INRIA

Project-Team LeD

Language and Dialog

Lorraine

THEME SYM

Activity

Report

2005

Table of contents

1. Team

Langue et Dialogue (Language and Dialog) 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

Patrick Blackburn [Research Director INRIA]

Vice-head of project team

Claire Gardent [Research Scientist SHS department CNRS]

Administrative assistant

Isabelle Blanchard [CNRS]

Research Scientists

Carlos Areces [Research Scientist INRIA]

Bertrand Gaiffe [Research Scientist INRIA]

Matthieu Quignard [Research Scientist SHS department CNRS (CR2, section 34)]

Laurent Romary [Research Director INRIA (temporary leave to CNRS)]

Jesse Tseng [Research Scientist SHS department CNRS]

Faculty Members

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

Daniel Coulon [Professor INPL]

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

Christine Fay-Varnier [Assistant Professor Ecole de Géologie INPL]

Jean-Marie Pierrel [Professor UFR STMIA UHP Nancy 1]

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

Ph.D. Students

Benoît Crabbé [UHP Nancy 1, INRIA grant (ATER IUT Nancy-Charlemagne) until 2005-07]

Alexandre Denis [INRIA grant]

Sébastien Hinderer [UHP Nancy 1]

Eric Kow [UHP Nancy 1, INRIA grant]

Cheng Yu Lu [INRIA joint with Taiwan from 2005-07-19]

Chia-Hung Lin [INRIA joint with Taiwan]

Thi-Minh-Huyen Nguyen [UFR STMIA UHP Nancy 1]

Yannick Parmentier [UFR STMIA UHP Nancy 1 (INRIA grant - UHP Nancy 1 Monitor)]

Joseph Roumier [ENSEM INPL (CIFRE grant financed par CEGELEC)]

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

Daniel Gorin [UHP Nancy 1, joint with Universidad de Buenos Aires]

Sergio Mera [UHP Nancy 1, joint with Universidad de Buenos Aires]

Visiting Scientists

Nancy Ide [Visiting Scientist (Vassar College (USA))]

Evelyne Jacquey [Visiting Scientist département STIC CNRS]

Susanne Salmon-Alt [Visiting Scientist CNRS (ATILF)]

Post-doctoral Fellows

Hyung Suk Ji [INRIA until 2005-11-30]

Anna Kupsc [CNRS until 2005-12-31]

Guillaume Pitel [from 2005-01-01 until 2005-12-31]

Technical Staff

Julien Ducret [INRIA]

Ingrid Falk [CNRS]

Ulrike Fleury [assistant INRIA until 2005-03-31, temporary agent CNRS from 2005-04-01]
Laurence Kbida [INRIA from 2005-04-01]
Isabelle Kramer [INRIA]
Philippe Sébire [INRIA until 2005-05-31]

Students

Mehdi Ben Abderrahmen [(SFAX, Tunisia) from 2005-09-01 until 2005-09-30]
Javier Fernandez Garcia [(University of Valladolida) until 2005-02-28]
Mohammed Ghribi [(SFAX, Tunisia) from 2005-09-01 until 2005-09-30]
Jacqueline Lai [INRIA (MIT) until 2005-06-30]
Phuong Le Hong [from 2005-04-01 until 2005-09-30]
Nourredine Loukil [(SFAX, Tunisia) from 2005-09-01 until 2005-09-30]
Jorge Pelizzoni [(Buenos Aires) from 2005-09-20]

2. Overall Objectives

2.1. Overall Objectives

The aim of the LED team can be summarised as follows: to investigate semantic phenomena (broadly construed) in natural languages from a computational perspective. Concretely, LED's goal is to develop wide coverage grammars (with a special emphasis on French) with a semantic dimension, to explore the linguistic and computational issues involved in putting these grammars to work (for both analysis and generation), and in particular to investigate the interplay between representation and inference.

Typical of the questions we wish to address are: what *types* of representations are needed (for example, to detect paraphrases or to represent temporal information)? What *level* of representation is needed (for some purposes, only shallow representations in weak logics are called for while for others, powerful logics seem to be required)? What kind of *knowledge* is needed to support the identified inferences? What type of inference is needed to guide the representation-building and reasoning processes? For example, when is statistical inference enough? When does it suffice to work with weak (decidable) logics? When are more powerful forms of inference (such as first- or higher-order theorem proving and model building) required?

To put it another way, LED focuses on *computational semantics*. This has always been an important theme at LED but, partly as a result of internal changes in the team, its role has become increasingly dominant during last years. First, on January 1st 2005, Patrick Blackburn took over as interim head of the LED team, and Claire Gardent became deputy head. Their areas of competence are mainstream computational linguistics, semantic analysis, and logics for knowledge representation. Their work, and the work of their students, has helped focus the research done within LED towards computational semantics. This focus on computational semantics has been further accentuated by the recruitment in the years 2002 and 2003 of Carlos Areces (CR1 INRIA), Matthieu Quignard (CR2, Section 34) and Jesse Tseng (CR2 CNRS Section 34). Like Patrick Blackburn, Carlos Areces has a background in interdisciplinary approaches to logic, language and information, with special emphasis on computational logic. Jesse Tseng reinforces the computational syntax side of LED. In particular, his arrival has greatly aided the development of linguistic resources in standard frameworks such as Head-Driven Phrase Structure Grammar (HPSG), and Lexical Functional Grammar (LFG). The development of such resources is an essential prerequisite for serious semantic work. Matthieu Quignard's work focuses on dialogue structure, and he has brought a strong cognitive science emphasis to the LED team. Pleasingly, the new members of the team have integrated well with the existing researchers. For example, the recent work on shallow anaphora resolution for French, which was developed and evaluated within the National Evaluation Campaign MeDIA, was carried out under the direction of Matthieu Quignard together with Bertrand Gaiffe (research scientist INRIA).

3. Scientific Foundations

3.1. Introduction

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

3.2. Computational Linguistics and Computational Logic

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

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

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

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

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

3.3. Semantics and Inference

Over the next decade, progress in natural language semantics will likely depend on obtaining a deeper understanding of the role played by inference. One of the simplest levels at which inference enters natural language is as a disambiguation mechanism. Utterances in natural language are typically highly ambiguous: inference allows human beings to (seemingly effortlessly) eliminate the irrelevant possibilities and isolate the intended meaning. But inference can be used in many other processes, for example, in the integration of new

information into a known context. This is important when generating natural language utterances. For this task we need to be sure that the utterance we generate is suitable for the person being addressed. That is, we need to be sure that the generated representations fit in well with the recipient's knowledge and expectations of the world, and it is inference which guides us in achieving this.

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

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

3.4. Linguistic Resources

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

Needless to say, in reality matters are not nearly so straightforward. For a start, for many languages (including French) there are no large-scale resources of the sort that exist for English. Furthermore even in the case of English, the idealised situation just sketched does not obtain. For example, the syntax/semantics interface cannot be regarded as a solved problem: phenomena such as gapping still offer challenging problems for semantic construction.

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

LED is involved in such aspects in a number of ways. For example, it participates in the development of an open source morpho syntactic lexicon for French, in an attempt to lay the ground for a French version of FrameNet; it collaborates with the development of a metagrammar compiler for TAG (to enable the construction of a wide coverage grammar for French); and, more recently, it is involved in the development of LFG and HPSG grammars (offering a syntax/semantics interface) for French.

