Activity Report 2018

Project-Team SEMAGRAMME

Semantic Analysis of Natural Language

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

1. **Team, Visitors, External Collaborators** .................................................... 1

2. **Overall Objectives** ................................................................................... 1
   2.1. Scientific Context 1
   2.2. Syntax-Semantics Interface 2
   2.3. Discourse Dynamics 3
   2.4. Common Basic Resources 3

3. **Research Program** ................................................................................... 4
   3.1. Overview 4
   3.2. Formal Language Theory 4
   3.3. Symbolic Logic 4
   3.4. Type Theory and Typed λ-Calculus 4

4. **Application Domains** ............................................................................... 5
   4.1. Deep Semantic Analysis 5
   4.2. Text Transformation 5

5. **New Software and Platforms** ................................................................. 5
   5.1. ACGtk 5
   5.2. Dep2pict 6
   5.3. Grew 6
   5.4. ZombiLingo 6

6. **New Results** ........................................................................................... 7
   6.1. Syntax-Semantics Interface 7
      6.1.1. Abstract Categorial Grammars 7
      6.1.2. Syntax-Semantics Interface as Graph Rewriting 7
      6.1.3. Lexical Semantics 7
   6.2. Discourse Dynamics 7
      6.2.1. Dynamic Logic 8
      6.2.2. Discourse Relations 8
      6.2.3. Dynamic Generalized Quantifiers 8
      6.2.4. Dialogue Modeling 8
      6.2.5. Pathological Discourse Modelling 9
   6.3. Common Basic Resources 9
      6.3.1. Application of Graph Rewriting to Natural Language Processing 9
      6.3.2. Building Linguistics Resources with Crowdsourcing 10
      6.3.3. Corpus Annotation 10
      6.3.4. FR-Fracas 10
      6.3.5. Large Coverage Abstract Categorial Grammars 10

7. **Partnerships and Cooperations** .............................................................. 10
   7.1. National Initiatives 10
   7.2. International Initiatives 11
   7.3. International Research Visitors 11

8. **Dissemination** ......................................................................................... 11
   8.1. Promoting Scientific Activities 11
      8.1.1. Scientific Events Selection 11
         8.1.1.1. Chair of Conference Program Committees 11
         8.1.1.2. Member of Conference Program Committees 11
         8.1.1.3. Reviewer 12
      8.1.2. Journal 12
         8.1.2.1. Member of Editorial Boards 12
         8.1.2.2. Reviewer - Reviewing Activities 12
8.1.3. Invited Talks 12
8.1.4. Leadership within the Scientific Community 12
8.1.5. Scientific Expertise 13
8.1.6. Research Administration 13
8.2. Teaching - Supervision - Juries 13
  8.2.1. Teaching 13
  8.2.2. Supervision 14
  8.2.3. Juries 15
8.3. Popularization 15
  8.3.1. Internal or External Inria Responsibilities 15
  8.3.2. Articles and Contents 15
  8.3.3. Education 15
  8.3.4. Interventions 15
  8.3.5. Creation of Media or Tools for Science Outreach 15
9. Bibliography ........................................................................15
Project-Team SEMAGRAMME

Creation of the Team: 2011 January 01, updated into Project-Team: 2013 July 01

Keywords:

**Computer Science and Digital Science:**
- A5.8. - Natural language processing
- A7.2. - Logic in Computer Science
- A9.4. - Natural language processing

**Other Research Topics and Application Domains:**
- B9.6.8. - Linguistics

1. Team, Visitors, External Collaborators

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

**Faculty Members**
- Maxime Amblard [Univ de Lorraine, Associate Professor, HDR]
- Guy Perrier [Univ de Lorraine, Emeritus, HDR]
- Michel Musiol [Inria secondment from Sep 2018]

**PhD Students**
- William Babonnaud [Univ de Lorraine, from Sep 2018]
- Timothée Bernard [Univ Diderot, Paris]
- Clément Beysson [Univ de Lorraine]
- Maria Boritchev [Inria]
- Pierre Ludmann [Univ de Lorraine]

**Technical staff**
- Nicolas Lefebvre [Inria, until Mar 2018]

**Interns**
- Mathieu Collet [Inria, from Apr 2018 until Aug 2018]
- Maxime Guillaume [Univ de Lorraine, from Apr 2018 until Jul 2018]
- Agnes Richard [CNRS, from Sep 2018]
- Valentin Richard [Ecole Normale Supérieure Paris, from Jun 2018 until Jul 2018]

**Administrative Assistants**
- Isabelle Herlich [Inria]
- Delphine Hubert [Univ de Lorraine]
- Annick Jacquot [CNRS, from Jul 2018]
- Martine Kuhlmann [CNRS]

2. Overall Objectives

2.1. Scientific Context

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 “highest” levels.

