INRIA

Project-Team L&D

Langue et Dialogue

Lorraine

THEME SYM

Activity Report

2004

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://led.loria.fr/.

Head of project-team

Laurent Romary [research director INRIA]

Vice-head of project team

Patrick Blackburn [research director INRIA]

Administrative assistant

Isabelle Blanchard [CNRS]

Research scientists

Carlos Areces [research fellow INRIA]

Bertrand Gaiffe [research fellow CNRS (STIC department)]

Claire Gardent [research fellow CNRS (SHS department)]

Matthieu Quignard [research fellow CNRS (SHS department CR2, section 34)]

Jesse Tseng [research fellow CNRS (SHS department)]

Faculty members

Nadia Bellalem [assistant professor University Nancy 2]

Daniel Coulon [professor INPL]

Samuel Cruz-Lara [assistant professor IUT Nancy-Charlemagne University 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 Nancy 2]

Post-doctoral fellow

Kornél Bangha [post-doctoral fellow INRIA from 2004-01-01 until 2004-12-31]

Visiting scientists

Johannes Bos [visiting scientist (University of Edinburgh) from 2004-06-07 until 2004-07-10]

Laurence Danlos [affiliated research scientist (Professor at University Paris 7)]

Juan Heguiabehere [visiting scientist (University of Bolzano) from 2004-08-01 until 2004-09-07]

Nancy Ide [visiting scientist (Vassar College - USA)]

Evelyne Jacquey [affiliated research scientist (STIC department) CNRS]

Susanne Salmon-Alt [affiliated research scientist CNRS (ATILF)]

Valentin Shehtmann [visiting scientist (University of Moscow) from 2004-10-07 until 2004-11-12]

Balder Ten Cate [visiting scientist (University of Amsterdam) from 2004-08-01 until 2004-09-07]

Amalia Todirascu [visiting scientist (assistant professor of the University of Troyes - UTT)]

Ph. D. students

Dalila Beckhouche [Ph. D. student (co-supervision)]

Benoît Crabbé [Ph. D. student, INRIA grant (Lecturer IUT Nancy-Charlemagne)]

Alexandre Denis [Ph. D. student, INRIA grant from 2004-10-01] [ESIEA - INRIA until 2004-02-13] [Graduate student intern INRIA until 2004-06-31]

Sébastien Hinderer [Ph. D. student UHP Nancy 1 from 2004-10-01]

Eric Kow [Ph. D. student UHP Nancy 1 from 2004-10-01] [Graduate student intern UHP Nancy 1 until 2004-09-24]

Hélène Manuélian [Ph. D. student UFR SDL University Nancy 2 (ATER at University of Metz) until 2004-06-30]

Thi-Minh-Huyen Nguyen [Ph. D. student UFR STMIA UHP Nancy 1 until 2004-11-15]

Yannick Parmentier [Ph. D. student UFR STMIA UHP Nancy 1 (INRIA grant - Teacher Assistant UHP Nancy 1)]

Joseph Roumier [Ph. D. student ENSEM INPL (CIFRE grant financed by CEGELEC)]

Djamé Seddah [Ph. D. student UFR STMIA UHP Nancy 1 until 2004-10-30 (ATER SDL at University of Metz 2003/2004)]

Kristina Striegnitz [Ph. D. student (co-supervision University of Saarbrück - Germany)]

Technical staff

Travis Choma [technical staff INRIA until 2004-03-31]

Julien Ducret [technical staff INRIA]

Pierre Dussaulx [technical staff INRIA until 2004-07-31]

Ingrid Falk [project technical staff CNRS (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))]

Laurence Kbida [project technical staff INRIA until 2004-10-31]

Isabelle Kramer [technical staff INRIA]

Satyendra Kumar Gupta [technical staff INRIA until 2004-11-30]

Frédéric Landragin [technical staff INRIA until 2004-08-30]

Erica Meena [technical staff INRIA until 2004-07-31]

Philippe Sébire [technical staff INRIA]

Jacky Tétu [technical staff INRIA until 2004-10-31]

student intern

Jean-Charles Fellet [student intern (ESIAL) from 2004-06-23 until 2004-08-08]

Javier Fernandez Garcia [student intern (University of Valladolid) from 2004-09-06]

Bilal Gargouri [student intern (SFEG - Sfax (Tunisie)) from 2004-05-31 until 2004-06-04]

Jacqueline Lai [student intern INRIA (MIT) from 2004-07-01]

Yvan Norsa [student intern UHP Nancy 1 (CEGELEC) until 2004-09-30]

Jérôme Perrin [student intern University Nancy 2 (MIAGE) from 2004-02-17 until 2004-09-16]

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:

- 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 examine 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 [68] or Relevance Theory [69]). 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 [64]), 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 [65]).

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 e1 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 e1 (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, 2002-2004) 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 Bos) 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.

Some investigation is also tried to use natural deduction, extend to i non classic logical logics as default logic in attempt to find gaps and inconsistances wich may explain where and how the speakers revise their argumentations during the dialog (see [67]).

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 is 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. 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 admitted 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. 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

5.3. 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).

Contact: Patrick Blackburn

5.4. Document Management plugin

Description : The Document Management plugin has been built as a demonstrator for the e-documentation server architecture designed for the PROTEUS project. It is developped as a plugin for the ontology editor Protégé and offers the users access to a document base through the following Access Points : conceptual, terminological, by MetaData and by Annotations. The generic WebServices for these Access Points were designed during the project to allow the switching between a software module and another offering the same features.

