



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team talaris

*Natural Language Processing:
representation, inference and semantics*

Nancy - Grand Est

Theme : Audio, Speech, and Language Processing

Activity
R *eport*

2010

Table of contents

1. Team	1
2. Overall Objectives	1
2.1. Background	1
2.2. Organization	1
2.3. Overall Objectives	2
2.4. Highlights	2
3. Scientific Foundations	2
3.1. Computational Linguistics and Computational Logic	2
3.2. Semantics and Inference	3
3.3. Linguistic Resources	3
3.4. Logic Engineering	4
3.5. Empirical Studies	4
4. Application Domains	5
4.1. Modular Grammar Building	5
4.2. Referential Expressions	5
4.3. Surface Realization	5
4.4. Textual Entailment Recognition	5
4.5. Computational Logics and Computational Semantics.	6
4.6. Hybrid Automated Deduction	6
4.7. Multimedia	6
4.8. Interfacing Virtual Worlds and Natural Language Processing	6
5. Software	7
5.1. AGREE	7
5.2. The eXtended Meta-Grammar (XMG) Compiler and Tools	7
5.3. Frolog	8
5.4. GenI surface realiser	8
5.5. HyLoRes, a Resolution Based Theorem Prover for Hybrid Logics	8
5.6. HTab, a Tableau Based Theorem prover for Hybrid Logics	9
5.7. HyLoBan, a Game Based Theorem Prover for Topological Hybrid Logics	9
5.8. hGEN, a Random Formula Generator	9
5.9. MEDIA	10
5.10. Nessie	10
5.11. DeDe Corpus	10
5.12. SemTAG	11
5.13. SemXTAG	11
5.14. WikipediaAnnotator	11
5.15. ATOOL	11
5.16. SRL-Web Annotation	11
5.17. PORT-MEDIA	12
5.18. WSMACI	12
5.19. 4LED	12
5.20. SLMC	12
5.21. DiacXis	13
5.22. IGNGF	13
5.23. TextClus	14
6. New Results	14
6.1. Dynamic Modal Logics	14
6.2. Automated Deduction for Hybrid Logics	14
6.3. Integrating inference with dialogue	14

6.4.	Giving Instruction in a Virtual World (GIVE)	14
6.5.	Using 3D worlds to teach French (I-FLEG)	15
6.6.	Semantic construction using rewriting	15
6.7.	Using Regular Tree Grammars to enhance surface realisation	15
6.8.	Verb classes and Semantic Role Labelling for French	15
6.9.	Deep analysis of interaction patterns in texts	15
6.10.	MLIF	15
6.11.	New clustering quality metrics for the analysis of complex textual data	16
6.12.	New diachronic method for the analysis of slowly time-varying textual data	16
6.13.	New incremental clustering algorithms for the analysis of highly time-varying textual data	16
6.14.	Building up evolving networks of wikis	16
7.	Other Grants and Activities	17
7.1.	Regional Initiatives	17
7.2.	National Initiatives	17
7.2.1.	CCCP-Prosodie	17
7.2.2.	PORT-MEDIA	17
7.2.3.	Passage	18
7.3.	European Initiatives	18
7.3.1.	Allegro	18
7.3.2.	EMOSPEECH	18
7.3.3.	METAVVERSE	18
7.3.4.	SEMbySEM: Services management by Semantics	19
7.4.	International Initiatives	19
8.	Dissemination	20
8.1.	PhD Theses	20
8.2.	Habilitations	20
8.3.	Service to the Scientific Community	20
8.4.	Editorial and Program Committee Work	21
8.5.	Teaching	22
9.	Bibliography	23

1. Team

Research Scientists

Patrick Blackburn [Team Leader, Research Director (DR2), INRIA, HdR]
Claire Gardent [Research Director (DR2), SHS Department CNRS, HdR]
Carlos Areces [Research Scientist (CR1), INRIA. On sabbatical leave 2009-04-01 / 2010-04-01]

Faculty Members

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, IUT Robert Schuman, University of Strasbourg, HdR]
Fabienne Venant [Assistant Professor, IUT Nancy-Charlemagne, University of Nancy 2. On maternity and parental leave from 2010-06-01]

Technical Staff

Mathieu Desnouveaux [INRIA Engineer on CPER MISN/TALC, 2008-09-01 until 2010-09-01]
Emmanuel Didiot [INRIA Engineer on CPER MISN/TALC, from 2010-12-01.]
Treveur Bretaudivere [INRIA Engineer on ADT MoViTAL, from 2010-12-01]
Tarik Osswald [Engineer, ITEA Project MetaVerse]

PhD Students

Corinna Anderson [Yale University, USA]
Paul Bedaride [UHP Nancy 1, Ministry grant, from 2006-10-15 until 2010-10-15]
Luciana Benotti [UHP Nancy 1, CORDIS 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]
Guillaume Hoffmann [UHP Nancy 1, Ministry grant, from 2007-10-15 until 2010-11-15]
Laura Perez [[University of Nancy 2, Ministry grant, from 2009-10-01 until 2012-10-01]
Alejandra Lorenzo [[UHP Nancy 1, Région/ALLEGRO Interreg IV A project, from 2010-10-15]

Administrative Assistant

Isabelle Blanchard

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 constructed) 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 design and implementation of incremental clustering techniques that can handle heterogeneous textual data collections.
- 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 integration and use of this framework in systems interfacing 3D worlds and Natural Language Processing (NLP) technologies e.g., extending a serious game with dialog capabilities or exploiting Natural Language Generation to automate the production of language learning exercices in a 3D setting.
- 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

- New results on dynamic modal logics (e.g., memory logics) in terms of lower and upper complexity bounds, tableaux algorithms, expressive power, axiomatization and interpolation.
- Mature expositions of inference frameworks for hybrid logics, together with stable implementations of inference systems.
- Talaris systems ranked first and third in the international GIVE (Giving Instruction in Virtual Environment) challenge [57].
- The 3 day Natal workshop, organized by Claire Gardent, which gathers together students and researchers from Nancy, Saarbrücken and neighbouring areas around current themes in NLP themes, was successfully held (for the third running) from 16–18 June 2010 <http://talc.loria.fr/NATAL-10-16-18-juin-2010-LORIA,178.html>. As in 2008 and 2009, there were two thematic days with invited speakers (this year, on Natural Language Processing and Computer Aided Language Learning) and a third Masters Student Day where Erasmus Mundus students from Nancy and Saarbrücken met and presented 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 exist 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 exist 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, it 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 information¹.

