INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *Project-Team L&D*

# *Langue et Dialogue*

*Lorraine*

THEME 3A

# Table of contents

# 1. Team

*Langue et Dialogue (Language and Dialogue) is a LORIA project (UMR 7503) common to INRIA, the CNRS, the University of Nancy 1 (Henri Poincaré), the University of Nancy 2, and the National Polytechnic Institute of Lorraine. For more details, we invite the reader to consult the team web site at http://www.loria.fr/equipes/led.*

**Head of project**

Laurent Romary [research director INRIA]

**Vice-head of project**

Patrick Blackburn [research director INRIA]

**Senior research scientists**

Carlos Areces [post-doctoral fellow INRIA from 2003-01-15, research fellow INRIA from 2003-10-01]
Marie-Dominique Devignes [research fellow CNRS (SDV department) until 2003-09-30[1] ]
Bertrand Gaiffe [research fellow CNRS (STIC department)]
Claire Gardent [research fellow CNRS]
Matthieu Quignard [research fellow CNRS (SHS department CR2, section 34)]
Jesse Tseng [research fellow CNRS (STIC department) from 2003-10-01]

**Faculty members**

Nadia Bellalem [assistant professor University of Nancy 2]
Daniel Coulon [professor INPL]
Samuel Cruz-Lara [assistant professor IUT Nancy-Charlemagne University of Nancy 2]
Christine Fay-Varnier [assistant professor "Ecole de Géologie" INPL]
Jean-Luc Husson [assistant professor IUT Saint-Dié-des-Vosges UHP Nancy 1]
Jean-Marie Pierrel [professor UFR STMIA UHP Nancy 1]
Azim Roussanaly [assistant professor UFR MI University of Nancy 2 (on temporary assignment to INRIA until 2003-09-01)]
Malika Smaïl [professor UFR STMIA UHP Nancy 1 until 2003-09-30]

**Affiliated scientists, visiting scientists, and post-doctoral fellows**

Laurence Danlos [affiliated research scientist (Professor University of Paris 7)]
Nancy Ide [visiting scientist (Vassar College, USA)]
Evelyne Jacquey [affiliated research scientist CNRS (STIC department)]
Susanne Salmon-Alt [affiliated research scientist CNRS (ATILF)]
Milena Slavcheva [visiting scientist INRIA from 2003-03-01 to 2003-05-31]
Koichi Takeuchi [post-doctoral fellow INRIA until 2003-04-30]
Amalia Todirascu [visiting scientist (assistant professor of the University of Troyes (UTT))]

**Doctoral students**

Benoît Crabbé [Ph. D. student, INRIA grant ]
Frédéric Landragin [Ph. D. student INRIA until 2003-03-31, project technical staff INRIA from 2003-04-01]
Hélène Manuélian [Ph. D. student UFR SDL University of Nancy 2 (ATER at the University of Metz)]
Thi-Minh-Huyen Nguyen [Ph. D. student UFR STMIA UHP Nancy 1 until 2003-12-31]
Yannick Parmentier [student intern (DEA "Informatique, Automatique et Productique" UTBM) from 2003-02-03 to 2003-07-18, project staff, INRIA grant from 2003-08-01 to 2003-09-30, Ph. D. student UFR STMIA UHP Nancy 1 (INRIA grant - graduate assistant UHP Nancy 1) from 2003-10-01]
Joseph Roumier [project staff INPL (MO) from 2003-09-01 to 2003-10-31, Ph. D. student ENSEM INPL (CIFRE grant financed by CEGELEC) from 2003-11-01]
Djamé Seddah [Ph. D. student, INRIA grant]
Kristina Striegnitz [Ph. D. student (with the University of Saarbrück, Germany)]

**Project staff, technical staff and administrative assistant**

[1] Research activities for both Marie-Dominique Devignes and Malika Smaïl appear in the annual report of the team Orpailleur

Charles Beiss [project technical staff UHP Nancy 1 until 2003-02-28]
Isabelle Blanchard [technician CNRS, administrative assistant]
Travis Choma [project technical staff INRIA from 2003-10-01]
Pierre Dussaulx [project technical staff INRIA]
Ingrid Falk [ [2] project technical staff CNRS]
Laurence Kbida [project staff INRIA (MO) from 2003-01-01 to 2003-03-31, project technical staff INRIA from 2003-08-01]
Eric Kow [project technical staff INRIA until 2003-09-24]
Ashwani Kumar [project technical staff INRIA until 2003-08-31]
Satyendra Kumar Gupta [project technical staff INRIA from 2003-11-13]
Erica Meena [project technical staff INRIA from 2003-01-01]
Etienne Petitjean [project technical staff until 2003-11-30]
Annalisa Ricci [project technical staff INRIA]
Nadia Viscogliosi [technical staff CNRS (attached to SeDRe) until 2003-08-31]

**Student interns**

Shireesh Annam [student intern, INRIA grant (Indian Institute of Technology (Kampur)) from 2003-05-05 to 2003-07-21]
Hervé de Palma [student intern (IUT Charlemagne Nancy 2) from 2003-04-07 to 2003-07-25]
Laurent Degardin [student intern UHP Nancy 1 (DESS Texte) from 2003-05-05 to 2003-08-31]
Alexandre Denis [student intern INRIA (ESIEA) from 2003-05-05]
Nicolas Dubois [student intern University of Nancy 2 (Cognitive science degree UFR Maths/info) from 2003-06-23 to 2003-08-23]
Julien Ducret [student intern INRIA (UFR Maths/Info Nancy 2) from 2003-04-01 to 2003-09-30]
Jonathan Faivre-Vuillin [student intern (DUT) from 2003-04-07 to 2003-07-31]
Jean-Charles Fellet [student intern (ESIAL) from 2003-08-01 to 2003-09-07]
Guillaume Georges [student intern (ITEA, Paris) from 2003-01-04 to 2003-02-16]
Sébastien Guérin [student intern INRIA (DESS "Industrie de la Langue", University of Metz) from 2003-04-01 to 2003-07-31, project staff INRIA (MO) from 2003-08-01]
Altaf Hussain [student intern (Imperial College, Londres) from 2003-04-01 to 2003-05-31]
Laurence Mullet [student intern (DEA "Sciences du Travail et de la Formation" Nancy 2) from 2003-04-01 to 2003-08-30]
Nicolas Neyrinck [student intern (DUT of computer science IUT Saint-Dié-des-Vosges) from 2003-04-07 to 2003-06-14]
Yvan Norsa [student intern UHP Nancy 1 (Computer science degree) from 2003-07-07 to 2003-09-07]
Loïs Rigouste [student intern DEA from 2003-06-09 to 2003-08-14]
Sabyasachi Roy [student intern, INRIA grant (Indian Institute of Technology (Kampur)) from 2003-05-01 to 2003-07-31]

# 2. Overall Objectives

## 2.1. Discussion

The introduction of natural language and speech into a man-machine interface requires extremely robust systems for language understanding and interpretation, systems which are well integrated into the application. The objective of the "Langue et Dialogue" (Language and Dialogue) team is to define and implement man-machine communication systems with robust and reliable language components. Our activities are expanding in three complementary directions:

[2] The technical activities of Ingrid Falk are more globally related to the State-Region program and are conducted under the joint supervision of Loria teams ("Langue et Dialogue", Orpailleur, Calligramme, Read, ATILF and INIST)

- Study of the **fundamental mechanisms of communication in natural language**, whether alone or accompanied by gestural designation (multimodal dialogue). This research is being undertaken in a cross-disciplinary context that draws on ideas from linguistics, logic, computer science, and cognitive science;
- Implementation of **effective dialogue systems**, particularly in the context of wider collaborations. This work provides us with experimental platforms for testing the various models that we design;
- Design of **generic methods and tools** for the refined study of real dialogue situations. These are obtained by transcribing experiments and simulations or by direct observation. This work is based on the experience acquired over a number of years on the standardization and manipulation of linguistic resources (texts, transcription of spoken data, lexica).

# 3. Scientific Foundations

## 3.1. Theoretical foundations

How does human language work? What are the processes involved when two people have a dialogue? How do they understand each other? Is it possible to model such interactions on a computer and, if so, what are the methods and tools we need? Such questions raise many scientific problems, such as how the sound systems of human languages function (phonology), and what is involved in the organization of grammatical structure (syntax). In the "Langue et Dialogue" Team, we are particularly interested in the semantic and pragmatic aspects of human language and with the computational problems they give rise to.

From a scientific point of view, It is an appropriate time to tackle semantic and pragmatic issues. Many fundamental problems of phonological and syntactic analysis have been satisfactorily solved, and there are now a wide range of tools which allow us to deal with speech and syntax automatically. To put it another way, semantics and pragmatics are the new frontier in computational linguistics, and developments in neighbouring disciplines make it plausible that progress can be made in these new areas. For a start, research in artificial intelligence has led to the creation of representation formalisms (such as description logics, temporal logics, and epistemic logics) and efficient reasoning tools for computing with them. Moreover, the automated reasoning community has made huge strides over the last decade in developing efficient tools for working with first order logic, and in particular, first order logic with equality. Furthermore, over the same period, research in formal semantics has given rise to sophisticated representation formalisms (such as DRT, SDRT, dynamic semantics, and various kinds of formalisms for semantic underspecification) which are particularly well adapted to the representation of semantic phenomena as encountered in human languages (such as anaphora, ambiguity, and presupposition) and the link between these formalisms and classical logic is well understood.

The orientation of "Langue et Dialogue" towards natural language semantics and pragmatics is based on such developments. The research program of the team is to develop and use logical, linguistic, and computational tools to tackle different problems in natural language. We place heavy emphasis on two themes:

- the importance of inference within natural language semantics and pragmatics, and
- language use in rich environments, particularly multimodal settings.

The approach taken by "Langue et Dialogue" is to examin such problems from the viewpoint of recent theories, to develop experimental implementations which allow us to test such theories and, finally, to link our research with empirical data.

## 3.2. Dialogue and cognition

Any dialogue situation reveals something of the complex relationships between language, cognition, and context. The approach adopted by the "Langue et Dialogue" team is to study and model such features in situations involving both humans and machines. In particular, both Man-Machine Dialogues (MMD), in which the machine is treated as an interlocutor, and Computer-Mediated Dialogues (CMD), in which human participants communicate with each other using machines as a medium, are important to our work. We could characterise this approach as belonging to the field of computational pragmatics, but with a cognitive orientation (that is, we take into consideration cognitive theories such as [52] or Relevance Theory [53]). This is because we believe that the design of dialogue systems should be guided by cognitive insights into human communication and collaboration.

Concerning Man-Machine Dialogue proper, our research is focused on multimodality: that is, communication that integrates communication channels besides speech and text. Speakers in dialogue situations are always situated at some particular place, over some particular period of time, and have a mutual awareness of their environment. The exploitation of such information in MMD (for example, via the use of gestures) could enable users to communicate more efficiently. This efficiency is present in human communication (which is naturally multimodal), and the design of dialogue systems should integrate it. Some contemporary semantic theories have started to integrate those factors (in particular, SDRT [54]), and ultimately we hope that research at "Langue et Dialogue" will lead to an understanding at an even deeper level. In particular, we aim to give a unique description for different information sources, in order to develop an interpretative model that will be independant of the modalities being used.

The research carried out on human computer-mediated dialogues also explores the relationships between language and cognition, but in a way that is the reverse of that found in MMD. In CMD we are not concerned to develop efficient dialogue systems inspired by human cognitive skills, rather we aim to study human cognition by means of dialogue systems. When we implement experimental communication interfaces for human dialogues, we want them to reveal which factors promote (or prevent) various forms of collaboration in dialogue. This pluridisciplinary research is based on the implementation of shared communication interfaces and the elaboration of methods for analysing collaboration in dialogues. Such ideas can be applied, for example, to the design of computer-supported collaborative learning environments, where the main objective is to support the development of certain cognitive processes from the interaction (for example, learning from processes like explanation, argumentation, confrontation, reformulation, and so on; see [19]).

## 3.3. 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 the simplest levels at which inference enters natural language is as a disambiguation mechanism. Utterances in natural language are typically extremely ambiguous: it is inference that 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 integratation 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, in effect, such work redefines the boundary between semantics and pragmatics. For example, the van der Sandt 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 "Langue et Dialogue" 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 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.

## 3.4. Linguistic engineering

The implementations developed by the team have an important role to play, a role that goes well beyond the exploration of new applications: they also allow us to gain insight into the complexity of real world semantic and pragmatic phenomena and how they can be managed. They enable us to experiment with the interactions between different data sources (for example, between the lexical content and the ontology characteristic of an application domain).

The team approach is to develop open platforms integrating state of the art components and tools. This approach makes heavy use of standard protocols (for example SOAP) and standardly specified data (for example, for the representation of multimodal content). By iterating this integration process we hope to develop well motivated systems that reflect the latest linguistic, logical, and computational insights. This approach depends on the existence of international standards for linguistic data representation, and this explains the deep involvement of the "langue et Dialogue" team in this area.

