing this architecture is that it lets us implement a notion of 'proof state'. This idea is again a fundamental characteristic of DL provers like *RACER*: the user should be able to 'load' a problem into the system, and then query it for answers. Perhaps many different queries will be posed to the prover about the same problem, and the prover can take advantage of previous results to answer future queries.

As an example, let us discuss points (2) and (3) above for the case of HYLORES. With respect to point (2), given an input formula φ, during the computation of the saturated set of clauses corresponding to φ, HYLORES

---

[13][96]

[14]We have already tested this intuition empirically: there are satisfiable formulas for which even the current version of HTAB (with only minimal optimizations) beats HYLORES, even though HYLORES is on average far better performing than HTAB, see [44].

can derive unit clauses which can be used by HTAB to close branches. Similarly, formulas which are common to all branches in a tableaux which is being constructed by HTAB, can be sent as unit clauses to be used by HYLORES. This idea is intuitively appealing, but care must be taken in coordinating the way that new nominals (which are used both by HYLORES and HTAB to decompose modal formulas into simpler cases) are introduced.

With respect to point (3), consider a formula φ and suppose that we want to decide whether $\models \varphi \rightarrow \psi_i$ for formulas $\psi_1, ..., \psi_n$. With the current version of HYLORES, we can only check formulas $\varphi \wedge \neg \psi_i$ for satisfiability one by one, reporting that $\models \varphi \rightarrow \psi_i$ everytime that the prover finds an inconsistency. In doing this, we are recomputing time and again the saturated set for φ. It would be much more efficient to first load the formula φ into HYLORES as our current problem, and then send the queries ψ_i one by one. HYLORES would then be able to compute only once the saturated set for φ, and just expand it as necessary for each query ψ_i.

5.4.2. *Higher Order Hybrid Logics*

Higher Order Logics is a classical formalism for natural language semantics semantics. In previous years, we have investigated how the addition of hybrid operators in the classical framework of higher order logics can improve langauge modelling[15].

This year we have taken up again this issue and we are currently working in a sound and complete axiomatization for higher order hybrid logics. We expect to obtain a general completeness result (i.e., covering extensions of the basic axiomatization with pure formulas and existential saturation rules) as it is the case for first-order hybrid logics. Such a result would be of interesting when providing semantics for different natural language phenomena (e.g., time and aspect) which assume special conditions on their formal models.

5.5. Multilinguality for Multimedia

Work in this domain is primarily carried out by Samuel Cruz-Lara, Nadia Bellelem, and Lotfi Bellelem. This is the most applied part of the TALARIS team and also the most independent. Their work centers around:

- MLIF (the Multi Lingual Information Framework) a generic ISO-based mechanism for representing and dealing with multilingual textual information.

- The W3C's SMIL (Synchronized Multimedia Integration Language), which allows an author to describe the temporal behaviour of a multi-media presentation.

This has been a transitional year for the group. The group was responsible for the PASSEPARTOUT project in interactive television which came (very successfully) to an end on the 31 May 2007. Since then the group has been involved in developing a number of research proposals though at the time of writing it was unclear whether these had been successful or not. One of the most interesting of these is the BIONORM project. This is the name of a regional project whose main objective is to develop and to test a platform allowing to manipulate in a standard and interoperable way, several results related to biological examinations published by biological medical laboratories.

In order to help the group achieve its goals, TALARIS recently expanded it: Lotfi Bellalem joined TALARIS in 30 September 2007. With the group thus increased in size, it is hoped that the following years will see a closer integration of their work in the mainstream of TALARIS research. It is also possible that Samuel Cruz-Lara will request temporary attachment to INRIA to aid this process.

6. Other Grants and Activities

6.1. International level

---

[15][59]

6.1.1. Nancy TEI-host

*Description:*On January 1st, 2005, Nancy became the fourth host of the TEI consortium (Text Encoding Initiative - http://www.tei-c.org). This resulted from the wish of ATILF, Loria (TALARIS) and INIST to act together in their contribution to standardisation activities for the encoding of textual information. In this context, TALARIS is particularly active in the domains of spoken corpora, terminology and dictionary encoding techniques.
*Partner(s):*ATILF, INIST

6.2. European Level

6.2.1. AMIGO: Ambient Intelligence for a Networked Home Environment

*Theme:*Ambient Intelligence

*Description:*Within the current trend of research in AI about ambient intelligence, the European AMIGO Project focuses on the design of middleware architecture supporting an optimal interoperability between devices and services for home care and family life. Amongst those services, a particular effort is planned for providing the most convenient way of interacting between systems and human users, and is based on use case scenarios (health and security; home information and entertainment; extended home activities such as working at home) and multimodal interfaces (voice, text messages, 2D and 3D gestures). The participation of TALARIS in the framework of this project is motivated by the design of an enhanced multimodal fusion module, which would extend the one designed in the former OZONE project (voice + 2D paths on a tablet PC) to process also 3D pointing gestures.

