

RESEARCH CENTRE

Nancy - Grand Est

IN PARTNERSHIP WITH:

CNRS, Université de Lorraine

2021

ACTIVITY REPORT

Project-Team

SEMAGRAMME

Semantic Analysis of Natural Language

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

DOMAIN

Perception, Cognition and Interaction

THEME

Language, Speech and Audio

Contents

Project-Team SEMAGRAMME	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
2.1 Scientific Context	3
2.2 Syntax-Semantics Interface	4
2.3 Discourse Dynamics	5
2.4 Common Basic Resources	5
3 Research program	5
3.1 Overview	5
3.2 Formal Language Theory	6
3.3 Symbolic Logic	6
3.4 Type Theory and Typed lambda-Calculus	6
4 Application domains	6
4.1 Deep Semantic Analysis	6
4.2 Text Transformation	7
5 Highlights of the year	7
6 New software and platforms	7
6.1 New software	7
6.1.1 ACGtk	7
6.1.2 Grew	8
6.1.3 SLODiM	8
7 New results	8
7.1 Syntax-Semantics Interface	8
7.1.1 Abstract Categorical Grammars	9
7.1.2 Lexical Semantics and Linguistic Knowledge	9
7.1.3 Anaphora Resolution	10
7.1.4 Graph-based Semantics	10
7.1.5 Effects and Handlers in Natural Language	10
7.1.6 Semantics of questions	10
7.2 Discourse Dynamics	11
7.2.1 Dialogue Modeling	11
7.2.2 Pathological Discourse Modelling	11
7.3 Common Basic Resources	11
7.3.1 FR-FraCas	12
7.3.2 Universal Dependencies and Surface Syntactic Universal Dependencies	12
7.3.3 DinG	12
7.3.4 Non-projective dependencies in French corpora	13
7.3.5 Interoperable Semantic Annotation	13
7.4 Ethics and biases	13
8 Bilateral contracts and grants with industry	14
8.1 Bilateral Grants with Industry	14
8.1.1 C&S Group	14
8.1.2 Yseop	14

9 Partnerships and cooperations	14
9.1 International initiatives	14
9.1.1 Participation in other International Programs	14
9.2 European initiatives	15
9.2.1 FP7 & H2020 projects	15
9.2.2 Other european programs/initiatives	15
9.3 National initiatives	15
9.3.1 ODiM	15
9.3.2 ANR CoDeinE	16
9.3.3 GDR LIFT	16
10 Dissemination	16
10.1 Promoting scientific activities	16
10.1.1 Scientific events: selection	16
10.1.2 Journal	17
10.1.3 Invited talks	17
10.1.4 Leadership within the scientific community	18
10.1.5 Scientific expertise	18
10.1.6 Research administration	18
10.2 Teaching - Supervision - Juries	19
10.2.1 Teaching	19
10.2.2 Supervision	20
10.2.3 Juries	21
10.3 Popularization	21
10.3.1 Internal or external Inria responsibilities	21
10.3.2 Articles and contents	21
10.3.3 Education	21
10.3.4 Interventions	21
11 Scientific production	21
11.1 Major publications	21
11.2 Publications of the year	22
11.3 Other	24
11.4 Cited publications	25

Project-Team SEMAGRAMME

Creation of the Project-Team: 2013 July 01

Keywords

Computer sciences and digital sciences

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 members, visitors, external collaborators

Research Scientists

- Philippe de Groote [Team leader, Inria, Senior Researcher, HDR]
- Bruno Guillaume [Inria, Researcher]
- Sylvain Pogodalla [Inria, Researcher]

Faculty Members

- Maxime Amblard [Univ de Lorraine, Associate Professor, HDR]
- Karèn Fort [Sorbonne Université, Associate Professor]
- Jacques Jayez [École Normale Supérieure de Lyon, Emeritus, HDR]
- Michel Musiol [Univ de Lorraine, Professor, HDR]
- Guy Perrier [Univ de Lorraine, Emeritus, HDR]

Post-Doctoral Fellow

- Marc Anderson [Univ de Lorraine]

PhD Students

- William Babonnaud [Univ de Lorraine]
- Maria Boritchev [Univ de Lorraine, ATER]
- Samuel Buchel [Inria]
- Hee Soo Choi [Univ de Lorraine, from Oct 2021]
- Maxime Guillaume [Yseop, CIFRE, from July 2021]
- Laurine Jeannot [CS Group, CIFRE]
- Amandine Lecomte [Inria]
- Chuyuan Li [Univ de Lorraine]
- Pierre Ludmann [Univ de Lorraine, ATER]
- Siyana Pavlova [Univ de Lorraine]
- Valentin Richard [Univ de Lorraine, from Sep 2021]
- Priyansh Trivedi [Inria]

Technical Staff

- Pierre Lefebvre [Inria, Engineer]

Interns and Apprentices

- Sina Ahmadi [Univ de Lorraine, from Jul 2021]
- Hee Soo Choi [Univ de Lorraine, from Apr 2021 until Sep 2021]
- Amandine Decker [Univ de Lorraine, from Jun 2021 until Sep 2021]
- Fanny Ducler [Univ de Lorraine, from May 2021 until Jul 2021]
- Léo Mangel [École Normale Supérieure Paris-Saclay, from Jun 2021 until Jul 2021]
- Valentin Richard [Univ de Paris, from Feb 2021 until Jul 2021]
- Camille Saran [Univ de Lorraine, from May 2021 until Jul 2021]
- Clara Serruau [Univ de Lorraine, until Aug 2021]
- Vincent Tourneur [EPITA, Paris, Intern, from Jan 2021 until Feb 2021]
- Tuan Anh Vo [Univ de Lorraine, from Mar 2021 until Aug 2021]

Administrative Assistants

- Isabelle Herlich [Inria]
- Delphine Hubert [Univ de Lorraine]

External Collaborator

- Mathieu Constant [Univ de Lorraine]

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 may be seen as basic inference systems, while categorial grammars are based on substructural logics specified by Gentzen sequent calculi. Finally, model-theoretic grammars amount to sets of logical constraints to be satisfied.

At the level of semantics, the most common approaches derive from Montague grammars, which are based on the simply typed λ -calculus and Church’s simple theory of types. 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 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. 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 gave rise to sophisticated 'dynamic' logics. 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.

Moreover, a transversal and transdisciplinary theme has been developed in the team in the past 6 years: ethics in NLP and more generally in AI.

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. This correspondence is the keystone of the syntax-semantics interface of modern type-logical grammars. It also motivated the definition of our own Abstract Categorial Grammars [44].

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 [45], [4].

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, considering 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 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, and the related theories of functional control operators.

We have indeed successfully applied such techniques in order to model the way quantifiers in natural languages may dynamically extend their scope [43]. 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 [46] and graph rewriting [1, 2] as models of natural language syntax. This includes the development of grammars for French [55], 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 λ -calculus and simple theory of types (also known as 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, 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.

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 Categorical 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.

5 Highlights of the year

Karèn Fort has been nominated as co-chair of the ACL ethics committee for five years, with Min Yen Kan (Singapore Univ.) and Yulia Tsvetkov (Univ. of Washington).

6 New software and platforms

6.1 New software

6.1.1 ACGtk

Name: Abstract Categorical Grammar Development Toolkit

Keywords: Natural language processing, NLP, Syntactic analysis, Semantics

Scientific Description: Abstract Categorical Grammars (ACG) are a grammatical formalism in which grammars are based on typed lambda-calculus. A grammar generates two 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 two software tools to develop and to use ACGs: `acgc`, which is a grammar compiler, and `acg`, which is an interpreter of a command language that allows one, in particular, to parse and realize terms.

Functional Description: ACGtk provides softwares for developing and using Abstract Categorical Grammars (ACG).

Release Contributions: This version fixes some bugs and adds some commands to the scripting language. Some internal modifications also prepare ACGtk to add extensions of ACGs such as weighting.

News of the Year: In addition to modifications to maintain the code (bug fixes, improvement of error messages, documentation, etc.), new functionalities have been added. They mainly consist in adding magic rewriting, automatic generation of abstract terms with a given type, and sorting parses when several solutions exist. Modifications to prepare extensions of ACGs (such as weighting) have also been added.