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 descriptions²; an algorithm for generating bridging definite descriptions has been specified and implemented which illustrates the interaction of realisation and inference³; 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 description⁴; and a shallow anaphora resolver for French has been developed and evaluated within the national evaluation campaign MeDIA.

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 surface realisers for French called GenI⁵ and RTGen⁶. While GenI manipulates full elementary trees to build derived trees, RTGen uses an RTG (Regular Tree Grammar) encoding of the Tree Adjoining Grammar to first build derivation trees. A preliminary comparison suggests that RTGen outperforms GenI. Current work concentrates on further optimising RTGen and on integrating top down control to reduce the search space and constrain the output.

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.

¹[63], [65], [64], [66], [67], [68], [69], [88], [85], [86], [87], [75], [84]

²[78], [82], [79], [81], [80], [83], [72], [73]

³[89], [74], [76], [59]

⁴[70], [59]

⁵[77], [71]

⁶[36], [38]

Textual entailment recognition is a difficult task. We work on developing linguistically principled approaches to tackle specific entailment sources such as syntactic variations or the compositional semantics of factive verbs [12], [20], [21].

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 construction⁷. The book “Representation and Inference for Natural Language: A First Course in Computational Semantics”⁸ 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 (first-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.

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.

4.8. Interfacing Virtual Worlds and Natural Language Processing

In 2010, Talaris addressed a new application domain namely, the integration of deep natural language processing (NLP) techniques with 3D worlds and games. A first foray into that theme has been the submission of two systems to the international GIVE (Giving instructions in a virtual environment). 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 developed by Talaris over the last 5 years for deep NLP and in particular, on the availability of an expressive grammar writing environment (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.

⁷[60], [61]

⁸[62]

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 Traitement 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 being 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 developed by the University of Pennsylvania. Both grammars are additionally equipped 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://talc.loria.fr/GenI-un-realisateur-de-surface.html>

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 $H(@, \downarrow)$, 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/content/tools.php>

Authors: Carlos Areces, Daniel Gorín and Juan Heguibehere

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: 1.5.4

License: GPL

Last update: 2010-11-10

Web site: <http://www.glyc.dc.uba.ar/content/tools.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/content/tools.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 $H(@, \downarrow, A, P)$. 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/content/tools.php>
Authors: Carlos Areces, Daniel Gorín, Juan Heguiabehere and Guillaume Hoffmann
Contact: Carlos Areces

5.9. 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 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

5.10. 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

5.11. 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 information⁹.

Authors: Claire Gardent, H el ene Manuelian
Web site: Distributed by the CNRTL <http://www.cnrtl.fr/>
Contact: Claire Gardent

⁹[72]

5.12. SemTAG

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

Contact: Claire Gardent

5.13. 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

5.14. WikipediaAnnotator

The WikipediaAnnotator program provides semantic annotation of Wikipedia discussion pages. It annotates French Wikipedia participants utterances on the connotation and subjective levels using deep syntax and shallow semantics.

Developing context: CCCP-Prosodie

Programming language: Java

Development effort: 18 man/month

Type of license: GPL

Partners: Telecom Paris Tech

Web site: – (in construction)

5.15. ATOOL

ATOOL is a semantic Annotation Tool for the High-Level Semantic Representation MMIL which takes as input the MEDIA corpus (TEI format according to the specifications document of September 2009)

Developing context: PORT-MEDIA

Programming language: Java

Development effort: 4 months (student) + 1 month of evaluation and improvements.

Expected users: Annotators

Type of license: Open Source.

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

5.16. SRL-Web Annotation

This tool allows the annotation of the utterances in the MEDIA corpus and stores all the information about predicates and arguments in a relational database.

Developing context: PORT-MEDIA

Programming language: JSP-JAVA

Development effort: 1 month.

Expected users: Annotators

Type of license: Open Source.

Web site: <http://talca.loria.fr/>

5.17. PORT-MEDIA

PORT-MEDIA is a framework supporting the automatic annotation of the MEDIA corpus with the high-level semantic representation MMIL. It is a blackboard architecture interfaced with the Tree Tagger, the Malt parser, the frames, the semantic role labeling and the HLSBuilder for the automatic annotation of the MEDIA Corpus. Additionally it is interfaced with RACER, two ontologies in owl and a relational database with all the information at different linguistic levels. Finally the evaluation software is also provided.

Developing context: PORT-MEDIA

Programming language: Java, MySql, XML, XSLT, OWL.

Development effort: 12 months.

Expected users: Annotators

Type of license: Open Source.

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

5.18. WSMACI

The Web Service for the Multilingual-Assisted Chat Interface program (WSMACI) is a linguistic assistant for virtual worlds. Its first version is dedicated to English assistance in such worlds. It has been developed in the context of the Metaverse1 project. It provide the end-users with MLIF-based provision of sentence analysis and word information (synonyms, definitions, translations) based on Google Translate, WordNet and the Brown Corpus.

Programming language: MLIF, PHP, SQL

Development effort: 6 man/month

Type of license: INRIA specific

Authors : Tarik Oswald, Samuel Cruz-Lara

Contact: Tarik Oswald

Web site: – (in construction)

5.19. 4LED

The 4 Layers Emotion Detection program (4LED) is an emotion detection tool. The emotions are extracted from texts in particular, from chat interfaces in virtual worlds. It has been developed in the context of the Metaverse1 project. The emotion detection process is based on SMILEY detection using WordNet-Domains and Tree-Tagger-based rules, WordNet-Affect, and keywords.

Programming language: MLIF, PHP, SQL

Development effort: 6 man/month

Type of license: INRIA specific

Partners: ArtefactO (for rendering)

Authors : Tarik Oswald, Samuel Cruz-Lara

Contact: Tarik Oswald

Web site: – (in construction)

5.20. SLMC

The Second Life Magic Carpet program (SLMC) is an assistant whose role is to guide people through virtual worlds with textual instructions. It has been developed in the context of the Metaverse1 project. It has been developed in the context of the Metaverse1 project. It analyses the instructions of the visitors in order to find where they want to go, using web services for the analysis, for synonyms retrieving and for path finding.

Developing context: Metaverse1
Programming language: LSL, PHP, SQL
Development effort: 8 man/month
Type of license: INRIA specific
Partners: Innovalia Spain, Utrecht University
Authors : Tarik Oswald, Samuel Cruz-Lara
Contact: Tarik Oswald
Web site: – (in construction)

5.21. DiacXis

The DiacXis program seeks to analyse the evolution over time of textual information. It has been developed in the context of the CPER TALC action McFiiD. DiacXis addresses the analysis of textual information evolving over time using a diachronic approach based both on the Multiview Data Analysis (MVDA) paradigm and on specific cluster labeling techniques especially developed for the statistical analysis of complex data, like textual data ¹⁰. Given a set of textual datasets stemming from different time periods, DiacXis provides the information analyst with lists of stable, appearing, disappearing, merging and splitting topics over these time periods.

