IN PARTNERSHIP WITH:
CNRS

Université de Lorraine

Activity Report 2012

Team SEMAGRAMME

Semantic Analysis of Natural Language

IN COLLABORATION WITH: Laboratoire lorrain de recherche en informatique et ses applications (LORIA)

RESEARCH CENTER
Nancy - Grand Est

THEME
Audio, Speech, and Language Processing

Table of contents

Team SEMAGRAMME

Keywords: Logics, Natural Language, Semantics, Linguistics

Creation of the Team: January 01, 2011 .

1. Members

Research Scientists
Philippe de Groote [Team Leader, Senior Researcher, Inria, HdR]
Bruno Guillaume [Researcher, Inria]
Sylvain Pogodalla [Researcher, Inria]

Faculty Members
Maxime Amblard [Associate Professor, Université de Lorraine]
Guy Perrier [Professor, Université de Lorraine, HdR]
Karën Fort [ATER, Université de Lorraine, since September 2012]

PhD Students
Ekaterina Lebedeva [Inria Cordi Grant, until April 2012]
Aleksandre Maskharashvili [Contrat doctoral, Inria, since November 2012]
Florent Pompigne [ATER, Université de Lorraine]
Sai Qian [ATER, Université de Lorraine]

Administrative Assistant
Marie-Françoise Loubressac

2. Overall Objectives

2.1. Overall objectives

Computational linguistics is a discipline at the intersection of computer science and linguistics. On the theoretical side, it aims to provide computational models of the human language faculty. On the applied side, it is concerned with natural language processing and its practical applications.

From a structural point of view, linguistics is traditionally organized into the following sub-fields:

- Phonology, the study of language abstract sound systems.
- Morphology, the study of word structure.
- Syntax, the study of language structure, i.e., the way words combine into grammatical phrases and sentences.
- Semantics, the study of meaning at the levels of words, phrases, and sentences.
- Pragmatics, the study of the ways in which the meaning of an utterance is affected by its context.

Computational linguistics is concerned by all these fields. Consequently, various computational models, whose application domains range from phonology to pragmatics, have been developed. Among these, logic-based models play an important part, especially at the "higher" levels.

At the level of syntax, generative grammars [24] may be seen as basic inference systems, while categorial grammars [33] are based on substructural logics specified by Gentzen sequent calculi. Finally, model-theoretic grammars [42] amount to sets of logical constraints to be satisfied.

At the level of semantics, the most common approaches derive from Montague grammars, [34], [35], [36] which are based on the simply typed λ-calculus and Church's simple theory of types. [25] In addition, various logics (modal, hybrid, intensional, higher- order...) are used to express logical semantic representations.

At the level of pragmatics, the situation is less clear. The word *pragmatics* has been introduced by Morris [38] to designate the branch of philosophy of language that studies, besides linguistic signs, their relation to their users and the possible contexts of use. The definition of pragmatics was not quite precise, and for a long time several authors have considered (and some authors are still considering) pragmatics as the wastebasket of syntax and semantics. [21] Nevertheless, as far as discourse processing is concerned (which includes pragmatic problems such as pronominal anaphora resolution), logic-based approaches have also been successful. In particular, Kamp's Discourse Representation Theory [31] gave rise to sophisticated 'dynamic' logics. [29] The situation, however, is less satisfactory than it is at the semantic level. On the one hand, we are facing a kind of logical "tower of Babel". The various pragmatic logic-based models that have been developed, while sharing underlying mathematical concepts, differ in several respects and are too often based on *ad hoc* features. As a conssence, they are difficult to compare and appear more as competitors than as collaborative theories that could be integrated. On the other hand, several phenomena related to discourse dynamics (e.g., context updating, presupposition projection and accommodation, contextual reference resolution...) are still lacking deep logical explanations. We strongly believe, however, that this situation can be improved by applying to pragmatics the same approach Montague applied to semantics, using the standard tools of mathematical logic.

Accordingly:

> *The overall objective of the Sémagramme project is to design and develop new unifying logic-based models, methods, and tools for the semantic analysis of natural language utterances and discourses. This includes the logical modelling of pragmatic phenomena related to discourse dynamics. Typically, these models and methods will be based on standard logical concepts (stemming from formal language theory, mathematical logic, and type theory), which should make them easy to integrate.*

The project is organized along three research directions (i.e., *Syntax-semantics interface*, *Discourse dynamics*, and *Common basic resources*), which interact as explained in the following paragraphs.

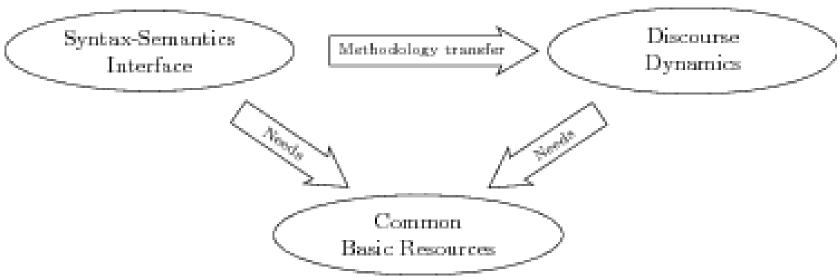

Figure 1.

2.1.1. Syntax-semantics interface

The Sémagramme project intends to focus on the semantics of natural languages (in a wider sense than usual, including some pragmatics). Nevertheless, the semantic construction process is syntactically guided, that is, the constructions of logical representations of meaning is based on the analysis of the syntactic structures. We do not want, however, to commit ourselves to such or such specific theory of syntax. Consequently, our approach should be based on an abstract generic model of the syntax-semantic interface.