URL: <http://acg.loria.fr/>

Publications: [hal-01242154](#), [hal-01328702](#), [tel-01412765](#), [inria-00112956](#), [inria-00100529](#)

Contact: Sylvain Pogodalla

Participants: Philippe De Groote, Pierre Ludmann, Jiri Marsik, Sylvain Pogodalla, Vincent Tourneur

6.1.2 Grew

Name: Graph Rewriting

Keywords: Semantics, Syntactic analysis, NLP, 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 2021, a few Grew software versions were released (1.8 is the latest one). The main changes concern the commands available for graph modification (a few commands were added).

The Grew-match tool (<http://match.grew.fr>) is an online service available where a user can query different corpora with graph matching requests. All UD corpora (217 in 122 different languages in v2.9) are available and data from several other projects can also be queried. In 2021, 130,000 requests were received on the Grew-match server.

A new web interface (available on <http://transform.grew.fr>) was developed to replace the old GTK based one. With this interface, it is possible to visualise the steps of the rewriting process and it can be used to debug rules, for demonstration or in a tutorial session for learning Grew.

In order to promote the tool and increase its usage, a demo was presented at the EACL 2021 conference (<https://hal.inria.fr/hal-03177701>) and two tutorials were organised during the TALN conference (<https://talnrecital2021.inria.fr/grew-match/> and <https://talnrecital2021.inria.fr/grew/>).

The development of Arborator-Grew (<https://arborator.github.io/>) is still active. The platform will be used in the new Autogramm ANR project.

URL: <https://grew.fr/>

Publications: [hal-03177701](https://hal.inria.fr/hal-03177701), [hal-01930591](https://hal.inria.fr/hal-01930591), [hal-01814386](https://hal.inria.fr/hal-01814386), [hal-03021720](https://hal.inria.fr/hal-03021720)

Contact: Bruno Guillaume

Participants: Bruno Guillaume, Guy Perrier, Guillaume Bonfante

6.1.3 SLODiM

Name: SLODiM

Keywords: NLP, Discourse, Dialogue, French

Functional Description: SLODiM is a software package for the analysis of oral French. It is more particularly developed to allow the analysis of interviews with clinicians in order to identify language behaviours characteristic of mental pathologies.

Release Contributions: first complete version

URL: <https://team.inria.fr/semagramme/odim/>

Contact: Maxime Amblard

Partners: Loria, Université de Lorraine, CNRS

7 New results

7.1 Syntax-Semantics Interface

Participants: Maxime Amblard, William Babonnaud, Philippe de Groote, Bruno Guillaume, Maxime Guillaume, Pierre Ludmann, Sylvain Pogodalla, Siyana Pavlova, Priyansh Trivedi.

7.1.1 Abstract Categorical Grammars

Feature Structure ACG has proven to be a powerful framework with well-defined theoretical properties. It was however lacking a facility which is useful and widely used for grammar engineering: feature structures. The latter are often used to express in a concise way some combinatorial properties related to morphosyntactic properties of expressions, for instance subject-verb agreement.

We worked on extending the ACG type system to provide such feature structures. This extension relies on a restricted addition of product (records) and dependent types. We also considered the reduction of grammars using this extension to Datalog programs (which is used to implement ACG parsing in ACGtk, see Sec. 6).

Multityped ACG (mACG) Symbolic parsing with large coverage grammars usually leads to combinatorial explosion of syntactic ambiguities (a single expression has many syntactic analysis). A widespread method to tackle this issue is to use statistics and probabilities, leading for instance to probabilistic Context Free Grammars (pCFGs) and probabilistic Tree Adjoining Grammars (pTAGs). An important goal is then to also extend ACGs with probabilities or weights.

Yet, ACGs come with features that make this extension non-trivial. In particular, ACGs can be composed by making the parse structures of a grammar the surface structures of another ACG. The resulting composition is a full-flavored ACG.

When aiming at integrating distinction between structures (that will eventually be associated with preferences, for instance expressed by weights) at each level of an ACG, we expect to have a well-behaved composition of these new grammatical objects. In particular the result should indeed be one of these new grammatical objects. To this end, we introduced multityped ACGs (mACGs). Multityped ACG are the underlying discrete mathematical structures that will support weighting extension. We also showed that a suitable notion of composition can be defined for multityped ACGs as well.

7.1.2 Lexical Semantics and Linguistic Knowledge

Lexical Coercion and Types 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 λ -calculus necessitates a calculus with notions of subtyping and type coercion.

In this context, William Babonnaud has studied possible solutions to the acknowledged incompatibility between subtyping and Montague-style λ -calculus. He has shown that choosing a topos as categorical semantics for such λ -calculi enables one to obtain the same properties as Modern Type Theories [28]. Moreover, he has developed a predicate calculus specifically designed for Montague semantics, which features covariant subtyping in a type safe way and relies on the powerfulness of toposes to ensure all the necessary properties [14].

Lexical Structure of Word Embeddings Word embeddings rely on the distributional hypothesis[47]: the meaning of a word is provided by the linguistic contexts in which it occurs and semantically related words should be represented by similar vectors. However, the exact nature of the semantic relatedness that word and sentence embeddings encode remains unclear. Context similarity mixes distinct relations together (e.g., synonymy, hyponymy, etc.)[54, 48]) and it depends on many heuristics and design choices[52] such as the choice of the similarity measure, the context size, the type of contexts[58, 53]. In order to explore the lexical structures learned by these models, we have related them to lexical knowledge, in particular as described in formal theories of lexical relatedness such as the theory of explanatory combinatorial lexicology, the lexicographical part of the Meaning-Text Theory[51], which provides a fine-grained characterization the lexical structure. Our first results show there is no systematic association between vector similarity and specific lexical relation.

7.1.3 Anaphora Resolution

Priyansh Trivedi has investigated methods for resolving bridging anaphora—an understudied subset of anaphora resolution, concerned with non-identity links between the anaphor and their antecedents. As a first step, numerous baselines were established, based on non-contextual embedding based approaches. These include the use of pretrained word embeddings, and of word net embeddings computed using relational graph convolution network[57] based models over the Wordnet graph. Forays into probing the self-attention matrices of common pretrained language models[42, 50, 56] provided a surprisingly strong baseline for the task as well, demonstrating that these language models are aware of the phenomenon of bridging anaphors, and the contextual representations for anaphors are influenced by antecedents, which in some cases may appear multiple sentences before the anaphor.

In [24], it was demonstrated that the neural models proposed for coreference resolution[49] (a sister task to bridging anaphora resolution, which has been historically extensively studied, and for which large gold-standard datasets are available) are ineffective in resolving bridging anaphors. Further analysis provides evidence to the hypothesis that a major hinderance in their use is the inability of these models to identify anaphoric noun phrases, possibly owing to much smaller training data.

7.1.4 Graph-based Semantics

Siyana Pavlova started her PhD in November 2020. She began to study and compare different existing semantic graph-based annotation frameworks (AMR, UCCA and DRS). The goal is to determine how these frameworks are compatible and if they encode the same level of semantic information. Clara Serruau is working on the same topic with a focus of DRS annotation available in the [Parallel Meaning Bank](#).

7.1.5 Effects and Handlers in Natural Language

We publish the long version article of our work on Effects and Handlers [7].

In formal semantics, logical meanings are assigned to natural language utterances with compositionality. The functions are often formalized using the λ -calculus. meanings are derived by processes that no longer correspond to pure mathematical functions but rather to context-sensitive procedures, much like the functions of a programming language that manipulate their context with side effects. We claim that by looking at these theories as theories of side effects, we can reuse results from programming language research to combine them.

Our work extends the λ -calculus with a monad of computations. The monad implements effects and handlers, a recent technique in the study of programming language side effects. We have proven some of the fundamental properties of our extended calculus: subject reduction, confluence and termination. We have then demonstrated how to use our calculus to implement treatments of several linguistic phenomena: deixis, quantification, conventional implicature, anaphora and presupposition.

7.1.6 Semantics of questions

Natural language statements are not only composed of declarative sentences but also of interrogative ones. Moreover, sentences cannot be categorized into purely declarative or purely interrogative sentences. Typically, a declarative statement may contain an indirect interrogative clause:

I do not know where is Mary.

Conversely, a direct interrogative clause may contain a declarative subordinate:

Do you know that Mary is here ?

This interaction between declarative and interrogative clauses is particularly present in dialogues, where the logical notion of *answerhood* is as significant as the one of *inference*.

In order to tackle this problematic from a formal standpoint, we investigated the properties and possible uses of inquisitive semantics, which is a formal semantic theory based on a logic that provides a uniform treatment of both declarative and interrogative expressions. In that vein, we have shown how a semantic framework developed from inquisitive logic and neo-Davidsonian event semantics gives a compositional account of the semantics of wh-questions [11]. We have also studied the connection

that exists between intensional and inquisitive models. In particular, we have shown how intensional semantics can be embedded into inquisitive semantics in a conservative way [38, 19].