Programming language: C + Java
Development effort: 6 man/month
Type of license: INRIA specific
Partners: INIST, ITT Kampur
Authors : Jean-Charles Lamirel, Navesh Pryankar
Contact: Jean-Charles Lamirel
Web site: – (in construction)

5.22. IGNGF

The IGNG-F program implements a new incremental clustering algorithm whose main domain of application is the statistical analysis of continuous flow of evolving textual data. It has been developed in the context of the CPER TALC (McFiiD action). It is based on a generic adaptation of the classical neural-based clustering approaches using gas of neurons with free topology. This adaptation resulted in a description space independent and parameter-free, neural clustering technique using Hebbian learning and labeling expectation maximization instead of classical Euclidean or correlation distances ¹¹. We showed that this approach is more efficient than the usual techniques for the analysis of highly polythematic textual data. Given a continuous flow of textual data, the IGNG-F tool provides the information analyst with precise online detection of diachronic topic changes.

Programming language: C
Development effort: 6 man/month
Type of license: INRIA specific
Partners: INIST, ITT Hyderabad
Authors : Jean-Charles Lamirel, Raghvendra Mall
Contact: Jean-Charles Lamirel
Web site: – (in construction)

¹⁰[49]

¹¹[41]

5.23. TextClus

The TextClus interface is a java interface whose role is to provide end-users with a whole set of clustering techniques applied to texts. It has been developed in the context of the CPER TALC McFiiD operation. The platform uses a vectorial representation of text data. It includes, in a federating interface, the management of different kinds of data preprocessing techniques (IDF, entropy, random mapping, ...), different kinds of clustering techniques (from standard methods to elaborated neural ones), the management of multiple experiments and comparison of their results by method-independent clustering quality measures specifically adapted to the analysis of textual data ¹². Each step can be precisely tuned by the end-user through the TextClus program interface to construct a global experimental process.

Programming language: C

Development effort: 6 man/month

Type of license: INRIA specific

Partners: INIST

Authors : Jean-Charles Lamirel, Pascal Cuxac

Contact: Jean-Charles Lamirel

Web site: – (in construction)

6. New Results

6.1. Dynamic Modal Logics

We are investigating modal logics that include operators that not only allow for exploration of the model in which they are evaluated, but that can also *modify* it. These logics could be specially suitable to describe, for example, the semantic of utterances, to model the fact that uttering a sentence changes the context it was uttered in.

We have investigated in detail a family of dynamic modal logics called *memory logics* and established lower and upper complexity bounds, mapped their expressive power, devised tableaux and model checking algorithms, and sound and complete axiomatizations.

6.2. Automated Deduction for Hybrid Logics

This year we have finally released stable versions of the two theorem provers (HyLoRes and HTab) for hybrid logics developed by the team. Moreover, the theoretical framework behind each of them has been described in detail in a PhD thesis and a journal article.

6.3. Integrating inference with dialogue

This work concentrates on integrating techniques from logic and artificial intelligence (notably planning) with work on pragmatics and the structure of dialogue [13], [22].

6.4. Giving Instruction in a Virtual World (GIVE)

We developed two generation systems for the international GIVE challenge [57], [27]. Interfaced with the 3D game provided by the challenge organisers, these systems guide the player with natural language instructions they generate in real time and according to the player's position in the game. The two systems were ranked first and third. Their development revived a long standing Talaris/Led interest in the generation of referring expressions and launched a new line of research on situated referring expression generation and on the interface between 3D game and NLP.

¹²[48], [46]

6.5. Using 3D worlds to teach French (I-FLEG)

Within the Interreg IV A Allegro project, we investigated how embedding text generation in a 3D game could help automating the generation of situated language exercises i.e., exercises whose content varies with the 3D world context, with the learner level and with the teaching goal. We developed a system (I-FLEG) illustrating this interaction and made contact with language teachers to arrange for learner use and teacher testing. I-FLEG will be deployed in 2011 in language learning classes and its usability for language teaching tested.

6.6. Semantic construction using rewriting

Paul Bédaride and Claire Gardent developed a new method for constructing semantic representations for textual data which is based on graph rewriting [12], [20], [21]. They tested it on artificial data with good results and showed that it permits constructing deep semantic representations from dependency structures.

6.7. Using Regular Tree Grammars to enhance surface realisation

Making use of the fact that Regular Tree Grammars can generate the derivation trees of a Tree Adjoining Grammar, we developed an alternative surface realiser to GenI called RTGen. Preliminary results suggest that RTGen outperforms GenI. Current work concentrates on further optimising RTGen and on integrating top down control to reduce the search space and constrain the output [36], [36].

6.8. Verb classes and Semantic Role Labelling for French

Although verb classes and Semantic Role Labelling (SRL) have been shown to be an essential component of semantic processing, there is to date no such resource and tool for French. Using Formal Concept Analysis and information from the LADL tables, we developed a method for classifying verbs based on their subcategorisation information; and a method for projecting thematic roles onto the Paris 7 dependency bank. We are currently extending the verb classification approach to integrate thematic grids and working on evaluating and improving the result. In the long run, we aim to provide a Propbank style corpus for French to and to develop a semantic role labeller based on this corpus [35], [33], [32].

6.9. Deep analysis of interaction patterns in texts

In the context of the CCCP-Prosodie project, we showed that connotation and subjective markers are good evidence for conflicts between participants on Wikipedia discussion pages. We observed striking patterns of interaction during conflicts, especially the alternation of 1st person and 2nd person subjective markers in negatively connotated utterances. Further work will involve examine the various conflicts type, aiming to automatically distinguish argument-based conflicts from personal ones (*ad hominem*) [34]

6.10. MLIF

TALARIS contributes to ISO TC 37 committee “Terminologies and other Language Resources”, and more specifically to the activities of its SC3 “Computer Applications in Terminology”, and SC4 “Linguistic Resources Management”. Within TC37/SC4, TALARIS is currently contributing, as project leader, to the definition and specification of the Multi Lingual Information Framework (MLIF) [ISO DIS 24616]. MLIF is being designed with the objective of providing a common abstract model being able to generate several formats used in the framework of translation and localization. MLIF will soon be released as FDIS (Final Draft International Standard) and it should finally be published as an official ISO Standard within the first semester of 2011 [31], [25], [24].