At the level of syntax, generative grammars [37] may be seen as basic inference systems, while categorial grammars [54] are based on substructural logics specified by Gentzen sequent calculi. Finally, model-theoretic grammars [66] amount to sets of logical constraints to be satisfied.

At the level of semantics, the most common approaches derive from Montague grammars [55], [56], [57], which are based on the simply typed λ-calculus and Church’s simple theory of types [38]. 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 [59] 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 [34]. 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 [52] gave rise to sophisticated ‘dynamic’ logics [48].

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 consequence, 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 modeling 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 below.

### 2.2. 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 are 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 and semantics could be connected using the Curry-Howard isomorphism [71]. This correspondence is the keystone of the syntax-semantics interface of modern type-logical grammars [58]. It also motivated the definition of our own Abstract Categorial Grammars [2].

Technically, an Abstract Categorial Grammar simply consists 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 [4], [69], [70], [62], [53], [65].

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 the same syntactic structure.

2.3. 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 [68], [35], [36], [67], and the related theories of functional control operators [45], [46].

We have indeed successfully applied such techniques in order to model the way quantifiers in natural languages may dynamically extend their scope [3]. We intend to tackle, in a similar way, other dynamic phenomena (typically, anaphora and referential expressions, presupposition, modal subordination...).

What characterizes 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?

2.4. 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 developed the notion of interaction grammar [49] and graph rewriting [22] as models of natural language syntax. This includes the development of grammars for French [60], 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 grammars for French, and provide our parsers 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. Research Program

3.1. Overview

The research program of Sémagramme aims 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.

3.2. Formal Language Theory

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.

3.3. Symbolic Logic

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 captured 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.

3.4. Type Theory and Typed $\lambda$-Calculus

Among the various possible logics that may be used, Church’s simply typed $\lambda$-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 $\lambda$-calculus, and so is our syntax-semantics interface model. On the other hand, as shown by Gallin [47], 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. Deep Semantic Analysis

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.

It seems clear, nowadays, that the need for semantics is ubiquitous. Nevertheless, according to the present state of the art, there are only a few applications for which a deep semantic analysis results in a real improvement over non-semantic-based techniques. This is due to the fact that most current application chains are such that their weakest links are not located at the semantic level.

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 \( \lambda \)-calculus makes it fit with our type-theoretic approach to discourse dynamics modeling.

5. New Software and Platforms

5.1. ACGtk

Abstract Categorial Grammar Development Toolkit

KEYWORDS: Natural language processing - NLP - Syntactic analysis - Semantics

SCIENTIFIC DESCRIPTION: Abstract Categorial Grammars (ACG) are a grammatical formalism in which grammars are based on typed lambda-calculus. A grammar generates languages: the abstract language (the language of parse structures), and the object language (the language of the surface forms, e.g., strings, or higher-order logical formulas), which is the realization of the abstract language.

ACGtk provides softwares to develop and to use ACGs: acgc, which is a grammar compiler, and acg, which is an interpreter of a command language that allows us, for instance, to parse and realize terms.

FUNCTIONAL DESCRIPTION: ACGtk provides softwares for developing and using Abstract Categorial Grammars (ACG).
NEWS OF THE YEAR: The new version extends the syntax for defining operators that can be used in grammars and removes dependencies to obsolete libraries. It also introduces some light optimizations compared to the previous one.

- Participants: Philippe De Groote, Jiri Marsik, Sylvain Pogodalla and Sylvain Salvati
- Contact: Sylvain Pogodalla
- URL: http://acg.loria.fr/

5.2. Dep2pict

KEYWORDS: Syntactic analysis - Semantics

FUNCTIONAL DESCRIPTION: Dep2pict is a program for drawing graphical representation of dependency structures of natural language sentences. Dep2pict takes into account the modified format mixing surface and deep syntactic information used in deep-sequoia.

