INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TALARIS

Traitement Automatique des Langues: Représentations, Inférences et Sémantique

Nancy - Grand Est

Theme : Audio, Speech, and Language Processing

Table of contents

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/.

1. Team

Research Scientist

Patrick Blackburn [Team Leader, Research Director (DR2), INRIA, HdR]
Claire Gardent [Research Director (DR2), SHS Department CNRS, HdR]
Carlos Areces [Research Scientist (CR1), INRIA]
Matthieu Quignard [Research Scientist (CR2), SHS Department CNRS. Finished 2009-04-01]

Faculty Member

Lotfi Bellalem [PRAG, ESIAL, UHP Nancy 1]
Nadia Bellalem [Assistant Professor, IUT Nancy-Charlemagne, 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]
Jean-Charles Lamirel [Assistant Professor, School of Geology, University of Strasbourg]
Fabienne Venant [Assistant Professor, IUT Nancy-Charlemagne, University of Nancy 2]

Technical Staff

Mathieu Desnouveaux [INRIA Engineer, 2008-09-01 until 2010-09-01]
Tarik Osswald [Engineer, ITEA Project MetaVerse]

PhD Student

Paul Bedaride [UHP Nancy 1, Ministry grant, from 2006-10-15 until 2010-10-15]
Luciana Benotti [UHP Nancy 1, CORDI grant, from 2006-06-10 until 2010-01-28]
Ingrid Falk [University of Nancy 2, SEMbySEM project, from 2008-10-01 until 2011-10-01]
Daniel Gorin [UHP Nancy 1, Co-tutelle Universidad de Buenos Aires. Finished 2009-31-09]
Guillaume Hoffmann [UHP Nancy 1, Ministry grant, from 2007-10-15 until 2010-11-15]
Sergio Mera [UHP Nancy 1, Co-tutelle Universidad de Buenos Aires. Finished 2009-31-09]
Laura Perez [[University of Nancy 2, Ministry grant, from 2009-10-01 until 2012-10-01]
Dmitry Sustretov [UHP Nancy 1, INRIA grant, from 2005-10-01 until 2010-04-01]

Post-Doctoral Fellow

Marilisa Amoia [2009-09-01 until 2010-12-31]
Lina Maria Rojas Barahona [2009-06-15 until 2009-11-30]
Alexandre Denis [2009-04-01 until 2010-12-31]

Visiting Scientist

Corinna Anderson [Yale University, PhD student until 2010-09-30]

Administrative Assistant

Isabelle Blanchard

2. Overall Objectives

2.1. Background

TALARIS stands for *Traitment 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. This 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 processing, 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.
- Integrating language technology and semantic resources into multimedia applications.

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

There were five main highlights in 2009. First, Daniel Gorin and Sergio Mera successfully defended their PhD theses, supervised by Carlos Areces and Patrick Blackburn. These were both co-tutelle theses (between the UHP Nancy 1 and the University of Buenos Aires in Argentina). The two defenses were held, one after the other, in Buenos Aires, on 9th December 2009, with examiners from France, Argentina and Germany; both theses were awarded the highest grade that the University of Buenos Aires can bestow. A second highlight was the completion of Luciana Benotti's PhD thesis, late in December 2009. As the thesis will not be defended until 2010, it is not listed in the bibliography, but it is safe to say it is an excellent thesis, and one that draws together the various research strands in TALARIS. Thirdly, the SEMbySEM project got into its stride, as did the new METAVERSE project; both are European funded ITEA projects led by Samuel Cruz-Lara. Apart from their intrinsic interest, both projects are playing an important role in TALARIS by providing a focus around which a number of activities are gelling. Fourth, a new theme is emerging headed by Claire Gardent which focuses on the integration of deep natural language processing (NLP) techniques with 3D worlds and games. A first foray into that theme has been the preparation of two systems for submission to the GIVE challenge. Two recently accepted EU funded projects (Interreg project Allegro and Eurostar project Emo-Speech) on that theme will permit a fully blown exploration of the research issues and of the technological problems arising in this area. This new theme builds on the tools and techniques developped by Talaris over the last 5 years for deep NLP and in particular, on the availability of an expressive grammar writing environnement (XMG), of wide coverage deep grammars for French and English (SemTAG and SemXTAG), of a grammar based surface realiser (GenI) and of parsers (LLP2, SemConst) using these grammars. Fifth, and finally, the

3 day Natal workshop, organized by Claire Gardent, which gathers together students and researchers from Nancy, Saarbrúcken and neighbouring areas around current themes in NLP themes, was successfully held (for the second running) from 17–19 June 2009. As in 2008 there were two thematic days (this year, Syntactic Lexicons and Verb Classes) and a third Masters Student Day where Erasmus Mundus students form Nancy and Saarbrúcken could meet and present their research.

3. Scientific Foundations

3.1. Computational Linguistics and Computational Logic

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.

3.2. Semantics and Inference

Over the next decade, progress in natural language semantics is likely to 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 recognition 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.

3.3. 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 and 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.

3.4. 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 useful 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.

3.5. 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.

Such work is important to our scientific goals. 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.

4. Application Domains

4.1. 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].

4.2. 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.