Here, an important idea of Montague comes into play, namely, the "homomorphism requirement": semantics must appear as a homomorphic image of syntax. While this idea is almost a truism in the context of mathematical logic, it remains challenged in the context of natural languages. Nevertheless, Montague's idea has been quite fruitful, especially in the field of categorial grammars, where van Benthem showed how syntax ans semantics could be connected using the Curry-Howard isomorphism. [51] This correspondence is the keystone of the syntax-semantics interface of modern type-logical grammars. [37] It also motivated the definition of our own Abstract Categorial Grammars. [45]

Technically, an Abstract Categorial Grammar consists simply of a (linear) homomorphism between two higher-order signatures. Extensive studies have shown that this simple model allows several grammatical formalisms to be expressed, providing them with a syntax-semantics interface for free. [46], [49], [50], [40], [32], [41]

We intend to carry on with the development of the Abstract Categorial Grammar framework. At the foundational level, we will define and study possible type theoretic extensions of the formalism, in order to increase its expressive power and its flexibility. At the implementation level, we will continue the development of an Abstract Categorial Grammar support system.

As said above, to consider the syntax-semantics interface as the starting point of our investigations allows us not to be committed to some specific syntactic theory. The Montagovian syntax-semantics interface, however, cannot be considered to be universal. In particular, it does not seem to be that well adapted to dependency and model-theoretic grammars. Consequently, in order to be as generic as possible, we intend to explore alternative models of the syntax-semantics interface. In particular, we will explore relational models where several distinct semantic representations can correspond to a same syntactic structure.

2.1.2. Discourse dynamics

It is well known that the interpretation of a discourse is a dynamic process. Take a sentence occurring in a discourse. On the one hand, it must be interpreted according to its context. On the other hand, its interpretation affects this context, and must therefore result in an updating of the current context. For this reason, discourse interpretation is traditionally considered to belong to pragmatics. The cut between pragmatics and semantics, however, is not that clear.

As we mentioned above, we intend to apply to some aspects of pragmatics (mainly, discourse dynamics) the same methodological tools Montague applied to semantics. The challenge here is to obtain a completely compositional theory of discourse interpretation, by respecting Montague's homomorphism requirement. We think that this is possible by using techniques coming from programming language theory, in particular, continuation semantics [44], [22], [23], [43] and the related theories of functional control operators. [26], [27]

We have indeed successfully applied such techniques in order to model the way quantifiers in natural languages may dynamically extend their scope. [47] We intend to tackle, in a similar way, other dynamic phenomena (typically, anaphora and referential expressions, presupposition, modal subordination...).

What characterize these different dynamic phenomena is that their interpretations need information to be retrieved from a current context. This raises the question of the modeling of the context itself. At a foundational level, we have to answer questions such as the following. What is the nature of the information to be stored in the context? What are the processes that allow implicit information to be inferred from the context? What are the primitives that allow a context to be updated? How does the structure of the discourse and the discourse relations affect the structure of the context? These questions also raise implementation issues. What are the appropriate datatypes? How can we keep the complexity of the inference algorithms sufficiently low?

2.1.3. Common basic resources

Even if our research primarily focuses on semantics and pragmatics, we nevertheless need syntax. More precisely, we need syntactic trees to start with. We consequently need grammars, lexicons and parsing algorithms to produce such trees. During the last years, we have developped the notion of interaction grammar [30] as a model of natural language syntax. This includes the development of grammar for French, [39] together with morpho-syntactic lexicons. We intend to continue this line of research and development. In particular, we want to increase the coverage of our French grammar, and provide our parser with more robust algorithms.

Further primary resources are needed in order to put at work a computational semantic analysis of utterances and discourses. As we want our approach to be as compositional as possible, we must develop lexicons annotated with semantic information. This opens the quite wide research area of lexical semantics.

Finally, when dealing with logical representations of utterance interpretations, the need for inference facilities is ubiquitous. Inference is needed in the course of the interpretation process, but also to exploit the result of the interpretation. Indeed, an advantage of using formal logic for semantic representations is the possibility of using logical inference to derive new information. From a computational point of view, however, logical inference may be highly complex. Consequently, we need to investigate which logical fragments can be used efficiently for natural language oriented inference.

3. Scientific Foundations

3.1. Fondation

The present proposal relies on deep mathematical foundations. We intend to develop models based on well-established mathematics. We seek two main advantages from this approach. On the one hand, by relying on mature theories, we have at our disposal sets of mathematical tools that we can use to study our models. On the other hand, developing various models on a common mathematical background will make them easier to integrate, and will ease the search for unifying principles.

The main mathematical domains on which we rely are formal language theory, symbolic logic, and type theory.

3.1.1. Formal language theory

studies the purely syntactic and combinatorial aspects of languages, seen as sets of strings (or possibly trees or graphs). Formal language theory has been especially fruitful for the development of parsing algorithms for context-free languages. We use it, in a similar way, to develop parsing algorithms for formalisms that go beyond context-freeness. Language theory also appears to be very useful in formally studying the expressive power and the complexity of the models we develop.

3.1.2. Symbolic logic

(and, more particularly, proof-theory) is concerned with the study of the expressive and deductive power of formal systems. In a rule-based approach to computational linguistics, the use of symbolic logic is ubiquitous. As we previously said, at the level of syntax, several kinds of grammars (generative, categorial...) may be seen as basic deductive systems. At the level of semantics, the meaning of an utterance is capture by computing (intermediate) semantic representations that are expressed as logical forms. Finally, using symbolic logics allows one to formalize notions of inference and entailment that are needed at the level of pragmatics.

3.1.3. Type theory and typed λ-calculus

Among the various possible logics that may be used, Church's simply typed λ-calculus and simple theory of types (a.k.a. higher-order logic) play a central part. On the one hand, Montague semantics is based on the simply typed λ-calculus, and so is our syntax-semantics interface model. On the other hand, as shown by Gallin, [28] the target logic used by Montague for expressing meanings (i.e., his intensional logic) is essentially a variant of higher-order logic featuring three atomic types (the third atomic type standing for the set of possible worlds).