7.2 Discourse Dynamics

Participants: Maxime Amblard, Maria Boritchev, Philippe de Groote, Bruno Guillaume, Pierre Ludmann, Michel Musiol.

7.2.1 Dialogue Modeling

Maxime Amblard and Maria Boritchev pursue the development of a dynamic model of dialogue for questions and answers.

Formal studies of discourse raise numerous interrogations on the nature and the definition of the way consecutive sentences combine with one another. The shift from discourse to dialogue brings forward even more specific issues. Dialogue acts are more intrinsically connected because of the dynamicity of the interaction.

They present in [32] a formal approach to compositional processing of questions and answers in the Schizophrenia and Language, Analysis and Modeling corpus (SLAM). They address dialogue lexicality issues starting from the formal definitions of so-called Düsseldorf Frame Semantics given. 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).

Maxime Amblard continues a common work with Chloé Braud on Formal and Statistical Modelling of dialogues. In the PhD thesis of Chuyuan Li, they design tools to automatically retrieve characteristic features of dialogues. They present results in [23]. To extend this work, Chuyuan Li is currently on sabbatical leave in University British Columbia, working in the Gioseppe Carenini group.

7.2.2 Pathological Discourse Modelling

Michel Musiol has obtained a full-time delegation in the Sémagramme team. This proximity makes it possible to set up a more active collaboration 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.

Maxime Amblard and Michel Musiol pursue the Inria Exploratory Action on this issues ODiM with the constitution of the resource and extension of the tool SLoDIM. The theoretical work focused on the formal definition of transactions in dialogue. To do so, with Samuel Buchel and Amandine Lecomte, they introduce a dynamic definition of back channel words which are used to classify the dialogue units. With Manuel Rebuschi they publish a book which summarizes the (In)Coherence of Discourse workshops [29], with a survey on Schizophrenia analysis [35]. Moreover they finish another survey on the specificities of the collection of data with patients suffering from mental diseases [30].

Maxime Amblard supervised the internship of Amandine Decker. During her previous internship they highlighted a more thematic reading of the dialogues that allowed us to understand the overall structure of the conversation. This work presents several representations that attempt to model the interaction between themes without completely overwriting the interaction within themes. Topoi annotations help to better define the sub-themes. They focus on pathological data.

7.3 Common Basic Resources

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

7.3.1 FR-FraCas

Maxime Amblard, Clement Beysson, Philippe de Groote, Bruno Guillaume and Sylvain Pogodalla carried on the development of FR-FraCas, a French version of the FraCas test suite [41] which is an inference test suite, in English, for evaluating the inferential competence of different NLP systems and semantic theories. There currently exists a multilingual version of the resource for Farsi, German, Greek, and Mandarin. Sémagramme completed the first translation into French of the test suite. The latter has been [publicly released](#).

During his internship, Leo Mangel builds the semantic annotation of a subset of the French version using Neo-Davidsonian event semantics and formalise it in the ACGtk framework.

7.3.2 Universal Dependencies and Surface Syntactic Universal Dependencies

The Universal Dependencies project (UD) aims at building a syntactic dependency scheme which allows for similar analyses for several different languages. Bruno Guillaume and Guy Perrier are active in the UD community, and participate to the development and the improvement of the French data in this international initiative.

During 2021, they continue working, in collaboration with Sylvain Kahane, Kim Gerdes and their teams on the promotion of the Surface Syntactic Universal Dependencies (SUD) framework. SUD is an annotation scheme for syntactic dependency treebanks, that is almost isomorphic to UD (Universal Dependencies). Contrary to UD, it is based on syntactic criteria (favoring functional heads) and the relations are defined on distributional and functional bases. In [18], they bring to the fore some advantages to first develop a new treebank in Surface-Syntactic Universal Dependencies (SUD) annotation scheme, even if the goal is to obtain a UD treebank. Theoretical benefits of SUD are presented, as well as UD-compatible SUD innovations. The twoway UD, SUD conversion is explained, as well as the possibility to customize the conversion for a given language. The paper concludes by a practical guide for the development of a SUD treebank.

A [website](#) was built to present the framework (guidelines, data). The Sémagramme teams is notably in charge of the GREW-based tools for conversion with the UD framework. These conversion tools are used both to produce the UD data for a few SUD native treebanks and of to produce the SUD version of all UD available data.

A new corpus was added to the SUD project in Beja, a Cushitic language spoken in Sudan. This is the first treebank for Beja in the UD/SUD project. It has been built from the conversion and enhancement of an Interlinear Glossed Text (IGT). The paper [22] presents this corpus and describes the choice to use a morph-based annotation and its consequences; the processing chain from an IGT to a morph-based dependency treebank and a word-based treebank; and several interesting constructions in Beja.

Following her internship in the summer of 2020, Hee-Soo Choi continued working on linguistic typology based on UD annotated data during her M2 research *mémoire*, with Karën Fort and Bruno Guillaume. She used GREW to enrich the UD annotations and studied the respective word order of verbs with their subjects and objects on 74 languages and compared with other linguistic works. The work was published at RANLP 2021 [17]. She then focused on four Greenberg's universal, providing results which constitute new typological information based on large amounts of data that can fill in gaps in the existing databases. The corpus-based approach allowed to evaluate the consistency between corpora of the same language and showed a great variation according to the corpus types: oral language, written language in newspapers, tweets, poetry, novels, grammars, etc. This last part of the work was published at the Syntax Fest (the event will happen in 2022 but the proceedings are dated in 2021 [16]).

Bruno Guillaume and Guy Perrier participated in the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (EUD) [21]. EUD is an enrichment of UD annotations with deep syntactic relations (subjects of infinitives, co-references in relative clause constructions, dependency propagation on conjuncts of coordinations ...). The originality of their approach is to use graph rewriting [20] to enrich UD annotations of corpora into EUD annotations.

7.3.3 DinG

Maria Boritchev and Maxime Amblard finished the development of DinG. DinG is a transcription corpus of oral French, based on multilogues between 3 to 4 people playing the board game *Catan*. It was created to

study human dialogue based on attested, spontaneous and unconstrained, without personal information, oral data in French. It allows the study of long interactions, going beyond informative exchanges. The recordings were then processed to produce a transcribed version of the games. To this end, a guide was developed and transcribers were recruited. Each recording was treated individually: segmentation into turns, transcription according to the guide, verification by a super annotator. Thus, the resource produced is of very good quality. It contains 14 hours of recording for 22k speaking turns and 115k words [26].

7.3.4 Non-projective dependencies in French corpora

Guy Perrier studied the non-projective syntactic dependencies in French corpora annotated according to two dependency syntax formats: *Universal Dependencies* and *Surface Syntactic Universal Dependencies* [8]. This study highlights the very local character of the configurations of two crossing dependencies, which is one of the ways to characterize non-projectivity. It also highlights four main linguistic sources of non-projectivity: clitic climbing, deep extraction, dependency length minimization and pairs of distant dependent words.

7.3.5 Interoperable Semantic Annotation

We participated in the challenge on quantification annotation of the [ISO Workshop on Interoperable Semantic Annotation](#), in order to assess the quantML annotation framework. In addition to the annotations, we also provided several tools, based on the software developed by the team, to make the task easier and improve the reliability of the annotation [13].

7.4 Ethics and biases

Participants: Karèn Fort, Maxime Amblard.

Maxime Amblard develop with the Orpailleur Team a work on biases. Unintended biases in machine learning (ML) models are among the major concerns that must be addressed to maintain public trust in ML. They address process fairness of ML models that consists in reducing the dependence of models on sensitive features, without compromising their performance. They revisit the framework FIXOUT that is inspired in the approach “fairness through unawareness” to build fairer models. They introduce several improvements such as automating the choice of FIXOUT’s parameters. Also, FIXOUT was originally proposed to improve fairness of ML models on tabular data. They also demonstrate the feasibility of FIXOUT’s workflow for models on textual data. They present several experimental results that illustrate the fact that FIXOUT improves process fairness on different classification settings [10].

Karèn Fort has been a member of the Sorbonne IRB (Comité d’Ethique de la Recherche) since 2019. The IRB members wrote a book chapter on their work during the COVID pandemic [33], in which they describe their work and their questioning about their action w.r.t the medical protocols in particular.