[1] [39], [41], [40], [42], [43], [44], [45], [66], [63], [64], [65], [53], [62]
[2] [56], [60], [57], [59], [58], [61], [50], [51]
[3] [67], [52], [54], [35]
[4] [46], [35]

4.3. 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.

4.4. Textual Entailment Recognition

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.

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.

4.5. Computational Logics and Computational Semantics.

Members of TALARIS have long actively proposed and developed 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 Patrick Blackburn and Johan Bos is nowadays an important reference in this area.

4.6. Hybrid Automated Deduction

TALARIS's main contribution in this topic has been the design of resolution and tableaux calculi for hybrid logics, calculi that were then implemented in the HyLoRes and HTab theorem provers. For example, TALARIS members have proved that the resolution calculus for hybrid logics can be enhanced with optimisations of order and selection functions without losing completeness. Moreover, a number of 'effective' (i.e., directly implementable) termination proofs for the hybrid logic $\mathcal{H}(@)$ has been established, for both resolution and tableaux based approaches, and the techniques are being extended to more expressive languages. Current work includes adding a temporal reasoning component to the provers, extending the architecture to allow querying against a background theory without having to explore again the theory with each new query, and testing the hybrid provers performance against dedicated state-of-the-art provers from other domains (firs-order logic, description logics) using suitable translations.

Moreover, we are interested in providing a range of inference services beyond satisfiability checking. For example, the current version of HyLoRes and HTab includes model generation (i.e., the provers can generate a model when the input formula is satisfiable).

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.

[5] [55], [49]
[6] [36], [37]
[7] [38]

4.7. Multimedia

MLIF (Multi Lingual Information Framework) is intended to be a generic ISO-based mechanism for representing and dealing with multilingual textual information. A preliminary version of MLIF has been associated with 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.

5. Software

5.1. AGREE

AGREE (Asynchronous Grounding of REferential Expressions) is a set of modules that manage the grounding process at the reference level. It contains an interpretation evaluation module that construes understanding judgments made by the system and those manifested in the dialogue by the user, a dialogue module that maintains a coherent state of the dialogue (adjacency pairs), and a generation module (GenI) in order to produce paraphrases of the understood referents. The whole system has been implemented in Java and uses the same semantic/referential representation that was used in the MEDIA project.

Version: 0.1
License: GPL
Last update: 2008-11-12
Web site: http://www.loria.fr/~denis/grounding.html
Authors: Alexandre Denis
Contact: Alexandre Denis

5.2. 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 Bertrand 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.

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 will deal with 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

5.3. 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 and Java, 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 Tulepa parser. 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 is intended to serve as a laboratory in order to test pragmatic theories about the phenomenon of accommodation. It is also result in the first integrated system to use SEMTAG (the LORIA toolbox for TAG-based Parsing and Generation).

Version: 1.0
License: GPL
Last update: 2008-11-07
Authors: Luciana Benotti, Alejandra Lorenzo, Laura Perez
Contact: Luciana Benotti

5.4. GenI surface realiser

The GENI surface realiser is a successor of the InDiGen realiser. Also based on a chart algorithm, it is implemented in Haskell 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.

GENI is compatible with both a grammar for French (SEMTAG) and for English (SEMXTAG), both grammars beeing produced using the MetaGrammar Compiler. SEMTAG covers the basic syntactic structures of French as described in Anne Abeillé's book "An Electronic Grammar for French". SEMXTAG has a coverage similar to that of XTAG, the TAG grammar for English developped by the University of Pennsylvannia . Both grammars are additionnally equiped with a compositional semantics supporting semantic construction (during parsing) and/or surface realisation.

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

Version: 0.20.2
License: GPL
Last update: 2009-11-16

Web site: http://projects.haskell.org/GenI
Project(s): GENI
Authors: Carlos Areces, Claire Gardent, Eric Kow
Contact: Claire Gardent

5.5. 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: 2.5
License: GPL
Last update: 2009-04-09
Web site: http://www.glyc.dc.uba.ar/intohylo/hylores.php
Authors: Carlos Areces, Daniel Gorín and Juan Heguiabehere
Contact: Carlos Areces

5.6. 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 efficient for unsatisfiable formulas (because a contradiction can be reported as soon as the empty clause is derived), tableaux methods are better suited to handle satisfiable formulas (because a saturated open branch in the tableaux represents a model for the input formula).

Version: 1.5.2
License: GPL
Last update: 2009-11-10
Web site: http://www.glyc.dc.uba.ar/intohylo/htab.php
Authors: Carlos Areces, Guillaume Hoffmann
Contact: Guillaume Hoffmann

5.7. 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 inefficient.

Version: 0.2
License: GPL
Last update: 2009-10-29
Web site: http://www.glyc.dc.uba.ar/intohylo/hyloban.php
Authors: Carlos Areces, Guillaume Hoffmann, Dmitry Sustretov
Contact: Guillaume Hoffmann

5.8. 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: 1.2
License: GPL
Last update: 2009-06-17
Web site: http://www.glyc.dc.uba.ar/intohylo/hgen.php
Authors: Carlos Areces, Daniel Gorín, Juan Heguiabehere and Guillaume Hoffmann
Contact: Carlos Areces

5.9. 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[8]. It is currently being validated using the Sylva web service and will be made available in 2009.