4. Application Domains

4.1. Introduction

Our applicative domains concern natural language processing applications that rely on a deep semantic analysis. For instance, one may cite the following ones:

- textual entailment and inference;
- dialogue systems;
- semantic-oriented query systems;
- content analysis of unstructured documents;
- text transformation and automatic summarization;
- (semi) automatic knowledge acquisition.

However, if the need for semantics seems to be ubiquitous. There is a challenge in finding applications for which a deep semantic analysis results in a real improvement over non semantic-based techniques.

4.2. Text Transformation

Text transformation is an application domain featuring two important sub-fields of computational linguistics:

- parsing: from surface form to abstract representation;
- generation: from abstract representation to surface form.

Text simplification or automatic summarization belong to that domain.

We aim at using the framework of Abstract Categorial Grammars we develop to this end. It is indeed a *reversible* framework that allows both parsing and generation. Its underlying mathematical structure of λ-calculus makes it fit with our type-theoretic approach to discourse dynamics modeling. The ANR project POLYMNIE(see section 7.2.1.1) is especially dedicated to this aim.

5. Software

5.1. Leopar

Participants: Bruno Guillaume [correspondant], Guy Perrier.

Interaction Grammar, parsing

5.1.1. Software description

Leopar is a parser for natural languages which is based on the formalism of Interaction Grammars [30]. It uses a parsing principle, called "electrostatic parsing" which consists in neutralizing opposite polarities. A positive polarity corresponds to an available linguistic feature and a negative one to an expected feature.

Parsing a sentence with an Interaction Grammar consists in first selecting a lexical entry for each of its words. A lexical entry is an underspecified syntactic tree, a tree description in other words. Then, all selected tree descriptions are combined by partial superposition guided by the aim of neutralizing polarities: two opposite polarities are neutralized by merging their support nodes. Parsing succeeds if the process ends with a minimal and neutral tree. As IGs are based on polarities and under-specified trees, Leopar uses some specific and non-trivial data-structures and algorithms.

The electrostatic principle has been intensively considered in Leopar. The theoretical problem of parsing IGs is NP-complete; the nondeterminism usually associated to NP-completeness is present at two levels: when a description for each word is selected from the lexicon, and when a choice of which nodes to merge is made. Polarities have shown their efficiency in pruning the search tree:

- In the first step (tagging the words of the sentence with tree descriptions), we forget the structure of descriptions, and only keep the bag of their features. In this case, parsing inside the formalism is greatly simplified because composition rules reduce to the neutralization of a negative feature-value pair $f \longleftarrow v$ by a dual positive feature-value pair $f \longrightarrow v$. As a consequence, parsing reduces to a counting of positive and negative polarities present in the selected tagging for every pair (f, v): every positive occurrence counts for $+1$ and every negative occurrence for -1, the sum must be 0.

- Again in the tagging step, original methods were developped to filter out bad taggings. Each unsaturated polarity p in the grammar induces constraints on the set of contexts in which it can be used: the unsaturated polarity p must find a *companion* (*i.e.* a tree description able to saturated it); and the set of companions for the polarity p can be computed statically from the grammar. Each lexical selection which contains an unsaturated polarity without one of its companions can be safely removed.
- In the next step (node-merging phase), polarities are used to cut off parsing branches when their trees contain too many non neutral polarities.

5.1.2. Current state of the implementation

Leopar is presented and documented at http://leopar.loria.fr; an online demonstration page can be found at http://leopar.loria.fr/demo.

It is open-source (under the CECILL License http://www.cecill.info) and it is developed using the InriaGforge platform (http://gforge.inria.fr/projects/semagramme/)

The main features of current software are:

- automatic parsing of a sentence or a set of sentences,
- dependency and parse-tree representation of sentences,
- interactive parsing (the user chooses the couple of nodes to merge),
- visualization of grammars produced by XMG-2 or of sets of description trees associated to some word in the linguistic resources,

5.2. ACG Development Toolkit

Participants: Sylvain Pogodalla [correspondant], Philippe de Groote.

In order to support the theoretical work on ACG, we have been developing a support system. The objectives of such a system are twofold:

1. to make possible to implement and experiment grammars the modeling of linguistic phenomena;
2. to make possible to implement and experiment results related to the ACG formalisms. Such results can concern parsing algorithms, type extensions, language extensions, etc.

The ACG Development toolkit development effort is part of the POLYMNIE project (see Section 7.2.1.1. It will support the experimentation and evaluation parts of the project.

The current version of the the ACG development toolkit prototype [1] issues from a first release published in October 2008. Further releases have been published before the ESSLLI 2009 course on ACG. It focuses on providing facilities to develop grammars. To this end, the type system currently implemented is the linear core system plus the (non-linear) intuitionistic implication, and a special attention has been paid to type error management. As a major limitation, this version only considers transformation from abstract terms to object terms, and not the other way around.

Enabling transformation from the object terms to the abstract terms is the first step of future development for the ACG support system. A parsing algorithm based on [32]'s methods is being implemented for second-order ACGs. It is based on a translation of ACG grammars into Datalog programs and is well-suited to fine-grained optimization.

However, since we're interested not only by recognizability (hence whether some fact is provable) but also by the parsing structure (hence the proof), the Datalog solver requires further adaptations. Note however that in the general case, the decidability of translating an object term to an abstract one is still an open problem.

5.3. Grew

Participants: Bruno Guillaume [correspondant], Guy Perrier.

[1] Available at http://acg.gforge.inria.fr with a CeCILL license.

Graph rewriting, Interface syntaxe-sémantique

Grew is a Graph Rewriting tools dedicated to applications in NLP. It is freely-available (from the page http://grew.loria.fr) and it is developed using the InriaGforge platform (http://gforge.inria.fr/projects/semagramme/)

We list below some of the major specificities of the GREW software.