Currently the software modules are :

- Conceptual and MetaData accesses : Protégé (protege.stanford.edu)

- Terminological Access : MyTerm from my-xML (http://www.my-xml.com, proprietary software)

- Annotations : Annotea (w3.org/2001/Annotea)

- Documents are stored in a native XML database : Berkeley DB:XML from Sleepycat Software (http://www.sleepycat.com/products/xml.shtml)

- WebServices are deployed using Tomcat (http://jakarta.apache.org/tomcat) and Axis (http://ws.apache.org/axis)

Last update : 2004-12-07
Author(s) : Yvan Norsa, Joseph Roumier
Contact : Laurent Romary, Yvan Norsa, Joseph Roumier
Project(s) : Proteus

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 system is "stand-alone" as we use the Glasgow Haskell compiler to obtain executable code for Windows, Solaris, Linux and Mac OS X.

The generator uses efficient datatypes and intelligent rule application to minimize the generation of redundant sentences. It also uses a notion of polarities (Perrier, Calligramme) as means first, of coping with lexical ambiguity and second, of selecting variants obeying given syntactic constraints.

The grammar used by the GenI generator is produced using a MetaGrammar compiler (namely the XMG system developed by Crabbé, Parmentier — Langue et Dialogue —, Duchier and Le Roux — Calligramme —) and covers the basic syntactic structures of French as described in Anne Abeillé's book "An electronic grammar for French".

Up to now, only a first prototype is available, which links to the description logic prover RACER for the required inference services.

Version: 1.1
Author(s): Carlos Areces, Eric Kow

Contact(s): Carlos Areces, Eric Kow

Project(s): GenI

5.6. Handling Multilingual Content in ITEA Jules Verne Project

Description: The main idea is to create specific Data Categories dedicated to generic RAMO scenes, objects and their related properties and to use MLIF as a metamodel.

The Multi Lingual Information Framework (MLIF) promotes the use of a common framework for the future development of several different formats dealing with the localisation process: TBX, TMX, XLIFF, Timed Text, TMF, etc. It does not create a complete new format from scratch, but suggests that the overlapping issues should be handled independently and separately.

Currently, within the framework of ITEA Jules Verne project, we are experimenting with some basic scenarios by using XMT (eXtensible MPEG4 Textual format) and SMIL (Synchronized Multimedia Integration Language).

Under this perspective, a catalogue of transformations will make easy for other formats to integrate with MLIF. Thus, it is important to have several XSLT (XML Stylesheet LanguageTransformation) transformations between the most important localization formats.

We have proposed MLIF (Multi Lingual Information Framework): a highlevel model for describing multilingual data. MLIF can be used in a wide range of possible applications in the translation/localization process in several multimedia domains.

In order to test MLIF in multimedia, we have proposed using MLIF and Data Categories for specifying RAMO (Reactive and Adaptive Multimedia Objects) within the ITEA Jules Verne Project. Currently, we are working on specifying the MLIF API through which a wide variety of formats (i.e. TBX, TMX, XLIFF, etc.) may be integrated into the core MLIF structure.

Web site: http://loreley.loria.fr:8888

Contact(s): Samuel Cruz-Lara

Project(s): Jules Verne

5.7. 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.

5.8. 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.

5.9. 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
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

5.10. Meta-Grammars Compilation

Description: A MetaGrammar Compiler is a device that generates automatically a wide-coverage grammar from a reduced description (the MetaGrammar). This description is composed of tree fragments to be combined.
Such a compiler aims at:

- avoiding inconsistencies due to redundancy in the trees of the grammar,
- easing the development of real scale grammars (thousands of trees),
- allowing scientists to capture linguistic phenomena represented in tree lexicons.

At Langue Et Dialogue, we have developed 2 MetaGrammar Compilers, namely MGC (MetaGrammar Compiler) and XMG (eXtended MetaGrammar). XMG can be seen as en extension of MGC on the following points:

- the language used to control the combination of tree fragments includes disjunction,
- you can specify a syntax / semantic interface,
- you can compile MetaGrammars for different syntactic formalisms (Tree Adjoining Grammars and Interaction Grammars currently).

Some features of these two compilers are given below.

5.10.1. MGC

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.

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

Contact: Bertrand Gaiffe

Project(s): RLT

5.10.2. XMG

XMG has been developed in collaboration with Denys Duchier and Joseph Le Roux of the Calligramme Project. XMG's control mechanism is similar to a logic program in the DCG (Definite Clause Grammar) paradigm, where the terminal symbols are fragments describing partial tree structures. The composition system is a model builder that allows to parametrize the structures output by the compiler (currently trees anr D-Trees (D-Trees are underspecified trees)). The model-theoretic approach used for describing the lexicon also allows linguists to further use an extensible library of constraints, called principles, ensuring additional well-formedness conditions on the output Trees. The tool is developed in Oz/Mozart.

Version: 1.0

Web site: http://sourcesup.cru.fr/xmg/

Author(s): Benoît Crabbé, Denys Duchier, Joseph Le Roux, Yannick Parmentier

Contact(s): Benoît Crabbé, Yannick Parmentier

Project(s): GenI

5.11. MIAMM-Ozone architecture

Description: an architecture for multimodal man-machine dialogue (previously called SoapMMIL)

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. This architecture uses tools from the Soapical project.

5.12. Morphalou

Description: Morphalou is a large coverage morphological XML lexicon for French (~540.000 inflected forms). It is based on the Lexical Markup Framework ISO (TC37/SC4) standard — a high level model for representing data in lexical resources — and thus guarantees a maximum of interoperability with multilingual computer applications. From a linguistic point of view, coherence and correctness are granted by an editorial committee located at the ATILF. The ATILF also will maintain and update the lexicon.

Morphalou is free for research and teaching purposes.

Morphalou has been generated from the TLFnome, the nomenclature of the 'Trésor de la Langue Française' (ATILF). The data has been re-structured and re-organized following the LMF principles, grammatical labels have been normalized.

Morphalou is now a well formed XML file, UTF-8 encoded and can be parsed by modern XML parsers.