Last update: 2005-10-14
Web site: Not yet available
Project(s): SynLex
Authors: Claire Gardent, Guy Perrier, Bruno Guillaume, Ingrid Falk
Contact: Claire Gardent

[8] [48], [47]

5.10. 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 LTAG parser (LLP2), a LTAG grammar, an OWL ontology and a set of rules in description logic, and works together with a reasoner such as RACER. The current version contains a reference resolution module (anaphora and deixis) which is based on the referential domains theory. The package also contains ways to project the semantic form (referentially solved) into the MEDIA formalism and to evaluate the accuracy of the representation using a test corpus. The whole system has been implemented in Java and communicates with other modules using TCP/IP.

Version: 0.5
License: GPL
Website: http://www.loria.fr/~denis/media.html
Last update: 12/11/2008
Project(s): MEDIA
Authors and Contact: Alexandre Denis

5.11. Nessie

Nessie is a semantic construction tool written in OCaml. It takes a lexicon and a syntax tree as input and produces a semantic representation taking the form of a simply typed lambda term. Simply typed lambda calculus is used not only as the target language, but also as the glue language for assembling the representations provided by the lexicon.

This tool has been successfully used in several applications, the most notable of which being the computation of discourse semantics according to two different theories, namely the compositional DRT (Muskens 95) and the compositional treatment of dynamicity (de Groote 2006).

Future developments of Nessie may include using richer typing systems, and interfacing it with inference and rewriting tools to simplify the representations it produces.

Last update: 2008-11-14
Authors: Sébastien Hinderer
Contact: Sébastien Hinderer

5.12. 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[9].

Authors: Claire Gardent, Hélène Manuelian
Web site: Distributed by the CNRTL http://www.cnrtl.fr/
Contact: Claire Gardent

5.13. SemFRaG

A TAG grammar developed with the XMG metagrammar compiler and which describes both the syntax and the semantics of a core fragment coverage of French. Syntactically, the grammars covers the constructions described in A. Abeillé 's book. Additionnally, it is equipped with a unification based compositional semantics which supports both semantic construction (using LLP2, Tulipa or SemCONST) and surface realisation (using GenI).

Authors: Claire Gardent, Benoit Crabbé

[9] [50]

Contact: Claire Gardent

5.14. SemXTAG

A TAG grammar for English developed with the XMG metagrammar compiler and which describes both the syntax and the semantics of English. Syntactically, the grammar has a coverage comparable to that of the XTAG grammar developed by the University of Pennsylvania. Additionnally, the grammar integrates a unification based compositional semantics. Used both for parsing (by LLP2 and SemCONST) and for generation (by GenI).

Authors: Claire Gardent, Katya Alahverdzhieva
Contact: Claire Gardent

6. New Results

6.1. New Results

One of the most pleasing aspects of the work in 2009 is that there has been a heavy convergence between the different parts of TALARIS. In previous years, it was fairly easy to distinguish the various lines of work; now, one year later, it is harder. This is an excellent shift, as it shows we are beginning to achieve the desired convergence.

Much of this convergence has been project driven. For example, the CCCP-Prosodie project is an ANR project intended to study online communities of practice using a multi-disciplinary approach. It gathers together ergonomists, psychologists, sociologists, lawyers, economists and computer scientists, and the scientific goal is to understand online communities of practice: that is, to study their mechanisms, their success or failure, and their interactions with other actors of their discipline (companies, institutions, etc.). The goal of the TALARIS team in this project is to analyze the participants interactive profile within a given community of practice, namely the astronomy project in Wikipedia. We are studying the contributions of participants in the Wikipedia discussion pages, and the relations between these contributions. However, previous approach (such as deep parsing and inference used in the MEDIA project) made the automatic analysis difficult because of the semi-finalized aspect of Wikipedia discussion pages. Therefore, we are currently investigating new analysis techniques to classify the participants' contributions, and thus, the participants' interactive profile.

As a second example, the PORT-MEDIA project is an ANR project continuing the MEDIA project in which we participated in an evaluation the comprehension abilities of dialogue systems. In particular, the goal was to compare the dialogue systems of several French laboratories (LIMSI, LIA, VALORIA, LORIA) along the semantic and referring dimensions. The technique of evaluation consisted of comparing the semantic/referring formal representation of each system produced for a given utterance, with a gold standard, annotated by a human in the MEDIA formalism of the same utterance. However the MEDIA formalism was too limited to represent correctly the semantic representations we want to have in dialogue systems. Moreover, the semantic forms produced by the TALARIS system interested other participants. This lead to the TALARIS team in PORT-MEDIA consists providing a new version of the evaluation corpus annotated with our own generic formalism MMIL (MultiModal Interface Language). The task is two-fold: first defining a clear version of MMIL compliant with TEI (Text Encoding Initiative, an ISO working group), and then providing an annotated corpus. We are currently investigating new statistical parsing techniques and semantic building in order to provide the MMIL annotation of the MEDIA corpus.