6.11. New clustering quality metrics for the analysis of complex textual data

Cluster quality evaluation is a key issue for many data analysis tasks. As we showed in previous work, the classical distance based quality indices are often strongly biased and highly dependent on the clustering method. To cope with such problems, we proposed in earlier work specific Macro-Recall/Precision and F-measures metrics that exploit the properties of cluster associated data. However, our more recent experiments showed that these new metrics failed to highlight degenerated clustering results when analyzing complex textual data [39], [48]. To remedy this shortcoming, we devised two extensions of these metrics namely Micro-Measures [26], [48], [47] and Cumulated Micro-measures [46]. We then experimentally showed the effectiveness of our extended approach by applying it to different documentary corpus of highly polythematic bibliographic records issued from the PASCAL CNRS scientific database.

6.12. New diachronic method for the analysis of slowly time-varying textual data

The literature taking into account the chronological aspect in textual information flows focuses on "DataStream" whose main idea is the "on the fly" management of incoming data. However the proposed algorithms are intended to treat very large volumes of data and are thus not optimal for detecting emergent topics such as, for example, the evolution of a research theme within bibliographical records. To address this issue, we proposed a new approach based on our Multiview Data Analysis Paradigm (MVDA) [46], [54], [44], in combination with specific cluster labeling techniques especially developed for the statistical analysis of complex data, like textual data. We applied our approach to the IST PROMTECH reference dataset related to optoelectronic research. When compared with state-of the art approaches, our method proved to provide very significant added-value by permitting to precisely highlight and quantify the observed evolutions, and their related context, ranging from vocabulary changes in a given topic to overall appearing/disappearing of topics, or even to splitting or merging between topics [49].

6.13. New incremental clustering algorithms for the analysis of highly time-varying textual data

Neural clustering algorithms show high performance in the general context of the analysis of homogeneous textual datasets. However, we showed that there is a drastic decrease of performance of these algorithms, as well as of the more classical algorithms, when applied to heterogeneous or polythematic textual datasets. Such result degradation indicates that most of the existing clustering methods, even those that are considered incremental, are not really able to deal with highly time-varying data [56]. We therefore proposed a new approach to incremental clustering based on a generic adaptation of the classical neural-based clustering methods relying on gas of neurons with free topology. This adaptation resulted in a description space independent and parameter-free, neural clustering technique using Hebbian learning and labeling expectation maximization instead of classical Euclidean or correlation distances. We have proved that our approach very significantly outperformed the existing ones in all experimental contexts, and specifically, in the case of highly time-varying data [41], [42], [43].

6.14. Building up evolving networks of wikis

The WICRI Project aims to explore a new concept of "wikis network" with adaptive capabilities. In particular, one aim is to propose strategies for enriching the content of the wiki network by dynamic integration and exploitation of Web data. Hence, Web mining represents an important challenge for enhancing the dynamicity, the flexibility and the scope of such a network. On the one hand, this process is mandatory for assisting the potential contributors with elaborated and reliable redaction guidelines during the network construction phase. On the other hand, it is essential for supplying end-users with external information whose added value is to maintain significant relationships with the semantic context of the wiki network. Although the WICRI project is still in his launching phase, our preliminary prototype of "network of wikis" is already acting as an on-line collaborative research platform [30].

7. Other Grants and Activities

7.1. Regional Initiatives

7.1.1. *McFIID*

Theme: Clustering; Statistical Analysis; Textual data; Time-evolving data; Distributed data

Description: The McFIID project is a CPER project continuing the CPER CLASSIF project. It concerns the development of incremental multi-clustering techniques for managing distributed and evolving flows of textual data. New approach of diachronic analysis based on the use of multiple viewpoints combined with unsupervised bayesian reasoning, as well as new online incremental clustering techniques based on non standard similarity measures, are tested in the course of these project.

Administrative context: CPER

Web site: <http://wikitalc.loria.fr/dokuwiki/doku.php?id=operations:mcfiid>

Period: start 2007-01-01 / 2011-12-31

Contact: Jean-Charles Lamirel

Partner(s): INIST, LORIA

7.2. National Initiatives

7.2.1. *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 Compiègne, UNSA (Univ. Nice Sophia-Antipolis), Univ. de Versailles St-Quentin

7.2.2. *PORT-MEDIA*

Theme: Corpus Linguistics, Semantic Annotation of Corpora. Discourse, Dialogue and Pragmatics; Natural Language Understanding and Knowledge Representation

Description: The PORT-MEDIA project is an ANR project that aims to collect linguistic data for multiple domains and to investigate the use of a high-level semantic representation for annotating dialogue corpora. TALARIS contributed to the high-level semantics specification for annotating the MEDIA corpus and to the development of tools for the manual annotation (e.g., ATOOL and SRL-Web Annotation) as well as to the development of the blackboard architecture for the automatic annotation of the MEDIA corpus. Additionally, Talaris provided the automatic annotation of the whole corpus and its evaluation.

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: Matthieu Quignard, Lina M. Rojas-Barahona

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

7.2.3. Passage

Theme: Computational Semantics

Description: The PASSAGE project has two main aims. 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 resulting 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

Partner(s): CEA-LIST, LIMSI, INRIA Rocquencourt, CNRS

7.3. European Initiatives

7.3.1. Allegro

Theme: Computational Semantics

Description: The Allegro project aims to develop NLP techniques that support language teaching for French and German.

Administrative context: INTERREG IV A

Web site: <http://talca.loria.fr/-ALLEGRO-Nancy-.html>

Period: start 2010-01-01 / end 2012-12-31

Contact: Claire Gardent

Partner(s): Saarbrücken University, Supelec Metz, INRIA Nancy Grand Est

7.3.2. EMOSPEECH

Theme: Computational Semantics

Description: The EMOSPEECH project aims to augment serious games with natural language (spoken and written dialog) and emotional abilities (gesture, intonation, facial expressions).

Administrative context: Eurostars

Period: start 2010-09-01 / end 2013-08-31

Contact: Claire Gardent

Partner(s): Artefacto, Acapella, INRIA Nancy Grand Est

7.3.3. 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.

7.3.4. **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. International Initiatives

7.4.1. **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)

Web site: http://led.loria.fr/dokuwiki/doku.php?id=intohylo_-_inria_equipes_associees

Period: start 2009-01 / end 2012-01

Contact: Carlos Areces

Partner(s): Universidad de Buenos Aires, Argentina.