- Graph structures can use a build-in notion of feature structures.
- The left-hand side of a rule is described by a graph called a pattern; injective graph morphisms are used in the pattern matching algorithm.
- Negative pattern can be used for a finer control on the left-hand side of rules.
- The right-hand side or rules is described by a sequence of atomic commands that describe how the graph should be modified during the rule application.
- Rules can be parametrized by lexical information.
- Filters can be used at the output of each module to control the structure produced are well-formed.
- Subset of rules are grouped in modules; the full rewriting process being a sequence of module applications.
- The Grew software has support both for confluent and non-confluent modules; when a non-confluent modules is used, all normal forms are returned and then ambiguity is handled in a natural way.
- Grew can be used on Corpus mode with statistics about rules usage or with an a Graphical User Interface which can show all intermediate graphs used during the rewriting process (useful either to debug rewriting system or for demonstrations).

A demonstration of the Grew Software was presented at the TALN conference in june in Grenoble.[15]

The Grew software were used for several kind of applications manipulating syntactic and/or semantic graph representations. It was used to build DMRS semantic representation from syntactic dependency trees in the French TreeBank [12], [14]. More recently, it was used on the Sequoia TreeBank, to produce deep syntax annotation and DMRS Semantic representations.

Another application of the Grew software which is currently investigated is the detection of annotation errors in corpora. Graph Rewriting is use to detect ill-formed structures that don't fit the annotation guide requirements. In collaboration with the Alpage team, this was applied to the Sequoia Corpus and the reported errors were corrected in version 3.2 and 3.3 of the corpus[2].

5.4. Other developments

Participant: Bruno Guillaume [correspondant].

Concordancer, Dependencies, Graphical tools Other peripheral developments of the team are available either as web service of as downloadable code:

- A concordancer named CONDOR which is usable online: http://condor.loria.fr. With Condor, ut is possible to search for all inflexions (given by a lexicon) of some lemma; it is possible to search for a couple of lemmas to find collocations.
- A program (named DEP2PICT) to build graphical representations (PNG, SVG or PDF) of dependency structures. It presented in http://dep2pict.loria.fr; it is usable online http://dep2pict.loria.fr/demo.

[2] https://www.rocq.inria.fr/alpage-wiki/tiki-index.php?page=CorpusSequoia

6. New Results

6.1. Syntax-Semantics Interface

6.1.1. Graph Rewriting

Bruno Guillaume and Guy Perrier have proposed a system for annotating the French Treebank with semantic dependencies [12], [14]. This system (Synsem_FTB) is based on Graph Rewriting. Graph Rewriting is a framework which is well-suited for syntax-semantic interface because it allows for a modular development of large systems. Each modelled linguistic phenomenon is described by a small set of local rewriting rules. The whole transformation is then described by a sequence of modules to apply successively to the input structure. Another benefit of the Graph Rewriting formalism is that it handles the ambiguity in a natural way with the use of non confluent rewriting systems.

The Synsem_FTB system produces a semantic annotation in the framework of DMRS starting from an annotation with surface syntactic dependencies. It contains 34 modules that can be split in two main parts; the first part produces a deep syntax annotation of the input and the second one rewrites deep syntax to semantics.

With respect to previous works, the system of rewriting rules itself has been improved: it has a larger coverage (causative constructions, rising verbs, ...) and the order between modules has been studied in a more systematic way.

The rewriting calculus has been enriched on two points: the use of rules to make a link with lexicons, especially with the lexicon of verbs Dicovalence, and the introduction of filters to discard inconsistent annotations at some computation steps.

This system has been experimented on the whole French Treebank with the Grew software, which implements the used rewriting calculus.

6.1.2. Passive Sentences

Chris Blom, Philippe de Groote, Yoad Winter, and Joost Zwarts have proposed a unified syntactic-semantic account of passive sentences and sentences with an unspecified object [18]. For both constructions, they use *option types* for introducing implicit arguments into the syntactic-semantic categorial mechanism. They show the advantages of this approach over previous proposals in the domains of scope and unaccusatives. Unlike pure syntactic treatments, option types immediately derive the obligatory narrow scope of existential quantification over an implicit argument's slot. Unlike purely semantic, event-based treatments, their solution naturally accounts for syntactic contrasts between passives and unaccusatives.

6.1.3. Intensionalization

Makoto Kanazawa and Philippe de Groote have defined a general *intensionalization* procedure that turns an extensional semantics for a language into an intensionalized one that is capable of accommodating *truly intensional* lexical items without changing the compositional semantic rules [48]. They have proved some formal properties of this procedure and have clarified its relation to the procedure implicit in Montague's PTQ.

6.1.4. Plural

Sai Qian and Maxime Amblard have modeled the semantics of plurality in continuation semantics [13]. Two types of discourse antecedents formations, inherited from the classical treatment, namely summation and abstraction, are studied in detail. Solutions for each phenomenon are provided respectively by introducing two new functions `Sum` and `Abs`, for obtaining the semantic interpretations.

6.2. Discourse Dynamics

In a joint work with a psycho-linguistist (Michel Musiol, ATILF) and a philosopher (Manuel Rebuschi, Archives Poincaré), are developing a formal analysis of pathological conversations involving schizophrenic speakers [16]. Such conversations give rise to manifest incongruities or ruptures that can be seen as mere contradictions by any "normal" speaker. Our analysis relies both on semantic and pragmatic features of conversation. We propose a SDRT-inspired [20] account of pathological conversations, and we apply it to two relevant excerpts. We conclude with a short discussion about the localization of inconsistencies by schizophrenics, either in semantics or in pragmatics, and its importance for our understanding of thought disorders.

7. Partnerships and Cooperations

7.1. Regional Initiatives

7.1.1. SLAM: Schizophrenia and Language, Analysis and Modeling

Participants: Maxime Amblard [coordinator], Sylvain Pogodalla.

Schizophrenia is well-known among mental illnesses for the strength of the thought disorders it involves, and for their widespread and spectacular manifestations: from deviant social behavior to delusion, not to speak about affective and sensitive distortions. It aims at exploring a specific manifestation, namely disorders in conversational speech. This is an interdisciplinary research, both empirical and theoretical from several domains, namely psychology, philosophy, linguistic and computer science.