But perhaps the best example of scientific collaboration between team members with different interests is provided by the Nancy-GIVE project. The GIVE project is a project lead by the university of the Saarland, in Saarbrücken, Germany, which aims to evaluate the generation abilities of dialogue system. In particular, it aims to evaluate the precision and correctness of instructions that are given by systems in virtual environments. The new version of GIVE, GIVE2, does this by evaluating the instructions within a maze where the user has to push buttons to reach a trophy while avoiding alarms. Two aspects are evaluated: the navigational part, and the referring part of the systems. Our participation in this project consists in testing our generation systems in such environment. Led by Claire Gardent, the Nancy-GIVE project developed two systems that will be submitted to the GIVE challenge in 2010. The first system developed by Alexandre Denis (Postdoc on the CCCP project) provides the most direct instructions to the user, while the second, developed by Marilisa Amoia (Postdoc co-funded by the ITEA Metaverse project and the Lorraine Region), aims to provide less direct but more friendly instructions. Both systems will then be compared to each other, and also with the systems of other participants in the GIVE challenge.

One individual effort is also worth noting. 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.

Giving an explicit computational model of part of this process is the underlying idea of Luciana Benotti's PhD thesis, which has been completed and which will be defended on 28th January 2010. 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. An implementation, called Frolog, of these ideas has been completed as part of her thesis work.

TALARIS work on the more logical aspects underlying inference was also extensively developed in several directions in 2009. On the one hand, HyLoRes, the resolution based theorem prover for hybrid logics, has finally arrived to a very stable stage of development. It has a modular architecture, it has been extensively tested, and it has a graphical interface. The main focus was to develop it to use parallel resolution algorithms.

On the other hand we have included a number of important optimizations into HTab, the tableaux based prover for hybrid logics. Actually, preliminary tests show that HTab outperforms HyLoRes, and the state of the art description logic provers FaCT and Pellet when tested over hybrid logic formulas.

Another important aim during this year was the development of inference services beyond satisfiability checking. Both HyLoRes and HTab currently includes model generation (i.e., they can generate a model when the input formula is satisfiable), and HTab includes instance retrieval (i.e., given a background theory, it can list the nominals which make certain formula true, and all atomic formulas true at a nominal, in all models of the theory).

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 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, bisimulation checking and instance retrieval. Initially, InToHyLo was 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 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, the tableau method performs better on satisfiable formulas.

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 some description logic provers 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 description logic provers: 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.

Exploring the effect of different optimizations in the behaviour of the theorem provers developed by the team is a complex task. One tool which is extensively used in the field is to use a random formula generator that can generate formulas according to different parameters that can tune the expected complexity of determining whether the formula is satisfiable or not. These tests provide extremely useful statistical data about the behaviour of the provers over different kinds of formulas. The main drawback of such tests is that they are very time consuming. To be statistically relevant, the provers have to be ran over thousands of formulas, and when each single formula becomes non trivial (say taking a couple of minutes of CPU time), running a single test might take over a week.

To solve this problem we developed a distributed testing framework that takes advantage of the Grid5000 (https://www.grid5000.fr). As every single run of a prover on a random formula is independent of each other, we can farm out each individual test to multiple machines in the grid obtaining a linear speed up (on the number of grid nodes used). By using this technique, testing time has been greatly reduced. This framework was presented at the Description Logic Workshop in Oxford, UK, this year.

This work has been done in the framework of the project PARLO, from the MISN TALC

When we are interested languages as tools for modeling behaviour, it is natural to look for extensions which are able to capture some notion of state. Good examples of such logics are the different epistemic logics with dynamic operators (often called Dynamic Epistemic Logics), which model the evolution of knowledge by accessing and changing the model structure through logic operators. Many other examples exist in the literature: update logics, XCTL, the freeze operator, etc.

We have investigated a new family of logics, which we call memory logics, to capture some common features which are shared by logics like the ones we just mentioned. The aim is to define a general framework where we can study how to add explicit state to a model, and how to access and modify it via logical operators. In a number of publications during this year we have investigated the expressive power of memory logics, the complexity of their satisfiability problem, and we have developed tableaux and model checking algorithms.

He have investigated in detail the expressive power of different members of this family, provided complete axiomatizations, and also tableaux calculi. These results have been presented at international conferences.

7. Other Grants and Activities

7.1. Introduction

The following section lists the projects and collaborations that TALARIS is involved in at international, European, national, and local levels. Each is classified according to which of the four major TALARIS themes (that is: Computational Semantics; Discourse, Dialogue and Pragmatics; Logics for Natural Language and Knowledge Representation; and Multilinguality for Multimedia) it mainly contributes towards. In some cases the work contributes to several themes.

7.2. International level

7.2.1. InSeDiSy: Inference Services for Dialogue Systems

Theme: Logics for Natural Language and Knowledge Representation; Discourse, Dialogue and Pragmatics

Description: This is a small pilot project carried out with the University of Buenos Aires to investigate the use of inference services for dialogue systems. The TALARIS team is involved because of its work in both areas.