3.5. Logic Engineering

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

For a start, most of the efforts of the first-order automated reasoning community have been devoted to theorem proving; model building, which is also a required technology for natural language processing, is nowhere nearly as well developed, and far fewer systems are available. Secondly, the first-order reasoning community has adopted a resolutely 'classical' approach to inference problems: their provers focus exclusively on the satisfiability problem. The description logic community has been much more flexible, offering architectures and optimisations which allow a greater range of problems to be handled more directly. One reason for this has been that historically, not all description logics offered full Boolean expressivity. So there

is a long tradition in description logic of treating a variety of inference problems directly, rather than via reduction to satisfiability. Thirdly, many of the logics for which optimised provers exists do not directly offer the kinds of expressivity required for natural language applications. For example, it is hard to encode temporal inference problems in implemented versions of description logics. Fourth, for very strong logics (notably higher-order logics) few implementations exists and their performance is currently inadequate.

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

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

3.6. Empirical Studies

The role of empirical methods (model learning, data extraction from corpora, evaluation) has greatly increased in importance in both linguistics and computer science over the last fifteen years. LED 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, and in the definition of standardised representation formats like TAG-ML. In addition, it has been worked on the applications of linguistic ideas in multimodal settings and multimedia.

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

4. Application Domains

4.1. Modular Grammar Building.

The development of large scale grammars is a complex task which usually involves factorising information as much as possible. While good grammar writing and factorisation environments exist for "non tree grammars" (e.g., HPSG, LFG), this is not the case for "tree based grammars" such as TAG, Interaction Grammars or Tree Description Grammars. The Extended Metagrammar Compiler (XMG) developed at LED remedies this shortcoming while additionally providing a clean and modular way to describe several linguistic dimensions thereby supporting the production of tree grammars with semantic information [16], [17], [18], [12], [15].

4.2. Deep Syntactic and Semantic Analysis.

LED is involved in the development of two computational grammars for deep syntactic and semantic processing of French using HPSG and LFG. These grammar implementation efforts are part of two international projects for multilingual grammar development, the LFG Parallel Grammar Project

http://www2.parc.com/istl/groups/nltt/pargram/ and the Delph-In consortium http://www.delph-in.net. Two examples of this work are the linguistic analysis of prepositions [19], with particular emphasis on the treatment of *de* in French and interface phenomena such as phrasal affixation and French liaison [4].

4.3. Referential expressions

LED has a longstanding interest in the semantics and the processing of referential expressions. In recent years, an extensive corpus annotation has been carried out on 5.000 definite descriptions [6]; an algorithm for generating bridging definite descriptions has been specified and implemented which illustrates the interaction of realisation and inference [2]; a constraint based algorithm for definite description has been proposed which differs from the standard one in that it uses constraints to produce a minimal description; and a shallow anaphora resolver for French has been developed and evaluated within the national evaluation campaign MeDIA.

4.4. Surface Realization

It is known that surface realization (i.e., the production from a fully specified semantics of the sentences verbalising that semantics) is computationally costly. LED has proposed a method that helps reducing the complexity of the task by adapting the idea of polarity marks originally used in parsing. The notion of polarities (developed by Guy Perrier from the Calligramme Team) can be used first as a means of coping with lexical ambiguity and second, as a mechanism to select different realizations obeying given syntactic constraints [14], [11].

4.5. Computational Logics and Computational Semantics.

Members of LED are among the main figures proposing the idea of using inference (and in particular, using computational tools like model builders and theorem provers) as an integral part of different tasks in computational semantics, mainly during semantic construction. The book [1] is nowadays an important reference in this area.

4.6. Higher Order Logics and Semantic Construction.

Montague's classic IL system for semantic construction is a higher order logic. It uses lambda calculus as a crucial ingredient to glue semantic pieces and achieve compositionality. The group has investigated an extension of Montague's calculus with hybrid and temporal operators aiming to provide a better account of tense and a better modelling of Reichenbach's core ideas [3].

4.7. Hybrid Automated Deduction

LED main contribution in this topic has been the design of resolution calculi for hybrid logics, that were then implemented into the HyLoRes theorem prover. In particular, LED members have proved that such calculi can be enhanced with optimisations of order and selection functions without losing completeness. Moreover, the first 'effective' (i.e., directly implementable) termination proof for $\mathcal{H}(@)$ has been recently established and the technique is being extended to more expressive languages [7]. Current work includes optimising performance by use of ordered resolution, adding a temporal reasoning component to HyLoRes, extending the HyLoRes architecture to allow querying against a background theory without having to re saturate the theory with each new query, and testing HyLoRes performance against dedicated first-order provers using a structure-preserving translation method.

4.8. Standardisation

LED strongly contributes to the domain of standardisation. From an organisational point of view, it created and chaired a new ISO technical committee (ISO/TC 37/SC 4; started in 2002); it established the mirror national (RNIL) and European community (Lirics); and it acted to make Nancy the fourth host of the TEI

consortium since January 2005. From a scientific point of view, LED contributed to the publication of ISO 16642 (Terminological Markup Framework, August 2003); to a recently adopted standard on the XML representation of feature structures (ISO 24620-1 - Feature Structures); to a proposal for the representation of lexical structures (ISO WD 24613 - Lexical Markup Framework); to the development of an international Data Category Registry in the domain of linguistic resources (see http://syntax.inist.fr); to the drafting of a proposal for reference annotation, and to the definition of the TEI ODD specification language in the context of the TEI council.

4.9. Multimedia

MLIF (Multi Lingual Information Framework) is being designed as a generic ISO-based mechanism for representing and dealing with multilingual textual information. In the framework of the ITEA Jules Verne and Passepartout projects [8], [13], [21], [20], a preliminary version of MLIF has been associated to digital media within the ISO/IEC MPEG context and dealing with subtitling of video content, dialogue prompts, menus in interactive TV, and descriptive information for multimedia scenes. MLIF comprises a flexible specification platform for elementary multilingual units that may be either embedded in other types of multimedia content or used autonomously to localise existing content.

5. Software

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

A metagrammar compiler generates automatically a grammar from a reduced description called a Meta-Grammar. This description captures the linguistic properties underlying the syntactical rules of a grammar. LED has been working on metagrammar compilation since 2001 and several tools have been developed within this framework starting with the MGC system of Gaiffe et al. to the newly developed XMG system of Crabbé et al.

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

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

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

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

Version: 1.1.4
License: CeCILL
Last update: 2005-09-27
Web site: http://sourcesup.cru.fr/xmg/
Documentation: http://sourcesup.cru.fr/xmg/#Documentation
Project(s): GenI

Authors: Benoît Crabbé, Denys Duchier, Joseph Le Roux, Yannick Parmentier
Contact: Benoît Crabbé, Yannick Parmentier

5.2. GenI Generator

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

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

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

The system can process the output of the Metagrammar compiler being developed by Crabbé (LED) and Duchier (Calligramme).

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

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

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

HyLoRes is implemented in Haskell (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.

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

Version: 2.0
License: GPL
Last update: 2005-09-01
Web site: http://www.loria.fr/~areces/HyLoRes
Documentation: http://www.loria.fr/~areces/HyLoRes
Authors: Carlos Areces, Daniel Gorín and Juan Heguiabehere
Contact: Carlos Areces

5.4. hGEN, a Random Formula Generator

hGen is a random CNF (conjunctive normal form) generator of formulas for sublanguages of H(@, ↓, A, P). It is an extension of the latest proposal of Patel-Schneider and Sebastiane, nowadays considered the standard

testing environment for classical modal logics. The random generator is used for assessing the performance of different provers.