## 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 15 years. The "Langue et Dialogue" team has 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 (in the SILFIDE project) or in the definition of standardized representation formats (for example, TAGML). Our works in this area goes hand-in-hand with our scientific projects both at the level of reusable lexical data descriptions (for example for the parametrization of Man-machine dialogue systems), the study of fine-grained phenomena in semantics (by the elaboration of tests suites), and corpus annotation in the context of studies in analysis or generation (e.g. reference study). Such work can only be significant if the data becomes perene by being encoded in standardized representation frameworks. Thus we recognize the importance of recent standardization initiatives (such as ISO committee TC37/SC4) which are linked to several projects of our team.

# 4. Application Domains

## 4.1. Syntactic Analysis

In order to cope with dialogue management (and in particular, reference resolution), we need automatic parsing tools. For several years now, the "Langue et Dialogue" team has worked on Tree Adjoining Grammars (TAGs). We work with TAGs the following reasons :

- the syntax/semantic interface is simpler in TAGs than in context free grammars, thanks to the extended locality domain provided by TAGs ;
- however the worst case complexity for TAG parsing remains polynomial (O(n6)).

We thus develop and maintain a TAG parser together with a morpho-syntactic annotator. It appears, however, that the main difficulty lies within the management of grammars. We have therefore worked on tools dedicated to managing and debugging grammars. In particular, the parser contains tools to:

- visualize the trees,

- test the anchoring of trees by a lexicon,
- visualize the results of parsing.

We also focus on abstract and compact representations of grammars. We have developed a meta-grammar compiler that allows linguists to hierarchically describe syntactic phenomena. These descriptions are then compiled into a TAG grammar.

The first version of this tool enabled us to manage our grammar, and has since given rise to a second version (developed in conjunction with the Calligramme project). In the new version, a novel mechanism (based on so-called "colored node") has been designed to enable the combination of sub-trees.

Of course, obtaining syntactic trees is not an aim in itself: it is only a step on the way to obtaining semantic representations. Accordingly, we also work on the syntax/semantic interface, and we do so following two approaches:

- one approach builds on the work of Claire Gardent and Laura Kallmeyer and makes use of derived trees. In this approach, the meta-grammar yields trees enriched with a logical formula that is coindexed with feature structures on the nodes of the trees;
- the other approach uses the parsing forest extracted from the chart of the parser and is designed to work simultaneously on the derivation and the derived trees.

## 4.2. Generation

In many computer applications, the data produced by the system can be complex and difficult to interpret by a non-expert. Using a text generator, which takes as input this data and produces text in natural language that expresses the content of the data, is one possible way to address this problem.

The generation task can be thought of as a complex constraint problem. A good solution to this problem is a natural language text which minimizes the violations of the various interacting linguistic constraints. For instance, given some entity e_1 to be described, the choice of the words and of the linguistic form used to describe that entity will depend on the information status of e_1 (for example, whether it has been mentioned in the previous discourse, and whether it is known to the hearer only or to both the speaker and the hearer). It will also depend on the semantics of the other words used in the surrounding linguistic context, and on world knowledge. The degree to which we can fulfil these constraints will affect the quality of the generated output, which will in turn determine how well we fulfil the initial communicative goal.

The generation of good quality texts also depends on a close interaction between knowledge representation, generation, and inference. This interaction is needed, for instance, to verify that presuppositions in the generated text are satisfied by the context of enunciation; to decide which are the best words expressing a certain concept in a given context (the lexicalization problem); or to produce "cooperative" responses, that is, responses that go beyond a simple "yes/no" that can repair misconceptions on the part of the user, for example by providing a partial answer when a full answer to the question is not possible.

The interaction between generation and inference constitute one of the research themes of "Langue et Dialogue" and lies at the heart of the scientific aims of the two projects InDiGen and GenI.

The first project (InDiGen, 2000-2004) deals with (i) the use of constraints and constraint programming for the generation of definite nominal groups and (ii) the use of theorem provers for the generation of associative anaphora. This research led to the development of a generator (implemented in the constraint based programming language MoZart/Oz) that interfaces a sentence planner with theorem provers for first-order logic (Spass, Otter) and for description logics (Racer) The InDiGen generator is available on the Web at http://www.coli.uni-sb.de/cl/projects/indigen/demo.

The second project (GenI, 2002-2003) involves the interaction of several French teams and aims to create resources and an architecture for the generation of French texts and to test this architecture on the problems of lexicalization, cooperative responses and associative anaphora.

## 4.3. Inference and speech

Intuitively, to perform an inference is to extract implicit information from explicit information. There are many different kinds of inferences. For example, statistical inference uses a large body of schemes found in a corpus in order to determine probable information about a small example. Another typical example of inference is logical inference. For example, if our knowledge base contains explicitly the information that "all contract killers are violent" and that "Vincent is a contract killer," then it implicitly contains the information that "Vincent is violent."

Logical and statistical inference (and, indeed other forms of inference) are important in discourse and dialogue. The "Langue et Dialogue" team is currently investigating the usage of logical inference. In the last years, the performance of automated reasoning tools (that is, software able to handle various kinds of logical inference) has increased considerably. Theorem provers have achieved performance levels that were unthinkable ten years ago. Moreover, the performance of model builders, even though this technology is less advanced than that of modern theorem provers, has achieved a level where they can be used as interesting experimental tools. Crucially, much of this progress in automated reasoning has been for logics which can be used for natural language semantic representation and inference, such as first-order logic with equality, description logics, and hybrid logics.

The most obvious use of logical inference in discourse is disambiguation. Sentences in natural language are frequently highly ambiguous. Indeed, the interaction between lexical ambiguity, syntactic ambiguity and scope ambiguity can produce sentences with hundreds of different interpretations, most of which are absurd given enough background knowledge. Theorem provers and model builders can examine the different interpretations and eliminate those which are incompatible with background knowledge and the previous discourse. Moreover, soemtimes it is also important to check if a sentence provides new information, or if it is redundant given the previous discourse. Again, theorem provers and model builders can be used to accomplish this kind of task. The CURT programs (Blackburn and Boss) constitute a collection of algorithms that illustrate these different possibilities.

The "Langue et Dialogue" team is also concerned with more experimental uses of these technologies. For example, we try to use model builders to "guess" what a given description stands for (by generating the smallest possible model for the described situation). This is an interesting approach to analyzing the role of inferences in the treatment of associative anaphora.

## 4.4. Multimodal dialogue

The design of multimodal systems requires a deep analysis of the way meaning is divided between language and gesture. In the type of MMD that is known as command dialogue, references to objects (how the words that have been produced by the user are linked to the objects of the application world), and references to events (how the words are linked to the corresponding software primitives), are the most important part of the interpretation process. Since we cannot study references to objects without studying how the user visually perceives these objects, we have to simultaneously analyse the role of language, gesture, and visual perception:

- Language: our interest is in referential expressions, and on the other parts of the utterance that intervene in the reference resolution process (for instance, the verb and its semantics);
- Gesture: our aim is to reconstitute a trajectory from the data that are given by a gestural device (mouse, touch screen, haptic device), and to identify the objects that are supposed to be pointed out;

- Visual perception: our interest is in the spatial configuration of the objects that are displayed on the screen, because this configuration has effects on both the form of the trajectory and on the choice of referential expression.

Thus the information that needs to be taken into account in highly heterogeneous: there are linguistic constraints, constraints on gestural trajectories (hopefully as device-independent as possible), and visual scene constraints arising from such notions as salience and perceptual grouping. Considering this heterogeneity, it is unsurprising that most current theories turn out to be ineffective. For a number of years, however, the "Langue et Dialogue" team has developed a model for integrating these diverse forms of information. This model, called the Reference Domains Model, is a unified framework for taking into account the constraints linked to each modality (through the notions of partitions and differentiation criteria). Its objective is to handle a variety of multimodal phenomena, for both comprehension than generation. It is the basis for the ongoing research activities of "Langue et Dialogue" concerning multimodal reference.

A last point should be made about the architecture of MMD systems. The implementation of a multimodal system relies on a set of modules that are dedicated to the realization of specific tasks, such as input treatment, semantic fusion of modalities, and dialogue management. These modules interact, which presupposes a constant flow of information between them, and therefore a homogeneous interface language to represent this information. Such a language has been designed for the MIAMM European project, with the representation of multimodal content as its objective. This language is called MMIL (MultiModal Interface Language) and is currently being updated and reused in the OZONE framework.

## 4.5. Collaborative environments

As a natural application of CMD, the Computer-Supported Collaborative Learning Environments developed in the "Langue et Dialogue" team aims to favor specific learning processes (that is, specific cognitive mechanisms) via dialogue situations whose structure, tasks and tools have been designed for this purpose. Building on our experience in the design of pedagogical situations on the Internet for Collaborative Argumentation-Based Learning (IST European project SCALE: Internet-based intelligent tool to Support Collaborative Argumentation-based LEarning. http://www.euroscale.net) and on the exploitation of the DREW platform elaborated in the same project (Dialogical Reasoning Educational Webtool. http://scale.emse.fr/drew), we are currently trying to improve and extend the functionalities of DREW for learning situations in schools and universities. We work in collaboration with teachers. Our approach is to conduct empirical studies (of pedagogical task sequences with DREW in the classroom; see [32], [33]), to design new modules (*drewlets*) for communication and coordination of specific collaborative tasks (such as collaborative text writing, or argumentative diagrams elaboration), and to give interaction analyses (for example, traces exploitation).

## 4.6. Linguistic and multimedia resources

Working with normalized linguistic resources enables us to validate, by observation, our theoretical models, and also provides us with a generic source of information (for example lexical information) for the prototypes which we develop. Thus we think that it is necessary to actively contribute to the definition of deeper norms in linguistic engineering domain and to participate in the spreading of existing normative frames.

Accordingly, the "Langue et Dialogue" team occupies an active place in the national and international community in the domain of the standardization of linguistic resources and their use. Moreover, the methods developed by "Langue et Dialogue", particularly in the context of the definition of the ISO 16642 standard (which gained 100% support at the DIS stage and has been published on 15 August 2003) seem to be emerging as a basis for the study of semi-structured document classes with similar properties.

Among the developments made in "Langue et Dialogue", we should mention :

- Development of a Man-Machine dialogue corpus completely standardized and with free access (ASILA project);

- Development of methods of annotation and evaluation of reference in dialogue corpora and texts, methods which can be reproduced (ANANAS project);
- Development of a generic model of multimodal content representation for Man-machine dialogue, with the aim of making an international normative contribution;
- Contribution to the coordination of international standardization activities within the TEI consortium and the TC37/SC4 standardization committee;
- Definition of a generic model of XML format specification, which allows to specify at anytime different types of linguistic annotations, and indeed cultural information (experimentations have been made with cinematographic describers (CINERGIE project) and even information about museum displays (in a collaboration with Taïwan).

# 5. Software

## 5.1. API for manipulating Feature Structures

Description: This Java package is dedicated to coping with untyped feature structures. It includes methods for dealing with unification and subsumption of such feature structures. The feature structures admited by the package are purely conjunctive (except for atomic values where disjunctions are possible). Two input/output formats are taken into account: one which closely ressembles the ususal bracketed notation, and another that is XML based. The latter will conform entirely to the forthcoming ISO proposition regarding feature structures. (ISO DIS 24610-1)

Author(s): Bertrand Gaiffe, Azim Roussanaly and Nicolas Dubois.

Contact: Bertrand Gaiffe

## 5.2. CINERGIE

Description: this platform handles the management and diffusion of multilingual documents and multimedia.

All software components have been developed. In the terminal stage of the project, the consortium has begun a deployment, base enrichment and evaluation phase. Moreover, our partners are partners are working on the development (data extension) and the diffusion of the project results via scientific publications and advertising medium:

*** Network level :**

- Relocated server network administration tools
- Tool for multicriteria consultation of public declared data of the network

*** Server level :**

- Users administration, their responsibilities, right access to data
- Databases administration (creation of a restricted model - subset of the reference model -, construction of an editorial committee, management of derivatives)

- Data edition. This is a generic tool, totally parametrized by the abstract model CINERGIE specification.

Web site: http://www.loria.projets/CINERGIE
–>
Contact: Jean-Luc Husson
Project(s): CINERGIE, SALT

## 5.3. Concordancier (Multilingual alignment)

Description: Tool enclosed with XAlign for multilingual alignment (see XAlign).
Web site: http://www.loria.fr/equipes/led/outils/ALIGN/align.html
Author(s): Thi Minh Huyen Nguyen, Sean O'Rourke
Contact: Thi-Minh-Huyen Nguyen
Project(s): Arcade II, Vietnam

## 5.4. CURT

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

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

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

Web site: http://www.comsem.org
Documentation: http://www.comsem.org
Contact: Patrick Blackburn

## 5.5. GenI generator

Description: The GenI generator is a successor of the InDiGen generator. Also based on a chart algorithm, it is implemented in Haskell (one of the leading functional programing languages available nowadays) and aims for modularity, re-usability and extensibility. The GenI generator uses efficient datatypes and intelligent rule application to minimize the generation of redundant sentences. Moreover, the system is "stand-alone" as we use the Glasgow Haskell compiler to obtain executable code for Windows, Solaris and Linux.

Up to now, only a first prototype is available, which links to the description logic prover RACER for the required inference services. The system can process the output of the Metagrammar compiler being developed by Crabbé (Langue et Dialogue) and Duchier (Calligramme).