NEWS OF THE YEAR: The software was adapted to some extensions of the CoNLL format. A new Graphical User Interface (based on PyQt5) was built to replace the previous one (https://gitlab.inria.fr/dep2pict/gui). It can be installed through PyPI (https://pypi.org/project/dep2pict-gui/)

- Contact: Bruno Guillaume
- URL: http://dep2pict.loria.fr/

5.3. Grew

Graph Rewriting

KEYWORDS: Semantics - Syntactic analysis - Natural language processing - Graph rewriting

FUNCTIONAL DESCRIPTION: Grew is a Graph Rewriting tool dedicated to applications in NLP. Grew takes into account confluent and non-confluent graph rewriting and it includes several mechanisms that help to use graph rewriting in the context of NLP applications (built-in notion of feature structures, parametrization of rules with lexical information).

NEWS OF THE YEAR: In 2018, the version 1.0 of Grew was released. The major novelties are a new implementation of lexical rules and the introduction of a Python binding (described in the book: Application of Graph Rewriting to Natural Language Processing)

- Participants: Bruno Guillaume, Guy Perrier and Guillaume Bonfante
- Contact: Bruno Guillaume
- Publications: Application de la réécriture de graphes au traitement automatique des langues - Application of Graph Rewriting to Natural Language Processing
- URL: http://grew.fr/

5.4. ZombiLingo

KEYWORDS: Syntactic analysis - Natural language processing - Lexical resource - Collaborative science

FUNCTIONAL DESCRIPTION: ZombiLingo is a prototype of a GWAP (Game With A Purpose) where gamers have to give linguistic information about the syntax of natural language sentence, currently in French, and later to other languages.
NEWS OF THE YEAR: The code was factorized and 3 independant librairies where built (available in the github project: https://github.com/gwaps4nlp/ to facilitate their usage in other projects.

- Authors: Bruno Guillaume, Karën Fort, Nicolas Lefebvre and Valentin Stern
- Contact: Karën Fort
- URL: http://zombilingo.org/

6. New Results

6.1. Syntax-Semantics Interface


6.1.1. Abstract Categorial Grammars

Although Abstract Categorial Grammars have well established formal properties that make them suitable for language modeling, some missing features hinder their practical use. For instance, in order to have a compact description of grammatical properties such as number agreement between the subject and the verb of a sentence, a very common approach is to have syntactic descriptions augmented with feature value matrices. Having such a mechanism in Abstract Categorial Grammars requires a lot of attention in order to avoid impacting their computational properties (a previous approach using dependent types showed that, if too general, the problem may become intractable [64]). We have been working on theoretical approaches to this problem from different perspectives: looking for a computationally adequate type extension of the formalisms, and using the composition capabilities of the framework.

We also have been working on a unifying and general framework, provided by a categorical generalization of Abstract Categorial Grammars [50]. The goal is to get a unified approach to several semantic modeling, and to add numerical methods to the formalism.

6.1.2. Syntax-Semantics Interface as Graph Rewriting

In their book (English version: [22] and French version: [21]), Guillaume Bonfante (LORIA, Université de Lorraine), Bruno Guillaume and Guy Perrier devote two chapters to the usage of the Graph Rewriting formalism in the modeling of Syntax-Semantics Interface. Chapter 4 presents two existing semantics formalisms and shows how they can be encoded as graphs: Abstract Meaning Representation (AMR) [33] and Dependency Minimal Recursion Semantics (DMRS) [43], [42]. Chapter 5 described two Graph Rewriting Systems proposed by the authors to build semantics graphs in these two formalisms from syntactic dependencies.

6.1.3. Lexical Semantics

The lexicon model underlying Montague semantics is an enumerative model that would assign a meaning to each atomic expression. This model does not exhibit any interesting structure. In particular, polysemy problems are considered as homonymy phenomena: a word has as many lexical entries as it has senses, and the semantic relations that might exist between the different meanings of a same word are ignored. To overcome these problems, models of generative lexicons have been proposed in the literature. Implementing these generative models in the realm of the typed \(\lambda\)-calculus necessitates a calculus with notions of subtyping and type coercion. William Babonnaud is currently developing such a calculus.

6.2. Discourse Dynamics

Participants: Maxime Amblard, Timothée Bernard, Clément Beysson, Maria Boritech, Philippe de Groote, Bruno Guillaume, Pierre Ludmann, Michel Musiol.
6.2.1. Dynamic Logic

We have revisited the type-theoretic dynamic logic introduced in [3]. We have shown how a slightly richer notion of continuation together with an appropriate notion of polarity results in a richer and more powerful framework. In particular, it allows new dynamic connectives and quantifiers to be defined in a systematic way. This work has been presented as an invited talk at the LACompLing 2018 symposium [11].