Administrative context: MINCYT (Argentinian Minstery of Science and Technology), CNRS, INRIA

Period: start 2008-01 / end 2010-01

Contact: Carlos Areces

Partner(s): University of Buenos Aires

7.2.2. InToHyLo: Inference Tools for Hybrid Logics

Theme: Logics for Natural Language and Knowledge Representation

Description: The main aim of the InToHyLo project is to investigate inference methods for hybrid logics, to develop highly optimized inference tools based on these methods, and to use these tools in natural language applications.

Talaris and GLyC are currently leaders in automated theorem proving for hybrid logics, and they are the developers of the two provers HyLoRes (based on resolution) and HTab (based on tableaux). With the InToHyLo project we want to investigate how to combine resolution and tableaux algorithms to allow our provers to collaborate and share partial results. We will integrate our tools in a platform suitable for inference in NLP applications (focusing on Dialogue Systems and Textual Entailment). This platform will include not only tools for satisfiability testing, but also for model building, model checking, bisimulation checking, and knowledge maintenance and retrieval. Finally, we want to develop parallel inference algorithms to improve performance, and distributed testing to speed up developing.
Administrative context: INRIA (Equipes Associées)

Period: start 2009-01 / end 2012-01

Contact: Carlos Areces

Partner(s): University of Buenos Aires

7.3. European Level

7.3.1. METAVERSE

Theme: Multilinguality for Multimedia

Description: Metaverse is an exciting project whose goal is to provide a standardized global framework enabling the interoperability between virtual worlds (for example Second Life, World of Warcraft, IMVU, Active Worlds, Google Earth and many others) and the Real world (sensors, actuators, vision and rendering, social and welfare systems, banking, insurance, travel, real estate and many others).

Administrative context: ITEA2 07016

Web site: http://www.metaverse1.org/

Period: start 2009-01-01 / end 2011-12-31

Contact: Samuel Cruz-Lara

Partner(s): Belgian partners: Alcatel-Lucent Bell N.V., Nazooka, IBBT-SMIT; French partners: Alcatel-Lucent France, Orange Labs, CEA List, Artefacto; Greek partners: Forthnew S.A., Ellinogermanki Agogi; Dutch partners: Philips Research, Philips I-Lab, DevLab, Technical University Eindhoven, University of Twente, Stg. EPN, VU Economics & BA, VU CAMeRA; Spanish partners: Innovalia, Ceeda, VirtualWare, CBT, Nextel, Corsa, Avantalia, I&IMS, VicomTECH, E-PYME, CIC Tour Game, UPF-MTH; Israeli partners: Metaverse Labs.

7.3.2. SEMbySEM: Services management by Semantics

Theme: Multilinguality for Multimedia

Description: The goal of the SEMbySEM project is to develop a new open source supervision system adapted to the increasing complexity of "systems of systems". This new supervisions system will be based on the extensive use of semantic technologies (notably ontologies). It will provide a set of tools allowing the set up of dedicated supervision systems according to the various stakeholders' needs and domain knowledge.

The TALARIS team's contribution to this project will center on providing language technology for developing, maintaining, and enriching ontologies and on developing ISO standards for multilingual user interfaces.

Administrative context: ITEA2 07021

Web site: http://www.sembysem.org/

Period: start 2008-07-31 / end 2010-12-31

Contact: Samuel Cruz-Lara

Partner(s): Finnish partners: Identoi, LogiNets, Oliotalo, VTT; French partners: Thales (Project Leader), ArcInformatique, CityPassenger, LISSI (Université de Paris 12), LIG (IMAG GRenoble); Spanish partners: Trimek, DataPixel, SQS, CBT, Innovalia; Turkish partners: AGM Lab, METU.

7.4. National Level

7.4.1. CCCP-Prosodie

Theme: Discourse, Dialogue and Pragmatics; Logics for Natural Language and Knowledge Representation

Description: The goal of CCCP-Prosodie is to empirically investigate the functioning of online communities (such as Wikipedia), and particular to link their activities and their use of language (as recorded in such corpora as email exchanges, for example). The TALARIS team is involved in this project for three reasons: to provide Natural language processing tools, to design an annotation scheme capable of dealing with information from both the social sciences (sociology and economics) and the humanities (psychology and ergonomics), and to provide help with inference technology.

Administrative context: ANR CONTINT

Web site: http://recherche.telecom-bretagne.eu/labo_communicant/cccp-prosodie/

Period: start 2008-01-12 / end 2011-31-06

Contact: Alexandre Denis

Partner(s): Institut Télécom, UTC Compiégne, UNSA (Univ. Nice Sophia-Antipolis), Univ. de Versailles St-Quentin

7.4.2. PORT-MEDIA

Theme: Discourse, Dialogue and Pragmatics; Logics for Natural Language and Knowledge Representation

Description: The PORT-MEDIA project is an ANR project continuing the MEDIA project in which TALARIS participated in an evaluation of understanding abilities of dialogue systems. The goal was to compare the dialogue systems of several French laboratories (LIMSI, LIA, VALORIA, LORIA) along the semantic and referential dimensions.