Maxime Amblard is coordinating the pre-project which ended at the end of 2012. A new application on this topic is send for a 2013-2015 project to the Maison des Sciences de l'Homme de Lorraine (MSH–Lorraine, USR 3261), with the same leader. While this year work was dedicated to the test protocol definition, the coming years will be devoted to building an open-access corpus of pathological uses of language. Other participants are: Denis Apotheloz (ATILF, Université de Lorraine),Valérie Aucouturier (Centre Léo Apostel, Université Libre de Bruxelles), Katarina Bartkova (ATILF, Université de Lorraine), Fethi Bretel (CHS Le Rouvray, Rouen), Michel Musiol (InterPSY, Université de Lorraine), Manuel Rebuschi (Archives Poincaré, Université de Lorraine).

7.2. National Initiatives

7.2.1. ANR

7.2.1.1. Polymnie: Parsing and synthesis with abstract categorial grammars. From lexicon to discourse

Participants: Maxime Amblard, Philippe de Groote, Aleksandre Maskharashvili, Sylvain Pogodalla [coordinator], Sai Qian.

POLYMNIE[3] is a research project funded by the French national research agency (ANR). It relies on the grammatical framework of Abstract Categorial Grammars (ACG). A feature of this formalism is to provide the same mathematical perspective both on the surface forms and on the more abstract forms the latter correspond to. As a consequence:

- ACG allows for the encoding of a large variety of grammatical formalisms such as context-free grammars, Tree Adjoining grammars (TAG), etc.
- ACG define two languages: an abstract language for the abstract forms, and an object language for the surface forms.

[3] http://semagramme.loria.fr/doku.php?id=projects:polymnie

Importantly, the notions of object language and abstract language are relative to each other. If we can naturally see surface forms as strings for instance and abstract forms as the associated syntactic trees, we can also consider to associate this abstract form to a first order logical formula as surface (object) form. This property it central in our project as it offers a unified approach to text analysis and text generation, in particular considering the underlying algorithms and their complexity.

ACG definition uses type-theory and lambda-calculus. From this point of view, they smoothly integrate formal semantics models issuing from Montague's proposal. Theories that extend to the discourse level such as Discourse Representation Theory (DRT) and Dynamic Predicate Logic (DPL) were not initially formulated using lambda-calculus. But such formulation have been proposed. In particular, a formulation based on continuation semantics allow them to be expressed quite naturally in the ACG architecture. Dynamic effects of discourse, in particular those related to anaphora resultion or rhetoretical relation inference, have then to be expressed by lexical semantics or computed from the syntactic rules as studied in the Inria Collaborative Research Project (ARC) CAuLD [4].

It has been shown that the discourse structure of texts play a key role in their understanding. This is the case not only for both for human readers but also for automatic processing systems. For instance, it can enhance text transformation systems such as the ones performing automatic summarization.

POLYMNIEfocuses on studying and implementing the modeling of sentences and discourses in a compositional paradigm that takes into account their dynamics and their structures, both in parsing and in generation. To that end, we rely on the ACG framework. The kind of processing we are interested in relate to the automatic construction of summaries or to text simplification. This has to be considered in the limits of the modelling of the linguistic processes (as opposed to inferential processes for instance) these tasks involve.

The complexity of the phenomena, of their formal description, and of their interactions, require to set up a testing and development environment for linguistic modelling. It will consist in extending and stabilizing a software implementing the functionnalities of the ACG framework. It will provide a tool for experimentation and validation of the approach.

Partners:

- Sémagramme people
- Alpage (Paris 7 university & Inria Paris-Rocquencourt): Laurence Danlos (local coordinator), C. Braud, C. Roze, Éric Villemonte de la Clergerie
- MELODI (IRIT, CNRS): Stergos Afantenos, Nicholas Asher (local coordinator), Juliette Conrath, Philippe Muller
- Signes (LaBRI, CNRS): Jérôme Kirman, Richard Moot, Christian Retoré (local coordinator), Sylvain Salvati, Noémie-Fleur Sandillon-Rezer

7.3. International Initiatives

7.3.1. Participation In International Programs

7.3.1.1. PHC: Partenariats Hubert Curien

The team collaborates with the Utrecht Institues of Linguistics OTS (Utrecht University) in the framework of a Van Gogh action (Hubert Curien program). This collaborations is concerned with conservative extensions of Montague semantics.

[4] http://www.loria.fr/~pogodall/cauld/

7.4. International Research Visitors

7.4.1. Visits of International Scientists

- Prof. A. Hadj-Salah (Académie Algérienne de la Langue Arabe) visited Philippe de Groote (January, 10).
- Chris Blom and Yoad Winter (University of Utrecht) visited Philippe de Groote (March, 28-30).

7.4.2. Visits to International Teams

- Philippe de Groote and Sylvain Pogodalla visited Prof. Makoto Kanazawa at the National Institute of Informatics (NII, Tokyo, Japan).
- Philippe de Groote was invited speaker at the workshop 'Properties and Optionality in Syntax and Semantics', Utrecht, February, 13-14.
- Philippe de Groote visited Prof. Yoad Winter at the Utrecht Institute of Linguistics (Utrecht University, the Netherlands), May 23-25.