6.2.2. Discourse Relations

A text as a whole must exhibit some coherence that makes it more than just a bag of sentences. This coherence hinges on discourse relations (DRs), that express the articulations between the different segments of the text. Typical DRs include relations of Contrast, Consequence or Explanation. The most direct and reliable way to express a DR is to use a discourse connective (e.g., because, instead, for example). These lexical items have specific syntactic, semantic and pragmatic properties, the study of which is the subject of Timothée Bernard’s PhD thesis.

Some discourse connectives (typically, adverbial connectives such as so or otherwise) have only one syntactic argument. It then seems natural to use an anaphora mechanism to retrieve the other argument from the context. This proposal has been formalized in [12] by means of continuation-based type theoretic dynamic logic. In this model, the semantic arguments of a DR are considered to be abstract entities akin to Davidsonian events. This approach raises difficulties when the argument of DR is a negative sentence. Indeed, according to the standard analysis of negation in event semantics, a negative sentence does not introduce any specific event. In order to circumvent this problem, we have developed a logical theory of negative events [13], [17], [29].

6.2.3. Dynamic Generalized Quantifiers

Clement Beysson has continued his work on dynamic generalized quantifiers as denotations of the (French) determiners. In this context, he has studied several issues raised by the modeling of plural determiners. In particular, the opposition between distributive and collective interpretations suggests that intrinsically dynamic plural determiners should introduce plural discourse referents that stand for collection of entities. In order to formalize this notion, he has studied several theories of plurality: mereology, plural logic, and second-order logic.

6.2.4. Dialogue Modeling

Maxime Amblard and Maria Boritchev develop a dynamic approach of dialogue modelling. One of the main difference between discourse and dialogue is the interactions between the speakers. To do so, they introduce a formal approach to compositional processing of questions and answers. They address dialogue lexicality issues starting from the formal definitions of so-called Düsseldorf Frame Semantics given in [51]. They introduce a view of dialogues as compositions of negotiation phases that can be studied separately one from another while linked by a common dialogue context (accessible to all participants of a dialogue). They apply Inquisitive Semantics [39] in that context.

Maxime Amblard and Maria Boritchev works on the categorisation of questions and answers and apply some machine learning approaches for automatic classification. They present the architecture of the model, especially how to handle these phenomena with logical representations in [14]. Their view is to narrow the problem of identifying incomprehension in dialogue to the one of finding logical incoherences in speech act combinations as the one we found in the SLAM project (ongoing project of the Sémagramme team on interviews with schizophrenics). They also start to build a new corpus - DinG (Discourse in Dialogue) - based on record and transcript plays to the settlers of Catan board game.

Maxime Amblard also started a cooperation with CLASP, especially with Robin Cooper, Ellen Breitholtz and Chris Howes. They work on the synchronisation of the representation of dialogue modelling with the previous proposals and Type-Theoretic-Records (TTR) [41]. They apply the solution on extracts from two corpora where patients with schizophrenia are involved.
6.2.5. Pathological Discourse Modelling

Michel Musiol obtained a part-time delegation in the Semagramme team. This proximity makes possible to set up a more active dialogue on the issue of pathological discourse modeling. He has worked on the development of the possibility of testing his conjectures on the cognitive and psychopathological profile of the interlocutors, in addition to information provided by the model of ruptures and incongruities in pathological discourse. This methodological system makes it possible to discuss, or even evaluate, the heuristic potential of the computational models developed on the basis of empirical facts.

Moreover, the diagnostic tools used today by the professional community (clinical and psychiatric) are of limited expertise for the effective identification of the signs of the pathology for at least two reasons: on the one hand, they are much too imprecise on the side of the recognition of Language Impairment and Thought Disorder (no underlying linguistic and psycholinguistic theories); on the other hand, they do not take into account (either theoretically or technically) the discursive structure within which these disorders are expressed. The objective of this research program is therefore also to anticipate the development of diagnostic tools for the psychiatric and psychological community.

As part of the work carried out in the SLAM project, Maxime Amblard, Michel Musiol and Manuel Rebuschi (Archives Henri-Poincaré, Université de Lorraine) continue to work on modelling interactions with schizophrenic patients. The project has progressed on three different operational levels: building new resources, editing a volume (Springer) on the SLAM project in 2019 and improving the representation model.