Although 2D and 3D devices provides more or less the same type of information (2D paths on a projection screen or display) and the same communicative intention (designation), the introduction of 3D gestures in our multimodal fusion will imply deep changes in our fusion algorithms. In the OZONE system, the moves of the pencil on the touch screen allowed users to select objects with very good accuracy. The low amount of ambiguities enabled us to process the fusion in a quite restricted verbal context. The introduction of massive ambiguities at the level of selected objects would need a better structuration of the dialogue history to eradicate those objects that are not salient in the current dialogue focus, and thus should not be relevant for the fusion.

*Administrative context:*IST European Program

*Web site:*http://www.extra.research.philips.com/euprojects/amigo

*Person(s) in charge:*Harmke de Groot (Philips, Eindhoven)

*Period:*start 2004-09-01 / end 2008-02-28

*Contact:*Matthieu Quignard

*Partner(s):*Philips Research (Eindhoven)

6.2.2. Passepartout: Interactive Television

*Theme:*Linguistic and multimedia resources

*Description:*Digitisation of society is always accelerating. A key factor in this acceleration is now software technology. This project focuses on the convergence of digital systems and applications in home media-centres in compliance with the ITEA roadmap "The Road towards Convergence" thus matching the vision of industries, institutions, SME and government partners. New technologies are expected to emerge from this project, that propel the European software industries on to convergence, over terminals and network towards the final goal of ambient intelligence.

The project aims at coupling home media-centers to home networks for rendering scalable content from high definition television (HDTV) to lower definitions in a seamless fashion. Integral to the content will be reactive access and interactivity of high-resolution graphics using ISO and W3C standards for object oriented TV. With the project's goal to make a step towards ambient intelligence through mass personalisation of reactive content (RAMPEG), implementation shall use the most practical elements of MPEG-4 and MPEG-7 with W3C standards such as SMIL, related content synthesis and syndication in XML. Implications will stretch far beyond infrastructure and basic services but will also affect content, human system interaction and engineering.

Implementation will be based on content access using a PVR media-centre as server to new generations of access networks, including Blu-ray optical storage and WIMAX wireless technology. These networks will support the creation of home media-centers that move beyond current STB and PVR-DVD players using MPEG-2 technology, to create true mass-customisation device for family entertainment, with the goals of content packaging and personalisation to match the cultural and linguistic needs of the states of the EU and their economies.

*Administrative context:*ITEA

*Period:*start 2005-01-01 / end 2007-03-31

*Contact:*Samuel Cruz-Lara

*Partner(s):*Cybercultus, Centre de Recherche Publique Henry Tudor, INT, Thomson, RTL, Philips, Telvent, Universidad Politecnica de Madrid, Universidad de Vigo, CharToon, Stoneroos, ETRI, VTT Electronics, V2, CWI, Technische Universiteit Eindhoven, Gradient

6.3. National Level

6.3.1. Lexsynt (ILF)

*Theme:*Lexicon: syntax and semantic

*Description:*The main aim of this project is to establish cooperation between several teams specialising in computational linguistics based on computational models of French. The accent is on the lexicon (both syntactical and semantic aspects are considered) but a global perspective which takes into account the interface between lexicon and grammar is emphasized.

*Web site:*http://lexsynt.inria.fr

*Period:*start 2005-01-01 / end 2007-31-12

*Contact:*Claire Gardent

6.3.2. Mosaïque

*Theme:*Computational linguistics and resources for French

*Description:*The principle motivation of this ARC is to see a detailed and motivated computational grammar of French emerge that is based on a high level formalism and is freely available to the computational linguistics community.

*Web site:*http://mosaique.labri.fr

*Period:*start start 2006-01-01 / end 2007-31-12

*Contact:*Claire Gardent

6.3.3. Passage

*Theme:*Computational linguistics and resources for French

*Description:*The PASSAGE project has two main aim. The first is to improve the robustness and precision of existing computational grammars for French, and to use them on large corpora (corpora containing several million words). The second is to exploit the result syntactical analyses to create richer linguistic resources (such as Treebanks) for the French language.

*Administrative context:*ANR MDCA

*Web site:*http://atoll.inria.fr/passage/home-fr.html

*Period:*start 2007-01-01 / end 2009-31-12

*Contact:*Claire Gardent

7. Dissemination

7.1. PhD Theses

- Eric Kow defended his University of Nancy 1 (UHP) PhD thesis entitled *Surface realisation: ambiguity and determinism*, supervised by Claire Gardent, on 13 November 2007.
- Yannick Parmentier defended his University of Nancy 1 (UHP) PhD thesis entitled *SemTAG : une plate-forme pour le calcul sémantique à partir de Grammaires d'Arbres Adjoints*, supervised by Claire Gardent, on 6 April 2007.