Version: 1.0
License: GPL
Last update: 2005-01-01
Authors: Carlos Areces and Juan Heguiabehere
Contact: Carlos Areces

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

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

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

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

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

SynLex is the result of joint work between LED, ATILF and Calligramme [10], [5]. It is currently being evaluated by the members of the ILF (Institut de la Langue Française) funded LexSynt project, and is partially available at http://www.loria.fr/~gardent/ladl.

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

5.6. Morphalou

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 ensured by an editorial committee located at the ATILF. ATILF also maintains and updates the lexicon. 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-organised following the LMF principles, and grammatical labels have been normalised using the Morphalou tools. 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. Morphalou is free for research and teaching purposes.

Version: 09.04
Last update: 2005-11-01
Project(s): CPER/ILDSTC
Authors: Susanne Salmon-Alt, Ingrid Falk
Contact: Ingrid Falk

5.7. Handling Multilingual Content in ITEA Jules Verne Project

Site presenting some useful documents and demonstrations related to Multilingual Content in the framework of ITEA Jules Verne and Passepartout projects. At the heart of these systems lies a flexible specification platform for elementary multilingual units that may be either embedded in other types of multimedia content or used autonomously to localise existing content.

Version: 0.1
Last update: 2004-12-08
Project(s): Jules Verne
Authors: Javier David Fernandez Garcia, Satyendra Gupta, Samuel Cruz-Lara
Contact: Samuel Cruz-Lara

5.8. MEDIA

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

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

5.9. CURT (Clever Use of Reasoning Tools)

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

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

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

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

Version: 1.0
License: GPL
Last update: 2005-10-11
Authors: Carlos Areces, Patrick Blackburn, Johan Bos, Sébastien Hinderer, Daniela Solomon
Contact: Carlos Areces, Patrick Blackburn, Sébastien Hinderer.

5.10. Nessie

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

For test purposes, this library has been interfaced with Curt's parser; this allowed us to discover a bug in Curt's beta-reduction engine, which is now fixed.

Curt has also been adapted to accept (and build semantic representations for) sentences written in Romanian, Spanish and French.

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

Last update: 2005-10-11
Authors: Sébastien Hinderer
Contact: Patrick Blackburn, Sébastien Hinderer

5.11. Term Extraction Tool

This tool allows the user to identify and extract multi-word units in domain-specific texts. The tool is a hybrid of linguistic and statistic techniques: the former identifies candidate terms by shallow parsing, whereas the latter ranks the extracted candidates (using the C/NC-value method of Frantzi et al., 2000). In order to improve performance, we extract only phrases built around words which are more frequent in the specialised domain rather than in general texts (a frequency list of a newspaper corpus has been used). The tool has been tested on medical texts written in Polish (diabetes patients' reports) and French (colonoscopy reports).

Version: 0.1
License: BSD
Last update: 2005-09-01
Authors and Contact: Anna Kupsc

5.12. DeDe Corpus

DeDe is a corpus of roughly 50.000 words where around 5.000 definite descriptions have been annotated as coreferential, contextually dependent, non referential or autonomous. The corpus consists of articles from the newspaper *Le Monde* and is annotated with Multext-based morphosyntactic information. It will be made available on the web in June 2006 at the latest [6].

Authors and Contact: Claire Gardent

5.13. SemFRaG

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

Authors: Claire Gardent, Benoit Crabbé
Contact: Claire Gardent

5.14. FRoG — French Resource Grammar

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

Version: 0.1

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

5.15. ParGram French Grammar

An LFG grammar of French developed with the XLE platform (parsing and generation). A large-coverage, robust implementation integrating a sizeable lexicon and a full finite-state morphology for French, with disambiguation via optimality-theory style markings on candidate analyses.

Version: 0.1
License: Pargram (Xerox)
Last update: 2005-11-01
Web site: http://www2.parc.com/istl/groups/nltt/pargram/
Documentation: http://www2.parc.com/istl/groups/nltt/pargram/
Project(s): LFG ParGram
Authors: Jesse Tseng
Contact: Jesse Tseng

5.16. vnLTAG - A Lexicalized Tree Adjoining Grammar for Vietnamese

This suite of tools comprises the first sizable grammar for Vietnamese LTAG and a parser for the grammar which is based on LLP2. The software includes a segmenter, a grammar for Vietnamese and a parser.

Version: 0.1
Last update: 2005-09-28
Web site: http://wiki.loria.fr/wiki/VnLTAG
Documentation: http://wiki.loria.fr/wiki/VnLTAG
Project(s): Vietnam
Authors: LE Hong Phuong, Thi-Minh-Huyen Nguyen
Contact: Thi-Minh-Huyen Nguyen

5.17. krLTAG: A Lexicalized Tree Adjoining Grammar for Korean

The agglutinative Korean language has several different syntax structures. For instance, while there is no agreement with respect to the subject for verb conjugation in English, in Korean, there is one depending on the relation between the speaker and interlocutor (speech levels). The different sentence structure (Subject-Object-Verb), the absence of prepositions and the existence of postpositions are further differences. The project aims to adapt LLP2 for parsing Korean, making LLP2, in this way, a more general linguistic tool.

Last update: 2005-11-01
Authors and Contact: Hyungsuk Ji

5.18. LLP2

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

Version: 1.0
Last update: 2005-05-31
Web site: http://www.loria.fr/~azim/LLP2/help/fr/
Documentation: http://www.loria.fr/~azim/LLP2/help/fr/

Project(s): EVALDA/EASY, Ozone, XMiner, MIAMM, MEDIA, Vietnam
Authors and Contact: Azim Roussanaly

5.19. Soapical

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

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

2. The Soapmill implementation for Java, which demonstrates how the Soapmill protocol could be used to build blackboard architectures. This implementation has used to build a multi-modal dialogue system in the MIAMM project.

3. The Soapmeter debugging tool, which helps developers to monitor the communication between two SOAP nodes. Soapmeter also provides custom message visualisation through the TAXI editor.

4. Soapkit, which provides a convenient library for working with XML and SOAP in Java. This library is meant to complement SOAP implementations such as Apache Axis.

5. The Soapcaster, which acts as a mediator for web-services in flexible architectures. It maintains a central table which is regularly updated as it receives new subscriptions (or unsubscriptions) and as it verifies the availability of services. All services receive regular updates of these tables, which allows them to know about, to keep track of, and to communicate with each other. Soapcaster is useful for large applications that consist of many, highly similar services.

Version: Soapmill 0.15.2, Soapmeter 0.5.2, Soapkit 0.11
License: BSD
Last update: 2004-05-01
Web site: http://soapical.sourceforge.net/
Documentation: http://soapical.sourceforge.net/
Project(s): MIAMM
Authors and Contact: Eric Kow

5.20. TAXI

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

Version: 1.2.1
License: BSD
Last update: 2002-06-05
Web site: http://taxi.sourceforge.net/
Documentation: http://taxi.sourceforge.net/
Authors: Christophe Coquet
Contact: Eric Kow

6. New Results

6.1. Introduction

The main recent contributions of LED can be organised into two categories. On the one hand a number of tools and resources have been developed and these have been described in Section 5.1. On the other hand, certain theoretical issues have been investigated, and new results established, usually in parallel to the tools and resources being developed, and these results are discussed below.

6.2. Computational Representation of Syntactic Formalisms

Participants: Benoit Crabbé, Claire Gardent.

LED has devised a new language for the representation of strongly lexicalised syntactic formalisms for natural language, and provided a methodology for grammatical development based on this language [9], [15]. This language was validated by developing a large sized tree adjoining grammar for French, which has been evaluated on a standard test suite (TSNLP, http://cl-www.dfki.uni-sb.de/tsnlp/).