8. Dissemination

8.1. Scientific Animation

- Maxime Amblard:
  - Head of the preoperative project SLAM: Schizophrenia and Language: Analyze and Modeling—MSH–Lorraine.
  - Vice-treasurer of the Association pour le Traitement Automatique des Langues (ATALA)
  - Head of the master Sciences Cognitives et Applications of University de Lorraine (until July 2012). He was responsible for the application to degree habilitation of the new version of the master mention Science Cognitive et ApplicationA, Université de Lorraine.
  - Member Opération Poste team [5] (until June 2012).
  - Member of the comité de sélection McF 4058 at LIPN – Université Paris 13
- Philippe de Groote:
  - Member of the editorial board of the journal *Higher-Order and Symbolic Computation*.
  - Area editor of the *FoLLI Publications on Logic, Language and Information* (LNCS subline).
  - Member of the ESSLLI standing committee.
  - Member of the scientific committee of the international conference on Logical Aspects of Computational Linguistics (LACL) 2012 and of the international conference on Traitement Automatique de Langues Naturelles (TALN) 2012.
- Bruno Guillaume:
  - Member (elected) of the “Pôle scientifique” IAEM of Université de Lorraine
  - Member of the COMPIERS Inria Nancy Grand-Est
- Guy Perrier:
  - Member of the editorial board of the journal *Traitement Automatique des Langues*.
  - Member of the scientific committee of the international conference on on Traitement Automatique de Langues Naturelles (TALN) 2012.
- Sylvain Pogodalla:

[5] http://postes.smai.emath.fr/

- Member of the editorial board of the journal *Traitement Automatique des Langues*.
- Head of the Commission des Développements Technologiques (CDT) of the Nancy Inria Research Center.
- Coordinator of the POLYMNIE ANR Project.
- Local coordinator of the Erasmus Mundus Master program Language and Communication Technologies for the Université de Lorraine.
- Member of the scientific committee of the International Conference on Logical Aspects of Computational Linguistics (LACL) 2012.
- Member of the jury of the ATALA Dissertation prize (dissertation prize of the French Association pour le Traitement Automatique des Langues).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence: Maxime Amblard, *Penser les algorithmes*, 25h, L1, Université de Lorraine, France

Licence: Maxime Amblard, *Introduction au Traitement Automatique des Langues*, 25h, L2, Université de Lorraine, France

Licence: Maxime Amblard, *Algorithmique avancée*, 62.5h, L2, Université de Lorraine, France

Licence: Maxime Amblard, *Formalismes de représentation et raisonnement*, 25h, L3, Université de Lorraine, France

Licence: Guy Perrier, *Outils conceptuels*, 37.5h, L3, Université de Lorraine, France

Master: Maxime Amblard, *Algorithmique pour l'Intelligence Artificielle*, 45h, M1, Université de Lorraine, France

Master: Maxime Amblard, *Discourse and Dialogue*, 25h, M2, Université de Lorraine, France

Master: Philippe de Groote (with Benoît Crabbé), *Structures informatiques et logiques pour la modélisation linguistique*, 18h, M2, Parisian Master of Research in Computer Science, France

Master: Philippe de Groote (with Kamel Smaïli), *Logique et statistiques pour la modélisation des langues*, 50h, M2, Université de Lorraine, France

Master: Bruno Guillaume, *Grammatical Formalisms*, 25h, M2, Université de Lorraine, France

Master: Guy Perrier, *Initiation au TAL*, 37.5h, M1, Université de Lorraine, France

Master: Guy Perrier, *Programmation pour le TAL*, 37.5h, M1, Université de Lorraine, France

Master: Guy Perrier, *Tools and algorithms for NLP*, 44h, M2, Université de Lorraine, France

Master: Sylvain Pogodalla, *Computational Semantics*, 25h, M2, Université de Lorraine, France

Master: Sai Qian, *Applications of Natural Language Processing*, 37.5h, M2, Université de Lorraine, France

Master: Sai Qian, *Algorithmique pour l'Intelligence Artificielle*, 22h, M1, Université de Lorraine, France

8.2.2. Supervision

Ekaterina Lebedeva, *Expression de la dynamique du discours à l'aide de continuations*, Université de Lorraine, defended on April 6 2012, Philippe de Groote [9]

Aleksandre Maskharashvili, *Generation and Discourse with Abstract Categorial Gramamrs*, since November 2012, Philippe de Groote and Sylvain Pogodalla

Florent Pompigne, *Modélisation logique de la langue et Grammaires Catégorielles Abstraites*, since September 2008, Philippe de Groote and Sylvain Pogodalla

Sai Qian, *Investigation of variable accessibility in natural language semantic modelization*, since September 2009, Maxime Amblard and Philippe de Groote

8.2.3. Juries

Guy Perrier was member of the PhD committee of François-Régis Chaumartin, Université Paris Diderot, September 25 2012

Guy Perrier was member of the PhD committee of Christian Gillot, Université de Lorraine, December 12 2012

Philippe de Groote was member of the PhD committee of Pierre Bourreau, University of Bordeaux 1, June 29 2012

8.3. Popularization

Maxime Amblard is member of the editorial board of)i(interstices.

Maxime Amblard supervised a group of Phd students for vulgarization of science with drama, Journée internationales Hubert Curien, Nancy, 3-4-5 September 2012.

Guy Perrier and Bruno Guillaume are interviewed in the scientific TV program “C à Savoir” on France 3 Lorraine channel. The program is available at http://tinyurl.com/semagramme-france3.

9. Bibliography

Major publications by the team in recent years

[1] N. Asher, S. Pogodalla. *A Montagovian Treatment of Modal Subordination*, in "20th Semantics and Linguistic Theory conference - SALT2010", Vancouver, Canada, N. Li, D. Lutz (editors), 2011, http://hal.inria.fr/inria-00565616/en.

[2] G. Bonfante, B. Guillaume, M. Morey, G. Perrier. *Modular Graph Rewriting to Compute Semantics*, in "9th International Conference on Computational Semantics - IWCS 2011", Oxford, Royaume-Uni, J. Bos, S. Pulman (editors), January 2011, p. 65–74, http://hal.inria.fr/inria-00579244/en/.

[3] G. Bonfante, B. Guillaume, G. Perrier. *Analyse syntaxique électrostatique*, in "Traitement Automatique des Langues (TAL)", 2003, vol. 44, n° 3, p. 93-120, Article dans revue scientifique avec comité de lecture nationale, http://hal.inria.fr/inria-00099582/en/.