8. Dissemination

8.1. PhD Theses

- Luciana Benotti defended her PhD thesis at the Université Henri Poincaré entitled *Implicature as an Interactive Process* supervised by Patrick Blackburn, on 28 January 2010.
- Dimitri Sustrétoff defended his PhD thesis at the Université Henri Poincaré entitled *Topological semantics for hybrid logic* supervised by Patrick Blackburn, on 9 July 2010.
- Paul Bedaride defended his PhD thesis at the Université Henri Poincaré entitled *Implication textuelle et réécriture* supervised by Claire Gardent, on 18 October 2010. He is now a Postdoc at Stuttgart University.
- Guillaume Hoffman defended his PhD thesis at the Université Henri Poincaré entitled *Taches de raisonnement en logiques hybrides* supervised by Patrick Blackburn and Carlos Areces, on 13 December 2010.

8.2. Habilitations

- Jean-Charles Lamirel defended his habilitation (HdR) at the Université Henri Poincaré entitled *Vers un approche systémique et multivues pour l'analyse de données et la recherche d'information : un nouveau paradigme* on 6 December 2010 [55].

8.3. Service to the Scientific Community

- Carlos Areces:
 - Member of the Management Board of the Association of Logic, Language and Information (FoLLI), 2005-2010.
- Patrick Blackburn
 - Liaison officer for the *Erasmus Mundus* Masters in *Language and Communication Technology*.
- Samuel Cruz-Lara
 - In charge, at the national level, of the reception of Mexican students in the “Professional Licences of Computer Science”.
 - Member of W3C’s SYNchronized MultiMedia Group <http://www.w3.org/AudioVideo/Group/>
 - Member of ISO’s TC37 “Terminologies and other Language Resources” / SC4 “Linguistic Resources Management”. Project leader of the Multi Lingual Information Framework (ISO DIS 24616).
- 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 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 <http://talcl.loria.fr/-TALC-Seminar-.html>
- Local organiser for the NaTAL 2010 workshop <http://talcl.loria.fr/NATAL-10-16-18-juin-2010-LORIA,178.html>
- Jean-Charles Lamirel:
 - Member of the Management Board of the Collnet international research group in Scientometrics/Informetrics/Webometrics, 2005-2010.
- Fabienne Venant
 - Member of the Administrative Council of ATALA, the French national organisation for computational linguistics (see <http://www.atala.org/>).

8.4. Editorial and Program Committee Work

- Carlos Areces
 - Member of the Editorial Board of the FoLLI Publications on Logic, Language, and Information (part of the Lecture Notes in Artificial Intelligence series published by Springer-Verlag). Since 2006.
 - Member of the Scientific Board of The Baltic International Yearbook of Cognition, Logic and Communication. Since 2005
 - Member of the Editorial Board of the Journal of Logic, Language and Information. Since 2004.
 - Member of the Editorial Board of the Journal of Applied Logics. Since 2004.
 - Member of the Organizing Committee of the *Workshop on NLP and Web-based technologies* held in conjunction with IBERAMIA 2010, Bahía Blanca, Argentina.
 - Member of the Program Committee of the *36th Latin American Conference of Informatics*, Asunción, Paraguay.
 - Member of the Program Committee of the *2010 Workshop on Hybrid Logics (HyLo 10)*, Edinburgh, United Kingdom.
 - Member of the Program Committee of the *Iera Escuela de Lingüística Computacional (ELiC-1)*, Buenos Aires, Argentina.
 - Member of the Program Committee of the *2010 International Workshop on Description Logics (DL2010)*, Waterloo, Canada.
 - Member of the Program Committee of the *International Joint Conference on Automated Reasoning (IJCAR10)* Edinburgh, United Kingdom.
 - Member of the Program Committee of *Advances in Modal Logic (AiML10)* Moscu, Russia.
- Patrick Blackburn
 - editor of Review of Symbolic Logic
 - editor of Notre Dame Journal of Formal Logic
 - editorial board of Logique et Analyse
 - subject editor (Logic and Language), Stanford Encyclopedia of Philosophy
 - Program committee of Advances in Modal Logic 2010 (AiML 2010)
 - Program committee of Hybrid Logic 2010 (HyLo 2010)
 - Program committee of Workshop on Theories of Information Dynamics and Interaction and their Application to Dialogue 2010 (TIDIAD@ESSLLI 10)

- Nadia Bellalem
 - PC Member for *ICEIS 2010 (The 12th International Conference on Enterprise Information Systems)*, Madeira, Portugal.
- Samuel Cruz-Lara
 - PC member for *KEOD 2010 (The International Conference on Knowledge Engineering and Ontology Development)*, Valencia, Spain.
 - Member of the Editorial Board of *Revista Iberoamericana de Tecnologías del Aprendizaje*
- Claire Gardent
 - PC member for *TALN 2010 (Traitement Automatique des Langues Naturelles) 2010*, Montréal, Canada.
 - PC member for *SemDial 2010 (14th Workshop on the Semantics and Pragmatics of Dialogue)*, Poznan, Poland.
 - PC member for *RFIA 2010 (17ème Colloque Francophone sur la Reconnaissance des Formes et l'Intelligence Artificielle)*, Caen, France.
 - PC member for *EMNLP 2010 (Conference on Empirical Methods in Natural Language Processing)*, MIT, USA.
 - PC member for *LREC 2010 (The seventh international conference on Language Resources and Evaluation)*, Malta.
 - PC member for *INLG 2010 (12th European Workshop on Natural Language Generation)*, Dublin, Ireland.
 - PC member for *ACL 2010 (Annual meeting of the Association for Computational Linguistics)*, Uppsala, Sweden.
- Jean-Charles Lamirel
 - Member of editorial board of the new international journal “COLLNET Journal of Scientometrics and Information Management”, Taru Publications, New Delhi, India (<http://www.tarupublications.com/>).
 - Reviewer for the Neural Networks, Geographical Information Systems and Collnet International Journals.
 - Program chair of VSST Technological and Strategic Survey Conference VSST 2010, Toulouse, France, October 2010.
 - Organizer of Special Session on Incremental clustering and novelty detection techniques and their application to intelligent analysis of time varying information in the framework of IEA/IAE International Conference, Syracuse, NY, USA, June 2011.
 - Co-organizer of the ECG Workshop: Clustering incrémental et méthodes de détection de nouveauté et leur application à l'analyse intelligente d'information évoluant au cours du temps, EGC 2011 Workshop, Brest, France, January 2011.
- Christine Fay-Varnier
 - Member of the organizing committee of TICE 2010 conference (<http://www.tice2010.nancy-universite.fr/>)

8.5. Teaching

- Carlos Areces
 - Invited one week course at ELiC-1, Buenos Aires, Argentina. Invited one week course at NASSLLI 2010, Indiana, USA. Invited talk at JCC 2010, Rosario, Argentina.