The language we developed builds on the observation that there are two methodological axes for the representation of grammatical information: structure sharing on the one hand and alternatives on the other hand. Existing languages generally account for this distinction with an inheritance hierarchy that handles structure sharing and with lexical rules for handling alternatives. The language provided here accounts for the two axis thanks to a composition operation and a disjunctive operator representing a choice. A nice property of this language is its declarativity, setting it apart from traditional lexical rule based languages. The formalisation of such a declarative language and its implementation have revealed that is eases considerably the development of grammars for natural language.

The XMG grammar compiler implements these ideas providing a clean and modular way to describe not only syntax but also semantics, thereby supporting the production of tree grammars with semantic information. The system allows the specification of a syntax/semantics interface at a metagrammatical level so that the produced grammars contain semantic information [12].

6.3. Building Dependency Trees from Parsing Structures

Participant: Djamé Seddah.

In TAG, semantic construction usually builds upon the derivations structures of a syntactic formalism (derivation trees in the TAG case).

A possible alternative is to build a dependency graph. In [16], [17], [18], we propose the introduction of pseudo-derivation operations that enable us to add edges to the derivation trees during parsing. A possible application of this is to use selectional restrictions (that is semantic information) during parsing. Moreover, the technique we propose enables us to build dependency structures directly from the parse forest, that is,

without extracting individual derivation trees (and this is an important issue because this extraction has a possible exponential cost).

Another solution consists in using the derived structures in the same way as other unification grammars. [12] explores this avenue of research and shows how to do semantic construction on the basis of the derived tree either during or after parsing.

6.4. Linguistic Analysis and Implementation

Participant: Jesse Tseng.

Significant efforts are made for the description and formal treatment of linguistic phenomena in a variety of languages. The topics addressed include:

- The analysis of the form *de* in French (ongoing collaboration with colleagues from University Paris 7 and University Paris 4). This year, particular attention has been devoted to the syntactic and semantic licensing of "non-oblique" phrases of the form [*de* N'], found in construction with degree quantifiers and negation. The analysis was presented at the the international workshop on "Indefinites" (Brussels, January 2005) and an article has been accepted for publication in an upcoming volume of *Travaux de Linguistique*.

- Grammatical interfaces. The conception of the grammar as a collection of independent modules (semantics, syntax, morphology, phonology, etc.) is called into question by the existence of phenomena that operate at the boundary between two or more modules. A richer model of grammatical interfaces is required to express this kind of data.

A classic example of a phenomenon at the syntax-phonology interface is liaison in French. In collaboration with colleagues from the University Nancy 2 and the University Paris 4, we have developed an integrated model of liaison (publication halshs-00004844), relying on the division of a word's phonological representation into a "body" and an "appendix", where the appendix is a generalisation of the traditional notion of the "latent consonant" appearing in liaison and in derivational morphology. By adopting a model of sign-based phonological construction, we developed an integrated model that determine the presence or absence of the appendix in liaison, subject to general constraints on obligatory, optional, and impossible liaison contexts. Another set of results involves data from Polish. The alternation between the long and short forms of the prepositions *w/we* and *z/ze* is shown to be determined by phonological and morphological properties at the left edge of the noun phrase with which they combine. This goes beyond the set of grammatical features that are generally assumed to be accessible in syntactic selection and calls for a limited weakening of the principle of syntactic locality. Similar data can be found in French, with the contraction of *à* and *de* with the definite article. These results have been presented at the 2nd ACL-SIGSEM Workshop on Prepositions (Colchester, April 2005, [19]) and at the "Morphonology" session of the 36th Poznan Linguistic Meeting (April 2005).

We have studied auxiliary constructions in Slavic (joint work with Anna Kupsc). In the Slavic languages, the past tense and conditional mood are derived historically from constructions involving the auxiliary verb "to be". In the modern languages, these auxiliary forms have evolved in various ways, providing a variety of structures that require distinct but related analyses. A first analysis of Polish auxiliary constructions was presented at the HPSG Conference (Lisbon, August 2005, with published proceedings to appear). A more typological approach to a wider range of Slavic languages will be presented at FDSL05 (Potsdam, December 2005).

Research in theoretical and descriptive linguistics yields results that can be used to enrich computational models of language. LED participates in two international projects for the development of large-scale computational grammars for a variety of languages (see International level activities). The details of the two French grammars currently under development are as follows:

1. Grammatical formalism: Head-driven Phrase Structure Grammar. Computational platform: Linguistic Knowledge Builder. Under development since 2001, now associated with the Delph-In consortium. Members of LED participated in the kick-off meeting (Lisbon, August 2005) and have attended a workshop in Saarbrücken in November 2005 to bring the current grammar into conformity with the Delph-In "Grammar Matrix", which includes in particular a component for producing (and generating from) semantic representations in the framework of Minimal Recursion Semantics.

2. Grammatical formalism: Lexical Functional Grammar. Computational platform: Xerox Linguistic Environment. LED has taken over the development of a wide-coverage grammar previously developed at XRCE Grenoble, including a complete finite-state morphological analyser and a sizeable lexicon. Members of LED participate in twice-yearly meetings (Dublin, September 2004; Palo Alto, March 2005; Gotemba, Japan, September 2005) that ensure quality control and parallelism of analyses across all the grammars, and also present applications of the grammatical resources developed. Once the current grammar is brought up to date, there will be numerous possibilities for incorporating it into French and multilingual NLP systems.

6.5. Optimising surface realisation using polarity based filtering

Participants: Claire Gardent, Eric Kow.

Surface realisation (the production, on the basis of a syntactico-semantic grammar, of all the phrases verbalising a given semantic content) is known to be exponential. Moreover, polarity based filtering has been demonstrated by Guy Perrier to efficiently prune the search space when parsing. We applied polarity filtering to the surface realiser GenI (developed during the INRIA ARC GenI) and showed that it significantly improves the performance of this realiser [11]. We have also developed techniques to account for zero-literal semantic lexical items such as function words and pronouns within the polarity filter framework [14]. For the purpose of wide-scale evaluation, we have developed an tree-anchoring interface which allows the realiser to use a common grammar, lexicon, tree selection and anchoring mechanism with parsers. The grammars we use are automatically generated using the XMG metagrammar compiler.

6.6. Task Independence in Finalised Dialogue Systems

Participants: Laurent Romary, Alexandre Denis, Matthieu Quignard.

LED has been involved in the design of different dialogue systems in projects like OZONE (reference resolution in a 2D scene), MEDIA (semantic understanding), AMIGO (reference resolution in a 3D scene), etc. The main problem of such finalised dialogue systems is their dependence to the task and the lack of reusability in other tasks. Hence, LED is investigating the issue of genericity in finalised dialogue systems. Certain aspects of dialogue management are clearly domain specific (ontology, mentioned objects, actions...), while others are more general (grammar, misunderstanding management, dialogue with application...). We concentrate on both generic task-independent mechanisms and the ability of the system to be configured in a particular domain.

6.7. A shallow parsing Approach to Natural Language Understanding in Dialogue Systems

Participants: Matthieu Quignard, Alexandre Denis, Bertrand Gaiffe, Guillaume Pitel.

In the framework of the MEDIA project, LED designed a shallow parsing approach that builds on the partial results produced by a deep parser and uses heuristics to produce a single overarching parse for each sentence. The partial analyses resulting from parsing with a very restricted TAG grammar are confronted to the domain ontology to process the relevant selection restrictions and eradicate inconsistent analyses. This inferential work is completed by projection on the expected output (the annotation format). This work is the first real-scale attempt in our group to evaluate our theoretical approaches to natural language understanding

(TAG grammars, reference domains, MMIL as a convenient internal representation formalism in dialogue systems), moreover in the challenging context of spoken dialogue understanding.