Version: 1.0
Author(s): Carlos Areces
Contact: Carlos Areces
Project(s): GenI

## 5.6. hGEN, a random formula generator

Description: hGen is a random CNF (Conjunctive Normal Form) generator for sublanguages of Hybrid Logic containing @, the universal modality A, and the downarrow binder. It is essentially an extension to full hybrid

logic of the latest proposal of Patel-Schneider and Sebastiane, which is nowadays considered the standard testing environment for classical modal and description logics.

Author(s): Carlos Areces and Juan Heguiabehere
Contact: Carlos Areces
Project(s): LED-LIT Research Alliance.

## 5.7. HyLoRes, A resolution based theorem prover for hybrid logics

Description: HyLores is a resolution based theorem prover for hybrid logics. 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.

HyLoRes is implemented in Haskell (ca. 3500 lines of code), and compiled with the Glasgow Haskell compiler (thus, users need no additional software to use the prover). We have also developed a graphical Tcl/Tk interface.

Version: 2.0
Web site: http://www.loria.fr/~areces/HyLoRes
Author(s): Carlos Areces, Daniel Gorin and Juan Heguiabehere
Contact: Carlos Areces
Project(s): LED-LIT Research Alliance.

## 5.8. ILD-ISTC (CPER)

Description: One of the goals of the project is to provide access to linguistic and terminological ressources across different software and hardware architectures. We try to achieve this by wrapping or adapting existing software into web services based on existing or emerging standards, like SOAP (an XML based message protocol). We also promote the coding of linguistic data using standards wherever possible, so that they can be easily used by different programs or computers.

The deployed webservices comprise:

- the TAXI service: transforming XML to SVG, SVG to PBM, GIF or JPEG
- the Calligramme service: parsing French
- the LLP2 service: parsing French
- the DILIB service: information extraction
- the Apache-Axis Dynamic Invoker Servlet: dynamically invoke a service from its WSDL
- the segmentation service: segment and tag input text from a MULTEXT word base

Detailed description:

- the TAXI service: transforms XML to SVG according to a specified style, visualises the svg as PBM, GIF or JPEG images. The core programs were developed in the "Langue et Dialogue" group, and we use the Java Servlet technology and SOAP messaging to make it available online.

- the Calligramme service is a syntax analyser for French based on interaction grammars, developped by the Calligramme group in CAML. It's been wrapped into a service using the perl module SOAP::Lite. The parse result can then be turned into SVG and viewed via the TAXI service.
- the LLP2 service: an LTAG parser for French developped in the "Langue et Dialogue" group using the TAGML2 description format: again we use the Java Servlet technology and the standard Java SOAP messaging APIs to turn the core application into a web service. http://loreley.loria.fr:8080/llp2-0.2a/
- the DILIB service: DILIB is a workbench for scientific or technical information and document engineering. Its kernel is a set of tools (C/Unix/XML), developped at the INIST, used for information extraction. The C-routines are wrapped into a perl(SOAP::Lite) web service. http://www.loria.fr/projets/ILD-ISTC/cgi-bin/callDilib.cgi
- the Apache-Axis Dynamic Invoker Servlet: a web application to dynamically build and invoke SOAP RPC calls starting from a WSDL (Web Services Description Language is a W3C recommendation) URL.http://loreley.loria.fr:8080/wsdi-0.1/
- Soon to come: the segmenter service is written in C, using the gSOAP implementation. It builds an automaton from a MULTEXT input file and then segments input lines.

All the services output xml, which can then be used as input for other services or applications (like TAXI, to visualise it).

All the services (except the segmentor) are accessible on the project's server loreley.loria.fr.

Last update: 2003-11-01

Web site: http://www.loria.fr/projets/ILD-ISTC/cgi-bin/webapps.cgi

Author(s): Ingrid Falk

Contact: Ingrid Falk

Project(s): CPER/ILDSTC

## 5.9. InDiGen

Description: The InDigen generator is a chart-based generation system iterfaced to state-of-the-art reasoning tools (e.g., the theorem prover SPASS). The reasoning tools are used to check the restrictions that linguistic expressions may impose on the context.

The system is implemented in Mozart Oz (a constrait programming language developed at the University of Saarbruecken) and, so far, focus on the pragmatic restrictions associated with the use of the definite article (e.g., the referred entity should be given by the discourse context and uniquely identifiable to the hearer).

Web site: http://www.coli.uni-sb.de/cl/projects/indigen/demo/

Contact: Claire Gardent

Project(s): InDiGen

## 5.10. LLP2: LTAG Parser

Description: the LTAG parser is based on the bottom-up algorithm described in the thesis of Patrice Lopez.

Restrictions :

- TIG

Properties :

- FB-TAG
- Resources standardization (TAGML2)
- Analysis of segment graph input
- Possibility of integrating an external segmentator with visualization and grammar debugging tools

Version: 0.2c
Last update: 2003-10-31
Web site: http://www.loria.fr/~azim/LLP2/help/fr/
Download: http://www.loria.fr/~azim/LLP2/help/fr/download.html
Contact: Azim Roussanaly, Djamé Seddah
Project(s): MIAMM, EVALDA/EASY, Ozone, XMiner

## 5.11. MGC (Meta-Grammar Compiler)

Description: MGC is a tool dedicated to TAG grammars management. It builds on the approach described by Marie-Hélène Candito in her PHD thesis, see also [Marie-Helene Candito. 1996. A Principle-Based Hierarchical Representation of LTAGs. In Proceedings of COLING-96, Copenhagen, Denmark.].

MGC enables linguists to describe classes hierarchies containing partial descriptions of trees in a logic which is interpreted on finite trees. The tool provides a compiler that produces a TAG grammar from descriptions that may be graphically edited. It is developped in Java.

Web site: http://www.loria.fr/equipes/led/outils/mgc/mgc.html

Author(s): Bertrand Gaiffe and many others, in particular, Benoit Crabbé, Azim Roussanaly and Kim Gerdes.

Contact: Bertrand Gaiffe
Project(s): RLT

## 5.12. Segment Server

Description: automata-based sentence segmentor.

This tool can be used independently or integrated in a syntactic parser (for example LLP2).

Version: 01
Last update: 2003-10-01
Download: http://www.loria.fr/~azim/
Contact: Azim Roussanaly
Project(s): EVALDA/EASY, MIAMM, Ozone, XMiner

## 5.13. Soapical

The Soapical project aims to improve software reusability by providing tools that help users make effective use of open standards such as XML and SOAP:

1. The Soapmill protocol which can be used for building architectures and complex applications out of reusable components. Soapmill is a non-RPC (Remote Procedural Call) use of the SOAP protocol.

2. The Soapmill implementation for Java demonstrates how the Soapmill protocol could be used to build blackboard architectures. This implementation was used to build a multi-modal dialogue system in the MIAMM project.
3. The Soapmeter debugging tool helps developers monitor the communication between two SOAP nodes. Soapmeter also provides custom message visualisation through the TAXI editor.
4. Soapkit provides a convenient library for working with XML and SOAP in Java. This library is meant to complement SOAP implementations such as Apache Axis.
5. The Soapcaster acts as a mediator for web-services in flexible architectures. It maintains a central table which is regularly updated as it receives new subscriptions (or unsubscriptions) and as it verifies the availability of services. All services receive regular updates of these tables which allows them to know about, to keep track of, and to communicate with each other. Soapcaster is useful for large applications that consist of many, highly similar services.

Web site: http://soapical.sourceforge.net/
Contact: Eric Kow

## 5.14. SoapMMIL

Description: an architecture for multimodal man-machine dialogue:

The MIAMM European project initiated in the "Langue et Dialogue" team the implementation of an architecture dedicated to multimodal man-machine dialogue systems. This architecture has been initially designed for the MIAMM multimedia database interrogation application. Our long-term aim is to provide a generic platform for tests and validation procedures. The extension of the MIAMM architecture for the OZONE European project (for a reservation dialogue application) during 2003 fits well with such concerns. All modules were updated to become more generic, i.e., more application-independent. The resulting architecture groups a wide set of competences from speech recognition to dialogue management, also including syntactic parsing, semantic analysis, multimodal fusion, and contextual interpretation of the user's utterances.

## 5.15. TAXI (The Adjustable XML edItor)

Description: The Adjustable XML edItor (TAXI) helps users work with XML by providing customised visualisations of their data. The visualisations are XML-stylesheets which anybody can modify to their needs. TAXI can also be used as an XML visualisation plugin in other applications.

Version: 1.2.1
Last update: 2002-06-05
Web site: http://taxi.sourceforge.net/
Contact: Eric Kow

## 5.16. Unannoy

Description: the Unannoy project provides tools to help programmers write software which is easier to use and install.

1. Kowey-generic provides a set of templates for small programming projects in single languages. The goal is to provide a way to make highly portable and nicely distributable software that "just works". These templates also help Java developers to adopt better coding habits, as the default build file auto-generates unit test skeletons (via JUnitDoclet).

2. Unannoyment is a small Java library of very small but frequently occuring fragments of code. Unannoyment provides a pre-condition mechanism to encourage defensive programming, functions for string manipulation, and more.

Web site: http://unannoy.sourceforge.net/
Contact: Eric Kow

## 5.17. vnQTAG - QTAG for Vietnamese

Description: QTAG is a probabilistic POS tagger developed by Oliver Mason (University of Birmingham), http://web.bham.ac.uk/O.Mason/software/tagger/

This tagger works on an annoted corpus that permits the computation of POS trigram probabilities and the definition of POS ambiguity classes for each word, with their probabilities. It also contains a predictor to recognize POS tags by morphological analysis.

vnQTAG is a modified version of QTAG for Vietnamese. The major modifications are the suppression of the morphology predictor and the use of a dictionary with POS information for each word. Some minor changes are also made to the data format and the character encoding.

The vnqtag.zip package also contains a tokenizer for Vietnamese texts.

Download: http://www.loria.fr/equipes/led/download/source/vnqtag.zip
Contact: Thi-Minh-Huyen Nguyen
Project(s): Vietnam

## 5.18. XAlign (Multilingual Alignement)

Description: Multilingual text alignment, i.e. the mapping from a text to its translation in another language at a certain granularity level (paragraph, sentence or expression), is one of the essential components of the researche carried out in the field of multilingual information extraction and to answer the more industrial concern of localization. The work undertaken in the team for several years has led to the development of an alignment tool (in Java) based on a technique using the hierarchical structure of documents. The texts are encoded in a XML format reflecting the hierarchy of divisions (recursively), paragraphs and sentences. This tool was tested on many language pairs (in particular Eastern-European languages) within the framework of the Telri project and on a French - Vietnamese corpus. Completely documented within the framework of the DHYDRO project, and accompanied by a multilingual concordancer making it possible to filter the texts according to constraints expressed on linguistic contexts specific to the considered source and target languages, it is now freely distributed.

Web site: http://www.loria.fr/equipes/led/outils/ALIGN/align.html
Author(s): Patrice Bonhomme, Thi Minh Huyen Nguyen, Sean O'Rourke
Contact: Thi-Minh-Huyen Nguyen
Project(s): Vietnam, Arcade II

# 6. New Results

## 6.1. Lexicon management for syntax and semantics

The interest in syntax within the "Langue et Dialogue" team has lead to the implementation of many tools intended for the syntactic analysis of the French ([13] [20]) with lexicalized Tag Adjoining Grammars (LTAG). The implementation non-trivial French grammars in a lexicalized system such as LTAG is highly bound to the development of tools for efficient syntactic lexicon management.

Moreover, one of the "Langue et Dialogue" teams objectives is to develop TAG grammars which allow us to get as a parsing output not merely a syntactic, but also a semantic representation of the input sentence ([14]).

However, in the LTAG context, no existing system of lexicon management allows to manage grammars with a semantic level.

Accordingly, we developed with the Calligramme project ([41], [50], [51]) a representation language that can express the structure of the lexicon in a compact way. This representation fulfils two needs:

1. Practical: it allows a compact and structured coding of the grammar. The compiler automatically expands the compacted representation.
2. Theoretical: this tool is designed to make easier the expression of lexical equivalences (for example, the passive/active equivalence). These relations are frequently formalized by a procedural system of lexical rules. By contrast, the system developed here is purely declarative.

The aim of this work is to create grammars with significant syntactic and semantic coverage that are usable for both analysis and generation.

## 6.2. Generation

In 2003, work in Natural Language Generation (NLG) has concentrated on:

- specifying and implementing a grammar framework supporting the fast prototyping of grammars describing both the syntax and the semantics of NL expressions.
- implementing an NLG system in which surface realization and description logic based inference are tightly integrated and
- identifying and modeling NL phenomena whose generation involve inference

GRAMMAR FRAMEWORK. As stated above, the aim of was to develop a grammar framework supporting the fast prototyping of grammars describing both the syntax and the semantics of NL.[14] provides a first step in that direction by specifying a syntax/semantic interface for Tree Adjoining Grammars; that is, by showing how to specify a TAG in such a way that it describes both the syntax and the semantics of any given natural language constituent. Next, the ideas presented in that paper were implemented by Yannick Parmentier (DEA student) using the TAG metagrammar compiler developed by Bertrand Gaiffe on the one hand and the metagrammar compiler developed by Denys Duchier (CALLIGRAMME) on the other. Currently both compilers are used to develop and test syntactic-semantic TAGs in generation (for instance, in the GenI generator) but also in analysis (e.g., in the OZONE and XMiner projects).