Karèn Fort originated a working group at LORIA for AI ethics (ethics@loria), involving researchers from various teams, including: Maxime Amblard, Marc Anderson, Armelle Brun (BIRD), Mathieu d’Aquin (Orpailleur), Christophe Cerisara (Synalp), Anne Bonneau (Multispeech), Slim Ouni (Multispeech) and Abdessamad Imine (Pesto). Aurore Coince helps managing the group. Ethics@loria met two times in 2021 and started working on a project of ethics seminar towards young researchers.

Karèn Fort led the creation of a French corpus of stereotyped/anti-stereotyped pairs of sentences (women cannot drive / men cannot drive) with Aurélie Névéol (LISN-CNRS), Yoann Dupont (Sorbonne Université) and Julien Bezançon (M2 student at Sorbonne). The corpus was created by translating and adapting an American corpus (CrowsPairs), then adding more specific French stereotypes (for example against Gypsies) obtained via a citizen science platform, LanguageArc. It was then used to evaluate the main masked models for French (including CamemBERT and FlauBERT). A paper on the subject has been submitted to ACL Rolling Review for ACL in October 2021.

Karèn Fort worked with Yves Lepage (Waseda University, Japan), Gaël Lejeune (Sorbonne Université) and Fanny Ducel (M1 student at Sorbonne Université) on the evaluation of the application of the Bender’s rule in NLP research papers. The Bender’s rule state that NLP researchers should name the language

they work on, but it is often not applied, especially when the language dealt with is English, giving the false impression that the research can apply to any language or that English is universal. A paper on a comparison between LREC and ACL proceedings has been submitted at LREC 2022.

In the context of the CODEINE ANR project and more specifically of Nicolas Hiebel's PhD thesis, Karën Fort worked with Aurélie Névéol (LISN-CNRS) and Olivier Ferret (CEA) on the creation of a sentence semantic similarity corpus for French on the clinical domain. A paper on the subject has been submitted at LREC 2022.

The collaboration between Karën Fort and Marc Anderson gave birth to two journal papers submissions. One is about ethics by design for real in the AI-Proficient project. The other criticizes the "human in" (the loop/command, etc.) terminology and proposes a new grid of analysis of the interaction between the human and the system. Both were conditionally accepted and the second version is to be submitted by end of January and mid-February 2022.

8 Bilateral contracts and grants with industry

8.1 Bilateral Grants with Industry

8.1.1 C&S Group

Participants: Maxime Amblard, Philippe de Groote, Laurine Jeannot.

The Sémagramme team has set up a Cifre thesis contract with C&S Group on the use of semantic and discourse representation and parser in order to automatically analyse the specification system. The Thesis started in 2021.

8.1.2 Yseop

Participants: Philippe de Groote, Maxime Guillaume.

The Sémagramme team has set up a Cifre thesis contract with Yseop on ACG extensions and use in an industrial environment.

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Participation in other International Programs

Participants: Maxime Amblard, Philippede Groote, Sylvain Pogodalla.

Sémagramme is part of the Inria-DFKI project **IMPRESS**. Its goals are to investigate the integration of semantic knowledge into embeddings and its impact on selected downstream tasks, to extend this approach to multimodal and mildly multilingual settings, and to develop open source software and lexical resources, focusing on video activity recognition as a practical testbed. The project is lead by Pascal Denis (**MAGNET**, Inria Lille-Europe), and **Multispeech** (Inria Nancy-Grand Est) member of this project.

Sémagramme is part of the Inria-DFKI project **MePheSTO**. It is an interdisciplinary research project that envisions a scientifically sound methodology based on artificial intelligence methods for the identification and classification of objective, and thus measurable, digital phenotypes of psychiatric disorders. MePheSTO has a solid foundation of clinically motivated scenarios and use-cases synthesized jointly with

clinical partners. Important to MePheSTO is the creation of a multimodal corpus including speech, video, and biosensors of social patient-clinician interactions, which serves as the basis for deriving methods, models and knowledge. Important project outcomes include technical tools and organizational methods for the management of medical data that implement both ELSI and GDPR requirements, demonstration scenarios covering patients' journeys including early detection, diagnosis support, relapse prediction, therapy support. The project is co-lead by François Bremond (Star, Inria Sophia Antipolis).

9.2 European initiatives

9.2.1 FP7 & H2020 projects

AI Proficient Sémagramme is part of the **AI Proficient ICT-38-2020 - Artificial intelligence for manufacturing project** (, coordinated by the CRAN laboratory of Université de Lorraine. By combining human knowledge with AI capabilities, the EU-funded AI-PROFICIENT project will develop proactive control strategies to improve manufacturing processes in terms of production efficiency, quality and maintenance. The overall goal is to increase the positive impact of AI technology on the manufacturing process as a whole, while keeping the human in a central position, assuming supervisory (human-on-the-loop) and executive (human-in-command) roles. By identifying the effective means for human-machine interaction, the project will assist Europe's manufacturing and process industry to improve production planning and execution.

Karèn Fort is the Project Ethics Officer and as such is responsible for the ethical dimensions of the project. Marc Anderson was hired as a post-doc researcher on the project is carrying out research on AI Ethics by Design in the Sémagramme team.

9.2.2 Other european programs/initiatives

enetCollect COST action Sémagramme is part of the **European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect) COST action**, which has been prolonged until September 2021. The action aims at unlocking a crowdsourcing potential available for all languages and at triggering an innovation breakthrough for the production of language learning material, such as lesson or exercise content, and language-related datasets such as, among others, NLP language resources. Karèn Fort is Management Committee member for France and was leading the Working Group 5 of the action (Application-oriented specifications for an ethical, legal and profitable solution) but she resigned in 2020 due to a potential conflict of interest with the AI Proficient external ethical advisor, Katerina Zdravkova (University of Skopje, Macedonia), who was vice leader of WG5.

LITHME COST action Sémagramme is part of the **Language In The Human-Machine Era (LITHME) COST action**. LITHME aims at shining a light on the ethical implications of emerging language technologies. Karèn Fort is Management Committee member for France.

9.3 National initiatives

9.3.1 ODiM

Outils informatisés d'aide au Diagnostic des Maladies mentales

2019 - 2022

Coordinator: Maxime Amblard

Participants: Maxime Amblard, Vincent-Thomas Barrouillet, Samuel Buchel, Amandine Lecomte, Chuyuan Li, Michel Musiol

Abstract:

ODiM is an interdisciplinary project, at the interface of psychiatry-psychopathology, linguistics, formal semantics and digital sciences. It aims to replace the paradigm of Language and Thought Disorders (LTD) as used in the Mental Health sector with a semantic-formal and cognitive model of Discourse Disorders (DD). These disorders are translated into pathognomonic signs, making them complementary diagnostic tools as well as screening for vulnerable people before the psychosis's trigger. The project has three main components.

The work is based on real data from interviews with patients with schizophrenia. A data collection phase in partner hospitals and with a control group, consisting of interviews and neuro-cognitive tests, is therefore necessary.

The data collection will allow the development of the theoretical model, both in psycholinguistic and semantic formalization for the identification of diagnostic signs. The success of such a project requires the extension of the analysis methodology in order to increase the model's ability to identify sequences with symptomatic discontinuities.

If the general objective of the project is to propose a methodological framework for defining and understanding diagnostic clues associated with psychosis, we also wish to equip these approaches by developing software to automatically identify these clues, both in terms of discourse and language behaviour.

9.3.2 ANR CoDeinE

The ANR project CoDeinE (artificial text CORpus DEsIgNed Ethically automatic synthesis of clinical documents) is coordinated by Aurélie Névéal (Limsi). Sémagramme is one the partner: Karèn Fort (local coordinator) and Bruno Guillaume are involved in the project.

Within the framework of this project, Karèn Fort is co-advisor of the PhD thesis of Nicolas Hiebel, with Aurélie Névéal and Olivier Ferret (CEA). The objective of the work is to develop synthetic clinical corpora to bypass the ethical issues generated by the use of real data. As resources for semantic similarity in the clinical domain in French are scarce, the efforts were focused on that point during the first months of the thesis (and during the Master 2 *mémoire* of N. Hiebel).

9.3.3 GDR LIFT

Sémagramme participates in GDR LIFT (Linguistique Informatique, Formelle et de Terrain). Karèn Fort is co-chair (with G. Wisniewski) of the axis 2: *Linguistique et évaluation des systèmes de traitement automatique des langues*. She is also the co-organizer of the LIFT summer school on annotation, planned in 2022.