- Patrick Blackburn
 - M2 course “Mathematics for Computer Science: Introduction to computability and computational complexity”. 15 hours, Erasmus Mundus Master “Language and Communication Technology”, University of Nancy 2, France, <http://webloria.loria.fr/~blackbur/courses/math/>.
 - M2 course “Discourse and Dialogue”. 15 hours, Erasmus Mundus Master “Language and Communication Technology”, University of Malta, <http://webloria.loria.fr/~blackbur/courses/dad/>
- Samuel Cruz-Lara
 - M2 course “Cognitive Sciences and Digital Media Technologies”. 22 hours, Cognitive Sciences, University of Nancy 2, France.
 - M2 course “Declarative Languages and Multimedia Applications”. 40 hours, Cognitive Sciences, University of Nancy 2, France.
 - M2 course “Video: Streaming and Captioning”. 12 hours, Cognitive Sciences, University of Nancy 2, France.
- Claire Gardent
 - Software project tutoring “Error mining and Surface Realisation”, Erasmus Mundus Master “Language and Communication Technology”, Nancy 2.
 - Invited tutorial on “Natural language processing and computer aided language learning”, 4th intensive Summer School and Collaborative Franco-Thai Workshop on Natural Language Processing, Kasetsart University, Bangkok, Thailand.
- Jean-Charles Lamirel
 - Master and PhD course on “Text Mining techniques applied to Linguistics: Introduction to the use of statistical methods for the analysis of literature. Case studies in French Literature. 20 hours, University of Alger.

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, S. FIGUEIRA. *Which Semantics for Neighbourhood Semantics?*, in "Proceedings 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", 2006, n^o 84, p. 277–322.
- [6] T. BOLANDER, P. BLACKBURN. *Termination for Hybrid Tableaus*, in "Journal of Logic and Computation", 2007, n^o 17, 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 Asynchronous 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", 2005, vol. 46, n^o 1.
- [11] C. GARDENT, K. STRIEGNITZ. *Generating Bridging Definite Descriptions*, in "Computing Meaning", H. BUNT, R. MUSKENS (editors), Studies in Linguistics and Philosophy Series, Kluwer Academic Publishers, 2007, vol. 3.

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [12] P. BEDARIDE. *Implication Textuelle et Réécriture*, Université Henri Poincaré - Nancy I, October 2010, <http://hal.inria.fr/tel-00541581/en>.
- [13] L. BENOTTI. *L'implicature comme un Processus Interactif*, Université Henri Poincaré - Nancy I, January 2010, <http://hal.inria.fr/tel-00541571/en>.
- [14] G. HOFFMANN. *Tâches de raisonnement en logiques hybrides*, Université Henri Poincaré - Nancy I, December 2010, <http://hal.inria.fr/tel-00541664/en>.

Articles in International Peer-Reviewed Journal

- [15] C. ARECES, S. FIGUEIRA, S. MERA. *The Expressive Power of Memory Logics*, in "Review of Symbolic Logic", August 2010, <http://hal.inria.fr/inria-00519040/en>.
- [16] C. ARECES, D. GORIN. *Coinductive models and normal forms for modal logics (or how we learned to stop worrying and love coinduction)*, in "Journal of Applied Logic", August 2010, <http://hal.inria.fr/inria-00519038/en>.
- [17] C. ARECES, D. GORIN. *Resolution with Order and Selection for Hybrid Logics*, in "Journal of Automated Reasoning", February 2010, <http://hal.inria.fr/inria-00519035/en>.
- [18] G. HOFFMANN. *Lightweight Hybrid Tableaux*, in "Journal of Applied Logic", 2010, vol. 8, n^o 4, p. 397–408, <http://hal.inria.fr/hal-00535976/en>.

International Peer-Reviewed Conference/Proceedings

- [19] C. ARECES, G. HOFFMANN, A. DENIS. *Modal Logics with Counting*, in "17th Workshop on Logic, Language, Information and Computation - WoLLIC 2010", Brazil Brasilia, 2010, <http://hal.inria.fr/hal-00482337/en>.
- [20] P. BEDARIDE, C. GARDENT. *Benchmarking for syntax-based sentential inference*, in "The 23rd International Conference on Computational Linguistics - COLING 2010", China Beijing, August 2010, <http://hal.inria.fr/inria-00536022/en>.
- [21] P. BEDARIDE, C. GARDENT. *Syntactic testsuites and Textual Entailment Recognition*, in "The seventh International Conference on Language Resources and Evaluation - LREC 2010", Malta Valletta, May 2010, <http://hal.inria.fr/inria-00536020/en>.
- [22] L. BENOTTI, P. BLACKBURN. *Negotiating causal implicatures*, in "Conference of the Special Interest Group in Dialogue", Japan Tokyo, September 2010, <http://hal.inria.fr/inria-00547508/en>.
- [23] C. CERISARA, C. GARDENT, C. ANDERSON. *Building and Exploiting a Dependency Treebank for French Radio Broadcast*, in "TLT9 – the ninth international workshop on Treebanks and Linguistic Theories", Estonia Tartu, November 2010, <http://hal.inria.fr/inria-00537147/en>.
- [24] S. CRUZ-LARA, G. FRANCOPOULO, L. ROMARY, N. SEMAR. *MLIF: A Metamodel to Represent and Exchange Multilingual Textual Information*, in "LREC (Language Resources and Evaluation Conference)", Malta Valletta, May 2010, <http://hal.inria.fr/inria-00518060/en>.
- [25] S. CRUZ-LARA, T. OSSWALD, J. GUINAUD, N. BELLALEM, L. BELLALEM. *Standards for communication and E-learning in virtual worlds: The Multilingual-Assisted Chat Interface*, in "12th International Conference on Enterprise Information Systems - ICEIS 2010", Portugal Funchal, J. CORDEIRO, J. FILIPE (editors), SciTePress, June 2010, <http://hal.inria.fr/inria-00512632/en>.
- [26] P. CUXAC, J.-C. LAMIREL, M. GHRIBI. *Les méthodes de classification non supervisées appliquées aux textes : mesure de la performance des résultats de clustering de documents*, in "Association Canadienne des Science de l'Information - ACSI 2010", Canada Montreal, 2010, <http://hal.inria.fr/inria-00535941/en>.
- [27] A. DENIS. *Generating Referring Expressions with Reference Domain Theory*, in "INLG 2010", Ireland Dublin, July 2010, p. 27-35, <http://hal.inria.fr/hal-00502414/en>.
- [28] A. DENIS. *Reference reversibility with Reference Domain Theory*, in "Proceedings of the 11th annual SIGdial Meeting on Discourse and Dialogue - SIGDIAL 2010", Japan Tokyo, September 2010, <http://hal.inria.fr/hal-00516998/en>.
- [29] A. DENIS, L. M. ROJAS BARAHONA, M. QUIGNARD. *Extending MMIL Semantic Representation: Experiments in Dialogue Systems and Semantic Annotation of Corpora*, in "Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation isa-5", China Hong-Kong, 2010, <http://hal.inria.fr/hal-00481868/en>.
- [30] J. DUCLOY, T. DAUNOIS, M. FOULONNEAU, A. HERMANN, J.-C. LAMIREL, S. SIRE, J.-P. THOMESSE, C. VANORBEEK. *Metadata for Wicri, a network of semantic Wikis for communities in research and innovation*, in "International Conference on Dublin Core and Metadata Applications - DC-2010", United States Pittsburgh, 2010, <http://hal.inria.fr/inria-00535962/en>.