Administrative context: ANR CONTINT

Web site: http://www.port-media.org/doku.php?id=start

Period: start 2009-03-01 / end 2012-03-01

Contact: Alexandre Denis

Partner(s): ELDA, LIG/GETALP, LIA, LIUM, LORIA

7.4.3. Passage

Theme: Computational Semantics

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 analyzes 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 2010-30-06

Contact: Claire Gardent

7.4.4. Rhapsodie

Theme: Discourse, Dialogue and Pragmatics; Computational Semantics

Description: The goal of this project is to provide a reference corpus for French which links syntactic and prosodic annotation. TALARIS is involved because of its experience in this area, and also to provide help on normalisation and integration with the TEI initiative.

Administrative context: ANR Corpus SHS

Web site: http://rhapsodie.risc.cnrs.fr/fr/index.html

Period: start 2008-01-01 / end 2011-31-06

Contact: Mathieu Quignard

Partner(s): Modyco, Ircam, Lattice, ERSS, LPL

7.4.5. Rhapsodis

Theme: Discourse, Dialogue and Pragmatics; Computational Semantics

Description: The aim of this project is to improve speech recognition performance by incorporating semantic and syntactic information into the recognition process. The TALARIS team is aims to provide suitable semantic and syntactic input.

Administrative context: ARC INRIA

Web site: http://rapsodis.loria.fr

Period: start 2008-01-01 / end 2009-31-12

Contact: Claire Gardent

Partner(s): ATILF, PAROLE, IRISA (Texmex), CEA (Metiss)

7.5. Local Level

7.5.1. Align

Theme: Discourse, dialogue and pragmatics

Description: The Align project deals with the problem of semi-automatically aligning written transcription of speech with the sound signal. In addition, it addresses the problems raised by anonymisation (suppressing proper names, for example, to ensure anonymity of the participants).

Administrative context: CPER MISN-TALC

Web site: http://www.loria.fr/~cerisara/TALC/index.html

Period: 2008-01-01 / end 2009-12-3

Contact: Matthieu Quignard

Partner(s): ATILF, PAROLE

8. Dissemination

8.1. PhD Theses

- Daniel Gorín defended his PhD thesis in cotutelle between the Université Henri Poincare and the Universidad de Buenos Aires entitled *Automated Reasoning Techniques for Hybrid Logics*, supervised by Patrick Blackburn and Verónica Becher, on 9 December 2009.
- Sergio Mera defended his PhD thesis in cotutelle between the Université Henri Poincare and the Universidad de Buenos Aires entitled *Modal Memory Logics*, supervised by Patrick Blackburn and Verónica Becher, on 9 December 2009.

8.2. Service to the Scientific Community

- Carlos Areces
  - Member of the Management Board of the Association of Logic, Language and Information (FoLLI), 2005–2009.
- Patrick Blackburn
  - Member of the INRIA Nancy-Grand Est steering committee.
  - Member of the Management Board of the Association of Logic, Language and Information, 2005–2009.
  - 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 SIGSEM board
  - Member of the LORIA steering committee.
  - Coordinator of the TALC theme (Computational Linguistics and Computational Approaches to Knowledge) for the MISN CPER (National and Regional Research Funding).
  - Organiser of the LORIA TALC seminar
  - Local organiser for the NaTAL workshop
  - Member of the recruiting committee for short term posts at INRIA Lorraine/LORIA
  - Member of Nancy 2 University selection committee

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

8.3. 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 FOLLI Editorial Board for the 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 Program Committee of the Automated Deduction: Decidability, Complexity, Tractability Workshop (ADDCTâ09) Montreal, Canada.
  - 2009 Member of the Program Committee of the 2009 International Workshop on Description Logics (DL2009), Dresden, Germany.
  - Member of the Program Committee of the 6th Methods for Modalities Workshop (M4M6), Copenhagen, Denmark.

- Patrick Blackburn:
  - Chief Editor of the *Journal of Logic, Language, and Information*, 2002 – Present.
  - Editor of the *Review of Symbolic Logic*, from 2007 – Present.
  - Editor of the *Notre Dame Journal of Formal Logic, 2005 – Present.*
  - Subject Editor (Logic and Language) for the *Stanford Encyclopedia of Philosophy*.
  - Foreign Correspondent of *Logique et Analyse*.

- Claire Gardent:
  - Program chair of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09), Athens, Greece.
  - Area Chair ESSLLI 2009, section Language and Computation
  - PC member for PRE-CogSci (Production of referring expressions) workshop 2009 of the Annual Meeting of the Cognitive Science Society, Amsterdam, The Netherlands.
  - PC member for TALN (Traitement Automatique des Langues Naturelles) 2009, Senlis, France.
  - PC member for DiaHolmia 2009 (Workshop on the semantics and pragmatics of dialogue), Stockholm, Sweden.

  - PC member for ENLG 2009 (12th European Workshop on Natural Language Generation), Athens, Greece.
  - PC member for NAACL-HLT 2009 (North American Chapter of the Association for Computational Linguistics - Human Language Technologie), Boulder, Colorado.