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: selection

Chair of conference ethics committees

- Karèn Fort co-chaired with Emily Bender (Univ. of Washington) the NAACL 2021 ethics committee.

Member of the conference program committees

- Maxime Amblard: Logical Aspects of Computational Linguistics (LACL 2021).
- Philippe de Groote: 17th Meeting on the Mathematics of Language (MOL 2021); Logical Aspects of Computational Linguistics (LACL 2021); International Conference on Computational Semantics (IWCS 2021); Conference on Reasoning in Interaction (ReInAct 2021); Fifth meeting of the Society for Computation in Linguistics (SCiL 2022).
- Sylvain Pogodalla: Logical Aspects of Computational Linguistics (LACL 2021).

Reviewer

- Maxime Amblard: ACL Rolling Review, [Association for Computational Linguistic](#), NAACL 2021 [2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics](#), [eacl 2021](#) [The 16th Conference of the European Chapter of the Association for Computational Linguistics](#), EMNLP 2021 [The 2021 Conference on Empirical Methods in Natural Language Processing](#).

- Marc Anderson: Special track APMS 2021 - "Human-centered AI in Smart Manufacturing for the Operator 4.0"
- Karèn Fort: ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT 2021), ACL Tutorials 2021, Journée du GDR LIFT, 2021
- Chuyuan Li: EMNLP 2021 [The 2021 Conference on Empirical Methods in Natural Language Processing](#)
- Sylvain Pogodalla: [ACL Rolling Review](#), 60th Annual Meeting of the Association for Computational Linguistics ([ACL 2022](#)), 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics ([NAACL 2021](#)), 2021 Conference on Empirical Methods in Natural Language Processing ([EMNLP 2021](#)), Conférence Nationale en Intelligence Artificielle collection ([CNIA 2021](#)), Workshop on Computing Semantics with Types, Frames and Related Structures 2021 ([CSTFRS 2021](#)), Logical Aspects of Computational Linguistics ([LACL 2021](#)), Traitement Automatique des Langues Naturelles 2021 ([TALN 2021](#)).

10.1.2 Journal

Member of the editorial boards

- Maxime Amblard: Member of the editorial board of the journal Traitement Automatique des Langues, in charge of the pdf pipeline.
- Philippe de Groote: area editor of the FoLLI-LNCS series.
- Sylvain Pogodalla: Member of the editorial board of the journal Traitement Automatique des Langues, in charge of the *Résumés de thèses* section.
- Michel Musiol: Psychological and educational sciences (Université d'ElOued Ed).

Reviewer - reviewing activities

- Maxime Amblard: Journal of Logic, Language and Information, PlosOne, Cognitive Systems Research, Academia Letters
- Marc Anderson: Transactions of the Charles S. Peirce Society
- Marc Anderson: Book Manuscript Reviewer for State University of New York Press, NY.
- Philippe de Groote: Journal of Functional Programming; Mathematical Structures in Computer Science; Studia Logica.

10.1.3 Invited talks

- Marc Anderson: "Human Where? A New Scale Defining Human Involvement in Technology Communities from an Ethical Standpoint" AI-MAN (ICT-38) Projects Cluster
- Marc Anderson: "Ethical and Legal Issues of Artificial Intelligence In Manufacturing" Workshops Series Online (25 November)
- Marc Anderson: "*Commentaires sur la Charte OLKi*," *Intelligence Artificielle et Vie Privée* (in the framework of the project Open Language and Knowledge for Citizens (OLKi)), France, June 10th, 2021.
- Karèn Fort: "Ethics in computational linguistics". DiLCo international research networking project workshop on "Ethics in digital language research". online/Hamburg, Germany, Oct. 8th, 2021.
- Karèn Fort: "Ethics and AI (a view from NLP)". Invited presentation at the seminar of the SAILS interdisciplinary program. online/Leiden, Netherlands. Sept. 27th, 2021

10.1.4 Leadership within the scientific community

- Maxime Amblard: Management Committee of the OLKI project (Lorraine Université d'Excellence project - PIA), co-leader of the workpackage 2 on NLP activities.
- Karën Fort: has been nominated as co-chair of the [ACL Ethics committee](#), with Min Yen Kan (Singapore Univ.) and Yulia Tsvetkov (Univ. of Washington).

10.1.5 Scientific expertise

- Maxime Amblard is member of the scientific board of the INJS - *Institut National des Jeunes Sourds*.
- Karën Fort was external evaluator for the special CFP Digital Humanism of the Vienna Science and Technology Fund and for a Discovery grant of the Natural Sciences and Engineering Research Council of Canada (NSERC)
- Philippe de Groote acted as an expert for the State Education Development Agency (SEDA) of Latvia.

10.1.6 Research administration

- Maxime Amblard
 - Member of *conseil scientifique of Université de Lorraine*
 - Standing invitee at the *pôle scientifique AM2I of Université de Lorraine*
 - Member of the *Sénat Académique of Université de Lorraine*
 - Member of the grade promotion committee of *Université de Lorraine*
 - Member of the administration council of the *Institut des sciences du digital, management et cognition*
 - Head of the master in Natural Language Processing (master 1 and 2) [12]. The supervision of different students project lead to [15]
- Karën Fort
 - Member of the CNU 27
 - Member of the IRB (CER) of Sorbonne Université
 - Member of the conseil of UFR SISH (*Sociologie et Informatique pour les Sciences Humaines*) at Sorbonne Université since 2018.
 - Member of the selection committee (*Comité de sélection*) for MCF section 27 at the UFR STN and *laboratoire Programmation et Informatique Fondamentale de l'Université Paris 8*.
- Philippe de Groote
 - Member the *bureau du comité des projets d'Inria Nancy – Grand Est*.
- Bruno Guillaume
 - Head of the Natural Language Processing and Knowledge Discovery department of the LORIA laboratory
 - Manager (with Alain Polguère) of the CPER (Contrat de Plan État-Région) "Langues, Connaissances et Humanités Numériques".
- Sylvain Pogodalla
 - Elected member of the comité de centre Inria Nancy – Grand Est,
 - In charge of the local commission IES (information et édition scientifique) of the Inria Nancy – Grand Est and LORIA.
 - Member of the national commission IES of Inria.

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

- Licence:
 - Valentin Richard, Langages de Scripts, 50h, *Université de Lorraine*, France.
 - Maxime Amblard, AI Introduction, 15h, L1, *Université de Lorraine*, France.
 - Maxime Amblard, Maria Boritchev and Chuyuan Li NLP for beginners, 20h, L2, *Université de Lorraine*, France.
 - Maxime Amblard, Maria Boritchev and Chuyuan Li Linguistic engineering, 20h, L3, *Université de Lorraine*, France.
 - Maria Boritchev, Formalisms and reasoning representations , 20h, L3, *Université de Lorraine*, France.
 - Maria Boritchev, Algorithmic 1, 26h, L1, *Université de Lorraine*, France.
 - Maria Boritchev, Tools and digital culture 1, 20h, L1, *Université de Lorraine*, France.
 - Hee-Soo Choi, Corpus processing (introduction to Python), 24h, *Université de Lorraine*, France
 - Karèn Fort, Relational databases, 55h, L3, *Sorbonne Université*, France.
 - Pierre Ludmann, Informatics 2, 20h, *Mines de Nancy*, France.
- Master:
 - Maxime Amblard and Siyana Pavlova, Python Programming, 30h, M1 NLP, *Université de Lorraine*, France.
 - Maria Boritchev MAmblard and Siyana Pavlova, Methods for NLP, 36h, M1 NLP, *Université de Lorraine*, France.
 - Maxime Amblard, NLP project, 20h, M1 NLP, *Université de Lorraine*, France
 - Maxime Amblard and William Babonnaud, Formalisms and Syntax, 24h, M2 NLP, *Université de Lorraine*, France.
 - Maxime Amblard and Valentin Richard, Discourse and Dialogue, 18h, M2 NLP, *Université de Lorraine*, France.
 - Maxime Amblard, software project, 10h, M2 NLP, *Université de Lorraine*, France
 - Maria Boritchev, Calculability and Complexity, 10h, M1, *Université de Lorraine*, France
 - Philippe de Groote, Formal Logic, 22h, M1 NLP, *Université de Lorraine*, France.
 - Philippe de Groote, Formal languages, 22h, M1 NLP, *Université de Lorraine*, France.
 - Philippe de Groote, Computational Semantics, 18h, M2 NLP, *Université de Lorraine*, France.
 - Philippe de Groote, Computational structures and logics for natural language modeling, 18h, M2 NLP, *Université Paris Diderot – Paris 7*, France.
 - Karèn Fort, Data ethics (English), 3h, M2 NLP and cog. Sces (IDMC), *Université de Lorraine*, France.
 - Karèn Fort, Formal Grammar, 39h, M1, *Sorbonne Université*, France.
 - Karèn Fort, Corpora, resources and tools for linguistics, 39h, M1, *Sorbonne Université*, France.
 - Karèn Fort, NLP platforms, 15h, M2, *Sorbonne Université*, France.
 - Karèn Fort, Collaborative annotation for NLP, 30h, M2, *Sorbonne Université*, France.
 - Karèn Fort, Crowdsourcing for NLP, 3h, M2, *Nanterre Université*, France.
 - Bruno Guillaume, Written Corpora TAL (English), 44h, M1 NLP, *Université de Lorraine*, France.
 - Chuyuan Li, Introduction to G5K (English), 4h, M1 NLP, *Université de Lorraine*, France.
- Doctorate:

- Karèn Fort, Scientific integrity, 6h, *École doctorale 5, Faculté des lettres, Sorbonne Université*.
- Karèn Fort, Ethics and NLP, 3h, *École d'été* of the GDR TAL (ETAL), June 2021, Lannion, France.
- Tutorials:
 - Karèn Fort: TALN 2021 tutorial on Reviewing Natural Language Processing Research co-organised with Kevin Cohen, Margot Mieskes, Aurélie Névéol et Anna Rogers.
 - Karèn Fort: Ethics and NLP : an (almost) tutorial, in the Tutorial on Ethics in AI and human-machine interaction, at the International Conference on Multimodal Interaction (ICMI) 2021, with Raja Chatila, Monique Morrow, Johanna Seibt, Mohamed Chetouani and David Cohen. October 2021.
 - Karèn Fort: Reviewing Natural Language Processing Research (Introductory) at EACL (rank A) co-organised with Kevin Cohen, Margot Mieskes, Aurélie Névéol and Anna Rogers. April 2021.
- International Summer School:
 - Philippe de Groote with Yoad Winter, Introduction to natural language formal semantics, at the 32nd European Summer School in Logic, Language and Information (ESSLLI 2021).

10.2.2 Supervision

- PhD:
 - Maria Boritchev, *Modélisation dynamique des dialogues*, Novembre 22th 2021, supervisors: Maxime Amblard and Philippe de Groote, Reviewers: Farah Benamara Zitoune (*IRIT, Université Paul Sabatier*, Toulouse, France) and Jonathan Ginzburg (*Université Paris Diderot – Paris 7*, France), Jury: Ellen Breitholtz (Göteborg University, Sweden), Iris Eshkol-Taravella, (*Université Paris Nanterre*, France), president: Miguel Couceiro (*Université de Lorraine*, France).
- PhD in progress:
 - William Babonnaud, *Lexical Semantics, Compositionality and Type Coercion*, since September 2018, Philippe de Groote.
 - Samuel Buchel, *Linguistic, Semantic and Cognitive Modelling of Dialogical Incongruities and Discontinuities in The Interaction with The Schizophrenic Patients*, since December 2019, Maxime Amblard and Michel Musiol.
 - Hee-Soo Choi, *Lier des ressources lexicales du français en vue d'une interopérabilité entre niveaux linguistiques*. Funding: Ecole Doctorale Sociétés, Langages, Temps, Connaissances, Univ. de Lorraine, since Oct. 2021. Karèn Fort and Mathieu Constant.
 - Maxime Guillaume, *Structures de traits pour les Grammaires Catégorielles Abstraites*, since July 2021, Philippe de Groote.
 - Laurine Jeannot, *Robust Semantic and Discourse Analysis of Natural Language System Specifications*, since January 2021, Maxime Amblard and Philippe de Groote.
 - Amandine Lecomte, *Analyse longitudinale de prise en charge psychothérapeutique de patients psychiatriques et de patients atteints de maladies neurodégénératives : informatisation et modélisation dialogique des indices comportementaux associés à l'efficacité (vs échec) des stratégies de prise en charge tentées par les thérapeute*, since October 2019, Michel Musiol and Alexandra König.
 - Chuyuan Li, *Formal and Statistical Modeling of Dialogue*, since October 2019, Maxime Amblard and Chloé Braud.
 - Pierre Ludmann, *Dynamic Construction of Discursive Structures*, since September 2017, Philippe de Groote and Sylvain Pogodalla.
 - Siyana Pavlova, *Tools and Methods for Semantic Annotation*, since November 2020, Maxime Amblard and Bruno Guillaume.

- Valentin Richard, *Aspects compositionnels et dynamiques de la sémantique inquisitrice*, since September 2021, Philippe de Groote
- Priyansh Trivedi, *Injecting Lexical and Semantic Knowledge into Word, Phrasal and Sentence Embeddings*, since November 2021, Philippe de Groote and Pascal Denis.

10.2.3 Juries

- Maxime Amblard and Philippe de Groote were members of the jury of the PhD thesis of Maria Boritchev (*Université de Lorraine*).

10.3 Popularization

10.3.1 Internal or external Inria responsibilities

- Maxime Amblard was the vice head of editorial board of Interstices.info until June and still member of the scientific committee

10.3.2 Articles and contents

- Maxime Amblard participated in a podcast on digital tools for mental health for Interstices.info
- Maxime Amblard has updated his unplugged activity [37].

10.3.3 Education

- Karèn Fort did a 2h seminar at *lycée Paul Valéry* in Paris, for 60 students of *première* specializing in computer science and science (physics and biology) on the work of a researcher (Paroles de Chercheuses et de Chercheurs).

10.3.4 Interventions

- Karèn Fort was interviewed in the MIT Technology Review, May 20th, 2021 : [The race to understand the exhilarating, dangerous world of language AI](#) (Karen Hao).
- Marc Anderson mentioned in "Pour un développement des IAs respectueux de la vie privée" Blog Binaire on "Le Monde" [website](#)

11 Scientific production

11.1 Major publications

- [1] G. Bonfante and B. Guillaume. ‘Non-size increasing Graph Rewriting for Natural Language Processing’. In: *Mathematical Structures in Computer Science* 28.08 (2018), pp. 1451–1484. DOI: [10.1017/S0960129518000178](#). URL: <https://hal.inria.fr/hal-00921038>.
- [2] G. Bonfante, B. Guillaume and G. Perrier. *Application of Graph Rewriting to Natural Language Processing*. Vol. 1. Logic, Linguistics and Computer Science Set. ISTE Wiley, 2018, p. 272. URL: <https://hal.inria.fr/hal-01814386>.
- [3] P. de Groote and M. Kanazawa. ‘A Note on Intensionalization’. In: *Journal of Logic, Language and Information* 22.2 (2013), pp. 173–194. DOI: [10.1007/s10849-013-9173-9](#). URL: <https://hal.inria.fr/hal-00909207>.
- [4] S. Pogodalla. ‘A syntax-semantics interface for Tree-Adjoining Grammars through Abstract Categorical Grammars’. In: *Journal of Language Modelling* 5.3 (2017), pp. 527–605. DOI: [10.15398/jlm.v5i3.193](#). URL: <https://hal.inria.fr/hal-01242154>.
- [5] R. Reinecke, T. A. Nazir, S. Carvallo and J. Jayez. ‘Factives at hand: When presupposition mode affects motor response’. In: *Journal of Experimental Psychology* (2022). DOI: [10.1037/xge0001167](#). URL: <https://hal.archives-ouvertes.fr/hal-03538732>.

11.2 Publications of the year

International journals

- [6] M. Dargnat and J. Jayez. ‘Mais et associés’. In: *Lexique* 29 (Dec. 2021). URL: <https://hal.archives-ouvertes.fr/hal-03538841>.
- [7] J. Maršík, M. Amblard and P. de Groote. ‘Introducing (λ), a λ -calculus for effectful computation’. In: *Theoretical Computer Science* 869 (Feb. 2021), pp. 108–155. DOI: 10.1016/j.tcs.2021.02.038. URL: <https://hal.inria.fr/hal-03200474>.
- [8] G. Perrier. ‘Study of non-projective dependencies in French’. In: *Revue TAL* 62.1 (2021). URL: <https://hal.inria.fr/hal-03389157>.
- [9] R. Reinecke, T. A. Nazir, S. Carvallo and J. Jayez. ‘Factives at hand: When presupposition mode affects motor response’. In: *Journal of Experimental Psychology* (2022). DOI: 10.1037/xge0001167. URL: <https://hal.archives-ouvertes.fr/hal-03538732>.