For evaluation purposes we also provide a web interface where lexical entries can be accessed by XPath queries:http://loreley.loria.fr/morphalou/morphalou_req.html

Version : 09.04

Web site : http://loreley.loria.fr/morphalou

Documentation : http://loreley.loria.fr/morphalou/morphalou_doc.html

Contact : Ingrid Falk

Project(s) : CPER/ILDSTC

5.13. Segment Server

Description: automata-based sentence segmentor.

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

Version: 01

Download: http://www.loria.fr/~azim/

Contact: Azim Roussanaly

Project(s): EVALDA/EASY, MIAMM, Ozone, XMiner

5.14. The SYNTAX Suite

Description: Syntax Suite is an implementation of the Data Category Registry (ISO12620) and it provides online tools to manipulate and create Data Categories.

During one year, the Syntax Suite has been developed according to the LAMP model (Lunix, Apache, MySql, PHP) and is now a powerful open-source framework for search and compare DC, makes Data Category Selection, can propose DC to ISO committee and is able to import and export DC in Generic Mapping Tool format.

Syntax Suite now constitutes the first online DCR and shares more than 1500 differents DC over differents domains such as Terminology, Morpho-syntax, Sementic, Meta-data, DC-Administration and so on. More than forty internationals experts in linguistics already use our framework and we continue developments to produce tool that allows experts to normalize any XML document into a meta-model and DC format (ex : Terminological Markup Framework ISO 16642).

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

Contact: Julien Ducret

5.15. 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

5.16. vnACCMS

Description: vnACCMS is a tool kit that manages all the processes of morpho-syntactic annotation for Vietnamese corpora. The choice of resource coding is especially to justify normalisation propositions in the framework of the ISO subcommitee TC 37 / SC 4.

Last update : 2003-11-30
Author(s) : Thanh Bon Nguyen (internship from IFI Hanoi)
Contact : Thi-Minh-Huyen Nguyen
Project(s) : Vietnam

5.17. vnQTAG - QTAG for Vietnamese

Description: QTAG is a probabilistic POS tagger developed by Oliver Mason (University of Birmingham),
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 research carried out in the field of multilingual information extraction to answer the more industrial concerns 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 an 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. Terminological Activities in SYNTAX

Context: Developed in ISO TC37/SC3 context (computer applications in terminology), TMF (Terminological Markup Framework) designs a platform for representing terminological data. For the short storie, TMF results from the competition between two formats TBX, now adopted by the localization industry (http://www.lisa.org) and Geneter.

TMF allows to express constraints on the representation of computerized terminologies. What is the underlying structure of computerized terminologies? Which data categories are used and under which conditions,

and finally how to maintain interoperability between representations and to provide a conceptual tool to compare two given formats.

TMF is a meta model, in fact it is a skeleton structure which can be decorated by some data categories. Data category is an single label to design a specific terminological concept of terminology. TMF with the ISO 12620 data categories allows better interoperability.

TERM'IC project: INIST (Institut National pour l'Information Scientifique et Technique.) provides indexing services of literature in Science, Technology, Medicine, Humanities and Social Sciences designed to contribute to bibliographic databases. INIST is a repository of large document collections available to the public through document delivery services.

The TERM'IC project joints project between INIST and Loria-INRIA, which should define a common framework for managing terminological data at INIST and background for better knowledge management activities (e.g. Semi-automatic indexing, Information extraction, Text mining,etc).

The principal objectives awaited by INIST are to improve at the same time indexing activities but also the model-based query language (in natural language) by the constitution of a knowledge base in order:

- to automate indexing while keeping an acceptable level of quality
- to associate with the terminological resources the linguistic information adapted to the NPL, so as to increase the share of the automatic processing
- to improve the methods of constitution and management terminology
- to capitalize it and share the knowledge on the terminologies and indexing

A base of terminological knowledge would be defined as a tank single and central terminological resources enriched with some linguistic informations (e.g. morphosyntaxic and semantic).

The standardization of the terminological products will ensure the continuation of a stock of knowledge, which in a generic format could be exploited on various bases (Web) and thus ensured INIST the image of one activity to the point of technicality and progress.

What we do: Created in 1992, Francis and Pascal databases propose some descriptors in few language to variety of disciplines, so a galaxie of bases is now so heterogeneous.

In order to have an unified view on terminological information at INIST, a terminological knowledge base proposes a concept based approach, based on application of ISO 704 principles and ISO 16642 modelling approach.

The single view on all terminological collections will present a modeling the informational coverage in the various databases which can be elaborate after identiying of generic usage scenarios.

In order to constitue a real a terminological knowledge base, we decide to Define a transitional model for unifying the databases, which lets time to INIST to clean and complete his bases. In a first time, the BCT will be virtual with a Dataflow management (hub) and at the end of the project we can merge all the bases in a same and single BCT.

Perpectives: publication of White paper, which detail « how to do TMF » with all SYNTAX experiences

- PROTEUS
- WORLDC
- ODEXTERM

2004 november 7th: Meeting with INIST staff will determine the next and final step : who and how clean or complete the bases.

Proposition of Terminology and IST project : how introduce some methodology in a terminological work.

6.2. Linguistic description and analysis

Current research (carried out, in part, in collaboration with colleagues at the LLF/Paris VII and at ATILF/Nancy 2) provides new insights into a number of major problems in the grammatical analysis of French and other languages. The topics addressed are:

- the treatment of the French functional elements _à_ and _de_, with particular attention to their non-prepositional uses (publications [21] and [22]). The proposed analysis relies on the notion of "weak head", a syntactic head that shares its syntactic category (part of speech) with its complement.

- the treatment of consonant liaison in French (publication [24]). The proposed approach allows the integration of the diverse constraints that determine the realization or non-realization of liaison in a given context: lexical properties, syntactic conditions, morphophonological features, prosodic structure, stylistic effects.

- a fully lexicalist analysis of phrase marking as edge affixation, for the treatment of morphosyntactic phenomena in a variety of languages (e.g., French, English, Turkish, Basque, and many others) (publications [62] and [63]).