NLG SYSTEM. Within the GenI project, a first prototype of a generator has been implemented in Haskell. This prototype builds on the experience acquired within InDiGen, and aims at providing a flexible tool able to generate text on the basis of reasonably large grammars. The new generator offers direct interfaces with RACER, a facility that will be exploited for implementing lexicalization as a description logic inference service as described in [35].

EMPIRICAL WORK In order to tailor automated reasoning systems to the needs of natural language, specific reasoning tasks must be identified which arise form the linguistic data and can be used to optimize or modify existing reasoners. From the generation perspective, three such tasks were identified namely, the generation of bridging definite descriptions ([17], [27], [18]); the generation of cooperative answers ([43]); and lexicalization ([34]).

For bridging definite descriptions, ([17]) specifies an algorithm showing how surface realization interact with inference during their generation, while ([27]) relates the results of a corpus study aiming at identifying the type of bridging relations use din real data. Related work was carried out within the PhD Thesis of

Helene Manuelian which provides an in depth corpus study of the form and content of both definite and demonstrative description and shows how existing generation algorithms should be modified to account for the variety of data thus encountered. In particular, ([30], [31]) identifies the various knowledge sources used in the production of definite and co-referential demonstrative descriptions. In her PhD thesis, Helene Manuelian uses the results of this corpus study to extend Gardent and Striegnitz's algorithm ([17]) in three directions: first, to generate a greater variety of bridging anaphora by using all involved knowledge sources, second, to generate co-referential definite descriptions integrating different types of information structure (i.e., proportion and distribution of given/new information) and using a structured context; and third, to choose between definite and demonstrative determiner based on syntactic and semantic properties of the produced descriptions.

The work on cooperative answers arose from a cooperation within the INRIA ARC GenI with Farah Bennamara and Patrick Saint-Dizier (IRIT, Toulouse). ([43]) identifies (based on a corpus analysis) the main types of cooperative answers possible and specifies the Description logic inference task implied by the generation of such answers. Further work within a CNRS TCAN project will aim at generalizing, implementing and testing the approach sketched in that first paper.

Finally, ([34]) argues that the lexicalization task in NLG should be viewed as a Description Logic Inference task. This research direction will be pursued further within a CNRS TCAN project with Toulouse, Paris 7 and Paris 3 which bears on the interaction between lexicalization, aggregation and surface realization.

## 6.3. Hybrid and description logics

An important research theme for the "Langue et Dialogue" team is the development of representation formalisms well suited for applications involving natural language processing.

The team is particularly interested in modal-like logics, such as description and hybrid logics. In particular, we are interested in applying modal logics to reason about syntactic structures, in the use of description logics for semantic representations (specially for generation) and also, on theoretical issues and formal properties of hybrid logics.

Our work shows that hybrid logics share more good properties with classical logics than standard modal logics. The paper [49] Tableaux for Quantified Hybrid Logics, Patrick Blackburn and Maarten Marx (in U. Egly and C. Fernmuller, editors, Automated Reasoning with Analytic Tableaux and Related Methods, International Conference, TABLEAUX 2002, Copenhagen Denmark, July/August, Proceedings, 2002, pages 38-52shows that very general completeness results for hybrid logics can be established in a constructive manner by a systematic translation between classical and hybrid tableaux. The paper [48] "Constructive Interpolation in Hybrid Logic", by Patrick Blackburn and Maarten Marx (in Journal of Symbolic Logic volume 68, pages 463-480, 2003), provides a constructive proof of interpolation for a wide class of hybrid logics. This is probably the most general constructive result concerning interpolation for modal-like logics.

Areces and Bernardi have investigated extensions of categorial logics (used as a formalism for the specification of lexical categories and of the rules that govern their interactions) with hybrid logical machinery ([38]). They have shown that the addition of hybrid operators to the base categorial logic enhanced its expressive power and allows the specification of properties leading, for example, to a proper treatment of quantifier phrases. Blackburn, Gaiffe, and Marx investigate the relationship of three modal languages proposed in the literature, and establishes precise complexity bounds for two of them.

The work of Areces et al. has investigated the use of resolution based techniques (the standard approach used by state-of-the-art first order theorem provers) for hybrid logics. In previous work was basic calculus developed, and recently it has been optimized by the use of ordering and selection functions, a key optimization also characteristic of first-order resolution based theorem provers.

In addition to these theoretical results, we are also working on implementations and applications. The theoretical results provide the basis for the development of automated theorem provers for modal-like logics:

- HyLoRes is the first prototype of a resolution based prover for hybrid logics (and indeed, one of the first examples of a resolution based prover directly applied to modal-like languages). To

test empirically the performance of HyLoRes, the random generator of hybrid formulas hGen was developed ([37]).

- In the related area of description logics (DL), a system has been developed that permits the extension of standard DL provers like RACER to permit inference in terms of Boolean A-Boxes ([44]). The system is capable of translating ALC assertions into an equisatisfiable knowledge base in a language that can be handled by modern provers. Extensive testing of this approach has been also performed. In collaboration with Carsten Lutz, Ulrike Sattler and Ian Horrocks we have also investigated the impact of the addition of "key boxes" to description logics with concrete domains ([39]).

Most of the research described in this section was developed in collaboration with the Language and Inference Technology (LIT) group, ILLC, University of Amsterdam. The LED-LIT collaboration is supported by INRIA's program of Equipes Associees.

In September 2003, Langue et Dialogue hosted an event which brought together many of the research themes we have just discussed: Nancy Inference Week (NIW), 22-26 September 2003. The first two days of NIW were taken up by the Methods for Modalities Workshop (M4M-3). Six invited and thirteen submitted papers on modal, hybrid and description logic were presented over this two day period, and it is likely that a special issue of a journal devoted to this event will appear. Wednesady 24 September was Description Logic Day (DLD). Six lectures devoted to description logic and its applications in natural language were presented. This event was designed to mark the collaboration between Langue et Dialogue and the University of the Saarland on research on description logic for natural language. The last two days of NIW were devoted to the Inference in Computational Semantics Workshop (ICoS-4). During these two days, three invited and fourteen submitted papers on computing inference problems for natural language were presented. NIW was partly designed in collaboration with the Language and Inference Technology (LIT) group, ILLC, University of Amsterdam. The LED-LIT collaboration is supported by INRIA's program of Equipes Associees. In addition to these three events, during NIW Langue et Dialogue also hosted the meeting of the SIGSEM working group on multimodal semantic representation. All in all, NIW reflected the wide variety of language and logic related work carried out at Langue et Dialogue.

## 6.4. Multimodal dialogue

**Extension of the Reference Domain Model**

Concerning referring to objects in a multimodal context, the main new result is the extension of the reference domain model, that consists in taking into account the specificities of visual perception and gesture. The PhD dissertation of Frédéric Landragin [12] sums up these aspects (see also [21]).

**Salience Modelling**

This research field deals with the notion of salience: every linguistic or multimodal utterance is built around salient entities, providing some points of departure for interpretation or generation. A really comprehensive dialogue system must be able to identify these entities, and thus to include a model of salience. The PhD dissertation of Frédéric Landragin [12] (see also [22]) provides such a model, with an integration of visual and linguistic factors.

**Taking haptic dimension into account in man-machine interaction**

One of the new research themes of the team is about taking haptic gesture into account in multimodal dialogue. The aim of Joseph Roumier's DEA training course was to develop on the laboratory haptic platform an application to show the main difficulties of integrating this new dimension. This work has made clear the difficulties which arise due to the strong correlation between visual perception and tactile perception, which shows the necessity of the use of virtual reality principles. The other conclusion of this work is that it is still too soon to integrate the sense of touch in Man-machine communication. Future work should should be based on cooperation with users groups in order to find situations where the use of haptics makes sense.

**Update of the common architecture of MIAMM and OZONE projects**

Concerning the multimodal dialogue systems architecture and the exchanges between the modules, a module dedicated to gestural trajectory recognition has recently been added, in the framework of the OZONE European project (http://www.extra.research.philips.com/euprojects/ozone). The MMIL language has also been used and updated.

## 6.5. Collaborative environments

Within the new application domain on collaborative environments, we focussed on the introduction of such technologies for learning philosophy.

Basically, teaching and learning philosophy remain mostly attached to a single learning and evaluation method that is the philosophical dissertation. Spoken activities (dialogues) are presented as valuable alternatives, since they promote students active participation and critical thinking (see [55] Michel Tozzi, "L'oral argumentatif en philosophie", CRDP Montpellier, 1999). Collaborative environments could be a means to favour the introduction of such dialogical practices in classrooms: they provide some facilities to organize and manage (silent) debates or collaborative activities (collaborative text writing) and some tools to exploit or evaluate them by processing their traces. Such a situation has been designed and evaluated by Laurence Mullet during her DEA work in "Sciences du Travail et de la Formation", Université Nancy 2, 2003.

## 6.6. MMIL, a multimodal content description language

In the context of the MIAMM project, we developed the MMIL language (Multimodal Interface Language), to allow various components of a dialogue system architecture to exchange information in an uniform way. This language, implemented as a XML schema, has been defined with specification methodology introduced in the context of more general works on the modelling of linguistic resources. This work integrate with a wider international initiative, which, under the aegis of the SIGDIAL group of ACL (Association for Computional Linguistic), launched by Harry Bunt (Tilburg Univ.), Kiyong Lee (KAIST) and Laurent Romary, aims to define a general context of multimodal content representation.

## 6.7. Linguistic resources

The thesis of Hélène Manuélian is based on a study of an annotated corpus. The aim of the annotation is to isolate all defined and demonstrative syntagms, in order to comparatively study anaphoric and coreferential use.

The procedure used to prepare the corpus was the following: we took the PAROLE Corpus, given by ATILF in the context of CPER, which is annotated at the morpho-syntactical level in the Multext format (SGML). It is composed of articles from Monde and includes 65 000 words, including 10 000 defined and demonstrative nominal syntagms.

The corpus was converted to the UIF format, adapted to be used with G-search, a tool identifying syntactic structures in non-parsed corpus, which allowed us to tag defined and demonstrative nominal groups in the corpus. Once the nominal groups were tagged, we converted the corpus in XML format, with required tags input for MMAX, a free software specifically designed for referential expressions annotation. The defined and demonstrative nominal groups annotation was done in XML, which allowed us to extract results by XSL stylesheets. They produced HTML tables in which results were sorted in order to treat them easily.

The study of the corpus allowed us to establish a new use of defined and demonstrative descriptions, to re-annotate the corpus with this new classification, and to extend existing algorithms of referential expressions generation.

The scripts, stylesheets and the applied procedure during the study have been made in order to be reused with this corpus or any other XML one, tagged with the same sketch.

Implementation of a service of layout of encoded dialogue transcription with the TEI guidelines (XML to PDF). http://loreley.loria.fr/~quignard/index.html

## 6.8. TMF (ISO 16642)

ISO (http://www.iso.org) has announced on 15 August 2003 the publication of ISO 16642 (Terminological Markup Framework), which aims at providing a platform for the interchange of computerized lexical data, as used in many kinds of applications (translation, technical writing, education etc.). This standard is the result of a work initiated 3 years ago in the context of the IST/SALT project in collaboration with two American teams (Univ. Of Kent and BYU) and more recently with the support of the INRIA national action Syntax. This standard describes means to specific application specific formats (TML - Terminological Markup Languague) while preserving the interoperability between these.

# 8. Other Grants and Activities

## 8.1. International level

### *8.1.1. CommonRefs*

Description: This project aims to integrate complementary work that has been developed recently by the partners on the problem of computational treatment of reference and co-reference. The Brazilian part of this project has developed work that is related to reference/co-reference in textual discourse for English and Portuguese. The French part has developed work on reference/co-reference in multimodal dialogue for French. The models we developed are complementary, and need now to be tested on real data. Therefore, another point in common between the teams is that at the present time, they are working on annotation schemes in order to study the modelled phenomena on linguistic corpora. The main goal of the project is the integration of our models into a unified approach to reference and co-reference which takes into account the linguistic variety of referring expressions in different languages and different kinds of data. Based on this unified approach, we intend to develop a computational tool for dealing with reference. The proposed work is based on Vieira and Poesio (2000), which follows the assumption that (co)reference is not primarily linking. This assumption is a common hypothesis worked out in previous work of the project's partners (Vieira, 1998) and (Salmon-Alt, 2001b). The work in this project will extend the previous studies with regard to three points: languages taken into account (English, Portuguese and French), expressions taken into account (definite descriptions and demonstratives), and text types taken into account (written texts and dialogues). As a result of this work, we will produce a unified annotation scheme for mark-up referring expressions in multilingual texts and dialogues, the corresponding annotated resources for Portuguese and French, an implemented system capable of dealing with reference and coreference resolution of definite descriptions and demonstratives for Portuguese and French, and comparative linguistic studies relating the results obtained for each language. These results will be made publicly available.