An agreement is being deployed with the psychiatric hospital of Aix-en-Provence. The on-site staff administered a test protocol to the entire test group of 60 people. Transcripts are in progress, which will provide a significant amount of data to work on for the project. Thanks to the involvement of a medical staff, the recovery of new data appears well advanced. In the same perspective, contacts are being made with the Psychotherapeutic Centre in Nancy.

In addition, Maxime Amblard carried out a one-week international mobility at CLASP thanks to a mobility grant from the French Embassy in Sweden. Discussions were initiated with these colleagues for the development of projects using formal semantic models for the analysis of interaction with schizophrenic patients.

6.3. Common Basic Resources

Participants: Maxime Amblard, Clément Beysson, Philippe de Groote, Bruno Guillaume, Maxime Guillaume, Guy Perrier, Sylvain Pogodalla, Nicolas Lefebvre.

6.3.1. Application of Graph Rewriting to Natural Language Processing

Guillaume Bonfante, Bruno Guillaume and Guy Perrier collected their work on the application of graph rewriting to Natural Language Processing (NLP) in a book written in French [21] and translated to English [22] by the editor. This book shows how graph rewriting can be used as a computational model adapted to NLP. Currently, there is no standard model for graph rewriting and, as such, the authors have conceived one that is specifically adapted to NLP, proposing their own implementation: the GREW system. In addition to the application to Syntax-Semantic Interface mentioned above, the book presents applications in syntactic parsing and in syntactic corpus conversion.

In [5], Guillaume Bonfante and Bruno Guillaume describe some mathematical properties of the Graph Rewriting framework used in GREW. The previous experiments on NLP tasks have shown that Graph Rewriting applications to Natural Language Processing do not require the full computational power of the general Graph Rewriting setting. The most important observation is that all graph vertices in the final structures are in some sense "predictable" from the input data and so, it is possible to consider the framework of Non-size increasing Graph Rewriting. The paper concerns the theoretical aspect of termination with respect to this calculus. It is shown that uniform termination is undecidable and that non-uniform termination is decidable. We define termination techniques based on weight, we prove the termination of weighted rewriting systems and we give complexity bounds on derivation lengths for these rewriting systems.
6.3.2. Building Linguistics Resources with Crowdsourcing

In the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions, Karên Fort (Sorbonne Université), Bruno Guillaume, Matthieu Constant (ATILF, Nancy), Nicolas Lefebvre and Yann-Alan Pilatte (Sorbonne Université) presented the results obtained in crowdsourcing French speakers’ intuition concerning multi-word expressions (MWEs) [15]. They developed a slightly gamified crowdsourcing platform, part of which is designed to test users’ ability to identify MWEs with no prior training. The participants perform relatively well at the task, with a recall reaching 65% for MWEs that do not behave as function words.

6.3.3. Corpus Annotation

Kim Gerdes (Sorbonne nouvelle, Paris 3), Bruno Guillaume, Sylvain Kahane (Université Paris Nanterre) and Guy Perrier proposed a surface-syntactic annotation scheme called Surface Universal Dependencies (SUD) that is near-isomorphic to the Universal Dependencies (UD) annotation scheme. The SUD scheme follows distributional criteria for defining the dependency tree structure and the naming of the syntactic functions [16]. Rule-based graph transformation grammars allow for a bi-directional transformation of UD into SUD. The back-and-forth transformation can serve as an error-mining tool to assure the intra-language and inter-language coherence of the UD treebanks. The UD corpora are available on gitlab.inria.fr.

Bruno Guillaume and Guy Perrier used the GREW system for the development of the French part of the Universal Dependencies project (UD) [32]. They focused in particular on correcting the annotation of two French corpora, UD_French-GSD and UD_French-Sequoia. For the correction, they first used the tool Grew-match (based on the pattern matching part of GREW) to detect error patterns, but also the GREW rewriting rule system to transform the annotation from one format to another one [19]. Version 2.3 of the UD corpora was released on 15 November 2018.

6.3.4. FR-Fracas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume and Sylvain Pogodalla continue their work on the FR-Fracas project. There are two major levels of processing that are significant in the use of a computational semantics framework: semantic composition, for the construction of meanings, and inference, either to exploit those meanings, or to assist the determination of contextually sensitive aspects of meanings. FraCas is an inference test suite for evaluating the inferential competence of different NLP systems and semantic theories. Providing an implementation of the inference level was beyond the scope of FraCaS, but the test suite nevertheless provides an overview of a useful and theory- and system-independent semantic tool [40].

