INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team LeD

Langue et Dialogue

Lorraine

THEME SYM

Table of contents

1. Team

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

Head of project-team
Patrick Blackburn [research director INRIA, HdR]

Vice-head of project team
Claire Gardent [research scientist SHS department CNRS, HdR]

Administrative assistant
Isabelle Blanchard [Project Assistant]

Research scientist
Carlos Areces [research scientist INRIA]
Bertrand Gaiffe [research scientist INRIA until 2006-12-01]
Matthieu Quignard [research scientist SHS department CNRS]
Laurent Romary [research director INRIA until 2006-08-15, HdR]
Jesse Tseng [research scientist SHS department CNRS]

Faculty member
Nadia Bellalem [assistant professor IUT Nancy-Charlemagne University of Nancy 2]
Daniel Coulon [professor University of Nancy 2, HdR]
Samuel Cruz-Lara [assistant professor IUT Nancy-Charlemagne University of Nancy 2]
Christine Fay-Varnier [assistant professor École de Géologie INPL]
Jean-Marie Pierrel [professor UFR STMIA UHP Nancy 1, HdR]
Azim Roussanaly [assistant professor UFR MI University of Nancy 2]

Ph. D. student
Alexandre Denis [scholarship INRIA, graduating in Nov/2007]
Sébastien Hinderer [UHP Nancy 1, graduating in Dec/2006]
Eric Kow [UHP Nancy 1, INRIA grant, graduating in Dec/2007]
Cheng Yu Lu [INRIA (with Taiwan), graduating in Jul/2008]
Chia-Hung Lin [INRIA (with Taiwan), graduating in Jul/2008]
Thi-Minh-Huyen Nguyen [UFR STMIA UHP Nancy 1, graduated on 01/Nov/2006]
Yannick Parmentier [UFR STMIA UHP Nancy 1 (INRIA grant - UHP Nancy 1 Monitor), graduating in March/2007]
Joseph Roumier [ENSEM INPL (CIFRE grant financed par CEGELEC), graduating in Nov/2007]
Dmitry Sustretov [UHP Nancy 1, INRIA grant, graduating in Oct/2008]
Daniel Gorín [UHP Nancy 1, Co-direction with the Universidad de Buenos Aires, graduating in Dec/2008]
Sergio Mera [UHP Nancy 1, Co-direction with the Universidad de Buenos Aires, graduating in Dec/2008]
Marilisa Amoia [UHP Nancy 1, Co-direction with the University of the Saarland, graduating in Dec/2007]
Phuong Le Hong [Nancy 2, Co-direction with the University of Hanoi, graduating in Oct/2009]
Luciana Benotti [UHP Nancy 1, graduating in Dec/2009]
Paul Bedaride [UHP Nancy, graduating in Oct/2009]

Post-doctoral fellow
Anna Kupsc [CNRS until 2006-08-31]
Guillaume Pitel [from 2005-01-01 until 2007-3-31]

Technical staff
Julien Ducret [INRIA until 2006-10-31]
Ingrid Falk [CNRS]
Laurence Kbida [INRIA until 2006-03-31]
Isabelle Kramer [INRIA until 2006-10-31]

Gil Francopoulo [until 2006-02-28]
Guillaume Rémi
Karen Fort [INRIA, LED & Calligramme, HdR]

Student intern

Mehdi Ben Abderrahmen [(University of Sfax) from 2006-08-15 until 2006-09-18]
Mohammed Ghribi [(University of Sfax) from 2006-08-31 until 2006-09-14]
Nourredine Loukil [(Sfax, Tunisia) from 2006-08-15 until 2006-09-19]
Jorge Pelizzoni [(Brasil) until 2006-08-31]
Diego Figueira [(Argentina) from 2006-06-07 until 2006-08-31]
Laura Infante Blanco [(Valladolid, Spain), from 2006-10-09]
Martin Delemotte [(IUT Charlemagne), from 2006-06-12 until 2006-08-25]
Sarah Maarek [(University of Paris VII), from 2006-03-06 until 2006-08-15]
Julien Vannesson [(UHP ESIAL), from 2006-06-19 until 2006-07-31]
Mohamed Jmaiel [(University of Sfax), from 2006-06-22 until 2006-06-27]
Keis Haddar [(University of Sfax), from 2006-08-31 until 2006-09-14]
Bilel Gargouri [(University of Sfax), from 2006-08-31 until 2006-09-14]

2. Overall Objectives

2.1. Overall Objectives

The aim of the LED team is 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 personnel changes in the team, its role has become increasingly dominant during last years. Two recent changes have further accentuated this trend. First, in September 2006, Laurent Romary left the team to work in Berlin, and in December 2006 Bertrand Gaiffe left to work at ATILF (the center for "Analyse et Traitement Informatique de la Langue Fran caise," a large research center in Nancy, with which LED has good contacts, devoted to the development of lexical resources).

3. Scientific Foundations

3.1. Introduction

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

3.1.1. Computational Linguistics, Computational Logic, Knowledge Representation:

The central research theme of LED is computational semantics (where "semantics" is broadly construed to cover various pragmatic, and discourse and dialog level phenomena). And research within LED is particularly focused on investigating the interplay between representation and inference. Another way of putting this idea 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; as we will soon discuss, LED is actively involved in carrying out the required background work. 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...

In the next sections we'll discuss in detail some examples of these ideas.

3.2. 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.

3.3. Linguistic Resources

In an ideal world, computational semanticists would not have to worry overly much about linguistic resources. Large scale lexica, treebanks, and wide coverage grammars (supported by fast parsers and offering a flexible syntax semantics interface) would be freely available and easy to combine and use. The semanticist could then focus on 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 Tree Adjoining Grammars, TAG, (to enable the construction of a wide coverage grammar for French); and, more recently, it is involved in the development of a Lexical Functional Grammar, LFG, and a Head-Driven Phrase Structure Grammar, HPSG, (both offering a syntax/semantics interface) for French.

3.4. Logic Engineering

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

For a start, most of the efforts of the first-order automated reasoning community have been devoted to theorem proving; model building, which is also a 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 exist do not directly offer the kind of expressivity required for natural language applications. For example, it is hard to encode temporal inference problems in

implemented versions of description logics. Fourth, for very strong logics (notably higher-order logics) few implementations exist and their performance is currently inadequate.

These problems are not insurmountable, and 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 automated tools for hybrid logics (which can be thought of as more expressive versions of description logics) used and developed by LED 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.

3.5. Empirical Studies

The role of empirical methods (model learning, data extraction from corpora, evaluation) has greatly increased in importance in both linguistics and computer science over the last fifteen years. 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. Application Domains

As we have said, much of the work of the LED team revolves around the development of wide coverage grammars that yield semantic representations, and linking such grammars with inference tools. That is, most of our current work is building both the theory and the resources and software needed to support detailed work on semantics. Until such resources are developed, it is somewhat premature to speak of the application domains — though (judging by international experience for better explored languages such as English) it is likely that our current foundational work will lead to applications in such domains as question answering, intelligent document retrieval, and so on.

In the meantime, LED is exploring the possibilities of using lighter methods (that is, methods making less use of linguistic knowledge and inference) for applications. Application domains we are currently exploring or developing include:

- Handling multilingual information in multimedia settings.
- Extraction of timelines and textual summaries from online encyclopedias.
- Applications of natural language processing to language education.
- Application of natural language processing to computing for the humanities.
- Applications of natural language processing in the publishing industry.