Person in charge: Susanne Salmon-Alt and Laurent Romary

Period: 2001-09-01 / 2003-09-01

Partner(s): Université UNISINOS São Leopoldo - Brésil

### *8.1.2. Digital Museum*

Distributed Structure of Content Management for Digital Museum

Theme: Linguistic and multimedia resources

Description: A digital library focuses on conserving, cataloguing, accessing, and tracking the usage of digitised material. On the other hand, a digital museum, other than being a simple digital archive, indeed puts more emphasis on providing users with highly educational and motivating exhibitions. Online exhibitions often consist of a variety of multimedia objects such as web pages, animation, and video clips. One can design different exhibitions on the same topic for children, experts, novices, high bandwidth users, low bandwidth users, all using the same set of digital artefacts. The difficulty here is that it is time-consuming to produce illustrative and intriguing online exhibitions. To transform efficiently the organized media objects deposited in the digital archive into educational experiences, there is a need of a novel content management framework for

organizing digital collections and for quickly selecting, integrating, and composing objects from the collection to produce exhibitions of different presentation styles.

Also, in order to retrieve the data distributed in various digital museums, we have to design a distributed software architecture through which it might be possible to access and to share multimedia resources, which would be spread among different servers. The general idea is that one should be able to bring together all existing multimedia resources, in order to provide any user with a global access to these. Obviously, there is a need for these resources to be created, and above all, maintained at a place where there is the competence to do so. But, there can also be specific constraints that can preclude some given resources to be deported to another site than the site that has originally created them. For example, we can have "strong" conditions on the actual distribution of "electronic versions" of some resources. It is thus more sensible not to take the risk of hampering the agreement by overly spreading the corresponding contents.

Starting from 1996, Taiwan has initiated a multi-year digital museum project to digitally archive precious cultural collections. The National Chi Nan University (NCNU) has actively participated to the project and has successfully constructed the "Butterfly Digital Museum", the "Lanyu Digital Museum" and the "Ali-Mountain Digital Museum" (http://dlm.ncnu.edu.tw). Over the years, the "Language and Dialog" Project LORIA / INRIA Lorraine) has devoted much effort on ELAN and SILFIDE projects. SILFIDE and ELAN are both distributed language resources systems, offering access to existing linguistic resources to their potential users throughout Europe. In the framework of SILFIDE and ELAN, all flowing data (i.e. requests, results, messages, ...), as well as all information about users and user's working spaces, have been encoded using XML (http://www.loria.fr/equipes/led).

The purpose of this project is to incorporate the NCNU and INRIA research effort for developing an XML-based distributed hypermedia digital museum content management framework. Design issues of the framework and prototype system implementation will be addressed in this paper.

Administrative context: National Science Council of Republic of China (Taiwan) and INRIA International Relations

Web site: http://www.loria.fr/projets/DigitalMuseum/

Contact: Samuel Cruz-Lara

Partner(s): National Chi Nan University - Taiwan

### *8.1.3. InDiGen*

Integrated Discourse Generation

Theme: natural language generation, associative anaphora, constraint based programming

Description: InDiGen stands for "Integrated Discourse Generation". The project concentrates on the generation of definite descriptions and in particular, of associative (or bridging) definite descriptions. It explored two main directions of research :

1. the use of constraint based programming for computing distinguishing descriptions and
2. the specification and implementation of a microplanner integrating surface realization and inference

Administrative context: Project funded by the Deutsche Forschungsgesellschaft

Web site: http://www.coli.uni-sb.de/cl/projects/indigen.html

Period: 2000-01-00 / 2004-01-01

Contact: Claire Gardent

Software: InDiGen

Partner(s): Dépt de Linguistique Informatique de l'Université de Sarrebrück, Sarrebrück - Allemagne

### 8.1.4. Hotaru

Description: Loria, University of Nantes and National Institute of Informatics start a joint research project called 'HOTARU' (firefly) that is study on Terminology from the view of practical analysis. As the first step of this theme, we investigate extraction of terms in Japanese based on a part-of-speech pattern match method comparing with French and English compounds.

Period: start 2002-10-01

Contact: Laurent Romary

### 8.1.5. ISO TC37 SC4

Theme: The standardization of language resources is an essential aspect of natural language processing since it allows one both to reuse linguistic data such as lexica or grammar from one application to another and to deploy interoperable linguistic component in complex processing chains (e.g. a man-machine dialogue systems).

Description: ISO committee TC 37/SC 4 (Language Resource Management) has been launched in 2002 to cover all standardization needs in the domain of language resources. Under the responsability of Laurent Romary (chair) and Prof. Key-Sun Choi (Secretary), the committee aims at providing ways of ensure a high level of interoperability within applications related human language technology. Beyond the admisnistrative responsibility of the group, the team is more specifically involved in and has contributed to the following work items :

- morpho-syntactic annotation;
- lexical data representation;
- feature structure representation;
- representation of data categories for language resources.

Administrative context: ISO (International Organization for Standardization)

Remark: Our participation to ISO/TC 37/SC 4 occurs in the context of the INRIA corporate action "Syntax", and more particularly the Technolangue/RNIL and RNTL/Outilex projects.

### 8.1.6. LED-LIT

Theme: Proof methods for logics for knowledge representation, with special emphasis on proof methods for hybrid logics and related systems.

Description: The research partnership between "Langue et Dialogue" with the Language and Inference Technology (LIT) group at the University of Amsterdam was established by INRIA in 2002 for a period of three years. LIT is the largest of the five groups making up the Institute for Logic, Language and Information (ILLC) at the University of Amsterdam. The goal of the research partnership is to develop logics for use in knowledge representation, with particular emphasis on logics for applications in natural language applications. Much of the work of the research partnership centers on the modal/hybrid/description family of logics, and deduction methods for them.

Web site: http://www.loria.fr/projets/ledlit

Period: 2001-01-01 / 2005-01-01

Contact: Patrick Blackburn

Partner(s): LIT, University of Amsterdam Language and Inference Technology, Amsterdam - Pays-Bas

## 8.2. European level

### *8.2.1. Intera*

Theme: One of the difficulties related to the dissemination and proper usage of language resources is to identify them wherever they are produced or maintained. To this purpose an infrastructure has to be defined for the creation and diffusion of meta-data for language resources.

Description: The INTERA project has essentially two goals: (1) to build an integrated European language resource area by connecting international, national and regional data centers and (2) to produce new multilingual language resources. "Langue et Dialogue" is involved in the first aspect by contributing to the standardization of meta-data descriptors for language resources. The work done within Intera has been the source of the current ISO work on Data Categories for Language Resources (ISO CD 12620-1).

Administrative context: EU e-Content project in collaboration with ELRA (FR, coord), DFKI (DE), MPI (NL), CNR-ILC (IT), ILSP (GR)

### *8.2.2. Jules Verne*

Interactive Television

Theme: Linguistic and multimedia resources

Description: The Broadcast industry is facing uncertain conditions and unproven business models for interactive television services. The Jules Verne project aims to bring two worlds together - to strengthen the foundations for object-oriented TV in Europe (MHP-MPEG4) and the Globe (GEM-MPEG4).

The prospects are promising. Our European ICT industry has the potential to create very advanced terminals using industry standards (MHP-Java), which utilize advanced video compression protocols (H264) and form the world standards for object-oriented TV media via DVB. Such advanced terminals could offer PC-like access to the web and even advanced on-line gaming. But it is unclear if the current TV broadcast business models can moulded into this new form. The deployment of such an advanced services in vertical networks must be based on more that a Sport channel Pay-TV monopoly, and must allow a competitive horizontal market for connected home networks, storage and terminals to come.

To achieve the goal of creating standards the Jules Verne project has formed important collaborations with the key players in this European-wide industry. This will facilitate the development of a strong growth market for each party's products.

Expected results :

- To develop and implement a plan for bridging the current gap between the needs of European content creators and the capabilities of terminals.
- To demonstrate a working model for an advanced content creation flow that matches the needs of the content industry and tools vendors, as well as open APIs for MHP and MPEG4 component vendors.
- To define a range of terminals that would facilitate European growth in the IDTV market (both Java-TV and home network appliances).
- To develop related applications for the entertainment and gaming industries.

**Work in Progress**

The RAMO (Reactive & Adaptive Multimedia Object) concept is based on the notion that a new dimension of interactivity can be achieved by enabling multimedia objects to fulfil the following criteria:

- To become fully autonomous,

- To be independent from predefined scenarios,
- To fully emulate the characteristics and behaviours of the represented entities.

Such objects are able to react and adapt themselves to any contextual situations resulting from interactions with other objects of the application and/or from users actions. Interactivity in such a concept has to be seen at two different levels: the object level and the system level (running environment). Interactivity is not deterministic at the system level. At the objects level it deals with two kinds of scenarios: the predefined and evolutive ones. Predefined scenarios consist in predefined behaviours that are initially set up and based on a stimulus/reaction model. The evolutive scenarios of an object refer to the notion of intelligent adaptive and learning entities. The dynamic evolution of the scene composition, the contextual situations and the objects adaptations create multiple and complex combinations that lead to unpredictable solutions. Mixed time-based and event-based multimedia scenarios will be designed.

A RAMO MPEG-7 based description schemes (RAMO DS) has to be developed in order to enable to completely describe RAMO objects in terms of states and behaviours. The states will be described based on "Sensorial / AV Representations" and in terms of "Temporal and Spatial Matches". The behaviours will be described in terms of "Processes" and of "Events". The MPEG-7 Multimedia Description Schemes (MPEG7-MDS) is an interesting technical choice to describe the RAMO Scenes and Objects. It brings standardized means in the description and the specification of the following main features:

- The structures of entities and their internal relations between elements;
- The multimedia content for attributes with their type definition;
- The multimedia resources for control and access;
- The references to external resources;
- The relations and interactions among entities;
- The granularity, the flexibility and the openness like what XML standard offers.

RAMO model has to specify the communication protocols to set up between the objects as well as the way RAMO components (Application, Scenes and Objects) are processed. This has to be achieved starting from the object management level up to the system level (application supervision).

The RAMO model aims at being a generic model usable for many kind of interactive multimedia presentation. It should be implementation independent and moreover opened to and compatible with extension profiles.

In the framework of RAMO, the main objective of the "Language et Dialogue" team is to add managing abilities to RAMO's model in two domains:

- Terminology
- Multilingual Content

As anticipated, the work conducted within "Work Package 1" is closely connected to recent or ongoing standardization within ISO and other standardization bodies. First it is an opportunity to apply the recently published (15 Aug. 2003) ISO 16642 (Terminological Markup Framework) to relate RAMO objects to terms describing and/or indexing them. Second, we work on the definition of a general model for describing multilingual information unit which should encompass the recent proposal for translation memories within Lisa (i.e. TMX - Translation Memory eXchange format see http://www.lisa.org), the work within Oasis on localization with the XLIFF format (see http://www.xliff.org), and more generally whatever type of multilingual information that could be encountered in a multimedia object (notices, subtitles, verbal interactions, etc.). This activity foresees a future new work item on this subject within ISO committee TC 37/SC 4. Finally, and probably most importantly from the point of view of making the Ramo model flexible with regards specific application domains and usages, we are experimenting the possible application of the ongoing work with ISO CD 12620-1 (within ISO/TC 37/SC 3) to describe the various data categories (whether generic or specific) to be used in a Ramo object (or rather type of object). To do so we forsee the deploiement of a specific Jules Verne Data Category registry to which any specification of a Ramo object will have to refer to. Note: ISO 12620-1 is closely related to ISO 11179 (Meta-data registries) defined in ISO-IEC/JTC 1/SC 32.

Administrative context: ITEA
Web site: http://webserver.tudor.lu/QuickPlace/julesverne/Main.nsf/
Person in charge: Keith Baker
Period: 2003-01-01 / 2004-12-31
Contact: Samuel Cruz-Lara

### *8.2.3. MIAMM*

Multimedia Information Access using Multiple Modalities

Theme: Specification and implementation of multimodal dialogue systems including haptic interaction.
Description: The "Langue et Dialogue" team is in charge of several tasks in the framework of this project:

- implementation of a parsing module for French. This task exploits previous works of the team dealing with syntactic analysis using tree adjoining grammars. The module provides a description logic output;
- implementation of a multimodal fusion module, that provides a contextual interpretation of multimodal utterances (involving speech and gesture);
- specification of MMIL language (MultiModal Interface Langage). All exchanges between the modules of the MIAMM system architecture are expressed using this language;
- software integration of all architecture modules.

This project is considered as strategic for the team, for two reasons. First, because it allows to experiment the various models that have been designed in the team for several years. Second, because it involves a strong collaboration between our team and "Parole" team which is in charge of the speech recognition aspects within MIAMM.

Administrative context: IST European project
Web site: http://www.loria.fr/projets/MIAMM/
Person in charge: Laurent Romary
Period: 2001-09-01 / 2004-02-28
Contact: Laurent Romary
Software: soapical: Soapmill, LLP2: Analyseur LTAG, Segment Server
Partner(s): CANON, SONY Europe, DFKI, TNO Human Factors

### 8.2.4. Ozone

O3, Offering an Open and Optimal roadmap towards ambient intelligence

Theme: Design of a multimodal dialogue system for ambient intelligence applications.

Description: The "Langue et Dialogue" team intervenes in the specification of a man-machine dialogue system in the framework of the OZONE architecture. The purpose is to build on an application-independent dialogue manager. The work that has been done during year 2002 was the specification of the various components of the dialogue system architecture. The work of year 2003 was the implementation of these components, in order to provide a demonstrator for December 2003.

