Lorraine Laboratory of **Research** in **Computer Science** and its **Applications**

ACTIVITY REPORT 2011 - 2016
PROSPECTIVES FOR 2017 - 2022

A research unit from the **research department AM2I** of **Lorraine University: Automatics, Mathematics, Computer Science and** their **Interactions**

Loria
Laboratoire lorrain de recherche en informatique et ses applications



Volume 5

Loria
cnrs
Inria
INVENTEURS DU MONDE NUMÉRIQUE
UNIVERSITÉ DE LORRAINE

Contents

01

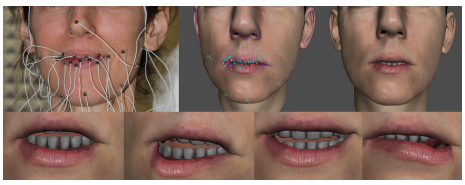
Activity Report

Department 4

Natural Language Processing & Knowledge Discovery

Department Head: Bruno Guillaume

Department 4 named *Natural Language Processing & Knowledge Discovery* (NLPKD) is composed of eight teams which are interested in the processing of all media used by humans for communication (speech, text and other kind of written document) and in the modeling of the content of these data (knowledge modeling, semantic of natural language). Speech processing and modeling is the main focus of **MULTISPEECH** and **SMART** (speech recognition, speech synthesis and translation are considered). Natural Language Processing is the main interest of **SYNALP** and **SÉMAGRAMME** with various activities about modeling, analysis and generation of several levels (syntax, semantics and discourse). **QGAR** and **READ** are active on document processing: segmentation, hand-written recognition, shape recognition. Knowledge discovery and knowledge processing is at the heart of **ORPAILLEUR** team: fundamental methods are developed and applied to areas like life science and data of the semantic web; **CELLO** is involved in knowledge modeling through extensions of modal logic. Despite different research areas, the eight teams of the department 4 share common background either on symbolic methods or on statistical approaches that are applied to most of these topics.



Overview of Department 4

1 Department Composition

Department leader

Bruno Guillaume (since September, 2013)
Yannick Toussaint (before September, 2013)

List of teams

- **CELLO**: Computational Epistemic Logic in Lorraine
- **ORPAILLEUR**: Knowledge Discovery and Knowledge Engineering (EPC Inria)
- **MULTISPEECH**: Speech Modeling for Facilitating Oral-Based Communication (EPC Inria)
- **QGAR**: Querying Graphics through Analysis and Recognition
- **READ**: Reconnaissance de l'Ecriture et Analyse de Documents (Writting recognition and document analysis)
- **SÉMAGRAMME**: Semantic Analysis of Natural Language (EPC Inria)
- **SMART**: Statistical Machine Translation
- **SYNALP**: Statistic and Symbolic Natural Language Processing

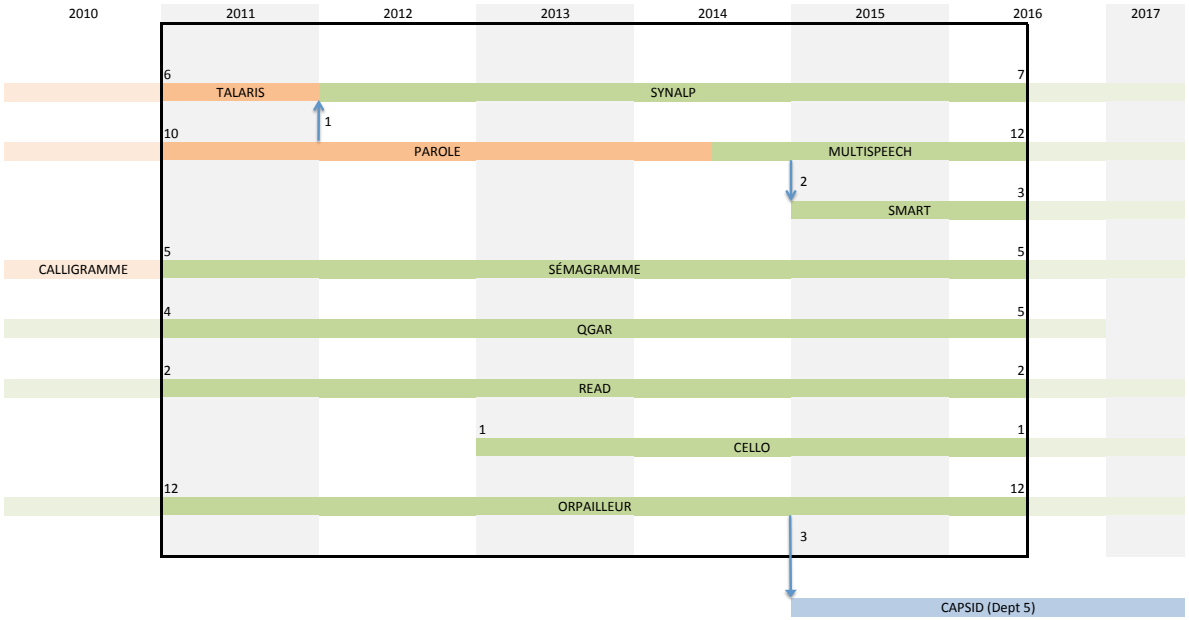
	PR	MCF	DR	CR	Total
2011	4	21	7	7	39
2016	7	22	6	12	47