The applications in multilinguality are well established (notably because of the work of LED member Samuel Cruz-Lara in the setting of the Passepartout and MLIF projects). The other four topics named are newer, and we are currently looking for industrial partners to collaborate with. It should be remarked that in previous years, LED was heavily involved with industry because of its focus on standardisation. However, with the departure of Laurent Romary, standardisation is no longer a core activity, hence our interest in finding new partners and new applications. We are currently discussing with Conitiv Synchron (a Nancy-based firm specialising in the development of software for accelerated language learning; see http://www.conitiv-synchron.com/) the possibility of using natural language processing techniques in this domain. A Masters student (supervised by Patrick Blackburn) will start working with Conitiv Synchron on this topic early in 2007.

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 MetaGrammar. 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. [57] to the newly developed XMG system of Crabbé et al. [55].

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 [56], with Éric de la Clergerie's DyALog parser [72] and with the GenI generator [61]. 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/her own target formalism.

Developed under the supervision of Denys Duchier, the XMG compiler is the result of an intensive collaboration with the INRIA team 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: 1.1.6
License: CeCILL
Web site: http://sourcesup.cru.fr/xmg/
Documentation: http://sourcesup.cru.fr/xmg/#Documentation
Project(s): XMG
Authors: Benoît Crabbé, Denys Duchier, Joseph Le Roux, Yannick Parmentier
Contact: Benoît Crabbé, Claire Gardent, 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 it uses 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 (developed by Guy Perrier of 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 (developed by Crabbé, previous LED member; Duchier, previous Calligramme member; Leroux, Calligramme; and Parmentier, LED) and covers the basic syntactic structures of French as described in Anne Abeillé's book "Une grammaire électronique du français" [50].

GenI now has support for multiple generation algorithms and has recently been tuned for better performance and memory consumption.

Also, we are starting work on interfacing GenI with the RACER description logic theorem prover to improve the quality and flexibility of lexicalization using a lexical ontology [60].
Version: 0.10
License: GPL
Web site: http://trac.loria.fr/~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 [51], [52] 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 sub-language 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 functions to prune the search on the space of possible clauses.

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.

We have recently extended the prover to handle also inverse operators and transitive relations. Our short term goal is to be able to deal with hybrid inference on models for linear time, so that HyLoRes can be used as inference module in other LED's projects which are involved with the extraction and consistency checking of temporal information.

For comparison, and to better test HyLoRes, we have developed an optimized satisfiability preserving translation from H(@,↓) into first-order logics [37]. Such a translation enables us to compare HyLoRes with state of the art resolution-based theorem provers.

HyLoRes is implemented in Haskell (ca. 3500 lines of code), and compiled with the Glasgow Haskell compiler http://www.haskell.org/ghc.
Version: 2.1
License: GPL
Web site: http://www.loria.fr/~areces/HyLoRes
Documentation: http://www.loria.fr/~areces/HyLoRes
Authors: Carlos Areces, Daniel Gorín, and Juan Heguiabehere (from the University of Buenos Aires)
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 Sebastiani, 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
Authors: Carlos Areces, and Juan Heguiabehere (from the University of Buenos Aires)
Contact: Carlos Areces

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

Maurice Gross' grammar lexicon [64] contains extremely rich and exhaustive information about the morphosyntactic and semantic properties of French syntactic functors (verbs, adjectives, nouns). Yet its use within natural language processing systems is still restricted.

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

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

Two sets of subcategorisation lexicons (called LADL-SynLex and NLP-SynLex) were extracted from the LADL tables (the LADL tables are a digitized version of Gross' grammar lexicon). The current SynLex contains the LADL- and NLP-SynLex lexicons for all the available LADL-tables, which amounts to roughly 22,462 entries and 5,548 verbs. 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 [59], [58]. 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.
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. MLIF

Linguistic information plays an essential role in the management of multimedia information, as it bears most of the descriptive content associated with more visual information. Depending on the context, it may be seen as the primary content (text illustrated by pictures or videos), as documentary content for multimedia information, or as one among several possible information components in specific contexts such as interactive multimedia applications. Linguistic information can appear in various formats: spoken data in an audio or video sequence, implicit data appearing on an image (caption, tags, etc.) or textual information that may be further presented to the user graphically or via a text to speech processor.

In this context, dealing with multilingual information is crucial to adapting the content to specific user targets. It requires one to consider potential situations where the linguistic information contained in a multimedia sequence is either already conceived in such a way that it can be adapted on the fly to the linguistic needs of the user, or by using an additional process where content should be adapted before presentation.

Finally, there are a wide variety of applications within which multilingual information may appear. MLIF supports development and implementation of a generic framework for dealing with multilingual content: subtitling of video content, dialogue prompts, menus in interactive TV, descriptive information for multimedia scenes, karaoke management, etc. Such information should be considered in the light of the experience of more specialised communities traditionally dealing with multilingual content, namely the translation and localisation industry.
Version: 0.1
Web site: http://mlif.loria.fr
Documentation: http://mlif.loria.fr
Project(s): MLIF
Authors: Guillaume Rémy, Julien Ducret, Samuel Cruz-Lara
Contact: Samuel Cruz-Lara

5.7. MEDIA

In the framework of the MEDIA project, software has been developed to process transcriptions of a spoken dialogue corpus and to provide a semantic representation of their task-related content. This software contains a tokeniser, a TAG parser (LLP2), a TAG grammar, an OWL ontology and a set of rules in description logic that can be used with an ontology 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 (its original application domain).
Version: 0.4
License: GPL
Project(s): MEDIA
Authors: Alexandre Denis
Contact: Alexandre Denis

5.8. CURT (Clever Use of Reasoning Tools)

The CURT (Clever Use of Reasoning Tools) family [54] 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 number of versions of CURT (covering French, Romanian and Spanish) have been also developed.

More recently, a Polish version of CURT has been developed. This improves on previous versions in a number of ways. First, the Prolog-based beta-reduction module has been replaced by the use of Nessie, which permits proper typing to lambda terms. Secondly, for the first time CURT has been extended to cover representation and inference for temporal phenomena in natural language. In fact, this new version updates CURT so thoroughly, that it is likely to eventually displace the old system entirely.
Version: 1.0
License: GPL
Web site: http://www.comsem.org
Documentation: http://www.comsem.org
Authors: Carlos Areces, Patrick Blackburn, Johan Bos (University of Bologna), Sébastien Hinderer

Contact: Carlos Areces, Patrick Blackburn, Sébastien Hinderer.

5.9. Nessie

Nessie is a semantic construction tool. Given a lexicon associating terms to lemmas and a text (represented by a syntax tree having lemmas as its leaves), Nessie computes a TY_n lambda term representing the semantics of the input.

Web site: http://trac.loria.fr/~nessie/
Documentation: http://trac.loria.fr/~nessie/
Authors: Sébastien Hinderer
Contact: Patrick Blackburn, Sébastien Hinderer

5.10. 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 was made available on the web in June 2006 through the CNRTL (Centre National de Ressources Textuelles et Lexicales) [62].
Project(s): FReeBank, ANANAS
Authors: Claire Gardent
Contact: Claire Gardent

5.11. Semantic Constructor

The Semantic Constructor computes a flat semantic representation of a sentence using a TAG grammar equipped with a syntax / semantics interface (e.g., a TAG compiled with XMG).

It uses the DyALog system developed by Éric De La Clergerie at INRIA Rocquencourt to perform syntactic parsing. The derivation forest output by DyALog is then used to compute the semantic representation using the technique given in [63].