International peer-reviewed conferences

- [10] G. Alves, M. Amblard, F. Bernier, M. Couceiro and A. Napoli. ‘Reducing Unintended Bias of ML Models on Tabular and Textual Data’. In: The 8th IEEE International Conference on Data Science and Advanced Analytics (DSAA). Porto (virtual event), Portugal, 20th Oct. 2021, pp. 1–10. URL: <https://hal.archives-ouvertes.fr/hal-03312797>.
- [11] M. Amblard, M. Boritchev and P. de Groote. ‘An inquisitive account of wh-questions through event semantics’. In: Logical Aspects of Computational Linguistics 2021. Montpellier (online), France, 14th Dec. 2021. URL: <https://hal.inria.fr/hal-03541222>.
- [12] M. Amblard and M. Couceiro. ‘A Broad NLP Training from Speech to Knowledge’. In: NAACL 2021 - Fifth Workshop on Teaching NLP at the Annual Conference of the North American Chapter of the Association for Computational Linguistics. Mexico / Virtual, Mexico, 6th June 2021. URL: <https://hal.inria.fr/hal-03200480>.
- [13] M. Amblard, K. Fort, B. Guillaume, P. de Groote, C. Li, P. Ludmann, M. Musiol, S. Pavlova, G. Perrier and S. Pogodalla. ‘The Annotators Did Not Agree on Some of the Guidelines Examples’. In: ISA-17, the Seventeenth Workshop on Interoperable Semantic Annotation. Groningen/Virtuel, Netherlands, 16th June 2021. URL: <https://hal.inria.fr/hal-03298940>.
- [14] W. Babonnaud. ‘Covariant Subtyping Applied to Semantic Predicate Calculi’. In: Logical Aspects of Computational Linguistics 2021. Montpellier (online), France, 14th Dec. 2021. URL: <https://hal.inria.fr/hal-03542057>.
- [15] E. Calò, L. Jacquemin, T. Rosemblatt, M. Amblard, M. Couceiro and A. Kulkarni. ‘GECKo+: a Grammatical and Discourse Error Correction Tool’. In: *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 3 : Démonstrations*. TALN 2021 - 28e Conférence sur le Traitement Automatique des Langues Naturelles. Lille / Virtual, France: ATALA, 2021, pp. 8–11. URL: <https://hal.archives-ouvertes.fr/hal-03265914>.
- [16] H.-S. Choi, B. Guillaume and K. Fort. ‘Corpus-based Language Universals Analysis using Universal Dependencies’. In: Quasy (Quantitative Syntax), SyntaxFest 2021. Proceedings of the Second Workshop on Quantitative Syntax (Quasy, SyntaxFest 2021). Sofia, Bulgaria, 2021. URL: <https://hal.inria.fr/hal-03501774>.
- [17] H.-S. Choi, B. Guillaume, K. Fort and G. Perrier. ‘Investigating Dominant Word Order on Universal Dependencies with Graph Rewriting’. In: Recent Advances in Natural Language Processing (RANLP2021). Online, Bulgaria, 1st Sept. 2021. URL: <https://hal.inria.fr/hal-03322613>.
- [18] K. Gerdes, B. Guillaume, S. Kahane and G. Perrier. ‘Starting a new treebank? Go SUD! Theoretical and practical benefits of the Surface-Syntactic distributional approach’. In: Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021). Sofia, Bulgaria, 2021. URL: <https://hal.inria.fr/hal-03509136>.

- [19] P. de Groote and V. D. Richard. ‘Embedding Intensional Semantics into Inquisitive Semantics’. In: 17th Meeting on the Mathematics of Language. Montpellier (online), France, 13th Dec. 2021. URL: <https://hal.inria.fr/hal-03541183>.
- [20] B. Guillaume. ‘Graph Matching and Graph Rewriting: GREW tools for corpus exploration, maintenance and conversion’. In: EACL 2021 - 16th conference of the European Chapter of the Association for Computational Linguistics. Kiev/Online, Ukraine, 19th Apr. 2021. URL: <https://hal.inria.fr/hal-03177701>.
- [21] B. Guillaume and G. Perrier. ‘Graph Rewriting for Enhanced Universal Dependencies’. In: IWPT 2021 - 17th International Conference on Parsing Technologies. Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021). Bangkok (Online), Thailand, 6th Aug. 2021. URL: <https://hal.inria.fr/hal-03326492>.
- [22] S. Kahane, M. Vanhove, R. Ziane and B. Guillaume. ‘A morph-based and a word-based treebank for Beja’. In: *SyntaxFest. TLT 2021 - 20th International Workshop on Treebanks and Linguistic Theories*. Sofia, Bulgaria, 2021. URL: <https://hal.archives-ouvertes.fr/hal-03494462>.
- [23] C. Li, M. Amblard, C. Braud, C. Demily, N. Franck and M. Musiol. ‘Investigating non lexical markers of the language of schizophrenia in spontaneous conversations’. In: *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*. 2nd Workshop on Computational Approaches to Discourse (CODI 2021). Punta Cana, Dominican Republic: Association for Computational Linguistics, 10th Nov. 2021, pp. 20–28. DOI: [10.18653/v1/2021.codi-main.3](https://doi.org/10.18653/v1/2021.codi-main.3). URL: <https://hal.inria.fr/hal-03537698>.
- [24] J. Renner, P. Trivedi, G. Maheshwari, R. Gilleron and P. Denis. ‘An End-to-End Approach for Full Bridging Resolution’. In: CODI-CRAC 2021 Shared-Task: Anaphora Resolution in Dialogues. Proceedings of the CODI-CRAC 2021 Shared Task on Anaphora, Bridging, and Discourse Deixis in Dialogue. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 48–54. URL: <https://hal.archives-ouvertes.fr/hal-03431313>.

National peer-reviewed Conferences

- [25] S. Ahmadi, M. Constant, K. Fort, B. Guillaume and J. P. McCrae. ‘Convertir le Trésor de la Langue Française en Ontolex-Lemon : un zeste de données liées’. In: Journées LIFT 2021 - Linguistique informatique, formelle et de terrain. Grenoble, France, 6th Dec. 2021. URL: <https://hal.inria.fr/hal-03463294>.
- [26] M. Boritchev and M. Amblard. ‘DinG – a corpus of transcriptions of real-life, oral, spontaneous multi-party dialogues between French-speaking players of Catan’. In: Journées LIFT 2021 - Linguistique informatique, formelle et de terrain. Grenoble, France, 6th Dec. 2021. URL: <https://hal.inria.fr/hal-03537970>.
- [27] H.-S. Choi, B. Guillaume and K. Fort. ‘Analyse orientée corpus d’universaux de Greenberg sur Universal Dependencies’. In: Journées LIFT 2021 - Linguistique informatique, formelle et de terrain. Grenoble, France, 6th Dec. 2021. URL: <https://hal.inria.fr/hal-03462112>.

Conferences without proceedings

- [28] W. Babonnaud. ‘On the Dual Interpretation of Nouns as Types and Predicates in Semantic Type Theories’. In: 2nd Workshop on Computing Semantics with Types, Frames and Related Structures, ESSLI 2021. Virtual, Netherlands, 26th July 2021. URL: <https://hal.archives-ouvertes.fr/hal-03468606>.

Scientific book chapters

- [29] M. Amblard, M. Musiol and M. Rebuschi. ‘Discourse coherence - From psychology to linguistics and back again’. In: *(In)coherence of discourse - Formal and Conceptual issues of Language*. Springer, June 2021. URL: <https://hal.inria.fr/hal-02269640>.