6.8. Integration of Vague Predicates and Reference Domains

Participant: Guillaume Pitel.

LED designed a model that integrates reference domain theory (RDT) with the functional reference resolution model (FRRM). This model has been described in the Situated Constructional Interpretation Model (SCIM), which derives from the framework of Embodied Construction Grammar, and allows for grammatical and semantical analysis to be described together with pragmatical processing.

The model of functional reference resolution describes the behaviour of vague and/or relational predicates used in referring expressions. Basically, in this model, a referential predicate is represented by three functions: one sorts two candidates (for instance "big" would rank the candidate with the greatest size before others), the two others split the whole list of candidates into three partitions: the preferred, the possible and the impossible.

The theory of the domains of reference describes the way a referring expression recruits candidates from a representation and restructures it. The base representation for the theory is a domain of reference, composed of a differentiation criterion, a logical predicate that partitions the candidates and an optional focus.

While the model of functional reference resolution was already described in the SCIM, the RDT was not. The main effort in the integration consisted in describing the RDT in SCIM, and in replacing the differentiation criterion and the logical predicates of the RDT by the functions of the FRRM.

The model still lacks an extensive description of all reference resolution modes, and needs further checking on real data. The next natural step in this line of work is to integrate results from (or into) the DRT, which is a logical model of reference resolution that do not describe the meaning of referring predicates.

6.9. Parallel Multilingual Semantic Lexicon Construction using Statistical Approaches

Participant: Hyung Suk Ji.

The construction of large scale semantic lexica is a very expensive task. Building a multilingual lexical semantic resource is currently as expensive as building a resource for each language, as the work carried out for one language usually cannot be reused for the others. The main reason for this is that even in aligned translated corpora, the lexical units are not aligned. Our proposal is to disregard aligned sentences, and rely instead on aligned semantic spaces computed from statistical approaches for automatic corpus pre-annotation. The process is decomposed into two steps. The first step is to build, for the target language, a list of possible lemmas for a given meaning. This is done by translation of lexical units from the source semantic lexicon followed by manual checking. The second step is to find the closest meaning for a lemma, depending on its arguments in the sentence, and on the "semantic location" of the arguments found in the sentences of the source corpus for each of the potential meaning. The semantic distance between the arguments of the meaning in the source corpus and the argument in the sentence of the target corpus will be computed using three semantic spaces built using LSA: a semantic space made from corpora in the source language, another one for the target language, and the third one is a bilingual semantic space, built from parallel aligned corpora. The two monolingual spaces are mandatory since few parallel aligned corpora are available. Our approach can be considered as the search for arguments similarity instead of similarity of lexical units in parallel corpora. It theoretically means that pre-annotation can be made on a corpus with no counterpart in the source language, provided that the coverage of the parallel semantic space is good enough.

6.10. Extracting a Syntactical Lexicon for French from the LADL Tables

Participants: Claire Gardent, Ingrid Falk, Ulrike Fleury.

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

processing systems is hampered both by its non standard encoding and by a structure that is partly implicit and partly underspecified.

Together with Calligramme, we developed a method for extracting an NLP oriented syntactic lexicon from the digitised version of the GL, namely the LADL tables. In essence, this method aims at making the table structure explicit and at translating the headings into standard practice feature structure notation. Specifically, it consists in the following three steps:

1. For each table, a **SynLex-graph** is (manually) produced which represents our interpretation of the table. This graph makes the table structure explicit and translates the headings into path equations.

2. A **graph traversal algorithm** is specified: given a SynLex-graph and a table, it produces for each entry in that table, the set of subcategorisation frames associated by the table with that entry. The resulting lexicon is called a **LADL-lexicon** and closely reflects the content of the LADL table. Some of the information it contains is not currently used by most NLP tools, in particular, by parsers and surface realisers.

3. A **simplification algorithm** is specified: given a LADL-lexicon, it produces an **NLP-lexicon**. The NLP lexicon is a simplified version of the LADL-lexicon where only features relevant for parsing/generating are preserved and which only partially reflects the content of the LAD-table. It is with this lexicon that NLP is expected to proceed.

This three step method has been applied to the available subset of the first set of tables compiled by Maurice Gross which describes what Harris called *operators* namely, verbs that take a (finite or infinitival) sentential complement. For these verbs, Maurice Gross specified 19 tables covering roughly 2 500 verbs. Through the LGPL-LR license, we obtained 12 of these tables thereby covering 1 936 verbs and 2 019 verb usages.

For each of these tables, we produced a SynLex-graph and computed the corresponding LADL- and NLP-lexicons. These are now available at http://www.loria.fr/~gardent/ladl/content/resultats.php.

6.11. Topological Semantics for Hybrid Languages

Participant: Dmitry Sustretov.

The most popular semantics for modal languages nowadays is frame semantics. It has a generalisation introduced by Dana Scott in the 1970s, called neighbourhood semantics. A particular case of neighbourhood structures are topological spaces, which means that modal languages can be used for spatial reasoning. Hybrid languages, being extensions of modal languages, can be used for that purpose too, offering additional expressive power. Some natural model-theoretic questions arise. LED has investigated the issue of definability for hybrid languages in topological semantics. In a paper, we prove a theorem that provides necessary and sufficient conditions for a topological property to be definable in hybrid languages (this is a topological analogue of Goldblatt-Thomason theorem).

6.12. Resolution Methods for Hybrid Logics

Participants: Carlos Areces, Daniel Gorín.

For a number of years now members of LED have been working on resolution methods for hybrid logics. HyLoRes, a resolution based prover for hybrid logics, is a direct result of this line of research. The resolution calculus used by HyLoRes has been recently enhanced with heuristics for order and selection functions and completeness is maintained even under these constraints, which help to greatly reduce the search space of the prover. In addition, by fixing a particular order and selection function, termination for resolution in H(@) was proved. This is the first proof of termination for an implementable decision method for H(@).

6.13. Logics with Concrete Domains

Participants: Carlos Areces, Sergio Mera.

Another topic within the work of LED is on special reasoning methods. We study how to combine the standard satisfiability problem (for modal, description and first-order logic) with some flavour of model checking. The core idea is that we want to consider a class of models where each model has an "abstract part" that can vary from model to model and a "concrete part" that is fixed and common to all models (we can think, for example, that the concrete part is the natural numbers with some operations, or a concrete model of the discourse up to the moment in a natural language dialogue system). An example of such logics are Description Logics with Concrete Domains. Work on what was called half order modal logics also fits under this view. The main result obtained is a general framework which evolved as an extension of previous work on hybrid logics and that provides a unified view on hybrid logics, description logics with concrete domains and half order logics.

6.14. Inference and Semantic Construction

Participants: Carlos Areces, Patrick Blackburn.

An important line of research in LED is semantic construction, both its theoretical and practical aspects, and the role of inference in it. Recently, the group has started to investigate higher order inference and the use of higher order provers in semantic construction. This is a project developed in collaboration with the Omega group from the University of Saarbrücken, one of the strongest research group on theorem proving from higher order logics.

6.15. Document Database Querying through Heterogeneous Resources

Participant: Joseph Roumier.

An increasing amount of domain-specific knowledge is being harvested for building domain-specific ontologies and terminologies. Document management systems have much to gain from these resources, collaboratively with meta-data and annotation. Based on a three-level architecture (from top to bottom: User-Interface, Acces-Points and Document Databases) a first implementation of such a system has been built for the European project PROTEUS-ITEA, whose objective was to set a framework for distributed industrial e-maintenance.