The significance of these results is primarily theoretical, but the proposals are formalized explicitly enough to be implemented computationally. Most of the analyses for French have been successfully incorporated into the HPSG/LKB grammar under development in LED. Although the linguistic phenomena studied play a relatively minor role in semantic interpretation (and therefore in the NLP task of parsing), they are of critical importance when it comes to the generation of well-formed linguistic output.

6.3. Synchronization between syntactic and semantic knowledge

Within the framework of applications having to react to orders expressed in natural language, the work to achieve to carry out the task expressed by a speaker is divided into 3 distinct phases: a speech recognition stage which aims at transforming the sound signal into segments of text, a stage of validation stage which determines if these statements can be analysed, and a translation stage which specifies the meaning of the action described by the initial signal.

The thesis of Djame Seddah lies between the second and third stages: : we have to first know if the sentences can be understood - from our perspective, this means verifying that it is syntactically correct and inferring a useful semantics for the application. - and then we have to infer a semantic meaning useful for the application.

In fact we should be able to answer some questions such as "who is doing what in this sentence ?", "what is its modality ?". Some answers to this questions deal with discourse semantic, enunciation syntax or a complex cognitive analysis. For the most part, the interest of an syntax-semantic interface comes from the fact that different sentences with different syntactic structures but with the same meaning must have the same logical structure.

Thus, the following sentences can be analysed such as they all expressed the fact that there exists an entity called John, an entity called Mary and that between them it exists a predicative link denoted by 'TO_LOVE'.

1. John loves Mary
2. Mary is loved by Jonh
3. John is in love with Mary

A single dependency graph can represent this predicative structure in a simple way..

We use Lexicalized Tree adjunct grammars (LTAG) as our syntactic framework. A LTAG grammar is a lexicon where each lexical element (called anchor) is associated to a set of trees which describes its subcategorization framework and its morpho-syntactic features. A LTAG analysis provides two structures of representation : a syntactic tree (ie the derived tree) and a derivation tree which is strictly the record of how the derived tree has been built. The derivation tree, following that the LTAG grammar follows well known

linguistic principles, is supposed to be closed to a predicative structure.Therefore, the choice of LTAG as syntactic back end consolidates the two last stages of analysis (validation and translation) into a single step.

But derivation trees have trees have three crucial problems to be solved:

- they can represent only one analysis

- the dependencies they include are sometimes incomplete

- and most importantly, not all argumental relations are always present.

In this work, it is argued that "shared forests" (as opposed to LTAG derivation and derived trees) are the optimal solution for building predicate-argument structure. This takes the form a graph that can then be used as a basis for further semantic analysis.

Through the study of linguistic phenomena induced by control-verbs, we define a new form of lexical information(called control canvas) needed to help the inference of argumental missing links. We also propose new inference rules expressed in a tabular parsing framework in order to build dependency graphs via a specific form of a shared forest called "derivation forest".

Gardent03 [66] presents a unification based syntax-semantic interface for Tree Adjoining Grammar. The theoretical proposal presented in this paper has since been integrated in the core grammar for French developed using the XMG metagrammar compiler so that a core grammar for French is now available which describes both the syntactic structure of a constituent and its compositional semantics. This grammar (called semFrenchTAG) is currently being tested both in analysis and in generation mode.

In generation, the GenI surface realiser has been optimised (DEA thesis Nancy I [14]) and is now using semFrenchTAG as a linguistic resource. The resulting system allows for the generation of the grammatical paraphrases of a given input that is, of all variations in the realisation and distribution of syntactic functors and in particular of verbs. Current work involves (i) a systematic evaluation of the results, (ii) an extension of the grammar to other types of paraphrases (cross categorial, morphologically based, etc.), (iii) the use of the generator for debugging the grammar (checking for over and undergeneration) and (iv) the integration of a morphological component.

In analysis, the grammar has been coupled with Eric de la Clergerie DYALOG TAG parser and two ways of doing semantic constructions are currently under investigation, one where semantic construction is done during derivation, the other where it is done afterwards on the basis of DYALOG derivation forest. The aim here is on the one hand, to compare the relative efficiency of both approaches and on the other hand, to investigate different ways of doing semantic construction using different semantic representation languages and different combining mechanism.

6.4. Lexicon management

An important work have been achieved in the field of wide-coverage grammars production. In collaboration with the Calligramme team, a robust MetaGrammar (MG) compiler for Tree Adjoining Grammars (TAG) and Interaction Grammars (IG) has been developed: the eXtensible MetaGrammar (available at http://sourcesup.cru.fr/xmg). The techniques used for the implementation of this compiler are presented in [31], and the representation language used to describe linguistic information with XMG is given in [28].

The main feature of this representation language corresponds to the fact that it includes disjunction and inheritance, and thus allows linguists to describe lexical information following two axes: "structure sharing" and "alternations".

Furthermore the XMG compiler offers additional linguistic principles (gathered in a library) and thus extend the compilation process with linguistic constraints on the produced TAG trees. XMG has been used to produce a wide-coverage TAG grammar for French. This grammar mainly concerns verbal and adjectival predicates. A documentation of this grammar has also been realised in the context of the EVALDA project (http://www.loria.fr/~crabbe/frenchtag/index.html).

Among ongoing work is the development of a wide lexicon adapted to such grammars. This will lead to the specification of a normalised interface (LMF). Another futur step consists of the evaluation of automatically produced grammars through specific tests.

6.5. 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.

7. Other Grants and Activities

7.1. International level

7.1.1. AMIGO Project

Description: The Amigo project will focus on the usability of a networked home system by developing open, standardized, interoperable middleware. The developed middleware will guarantee automatic dynamic configuration of the devices and services within this home system by addressing autonomy and composability aspects. The second focus of the Amigo project will be on improving the end-user attractiveness of a networked home system by developing interoperable intelligent user services and application prototypes. The LORIA will provide an expertise on multimodality needed in these user services.