- [31] I. FALK, S. CRUZ-LARA, N. BELLALEM, T. OSSWALD, V. HERRMANN. *Multilingual Lexical Support for the SEMbySEM project.*, in "LREC Workshop Language Resource and Technology Standards", Malta La Valetta, May 2010, <http://hal.inria.fr/inria-00509763/en>.
- [32] I. FALK, C. GARDENT. *Bootstrapping a Classification of French Verbs Using Formal Concept Analysis.*, in "Interdisciplinary Workshop on Verbs", Italy Pisa, November 2010, 6, <http://hal.inria.fr/hal-00521268/en>.
- [33] I. FALK, C. GARDENT, A. LORENZO. *Using Formal Concept Analysis to Acquire Knowledge about Verbs*, in "7th International Conference on Concept Lattices and Their Applications - CLA 2010", Spain Sevilla, October 2010, 12, <http://hal.inria.fr/hal-00516789/en>.
- [34] D. FRÉARD, A. DENIS, F. DÉTIENNE, M. BAKER, M. QUIGNARD, F. BARCELLINI. *The role of argumentation in online epistemic communities: the anatomy of a conflict in Wikipedia*, in "European Conference on Cognitive Ergonomics - ECCE 2010", Netherlands Delft, August 2010, p. 91-98, <http://hal.inria.fr/hal-00516994/en>.
- [35] C. GARDENT, C. CERISARA. *Semi-Automatic Propbanking for French*, in "TLT9 - The Ninth International Workshop on Treebanks and Linguistic Theories", Estonia Tartu, November 2010, <http://hal.inria.fr/inria-00537148/en>.
- [36] C. GARDENT, B. GOTTESMAN, L. PEREZ-BELTRACHINI. *Comparing the performance of two TAG-based surface realisers using controlled grammar traversal*, in "Coling 2010: Posters", China Beijing, August 2010, <http://hal.inria.fr/inria-00537017/en>.
- [37] C. GARDENT, A. LORENZO. *Identifying Sources of Weakness in Syntactic Lexicon Extraction*, in "The seventh international conference on Language Resources and Evaluation - LREC'10", Malta Malta, N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER, D. TAPIAS (editors), European Language Resources Association (ELRA), November 2010, ISBN : 2-9517408-6-7, <http://hal.inria.fr/inria-00537150/en>.
- [38] C. GARDENT, L. PEREZ-BELTRACHINI. *RTG based surface realisation for TAG*, in "Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)", China Beijing, August 2010, p. 367-375, <http://hal.inria.fr/inria-00537159/en>.
- [39] M. GHRIBI, P. CUXAC, J.-C. LAMIREL, A. LELU. *Mesures de qualité de clustering de documents : prise en compte de la distribution des mots clés.*, in "Évaluation des méthodes d'Extraction de Connaissances dans les Données- EvalECD'2010", Tunisia Hammamet, N. BÉCHET (editor), Fatiha Saïs, January 2010, 14, <http://www.lirmm.fr/~bechet/EvalECD2010/Actes.html>, <http://hal.inria.fr/inria-00442953/en>.
- [40] G. JACQUET, J.-L. MANGUIN, F. VENANT, B. VICTORRI. *Construction dynamique du sens : application à la prédication verbale*, in "Rochebrune 2010", France Rochebrune, 2010, <http://hal.inria.fr/hal-00526804/en>.
- [41] J.-C. LAMIREL, Z. BOULILA, M. GHRIBI, P. CUXAC, C. FRANÇOIS. *A new incremental growing neural gas algorithm based on clusters labeling maximization: application to clustering of heterogeneous textual data*, in "23rd International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA-AIE 2010)", Spain Cordoba, 2010, <http://hal.inria.fr/inria-00535942/en>.
- [42] J.-C. LAMIREL, Z. BOULILA, M. GHRIBI, P. CUXAC, C. FRANÇOIS. *A new incremental neural clustering approach for performing reliable large scope scientometrics analysis*, in "6th International Conference on