There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the translation into French of the test suite. All translations were subject to a bidding phase by two project members. Then the cases that were identified as difficult were discussed by all project members. An adjudication step finally ensured the quality of the translation. In order to evaluate the inference mechanism triggered by the translated sentences, a web interface is being developed.

6.3.5. Large Coverage Abstract Categorial Grammars

Maxime Amblard, Maxime Guillaume, and Sylvain Pogodalla have worked on the automatic translation of large coverage Tree-Adjoining grammars into Abstract Categorial Grammars. On the theoretical side, this work hinges on the encoding proposed by Philippe de Groote and Sylvain Pogodalla [69], [63]. On the implementation side, the starting point are TAG grammars generated from meta-grammars by XMG [44], [61]. This generates Abstract Categorial grammars containing about 23,000 entries, and was used as a test bed for the ACGtk toolkit, some parts of which have been rewritten to scale up.

7. Partnerships and Cooperations
7.1. National Initiatives

7.1.1. PLURAL

- Program: Langues et Numérique 2018 (DGLFLF: Délégation générale à la langue française et aux langues de France)
- Project acronym: PLURAL
- Project title: Production L'Udique de Ressources Annotées pour les Langues de France (Gamified production of annotated resources for Languages of France)
- Duration: October 2017 - June 2018
- Coordinator: Bruno Guillaume
- Other partners: Université Paris-Sorbonne (Karën Fort, Alice Millour, André Thibault) and Université de Strasbourg (Delphine Bernhard).
- Abstract: The objective of the PLURAL project is to build linguistic resources with GWAPs (Game With A Purpose) for poorly endowed languages. Unlike other languages, poorly endowed languages lack freely available raw corpora. The goal of the PLURAL project is to provide a web interface to gather corpora in poorly endowed languages of France. First target languages are Alsacian and Guadeloupean creole. The main difficulty is to take into account orthographic diversity and regional diversity for these languages.

Nicolas Lefebvre was employed as an engineer in the PLURAL project from October 2017 to March 2018.

7.2. International Initiatives

7.2.1. Informal International Partners

Maxime Amblard continues discussing with the Centre for Linguistic Theory and Studies in Probability (CLASP, University of Gothenburg, Sweden), especially with Robin Cooper, Ellen Breitholtz and Chris Howes. The discussions are about computational treatments of dialogues modelling. We have common issues about the management corpora and models of dialogue. As for now, ongoing discussions have not yet been turned into a formal project.

7.3. International Research Visitors

7.3.1. Visits to International Teams

7.3.1.1. Explorer programme

Maxime Amblard visited Gotenborg University, Sweden, from October 21 to October 26, 2018.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Selection

8.1.1.1. Chair of Conference Program Committees

- Sylvain Pogodalla: co-chair of FG 2018 23rd Conference on Formal Grammar [23].

8.1.1.2. Member of Conference Program Committees

- Maxime Amblard: 25ème conférence sur le Traitement Automatique des Langues Naturelles (TALN 2018)
8.1.1.3. Reviewer


8.1.2. Journal

8.1.2.1. Member of Editorial Boards

- Philippe de Groote: area editor of the FoLLI-LNCS series; associate editor of Higher-Order and Symbolic Computation; member of the editorial board of Cahiers du Centre de Logique.
- Sylvain Pogodalla: Member of the editorial board of the journal Traitement Automatique des Langues, in charge of the Résumés de thèses section.
- Maxime Amblard: Member of the editorial board of the journal Traitement Automatique des Langues, in charge of the hard copy editorial process, Editor of the (In)coherence of Discourse (upcoming volume of the Language, Cognition, and Mind series)

8.1.2.2. Reviewer - Reviewing Activities


8.1.3. Invited Talks

- Maxime Amblard gave four invited talks:
  - at the Colloque Cathy Dufour, November 2018, Nancy [6];
  - at the EMLeX lecture series/Séminaire de l’ATILF, March 2018, Nancy [7];
  - at the Recherches linguistiques et corpus, séminaire STIH de la Faculté des Lettres de Sorbonne Université, March 2018, Paris [9];
  - at the Linguistique textuelle, linguistique de corpus, April 2018, Metz [8].
- Maria Boritchev gave a talk to the Séminaire des doctorantes et doctorants en informatique, June 2018, Paris [10].