This software is being used to perform an evaluation of the SemFRAG (semantic French TAG grammar) of [33] in terms of semantic coverage of the Test Suite for NLP.
Version: 1.2
License: CeCILL
Web site: http://trac.loria.fr/~semconst
Documentation: http://wiki.loria.fr/wiki/SemConst/Documentation
Project(s): GenI
Authors: Claire Gardent, Yannick Parmentier
Contact: Claire Gardent, Yannick Parmentier

5.12. SemFRaG

SemFRaG is a TAG grammar developed with the XMG metagrammar compiler and which describes both the syntax and the semantics of natural language expressions. The grammar covers the TSNLP (“Test Suites for Natural language Processing” [53]) testsuite both syntactically and semantically. It is used both for parsing and for generation.
Authors: Claire Gardent, Benoit Crabbé
Contact: Claire Gardent

5.13. FRoG — French Resource Grammar

FRoG is an HPSG grammar of French developed with the LKB platform (Lexical Knowledge Builder, http://wiki.delph-in.net/moin/LkbTop). 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
Project(s): Delph-In
Authors: Jesse Tseng
Contact: Jesse Tseng

5.14. ParGram French Grammar

ParGram is an LFG grammar of French developed with the XLE platform (Xerox Linguistics Environment, http://www2.parc.com/isl/groups/nltt/) for parsing and generation. A large-coverage, robust implementation integrating a sizable 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)
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

5.15. LLP2

LLP2 is an LTAG (Lexicalized Tree Adjoining Grammars) parser based on the bottom up algorithm described in Patrice Lopez's thesis [67]. The present version is restricted to TIG (Tree Insertion Grammar), a formalism related to Tree Adjoining Grammars where the adjunction rule is sufficiently restricted to ensure that TIGs only derive context free languages. 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
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

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, resources and programs directly associated with natural language processing have been developed, and on the other hand, certain theoretical and computational results concerning inference in the family of languages known as hybrid logics have been obtained (hybrid logics are languages with weaker expressive power than first-order logic, and related to the description logics used in Artificial Intelligence, that LED members want to apply to problems in natural language semantics).

Given the diversity of these topics, we think it would be misleading to point to just *one* achievement by LED in 2006. Rather, the interdisciplinary nature of the work conducted in LED is better illustrated by drawing draw special attention to *two* of the topics discussed below. First, on the linguistic side, we note the work (led by Claire Gardent) on the conversion of Maurice Gross's LADL tables (which give information about the morphosyntactic and semantic properties of French verb) into a computational format [34], [33]. This is a significant addition to the computational resources available to the francophone computational linguistics community. Second, on the inferential side, we draw attention to the work (led by Carlos Areces and Patrick Blackburn) that has been conducted on inference in hybrid logic. In 2006, a number of outstanding issues concerning inference for hybrid logic were resolved: general results on axiomatics (and the limits of axiomatics) for hybrid logic were published [17], [43], termination results for a number of hybrid languages of varying strength were obtained (this work has been submitted for journal publication), and the resolution method for hybrid logic was further refined [37].

Moreover, we would also like to mention that two research themes are emerging which should draw further together the linguistic- and logic-based work at LED, namely work on textual entailment [48] and dialogue [23]. Two PhD students (Paul Bedaride and Luciana Benotti) whose work bears on these topics began working at LED towards the end of 2006, so the following years should see advances in these areas.

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

Participants: Claire Gardent, Ingrid Falk, Guy Perrier, Bruno Guillaume.

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 which 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 Gross' grammar lexicon, 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. A SynLex-graph and a LADL table are the input of the next step.
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 obtained in this way is superfluous for most current NLP tools, in particular, by parsers and surface realisers. Hence, a third step is required.
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 LADL-table. It is with this lexicon that NLP is expected to proceed.

6.3. Constraint-based Linguistic Analysis and Implementation

Participant: Jesse Tseng.

Major efforts are devoted to the precise description and formal analysis of linguistic phenomena from various languages in constraint-based grammatical frameworks, particularly Head-Driven Phrase Structure Grammar, HPSG. Theoretical results for the analysis of the French prepositional forms *à* and *de* are presented in [13]. An HPSG analysis of prepositional passives (or "pseudopassive" constructions) in English is proposed in [47]. Results from an ongoing comparative study of compound verb constructions across the Slavic languages are reported in [46].

Insights from theoretical work on French are incorporated into La Grenouille, the HPSG-style French resource grammar implemented using the LKB grammar and lexicon development platform. The latest versions of this grammar were presented at meetings of the international consortium DELPH-IN in Fefor, Norway, and at the Workshop on Large-scale Grammar Development and Grammar Engineering in Haifa, Israel (both in June 2006). Major developments for this year included the incorporation of richer phonemic representations and mechanisms for handling phonological information, with a view to eventual interaction with speech-based resources and applications. This work, like Claire Gardent's work on the LADL tables, is intended to increase the number of resources available for creating large-scale grammars for French.

6.4. 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. However, 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 [61]. We have also developed techniques to account for zero-literal semantic lexical items such as function words and pronouns within the polarity filter framework [66].

For the purpose of wide-scale evaluation, we have developed a tree-anchoring interface which allows the realiser to use a common grammar, lexicon, tree selection and anchoring mechanism with parsers, and testing is well under way. The grammars we use are automatically generated using the XMG metagrammar compiler.

This work was presented at the TAG+8 workshop [39], [35] and the Haskell Workshop [38].

6.5. Reference Resolution in Dialogue Systems

Participants: Laurent Romary, Alexandre Denis, Matthieu Quignard.

Reference resolution is a critical ability in a dialogue system because the natural behaviour of a speaker is to use short expressions designating the objects instead of full descriptions. The difficulty lies in the diversity of referring expressions (pronominals, definites, demonstratives, indefinites...) and the different modalities used to refer: gestural or linguistic. The Reference Domain model [68], integrates the range of referring expressions in a multimodal framework, unlike the discourse representation theory, which focuses on a monomodal representation [65].

However, the Reference Domain model had to be extended in order to manage plural referents and groupings. The problem was that the theory did not include a high level reference domain, acting as a container, for plural referents or newly created groupings. We added such upper-level domain and modelled the resolution of particular expressions like ordinals or alterities within the extended model. This work has been validated in the MEDIA understanding evaluation campaign, which gave a chance to test the model on a large spoken corpus (more than 1,200 dialogues). Indeed, the model was applied to several complex cases involving plural anaphoric or generic expressions which were hard to resolve, but proved to be efficient on the more usual expressions.

6.6. Towards a FrameNet for French

Participant: Guillaume Pitel.

Following the collaboration with the Berkeley's FrameNet team, we have achieved some important tasks toward the definition of a large scale effort for building a FrameNet for French. Thus, like Claire Gardent's work on the LADL tables, the work reported here is another component of LED's effort to create lexical semantic resources for French.