- Webometrics, Informetrics and Scientometrics and 11th COLLNET Meeting", India Mysore, 2010, <http://hal.inria.fr/inria-00535967/en>.
- [43] J.-C. LAMIREL, Z. BOULILA, M. GHRIBI, P. CUXAC, C. FRANÇOIS. *Un nouvel algorithme incrémental de gaz neuronal croissant basé sur des l'étiquetage des clusters par maximisation de vraisemblance : application au clustering des gros corpus de données textuelles hétérogènes*, in "Technological and Strategic Survey Conference – VSST 2010", France Toulouse, 2010, <http://hal.inria.fr/inria-00535968/en>.
- [44] J.-C. LAMIREL, P. CUXAC, C. HAYDAR, C. FRANÇOIS. *A new general paradigm for mining research topics evolving over time*, in "3rd International Conference on Information Systems and Economic Intelligence - SIIE'2010", Tunisia Sousse, 2010, <http://hal.inria.fr/inria-00535938/en>.
- [45] J.-C. LAMIREL, M. GHRIBI, P. CUXAC. *Exploitation d'une mesure de micro-précision cumulée non supervisée pour l'évaluation fiable de la qualité des résultats de clustering*, in "42èmes Journées de Statistique", France Marseille, Société Française de Statistique (SFdS), 2010, <http://hal.inria.fr/inria-00494723/en>.
- [46] J.-C. LAMIREL, M. GHRIBI, P. CUXAC. *Micro rappel précision non supervisés : vers de nouvelles mesures de qualité de clustering*, in "XVIIèmes Rencontres de la Société Francophone de Classification- SFC'10", France Saint-Denis de La Réunion, 2010, <http://hal.inria.fr/inria-00535940/en>.
- [47] J.-C. LAMIREL, M. GHRIBI, P. CUXAC. *Unsupervised recall and precision measures: a step towards new efficient clustering quality indexes*, in "19th International Conference on Computational Statistics - COMPSTAT'2010", France Paris, 2010, <http://hal.inria.fr/inria-00535961/en>.
- [48] J.-C. LAMIREL, G. SAFI. *Use of distance-based indexes might well lead to misinterpretation of clustering quality results*, in "1st International Workshop on Validation Statistics", Germany Berlin, 2010, <http://hal.inria.fr/inria-00535963/en>.
- [49] J.-C. LAMIREL, G. SAFI, N. PRYANKAR, P. CUXAC. *Mining research topics evolving over time using a diachronic multi-source approach*, in "The Fourth International Workshop on Mining Multiple Information Sources - ICDM 2010", Australia Sydney, 2010, <http://hal.inria.fr/inria-00535969/en>.
- [50] C. LARIZZA, M. GABETTA, L. M. ROJAS BARAHONA, G. MILANI, E. GUASCHINO, G. SANCES, C. CEREDA, R. BELLAZZI. *Extraction of Clinical Information from Clinical Reports: an Application to the Study of Medication Overuse Headaches in Italy*, in "AMIA Summit on Translational Bioinformatics", United States San Francisco, 2010, <http://hal.inria.fr/inria-00519826/en>.
- [51] F. TANTINI, C. CERISARA, C. GARDENT. *Memory-Based Active Learning for French Broadcast News*, in "INTERSPEECH 2010", Japan Tokyo, September 2010, p. 1377-1380, <http://hal.inria.fr/inria-00540423/en>.
- [52] F. VENANT. *Meaning representation: from continuity to discreteness*, in "Seventh international conference on Language Resources and Evaluation (LREC)", Malta Malte, 2010, <http://hal.inria.fr/hal-00526811/en>.
- Scientific Books (or Scientific Book chapters)**
- [53] C. ARECES, P. BLACKBURN. *Special Issue on Hybrid Logics*, Elsevier, 2010, vol. 8, Special Issue of the Journal of Applied Logic. C. Areces and P. Blackburn (editors)., <http://hal.inria.fr/inria-00549320/en>.

- [54] J.-C. LAMIREL. *A new multi-viewpoint and multi-level clustering paradigm for efficient data mining tasks*, INTECH Open Access Publisher, 2010, <http://hal.inria.fr/inria-00535971/en>.
- [55] J.-C. LAMIREL. *Vers une approche systémique et multivues pour l'analyse de données et la recherche d'information : un nouveau paradigme*, 2010.
- [56] J.-C. LAMIREL, P. CUXAC, C. FRANÇOIS. *Recherche des évolutions technologiques à partir de bases de données bibliographiques : apport de la classification incrémentale*, in "Technologies de la Connaissance et Recherche d'Information en Contexte", S. COLLECTION (editor), Hermes Science Publishing Ltd, 2010, <http://hal.inria.fr/inria-00535972/en>.

Research Reports

- [57] A. DENIS, M. AMOIA, L. BENOTTI, L. PEREZ-BELTRACHINI, C. GARDENT, T. OSSWALD. *The GIVE-2 Nancy Generation Systems NA and NM*, INRIA Nancy Grand Est, July 2010, <http://hal.inria.fr/inria-00541578/en>.

Other Publications

- [58] D. SUSTRETOV. *Topological semantics for hybrid logic*, July 2010, <http://hal.inria.fr/inria-00541569/en>.

References in notes

- [59] 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.
- [60] C. ARECES, R. BERNARDI. *Analyzing the Core of Categorical Grammar*, in "Journal of Logic, Language and Information", 2004, vol. 13, n^o 2, p. 121–137.
- [61] P. BLACKBURN, J. BOS. *Computational Semantics*, in "Theoria", Jan 2003, vol. 18, n^o 46, p. 27-45.
- [62] P. BLACKBURN, J. BOS. *Representation and Inference for Natural Language: A First Course in Computational Semantics*, CSLI Press, Aug 2005.
- [63] 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.
- [64] 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.
- [65] B. CRABBÉ, B. GAIFFE, A. ROUSSANALY. *Representation et gestion de grammaires d'arbres adjoints lexicalises*, in "Traitement Automatique des Langues", Dec 2003, vol. 44, n^o 3, p. 67-91.
- [66] 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.

-
- [67] B. CRABBÉ, D. DUCHIER. *Metagrammar Redux*, in "Proc. of the International Workshop on Constraint Solving and Language Processing - CSLP 2004, Copenhagen, Norway", Sep 2004.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] 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.
- [72] C. GARDENT, H. MANUÉLIAN. *Cratation d'un corpus annot pour le traitement des descriptions dfinies*, in "Traitement automatique des langues", 2005, vol. 46, n^o 1.
- [73] 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.
- [74] 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.
- [75] 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, Springer, Apr 2005, vol. 3492, p. 131-146.
- [76] C. GARDENT, K. STRIEGNITZ. *Generating Bridging Definite Descriptions*, in "Computing Meaning", H. BUNT, R. MUSKENS (editors), Studies in Linguistics and Philosophy Series, Kluwer Academic Publishers, Dec 2003, vol. 3.
- [77] 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.
- [78] H. MANUÉLIAN. *Annotation des descriptions dfinies : le cas des reprises par les rles thmatiques*, in "Rencontre des tudients Chercheurs en Informatique pour le Traitement Automatique des Langues - RECITAL'2002, Nancy, France", Jun 2002.
- [79] 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.
- [80] H. MANUÉLIAN. *Descriptions Dfinies et Dmonstratives : Analyses de corpus pour la gnration*, Universit de Nancy 2, Nov 2003.

-
- [81] H. MANUÉLIAN. *Generation de descriptions définies et démonstratives*, in "Huitième Atelier des doctorants en linguistique - ADL'2003, Paris, France", Jul 2003.
- [82] H. MANUÉLIAN. *Une analyse des emplois du démonstratif en corpus*, in "Traitement Automatique des Langues Naturelles - TALN 2003, Batz sur Mer, France", Jun 2003.
- [83] H. MANUÉLIAN. *Generating Coreferential Descriptions from a Structured Model of the Context*, in "Proc. of the International Conference on Language Resources and Evaluation - LREC 2004, Lisbon, Portugal", May 2004.
- [84] 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.
- [85] 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.
- [86] 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 Computational Linguistic - LACL'05, Bordeaux, France", Apr 2005.
- [87] 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.
- [88] 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.
- [89] K. STRIEGNITZ. *Génération d'expressions anaphoriques. Raisonnement contextuel et planification de phrases*, Université Henri Poincaré, Nov 2004.