[4] B. Guillaume, G. Perrier. *Interaction Grammars*, in "Research on Language & Computation", 2009, vol. 7, p. 171–208.

[5] P. de Groote. *Towards abstract categorial grammars*, in "Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter", Toulouse, France, none, July 2001, p. 148-155, Colloque avec actes et comité de lecture. internationale., http://hal.inria.fr/inria-00100529/en.

[6] P. de Groote, S. Pogodalla. *On the expressive power of abstract categorial grammars: Representing context-free formalisms*, in "Journal of Logic, Language and Information", 2004, vol. 13, n° 4, p. 421–438.

[7] P. de Groote, S. Pogodalla, C. Pollard. *About Parallel and Syntactocentric Formalisms: A Perspective from the Encoding of Convergent Grammar into Abstract Categorial Grammar*, in "Fundamenta Informaticae", 2011, vol. 106, n° 2-4, p. 211-231 [*DOI :* 10.3233/FI-2011-384], http://hal.inria.fr/inria-00565598/en.

[8] P. DE GROOTE. *Towards a Montagovian Account of Dynamics*, in "16th Semantics and Linguistic Theory conference - SALT2006", Tokyo, Japan, M. GIBSON, J. HOWELL (editors), 2006, http://elanguage.net/journals/index.php/salt/article/view/16.1.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[9] E. LEBEDEVA. *Expression de la dynamique du discours à l'aide de continuations*, Université de Lorraine, 2012, In English, https://sites.google.com/site/katyaspage/publications/lebedeva-thesis.pdf.

International Conferences with Proceedings

[10] K. FORT, A. NAZARENKO, S. ROSSET. *Modeling the Complexity of Manual Annotation Tasks: a Grid of Analysis*, in "International Conference on Computational Linguistics (COLING)", Mumbai, India, December 2012.

[11] Y. MATHET, A. WIDLÖCHER, K. FORT, C. FRANÇOIS, O. GALIBERT, C. GROUIN, J. KAHN, S. ROSSET, P. ZWEIGENBAUM. *Manual Corpus Annotation: Evaluating the Evaluation Metrics*, in "International Conference on Computational Linguistics (COLING)", Mumbai, India, December 2012, Poster.

[12] G. PERRIER, B. GUILLAUME. *Semantic Annotation of the French Treebank with Modular Graph Rewriting*, in "META-RESEARCH Workshop on Advanced Treebanking, LREC 2012 Workshop", Istanbul, Turquie, J. HAJIC (editor), META-NET, May 2012, http://hal.inria.fr/hal-00760577.

[13] S. QIAN, M. AMBLARD. *Accessibility for Plurals in Continuation Semantics*, in "the Ninth International Workshop of Logic and Engineering of Natural Language Semantics 9 (LENLS 9)", Myasaki, Japon, 978-4-915905-51-3 C3004 (JSAI), November 2012, p. 52-65, http://hal.inria.fr/hal-00762203.

National Conferences with Proceeding

[14] G. PERRIER, B. GUILLAUME. *Annotation sémantique du French Treebank à l'aide de la réécriture modulaire de graphes*, in "Conférence annuelle sur le Traitement Automatique des Langues - TALN'12", Grenoble, France, G. ANTONIADIS, H. BLANCHON, G. SÉRASSET (editors), ATALA, June 2012, p. 293-306, http://hal.inria.fr/hal-00760629.

[15] G. PERRIER, B. GUILLAUME. *Grew : un outil de réécriture de graphes pour le TAL*, in "12ième Conférence annuelle sur le Traitement Automatique des Langues (TALN'12)", Grenoble, France, G. ANTONIADIS, H. BLANCHON, G. SÉRASSET (editors), ATALA, June 2012, http://hal.inria.fr/hal-00760637.

Conferences without Proceedings

[16] M. AMBLARD, M. MUSIOL, M. REBUSCHI. *Schizophrénie et Langage : Analyse et modélisation. De l'utilisation des modèles formels en pragmatique pour la modélisation de discours pathologiques*, in "Congrès MSH 2012", Caen, France, December 2012, http://hal.inria.fr/hal-00761540.

Scientific Books (or Scientific Book chapters)

[17] M. AMBLARD, S. POGODALLA. *Modeling the Dynamic Effects of Discourse: Principles and Frameworks*, in "Dialogue, Rationality, Formalism", G. HEINZMANN, M. MUSIOL, M. REBUSCHI, A. TROGNON (editors), Logic, Epistemology, and the Unity of Science, Springer, 2013, http://hal.inria.fr/hal-00737765.

[18] P. DE GROOTE, C. BLOM, Y. WINTER, J. ZWARTS. *Implicit Arguments: Event Modification or Option Type Categories?*, in "Amsterdam Colloquium on Logic, Language and Meaning", M. ALONI, V. KIMMELMAN, F. ROELOFSEN, G. W. SASSOON, K. SCHULZ, M. WESTERA (editors), Lecture Notes in Computer Science, Springer, Amsterdam, Pays-Bas, 2012, vol. 7218, p. 240-250 [*DOI :* 10.1007/978-3-642-31482-7_25], http://hal.inria.fr/hal-00763102.

Books or Proceedings Editing

[19] P. DE GROOTE, M.-J. NEDERHOF (editors). *Formal Grammar - 15th and 16th International Conferences, FG 2010, Copenhagen, Denmark, August 2010, FG 2011, Ljubljana, Slovenia, August 2011, Revised Selected Papers*, Springer, 2012, vol. 7395, 307 [*DOI :* 10.1007/978-3-642-32024-8], http://hal.inria.fr/hal-00763203.

References in notes

[20] N. ASHER, A. LASCARIDES. *Logics of conversation*, Cambridge University Press, 2003.

[21] Y. BAR-HILLEL. *Out of the pragmatic wastebasket*, in "Linguistic Inquiry", 1971, vol. 2, n° 3, p. 401–407.