- An analysis of the semantic cohesion of Frame Elements contents has been conducted ("semantic" here is based on co-occurrence information, as in Latent Semantic Analysis). Results can be found at http://guillaume.pitel.free.fr/Frames.en.3/index.html, showing how an important part of the Frame Element annotated contents are actually in limited semantic areas, allowing theoretically for a bilingual projection of Frame Elements (that is, determining the Frame Element of a sentence segment in a language lacking manual FrameNet annotation).
- An annotation in French of 1076 sentences extracted from the Europarl corpus (corresponding to a subcorpus already annotated for English and German by Sebastian Pado and Katrin Erk)
- An evaluation of the expected necessary time for production of an assisted manually-produced list of lexical units for French, using translations from the Semantic Atlas Project (Sabine Ploux and Hyung-suk Ji) and translations from the online WordReference tool.
- A set of monolingual and bilingual LSA spaces built using the Infomap tool, using corpora from the Europarl set as well as the BNC, Frantext, and several texts from the Gutenberg project.
- A clustering tool that can be integrated in the Infomap system, used to build more precise subsets of closely related terms in the Frames and Frame Elements annotation contents.
- A set of tools for automatic Frame and Frame Element attribution based on those tools and the data they produce. Results seem to be good enough for the system to be used as an annotation assistant. Level 1 (first proposal) results for correct attribution are 54.5% for Frames and 63% for Frame Elements. Level 4 (one of the first 4 proposals) are 72% and 92%. Baseline for the Frame annotation is close to zero in our resource-free model since there are 500+ frames that can be attributed, baseline for the Frame Element is much higher : 41%, since a Frame has 7.65 Frame Elements on average.

A chapter in a FrameNet anniversary book edited by Hans Boas is currently being written.

6.7. Mutual Understanding in Dialogue as a Collaborative Process

Participants: Luciana Benotti, Patrick Blackburn, Daniel Coulon, Alexandre Denis, Bertrand Gaiffe, Guillaume Pitel, Matthieu Quignard.

In common human conversations or human-machine interactions, misunderstandings are often the main causes for dialogues to fail. A reason for that is probably that participants actually interpret their interlocutor's utterance from their own point of view without being able to appreciate whether their points of view diverge from each other. To prevent such phenomena to arise, one should consider that mutual understanding is a collaborative process achieved across the dialogue together with the goal of the dialogue itself. This theory has been elaborated by Herbert H. Clark and colleagues, and is called *grounding*. This analytic method does not focus on the way a task is achieved within a dialogue (how each utterance contributes to the common goal) but on how utterances give evidence that some previous utterance has been mutually understood.

Within the two last decades, the model of grounding has been applied, tested, formalised and sometimes implemented. Our working group aims to read those works in details and confront each other to discover their limits and how they can be improved.

6.8. Neighbourhood and Topological Semantics for Hybrid Languages

Participants: Carlos Areces, Patrick Blackburn, Diego Figueira, Dmitry Sustretov.

The most popular semantics for modal languages nowadays is relational 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.

Recently, members in LED had started a semantic exploration of neighbourhood semantics. Two different semantics proposals have been introduced in the literature and we have found characterizations, via appropriate notions of bisimulations, for them. We also proposed extensions that lead to more natural languages. Finally, we have also investigated the computational complexity of these logics, extending the work of Vardi et al. [71]. This work will soon be submitted for publication.

LED has also investigated the issue of definability for hybrid languages in topological semantics. In a paper [69], 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).

More recently an investigation has been done on the computational properties of logics of different classes of topological spaces. It has been proven that basic modal logics of separation axioms T_0, T_1 and T_2 coincide and are decidable. The hybrid logic of T_1 has also been proven decidable.

6.9. Inference Methods for Hybrid Logics

Participants: Carlos Areces, Patrick Blackburn, Thomas Bolander [Technical University of Denmark], Daniel Gorín, Balder ten Cate [University of Amsterdam].

For a number of years now, members of LED have been working on inference methods for hybrid logics, both theoretically and developing applications.

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, that 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(@).

We have also explored efficient, satisfaction preserving translations into first-order logic. Such translations allow the use of modern first-order theorem provers to check for satisfiability of hybrid logic formulas. Intensive empirical testing shows this to be a viable option. The nature of the formulas obtained under our optimised translation of the hybrid input gives rise to a 'layering' effect that helps the first-order prover to find a solution [37]

In addition to resolution methods, LED members have recently obtained results in more classical approaches to inference like axiomatics and tableaux methods. A detailed investigation of axiomatic systems for hybrid languages have recently appeared in Studia Logica [17], and an article on termination techniques for hybrid logic tableaux is under way.

6.10. 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 [43].

6.11. Inference, Discourse and Dialogue

Participants: Carlos Areces, Luciana Benotti, Patrick Blackburn, Sébastien Hinderer.

An important line of research in LED is natural language semantic and in particular in Dialogue Systems. We are interested in both its theoretical and practical aspects, and the role that different kinds of inferences play in it. On one hand, 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 groups on theorem proving from higher order logics. The Calligramme team is also starting to be interested in these issues.

More recently, a Polish version of the dialogue system CURT has been developed. The new version improves on previous versions in a number of ways, but the most important is that it covers representation and inference for temporal phenomena in natural language, and further extensions to be able to handle also aspect are being developed.

We have also started experimenting on the use of non-standard inference techniques (e.g., hybrids of inference and planning) in the framework of Dialogue Systems, in particular to help handling tacit actions [23]. When dialogue systems are developed as interfaces for applications, users often left implicit certain actions (which we call, tacit actions) when asking the system to perform a particular action for them. The system has to infer these tacit actions which are preconditions for the execution of the action requested by the user. The team is investigating the use of state of the art planners in this kind of scenarios. This research is related to recent work of Matthew Stone on presupposition and other pragmatic phenomena [70].

6.12. Textual Entailment

Participants: Carlos Areces, Paul Bedaride, Claire Gardent.

One important characteristic of natural language is the huge flexibility on the ways in which we can express the same information. Many NLP applications like Question/Answering, information retrieval, and so on, need to deal with this diversity efficiently and accurately. Recognizing textual entailment, i.e., the question of whether the information contained in a given text T_1 can be inferred from the information provided by another text T_2, is the core inference task for such systems.

Textual entailment recognition is a very difficult task, and systems presented at conferences like the RTE Challenge (http://www.pascal-network.org/Challenges/RTE2) still perform very poorly.

LED has recently started investigating this issue, and Paul Bedaride's Master Thesis [48] investigates how lexical information encoded as a description logic ontology can be used in this task.

6.13. 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, Access-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.

With the aim of specifying an unified documentation query language based both on the described framework and on data models of the resources, the focus has been set on modeling these resources – terminology, meta-data and documents – using OWL-DL. Immediate results of this approach include automated classification and validation of models using an inference engine.

7. Other Grants and Activities

7.1. International level

7.1.1. ISO TC37 SC4

Theme: Standardisation of language resources

Description: 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 components in complex processing chains (e.g., a man-machine dialogue system).

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 Organisation for Standardisation)
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, RNTL/Outilex, and (since August 2006) the MLIF projects.

7.1.2. FR.FrameNet

Theme: Lexical Semantics (Frame Semantics, Cognitive Linguistics, Lexicography)
Description: The FR.FrameNet project aimed at exploring methodologies, resources and techniques for the construction of a French semantic lexicon in the FrameNet family. The final purpose was to establish conditions under which a real-scale project of lexicon construction could be carried on, depending on available funding and partnerships, with an optimal cost/coverage ratio. More detail on this project (which finished at the end of 2006) can be found elsewhere in this document.
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

7.1.3. 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.
Administrative context: Open international collaboration network
Web site: http://www.delph-in.net
Person(s) in charge: Hans Uszkoreit (DFKI) and Dan Flickinger (Stanford)
Period: start 2005-08-01
Contact: Jesse Tseng
Partner(s): DFKI, Stanford University, Saarland University, University of Washington, University of Oslo, NTNU (Trondheim, Norway), Cambridge University, NTT (Japan)

7.1.4. 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.
Administrative context: Text Encoding Initiative
Web site: http://www.tei-c.org
Period: start 2005-08-01
Contact: Laurent Romary
Partner(s): ATILF, INIST

7.1.5. Passepartout: Interactive Television

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-centres 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-centres 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
Web site: http://www.citi.tudor.lu/passepartout
Period: start 2005-01-01 / end 2007-03-31
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 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 provide 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.amigo-project.org
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. LIRICS: Linguistic Infrastructure for Interoperable Resources and Systems

Theme: Standardization of linguistic resources
Description: 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.