Administrative context: IST European project

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

Period: 2001-10-01 / 2004-05-01

Contact: Laurent Romary

Software: soapical: Soapmill, LLP2: Analyseur LTAG, Segment Server

Partner(s): MAIA, INRIA Nancy, PAROLE (INRIA Nancy), Philips, Interuniversity Micro Electronics Center, Epictoïd, Technical University Eindhoven, Thomson Multimedia

## 8.3. National level

### 8.3.1. ANANAS

"Annotation Anaphorique pour l'Analyse Sémantique de Corpus"

Description: This project aims at creating a base of french semantic corpora, annotated with anaphoric links following a normalized annotation schema. The need of such corpora is indeed obvious for researchers in linguistics, NLP, information retrieval and cognitive science, for example for testing algorithms for generation and analysis of referring expressions or anaphora. However, the available ressources for French are of insufficient size and heterogeneous concerning annotated phenomena and coding schemes. Furthermore, none of them has been annotated following the stand-off principle. Within this project, we selected 23 corpora of different genres (litterature, journal, low and technical texts, dialogue samples), which have been tagged, parsed and prepared for noun phrase extraction. We are currently defining a suitable annotation scheme in form of a meta-schema which will be submitted as a proposion for ISO normalisation. We also implemented tools for evaluating on-line the quality of inter-annotator agreement for coreference annotation. Next steps (2004-2005) will be the effective annotation of anaphora in the coropora and implementation of the web interface for corpus search and distribution.

Web site: http://www.atilf.fr/ananas/

Person in charge: Susanne Salmon-Alt

Period: 2002-01-01 / 2005-01-01

Contact: Susanne Salmon-Alt

Partner(s): ATILF, Grenoble III, ISSCO

### 8.3.2. Evalda/Arcade II

"Action de Recherche Concertée sur l'Alignement de Documents et son Evaluation"

Theme: Multilingual Alignment

Description: The project aims at exploring the techniques of parallel text alignment through a fine evaluation of the existing techniques and the development of new alignment methods. This project follows a preceding evaluation campaign of techniques for parallel text alignment (Arcade I).

LORIA is a participant with its alignment system.

Administrative context: Sponsored by Technolangue (National project, Ministry of Industry)

Web site: http://www.up.univ-mrs.fr/~veronis/arcade/index.html

Person in charge: ELDA, DELIC (Organisateurs)

Period: 2002-10-01 / 2005-09-30

Software: Concordancier (Alignement multilingue), XAlign (Alignement multilingue)

### *8.3.3. AS COMETE*

"Co-construction éMergente d'Expérience par inTEraction"

Theme: Individual and collective experience. Knowledge Sharing, exchange and co-construction.

Description: This project addresses the problem of collective experience emerging in a group of users in computer-supported collaborative work. This research has the following issues from both point of view of computer and human sciences (the list is not exhaustive):

- identify various forms of experience which can be traced along a collective task;
- provide models or representations able to process such experience in a computational context;
- provide an assistance paradigm, based on experience capitalization (ability to transfer experience in other individual or collective tasks contexts)
- study the co-construction processes of common knowledge, in the dynamics of group interactions
- study the co-evolution and learning of users engaged in the collective process of experience sharing.

Administrative context: CNRS
Web site: http://www-rtp38.univ-lyon1.fr/rtp38/site_AS_Comete/
Person in charge: Salima Hassas, Alain Mille (LIRIS)
Period: start 2003-07-11
Contact: Matthieu Quignard
Partner(s): EMSE-SIMMO, IRIN, CRISTO, LIRIS

### *8.3.4. ASILA*

"Action Spécifique Interaction Langagière et Apprentissage"

Theme: Research network for prospective in the field of 'Dialogue and Learning'.

Description: Goals of the project:

1. elaborate a research network between labs or researchers working on Dialogue and Learning from the point of view of Computer or Human Sciences
2. draw perspectives for further research or project goals that could be worth to support or to develop in workgroups.

This work is based on the expertise in the team about normalization and processing of linguistic resources, and notably of dialogue corpora.

Administrative context: CNRS
Web site: http://www.loria.fr/projets/asila
Person in charge: Daniel Luzzati (LIUM), Laurent Romary (LORIA)
Period: 2002-01-20 / 2003-01-20
Contact: Christine Fay-Varnier, Laurent Romary, Matthieu Quignard
Partner(s): LIMSI, GREYC, LIUM, IRISA, ICAR

### 8.3.5. CINERGIE

Description: The promotion of the heritage of European film increasingly depends on information technology, but these new tools have not yet realized their full potential. Many projects are still confronted with the problems posed by heterogeneous resources, working in isolation and language barriers.

CINERGIE, a project jointly initiated by academic research and industry, solves these problems by developing a collaborative IT platform intended to manage and disseminate multimedia film resources - which are especially rich in Europe.

Web site: http://www.loria.fr/projets/CINERGIE/
Person in charge: Jean-Luc Husson
Period: 2001-03-20 / 2003-02-20
Contact: Jean-Luc Husson
Software: CINERGIE
Partner(s): Lucid'it, CNC, CERIIME

### 8.3.6. Evalda/Easy

Evaluation of the syntactical parser

Theme: evaluation, syntactical analysis

Description: The EVALDA project is financed by the French Research Ministry, in the context of the Technolangue program. This project aims to implement an infrastructure dedicated to the evaluation of language technologies in France, for the French language.

One of the campaigns, in which "Langue et Dialogue" team works, of this project is the evaluation of the syntactic parsers (EASY)

Administrative context: Technolangue program
Web site: http://www.technolangue.net/
Person in charge: Khalid Choukri
Period: 2003-01-01 / 2004-12-31
Contact: Azim Roussanaly
Software: LLP2: Analyseur LTAG, Segment Server
Partner(s): ELRA/ELDA, LIMSI, ATILF, ATOLL, GREYC, LLF/Paris 7, ERSS/Toulouse, TAGMATICA, DELIC, LPL, XRCE, DIAM/ Paris 6

### 8.3.7. GenI

Theme: Generation and Inference

Description: GenI stands for Generation and Inference. The project aims at :

1. investigating the natural language generation subtasks (le.g., lexicalisation, associative anaphora) that involve inference and
2. developing and integrating the tools (grammar, generator, theorem prover) necessary to carry out these inferences

Administrative context: "Action de Recherche Concertée INRIA"
Web site: http://www.loria.fr/projets/geni/
Person in charge: Claire Gardent
Period: 2002-03-01 / 2004-04-01
Contact: Claire Gardent
Software: GenI Generator
Partner(s): ATOLL, Lattice (Paris 7), ILPL

### *8.3.8. GDR Sémantique*

"Sémantique et Modélisation"

Theme: Natural Language Semantics.

Description: This GDR is concerned with natural language semantics. Roughly speaking, semantics in this GDR is understood in the sense of Tarski and Frege: that is, it is concerned with what is often called 'formal semantics'.

Web site: http://semantique.free.fr/

Person in charge: Francis Corblin

Period: 2002-01-01 / 2005-12-31

Contact: Patrick Blackburn

Partner(s): Jean Nicod Institute, Institute of Cognitive Science, ATILGF (University of Paris VII), "Logique Interaction langue Computation" (Toulouse)

### *8.3.9. Guirlande-FR*

Gestion et Usages Informatiques des Ressources Langagières pour la Diffusion et l'Etude du Français

Description: the current development of language engineering and corpus linguistics needs textual corpora, and a wide range of tools for using them effectively. The aim of GUIRLANDE-fr is to set up a network of distributed servers allowing sharing of coded and annotated French language texts, and shared tools for working with them. These resources will be made available in a completely transparent fashion; that is, the user will not need to be aware of the actual location of the resource.

Administrative context: Incentive Concerted Action GRID 2001 (INRIA)

Web site: http://www.loria.fr/projets/Guirlande/

Partner(s): ATILF, ILF

### *8.3.10. Evalda/Media*

Theme: Contextual evaluation of man-machine dialogue systems

Description: The Media project aims is a) to constitute a corpus of semantically annotated data and b) to conduct, using this data as a reference, an evaluation campaign of existing dialogue manager modules to compare their understanding capabilities in the context of preceding interactions.

In this project, we contribute to both the definition of the annotation scheme and to the evaluation proper by adapting the dialogue module designed within the MIAM and Ozone projects.

Administrative context: Media is one of the component of the Evalda project in the French Technolangue program

Contact: Laurent Romary

### *8.3.11. Outilex*

Theme: The design of lexical formats and the delivery of generic tools to manipulate lexical data is a core issue of language engineering at large. The Outilex project aims putting in common the experience of several industrial and academic partner to deliver an open source platform of interoperable lexical tools.

Description: "Langue et Dialogue" is in charge of coordinating the design of a generic format for the representation of multilingual lexical data. In close interaction with the RNIL committee and the INRIA national action Syntax, the following results have been acheived:

- on the basis of a set of lexical samples provided by the various partners of the project, we identified some first principles as well as a selection of data categories covering all the elementary features (morpho-syntactic, syntactic or semantic) contained in the lexica;
- we launched a wider working group on lexical structures for NLP that gathers up around 35 experts from the field and produced a document outlining a possible model to be submitted to ISO committee TC 37/ SC 4.

The objective for next year is to closely interact with the proponent of a new work item on lexical representation (USA) to promote the results from the project.

Administrative context: RNTL platform project

### *8.3.12. RNIL*

Theme: As a mirror group to ISO/TC 37/SC 4 (Language Resource Management), the RNIL committee (Ressources Normalisées pour l'Industrie des Langues) was established in May 2002.

Description: "Langue et Dialogue" has been particularly active this year on three aspects: a) it has been the initiator of the ISO CD 12620-1 proposal aiming at defining a framework for describing and registering data categories in language engineering; b) as a member of the working group on feature structure representation (ISO DIS 24610-1), testing and adapting our own implementation to make it close to the most recent decisions; c) morpho-syntactic annotation for which we gathered a core set of around 200 categories, including the results of the analysis conducted within Outilex as well as previous projects like Eagles and Multext/Multext-east.

Administrative context: Technolangue program of the French Ministry for Research

### *8.3.13. XMiner*

Description: XMiner is an RNTL project involving EADS, LIP6, Gerdoss and Loria. The aim of the project is to cope with medical files and more precisely to annotate such documents so that they can be indexed by illness types, dates, etc. In this context, the "Langue et Dialogue" team proposes a Tree Adjoining Parser and a grammar that detects linguistic structures introducing illnesses.

Web site: http://www.industrie.gouv.fr/rntl/AAP2001/Fiches_Resume/XMINER.htm

Contact: Bertrand Gaiffe, Laurent Romary

Software: LLP2: Analyseur LTAG, Segment Server

Partner(s): Matra Systèmes et Information, GERDOSS, LIP6, Lucid'it

## 8.4. Regional level

### *8.4.1. CPER/ILDSTC*

"Contrat Plan Etat Région": Language, Document and Scientific, Technical and Cultural Information Engineering.

Theme: language treatment, management of scientific, technical and cultural information.

Description: One of the main lines of the CPER, in which the "Langue et Dialogue" team is involved, is Languages, Document and Scientific, Technical and Cultural Information Engineering. Objectives are :

- to reinforce and create synergies between research (LORIA, INIST, ATILF and university laboratories), content offerers, solutions developers (INIST, Berger-Levrault), teaching institutions (Metz and Nancy Universities) and users (the worlds of economics and professional training) ;
- To structure a competences center, to propose and to diffuse innovative technical solutions, and to evaluate the results of research in languages, document and scientific and technical information engineering;
- To reinforce economic activities in this domain, giving an additional asset to the industries of Lorraine in the context of information society;
- To create new activities from the fields of competences of the various laboratories.

In concrete terms, an action to implement a double computing platform of management, distribution and access to textual information is in progress :

- One for the scientific and technical information ;
- The other for the management of linguistic resources (annotated texts, lexicons, dictionaries, linguistic engineering tools)

A part of these basic tools software will certainly be shared by the two platforms. To allow an easy integration of various software components, we must use exchange formats between the different components based on clearly defined standard (XML, TEI, etc.) and integrate the competences of the participating teams.