7.2. Service to the Scientific Community

- Carlos Areces
    - Member of the Management Board of the Association of Logic, Language and Information, 2005–2008.
    - Liason officer for the *Erasmus Mundus* Masters in *Language and Communication Technology*.
- Patrick Blackburn
    - Member of the INRIA Nancy-Grand Est steering committee.
    - Member of the Management Board of the Association of Logic, Language and Information, 2005–2008.
    - Liason officer for the *Erasmus Mundus* Masters in *Language and Communication Technology*.
- Samuel Cruz-Lara
    - Samuel Cruz-Lara: Person in charge, at the national level, of the reception of Mexican students in the "Professional Licences of Computer Science".
- Christine Fay-Varnier
    - Vice president of the Council of studies and university life of the INPL.
    - Representative of the INPL for the steering committee TICE (Information and Communication Technology for Education) for Nancy University.
- Claire Gardent
    - Member of the nominating committee of the European Chapter of the Association for Computational Linguistics (EACL)

– Member of the FOLLI Editorial Board for the new series of books in Logic Language and Information to be published with Springer-Verlag as Lecture Notes in Computer Science (LNCS) and/or Lecture notes in Artificial Intelligence (LNCS/LNAI).

– Member of the ESSLLI Standing committee

– Member of the LORIA steering committee.

– Coordinator of the theme TALC (Computational Linguistics and Computational Approaches to Knowledge) for the CPER-MISN (National and Regional Research Funding).

– Organiser of the LORIA seminar on Computational Linguistics.

– Member of the recruiting committee for short term posts at INRIA Lorraine/LORIA.

- Matthieu Quignard

  – Coordinator for TEI Nancy (Text Encoding Initiative, Nancy branch) concerning spoken corpus annotation normalisation.

- Fabienne Venant

  – Member of the Administrative Council of ATALA, the French national organisation for computational linguistics (see http://www.atala.org/).

7.3. Editorial and Program Committee Work

- Carlos Areces:

  – Editor of the *Journal of Logic, Language, and Information*, 2005 – Present.

  – Editor of *Journal of Applied Logic*, 2004 – Present.

  – Member of the Program Committee of the 2007 Eurolan Summer School (EUROLAN07), Iasi, Romania.

  – Co-chair of the Program Committee of the 5th Methods for Modalities Workshop (M4M5), Chachan, France.

  – Member of the Program Committee of the 2007 International Workshop on Description Logics (DL2007), Brixen-Bressanone, Italy.

  – Member of the Program Committee of the Noveno Simposio Argentino de Inteligencia Artificial (ASAI 2007), Mar del Plata, Argentina.

  – Member of the Program Committee of the 2007 Workshop on Hybrid Logics (HyLo 07), Dublin, Ireland.

- Patrick Blackburn:

  – Chief Editor of the *Journal of Logic, Language, and Information*, 2002 – Present.

  – Editor of the *Journal of Philosophical Logic, 2004 – Present.*

  – Editor of the *Notre dame Journal of Formal Logic, 2005 – Present.*

  – Editor of the *Review of Symbolic Logic*, from 2007.

  – Subject Editor (Logic and Language) for the *Stanford Encyclopedia of Philosophy*.

  – Foreign Correspondent of *Logique et Analyse*.

- Claire Gardent

  – Member of the Program Committee for the 4th International workshop on Constraint Solving and Language Processing is held together with CONTEXT'07 in Roskilde (Denmark), August 20-21, 2007.

    – Member of the Program Committee for the Lexis & Grammar Conference, Bonifacio (France), October 2-6, 2007.

    – Member of the Program Committee for the Joint Meeting of the Conference on Empirical Methods on Natural Language Processing (EMNLP) and the Conference on Natural Language Learning (CoNLL), Prague, June 28-30, 2007,

    – Member of the Program Committee for the atelier TALN "Formalismes syntaxiques de haut niveau", TALN, Toulouse, June 2007.

    – Member of the Program Committee for the Traitement Automatique des Langues Naturelles (TALN) 2007, Toulouse, 12-15 juin 2007.

    – Member of the Program Committee for the 19th European Summer School in Logic, Language and Information (ESSLLI), Student Session - Language & Computation, Dublin (Ireland), 6-17 August, 2007.

    – Member of the Program Committee for the Journées Sémantique et Modélisation (JSM) 07, Paris, March 29-30 2007.

    – Member of the Program Committee for the 6th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC'2007), Lagos (Portugal), March 29-30, 2007.

    – Member of the Program Committee of the 2007 Eurolan Summer School (EUROLAN07), Iasi, Romania.

7.4. Invited presentations

- Carlos Areces:
  - Invited lecture, *ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*, Universidad del Pais Vasco, 28 November, 2007.

- Patrick Blackburn:
  - Invited course, *Linguistics Institute*, Stanford University, 1–27 July, 2007.
  - Invited lecture, *16th Amsterdam Colloquium*, 17 December 2007.

- Claire Gardent
  - Invited Tutorial. Semantics in NLP University of Buenos Aires, (Argentina) , November 2006. 15 hours, Postgraduate.
  - Invited Tutorial. Tree Adjoining Grammar: Theory and Practice. LAICS Summer School (Language, Artificial Intelligence and Computer Science for Natural Language Processing applications), Bangkok, (Thailand) , October 2006. 4 hours.
  - Invited Tutorial. Natural Language Generation. LAICS Summer School (Language, Artificial Intelligence and Computer Science for Natural Language Processing applications) Bangkok, (Thailand) , October 2006. 3 hours.