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:*European eContent Program, Project No.22236
*Web site:*http://lirics.loria.fr
Contact: Laurent Romary
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

7.3. National level

7.3.1. Acquiring and validating subcategorisation lexicons for French and Polish

Description: The aim of the project is to develop and evaluate subcategorisation lexicons for French and Polish and to do this by combining the complementary strengths of both partners. Indeed while the french side has extensive expertise in the design, formatting and normalisation of linguistic resources, the polish side is already engaged in the exploration of statistical and corpus based techniques for the acquisition of subcategorisation information. The project will aim at producing, and then evaluating, the lexicon using a blend of symbolic and statistical methods. While the french side will provide specifications concerning the content and format of the lexicons, the polish side will provide expertise on the statistical acquisition of subcategorisation information from corpora.

Administrative context: Projet PAI Polonium
Period: start 2006-01-01 / end 2007-31-12
Web site:
Partner(s): LED/LORIA and Institute of Computer Science PAS, Warszawa, Poland
Contact: Claire Gardent

7.3.2. Lexsynt

Description: The aim of the project is to develop a syntactic and semantic lexicon for french that is open source and adapted for Natural Language Processing.
Administrative context: Projet ILF (Institut de la Langue Francaise)
Period: start 2005-01-01 / end 2008-31-12
Web site: http://lexsynt.inria.fr
Partner(s): LED/LORIA, Calligramme/LORIA, Atoll/INRIA Rocquencourt, Laboratoire de Linguistique Formelle/Paris 7, Laboratoire Parole et Langage/U. de Provence, Signes/INRIA Futurs, ATILF/UMR 7118, ERSS/UMR 5610, IGM/UMR 8049, LPL/UMR 6057, Lattice/UMR 8094, Modyco/UMR 7114, ATV/K. U. Leuven, OLST/Université de Montréal
Contact: Claire Gardent

7.3.3. MEDIA

Description: Within the framework of the French Technolangue project, several campaigns for evaluating different approaches to natural language processing were carried out on various topics (for example, parsing and natural language understanding). One of these campaigns, the EVALDA-MEDIA project, aimed to evaluate the ability of a dialogue system to understand spoken utterances produced in a real dialogue context. For a given task (here, hotel-booking transactions), a consortium of eight academic or industrial research laboratories carried out the transcriptions of 1200 dialogs collected using 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 with the dialogue context, and referential expressions have to be resolved. Neither evaluation explicitly concerns pragmatic understanding, such as for example speech act recognition or dialogue structuring.

For these evaluations, the LED team developed a Natural Language Interpretation system using a fully symbolic approach. Utterances are first parsed with a flexible TAG grammar, then interpreted against the task ontology. Finally, the internal semantic representation (a graph in the Multimodal Interface Language) is converted to a MEDIA semantic formalism. The output is compared with the manual annotation.

Our participation in the MEDIA campaign enabled us to validate the relevance of our fully symbolic approach for the design of robust dialog systems.
Administrative context: Programme Technolangues, Campagne EVALDA
Period: start 2002-12-04 / end 2006-12-04
Contact: Matthieu Quignard

7.3.4. Mosaique

Description: The project aims to develop a high level formalism for describing the syntax and semantics of natural language; to provide compilers for this formalism that supports the compilation of the high level formalism into several operational ones as needed for the various existing theories of syntax; and to validate the grammars thus obtained on hand built and real text testsuites.
Administrative context: Action de Recherche Concerté INRIA
Period: start 2006-01-01 / end 2007-31-12
Web site: http://mosaique.labri.fr
Partner(s): LED/LORIA, Calligramme/LORIA, Atoll/INRIA Rocquencourt, Laboratoire de Linguistique Formelle/Paris 7, Laboratoire Parole et Langage/U. de Provence, Lattice/Paris 7, Modyco/Paris X, Signes/INRIA Futurs, TALN/Nantes

Contact: Claire Gardent

8. Dissemination

8.1. Service to the Scientific Community

- Patrick Blackburn:
  - President of the SIGSEM, ACL Special Interest Group in Computational Semantics.
  - Member of the Management Board of the Association of Logic, Language and Information (FoLLI). 2002–2006.
- Claire Gardent:
  - Member of the ESSLLI (European Summer School in Logic, Language, and Information) Standing Committee, 2004–2006.
  - Member of the EACL (European Chapter of the Association for Computational Linguistics) nominating committee.
  - Member of the commission of specialist, University of Nancy 2, Section 27.
- Carlos Areces: Member of the Management Board of the Association of Logic, Language and Information (FoLLI). 2005–Present.
- 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).
- Samuel Cruz-Lara: Elected member of the Scientific Council of the University of Nancy 2.
- 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.
- Laurent Romary:
  - Member of the Scientific Council of TEI (Text Encoding Initiative).
  - Chairman of the ISO committee TC 37/SC 4 on Language Resource Management.

8.1.1. Management Responsibilities

- Claire Gardent:
  - Member of the recruiting committee for short term posts at INRIA Lorraine/LORIA.
  - Manager of the PAI POLONIUM project, "Acquiring and validating subcategorisation lexicons for French and Polish". Partners: LORIA and the Linguistic Engineering Group (Department of Artificial Intelligence at the Institute of Computer Science), Polish Academy of Sciences, Warsaw. 2006 - 2007.
  - Organiser of the Nancy TAL seminar (the LORIA-based seminar on computational linguistics).
- Samuel Cruz-Lara:
  - Person in charge, at the national level, of the reception of Mexican students in the "Professional Licenses of Computer Science".
  - MLIF project leader [ISO AWI 29616].

- Member of the SYMM [Synchronised Multi-Media] group of W3C.

8.1.2. Editorial and Program Committee Work

- Patrick Blackburn:
  - Chief Editor of the *Journal of Logic, Language, and Information*, 2002–Present.
  - Editor of the *Journal of Philosophical Logic*, 2004–Present.
  - Member of Editorial Board of the *Notre Dame Journal of Formal Logic*, 2005–Present.
  - Foreign Correspondent for *Logique et Analyse*, 2005–Present.
  - Subject Editor (Logic and Language) for the *Stanford Encyclopedia of Philosophy*, 2004–Present.
  - Editor of *Handbook of Modal Logic*, 2006.
  - Member of the Programme Committees:
    * The 19th International FLAIRS Conference, Special track on "Trends in Natural Language Processing". May 11-13, 2006. Melbourne Beach, Florida (USA).
    * The 5th Conference on Language Resources and Evaluation (LREC 2006). May 22-28, 2006. Genoa (Italy).
    * 5th International workshop on Inference in Computational Semantics (ICoS-5). April 20-21, 2006. Buxton (England).
    * 2006 Workshop on Hybrid Logics (HyLo 06), LICS 2006 affiliated workshop, Seattle (USA).
    * Advances in Modal Logic (AiML) 2006, Noosa, Queensland, (Australia), September 2006.
- Claire Gardent:
  - Tutorial chair, Coling/ACL-2006 conference, Sydney, Australia, 17-21 July 2006.
  - Member of the Programme Committees:
    * The 11th Conference of the European Chapter of the Association of Computational Linguistics (EACL 2006). April 3-7, 2006. Trento (Italy).
    * The 13th Conference on Natural Language Processing (TALN 2006). April 10-13, 2006. Leuven (Belgium).
    * The 19th International FLAIRS Conference, Special track on "Trends in Natural Language Processing". May 11-13, 2006. Melbourne Beach, Florida (USA).
    * The 5th Conference on Language Resources and Evaluation (LREC 2006). May 22-28, 2006. Genoa (Italy).
    * 5th International workshop on Inference in Computational Semantics (ICoS-5). April 20-21, 2006. Buxton (England).
    * Human Language Technology Conference/North American chapter of the Association for Computational Linguistics annual meeting (HLT/NAACL 2006). June 4-9, 2006. New York (USA).
    * 4th Journées de Sémantique et Modélisation (JSM06). Avril 6-7, 2006. Bordeaux (France)
    * Brandial06, 10th 'Semantics and Pragmatics of Dialogue (SEMDIAL)' workshop series, Potsdam, Germany, September 2006.
- Carlos Areces:

  - Editor of the *Journal of Logic, Language, and Information*, 2005 – Present.
  - Member of Editorial Board of the *Journal of Applied Logics*, 2004 – Present.
  - Program Chair of the ESSLLI 2006 (European Summer School in Logic, Language, and Information), Málaga, Spain.
  - Member of the Programme Committees:
    * 2006 International Workshop on Description Logics - DL2006 Lake District, United Kingdom, 2006.
    * Second International Congress on Tools for Teaching Logic, Salamanca, Spain; of the Campus Multidisciplinar en Percepción e Inteligencia (CMPI-2006), Albacete, Spain.
    * Twenty-First National Conference on Artificial Intelligence (AAAI-06), Boston, USA; of the 5th Inference in Computational Semantics (ICoS-5) Workshop, Buxton, England.
    * 2006 Workshop on Hybrid Logics (HyLo 06), LICS 2006 affiliated workshop, Seattle (USA).

- Daniel Coulon: Member of the Editorial Committee for *Intellectica*.
- Guillaume Pitel: Member of Program Committee of RECITAL 2006.
- Matthieu Quignard: Member of the Review Committee of the Journal "STICEF"
- Laurent Romary
  - Member of the Editorial Boards:
    * *Language Resources and Evaluation*, 2004 – Present.
    * *ACM Transactions on Asian Language Processing*, TALIP.
    * *Computer and the Humanities*.
- Jesse Tseng:
  - Member of the Programme Committees:
    * HPSG 2006 (July 2006, Varna, Bulgaria) Formal Grammar 2006 (July 2006, Malaga, Spain)
    * 3rd ACL-SIGSEM Workshop on Prepositions (EACL, Trento, April 2006) Workshop on Typed Feature Structure Grammars (Scandinavian Conference of Linguistics, June 2006, Alborg, Denmark)
    * Reviewer for *Computational Linguistics*. Special Issue: Prepositions in Applications.

8.1.3. Conference and Workshop Organisation

Guillaume Pitel of LED was the co-organiser (with Susanne Alt of ATILF) of a seminar in the "Journées de l'ATALA" series which was held on May 13th 2006. The theme was "Des ressources sémantiques existantes à un FrameNet français (Contre-)arguments, ressources, methodes et outils".

8.1.4. Seminars and Invited Talks

- Patrick Blackburn:
  - Invited seminar at the Universidad de Salamanca, September 2006.
  - ESSLLI 2006 (European Summer School in Logic, Language, and Information), Malaga, Spain, August 2006:
    * Working with Discourse Representation Theory (postgraduate course).
    * Formal Semantics (postgraduate course).
- Claire Gardent:
  - Semantics in NLP, University of Buenos Aires (Argentina), November 2006. 15 hours, Invited Postgraduate Tutorial.
  - Tree Adjoining Grammar: Theory and Practice. LAICS Summer School (Language, Artificial Intelligence and Computer Science for Natural Language Processing applications), Bangkok (Thailand), October 2006. 4 hours. Invited Tutorial.
  - Natural Language Generation. LAICS Summer School (Language, Artificial Intelligence and Computer Science for Natural Language Processing applications) Bangkok, (Thailand), October 2006. 3 hours. Invited Tutorial.
- Carlos Areces:
  - "Description Logics," invited postgraduate course at the Universidad de Salamanca, January 2006.
  - Invited talk at the University of Buenos Aires, February 2006.
  - Invited talk at the University of Cordoba, February 2006.
  - Invited talk at the University of Manchester, March 2006.
  - "Lógica Modal Computacional," invited postgraduate course at the Universidad de Buenos Aires, Argentina, October to December 2006.

8.2. University Teaching

In 2006, Carlos Areces, Patrick Blackburn, and Claire Gardent in collaboration with Guy Perrier (Calligramme), succeeded in obtaining *Erasmus Mundus* recognition for the Masters degree track in Computational Linguistics at the University of Nancy 2. This links the Nancy 2 Masters to the Masters degrees in computational linguistics at the Universities of Bolzano, Malta, Prague, Saarbrücken, and Groningen. In effect, this means that the joint Masters program offered by these institutes is now the most important point of entry for students wishing to study computational linguistics in Europe. Inclusion in the Erasmus Mundus program should increase the visibility and international recognition of the work in computational linguistics being conducted in Nancy.

Members of LED make a substantial contribution to university teaching at Nancy. Each professor and assistant professor associated with LED teaches at least 96 and 192 hours, respectively, per year in the universities of Nancy. This covers the following LED members: Daniel Coulon, Christine Fay-Varnier, Nadia Bellalem, Samuel Cruz-Lara, and Azim Roussanaly. In addition the following LED-based INRIA and CNRS researchers have contributed the following hours to the Erasmus Mundus Masters track in Computational Linguistics:

- 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. 10 hours.

9. Bibliography

Major publications by the team in recent years

[1] C. ARECES, P. BLACKBURN, M. MARX. *Repairing the interpolation Theorem in Quantified Modal Logic*, in "Annals of Pure and Applied Logics", vol. 123, n^{o} 1–3, 2003, p. 287–299.

[2] C. ARECES, B. TEN CATE. *Hybrid Logics*, in "Handbook of Modal Logic", P. BLACKBURN, F. WOLTER, J. VAN BENTHEM (editors). , chap. 14, Elsevier, 2006, http://hal.inria.fr/inria-00104513/en/.

[3] P. BLACKBURN, J. BOS. *Representation and Inference for Natural Language: A First Course in Computational Semantics*, CSLI Press, 2005, http://hal.inria.fr/inria-00001208/en/.

[4] P. BLACKBURN, B. TEN CATE. *Pure Extensions, Proof Rules, and Hybrid Axiomatics*, in "Studia Logica", vol. 84, 2006, p. 281-326, http://hal.inria.fr/inria-00119855/en/.

[5] P. BLACKBURN, J. VAN BENTHEM, F. WOLTER. *Handbook of Modal Logic*, Springer, 2006, http://hal.inria.fr/inria-00120237/en/.

[6] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Extracting subcategorisation information from Maurice Gross' Grammar Lexicon*, in "Archives of Control Sciences", vol. 15, n^{o} LI, 2005, p. 253-264.

[7] C. GARDENT, E. KOW. *Generating and selecting grammatical paraphrases*, in "Proceedings of the 10th European Workshop on Natural Language Generation, Aberdeen, Scotland", 2005.