Currently, focus is set on modelling the resources – terminology, meta-data and also documents – using one unifying framework: ontology (OWL). The output will be used to specify an unified documentation query language based both on the framework and the on data models of the resources.

6.16. Using Contextual Information for Natural Language Related Tasks

Participant: Hyungsuk Ji.

Statistical approaches are useful for several natural language related tasks such as anaphora resolution, word sense disambiguation, information retrieval, etc. Indeed, very sophisticated machine translation tasks or metaphor detecting/analysing tasks do require a reliable statistical method for a detailed semantic analysis. Explicit related-words-generation models such as ACOM (Automatic Contexonym Organization Model) developed by Hyungsuk Ji can be used for that purpose: unlike most statistical models, this model produces explicitly contextually related words (contexonyms) for a given word. An attempt was made to integrate this statistical approach to several projects in LED: using contexonyms for a word in a dialogue corpus for seeking and examining relevant key words to related topics; using double-symmetric contexonym producing method in bitext for use in machine translation tasks such as example-based machine translation, etc.

Extensions to this work are being developed, going beyond the work on a small dialogue/bitext. In particular LED is investigating the application of these ideas in very large – but not bitextual – French-English corpora (literature and newspapers) in a equivalent domain to obtain more robust results.

7. Other Grants and Activities

7.1. International level

7.1.1. ISO TC37 SC4

*Theme:*The standardisation of language resources is an essential aspect of natural language processing since it allows one both to reuse linguistic data such as lexica or grammars 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 standardisation needs in the domain of language resources. Under the responsibility of Laurent Romary (chair) and Prof. Key-Sun Choi (Secretary), the committee aims at providing ways to ensure a high level of interoperability in applications related to human language technology. Beyond the administrative 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)

Remarks: 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.

7.1.2. 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 a variety of languages (to date: Arabic, Chinese, English, French, German, Japanese, Malagasy, Norwegian, Urdu, and Welsh). These are written collaboratively within the linguistic framework of Lexical Functional Grammar (LFG) using the Xerox Linguistic Environment (XLE). The grammars adopt a common set of grammatical features to maximise parallelism in output representations.

The members of the project meet twice a year to compare new results and to discuss NLP tools and applications that use ParGram grammar resources.

LED is responsible for the development of the French grammar.

*Web site:*http://www2.parc.com/istl/groups/nltt/pargram/

*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

7.1.3. FR.FrameNet

*Theme:*Lexical Semantics (Frame Semantics, Cognitive Linguistics, Lexicography)

*Description:*The FR.FrameNet project aims at exploring methodologies, resources and techniques for the construction of a French semantic lexicon in the FrameNet family. The final purpose is to establish conditions under which a real-scale project of lexicon construction could be carried on, depending on available fundings and partnerships, with an optimal cost/coverage ratio.

The first stage of the project, has lasted three months, and ended up with the visit of a member of ICSI in Nancy. It enabled fixing the limits of our exploration around two main axes:

- semi-automatic annotation using statistical and symbolical translation resources,

- manual annotation following a mixed approach.

During the next nine months, these approaches will be evaluated in order to choose what approach should be advised for the real-scale construction.

*Administrative context:*France-Berkeley Fund

*Web site:*http://libresource.inria.fr/projects/framenet/

*Person(s) in charge:*Laurent Romary (Nancy), Charles Fillmore (Berkeley)

*Period:*start 2005-07-01 / end 2006-06-30

*Contact:*Guillaume Pitel

*Partner(s):*ICSI, ATILF, Saarbrücken University, NLT team (ICSI), Strasbourg University

7.1.4. *Delph-In: Deep Linguistic Processing with HPSG*
*Theme:*multilingual grammar development

*Description:*Delph-In is a consortium of researchers committed to open-source development of grammars, parsers, and other linguistic resources in the framework of Head-Driven Phrase Structure Grammar (HPSG). The members of the project use the Linguistic Knowledge Builder (LKB) platform and the grammars are developed in accordance with the principles of the Grammar Matrix, in particular with respect to semantic representation using Minimal Recursion Semantics (MRS).

At present, grammar development projects are underway for the following languages: English, Catalan, Danish, French, German, Modern Greek, Japanese, Korean, Norwegian, Portuguese, Spanish, Swedish. LED is responsible for the development of the French grammar.

*Period:*start 2005-08-01

*Partner(s):*DFKI, Stanford University, Saarland University, University of Washington, University of Oslo, NTNU (Trondheim, Norway), Cambridge University, NTT (Japan)

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

*Partner(s):*ATILF, INIST

7.1.6. *Passepartout: Télévision Interactive*
*Theme:*Linguistic and multimedia resources

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

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

stretch far beyond infrastructure and basic services but will also affect content, human system interaction and engineering.

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

*Administrative context:*ITEA

*Period:*start 2005-01-01 / end 2006-12-31

*Contact:*Samuel Cruz-Lara

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

7.2. European Level

7.2.1. *AMIGO: Ambient Intelligence for a Networked Home Environment*

*Theme:*Ambient Intelligence

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

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

*Administrative context:*IST European Program

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

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

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

*Contact:*Matthieu Quignard

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

7.2.2. *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. LED is involved in the first aspect by contributing to the standardisation 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).

7.2.3. *LIRICS: Linguistic Infrastructure for Interoperable Resources and Systems*

*Theme:*The LIRICS Consortium brings together experts in the field of NLP and related standards development via participation in ISO committee and National Standardisation committees. The Consortium has strong Industry support and involvement through the 21 members of the LIRICS Industry Advisory Group.

*Description:*Beyond its role as coordinator, LED is particularly involved in Lirics through several technical activities:

- co-edition of the LMF (Lexical Markup Framework) standard;
- maintenance of a central data category registry for all linguistic descriptors used within ISO committee TC 37/SC 4;
- proposal for a standard representation of reference annotation (collaboration with ATILF).

*Administrative context:*Programme e-content

*Web site:*http://lirics.loria.fr

Period: start 2005-01-01 / end 2007-06-30

*Partner(s):*DFKI, USFD, CNR-ILC, University of Vienna, UTiL, MPI, IULA-UPF, University of Surrey

7.2.4. *Proteus: Fully Integrated Platform to Support Broad E-maintenance Strategies*

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

*Description:*The objective of the Proteus project is to propose an integration platform for existent software tools in the field of industrial maintainance originally not designed to work together and, as such, not adapted to form a coherent system with disciplined information transformation and exchange. The platform deals with the following software tools: CMMS, ERP, SCADA and e-Documentation server. To make them communicate, a set of generic WebServices for Maintenance has been designed. LED was originally part of the e-Documentation Work Package. The group also proposed the design and use of a maintenance ontology compatible with OWL-DL in accordance with its original goal, to integrate a domain description as an Access Point into the e-Documentation server architecture. (This maintenance ontology is a common definition of the manipulated objects which facilitates the co-operation of the different software tools listed above.) A Terminology for maintenance normalised with TMF and a Set of Metadata extracted and adapted from the Dublin Core Document Metadata Element Set are also integrated as Access Points. The e-Documentation server architecture is built mainly to handle structured XML documents and to allow multilingual access.