8. Bibliography

Major publications by the team in recent years

[1] P. BLACKBURN, J. VAN BENTHEM, F. WOLTER (editors). *Handbook of Modal Logic*, Elsevier, 2007.

[2] C. ARECES, B. TEN CATE. *Hybrid Logics*, in "Handbook of Modal Logics", P. BLACKBURN, F. WOLTER, J. VAN BENTHEM (editors), Elsevier, 2006.

[3] L. BENOTTI. *Incomplete Knowledge and Tacit Action: Enlightened Update in a Dialogue Game*, in "DECA-LOG 2007 Workshop on the Semantics and Pragmatics of Dialogue, Rovereto, Italy", 2007.

[4] P. BLACKBURN, B. TEN CATE. *Pure Extensions, Proof Rules, and Hybrid Axiomatics*, in "Studia Logica", n[o] 84, 2006, p. 277–322.

[5] T. BOLANDER, P. BLACKBURN. *Termination for Hybrid Tableaus*, in "Journal of Logic and Computation", n[o] 17, 2007, p. 517–554.

[6] D. C. A. BULTERMAN, A. J. JANSEN, P. CESAR, S. CRUZ-LARA. *An efficient, streamable text format for multimedia captions and subtitles*, in "DocEng '07: Proceedings of the 2007 ACM symposium on Document engineering, New York, NY, USA", ACM, 2007, p. 101–110.

[7] A. DENIS, G. PITEL, M. QUIGNARD, P. BLACKBURN. *Incorporating Asymmetric and Asychronous Evidence of Understanding in a Grounding Model*, in "DECALOG 2007 Workshop on the Semantics and Pragmatics of Dialogue, Rovereto, Italy", 2007.

[8] C. GARDENT, E. KOW. *A symbolic approach to near-deterministic surface realisation using Tree Adjoining Grammar*, in "Proceedings of ACL, Prague", 2007.

[9] C. GARDENT, H. MANUÉLIAN. *Création d'un corpus annoté pour le traitement des descriptions définies*, in "Traitement Automatique des Langues", vol. 46, n° 1, 2005.

[10] C. GARDENT, K. STRIEGNITZ. *Generating Bridging Definite Descriptions*, in "Computing Meaning", H. BUNT, R. MUSKENS (editors), Studies in Linguistics and Philosophy Series, vol. 3, Kluwer Academic Publishers, 2007.

Year Publications

Books and Monographs

[11] P. BLACKBURN, J. VAN BENTHEM, F. WOLTER (editors). *Handbook of Modal Logic*, Elsevier, 2007.

[12] P. BLACKBURN, J. BOS, K. STRIEGNITZ. *Prolog, tout de suite!*, College Publications, 2007.

Doctoral dissertations and Habilitation theses

[13] E. KOW. *Surface realisation: ambiguity and determinism*, Ph. D. Thesis, Université Henri Poincaré, 2007.

[14] Y. PARMENTIER. *SemTAG : une plate-forme pour le calcul sémantique à partir de Grammaires d'Arbres Adjoints*, Ph. D. Thesis, Université Henri Poincaré, 2007.

Articles in refereed journals and book chapters

[15] T. BOLANDER, P. BLACKBURN. *Termination for Hybrid Tableaus*, in "Journal of Logic and Computation", n° 17, 2007, p. 517–554.

[16] P. CESAR, K. BAKER, D. BULTERMAN, L. SOARES, S. CRUZ-LARA, A. KAPTEIN. *Technologies and applications*, in "Interactive Digital Television", G. LEKAKOS, K. CHORIANOPOULOS, G. DOUKIDIS (editors), IGI Publishing, Hershey (PA), USA, 2007, p. 91–111.

[17] D. COULON, M. MUSIOL, D. BRIXHE. *Perturbations et Réajustements Langue et Langage*, B. VAXELAIRE, R. SOCK, G. KLEIBER, F. MARSAC (editors), Du rôle des perturbations de compréhension dans la conduite des interlocutions, 2007, p. 227–236.

[18] G. DURÁN, M. C. LIN, S. MERA, J. SZWARCFITER. *Algorithms for finding clique-transversals of graphs*, in "Annals of Operations Research", in press, 2007.

[19] C. GARDENT, K. STRIEGNITZ. *Generating Bridging Definite Descriptions*, in "Computing Meaning", H. BUNT, R. MUSKENS (editors), Studies in Linguistics and Philosophy Series, vol. 3, Kluwer Academic Publishers, 2007.

[20] M. MUSIOL, A. TROGNON, D. COULON, C. BOCEREAN. *Structure de l'interaction verbale et rationalité argumentative chez l'adolescent polyhandicapé*, in "Le Langage et l'Homme", vol. XXXXI, 2007, p. 27–43.