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

[9] C. GARDENT, Y. PARMENTIER. *Large scale semantic construction for Tree Adjoining Grammar*, in "Proceedings of Logical Aspects in Computational Linguistics, Bordeaux, France", 2005.

Year Publications

Books and Monographs

[10] P. BLACKBURN, J. BOS, K. STRIEGNITZ. *Learn Prolog Now!*, Texts in Computing, vol. 7, College Publications, 2006, http://hal.inria.fr/inria-00119862/en/.

[11] P. BLACKBURN, J. VAN BENTHEM, F. WOLTER. *Handbook of Modal Logic*, Springer, 2006, http://hal.inria.fr/inria-00120237/en/.

Doctoral dissertations and Habilitation theses

[12] T. M. H. NGUYEN. *Outils et ressources linguistiques pour l'alignement de textes multilingues français-vietnamiens*, Ph. D. Thesis, Université Henri Poincaré - Nancy I, 2006, http://tel.archives-ouvertes.fr/tel-00105592/en/.

Articles in refereed journals and book chapters

[13] A. ABEILLÉ, O. BONAMI, D. GODARD, J. TSENG. *The syntax of French à and de: An HPSG analysis*, in "Syntax and Semantics of Prepositions, Dordrecht", P. SAINT-DIZIER (editor). , Springer, 2006, p. 147–162.

[14] C. ARECES, B. TEN CATE. *Hybrid Logics*, in "Handbook of Modal Logic", F. W. P. BLACKBURN, J. VAN BENTHEM (editors). , chap. 14, Elsevier, 2006, http://hal.inria.fr/inria-00104513/en/.

[15] M. BAKER, A. BRONNER, J. LEHUEN, M. QUIGNARD. *Gestion de l'interaction et modèles de dialogues*, in "Environnements informatiques et apprentissages humains, Paris", M. GRANDBASTIEN, J.-M. LABAT (editors). , chap. 5, Hermès Science Publications, 2006, p. 117–138.

[16] P. BLACKBURN. *Arthur Prior and Hybrid Logic*, in "Synthese", vol. 150, 2006, p. 329-372, http://hal.inria.fr/inria-00119854/en/.

[17] P. BLACKBURN, B. TEN CATE. *Pure Extensions, Proof Rules, and Hybrid Axiomatics*, in "Studia Logica", vol. 84, 2006, p. 281-326, http://hal.inria.fr/inria-00119855/en/.

[18] P. BLACKBURN, J. VAN BENTHEM. *Modal logic: a Semantic Perspective*, in "Handbook of Modal Logic", P. BLACKBURN, J. VAN BENTHEM, F. WOLTER (editors). , chap. 1, Elsevier, 2006, http://hal.inria.fr/inria-00119856/en/.

[19] P. CESAR, K. BAKER, D. BULTERMAN, S. CRUZ-LARA, L. FERNANDO GOMES SOARES, A. KAPTEIN. *Present and Future of Software Graphics Architectures for IDTV*, in "Interactive Digital Television: technologies and applications", G. LEKAKOS, K. CHORIANOPOULOS (editors). , IDEA Group, 2006, http://hal.inria.fr/inria-00105654/en/.

[20] B. GAIFFE. *Références*, in "Compréhension des langues et interaction", G. SABAH (editor). , Traité IC2, Série Cognition et traitement de l'information, Lavoisier, 2006, http://hal.inria.fr/inria-00069441/en/.

[21] M. MUSIOL, A. TROGNON, D. COULON, C. BOCÉRÉAN. *Structures de l'interaction verbale et rationalité argumentative chez l'adolescent polyhandicapé*, in "Le Langage et l'Homme", vol. XXXXI, n^{o} 2, 2006.

Publications in Conferences and Workshops

[22] M. AMOIA, C. GARDENT. *Adjective based inference*, in "Proceedings of KRAQ'06: Knowledge and Reasoning for Answering Questions, Trento, Italy", 2006.

[23] L. BENOTTI. *'Drink me': Handling Actions through Planning in a Text Game Adventure*, in "Proceedings of the 11th ESSLLI Student Session, Málaga, Spain", J. HUITINK, S. KATRENKO (editors). , 2006.

[24] F. BLANQUI, S. COUPET-GRIMAL, W. DELOBEL, S. HINDERER, A. KOPROWSKI. *CoLoR: a Coq library on rewriting and termination*, in "Eighth International Workshop on Termination - WST 2006, Seattle, Washington/USA", 2006, http://hal.inria.fr/inria-00084835/en/.

[25] Y.-C. CHIAO, O. KRAIF, D. LAURENT, T. M. H. NGUYEN, N. SEMMAR, F. STUCK, J. VÉRONIS, W. ZAGHOUANI. *Evaluation of multilingual text alignment systems: the ARCADE II project*, in "5th international Conference on Language Resources and Evaluation - LREC'06, Genoa/Italy", 2006, http://hal.inria.fr/inria-00115670/en/.

[26] S. CRUZ-LARA, N. BELLALEM, J. DUCRET, I. KRAMER. *Interactive Handling of Multilingual Content within Digital Media*, in "4th European Interactive TV Conference - EuroITV 2006 - Beyond Usability, Broadcast, and TV, Athens/Greece", 2006, http://hal.inria.fr/inria-00001119/en/.

[27] S. CRUZ-LARA, N. BELLALEM, J. DUCRET, I. KRAMER. *Interactive Handling of Multilingual Content within Digital Media*, in "EuroiTV 2006 Beyond Usability, Broadcast, and TV, Athens/Greece", 2006, http://hal.inria.fr/inria-00105650/en/.

[28] S. CRUZ-LARA, N. BELLALEM, J. DUCRET, I. KRAMER. *Interoperability between translation memories and localization tools by using the MultiLingual Information Framework*, in "European Association for Machine Translation - EAMT 2006, Oslo/Norway", 2006, http://hal.inria.fr/inria-00105651/en/.

[29] S. CRUZ-LARA, N. BELLALEM, J. DUCRET, I. KRAMER. *Standardizing the management and the representation of multilingual data: the MultiLingual Information Framework*, in "Third International Workshop on Language Resources for Translation Work, Research & Training, Genoa/Italy", 2006, http://hal.inria.fr/inria-00105653/en/.

[30] A. DENIS, G. PITEL, M. QUIGNARD. *A model of grouping for plural and ordinal references*, in "18th European Summer School in Logic, Language and Information - ESSLLI 2006, Malaga/Spain", 2006, http://hal.inria.fr/inria-00110858/en/.

[31] A. DENIS, G. PITEL, M. QUIGNARD. *Resolution of Referents Groupings in Practical Dialogues*, in "7th SIGDial Workshop on Discourse and Dialogue - SIGdial'06, Sydney/Australie", 2006, http://hal.inria.fr/inria-00112888/en/.

[32] A. DENIS, M. QUIGNARD, G. PITEL. *A Deep-Parsing Approach to Natural Language Understanding in Dialogue System: Results of a Corpus-Based Evaluation*, in "Proceedings of Language Resources and Evaluation Conference, Malaga/Spain", 2006, http://hal.inria.fr/inria-00112897/en/.

[33] C. GARDENT. *Intégration d'une dimension sémantique dans les grammaires d'arbres adjoints*, in "Actes de Traitement Automatique des Langues Naturelles, Louvain, Belgique", 2006.