Administrative context: "Contrat de Plan Etat-Région"
Web site: http://www.loria.fr/projets/ILD-ISTC
Person in charge: Jean-Marie Pierrel
Period: 2000-01-01 / 2006-12-31
Contact: Ingrid Falk
Software: ILD-ISTC (dans le cadre du CPER), HyLoRes, A resolution based theorem prover for hybrid logics
Partner(s): ATILF, INIST/CNRS, CALLIGRAMME, ORPAILLEUR (INRIA Nancy), READ/LORIA

# 9. Dissemination

## 9.1. Service to the scientific community

*9.1.1. Management responsibilities*

- Co-director of the UMR ATILF/University of Nancy 2 (Jean-Marie Pierrel)
- Chairmanship of the TC37/SC4 committee of the ISO (Laurent Romary)
- Presidency of the EACL board (European Chapter of the Association for Computational Linguistics) (Claire Gardent)
- Presidency of SIGSEM, ACL Special Interest Group in Computational Semantics (Patrick Blackburn)
- Member of the technical council of the TEI (Laurent Romary)
- Member of the EACL board (European Chapter of the Association for Computational Linguistics) (Claire Gardent)
- Member of the scientific board of the ACI "Travail" of MENRT (Jean-Marie Pierrel)
- Nominated member of the working group on "traitement informatique du langage auprès du Conseil Consultatif de la Langue Française" (Jean-Marie Pierrel)
- Member of the council of the "Ecole doctorale de Linguistique" of the University of Paris 7 (Jean-Marie Pierrel)

- Member of the scientific board of the "Pôle de Recherche Scientifique et Technologique (PRST) Intelligence Logicielle" (Jean-Marie Pierrel)
- Member of the scientific council of the ACI GRID "Grille de ressources informatiques et de données"(Jean-Marie Pierrel)
- Member of the steering committee of the Thematic Pluridisciplinary Network (RTP) "Communication et Dialogue " of the STIC department CNRS (Jean-Marie Pierrel)
- Elected member of the Scientific Council of the UHP, University of Nancy 1 (Jean-Marie Pierrel)
- Elected member of the Scientific Council of the University of Nancy 2 (Daniel Coulon, Samuel Cruz-Lara)
- Elected member of the Council of studies and university life of the INPL (Christine Fay-Varnier)
- Members of the specialist commissions of the 3 Universties of Nancy (Daniel Coulon, Christine Fay-Varnier, Jean-Marie Pierrel, Samuel Cruz-Lara)
- Elected member of the board of directors of the IUT of St Dié (Jean-Luc Husson)
- Member of the selection commission of the IUT of St Dié (Jean-Luc Husson, Jean-Marie Pierrel)
- Co-opted member of the scientific and technical council of "Nancy 2005" (Jean-Marie Pierrel)
- Member of the local commission temporary assignment/delegation INRIA Lorraine (Claire Gardent)
- Member of the editorial committee of "la lettre du LORIA" (Claire Gardent)
- Representative to follow through the social affairs of the INPL (Christine Fay-Varnier)
- Project leader for the registrar's office of the IUT Nancy-Charlemagne (Nadia Bellalem)
- In charge of the training for the CNRS and INRIA at the Loria (Laurent Romary)
- Member of scientific council PI CNRS TCAN (Claire Gardent)
- Program Co-Chair, Context 03, Fourth International and Interdisciplinary Conference on Modeling and Using Context. Stanford, California (USA), June 23-25, 2003(Patrick Blackburn)
- Program Committee, Proceedings of Mathematics of Language (MOL-8), Bloomington, Indiana (USA), 2003(Patrick Blackburn)
- Program committee membership of ICEIS-2003(Nadia Bellalem)

- Manager of the project "Ingénierie des langues, du document et de l'information scientifique, technique et culturelle" in the context of the CPER of Lorraine region (Jean-Marie Pierrel)
- In charge of the INRIA Concerted Research Action "Génération et Inférence" (Claire Gardent)
- Manager of the Franco-German project InDiGen. With the Department of Computational Linguistics, Universität des Saarlandes, Saarbrücken (Claire Gardent)
- Scientific manager of the project "Inference et Generation automatique de la Langue Naturelle". Region Lorraine (Claire Gardent)
- Coordinator of the INRIA funded Research Alliance between the Language and Inference Technology Group, University of Amsterdam, and "Langue et Dialogue", LORIA, Nancy (Patrick Blackburn)
- Expert for the AFNOR X03A-GE1 on the redefination of the standard ISO 1951 "Présentation/Représentation des articles dans les dictionnaires" (Jean-Marie Pierrel)
- Expert for the research ministry on the national prospects for the calculation and intensive storage for the social science (Jean-Marie Pierrel)

*9.1.2. Editorial and program committee work*

- Head-editor of Journal of Logic, Language, and Information (Patrick Blackburn)
- Editor of the EACL Newsletter du bulletin du chapitre européen de l'Association pour la Linguistique Informatique (EACL Newsletter) (Claire Gardent)
- Co-editor of the T.A.L. Journal (Traitement Automatique des Langues)(Claire Gardent)
- Assistant editor for the review: "Revue d'Intelligence Artificielle" (RIA), Editions Hermès (Jean-Marie Pierrel)
- Member of the editorial committee for: Journal of Semantics (Claire Gardent)
- Member of the editorial committee for: Traitement Automatique des Langues (Claire Gardent, Jean-Marie Pierrel)
- Member of the editorial committee for: Computer and the Humanities (Laurent Romary)
- Member of the editorial committee for: Information-Interaction-Intelligence (Jean-Marie Pierrel)
- Member of the SCientific Council for "Hermès Sciences Publications" (Jean-Marie Pierrel)
- Member of the editorial committee of TAL, RIA, I3 (Jean-Marie Pierrel)

- Member of the program committee of the Journal "STICEF" (Matthieu Quignard)
- Member of the program committee of "EACL Workshop on Natural Language Processing (NLP) for Question-Answering", April 14th 2003, Budapest, Hungary (Claire Gardent)
- Member of the program committee of "EACL Workshop on the Computational Treatment of Anaphora", April 14th 2003, Budapest, Hungary (Claire Gardent)
- Member of the program committee of "9th European Workshop on Natural Language Generation", April 13-14th 2003, Budapest, Hungary (Claire Gardent)
- Member of the program committee of ACL03 (Claire Gardent)
- Member of the program committee of "Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications" (Claire Gardent)
- Member of the program committee of the International Workshop on Description Logics - DL2003 Università di Roma "La Sapienza", Rome, Italy, September 5-7, 2003 (Carlos Areces)

*9.1.3. Conference and workshop organization*

- EACL 03 (Tenth Conference of the European Association for Computational Linguistics)
- 16th European Summer School in Logic, Language and Information (ESSLLI 2004), Nancy, France, 9 - 20 August, 2004 (Carlos Areces, Patrick Blackburn)
- Nancy Inference Week (NIW), LORIA, Nancy, 22-26 September, 2003 (Carlos Areces, Patrick Blackburn)
- Methods for Modalities 3 (M4M-3), 22-23 September, 2003 (Carlos Areces, Patrick Blackburn)
- Description Logic Day (DLD), 24 September, 2003 (Carlos Areces, Patrick Blackburn)
- Inference in Computational Semantics (ICoS-4), 25-26 of September, 2003 (Patrick Blackburn, Johan Bos)
- Prospects and Advances in the Syntax/Semantics Interface, LORIA (Nancy), France, October 20-21, 2003 (Carlos Areces)
- 3rd Annual Members Meeting of the TEI, Nancy, 7-8 november 2003 (Laurent Romary)
- IWPT 03, Nancy, France, 23-25 april 2003, co-organization (Bertrand Gaiffe, Jean-Marie Pierrel, Laurent Romary, Azim Roussanaly, Nadia Viscogliosi)

*9.1.4. Seminars and invited talks*

- "FTAG et construction sémantique", Calligramme Seminar, LORIA, Nancy, 23 january 2003
- "FTAG et construction sémantique", LABRI, Bordeaux, 30 january 2003
- Computational semantics meets computer science Third International Workshop on Computational Semantics (ICOS) - Invited talk, Nancy, 25 September 2003
- Arthur Prior and Hybrid Logic. Logic of Time and Modality Conference, Roskilde University, Denmark. 1 November 2003 (Patrick Blackburn)
- An Introduction to Contemporary Hybrid Logic, Logic of Time and Modality Conference, Roskilde University, Denmark. 31 October 2003 (Patrick Blackburn)
- Hybrid Logic. British Logic Colloquium, St Andrews,Scotland. 4 September, 2003 (Patrick Blackburn)
- Coping with Content. 15th European Summer School in Logic, Language and Information (ESSLLI 2003), Vienna, Austria. 25 August, 2003 (Patrick Blackburn)
- "Inférence et Sémantique Computationnelle". TALN 2003, Batz-sur-Mer, France. 11 June, 2003 (Patrick Blackburn)
- "The standardized tagging of concepts and documents with the EAD standards. Le balisage normalisé des concepts et documents en liaison avec les normes de l'EAD". "Journée Internationale de réflexion Normes et standards pour l'apprentissage en ligne" (AUF-AFNOR), Versailles, 19 March 2003. (http://www.initiatives.refer.org/Initiatives-2003/) (Laurent Romary et Henri Hudrisier)
- A Metamodel to Represent Terminology Data Collections, Open Forum 2003 on Metadata Registries, Terminology and Ontologies Track, Santa Fe, 20-24 January 2003 (Laurent Romary)
- "Métadonnées : des contenus aux structures", "Les rencontres des professionnels de l'IST, Paris, Palais des Congrès", 17-19 June 2003 (Laurent Romary)
- "Organiser l'information pour mieux la trouver : XML, les métadonnées et le web sémantique, Conférence i-expo", Paris, 18 June 2003. (Laurent Romary)
- "Annotation - the ISO perspective". ECHO IT days, Lund (Sweden), 18-20 September - http://www.ling.lu.se/projects/echo/contributors/itdays/itdays_report.html (Laurent Romary)
- Modelling language resources - towards reference standards in NLP, seminar at the Computer Laboratory, Natural Language and Information Processing Group, Cambridge, 4 March 2003 - http://www.cl.cam.ac.uk/~pjb48/seminars/Lent.html (Laurent Romary)

- Lectures on Language Resources and Standardisation, University of Malta, 10-14 March 2003 - http://www.um.edu.mt/noticeboard/lrs.html (Laurent Romary)
- "Normalisation d'annotations langagières et XML" , seminar at LIMSI-CNRS, 1st April 2003 (Laurent Romary)

## 9.2. University teaching

- Joint responsibility for the course: "Perception, raisonnement et Traitement automatique des langues" of the computer science DEA of Lorraine (Jean-Marie Pierrel)
- Responsibility for the DEA "Modélisation et simulation des espaces bâtis" (Jean-Marie Pierrel)
- Responsibility for the degree and the master degree of "Sciences Cognitives" at the University of Nancy 2 (Daniel Coulon)
- Responsibility for the computing commission of the SRC department at the IUT of St Dié (Jean-Luc Husson)
- Responsibility for the computing and multimedia department at the ENSG (Christine Fay-Varnier)
- Directorship of the industrial relations of the SRC department at the IUT of St Dié (Jean-Luc Husson)
- Participation in teachings of bio-computing of the DESS RGTI "Ressources Génomiques et Traitements Informatiques" and the master degree MGMC "Mention Génétique Moléculaire et Cellulaire" (Marie-Dominique Devignes, Malika Smaïl)
- Participation in the course of PRTAL of computer science DEA d'informatique (Bertrand Gaiffe, Laurent Romary)
- Participation in the course of "Sciences Cognitives" at the University of Nancy 2 (Daniel Coulon, Benoît Crabbé, Evelyne Jacquey, Hélène Manuélian, Azim Roussanaly)
- Participation in the DESS "TEXTE" at the University of Nancy 2 (Claire Gardent, Jean-Marie Pierrel, Laurent Romary, Djamé Seddah)
- Participation in the DESS "Industries de la langue"at the University of Metz (Daniel Coulon, Samuel Cruz-Lara)
- Participation in the DESS "Systèmes d'Information Distribués" at the University of Nancy 2 (Samuel Cruz-Lara)
- Participation in the technical degree of "Concepteur-Intégrateur de Systèmes Intranet / Internet" at the University of Nancy (Samuel Cruz-Lara)

- Participation in the "Sciences du langage" course at the University of Nancy 2 (Evelyne Jacquey, Hélène Manuélian)
- Participation in the "Sciences du langage" course at the University of Metz (Hélène Manuélian)
- Participation in the course of "Mathématiques, Informatique et Statistiques Appliquées aux Sciences Humaines et Sociales" at the University of Nancy 2 (Benoît Crabbé, Azim Roussanaly)
- Participation in the "MIAGE" course et the University of Nancy 2 (Benoît Crabbé)
- Participation in the master's degree of "Sciences de l'Information et de la Documentation" at the University of Nancy 2 (Nadia Viscogliosi)
- Member of the international relations service of the IUT Nancy-Charlemagne (University of Nancy 2 (Samuel Cruz-Lara)
- Computational Semantics for Natural Language, PRTAL4, DEA course, University of Nancy I (Henri Poincare), 20 hours October-December 2003. (Patrick Blackburn)
- In charge of the teachings "Algorithmique et Programmation" of the first-/second-year at the ENSG (Christine Fay-Varnier)
- In charge of the optional module: "I.A. et techniques émergentes" of the third year at the ENSG (Christine Fay-Varnier)
- Member of the steering committee of the TICE projects at the INPL (Christine Fay-Varnier)
- Project leader for the deployment of the "Espace Numérique de Travail ESUP - Portail" for the INPL (Christine Fay-Varnier)
- Director of the studies of the computer science department at the IUT Nancy-Charlemagne (Nadia Bellalem)
- Participation in the technical degree of "concepteur-intégrateur de systèmes intranet-internet" University of Nancy 2 (Nadia Bellalem)

## 9.3. Other teaching

- Course and workshop on "Knowledge and Interaction" during the CNRS thematic summershool on Computer-based Learning Environment (EIAH, Autrans, 2003) (Matthieu Quignard)
- Invited tutorial, A brief introduction to computational linguistics, Tbilissi Symposium, 06-10 September 2003 (Claire Gardent)

- Invited course, Computational Semantics for Natural Language. North American Summer School in Logic, Language and Information, (NASSLLI), Bloomington, Indiana (USA). 17-21 June 2003 (Patrick Blackburn)
- Invited course, Computational Modal Logic, Introductory Course. Lecturers: Carlos Areces and Maarten de Rijke. At: 15th European Summer School in Logic Language and Information, Vienna University of Technology, Vienna, Austria 18th - 29th August 2003 (Carlos Areces)

# 10. Bibliography

## Major publications by the team in recent years

[1] P. BLACKBURN, M. DE RIJKE, Y. VENEMA. *Modal Logic.* series Cambridge Tracts in Theoretical Computer Science, volume 53, Cambridge University Press, juin, 2001.