[21] J. TSENG, A. KUPŚĆ. *A cross-linguistic approach to Slavic past tense and conditional constructions*, in "Linguistic Investigations into Formal Description of Slavic Languages, Frankfurt am Main", P. KOSTA, L. SCHÜRCKS (editors), Peter Lang, 2007, p. 427–439.

Publications in Conferences and Workshops

[22] M. AMOIA, C. GARDENT. *A first order semantic approach to adjectival inference*, in "Proceedings of ACL-Pascal Workshop on Textual Entailment and Paraphrasing, Prague, Czech Republic", 2007.

[23] C. ARECES. *Hybrid Logics: The Old and the New*, in "Proceedings of LogKCA-07", November 2007.

[24] P. BEDARIDE. *Using description logics for recognising textual entailment*, in "19th European Summer School in Logic, Language and Information - ESSLLI'07", 2007, http://hal.inria.fr/inria-00179593/fr/.

[25] A. BELAÏD, I. FALK, Y. RANGONI. *Représentation des données en XML pour l'analyse d'images de documents*, in "10 ème Colloque International sur le Document Electronique, CIDE'07, Nancy", 2007.

[26] A. BELAÏD, I. FALK, Y. RANGONI. *XML Data Representation in Document Image Analysis*, in "9th. International Conference on Document Analysis and Recognition, ICDAR'07", IEEE Computer Society Press, 2007.

[27] L. BENOTTI. *Incomplete Knowledge and Tacit Action: Enlightened Update in a Dialogue Game*, in "DECALOG 2007 Workshop on the Semantics and Pragmatics of Dialogue, Rovereto, Italy", 2007, http://hal.inria.fr/inria-00179646/fr/.

[28] P. BLACKBURN, T. BOLANDER. *Terminating Tableau Calculi for Hybrid Logics extending K*, in "Proceedings of Methods for Modalities 5", November 2007.

[29] P. BLACKBURN, S. HINDERER. *From TY_n to DRT: an implementation*, in "3rd International Language & Technology Conference - L&TC'07, Poznam Pologne", 2007, p. 384-388, http://hal.inria.fr/inria-00179297/fr/.

[30] P. BLACKBURN, S. HINDERER. *Generating models for temporal representations*, in "Recent Advances in Natural Language Processing - RANLP 2007 International conference recent advances in natural language processing, Borovets Bulgarie", 2007, p. 69-75, http://hal.inria.fr/inria-00179485/fr/.

[31] D. BULTERMAN, J. JANSEN, P. CESAR, S. CRUZ-LARA. *An Efficient, Streamable Text Format for Multimedia Captions and Subtitles*, in "ACM Symposium on Document Engineering, Winnipeg Canada", 2007, http://hal.inria.fr/inria-00192467/fr/.

[32] P. CESAR, DICK C. A. BULTERMAN, Z. OBRENOVIC, J. DUCRET, S. CRUZ-LARA. *Non-Intrusive User Interfaces for Interactive Digital Television Experiences*, in "5th European Interactive TV Conference, Amsterdam Pays-Bas", 2007, http://hal.inria.fr/inria-00192461/fr/.

[33] A. DENIS, F. BÉCHET, M. QUIGNARD. *Résolution de la référence dans des dialogues homme-machine : évaluation sur corpus de deux approches symbolique et probabiliste*, in "Traitement Automatique des Langues Naturelles - TALN 2007, Toulouse France", 2007, http://hal.inria.fr/inria-00179697/fr/.

[34] A. DENIS, G. PITEL, M. QUIGNARD, P. BLACKBURN. *Incorporating Asymmetric and Asynchronous Evidence of Understanding in a Grounding Model*, in "11th Workshop on the Semantics and Pragmatics of Dialogue - DECALOG 2007, Trento Italie", R. ARTSTEIN, L. VIEU (editors), 2007, http://hal.inria.fr/inria-00179694/fr/.

[35] I. FALK, G. FRANCOPOULO, C. GARDENT. *Evaluer SynLex*, in "Traitement Automatique des Langues Naturelles, TALN'07, Toulouse", 2007.

[36] K. FORT, B. GUILLAUME. *PrepLex : un lexique des prépositions du français pour l'analyse syntaxique*, in "Actes de Traitement Automatique des Langues Naturelles, Toulouse, France", 2007, http://hal.inria.fr/inria-00186761/en/.

[37] K. FORT, B. GUILLAUME. *PrepLex: a lexicon of French prepositions for parsing*, in "Fourth ACL-SIGSEM Workshop on Prepositions, Prague, République Tchèque", 2007, http://hal.inria.fr/inria-00186777/en/.

[38] C. GARDENT. *Tree Adjoining Grammar, Semantic Calculi and Labelling Invariant*, in "Proceedings of the International Workshop on Computational Semantics (IWCS-7), Tilburg, The Netherlands", 2007.

[39] C. GARDENT, E. KOW. *A symbolic approach to Near-Deterministic Surface Realisation using Tree Adjoining Grammar*, in "Proceedings of ACL'07, Prague", 2007, http://hal.inria.fr/inria-00149366/fr/.