[22] C. BARKER. *Continuations and the Nature of Quantification*, in "Natural Language Semantics", 2002, vol. 10, n° 3, p. 211–242.

[23] C. BARKER. *Continuations in Natural Language*, in "Proceedings of the Fourth ACM SIGPLAN Continuations Workshop (CW'04)", 2004.

[24] N. CHOMSKY. *Three models for the description of language*, in "Information Theory, IRE Transactions on", 1956, vol. 2, n° 3, p. 113–124.

[25] A. CHURCH. *A formulation of the simple theory of types*, in "The journal of symbolic logic", 1940, vol. 5, n° 2, p. 56–68.

[26] M. FELLEISEN, D. P. FRIEDMAN, E. KOHLBECKER, B. DUBA. *A syntactic theory of sequential control*, in "Theor. Comput. Sci.", June 1987, vol. 52, p. 205–237, http://dx.doi.org/10.1016/0304-3975(87)90109-5.

[27] M. FELLEISEN, R. HIEB. *The revised report on the syntactic theories of sequential control and state*, in "Theoretical Computer Science", 1992, vol. 103, n° 2, p. 235 - 271 [*DOI :* 10.1016/0304-3975(92)90014-7], http://www.sciencedirect.com/science/article/pii/0304397592900147.

[28] D. GALLIN. *Intensional and higher-order modal logic: with applications to Montague semantics*, North-Holland mathematics studies, North-Holland, 1975, http://books.google.fr/books?id=jkiNo5Rok7cC.

[29] J. GROENENDIJK, M. STOKHOF. *Dynamic predicate logic*, in "Linguistics and philosophy", 1991, vol. 14, n° 1, p. 39–100.

[30] B. GUILLAUME, G. PERRIER. *Interaction Grammars*, in "Research on Language & Computation", 2009, vol. 7, p. 171–208.

[31] H. KAMP, U. REYLE. *From discourse to logic: Introduction to modeltheoretic semantics of natural language, formal logic and discourse representation theory*, Kluwer Academic Dordrecht,, The Netherlands, 1993, vol. 42.

[32] M. KANAZAWA, S. SALVATI. *Generating Control Languages with Abstract Categorial Grammars*, in "Proceedings of the 12th conference on Formal Grammar", 2007.

[33] J. LAMBEK. *The mathematics of sentence structure*, in "The American Mathematical Monthly", 1958, vol. 65, n^{o} 3, p. 154–170.

[34] R. MONTAGUE. *English as a formal language*, in "Linguaggi nella societae nella tecnica", 1970, p. 189–224.

[35] R. MONTAGUE. *Universal grammar*, in "Theoria", 1970, vol. 36, n^{o} 3, p. 373–398.

[36] R. MONTAGUE. *The Proper Treatment of Quantification in Ordinary English*, in "Formal Semantics", 1973, p. 221–242.

[37] M. MOORTGAT. *Categorial type logics*, in "Handbook of logic and language", 1997, p. 93–177.

[38] C. MORRIS. *Foundations of the Theory of Signs*, University of Chicago Press, 1938, vol. 1.

[39] G. PERRIER. *A French Interaction Grammar*, in "International Conference on Recent Advances in Natural Language Processing - RANLP 2007", Borovets, Bulgarie, G. ANGELOVA, K. BONTCHEVA, R. MITKOV, N. NICOLOV, K. SIMOV (editors), INCOMA Ltd, Shoumen, Bulgaria, 2007, p. 463-467, http://hal.inria.fr/inria-00184108/en/.

[40] S. POGODALLA. *Computing semantic representation: Towards ACG abstract terms as derivation trees*, in "Proceedings of the Seventh International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+7)", 2004.

[41] C. RETORÉ, S. SALVATI. *A Faithful Representation of Non-Associative Lambek Grammars in Abstract Categorial Grammars*, in "Journal of Logic Language and Information", April 2010, vol. 19, n^{o} 2, p. 185–200, http://hal.inria.fr/inria-00409557/en/.

[42] J. ROGERS. *A model-theoretic framework for theories of syntax*, in "Proceedings of the 34th annual meeting on Association for Computational Linguistics", Association for Computational Linguistics, 1996, p. 10–16.

[43] C.-C. SHAN. *Delimited continuations in natural language: quantification and polarity sensitivity*, in "CoRR", 2004, vol. cs.CL/0404006.

[44] C. STRACHEY, C. P. WADSWORTH. *Continuations: A Mathematical Semantics for Handling Full Jumps*, in "Higher-Order and Symbolic Computation", 1974, vol. 13, p. 135–152.

[45] P. DE GROOTE. *Towards abstract categorial grammars*, in "Proceedings of the 39th Annual Meeting on Association for Computational Linguistics", Association for Computational Linguistics, 2001, p. 252–259.

[46] P. DE GROOTE. *Tree-Adjoining Grammars as Abstract Categorial Grammars*, in "TAG+6, Proceedings of the sixth International Workshop on Tree Adjoining Grammars and Related Frameworks", Università di Venezia, 2002, p. 145–150.

[47] P. DE GROOTE. *Towards a montagovian account of dynamics*, in "Semantics and Linguistic Theory (SALT) 16", M. GIBSON, J. HOWELL (editors), 2006, p. 1-16.

[48] P. DE GROOTE, M. KANAZAWA. *A note on intensionalization*, submitted, http://semanticsarchive.net/Archive/WMyNTA2N/intensionalization.pdf.

[49] P. DE GROOTE, S. POGODALLA. *m-linear context-free rewriting systems as abstract categorial grammars*, in "oceedings of Mathematics of Language - MOL-8", 2003.

[50] P. DE GROOTE, S. POGODALLA. *On the expressive power of abstract categorial grammars: Representing context-free formalisms*, in "Journal of Logic, Language and Information", 2004, vol. 13, n^{o} 4, p. 421–438.

[51] J. VAN BENTHEM. *Essays in logical semantics*, Reidel, 1986, n^{o} 29.