PhD's defended	40	On-going PhD's	27

In the 67 PhD Thesis, 9 are co-supervised PhD funded outside the department. The 58 PhD remaining theses are funded as follows: 12 by MESR, 8 by Inria, 8 by industrial contracts, 7 by ANR projects, 4 by European projects, 2 by Foreign fundings (Vietnam and Algeria). Other funding are given by INCa, CNRS, FUI, DGA, Campus France, École Polytechnique.

During the evaluation period, 26 post-docs and 28 engineers were in the teams of the department.

Department evolution



The figure above gives an overview of the teams and their evolution during the evaluation period. Arrows correspond to creation of a new team where several members of the department move from an old team to the new one.

2 Life of the department

The department organizes a seminar with invited talks every two months both for members of the department and for students of the Erasmus Mundus master program in NLP (Natural Language Processing). Since last year, a department day is organized each year where all PhD students are invited to present their works to members of other teams. The department is also in charge of the ranking of PhD applications and of the profiling of job for permanent positions.

There are many collaborations between the different teams of the department, in particular through two national funding PIA (Programme d'Investissment d'Avenir): **ORPAILLEUR** and **SYNALP** are involved in the PIA Istex; **SYNALP**, **SÉMAGRAMME** and **MULTISPEECH** are involved in the PIA Ortolang (two engineers were shared by these teams).

Combined national and regional fundings through CPER (Contrat de Plan État-Région) were used notably to construct common infrastructures like clusters and experimental platforms ("CPER TALC" until 2013 and "CPER LCHN" since 2014).

In terms of collaborations with other departments, we mainly have interactions with Department 5 (common projects, co-supervised theses) and also a few interactions with Department 2.

3 Research topics

Keywords: Speech processing, Natural Language Processing, Document Processing, Knowledge Processing and Knowledge Discovery.

The teams of the NLPKD department have a wide scope, the main centers of interest are:

- **Speech Processing.** Many different aspects are covered by teams **MULTISPEECH**, **SMART** and **SYNALP**. Main topics are: Speech Recognition, Speech Synthesis, Multimodal Speech Processing, Source Separation, Articulatory Modeling, Speech-to-speech Translation.

- **Natural Language Processing.** In the department, the teams **SMART**, **SYNALP**, **SÉMAGRAMME** and **ORPAILLEUR** are working on NLP. They are active in: Language Modeling (Formal semantics, Syntax-Semantic Interface, Discourse Dynamics), Natural Language Processing (Syntax and Semantics Parsing, Dialog Processing, Text Clustering), Natural Language Generation and Text Mining.

- **Document Processing.** The two teams **READ** and **QGAR** focus on Handwriting Recognition, Document Image Analysis and Indexing, Pattern Recognition, Feature Extraction, Segmentation and Incremental Classification.

- **Knowledge Processing and Knowledge Discovery.** The **ORPAILLEUR** team has a strong focus on Knowledge with topics like: Knowledge Discovery in Databases, Formal Concept Analysis, Knowledge Engineering or Web of Data. The **CELLO** team is involved in Knowledge and Belief, Knowledge progression and its formalization in Modal Logic.

The research activities in the department also cover a large range of methods:

- **Statistical or Learning based methods** (HMM, BIRL, SVM, CRF, DNN, DBN) are used in all area described above, i.e. speech, text, document and Knowledge processing.

- **Symbolic based methods** are also widely used in **SÉMAGRAMME**, **SYNALP**, **ORPAILLEUR** and **CELLO** with a focus on Logic in **SÉMAGRAMME** and **CELLO**, on Formal Concept Analysis and Case-based Reasoning in **ORPAILLEUR** and on hybrid methods in **SYNALP**.