- Carlos Areces:
  - One week course on “Description Logics” at the Universidad de Salamanca, Spain.
  - Invited tutorial on “Computational Logic” and invited lecture at the Tbilisi Symposium, Georgia.
  - Invited tutorial on “Computational Modal Logics” at the FIRST Autumn School, Copenhagen, Denmark.
- Patrick Blackburn:
  - Invited Speaker at *Tableaux 2009, The 18th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods*, Oslo, Norway, 9 July 2009.
  - Invited tutorial on “Computational Modal Logics” at the FIRST Autumn School, Copenhagen, Denmark.
- Claire Gardent:
  - Invited Speaker at *Rencontres Grammaires Electroniques*, Marseilles, 29-31 May 2009.
  - Invited tutorial on “Complex matching situations via paraphrasing” at the intensive summer school on Question Answering, Kasetsart University, Bangkok, Thailand, 18-19 October 2009.

9. 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, D. FIGUEIRA. *Which Semantics for Neighbourhood Semantics?*, in "Proceedigns of IJCAI 09, Pasadena, California, USA", 2009, p. 671–676.

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

[4] 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.

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

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

[7] 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, http://doi.acm.org/10.1145/1284420.1284451.

[8] 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.

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

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

[11] 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

Articles in International Peer-Reviewed Journal

[12] H. BONNEAU-MAYNARD, M. QUIGNARD, A. DENIS. *MEDIA: a semantically annotated corpus of task oriented dialogs in French*, in "Language Resources and Evaluation", 2009, http://hal.inria.fr/inria-00424619/en/.

[13] S. CRUZ-LARA, N. BELLALEM, L. BELLALEM, T. OSSWALD. *Immersive 3D Environments and Multilinguality: Some Non-Intrusive and Dynamic e-learning-oriented Scenarios based on Textual Information*, in "Journal of Virtual Worlds Research", vol. 2, n^{o} 3, 2009, ., http://hal.inria.fr/inria-00426562/en/.

[14] L. KALLMEYER, Y. PARMENTIER. *Convertir des grammaires d'arbres adjoints à composantes multiples avec tuples d'arbres (TT-MCTAG) en grammaires à concaténation d'intervalles (RCG)*, in "Traitement Automatique des Langues", 2009, http://hal.inria.fr/inria-00380770/en/DE.

[15] C.-Y. LU, S.-H. LIN, J.-C. LIU, S. CRUZ-LARA, J.-S. HONG. *Automatic event-level textual emotion sensing using mutual action histogram between entities.*, in "Expert Systems with Applications", 7 2009, http://hal.inria.fr/inria-00426567/en/TW.

[16] D. SUSTRETOV. *Hybrid logics of separation axioms.*, in "Journal of Logic Language and Information", vol. 18, n^{o} 4, 2009, p. 541-558, http://hal.inria.fr/inria-00424685/en/.

[17] B. TEN CATE, G. DAVID, D. SUSTRETOV. *Modal languages for topology: Expressivity and definability*, in "Annals of Pure and Applied Logic", vol. 159, n^{o} 1-2, 2009, p. 146-170, http://hal.inria.fr/inria-00424693/en/GENL.

International Peer-Reviewed Conference/Proceedings

[18] C. ARECES, D. FIGUEIRA. *Which Semantics for Neighbourhood Semantics?*, in "Twenty-first International Joint Conference on Artificial Intelligence - IJCAI 09, États-Unis d'Amérique Pasadena, California", 2009, p. 671–676, http://hal.inria.fr/inria-00423047/en/.

[19] C. ARECES, D. FIGUEIRA, D. GORIN, S. MERA. *Tableaux and Model Checking for Memory Logics*, in "18th International Conference on Automated Reasoning with Analytic Tableaux and Related Methods -

TABLEAUX 2009, Norvège Oslo", M. GIESE, A. WAALER (editors), vol. 5607, Springer-Verlag, 2009, p. 47–61, http://hal.inria.fr/inria-00423049/en/.

[20] C. ARECES, D. GORIN, A. LORENZO, M. PEREZ RODRIGUEZ. *Testing Provers on a Grid - Framework Description*, in "Proceedings of the 22nd International Workshop on Description Logics - DL 2009, Royaume-Uni Oxford", B. CUENCA GRAU, I. HORROCKS, B. MOTIK, U. SATTLER (editors), vol. 477, CEUR-WS.org, 2009, http://hal.inria.fr/inria-00423068/en/.

[21] P. BEDARIDE, C. GARDENT. *Semantic Normalisation : a Framework and an Experiment*, in "8th International Conference on Computational Semantics - IWCS 2009, Pays-Bas Tilburg", 2009, http://hal.inria.fr/inria-00413636/en/.

[22] L. BENOTTI. *Clarification potential of instructions*, in "The 10th Annual SIGDIAL Meeting on Discourse and Dialogue, Royaume-Uni London", 2009, http://hal.inria.fr/inria-00424676/en/.

[23] L. BENOTTI. *Frolog: An accommodating text-adventure game*, in "The 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09), Grèce Athens", 2009, http://hal.inria.fr/inria-00424678/en/.

[24] L. BENOTTI, D. TRAUM. *A computational account of comparative implicatures for a spoken dialogue agent*, in "Eighth International Conference on Computational Semantics - IWCS-8 2009, Pays-Bas Tilburg", 2009, http://hal.inria.fr/inria-00424703/en/US.

[25] I. FALK, S. CRUZ-LARA, N. BELLALEM, L. BELLALEM. *Linguistic Information For Multilinguality In The Sembysem Project*, in "The International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - IC3K, Portugal Madeira", 2009, http://hal.inria.fr/inria-00426572/en/.