Contact(s): Matthieu Quignard, Alexandre Denis

Context: IST European Program

Responsible: Harmke deGroot (Philips, Eindhoven)

Start: 2004-09-01

2008-03-01

Partners: Philips Research, Eindhoven (+ other partners, see the project website)

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

7.1.2. Comparative syntax of French and Polish: Formalization and Implementation

Theme: syntax of French and Polish, computational grammars

Description: This project considers several aspects of the comparative syntax of French and Polish in the formal framework of Head-Driven Phrase Structure Grammar (HPSG) with the dual aim of enriching theoretical grammars for the two languages and extending NLP applications. The phenomena studied include interrogative and relative clause constructions, comparative constructions, coordination, and negation.

Administrative context: French-Polish collaboration jointly supported by the CNRS and the Polish Academy of Sciences (DRI/CNRS project number 14482)

Period: start 2003-01-01 / end 2004-12-31

Contact: Jesse Tseng

Partner(s): Laboratoire de Linguistique Formelle (Paris 7 University), Insitute of Computer Science of the Polish Academy of Sciences (Warsaw)

7.1.3. 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 digitized 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 for 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 the 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

7.1.4. 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

7.1.5. 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

7.1.6. 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.

7.1.7. 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.

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

Contact: Patrick Blackburn

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

7.1.8. LFG ParGram

Lexical Functional Grammar Parallel Grammar Project

Theme: parallel grammar development, multilingual systems

Description: The aim of the project is to produce wide coverage grammars for Chinese, English, French, German, Japanese, Malagasy, Norwegian, Urdu, and Welsh. These are written collaboratively within the linguistic framework of LFG and with a commonly-agreed-upon set of grammatical features.

In 2004 "Langue et Dialogue" becomes the new center for French grammar development in the ParGram project.

Web site: http://www2.parc.com/istl/groups/nltt/pargram/
Contact: Jesse Tseng
Period: start 2004-11-01
Partner(s): Palo Alto Research Center, Fuji Xerox, University of Essex, Oxford University, UMIST, University of Stuttgart, University of Bergen, University of Konstanz

7.2. European level

7.2.1. *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)

7.2.2. *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

7.2.3. 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

7.2.4. *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

7.2.5. *Proteus*

Fully integrated platform designed to support any broad e-maintenance strategy.

Theme : Knowledge Management. Document engineering. Industrial e-maintenance.

Description: The Proteus project is aimed at proposing an integration platform which federates a collection of existent software tools not designed originally to work in co-operation and as such not adapted to form a coherent system allowing for disciplined information transformation and exchange. The various software tools the platform has to deal with are CMMS, ERP, SCADA, e-Documentation server. To make them communicate a set of generic WebServices for Maintenance has been designed. "Langue et Dialogue" was originally part of the e-Documentation Work Package but proposed also the design and the use of an ontology of maintenance (compatible with OWL-DL) to ease the co-operation of the different tools through a common definition of the objects manipulated. This proposal was based on its own will to integrate a domain description as an Access Point into the e-Documentation server architecture. Also integrated as Access Points are a Terminology for maintenance normalized with TMF and a set of Metadata extracted and adapted from Dublin Core Document Metadata Element Set. The e-Documentation server architecture is built mainly for handling structured XML documents and to allow multilingual access.

Administrative context: ITEA

Web site : http://www.proteus-iteaproject.com"

Person in charge : Denis Chevé

Period : start 2002-10-01 / end 2005-02-27

Contact : Laurent Romary, Joseph Roumier

Software : Document Management plugin

Partner(s) : CEGELEC, CEGELEC AT, AKN, Arc Informatique, BiKit, TIL Technologies, IFAK, IML Fraunhofer, LIP6, LAB, Pertinence, Schneider Electric, TUM, VARTEC, MAIA, INRIA Nancy, TRIO, INRIA Nancy

7.3. National level

7.3.1. *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 corpora 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

7.3.2. *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)

7.3.3. *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.

This project has been closed by a national workshop in Paris (Dec. 2003). Many of its scientific activities are still carried out within a new national workgroup on Animated Conversational Agents.

Administrative context: CNRS

Person in charge: Salima Hassas, Alain Mille (LIRIS)

Period: start 2003-07-11

Contact: Matthieu Quignard

Partner(s): EMSE-SIMMO, IRIN, CRISTO, LIRIS

7.3.4. 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

7.3.5. FReeBank, vers une base libre de corpus annotés

Theme: Linguistic resources

Description: The few available French resources for evaluating linguistic models or algorithms on other linguistic levels than morpho-syntax are either insufficient from quantitative as well as qualitative point of view or not freely accessible. Based on this fact, the FREEBANK project intends to create French corpora constructed using manually revised output from a hybrid Constraint Grammar parser and annotated on several linguistic levels (structure, morphosyntax, syntax, coreference), with the objective to make them available on-line for research purposes. Therefore, we will focus on using standard annotation schemes, integration of existing resources and maintenance allowing for continuous enrichment of the annotations.

Context: CNRS

Responsable:Susanne Salmon-Alt

Partners: University of Southern Denmark, Odense (Danemark) - ATILF, Laboratoire Analyses et Traitement Informatiques du Lexique Français, Nancy (France)

7.3.6. 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

7.3.7. 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)

7.3.8. 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

7.3.9. 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

7.3.10. 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

7.3.11. 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

7.3.12. 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

7.4. Regional level

7.4.1. 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"

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

8. Dissemination

8.1. Service to the scientific community

8.1.1. 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)
- 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)
- 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 "Inférence et Génération 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 Ministry of Research on the national prospects for computing and intensive storage for the social science (Jean-Marie Pierrel)

8.1.2. 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 journal: "Revue d'Intelligence Artificielle" (RIA), Editions Hermès (Jean-Marie Pierrel)
- Member of the editorial committee of "La lettre du LORIA" (Claire Gardent)
- 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 editorial committee for: Intellectica (Daniel Coulon)
- 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 "Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications" (Claire Gardent)
- Member of the reading committee for: ARCo'04 (Daniel Coulon)

8.1.3. Conference and workshop organization

- 16th European Summer School in Logic, Language and Information (ESSLLI 2004 - http://esslli2004.loria.fr), Nancy, France, 9 - 20 August, 2004 (Carlos Areces, Patrick Blackburn)

ESSLLI is the yearly summer school of FoLLI, the Association for Logic, Language and Information (http://www.folli.org) and has been organized since 1989 in different countries in Europe.