[2] P. BLACKBURN, M. KOHLHASE. *Inference in Computational Semantics.* juin, 2001, ICoS-3 Workshop Proceedings, June 18-19, Siena, Italy.

[3] P. BLACKBURN, M. DE RIJKE, Y. VENEMA. *Modal Logic.* Cambridge University Press, 2001.

[4] D. DUCHIER, C. GARDENT. *Tree Descriptions, Constraints and Incrementality.* R. M. H. BUNT, E. T. (EDS), editors, in « Computing Meaning », series Studies in Linguistics and Philosophy Series, volume 2, Kluwer Academic Publishers, 2001.

[5] N. IDE, L. ROMARY. *Encoding syntactic annotation.* A. ABEILLÉ, editor, in « Treebanks », Kluwer Academic Publisher, 2001.

[6] P. LOPEZ. *Analyse d'énoncés oraux pour le dialogue homme-machine à l'aide de grammaires lexicalisées d'arbres.* thèse de doctorat, Université Henri Poincaré, Nancy 1, 1999.

[7] J.-M. PIERREL. *Ingénierie des Langues.* series Traité IC2 (Information, communication et commande, Hermes, octobre, 2000.

[8] L. ROMARY, P. BONHOMME. *Parallel Alignment of Structured Documents.* J. VÉRONIS, editor, Kluwer Academic Press, 2000.

[9] L. ROMARY. *Langue et dialogues - perspectives de recherche.* habilitation à diriger des recherches, décembre, 1999.

[10] G. SABAH, J. VIVIER, A. VILNAT, J.-M. PIERREL, L. ROMARY. *Machine, langage et dialogue.* L'Harmattan, 1998.

[11] S. SALMON-ALT. *Référence et dialogue finalisé: de la linguistique à un modèle opérationnel.* thèse d'université, Nancy, mai, 2001.

## Doctoral dissertations and "Habilitation" theses

[12] FRÉDÉRIC LANDRAGIN. *Modélisation de la communication multimodale. Vers une formalisation de la pertinence.* Université Henri Poincaré, 2003, april, thèse d'université.

## Articles in referred journals and book chapters

[13] BENOIT CRABBÉ, BERTRAND GAIFFE, AZIM ROUSSANALY. *Représentation et gestion de grammaires TAG lexicalisées.* in « Traitement Automatique des Langues », December, 2003.

[14] CLAIRE GARDENT, LAURA KALLMEYER. *Semantic construction in feature-based TAG.* in « 10th meeting of the European Chapter of the Association for Computational Linguistics - EACL'03, Budapest, Hungary », 2003, April.

[15] PATRICK BLACKBURN, JOHAN BOS. *Computational Semantics.* in « Theoria », January, 2003, volume 18, number 46, pages 27-45.

[16] DANIEL COULON , MICHEL MUSIOL. *L'influence du processus conversationnel sur la dynamique de la croyance assertive.* in « Psychologie de l'interaction », number 17-18, 2003.

[17] CLAIRE GARDENT, KRISTINA STRIEGNITZ. *Generating Bridging Definite Descriptions.* in « Computing Meaning », Kluwer Academic Publishers, volume 3, December, 2003, Dordrecht.

[18] CLAIRE GARDENT, HÉLÈNE MANUÉLIAN, KRISTINA STRIEGNITZ, MARILISA AMOIA. *Generating Definite Descriptions: Non incrementality, inference and data.* in « Speech production », Walter de Gruyter, Berlin., December, 2003.

[19] MICHAEL BAKER, ERICA DE VRIES, KRISTINE LUND, MATTHIEU QUIGNARD. *Interactions épistémiques médiatisées par ordinateur pour la co-élaboration des notions scientifiques.* in « Collaborer pour apprendre et faire apprendre, la place des outils technologiques », Presses Universitaires du Québec, C. DEAUDELIN, THÉRÈSE NAULT, editors, volume 8, series Education Recherche, chapter 6, pages 121-134, January, 2003.

## Publications in Conferences and Workshops

[20] BENOÎT CRABBÉ, BERTRAND GAIFFE, AZIM ROUSSANALY. *Une plateforme de conception et d'exploitation de grammaire d'arbres adjoints lexicalisés.* in « Traitement Automatique du Langage Naturel 2003 - TALN 2003, Batz-sur-Mer, France », June, 2003.

[21] FRÉDÉRIC LANDRAGIN, LAURENT ROMARY. *Referring to Objects Through Sub-Contexts in Multimodal Human-Computer Interaction.* in « Seventh Workshop on the Semantics and Pragmatics of Dialogue - DiaBruck'03, Saarbrücken, Germany », 2003, September, pages 67-74.

[22] FRÉDÉRIC LANDRAGIN. *La saillance comme point de départ pour l'interprétation et la génération.* in « Journée d'étude de l'Association pour le Traitement Automatique des LAngues (ATALA) sur le thème: structure communicative/structure informationnelle, Paris, France », 2003, March.

[23] SAMUEL CRUZ-LARA, HONG JEN-SHIN. *A distributed framework for digital museum exhibitions.* in « DigiCULT Forum », February, 2003, volume 3, pages 26-28.

[24] FRÉDÉRIC LANDRAGIN. *Une caractérisation de la pertinence pour les actions de référence.* in « Traitement Automatique des Langues Naturelles - TALN 2003, Batz-sur-Mer, France », June, 2003.

[25] BENOIT CRABBÉ . *Alternations, monotonicty and the lexicon: an application to factorising information in a Tree Adjoining Grammar.* in « European Summer School of Logic Language and Computation, Student Workshop - ESSLLI'03, Vienna, Austria », August, 2003, pages 69-80.

[26] THI MINH HUYEN NGUYEN, LAURENT ROMARY, XUAN LUONG VU. *Une étude de cas pour l'étiquetage morpho-syntaxique de textes vietnamiens.* in « Traitement Automatique des Langues Naturelles - TALN'2003, Batz-sur-mer, France », ATALA (Association pour le Traitement Automatique des LAngues), June, 2003.

[27] CLAIRE GARDENT, HÉLÈNE MANUÉLIAN, ERIC KOW. *Which bridges for bridging definite descriptions?.* in « 4th International Workshop on Linguistically Interpreted Corpora - LINC'03, Budapest, Hungary », April, 2003.

[28] FRÉDÉRIC LANDRAGIN. *Clues for the Identification of Implicit Information in Multimodal Referring Actions.* in « International Conference on Human-Computer Interaction (HCI International), Heraklion, Crete, Greece », Lawrence Erlbaum Associates, volume 2, pages 711-715, June, 2003.

[29] SAMUEL CRUZ-LARA, HONG JEN-SHIN. *Using WebServices for Accessing and Sharing Multimedia Resources on a Distributed Software Architecture.* in « IEEE Multimedia Software Engineering 2003 - MSE 2003, Taichung, Taiwan », IEEE, December, 2003.

[30] HÉLÈNE MANUÉLIAN. *Une analyse des emplois du démonstratif en corpus.* in « Traitement Automatique des Langues Naturelles - TALN 2003, Batz sur Mer, France », Université de Nantes, June, 2003.

[31] HÉLÈNE MANUÉLIAN. *Coreferential Definite and Demonstrative Descriptions in French: A Corpus Study for Text Generation.* in « ESSLLI Student Session, Vienna, Austria », August, 2003.

[32] MICHAEL BAKER, MATTHIEU QUIGNARD, KRISTINE LUND, ARNAULD SÉJOURNÉ. *Computer-supported collaborative learning in the space of debate.* in « Fifth International Conference on Computer-Supported Collaborative Learning - CSCL'2003, Bergen, Norvège », June, 2003.

[33] MATTHIEU QUIGNARD, MICHAEL BAKER, KRISTINE LUND, ARNAULD SÉJOURNÉ. *Conception d'une situation d'apprentissage médiatisée par ordinateur pour le développement de la compréhension de l'espace du débat.* in « Environnements Informatiques d'Apprentissage Humain - EIAH'2003, Strasbourg, France », C. DESMOULINS, P. MARQUET, D. BOUHINEAU, editors, pages 355-366, INRP, April, 2003.

[34] CLAIRE GARDENT, EVELYNE JACQUEY. *Lexicalisation as a Description Logic Inference Task.* in « 4th International Workshop on Computational Semantics, Nancy, France », September, 2003.

[35] CLAIRE GARDENT, EVELYNE JACQUEY. *Lexical Reasoning.* in « International Conference on Natural Language Processing, Mysore, India », December, 2003.

[36] LAURENT ROMARY. *Annotation - the ISO (and NLP) perspective.* in « ECHO IT Days, Lund, Suède », Lund University, September, 2003.

[37] CARLOS ARECES, JUAN HEGUIABEHERE. *hGen: A Random CNF Formula Generator for Hybrid Languages.* in « Methods for Modalities 3 - M4M-3, Nancy, France », September, 2003.

[38] CARLOS ARECES, RAFFAELLA BERNARDI. *In Situ Binding: A Hybrid Approach.* in « Inference in Computational Semantics 4 - ICoS -4, Nancy, France », September, 2003.

[39] CARSTEN LUTZ, CARLOS ARECES, IAN HORROCKS, ULRIKE SATTLER. *Keys, Nominals, and Concrete Domains.* in « Eighteenth International Joint Conference on Artificial Intelligence - IJCAI'03, Acapulco, Mexico », August, 2003.

[40] AMALIA TODIRASCU, LAURENT ROMARY, DALILA BEKHOUCHE. *Extraction d'information à base d'ontologies dans une application de veille.* in « 5ème Rencontre Terminologie et Intelligence Artificielle - TIA'2003, Strasbourg, France », pages 205-208, March, 2003.

[41] BENOIT 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 », October, 2003.

[42] CARLOS ARECES, PATRICK BLACKBURN, BERNADETTE MARTINEZ HERNANDEZ, MAARTEN MARX. *Handling Boolean ABoxes.* in « International Workshop on Description Logics - DL'2003, Rome, Italy », September, 2003.

[43] FARAH BENAMARA, CLAIRE GARDENT, PATRICK SAINT-DIZIER. *Représentation et inférence dans un système pour la génération de réponses coopératives.* in « Journées scientifiques Sémantique et Modélisation, Paris », March, 2003.

[44] C. ARECES, P. BLACKBURN, B. MARTINEZ HERNANDEZ, M. MARX. *Handling Boolean ABoxes.* in « Proceedings of the 2003 International Workshop on Description Logics (DL2003) », D. CALVANESE, G. DE GIACOMO, E. FRANCONI, editors, volume 81, series CEUR - Workshop Proceedings, Rome, Italy, September, 2003.

[45] HÉLÈNE MANUÉLIAN. *Génération de descriptions définies et démonstratives.* in « Huitième Atelier des doctorants en linguistique - ADL'2003, Paris, France », July, 2003, Association des doctorants en linguistique de Paris.

[46] PATRICK BLACKBURN, BERTRAND GAIFFE, MARTEEN MARX. *Variable-free reasoning on finite trees.* in « Eighth Meeting on Mathematics of language - MOL8, Bloomington, Indiana, USA », June, 2003.

[47] PATRICK BLACKBURN, BERTRAND GAIFFE, MAARTEN MARX. *Variable free reasoning on finite trees.* in « Proceedings of Mathematics of Language - MOL-8, Bloomington, Indiana, USA », June, 2003.

## Bibliography in notes

[48] PATRICK BLACKBURN, MAARTEN MARX. *Constructive Interpolation in Hybrid Logic.* in « Journal of Symbolic Logic », August, 2002.

[49] PATRICK BLACKBURN, MAARTEN MARX. *Tableaux for Quantified Hybrid Logic.* in « International Conference on Automated Reasoning with Analytic Tableaux and Related Methods - TABLEAUX 2002, Copenhagen, Denmark », U. EGLY, C.G. FERMÜLLER, editors, volume 2381, series Lecture Notes in Artificial Intelligence, pages 38-52, Springer, July, 2002.

[50] DENYS DUCHIER. *A Metagrammatical formalism for Lexicalized TAGs.* in « Lorraine-Saarland Workshop Series: Prospects and advances in the Syntax/Semantics Interface », 2003, Nancy, october.

[51] YANNICK PARMENTIER. *Méta-grammaire et calcul sémantique pour les Grammaires d'Arbres Adjoints.* Mémoire de DEA, Université de Franche-Comté, Besançon, 2003.

[52] ANNE REBOUL, JACQUES MOESCHLER. *Pragmatique du discours: de l'interprétation des énoncés à l'interprétation du discours.* Paris, Armand Colin, 1998.

[53] DAN SPERBER, DEIRDRE WILSON. *Relevance: Communication and Cognition. (second edition).* Harvard University Press, Cambridge, MA, 1995.

[54] NICHOLAS ASHER, ALEX LASCARIDES. *Logics of Conversation.* Cambridge University Press, 2003.

[55] MICHEL TOZZI. in « L'oral argumentatif en philosophie », CRDP Montpellier, 1999.