The eight teams of the department are working in a large set of application domains: Foreign Language Learning, Sentiment Analysis and Opinion mining, Preference Modeling, Pathological speech or Pathological language, Life Science, Dialog System, Serious Games, Chemistry, Typography.

We summarize below the research topics and application domains of these 8 teams as they exist in 2016. A more detailed description is given in the sections dedicated to each team.

During the evaluation period, the **PAROLE** team has become **MULTISPEECH** and has now a stronger focus on speech. The main objective of the team is to better understand how humans produce and perceive speech. Statistical approaches are common for processing speech and they are used in actual applications but their performance drops significantly when dealing with degraded speech, such as noisy signals and spontaneous speech. **MULTISPEECH** has used a wide range of statistical methods (with a focus on Deep Neural Networks) and applied them to audio source separation, speech enhancement, acoustic modeling, pronunciation modeling, language modeling and speech generation. One of the interests of the team is to investigate uncertainty estimation, *i.e.* to quantify the confidence in the output of a given speech processing technique and to exploit it for further processing. Uncertainty estimation is applied to acoustic modeling, speech recognition, phonetic segmentation and prosody. Another research axis is more oriented toward synthesis. In this axis, the team has worked on articulatory modeling, audio-signal speech synthesis (animation of a 3D model of human face) and categorization of sounds and prosody. Applications for hard-of-hiring children or for feedback for foreign language learners are considered.

Since 2015 a part of the activities previously developed in the **PAROLE** team have led to a new team called **SMART**. The objective of **SMART** team is to develop statistical approaches for Natural Language Processing applications. Methods are corpus-based and researches are focused on multi-lingual aspects. Machine translation is one of the main area for the team. More precisely, application domains are speech-to-speech translation, translation of modern Arabic to different Arabic dialects and quality estimation for

machine translation. Other notable activities of **SMART** concern cross-lingual sentiment analysis and opinion mining but also vocabulary and data selection for speech recognition.

The **SYNALP** team research concerns hybrid methods in Natural Language Processing. They investigate the interplay between symbolic and stochastic processing; to explore supervised, semi-supervised and unsupervised approaches. In Natural Language Generation, the team developed robust techniques; they participate to the creation of an international shared task and applied these techniques to practical applications in computer aided language learning applications. Parsing was also investigated at the syntactic level (dependency structures) and at the semantic level (Named Entities Recognition, Semantic Role Labeling); in both cases, a large range of statistical techniques was explored. Text clustering and mining is also considered in **SYNALP**; with a focus on incremental clustering on multidimentional, unbalanced, noisy and sparse data.

The objective of the **SÉMAGRAMME** team is to design and develop logic-based models for the semantic analysis of natural language utterances and discourses (including pragmatic phenomena). The team focuses on the semantics of natural languages with an approach based on abstract generic models (Graph Rewriting, Abstract Categorial Grammars) of the syntax-semantic interface, they study the formal properties of the models and develop large scale grammars. To take into account the dynamicity or discourse interpretation, **SÉMAGRAMME** generalises the methodological tools applied by Montague to semantics; these tools are based on compositionality . The approach is applied to model the way quantifiers in natural languages may dynamically extend their scope. The team also contributes to the production of linguistics tools and resources such as syntactic lexicons, grammars, treebanks and annotated corpora.

CELLO is a small team which started in 2013 when Hans van Ditmarsch arrived at Loria, granted by the ERC Starting Grant EPS 313360. The research area of the team is Dynamic Epistemic Logic and Temporal Modal Logic with a focus on Epistemic protocol synthesis. Theoretical research on extensions of modal logic like public announcement logic and refinement modal logic were conducted. Both expressivity and complexity of these extensions are studied. Applications to security of communication protocols and to cryptography were also considered. The team is active in generalization of the cards cryptography problem and of the gossip protocol in synchronous, asynchronous or dynamic context.

The **ORPAILLEUR** team focuses on KDD (Knowledge Discovery in Databases) and investigates both symbolic and numerical data mining methods. Symbolic methods are based on pattern mining, FCA (Formal Concept Analysis) and extensions of FCA such as Pattern Structures and Relational Concept Analysis. Numerical methods are based on probabilistic approaches such as second-order Hidden Markov Models which are well adapted to the mining of temporal and spatial data. Domain knowledge can improve and guide the KDD process at each step of the process, this leads to KDDK (Knowledge Discovery guided by Domain Knowledge). **ORPAILLEUR** is active both on Fundamentals of KDDK and on applications to other domains like Life Science. KDD methods are also applied to the web of data; the challenge is to build knowledge bases and efficient knowledge systems using mining and knowledge discovery technique on the data of the semantic web. The part of these activities more focused on bioinformatics is now developed in the CAPSID team of the department 5.