*Administrative context:*ITEA

*Person(s) 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. *EPML Corpus: Equipe Projet Multi Laboratoires*

*Description:*A prototypical corpus has been collected which should be transcribed by the partners, (an 8 minute dialogue between 3 people)
*Person(s) in charge:*Anne Larcheret, Daniel Luzzati
*Period:*start 2004-01-01 / end 2005-12-31

7.3.2. *MEDIA*

*Theme:*Contextual evaluation of man-machine dialogue systems
*Description:*Within the framework of the French Technolangue project, several campaigns for evaluating different approaches to natural language processing are being carried out on various topics (e.g. parsing, natural language understanding, ...). One of these campaign, the EVALDA-MEDIA project, aims to evaluate the ability for a dialogue system to understand spoken utterances produced in a finalised dialogue context. For a given task (here, hotel booking transactions), a consortium of eight academic or industrial research laboratories has carried out the transcription of 1200 dialogs collected with the Wizard of Oz protocol. The corpus has been manually annotated and verified and two evaluation protocols have been elaborated. The context-independent evaluation consists in producing the semantic annotation of isolated utterances extracted from their dialogue history (but still interpreted within the transactional context). For the context-dependent evaluation, each utterance has to be interpreted within the dialogue context, and referential expressions have to be solved. Both evaluations do not explicitly concern pragmatic understanding, such as for example speech act recognition or dialogue structuring.

At this stage of the project, only the first evaluation has been carried out and the system designed in our team produced very interesting results (see new results).
*Administrative context:*Media is one of the components of the Evalda project in the French Technolangue program
*Period:*start 2002-12-04 / end 2005-12-04
*Contact:*Matthieu Quignard
Software: MEDIA

7.3.3. *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 at putting in common the experience of several industrial and academic partner to deliver an open source platform of interoperable lexical tools.
*Description:*LED 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 achieved:

- on the basis of a set of lexical samples (that is, words) 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 has produced a document outlining a possible model to be submitted to ISO committee TC 37/ SC 4.

The objective is to closely interact with the proponent of a new work item on lexical representation (USA) to promote the results of the project.
*Administrative context:*RNTL platform project
*Period:*start 2002 / end 2005-10-01

7.3.4. 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 as one component of the Technolangue/Normalangue project.

*Description:*LED has been particularly active 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), it has tested and adapted our own implementation to make it close to the most recent decisions; c) it has proposed a morpho-syntactic annotation for which it 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

*Period:*start 2003-01-23 / end 2005-12-04

7.3.5. Arcade II: Action de Recherche Concertée sur l'Alignement de Documents

*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 an evaluation campaign of techniques for parallel text alignment (Arcade I).

Current status: Two tasks have been defined concerning two kinds of alignments: phrase alignment and named entity alignment. The first task consists in aligning French texts with the texts in two groups of languages (one contains Indo-Europeans languages: English, German, Italian and Spanish, the other contains various more "distant" languages: Arabian, Chinese, Greek, Japanese and Russian). The second tasks consists of detecting named entities in Arabian text corresponding to annotated French named entities, in a bilingual Arabian-French corpus.

LED has confirmed its participation to the first task: alignment of French texts with texts in 10 other languages. The evaluation is underway.

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

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

*Person(s) in charge:*ELDA, DELIC

*Period:*start 2002-10-01 / end 2005-11-30

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

8. Dissemination

8.1. Service to the Scientific Community

- Patrick Blackburn: President of the SIGSEM, ACL Special Interest Group in Computational Semantics.

- Jean-Marie Pierrel

  – Member of the scientific board of the ACI "Travail" of MENRT; of the "Pôle de Recherche Scientifique et Technologique (PRST) Intelligence Logicielle".

  – Nominated member of the working group on "traitement informatique du langage auprès du Conseil Consultatif de la Langue Française".

– Member of the council of the "Ecole doctorale de Linguistique" of the University of Paris 7.

– Member of the scientific council of the ACI GRID "Grille de ressources informatiques et de données".

– Member of the steering committee of the Thematic Pluridisciplinary Network (RTP) "Communication et Dialogue" of the STIC department CNRS.

– Elected member of the Scientific Council of the UHP, the University of Nancy 1.

– Member of the selection commission of the IUT of St Dié.

– Co-opted member of the scientific and technical council of "Nancy 2005".

– Expert for the AFNOR X03A-GE1 on the redefinition of the standard ISO 1951 "Présentation/Représentation des articles dans les dictionnaires"; and for the Ministry of Research on the national prospects for computing and intensive storage for the social science.

● Claire Gardent

– Member of the ESSLLI Standing Committee, 2004–2005.

– Member of the EACL board (European Chapter of the Association for Computational Linguistics).

– Member of scientific council PI CNRS TCAN.

● Laurent Romary

– Director of IST (Information Scientifique et Technique du CNRS).

– Member of the Scientific Council of TEI (Text Encoding Initiative).

– Chairman of the ISO committee TC 37/SC 4 on Language Resource Management.

● Christine Fay-Varnier

– Elected member of the Council of studies and university life of the INPL.

– Representative to follow the social affairs of the INPL.

● Carlos Areces: Member of the Management Board of the Association of Logic, Language and Information. 2005–2006.

● Nadia Bellalem: Project leader for the registrar's office of the IUT Nancy-Charlemagne

● Nadia Bellalem, Samuel Cruz-Lara, Julien Ducret, Isabelle Kramer: Member of the "Normalangue-RNIL" group (Working group "Translation Memories"): Definition of MLIF (Multi Lingual Information Framework) (.

● Daniel Coulon, Samuel Cruz-Lara: Elected member of the Scientific Council of the University of Nancy 2

● Daniel Coulon, Christine Fay-Varnier, Jean-Marie Pierrel, Samuel Cruz-Lara: Members of the specialist commissions of the 3 Universities of Nancy

8.1.1. Management Responsibilities

- Jean-Marie Pierrel: Co-director of the UMR ATILF/University of Nancy 2.
- Claire Gardent: Member of the recruiting committee for short term posts at INRIA Lorraine/LORIA.
- Yannick Parmentier, Laurent Romary, Claire Gardent: Member of Semantic and Syntactic Processing. Department of Computer Science, University of Nancy 1.
- Samuel Cruz-Lara: Person in charge, at the national level, of the reception of Mexican students in the "Professional Licences of Computer Science".

8.1.2. Editorial and Program Committee Work

- Jean-Marie Pierrel

  – Member of the Editorial Board of Traitement Automatique des Langues; of Information-Interaction-Intelligence; of the Revue d'Intelligence Artificielle.
  – Member of the Scientific Council for "Hermès Sciences Publications".

- Matthieu Quignard: Member of the Review Committee of the Journal "STICEF" .
- Guillaume Pitel: Co-president of RECITAL2005 (French Young Researchers NLP Conference).
- Nadia Bellalem: Member of the Programme Committee of ICEIS-2005.
- Samuel Cruz-Lara: Co-Chair of the Scientific and Organization Committee of TFIT 2006 (3rd Taiwanese-French Conference on Information Technologies).
- Claire Gardent

  – Member of the Editorial Board of the Lettre du LORIA.
  – Member of the Programme Committees

    * 2nd International Joint Conference on Natural Language Processing, IJCNLP-05. Jeju Island, Korea, 11-13 October 2005. Area chair "Dialogue and Discourse";
    * Traitement Automatique de la Langue Naturelle. 2005;
    * Logical Aspects of Computational Linguistics. Bordeaux, France, 2005;
    * SEM-05. Symposium sur l'étude du Sens: Exploration et Modélisation, Biarritz, France, 2005;
    * Constraints in Discourse, Dortmund, Germany. June 3-5, 2005.
    * The 6th Tbilisi Symposium on Language, Logic and Computation. September 12 — 16, 2005. Batumi, Georgia.
    * Constraint Solving and Language Processing. Madrid, Spain, 2005.
  – Program Chair of the 9th Workshop on the Semantics and Pragmatics of Dialogue (DIALOR). Nancy, June 2005.
  – Editor of the EACL Newsletter "Bulletin du chapitre européen de l'Association pour la Linguistique Informatique" (EACL Newsletter).
  – Member of the Editorial Board of the Journal of Semantics.