8.1.4. Leadership within the Scientific Community

- Philippe de Groote: president of SIGMOL, Association for Mathematics of Language, a Special Interest Group of the Association for Computational Linguistics; member of the LACL steering committee.
- Bruno Guillaume: Management Committee Substitute of the COST Action CA16105 "European Network for Combining Language Learning with Crowdsourcing Techniques" (http://www.cost.eu/COST_Actions/ca/CA16105).
- Sylvain Pogodalla: member of the LACL steering committee; member of the Formal Grammar standing committee.
• Maxime Amblard: head of the work package on Natural Language Processing of the OLKI project (PIA funding), leader of the semantics topic for the national pre-GDR on Traitement Automatique des Langues.

8.1.5. Scientific Expertise

• Philippe de Groote: member of the scientific council of the LIRMM, Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier; member of the scientific council of the AREN e-FRAN project, ARGumentation Et Numérique.
• Sylvain Pogodalla: expert for the Research Executive Agency (REA) of the EU.
• Maxime Amblard: expert for the Agence Nationale pour la Recherche (ANR), expert for the Haut Conseil de l’évaluation de la recherche et de l’enseignement supérieur (HCERES).

8.1.6. Research Administration

• Philippe de Groote:
  – Member of the bureau du comité des projets d’Inria Nancy – Grand Est.
• Sylvain Pogodalla:
  – Elected member of the comité de centre d’Inria Nancy – Grand Est, in charge of the commission IES (information et édition scientifique du centre d’Inria Nancy – Grand Est.
• Bruno Guillaume:
  – Head of the Loria department NLPKD (Natural Language Processing and Knowledge Discovery).
  – Animator of the CPER 2015-2020 project Langues, Connaissances et Humanités Numériques (Languages, Knowledge and Digital Humanities) in which ten laboratories of the Université de Lorraine participate.
  – Member of the Comipers (Inria committee for PhD and Post-doctoral selection).
• Maxime Amblard:
  – Member of conseil scientifique of Université de Lorraine, in charge of the working group on publiants.
  – Standing invitee at the pôle scientifique AM2I of Université de Lorraine.
  – Member of the Sénat Académique of Université de Lorraine.
  – Member of the progress commission of Université de Lorraine.
  – Member of the administration council of the Institut des sciences du digital, management et cognition.
  – Member of the board of the Maison des sciences de l’homme, MSH-Lorraine.
  – Head of the master in Natural Language Processing (master 1 and 2).
  – Member of the McF selection committee 4373 (section 7 and 27), Université Paris Sorbonne.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence:

Maxime Amblard, Introduction au TAL, 4h, L1, Université de Lorraine, France
Maxime Amblard, Ingénierie linguistique, 20h, L3, Université de Lorraine, France
Timothée Bernard, Algorithmique, 24h, L3, Université Paris Diderot, France.
Clement Beysson, Représentation Avancée de Données, 20h, L2, Université de Lorraine, France.
Clement Beysson, NUMOC, 40h, L1, Université de Lorraine, France.
Clement Beysson, Algorithmique et Programmation 1, 16h, L1, Université de Lorraine, France.
Clement Beysson, Introduction aux bases de données, 8h, L2, Université de Lorraine, France.
Clement Beysson, Système 2, 12h, L3, Université de Lorraine, France.
Maria Boritchev, Formalismes et représentations de raisonnements, 20h, L3, Université de Lorraine, France.
Maria Boritchev, Algorithmique 1, 22h, L1, Université de Lorraine, France.
Pierre Ludmann, Informatique 1, 20h, L3, Mines Nancy, France.
Pierre Ludmann, Informatique 2, 40h, L3, Mines Nancy, France.

Master:
Maxime Amblard, Python Programming (english), 44h, M1, Université de Lorraine, France.
Maxime Amblard, Algorithm (english), 30h, M1, Université de Lorraine, France.
Maxime Amblard, Methods for NLP (english), 36h, M1, Université de Lorraine, France.
Maxime Amblard, Remise à niveau (english), 3h, M1/M2, Université de Lorraine, France.
Maxime Amblard, Formalisms (english), 24h, M2, Université de Lorraine, France.
Maxime Amblard, Programming Project (english), 3h, M2, Université de Lorraine, France.
Timothée Bernard, Projet TAL, 18h, L3/M1, Université Paris Diderot, France.
Timothée Bernard, Sémantique computationnelle, 20h, M1, Université Paris Diderot, France.
Clement Beysson, Algorithmique et Complexité, 22h, M1, Université de Lorraine, France.
Maria Boritchev, Technology and Innovation (english), 15h, M1, Université de Lorraine, France.
Maria Boritchev, Remise à niveau (english), 3h, M1/M2, Université de Lorraine, France.
Philippe de Groote, Computational Semantics, 18h, M2, Université de Lorraine, France.
Philippe de Groote, Computational structures and logics for natural language modeling, 18h, M2, Université Paris Diderot – Paris 7, France.
Bruno Guillaume, Remise à niveau TAL (english), 6h, M1/M2, Université de Lorraine, France.
Bruno Guillaume, Written Corpora TAL (english), 30h, M1, Université de Lorraine, France.