The **READ** team is interested in handwriting recognition, analysis of scanned document content and their incremental classification and segmentation. For handwriting recognition, the team propose several statistical based methods with applications to Arabic and Urdu languages and to signature verification. In document segmentation (Oséo DOD contract with 2 companies Itesoft and Sagecom), focus was put on extraction of tables, separation of handwritten/printed script and extraction of named entities.

The **QGAR** team work on graphics-rich image analysis and document recognition with a focus on indexation, segmentation of images and shape recognition. The team has proposed several new methods based on noise removal to improve image segmentation and text detection with application in typography to detect typefaces. **QGAR** proposed generic polar harmonic transformations and extensions of the Radon transform to improve robustness of the detection of shapes in noisy document where shapes are subject to variation.

4 Main Results

We shortly present some results obtained by the teams of the department. More detailed descriptions of these results are given in the sections dedicated to each team.

Speech Processing

Statistical modeling of speech (**MULTISPEECH**). Audio source separation is a key issue for the enhancement of speech recognition in noise or singing voice. Recently, we introduced new modeling frameworks for source separation based on alpha-stable distributions [163, 164], deep neural networks [58], and fusion of multiple separation systems [140, 54, 34]. Our multichannel deep neural network-based system [58] led to good results in 2015 evaluation campaign [179, 206]: we stand among the top teams worldwide in the field of source separation and speech enhancement thanks to the combination of deep learning with our expertise in signal processing.

Articulatory Inversion (**MULTISPEECH**). For articulatory inversion, we have extended the approach based on the analysis by synthesis paradigm to the use of cepstral coefficients as input [98]; this has shown the importance of improving the quality of the direct model. The Xarticulators software was developed to construct articulatory models from images of the mid-sagittal slices (MRI) or projections (X-ray) of the vocal tract. The resulting 2D articulatory models [159] are far more complete than other models.

Explicit modeling of speech production (**MULTISPEECH**). Based on our text-to-speech synthesis system (SoJA), we have developed a bimodal acoustic-visual synthesis technique that concurrently generates the acoustic speech signal and a 3D animation of the speaker's face [59]. In the visual domain, we had mainly focused on the dynamics of the face and we have developed an algorithm to animate the 3D model of human face from a limited number of markers [259]. Applications regarding acquisition of language by hard-of-hearing children and foreign language learning are also considered.

Uncertainty estimation in speech processing (**MULTISPEECH**). Uncertainty estimation of the output of a given speech processing technique is important to improve it. On speech enhancement, we achieved 30% relative word error rate reduction with respect to conventional decoding (without uncertainty) on audio speech recognition vocabulary tasks, that is about twice as much as the best existing uncertainty estimator and propagator [225, 70]. Uncertainty estimation was also investigated in systems for helping hard-of-hearing people (RAPSODIE project, [193]) and in the phonetic segmentation domain (non-native speech-text alignment in the ALLEGRO project).

Text Processing

Syntax (**SYNALP**, **SÉMAGRAMME**). Dependency parsing were explored with supervised methods [967, 970] and hybrid methods (weakly supervised and rule-based) [969]. New approach to dependency syntax parsing based on Graph Rewriting were explored [846].

Semantics (**SYNALP**, **SÉMAGRAMME**). We designed a semisupervised model that combines a Bayesian model with a maximum entropy model for Semantic Role Labelling [1025]. We have derived a new method to train a linear classifier in an unsupervised way and applied it to predicate identification and Named Entity Recognition [1037]. Our main contributions are (a) a formal solution to the integration of the risk in the binary case, which alleviates the need for expensive numerical integration; (b) the development of features selection methods to focus the stochastic gradient onto the most interesting part of the search space; (c) new and faster Gaussian Mixture Model training algorithms that are adapted to the specificities of this model. In order to provide a modular approach to the semantic modeling of several phenomena (quantification, intensionality), we defined a general intensionalization procedure that turns an extensional semantics into an intensionalized one that is capable of accommodating truly intensional lexical items without changing the compositional semantic rules [814, 837].