ESSLLIs offer an opportunity to attract the world's leading specialists in the areas of Logic, Language and Information, and is attended by students from many countries. In this year's instance, ESSLLI reported about 500 participants from 40 different countries. 48 one week long courses were organized in three parallel sesions during the 2 weeks of the Summer School, by lecturers of world renown as Ruth Kempson, Rob van der Sandt, Vladimir Lifschitz or Andrei Voronkov. Moreover, four plennary evening lectures, open to the general public were organized with talks by Gerard Huet, Vladimir Lifschitz, Karen Sparck Jones and Elisabeth Andre.

A number of members of LED were involved in the organization: Carlos Areces, Patrick Blackburn, Benoît Crabbé, Bertrand Gaiffe, Satyendra Gupta, Evelyne Jacquey, Laurence Kbida, Eric Kow, Jacqueline Lai, Frédéric Landragin, Hélène Manuélian, Thi Minh Huyen Nguyen, Yannick Parmentier, Matthieu Quignard, Djamé Seddah and Jesse Tseng. The event received as well a strong support from LORIA, INRIA and the University Henri Poincare.

ESSLLI 2004 in Nancy was very sucessful. As all ESSLLIs, it offered a selection of courses in the field of Logic, Language and Information impossible to find in any single University or Institute. And, judging from the registration (70 people registered were affiliated to LORIA) many researchers of LORIA took advantage of this possibility. But, most importantly, the event draw the attention of an important number of researchers and students all over the world to the activities carried out in LORIA.

8.2. 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 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 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, Hélène Manuélian, Azim Roussanaly)

- Participation in the DESS "TEXTE"; at the University of Nancy 2(Benoît Crabbé, Claire Gardent, Yannick Parmentier, Jean-Marie Pierrel, Laurent Romary)

- 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 Metz (Hélène Manuélian)
- Participation in the course of "Mathématiques, Informatique et Statistiques Appliquées aux Sciences Humaines et Sociales" at the University of Nancy 2 (Azim Roussanaly)
- Participation in the course of "Mathématiques, Informatique et Application aux Sciences" at the University of Nancy 1 (Yannick Parmentier)
- Member of the international relations service of the IUT Nancy-Charlemagne (University of Nancy 2 (Samuel Cruz-Lara)
- 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 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. Bibliography

Major publications by the team in recent years

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

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

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

[4] D. DUCHIER, C. GARDENT. *Tree Descriptions, Constraints and Incrementality*, in "Computing Meaning", Studies in Linguistics and Philosophy Series, vol. 2HARRY BUNT, REINHARD MUSKENS, ELIAS THIJSSE (editors)., Kluwer Academic Publishers, 2001.

[5] NANCY IDE, LAURENT ROMARY. *Encoding syntactic annotation*, in "Treebanks"ANNE ABEILLÉ (editor)., 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*, Traité IC2 (Information, communication et commande, Hermes, octobre 2000.

[8] LAURENT ROMARY, PATRICE BONHOMME. *Parallel Alignment of Structured Documents*JEAN 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.

Books and Monographs

[12] FRANCIS CORBLIN, CLAIRE GARDENT. *Interpréter en contexte*, Traité Sciences Cognitives, Hermès, Dec 2004.

Doctoral dissertations and Habilitation theses

[13] CLAIRE GARDENT. *Traitement automatique des langues et inférence*, habilitation à diriger des recherches, Nancy 1, Sep 2004.

[14] ERIC KOW. *Optimising a surface realiser*, Mémoire de DEA, Université Henri Poincaré - Nancy 1, Jun 2004.

[15] HÉLÈNE MANUÉLIAN. *Descriptions Définies et Démonstratives : Analyses de corpus pour la génération*, thèse d'université, Université Nancy 2, november 2003.

[16] DJAMÉ SEDDAH. *Synchronisation des connaissances syntaxiques et sémantiques pour l'analyse d'énoncés en langage naturel à l'aide des grammaires d'arbres adjoints lexicalisées*, Thèse de Doctorat, Université Henri Poincaré - Nancy 1, Nov 2004.

Articles in referred journals and book chapters

[17] CARLOS ARECES, PATRICK BLACKBURN, MARTEEN MARX. *Repairing the interpolation theorem in Quantified Modal Logic*, in "Annals of Pure and Applied Logic", Mar 2004.

[18] FRANCIS CORBLIN, CLAIRE GARDENT. *Contextes et interprétationInterpréter en contexte*, Traité Sciences Cognitives, Hermès, Dec 2004.

[19] CLAIRE GARDENT. *Interprétations minimales et interprétations maximales*, in "Interpréter en contexte", Hermès, .

[20] FRÉDÉRIC LANDRAGIN. *Compte rendu de lecture de l'ouvrage "Les actes de langage dans le discours" de C. Kerbrat-Orecchioni*, in "Bulletin de la Société Linguistique de Paris", vol. XCVIII, nº 2, p. 209-214, Apr 2003.

Publications in Conferences and Workshops

[21] ANNE ABEILLÉ, OLIVIER BONAMI, DANIÈLE GODARD, JESSE TSENG. *The syntax of French "à" and "de": An HPSG analysis*, in "ACL-SIGSEM Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications"Toulouse, France, Sep 2003.