[40] C. GARDENT, E. KOW. *Spotting overgeneration suspect*, in "11th European Workshop on Natural Language Generation (ENLG), Schloss Dagstuhl, Germany", 2007, http://hal.inria.fr/inria-00149372/fr/.

[41] C. GARDENT, E. KOW. *Une réalisateur de surface basé sur une grammaire réversible*, in "TALN (Poster)", 2007.

[42] C. GARDENT, Y. PARMENTIER. *SemTAG, une architecture pour le développement et l'utilisation de grammaires d'arbres adjoints à portée sémantique*, in "Conference sur le Traitement Automatique des Langues Naturelles (TALN'2007), Toulouse", 2007, http://hal.inria.fr/inria-00160393/fr/.

[43] C. GARDENT, Y. PARMENTIER. *SemTAG: a platform for specifying Tree Adjoining Grammars and performing TAG-based Semantic Construction*, in "Proceedings of the Conference of the Association for Computational Linguistics, ACL 2007, Praha", 2007, http://hal.inria.fr/inria-00160387/fr/.

[44] G. HOFFMANN, C. ARECES. *HTab: A Terminating Tableaux System for Hybrid Logic*, in "Proceedings of Methods for Modalities 5", November 2007, http://hal.inria.fr/inria-00187300/fr/.

[45] S.-H. HUNG, P.-H. CHEN, J.-S. HONG, S. CRUZ-LARA. *Context-based image retrieval: A case study in background image access for Multimedia presentations*, in "IADIS International Conference WWW/Internet 2007, Vila Real Portugal", 2007, http://hal.inria.fr/inria-00192463/fr/.

[46] C.-H. LIN, C.-W. YEN, J.-S. HONG, S. CRUZ-LARA. *Event-based textual document retrieval by using semantic role labeling and coreference resolution*, in "IADIS International Conference WWW/Internet 2007, Vila Real Portugal", 2007, http://hal.inria.fr/inria-00192465/fr/.

[47] Y. PARMENTIER, L. KALLMEYER, T. LICHTE, W. MAIER. *XMG: eXtending MetaGrammars to MCTAG*, in "Actes de l'atelier sur les formalismes syntaxiques de haut niveau, Conference sur le Traitement Automatique des Langues Naturelles (TALN'2007), Toulouse", 2007, http://hal.inria.fr/inria-00160400/fr/.

[48] B. SAGOT, K. FORT. *Améliorer un lexique syntaxique á l'aide des tables du lexique-grammaire - Adverbes en -ment.*, in "Actes du Colloque Lexique et Grammaire, Bonifacio, France", 2007, http://hal.inria.fr/inria-00186779/fr/.

[49] S. SCHMITZ. *An Experimental Ambiguity Detection Tool*, in "LDTA'07: 7th Workshop on Language Descriptions, Tools and Applications", A. SLOANE, A. JOHNSTONE (editors), Electronic Notes in Theoretical Computer Science, 2007.

[50] S. SCHMITZ. *Conservative Ambiguity Detection in Context-Free Grammars*, in "ICALP'07: 34th International Colloquium on Automata, Languages and Programming", L. ARGE, C. CACHIN, T. JURDZIŃSKI, A. TARLECKI (editors), Lecture Notes in Computer Science, vol. 4596, Springer, 2007, p. 692–703.

[51] D. SUSTRETOV, G. HOFFMANN, C. ARECES, P. BLACKBURN. *Experiments in Theorem Proving for Topological Hybrid Logic*, in "Proceedings of Methods for Modalities 5", November 2007, http://hal.inria.fr/inria-00187303/fr/.

[52] D. SUSTRETOV. *Topological semantics and decidability*, in "Proceedingo of International Workshop on Hybrid Logic (HyLo'2007)", 2007.

[53] S. THIBAULT, S. HINDERER. *BrlAPI: Simple, Portable, Concurrent, Application-level Control of Braille Terminals*, in "The First International Conference on Information and Communication Technology and Accessibility - ICTA 2007, Hammamet/Tunisia", K.: Computing Milieux/K.4: COMPUTERS AND SOCIETY/K.4.2: Social Issues/K.4.2.1: Assistive technologies for persons with disabilities, 04 2007, p. 27–31, http://hal.inria.fr/inria-00135946/fr/.

[54] J. TSENG. *English prepositional passive constructions*, in "The Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford", S. MÜLLER (editor), CSLI Publications, 2007, p. 271–286.

[55] J. TSENG. *L'abstraction de l'extraction*, in "TALN '07: Actes de la 14e conférence sur le Traitement Automatiques des Langues Naturelles, Toulouse", N. HATHOUT, P. MULLER (editors), IRIT Press, 2007, p. 463–472.

[56] F. VENANT. *La construction du sens: un système dynamique complexe*, in "ARCO'07", 2007.