Discourse and dialog (**SYNALP**, **SÉMAGRAMME**). We refine previous works on continuation semantics in order to model more precisely the accessibility constraints that exist for a referring expression [859], in particular in the case of presupposition induced by the use of proper names [807] (E. W. Beth Dissertation Prize). We also studied discourse dynamics of morbid discourses. In collaboration with a psycho-linguistist and an epistemologist, we performed a formal analysis of pathological conversations involving a schizophrenic patient and a psychiatrist. Such dialogs are characterized by so-called discourse ruptures. Our analysis relies both on semantic and pragmatic features [811, 817]. We developed several dialog systems (symbolic, supervised and hybrid symbolic/supervised) within the framework of the Emospeech project [1039].

Natural Language Generation (**SYNALP**, **SÉMAGRAMME**). The approach being pursued in Natural Language Generation combines linguistically informed models with statistical models [980]. We developed both symbolic [984] and statistical [986, 1029] error mining approaches. Applying these methods to a generation corpus of an international shared task on surface realization, we increase coverage from 38.5% to 81% [943]. To deal with the combinatorial complexity of symbolic grammars, we developed a surface realization algorithm which allows for generation of the Penn Tree Bank sentences in an average of 2 seconds per input [1030]. Using a shared task data (created in collaboration with Stanford SRI [962]), we defined a linguistically driven grammar induction method that outperformed a competing statistical system [995]. Reversibility of ACG (Abstract Categorial Grammars) were also considered in generation tasks in the Polymnie project [832, 833, 834, 851].

Document Processing

Handwritting recognition (**READ**). We have developed several systems, based on generative and discriminative models: Arabic/Latin and machine-printed/handwritten word discrimination using [786, 801], Probabilistic Graphical Models for Arabic handwritten word recognition [793], Collaborative combination of neuron-linguistic classifiers for large Arabic word vocabulary recognition [761], Fuzzy based preprocessing for online Urdu script recognition [759], Signature verification [796].

Feature extraction, Segmentation and indexing (**READ**, **QGAR**). Through the Oséo ITESOFT project, we investigated the global problem of document flow segmentation, focusing on labeling logical structures [764, 799], Entity Recognition[794, 795] and extraction of tables and forms [790, 797]. We consider incremental classification in case of uncertain labeling knowledge and evolving data (novel class detection); we have developed an adaptive streaming active learning strategy

used by the ITESOFT company for mail sorting [758, 776]. We also consider feature extraction and segmentation in graphics-rich documents. In this context, we consider edge noise removal both when the model of noise is know [665] or unknown [698] and we propose interactive segmentation methods[662] and adapt them to tablets [720]. Text detection in graphics document is also difficult, we present a new hybrid page segmentation approach which is able to segment and identify lines, backgrounds, photo regions and multiscale text [739, 703].

Knowledge Processing

Fundamentals of KDDK (**ORPAILLEUR**). We have contributed to an extension of FCA (Formal Concept Analysis) called RCA (Relational Concept Analysis) which is able to analyze objects described both by binary and relational attributes and can play an important role in text classification and text mining [395, 530, 537]. We improved standard algorithms for building lattices from large data and for completing the algorithm collection of the Coron platform [398]. We also worked on temporal data and healthcare patient trajectories [354, 506, 355]. On privacy challenges, we introduced different data anonymization methodologies based on different usability scenarios [556, 554]. These algorithms are applied to monitor and analyze mobile phone data while taking into account the privacy issues.

KDDK in Life Sciences (**ORPAILLEUR**). Biological data are complex from many points of view, e.g. huge size, high-dimensionality, deep inter-connections, and quick evolutions. We explored the potential complementarity of ILP (Inductive Logic Programming) and FCA for analyzing biological data. ILP can be combined with FCA for interpreting theories w.r.t. domain knowledge, e.g. for understanding drug side-effects [310]. We proposed some approaches for selecting, integrating, and mining Linked Open Data with the objective of discovering genes responsible for a disease [521, 539]. We proposed a generic method based on FCA and pattern structures to explore and complete hand-made annotations (links between data and ontologies) in biomedical document databases [490]. Finally, in the Hybride project, we focused on extracting relations between named entities using graph mining methods over a collection of dependency graphs. We also applied this idea to learn syntactic patterns relating diseases with symptoms which are used for discovering relationships in biomedical texts [518]. Since 2015, activities centered on bio-infomatics are conducted in the CAPSID team in department 5.