[22] ANNE ABEILLÉ, OLIVIER BONAMI, DANIÈLE GODARD, JESSE TSENG. *The Syntax of French de-N' Phrases*, in "11th International Conference on Head-Driven Phrase Structure Grammar - HPSG'2004"Leuven, Belgium, p. 6-26STEFAN MÜLLER (editor)., CSLI Publications, Nov 2004.

[23] ABDEL BELAÏD, ISABELLE TURCAN, JEAN-MARIE PIERREL, YOLANDE BELAÏD, YVES RANGONI, HASSEN HADJAMAR. *Automatic indexing and reformulation of ancient dictionaries*, in "First International Workshop on Document Image Analysis for Libraries - DIAL'2004"Palo Alto, USA, p. 342-354, IEEE, Jan 2004.

[24] OLIVIER BONAMI, GILLES BOYÉ, JESSE TSENG. *An integrated approach to French liaison*, in "9th conference on Formal Grammar 2004 - FGNancy"Nancy, France, p. 29-45GERHARD JÄGER, PAOLA MONACHESI, GERALD PENN, SHULY WINTNER (editors)., Aug 2004.

[25] DAAN BROEDER, THIERRY DECLERCK, LAURENT ROMARY, ERIC DE LA CLERGERIE, MARKUS UNESON, SVEN STRÖMQVIST, PETER WITTENBURG. *A large Metadata Domain for Language Resources*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbon, Portugal, May 2004.

[26] HARRY BUNT, LAURENT ROMARY. *Standardization in Content Representation: some methodological issues*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbon, Portugal, May 2004.

[27] LOU BURNARD, LAURENT ROMARY. *Kompetenzzentrum für EDV-Philologie*, in "TEI Workshop"Würzburg, Allemagne, Oct 2004.

[28] BENOÎT CRABBÉ, DENYS DUCHIER. *Metagrammar Redux*, in "International Workshop on Constraint Solving and Language Processing - CSLP 2004"Roskilde, Denmark, Sep 2004.

[29] SAMUEL CRUZ-LARA, SATYENDRA GUPTA, LAURENT ROMARY. *Handling Multilingual Content in Digital Media\ : the multilingual information framework*, in "IEE European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology"London, UK, Nov 2004.

[30] EMMANUEL DESVEAUX, LAURENT ROMARY. *Ethnonymie et ingénierie sémantique*, in "Sources et ressources pour les sciences sociales"Paris, France, Dec 2004.

[31] DENYS DUCHIER, JOSEPH LE ROUX, YANNICK PARMENTIER. *The MetaGrammar Compiler : An NLP Application with a Multi-paradigm Architecture*, in "Second International Mozart/Oz Conference - MOZ 2004"Charleroi, Belgique, Oct 2004.

[32] BERTRAND GAIFFE, FRÉDÉRIC LANDRAGIN, MATTHIEU QUIGNARD. *Le dialogue naturel comme un service dans un contexte multi-applicatif*, in "Workshop of the Association pour le Traitement Automatique des LAngues (ATALA) journée AGENTAL : "Agents et Langue""Paris, France, Mar 2004.

[33] CLAIRE GARDENT, MARILISA AMOIA, EVELYNE JACQUEY. *Paraphrastic grammars*, in "2nd International Workshop on Text and Meaning and Interpretation"Barcelona, Spain, Jul 2004.

[34] NANCY IDE, LAURENT ROMARY. *A Registry of Standard Data Categories for Linguistic Annotation*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbon, Portugal, May 2004.

[35] FRÉDÉRIC LANDRAGIN. *L'utilisation de scores numériques en sémantique computationnelle*, in "Journées Scientifiques Sémantique et Modélisation - JSM'04"Lyon, France, Mar 2004.

[36] FRÉDÉRIC LANDRAGIN. *Interface sémantique-pragmatique et domaines de référence*, in "Quatrièmes Journées d'Etudes Linguistiques de Nantes - JEL'2004"Nantes, France, May 2004.

[37] FRÉDÉRIC LANDRAGIN, ALEXANDRE DENIS, ANNALISA RICCI, LAURENT ROMARY. *Multimodal Meaning Representation for Generic Dialogue Systems Architectures*, in "4th International Conference on Language Resources and Evaluation - LREC'2004"Lisbon, Portugal, May 2004.

[38] FRÉDÉRIC LANDRAGIN, LAURENT ROMARY. *Dialogue History Modelling for Multimodal Human-Computer Interaction*, in "Eighth Workshop on the Semantics and Pragmatics of Dialogue - Catalog'04"Barcelona, Spain, p. 41-48, Jul 2004.

[39] KIYONG LEE, SYD BAUMAN, LOU BURNARD, HARRY BUNT, LIONEL CLÉMENT, ERIC DE LA CLERGERIE, THIERRY DECLERCK, TOMAZ ERJAVEC, LAURENT ROMARY. *Towards an international standard on feature structures representation*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbon, Portugal, May 2004.

[40] HÉLÈNE MANUÉLIAN. *Generating Coreferential Descriptions from a Structured Model of the Context*, in "4th International Conference on Language Ressources and Evaluation - LREC 2004"Lisbon, Portugal, May 2004.

[41] K MCTAIT, H MAYNARD, D MOSTEFA, L DEVILLERS, S ROSSET, P PAROUBEK, K CHOUKRIL, J Y ANTOINE, F BÉCHET, C BOUSQUET, N VIGOUROUX, O BONTRON, L CHARNAY, L ROMARY, M VERGNES. *Constitution d'un corpus de dialogue oral pour l'évaluation automatique de la compréhension hors- et en- contexte du dialogue*, in "JEP 2004"Fès, Maroc, Apr 2004.

[42] ERICA MEENA, ASHWANI KUMAR, LAURENT ROMARY. *An Extensible Framework for Efficient Document Management Using RDF and OWL*, in "Workshop ACL on RDF and XML"Barcelona, Spain, Jul 2004.