[26] I. FALK, S. CRUZ-LARA, N. BELLALEM, L. BELLALEM. *Linguistic Information for Multilinguality in the SEMbySEM Project.*, in "IC3K - International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, Portugal Funchal, Madeira", 2009, http://hal.inria.fr/inria-00423602/en/.

[27] I. FALK, C. GARDENT, E. JACQUEY, F. VENANT. *Grouping Synonyms by Definitions*, in "Recent Advances in Natural Language Processing (RANLP), Bulgarie Borovets", R. MITKOV (editor), University of Wolverhampton, UK Institute for Parallel Processing, BAS, Bulgaria Incoma Ltd, Shoumen, Bulgaria, 2009, 6, http://hal.inria.fr/inria-00418458/en/.

[28] L. KALLMEYER, W. MAIER, Y. PARMENTIER. *An Earley Parsing Algorithm for Range Concatenation Grammars*, in "Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009), Singapour Suntec", 2009, http://hal.inria.fr/inria-00393980/en/DE.

[29] L. KALLMEYER, W. MAIER, Y. PARMENTIER. *Un Algorithme d'Analyse de Type Earley pour Grammaires à Concaténation d'Intervalles*, in "Conférence sur le Traitement Automatique des Langues Naturelles - TALN'09, France Senlis", ATALA, 2009, http://hal.inria.fr/inria-00384556/en/DE.

[30] F. VENANT. *Géométriser le sens lexical La synonymie comme accès à la sémantique*, in "Quatrième conférence internationale sur la Théorie Sens-Texte (MTT'09), Canada montréal", 2009, xx, http://hal.archives-ouvertes.fr/hal-00429219/en/.

National Peer-Reviewed Conference/Proceedings

[31] C. CERISARA, C. GARDENT. *Analyse syntaxique du français parlé*, in "Journée ATALA, France Paris", 2009, http://hal.inria.fr/inria-00432754/en/.

[32] A. DENIS, M. QUIGNARD. *Exploitation du terrain commun pour la production d'expressions référentielles dans les systèmes de dialogue*, in "Conférence sur le Traitement Automatique des Langues Naturelles - TALN'09, France Senlis", 2009, http://hal.inria.fr/inria-00424625/en/.

[33] I. FALK, C. GARDENT, E. JACQUEY, F. VENANT. *Sens, synonymes et définitions*, in "Conférence sur le Traitement Automatique du Langage Naturel - TALN'2009, France Senlis", Laboratoire d'Informatique de Paris-Nord, Université Paris 13 & CNRS (LIPN), 2009, http://hal.inria.fr/inria-00403572/en/.

Workshops without Proceedings

[34] I. FALK, C. GARDENT, E. JACQUEY, F. VENANT. *A method for grouping synonyms.*, in "eLexicography in the 21st century (eLex 2009), Belgique Louvain-la-Neuve", Centre for English Corpus Linguistics (CECL) under the aegis of the European Association for Lexicography (EURALEX), 2009, http://hal.inria.fr/inria-00424897/en/.

References in notes

[35] 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.

[36] C. A. R. BERNARDI. *Analyzing the Core of Categorial Grammar*, in "Journal of Logic, Language and Information", vol. 13, n^{o} 2, 2004, p. 121–137.

[37] P. BLACKBURN, J. BOS. *Computational Semantics*, in "Theoria", vol. 18, n^{o} 46, Jan 2003, p. 27-45.

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

[39] 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.

[40] 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.

[41] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Reprsentation et gestion de grammaires d'arbres adjoints lexicalises*, in "Traitement Automatique des Langues", vol. 44, n^{o} 3, Dec 2003, p. 67-91.

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

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

[44] 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.

[45] 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.

[46] 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.

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

[48] 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.

[49] 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.

[50] C. GARDENT, H. MANULIAN. *Cration d'un corpus annot pour le traitement des descriptions dfinies*, in "Traitement automatique des langues", vol. 46, n^{o} 1, 2005.

[51] C. GARDENT, H. MANULIAN, 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.

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

[53] 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.

[54] 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.

[55] 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.

[56] H. MANULIAN. *Annotation des descriptions dfinies : le cas des reprises par les rles thmatiques*, in "Rencontre des tudiants Chercheurs en Informatique pour le Traitement Automatique des Langues - RECITAL'2002, Nancy, France", Jun 2002.

[57] H. MANULIAN. *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.

[58] H. MANULIAN. *Descriptions Dfinies et Dmonstratives : Analyses de corpus pour la gnration*, Universit de Nancy 2, Nov 2003, Ph. D. Thesis.

[59] H. MANULIAN. *Gnration de descriptions dfinies et dmonstratives*, in "Huitime Atelier des doctorants en linguistique - ADL'2003, Paris, France", Jul 2003.

[60] H. MANULIAN. *Une analyse des emplois du dmonstratif en corpus*, in "Traitement Automatique des Langues Naturelles - TALN 2003, Batz sur Mer, France", Jun 2003.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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*, Université Henri Poincaré - Nancy 1, Nov 2004, Ph. D. Thesis.

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