- [30] M. Amblard, M. Rebuschi and M. Musiol. ‘Corpus et pathologie mentale : particularités dans la constitution et l’analyse d’une ressource’. In: *Les corpus en sciences humaines et sociales*. Presses Universitaires de Nancy, Feb. 2021. URL: <https://hal.inria.fr/hal-02269622>.
- [31] M. M. Anderson. ‘Personality in Seagoing Ships’. In: *The Cultural Power of Personal Objects: Traditional Accounts and New Perspectives*. SUNY series in American Philosophy and Cultural Thought. Jared Kemling, Dec. 2021. URL: <https://hal.inria.fr/hal-03536586>.
- [32] M. Boritchev and M. Amblard. ‘Picturing Questions and Answers - a formal approach to SLAM’. In: *(In)coherence of discourse - Formal and Conceptual issues of Language*. (In)coherence of discourse Formal and Conceptual issues of Language. Springer, June 2021. URL: <https://hal.inria.fr/hal-02269631>.
- [33] R. Chatila, M. Chetouani, R. Debailly, K. Fort, F. Dhombres, E. Fournier, L. Khellaf, S. Lavault, S. Payan, É. Pesce, A. Roby-Brami and M. Van Donzel. ‘L’éthique de la recherche en temps de pandémie : un retour d’expérience’. In: *Sorbonnavirus - Regards sur la crise du coronavirus*. 27th May 2021. URL: <https://hal.inria.fr/hal-03237653>.
- [34] A. Kuyumcuyan and M. Musiol. ‘L’entretien clinique avec la personne polyhandicapée : un terrain commun sciences du langage / psychiatrie’. In: *Les sciences du langage face aux défis de la disciplinarisation et de l’interdisciplinarité*. Malika Temmar, Marina Krylyschin, Guy Achard-Bayle (éds). Jan. 2021. URL: <https://hal.inria.fr/hal-03439236>.

Edition (books, proceedings, special issue of a journal)

- [35] M. Amblard, M. Musiol and M. Rebuschi. *(In)coherence of Discourse. Formal and conceptual issues of language*. Language, Cognition and Mind (Chungmin Lee “Editor Springer book series). Springer, June 2021. URL: <https://hal.archives-ouvertes.fr/hal-02501028>.

Doctoral dissertations and habilitation theses

- [36] M. Boritchev. ‘Dialogue Modeling in a Dynamic Framework’. Université de Lorraine (UL), Vandoeuvre-lès-Nancy, FRA., 22nd Nov. 2021. URL: <https://hal.archives-ouvertes.fr/tel-03541628>.

Reports & preprints

- [37] M. Amblard. *Les lapins magiciens*. 26th June 2021. URL: <https://hal.inria.fr/hal-02414345>.

Other scientific publications

- [38] V. D. Richard. ‘Embedding Intentional Semantics into Inquisitive Semantics’. Université de Paris, 23rd June 2021. URL: <https://hal.inria.fr/hal-03280889>.
- [39] D. Sayers, R. Sousa-Silva, S. Höhn, L. Ahmedi, K. Allkivi-Metsoja, D. Anastasiou, Š. Beňuš, L. Bowker, E. Bytyçi, A. Catala, A. Çepani, R. Chacón-Beltrán, S. Dadi, F. Dalipi, V. Despotovic, A. Doczekalska, S. Drude, K. Fort, R. Fuchs, C. Galinski, F. Gobbo, T. Gungor, S. Guo, K. Höckner, P. Láncoš, T. Libal, T. Jantunen, D. Jones, B. Klimova, E. Korkmaz, S. Maučec Mirjam, M. Melo, F. Meunier, B. Migge, B. Mititelu Verginica, A. Névéol, A. Rossi, A. Pareja-Lora, C. Sanchez-Stockhammer, A. Şahin, A. Soltan, C. Soria, S. Shaikh, M. Turchi and S. Yildirim Yayilgan. *The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies*. 19th May 2021. URL: <https://hal.archives-ouvertes.fr/hal-03230287>.

11.3 Other

Scientific popularization

- [40] M. Amblard and J. Jongwane. ‘L’intelligence artificielle au service des maladies mentales’. In: *Interstices* (30th Apr. 2021). URL: <https://hal.inria.fr/hal-03220415>.

11.4 Cited publications

- [41] R. Cooper, D. Crouch, J. Van Eijck, C. Fox, J. Van Genabith, J. Jaspars, H. Kamp, D. Milward, M. Pinkal, M. Poesio, S. Pulman, T. Briscoe, H. Maier and K. Konrad. *Using the framework*. Tech. rep. Technical Report LRE 62-051 D-16. The FraCaS Consortium, 1996.
- [42] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova. ‘Bert: Pre-training of deep bidirectional transformers for language understanding’. In: *arXiv preprint arXiv:1810.04805* (2018).
- [43] P. de Groote. ‘Towards a Montagovian Account of Dynamics’. In: *16th Semantics and Linguistic Theory conference - SALT2006*. Ed. by M. Gibson and J. Howell. Tokyo, Japan, 2006. URL: <https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/2952/0>.
- [44] P. de Groote. ‘Towards abstract categorial grammars’. In: *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter*. Colloque avec actes et comité de lecture. internationale. Toulouse, France: Association for Computational Linguistics, July 2001, pp. 148–155. URL: <http://hal.inria.fr/inria-00100529/en>.
- [45] P. de Groote and S. Pogodalla. ‘On the expressive power of abstract categorial grammars: Representing context-free formalisms’. In: *Journal of Logic, Language and Information* 13.4 (2004), pp. 421–438.
- [46] B. Guillaume and G. Perrier. ‘Interaction Grammars’. In: *Research on Language & Computation* 7 (2009), pp. 171–208.
- [47] Z. S. Harris. ‘Distributional Structure’. In: *Word* 10.2-3 (1954), pp. 146–162. DOI: [10.1080/00437956.1954.11659520](https://doi.org/10.1080/00437956.1954.11659520).
- [48] K. Heylen, Y. Peirsman, D. Geeraerts and D. Speelman. ‘Modelling Word Similarity: an Evaluation of Automatic Synonymy Extraction Algorithms.’ In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*. Marrakech, Morocco: European Language Resources Association (ELRA), May 2008. URL: http://www.lrec-conf.org/proceedings/lrec2008/pdf/818_paper.pdf.
- [49] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer and O. Levy. ‘Spanbert: Improving pre-training by representing and predicting spans’. In: *Transactions of the Association for Computational Linguistics* 8 (2020), pp. 64–77.
- [50] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov. ‘Roberta: A robustly optimized bert pretraining approach’. In: *arXiv preprint arXiv:1907.11692* (2019).
- [51] I. Mel’čuk and A. Polguère. ‘La définition lexicographique selon la Lexicologie Explicative et Combinatoire’. In: *Cahiers de lexicologie* 109 (2016), pp. 61–91. DOI: [10.15122/isbn.978-2-406-06861-7.p.0061](https://doi.org/10.15122/isbn.978-2-406-06861-7.p.0061).
- [52] S. Padó and M. Lapata. ‘Dependency-Based Construction of Semantic Space Models’. In: *Computational Linguistics* 33.2 (2007), pp. 161–199. DOI: [10.1162/coli.2007.33.2.161](https://doi.org/10.1162/coli.2007.33.2.161). URL: <https://www.aclweb.org/anthology/J07-2002>.
- [53] M. Padró, M. Idiart, A. Villavicencio and C. Ramisch. ‘Comparing Similarity Measures for Distributional Thesauri’. In: *Proceedings of LREC 2014*. Ed. by N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk and S. Piperidis. 2014. URL: <https://www.aclweb.org/anthology/L14-1496/>.
- [54] Y. Peirsman, K. Heylen and D. Speelman. ‘Finding semantically related words in Dutch: co-occurrences versus syntactic contexts’. In: *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*. Ed. by M. Baroni, A. Lenci and M. Sahlgren. 2007, pp. 9–16. URL: <https://bibliotek.dk/eng/moreinfo/netarchive/870970-basis:28214510>.
- [55] G. Perrier. ‘A French Interaction Grammar’. In: *International Conference on Recent Advances in Natural Language Processing - RANLP 2007*. Ed. by G. Angelova, K. Bontcheva, R. Mitkov, N. Nicolov and K. Simov. IPP & BAS & ACL-Bulgaria. Borovets, Bulgarie: INCOMA Ltd, Shoumen, Bulgaria, 2007, pp. 463–467. URL: <http://hal.inria.fr/inria-00184108/en/>.

-
- [56] M. E. Peters, M. Neumann, R. L. Logan IV, R. Schwartz, V. Joshi, S. Singh and N. A. Smith. ‘Knowledge enhanced contextual word representations’. In: *arXiv preprint arXiv:1909.04164* (2019).
- [57] S. Vashishth, S. Sanyal, V. Nitin and P. Talukdar. ‘Composition-based multi-relational graph convolutional networks’. In: *arXiv preprint arXiv:1911.03082* (2019).
- [58] J. Weeds, D. Weir and D. McCarthy. ‘Characterising Measures of Lexical Distributional Similarity’. In: *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. Geneva, Switzerland: COLING, 23rd–27th Aug. 2004, pp. 1015–1021. URL: <https://www.aclweb.org/anthology/C04-1146>.