- Laurent Romary

  – Member of the Editorial Boards

    * Language Resources and Evaluation, 2004 – Present.
    * ACM Transactions on Asian Language Processing, TALIP.
    * Computer and the Humanities.

  – Member of the Programme Committees

    * IJCNP 2005 Workshop and satellite event. Jeju, Korea. Co-Chair.
    * the Workshop "Linguistic engineering meets cognitive engineering in the design of multimodal interfaces" in ICMI05 October 04-06, Trento, Italy, 2005.
    * 6th International Workshop on Linguistically Interpreted Corpora LINC-2005, in conjunction with IJCNLP, Jeju, Korea, 15 October 2005.
    * Cross-Language Knowledge Induction Workshop, International Workshop, held as part of the Eurolan 2005 Summer School, Cluj-Napoca, Romania, 2005.
    * ACH/ALLC Conference 2005 June 15–18, Victoria, Canada, 2005.

- Jesse Tseng: Member of the Programme Committees

  – CSSP 2005 (Colloque de Syntaxe et Sémantique à Paris, France September 2005).
  – HPSG 2005, Lisbon, Portugal, August 2005.
  – The CL-SIGSEM Workshop on Prepositions, Colchester, UK, April 2005.

- Carlos Areces:

  – Member of the Editorial Board of the Journal of Logic, Language, and Information, 2005 – Present; and of the Journal of Applied Logics, 2004 – Present.
  – Member of the Programme Committee of 2005 International Workshop on Description Logics - DL2005 Edinburgh, United Kingdom, July 26-28, 2005; and of Methods for Modalities (M4M-4), Berlin, Germany, 2005
  – Program Chair of the ESSLLI 2006 Program Committee. Málaga, Spain.

- Patrick Blackburn:

  – Chief Editor of the Journal of Logic, Language, and Information, 2002 – Present.
  – Member of the Editorial Board of the Journal of Philosophical Logic, 2004 – Present.
  – Subject Editor (Logic and Language) for the Stanford Encyclopedia of Philosophy .
  – Member of the Programme Committee and of Methods for Modalities (M4M-4), Berlin, Germany, 2005

- Daniel Coulon: Member of the editorial committee for Intellectica.

8.1.3. Conference and Workshop Organization

- DIALOR'05 - Ninth Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)
DIALOR'05 was the ninth in a series of workshops that aims to bring together researchers working on the semantics and pragmatics of dialog in fields such as artificial intelligence, formal semantics and pragmatics, computational linguistics, philosophy, and psychology. The main topics of interest for the community participating in DIALOR are an excellent match for LED's research interests. Moreover, the meeting was a valuable opportunity to establish links for future collaboration.
The keynote speakers at DIALOR'05 were Justine Cassell (Northwestern University), Gerhard Jaeger (University of Bielefeld), and Arthur Graesser (University of Memphis).
Participants: Bertrand Gaiffe, Claire Gardent, Eric Kow, Yannick Parmentier, Laurent Romary, Matthieu Quignard.

8.1.4. Seminars and Invited Talks

- Claire Gardent: "Linguistique informatique: implantation et modélisation". Invited talk at the "Journées de l'école doctorale" "Langage, Cognition et Interaction". University of Paris 8.
- Carlos Areces: Lógicas para la Descripción. In the Escuela de Ciencias Informáticas (ECI'05), Universidad de Buenos Aires, Buenos Aires, Argentina. July 2005.
- Patrick Blackburn: Invited lecture series: Hybrid Logic, Ecole Normale supérieure, Paris, France. 10–14 May 2005.

8.2. University Teaching

- Each professor and assistant professor teaches at least 192 hours per year in the universities of Nancy. This covers the following LED members: Jean-Marie Pierrel, Daniel Coulon, Christine Fay-Varnier, Nadia Bellalem, Samuel Cruz-Lara, and Azim Roussanaly.
- Carlos Areces: University of Nancy 1 and Nancy 2, Masters in computer-science and cognitive sciences, second year. 15 hours.
- Patrick Blackburn: 2005/2006: University Nancy 1 and Nancy 2, Masters in computer-science and cognitive sciences, first year. 30h.
- Claire Gardent: 2005/2006: University Nancy 1 and Nancy 2, Masters in computer-science and cognitive sciences, second year. 30h.
- Bertrand Gaiffe: University Nancy 1 and Nancy 2, Masters in computer-science and cognitive sciences, second year. 10 hours.
- Jesse Tseng: University Nancy 1 and Nancy 2, Masters in computer-science and cognitive sciences, second year. 15 hours.

- Carlos Areces, Patrick Blackburn, Claire Gardent and Jean-Marie Pierrel, in collaboration with Guy Perrier (Calligramme), have been responsible for defining a new Masters degree track in Computational Linguistics, and linking it to other European Masters degree in computational linguistics (bilateral agreements were signed with the Universities of Amsterdam, Bolzano, Copenhagen, Malta, Prague, Roskilde, Saarbrücken, and Utrecht). This is an important development. Prior to 2005, it was only possible to take two courses in computational linguistics at Nancy universities, and both courses were fifth year courses. The new Masters track proposes two courses in computational linguistics at the fourth year level, and that the entire fifth year can now be devoted to computational linguistics. The bilateral agreements will make it possible to exchange students with other European programs in computational linguistics. It is hoped that the masters degree will increase number of students doing PhD studies in computational linguistics in Nancy.

9. Bibliography

Books and Monographs

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

Doctoral dissertations and Habilitation theses

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

Articles in refereed journals and book chapters

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

[4] O. BONAMI, G. BOYÉ, J. TSENG. *Sur la grammaire des consonnes latentes*, in "Langages", nᵒ 158, 2005, p. 89–100.

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

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

Publications in Conferences and Workshops

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

[8] S. CRUZ-LARA, S. GUPTA, F. GARCIA, J. DAVID, L. ROMARY. *Multilingual Information Framework for Handling Textual Data in Digital Media*, in "Proc. of the IEEE International Conference on Active Media Technology - AMT 2005, Takamatsu, Japan", May 2005.

[9] D. DUCHIER, J. LE ROUX, Y. PARMENTIER. *XMG: Un Compilateur de Méta-Grammaires Extensible*, in "12e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles - TALN 2005, Dourdan, France", Jun 2005.

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

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

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

[13] S. GUPTA, S. CRUZ-LARA, L. ROMARY. *Implementing Multilingual Information Framework in Applications using Textual Display*, in "Proc. of the 7th International Conference on Enterprise Information Systems - ICEIS 2005, Miami, USA", May 2005.

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

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

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

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

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

[19] J. TSENG. *Prepositions and complement selection*, in "Proc. of the 2nd ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications, Colchester, England, UK", Apr 2005, p. 11-19.

Internal Reports

[20] S. CRUZ-LARA. *Handling multilingual content in digital media: a critical analysis*, Technical report, Loria, Jan 2005.

[21] S. CRUZ-LARA, S. GUPTA, J. FERNENDEZ-GARCIA, L. ROMARY, P. SÉBIRE, F. PRATEUX, T. ZAHARIA, M. PREDA, S. TRAN. *Handling Multilingual Content in Digital Media: A Critical Analysis*, INRIA Technical Report (ITEA Jules Verne Project), INRIA, May 2005.