8.2.2. Supervision

PhD in progress:
William Babonnaud, Sémantique lexicale, compositionnalité et coercition de types, since September 2018, Philippe de Groote.
Clement Beysson, Quantificateurs généralisés dynamiques pour l’analyse discursive, since September 2015, Philippe de Groote and Bruno Guillaume.
Pierre Ludmann, Construction dynamique des structures discursives, since September 2017, Philippe de Groote and Sylvain Pogodalla.

8.2.3. Juries

- Maxime Amblard was member of the jury of the master thesis of the master of NLP (12 students).
- Maxime Amblard was member of the jury PhD thesis of Mehdi Mirzapour, Modeling Preferences for Ambiguous Utterance Interpretations, September 28th, 2018, Université de Montpellier.
- Michel Musiol was member of the jury PhD thesis of Marine Labalestra November 21th, 2018, Université de Reims Champagne Ardenne
- Michel Musiol was member of the jury PhD thesis of Sarah Del Goleto, Du déficit de l’intégration contextuelle au trouble de la mentalisation dans la schizophrénie - études électrophysiologiques et comportementales de la compréhension de l’ironie., December 4th 2018, Université Paris 8

8.3. Popularization

8.3.1. Internal or External Inria Responsibilities

- Maxime Amblard is the vice head of editorial board of Interstices.info

8.3.2. Articles and Contents

- Karën Fort and Bruno Guillaume have published an article in Interstices [27] about ZombiLingo. The paper explain to non-specialists how crowdsourcing may be used to build resources that are needed for Natural Language Processing applications. They explain the aim of the project.
- Maxime Amblard has written articles for Interstices.info as Idée reçue : Les algorithmes prennent-ils des décisions ?

8.3.3. Education

- Maxime Amblard participates in the event S’orienter et se réorienter dans les sciences numériques under the patronage of FAFIEC, ONISEP Grand Est and Académie de Nancy-Metz, November 14th, 2018

8.3.4. Interventions

- Maria Boritchev participated in Samedi des curieux, March 2018, Paris France: Comment confondre de faux témoignages ?
- Maxime Amblard participates in the Ada Lovelace day in Nancy Grand-Est
- Maxime Amblard gave a talk about ethics in Artificial Intelligence at the Café des Sciences of the Cognitive science student association
- Maxime Amblard was organizer of the forum des sciences cognitives 2018 in Nancy
- Maxime Amblard has organized and supervised two panel discussions (new trend in IT recruitment and New Challenge for Neural Networks in Artificial Intelligence and NLP) for the forum des sciences cognitives 2018

8.3.5. Creation of Media or Tools for Science Outreach

- Maxime Amblard was the leader of the project Happy Family card Games, where a group of 10 researchers develop such a game for promoting computer science as a scientific field and give epistemological perspectives. More than 10 000 copies are printed and the promotion of the project will be in 2019.

9. Bibliography

Major publications by the team in recent years


**Publications of the year**

**Articles in International Peer-Reviewed Journals**


**Invited Conferences**


**International Conferences with Proceedings**


Conferences without Proceedings


Scientific Books (or Scientific Book chapters)


Books or Proceedings Editing


**Scientific Popularization**


**Other Publications**

[28] T. Bernard. *Continuations as a semantics-pragmatics interface for presuppositions*, September 2018, Sinn und Bedeutung (SuB 23), Poster, https://hal.inria.fr/hal-01931283


[31] M. Boritchev. *Approaching dialogue modeling in a dynamic framework*, Université de Lorraine (Nancy), January 2018, https://hal.inria.fr/hal-01684145

References in notes


[42] A. COPESTAKE. *Dependency and (R)MRS*, 2008, Introductory draft to RMRS and DMRS