[43] THOMAS MONTROIG, JOSEPH ROUMIER, NICOLAS BLANC, LAURENT ROMARY, JACEK SZYMANSKI, DENIS REBOUL. *Building Conceptual Models by Knowledge Management Methodology*, in "IEEE Mechatronics & Robotics - MechRob'2004"Aachen, Germany, vol. 3, p. 1140-1145PAUL DREWS (editor)., IEEE, Sep 2004.

[44] THANH BON NGUYEN, THI MINH HUYEN NGUYEN, LAURENT ROMARY, XUAN LUONG VU. *Lexical descriptions for Vietnamese language processing*, in "The 1st International Joint Conference on Natural Language Processing - IJCNLP'04 / Workshop on Asian Language Resources"Sanya, Hainan Island, China, Mar 2004.

[45] THANH BON NGUYEN, THI MINH HUYEN NGUYEN, LAURENT ROMARY, XUAN LUONG VU. *Developing Tools and Building Linguistic Resources for Vietnamese Morpho-Syntactic Processing*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbonne, Portugal, May 2004.

[46] ANDRÉI POPESCU-BELIS, LOÏS RIGOUSTE, SUSANNE SALMON-ALT, LAURENT ROMARY. *Online Evaluation of Coreference Resolution*, in "4th International Conference on Language Resources and Evaluation - LREC'04"Lisbon, Portugal, May 2004.

[47] LAURENT ROMARY, AMALIA TODIRASCU, DAVID LANGLOIS. *Experiments on Building Language Resources for Multi-Modal Dialogue Systems*, in "4th International Conference on Language Resources and Evaluation - LREC'2004"Lisbon, Portugal, May 2004.

[48] LAURENT ROMARY. *ISO and TEI*, in "TEI Annual Members' Meeting"Baltimore, Maryland, Oct 2004.

[49] LAURENT ROMARY. *Position paper on data categories for data modelling*, in "DELOS brainstorming meeting", Jul 2004.

[50] LAURENT ROMARY. *Apports possible de la normalisation à la linguistique de corpus*, in "Ecole thématique linguistique de corpus : constitution, archivage, évaluation"Caen, France, June 2004.

[51] LAURENT ROMARY. *Computational morphology workshop*, in "Série de 5 conférences invitées à l'UNISA (University of South Africa) et l'Université de Stellenbosch"Pretoria, Afrique du Sud, Sep 2004.

[52] LAURENT ROMARY. *Perspectives de normalisation pour les corpus et lexiques multilingues*, in "Lexicométrie et corpus multilingues - JADT 2004"Louvain-la-Neuve, Belgique, Mar 2004.

[53] LAURENT ROMARY. *Normalisation, formation et recherche*, in "La normalisation est-elle une discipline ?", Jan 2004.

[54] LAURENT ROMARY. *Construction of Grammar Based Term Extraction Model for Japanese*, in "3rd International Workshop on Computational Terminology - CompuTerm 2004"Geneva, Switzerland, Aug 2004.

[55] LAURENT ROMARY, SUSANNE SALMON-ALT, GIL FRANCOPOULO. *Standards going concrete : from LMF to Morphalou*, in "Workshop Enhancing and Using Electronic Dictionaries"Geneva, Switzerland, Aug 2004.

[56] LAURENT ROMARY, NANCY IDE, PETER WITTENBURG, THIERRY DECLERCK. *An on-line data category registry for linguistic resource management*, in "LREC Workshop"Lisbon, Portugal, Sep 2004.

[57] LAURENT ROMARY, ISABELLE KRAMER, PHILIPPE SÉBIRE. *Terminology standards at work: exploring several scenarios to implement ISO 16642*, in "Terminology in Advanced Management Applications. Multilingual Content Integration"Cologne, Allemagne, Nov 2004.

[58] SUSANNE SALMON-ALT, LAURENT ROMARY. *Data Categories for a Normalized Reference Annotation Scheme*, in "5th Discourse Anaphora and Anaphor Resolution Colloquium - DAARC'2004"São Miguel (Azores), Portugal, Sep 2004.

[59] SUSANNE SALMON-ALT, ECKHARD BICK, LAURENT ROMARY, JEAN-MARIE PIERREL. *La FReeBank : vers une base libre de corpus annotés*, in "Traitement Automatique des Langues Naturelles - TALN'04"Fès, Maroc, April 2004.

[60] SUSANNE SALMON-ALT, LAURENT ROMARY. *Towards a Reference Annotation Framework*, in "LREC'04"Lisbon, Portugal, May 2004.

[61] JESSE TSENG. *LKB grammar development: French and beyond*, in "Workshop on Ideas and Strategies for Multilingual Grammar Development"Vienna, Austria, p. 91-97EMILY BENDER, DAN FLICKINGER, FREDERIK FOUVRY, MELANIE SIEGEL (editors)., Aug 2003.

[62] JESSE TSENG. *Phrasal affixes and French morphosyntax*, in "Proceedings of Formal Grammar 2003 FGVienna"Vienne, Autriche, p. 177-188GERHARD JÄGER, PAOLA MONACHESI, GERALD PENN, SHULY WINTNER (editors)., May 2003.

[63] JESSE TSENG. *Case marking on heads and at edges*, in "proceedings of the 2004 International Conference of the Linguistic Society of Korea - LSK'2004"Seoul, South Korea, vol. 1, p. 230-239, Jul 2004.

Bibliography in notes

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

[65] 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", Education RechercheCOLETTE DEAUDELIN, THÉRÈSE NAULT (editors)., vol. 8, chap. 6, p. 121-134, Presses Universitaires du Québec, January 2003.

[66] 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, April 2003.

[67] MICHEL MUSIOL, DANIEL COULON . *L'influence du processus conversationnel sur la dynamique de la croyance assertive*, in "Psychologie de l'interaction", n° 17-18, 2003.

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

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