References in notes

[57] M. AMOIA, C. GARDENT, S. THATER. *Using set constraints to generate distinguishing descriptions*, in "Proc. of the 7th International Workshop on Natural Language Understanding and Logic Programming - NLULP'02, Copenhagen, Denmark", Jul 2002, http://www.loria.fr/publications/2002/A02-R-459/A02-R-459.ps.

[58] C. ARECES, R. BERNARDI. *Analyzing the Core of Categorial Grammar*, in "Journal of Logic, Language and Information", vol. 13, n° 2, 2004, p. 121–137.

[59] C. ARECES, P. BLACKBURN. *Reichenbach, Prior and Montague: A semantic get-together*, in "Articles in Honor of Dov Gabbay's 60th Birthday", Kings College Press, 2005.

[60] C. ARECES, D. GORÍN. *Ordered Resolution with Selection for H(@)*, in "Proc. of the 11th International Conference on Logic for Programming Artificial Intelligence and Reasoning - LPAR 2004, Montevideo, Uruguay", F. BAADER, A. VORONKOV (editors), Lecture Notes in Computer Science, vol. 3452, Springer, Mar 2005, p. 125-141.

[61] C. ARECES, J. HEGUIABEHERE. *hGen: A Random CNF Formula Generator for Hybrid Languages*, in "Proc. of Methods for Modalities 3 - M4M-3, Nancy, France", Sep 2003, http://www.loria.fr/publications/2003/A03-R-306/A03-R-306.ps.

[62] P. BLACKBURN, J. BOS. *Computational Semantics*, in "Theoria", vol. 18, n° 46, Jan 2003, p. 27-45.

[63] P. BLACKBURN, J. BOS. *Representation and Inference for Natural Language: A First Course in Computational Semantics*, CSLI Press, Aug 2005.

[64] B. CRABBÉ, D. DUCHIER. *Metagrammar Redux*, in "Proc. of the International Workshop on Constraint Solving and Language Processing - CSLP 2004, Copenhagen, Norway", Sep 2004.

[65] B. CRABBÉ. *Alternations, monotonicity and the lexicon: an application to factorising information in a Tree Adjoining Grammar*, in "Proc. of the European Summer School of Logic Language and Computation, Student Session - ESSLLI'03, Vienna, Austria", Aug 2003, p. 69-80.

[66] B. CRABBÉ. *Lexical Classes for structuring the lexicon of a TAG*, in "Lorraine-Saarland Workshop series: prospects and advances in the syntax semantics interface, Nancy, France", Oct 2003, http://www.loria.fr/publications/2003/A03-R-396/A03-R-396.ps.

[67] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Représentation et gestion de grammaires d'arbres adjoints lexicalisées*, in "Traitement Automatique des Langues", vol. 44, n° 3, Dec 2003, p. 67-91.

[68] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Une plateforme de conception et d'exploitation de grammaire d'arbres adjoints lexicalisés*, in "Traitement Automatique du Langage Naturel 2003 - TALN 2003, Batz-sur-Mer, France", Jun 2003.

[69] D. DUCHIER, J. LE ROUX, Y. PARMENTIER. *The MetaGrammar Compiler: An NLP Application with a Multi-paradigm Architecture*, in "Proc. of the 2nd International Mozart/Oz Conference - MOZ 2004, Charleroi, Belgium", Oct 2004.

[70] B. GAIFFE, B. CRABBÉ, A. ROUSSANALY. *A New Metagrammar Compiler*, in "Proc. of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks - TAG+6, Venice, Italy", May 2002, http://www.loria.fr/publications/2002/A02-R-495/A02-R-495.ps.

[71] C. GARDENT. *Generating minimal definite descriptions*, in "Proc. of the 40th Annual Meeting of the Association for Computational Linguistic - ACL'02, Philadelphia, USA", Jul 2002, http://www.loria.fr/publications/2002/A02-R-460/A02-R-460.ps.

[72] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Extracting a subcategorisation lexicon from Maurice Gross Grammar Lexicon*, in "Archive of Control Sciences", 2005.

[73] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Extracting subcategorisation information from Maurice Gross' Grammar Lexicon*, in "Archives of Control Sciences", vol. 15, nᵒ LI, 2005, p. 253-264.

[74] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Maurice Gross' Grammar Lexicon and Natural Language Processing*, in "Proc. of the 2nd Language and Technology Conference - L&T'05, Poznan, Poland", Apr 2005.

[75] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Maurice Gross' Grammar Lexicon and Natural Language Processing*, in "Proceedings of the 2nd Language and Technology Conference, Poznan, Poland", 2005.

[76] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Extraction d'information de sous-catégorisation à partir des tables du LADL*, in "Actes de La 13ème édition de la conférence sur le Traitement Automatique des Langues Naturelles (TALN 2006)", 2006.

[77] C. GARDENT, E. KOW. *Generating and selecting paraphrases*, in "Proc. of the 10th European Workshop on Natural Language Generation - ENLG 05, Aberdeen, Scotland", Aug 2005, p. 191-196.

[78] C. GARDENT, H. MANUÉLIAN. *Création d'un corpus annoté pour le traitement des descriptions définies*, in "Traitement automatique des langues", vol. 46, nᵒ 1, 2005.