[34] C. GARDENT, B. GUILLAUME, G. PERRIER, I. FALK. *Extraction d'information de sous-catégorisation à partir des tables du LADL*, in "Traitement Automatique de la Langue Naturelle - TALN 2006, Leuven/Belgique", Actes de la 13ème conférence sur le Traitement Automatique de la Langue Naturelle, 2006, http://hal.inria.fr/inria-00103163/en/.

[35] C. GARDENT, E. KOW. *Three reasons to adopt TAG-based surface realisation*, in "The Eighth International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+8), Sydney/Australia", 2006, http://hal.inria.fr/inria-00088784/en/.

[36] C. GARDENT, Y. PARMENTIER. *Coreference handling in XMG*, in "International Conference on Computational Linguistics - COLING 2006, Sydney/Australie", 2006, http://hal.inria.fr/inria-00083546/en/.

[37] D. GORÍN. *Hybrid Layering*, in "Proceedings of the 11th ESSLLI Student Session, Málaga, Spain", J. HUITINK, S. KATRENKO (editors). , 2006.

[38] E. KOW. *GenI: Natural language generation in Haskell*, in "Proceedings of Haskell'06, Portland/USA", 2006, http://hal.inria.fr/inria-00088787/en/.

[39] E. KOW, Y. PARMENTIER, C. GARDENT. *SemTAG, the LORIA toolbox for TAG-based Parsing and Generation*, in "The Eighth International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+8), Sydney, Australia", 2006, http://hal.inria.fr/inria-00083555/en/.

[40] H. P. LE, T. M. H. NGUYEN, L. ROMARY, A. ROUSSANALY. *A Lexicalized Tree-Adjoining Grammar for Vietnamese*, in "International Conference on Language Resources and Evaluation - LREC 2006, Gène/Italie", 2006, http://hal.inria.fr/inria-00106159/en/.

[41] J. LE ROUX, B. CRABBÉ, Y. PARMENTIER. *A constraint driven metagrammar*, in "The Eighth International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+8), Sydney/Australie", 2006, http://hal.inria.fr/inria-00083550/en/.

[42] C.-Y. LU, J.-S. HONG, S. CRUZ-LARA. *Emotion Detection in Textual Information by Semantic Role Labeling and Web Mining Techniques*, in "Third Taiwanese-French Conference on Information Technology - TFIT 2006, Nancy/France", 2006, http://hal.inria.fr/inria-00105649/en/.

[43] S. MERA. *Hybrid Logics with Concrete Domains*, in "Proceedings of the 11th ESSLLI Student Session, Málaga, Spain", J. HUITINK, S. KATRENKO (editors). , 2006.

[44] Y. PARMENTIER, J. LE ROUX, B. CRABBÉ. *XMG - An expressive formalism for describing tree-based grammars*, in "11th Conference of the European Chapter of the Association for Computational Linguistics (Demo Session) - EACL 2006, Trento/Italy", 2006, http://hal.inria.fr/inria-00001133/en/.

[45] G. PITEL. *Scaling Construction Grammar up to Production Systems: the SCIM*, in "Third International Workshop on Scalable Natural Language Understanding - ScaNaLU 2006, New-York/Etats-Unis", 2006, http://hal.inria.fr/inria-00114040/en/.

[46] J. TSENG, A. KUPSC. *A cross-linguistic approach to Slavic past tense and conditional constructions*, in "Proceedings of FDSL6, Potsdam", to appear, 2006.

[47] J. TSENG. *English prepositional passives in HPSG*, in "FG-2006: Proceedings of the 11th conference on Formal Grammar, Málaga", G. JÄGER, P. MONACHESI, G. PENN, S. WINTNER (editors). , 2006, p. 147–159.

Internal Reports

[48] P. BEDARIDE. *Implication textuelle et logiques de description*, Masters Thesis, LORIA, 2006, http://hal.inria.fr/inria-00119768/en/.

[49] S. CRUZ-LARA, S. KUMAR GUPTA, J. DAVID FERNÁNDEZ GARCÍA, L. ROMARY, P. SÉBIRE, F. PRÊTEUX, T. ZAHARIA, M. PREDA, S.-M. TRAN. *Handling multilingual content in digital media: A critical analysis*, Rapport de recherche, LORIA, 2006, http://hal.inria.fr/inria-00001120/en/.

References in notes

[50] A. ABEILLÉ. *Une grammaire électronique du français*, CNRS Editions, 2002.

[51] C. ARECES, D. GORÍN. *Ordered Resolution with Selection for H(@)*, in "Proceedings of LPAR 2004, Montevideo, Uruguay", F. BAADER, A. VORONKOV (editors). , LNCS, vol. 3452, Springer, 2005, p. 125-141.

[52] C. ARECES, J. HEGUIABEHERE. *HyLoRes 1.0: Direct Resolution for Hybrid Logics*, in "Proceedings of CADE-18, Copenhagen, Denmark", A. VORONKOV (editor). , July 2002, p. 156–160.

[53] L. BALKAN, K. NETTER, D. ARNOLD, S. MEIJER. *TSNLP — Test Suites for Natural Language Processing*, in "Proceedings of Language Engineering Convention, Edinburgh", ELSNET, 1994, p. 17–22.

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

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

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

[57] B. GAIFFE, B. CRABBÉ, A. ROUSSANALY. *A New Metagrammar Compiler*, in "Proc. of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks - TAG+6, Venice, Italy", May 2002, http://www.loria.fr/publications/2002/A02-R-495/A02-R-495.ps.

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

[59] 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.

[60] C. GARDENT, E. JACQUEY. *Lexicalisation as a Description Logic Inference Task*, in "Proceedings of ICOS'03 (Inference in Computational Semantics), Nancy, France", 2003.

[61] 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.

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

[63] 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.

[64] M. GROSS. *Lexicon-grammar: the representation of compound words*, in "Proceedings of the 11th coference on Computational linguistics, Morristown, NJ, USA", Association for Computational Linguistics, 1986, p. 1–6.

[65] H. KAMP, U. REYLE. *From Discourse to Logic*, Kluwer, Dordrecht, 1993.

[66] 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.

[67] P. LOPEZ. *Analyse d'énoncés oraux pour le Dialogue Homme-Machine à l'aide de Grammaires Lexicalisées d'Arbres*, Ph. D. Thesis, Université Henri Poincaré – Nancy 1, Nancy, France, 1999.

[68] S. SALMON-ALT. *Référence et dialogue finalisé: de la linguistique à un modèle opérationnel*, Ph. D. Thesis, Université H. Poincaré - Nancy 1, Nancy, France, May 2001.

[69] D. SUSTRETOV. *Hybrid Definability in Topological Spaces*, ILLC Technical Report PP-2005-12, Technical report, University of Amsterdam, 2005, http://www.illc.uva.nl/Publications/reportlist.php?Series=PP.

[70] R. THOMASON, M. STONE, D. DEVAULT. *Enlightened update: A computational architecture for presupposition and other pragmatic phenomena*, in "Presupposition Accommodation", D. BYRON, C. ROBERTS, S. SCHWENTER (editors). , 2006.

[71] M. VARDI. *On the Complexity of Epistemic Reasoning*, in "Proceedings 4th IEEE Symposium on Logic in Computer Science", 1989, p. 243–252.

[72] É. VILLEMONTE DE LA CLERGERIE. *DyALog: a Tabular Logic Programming based environment for NLP*, in "Proceedings of CSLP'05, Barcelona", 2005, ftp://ftp.inria.fr/INRIA/Projects/Atoll/Eric.Clergerie/CSLP05.pdf.