[79] C. GARDENT, H. MANUÉLIAN, E. KOW. *Which bridges for bridging definite descriptions?*, in "Proc. of the 4th International Workshop on Linguistically Interpreted Corpora - LINC'03, Budapest, Hungary", Apr 2003, http://www.loria.fr/publications/2003/A03-R-067/A03-R-067.ps.

[80] C. GARDENT, H. MANUÉLIAN, K. STRIEGNITZ, M. AMOIA. *Generating Definite Descriptions: Non incrementality, inference and data*, in "Speech production", Walter de Gruyter, Berlin, Dec 2003.

[81] C. GARDENT, Y. PARMENTIER. *Large scale semantic construction for Tree Adjoining Grammars*, in "Proc. of Logical Aspects of Computational Linguistics - LACL'05, Bordeaux, France", P. BLACHE, E. STABLER, J. BUSQUETS, R. MOOT (editors), Lecture Notes in Computer Science, vol. 3492, Springer, Apr 2005, p. 131-146.

[82] C. GARDENT, K. STRIEGNITZ. *Generating Bridging Definite Descriptions*, in "Computing Meaning", H. BUNT, R. MUSKENS (editors), Studies in Linguistics and Philosophy Series, vol. 3, Kluwer Academic Publishers, Dec 2003.

[83] E. KOW. *Adapting polarised disambiguation to surface realisation*, in "Proc. of the 17th European Summer School in Logic, Language and Information - ESSLLI'05, Edinburgh, United Kingdom", Aug 2005.

[84] H. MANUÉLIAN. *Generating Coreferential Descriptions from a Structured Model of the Context*, in "Proc. of the International Conference on Language Ressources and Evaluation - LREC 2004, Lisbon, Portugal", May 2004, http://www.loria.fr/publications/2004/A04-R-060/A04-R-060.ps.

[85] H. MANUÉLIAN. *Annotation des descriptions définies : le cas des reprises par les rôles thématiques*, in "Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues - RECITAL'2002, Nancy, France", Jun 2002, http://www.loria.fr/publications/2002/A02-R-076/A02-R-076.ps.

[86] H. MANUÉLIAN. *Coreferential Definite and Demonstrative Descriptions in French: A Corpus Study for Text Generation*, in "Proc. of the ESSLLI'03 Student Session, Vienna, Austria", Aug 2003, http://www.loria.fr/publications/2003/A03-R-108/A03-R-108.ps.

[87] H. MANUÉLIAN. *Descriptions Définies et Démonstratives : Analyses de corpus pour la génération*, Ph. D. Thesis, Université de Nancy 2, Nov 2003, http://www.loria.fr/publications/2003/A03-T-527/A03-T-527.ps.

[88] H. MANUÉLIAN. *Génération de descriptions définies et démonstratives*, in "Huitième Atelier des doctorants en linguistique - ADL'2003, Paris, France", Jul 2003.

[89] H. MANUÉLIAN. *Une analyse des emplois du démonstratif en corpus*, in "Traitement Automatique des Langues Naturelles - TALN 2003, Batz sur Mer, France", Jun 2003, http://www.loria.fr/publications/2003/A03-R-107/A03-R-107.ps.

[90] Y. PARMENTIER, J. LE ROUX. *XMG: a Multi-formalism Metagrammatical Framework*, in "Proc. of the 17th European Summer School in Logic, Language and Information - ESSLLI '05, Edinburgh, United Kingdom", Aug 2005.

[91] D. SEDDAH, B. GAIFFE. *Des arbres de dérivation aux forêts de dépendance : un chemin via les forêts partagées*, in "Traitement automatique des langues Naturelles - TALN'05, Dourdan, France", Jun 2005.

[92] D. SEDDAH, B. GAIFFE. *How to Build Argumental graphs Using TAG Shared Forest: a view from control verbs problematic*, in "Proc. of the 5th International Conference on the Logical Aspect of Computional Linguistic - LACL'05, Bordeaux, France", Apr 2005.

[93] D. SEDDAH, B. GAIFFE. *Using both Derivation tree and Derived tree to get dependency graph in derivation forest*, in "Proc. of the 6th International Workshop on Computational Semantics - IWCS-6, Tilburg, The Netherlands", Jan 2005.

[94] D. SEDDAH. *Synchronisation des connaissances syntaxiques et sémantiques pour l'analyse d'énoncés en langage naturel à l'aide des grammaires d'arbres adjoints lexicalisées*, Ph. D. Thesis, Université Henri Poincaré - Nancy 1, Nov 2004.

[95] K. STRIEGNITZ. *Génération d'expressions anaphoriques. Raisonnement contextuel et planification de phrases*, Ph. D. Thesis, Université Henri Poincaré, Nov 2004.

[96] J. VAN EIJCK. *HyLoTab — Tableau-based Theorem Proving for Hybrid Logics*, 2002, http://www.cwi.nl/~jve/hylotab.