Knowledge Engineering (**ORPAILLEUR**). We investigate knowledge discovery in the web of data and mainly on data represented in the RDF (Resource Description Framework) format. We are interested in the completeness of the data and the their potential to provide concept definitions in terms of necessary and sufficient conditions that can also be considered as a search for descriptions [415]. Other important aspects are concerned with data access, data visualization w.r.t. the SPARQL query language [419]. CBR (Case-based reasoning) is a reasoning paradigm based on experience reuse. A classical case-based inference consists in selecting in a case base a case similar to the query (retrieval step) and in modifying it in order to meet the query (adaptation step). Theoretical and practical researches have led to the development of tools for building and improving CBR systems [608, 352], such as a case-based inference engine and a library for revision-based adaptation. In addition, some studies have addressed the issue of knowledge acquisition for CBR, in order to feed the knowledge base [514] or to manage knowledge reliability [510].

5 Synthesis of publications

	2011	2012	2013	2014	2015	2016	Total
PhD Thesis	9	5	9	8	9	1	41
H.D.R	2		4	1			7
Journal	29	33	38	57	23	16	196
Conference proceedings	124	124	107	134	117	17	623
Book chapter	16	13	4	12	2	1	48
Book (written)							
Book or special issue (edited)	4	2	4	8	4	3	25
Patent						1	1
General audience papers	1			1		1	3

The teams in department 4 have different expertise areas and they publish in quite different journals and conferences. Hence, there are only a few overlaps between top journals and top conferences of different teams.

List of top journals in which we have published

- Artificial Intelligence [7, 400]
- IEEE Signal Processing Magazine [71]
- Theoretical Computer Science [3, 2, 306]
- IEEE Transactions on Audio, Speech and Language Processing [65, 68, 70]
- Journal of Logic, Language and Information [814]
- Computational Linguistics [935, 943]
- Neurocomputing [948, 664]
- Pattern Recognition Letters [672, 757]
- Traitement Automatique des Langues [813, 811]
- Information Science [378]

Around 30% (56 of 189) of the publications in journals are in the list of top journals of the corresponding teams.

List of top conferences in which we have published

- International Joint Conference on Artificial Intelligence (IJCAI)[17, 470, 526]
- Annual Conference of the International Communication Association (INTERSPEECH) [83, 92, 107, 118, 122, 128, 131, 139, 144, 148, 149, 160, 169, 182, 184, 188, 190, 198, 199, 204, 220, 221, 229, 906, 905, 970, 969, 1060, 967, 968, 991]
- International Conference in Computational Linguistics (CoLing) [1030, 852]
- Annual Meeting of the Association for Computational Linguistics (ACL) [980, 1031, 995]
- Conference of the European Chapter of the Association for Computational Linguistics (EACL)[994]
- IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP) [106, 109, 116, 146, 152, 163, 165, 196, 197, 201, 209, 216, 222, 223, 224, 225]
- Semantics and Linguistic Theory (SALT) (1)
- Advances in Modal Logic (AiML) [24, 21]
- International Conference on Data-Mining (ICDM) [547, 520, 532]

Around 40% (240 of 614) of the publications in conferences are in the list of top conferences of the corresponding teams.

6 Software

We highlight here some of our principal software and refer to the team's sections for details.

- **FASST (Flexible Audio Source Separation Toolbox)** is a toolbox for audio source separation. It offers the possibility for users to specify easily a suitable algorithm for their use case thanks to the general modeling and estimation framework. It forms the basis of most of our current research in audio source separation. Version 2.0 was released in 2014; it is distributed under the Q Public License. To the best of our knowledge, this is the only public implementation of multichannel NMF (Non-negative Matrix Factorization).
- **VisArtico: Visualization of EMA Articulatory data**. VisArtico is a user-friendly software which allows visualizing EMA data acquired by an articulograph. This visualization software has been designed to display the articulatory coil trajectories, synchronized with the corresponding acoustic recordings. Moreover, VisArtico not only allows viewing the coils but also enriches the visual information by indicating clearly and graphically the data for the tongue, lips and jaw. In addition, it is possible to insert images (MRI or X-Ray, for instance) to compare the EMA data with data obtained through other acquisition techniques. The software is freely available for research and it has been downloaded by more than 130 researchers around the world.
- Within the Empathic project, we developed the **SATI API** (Sentiment Analysis from Textual Information). SATI is a web API that enables to analyze the sentiment or emotion in a sentence. It makes use of parsing techniques and supervised machine learning as well. The web API offers a flexible interface that returns, given a sentence, its most salient sentiment or emotion encoded in the EmotionML format, a new W3C standard for emotion representation. The SATI software has also been released as a mobile application that is distributed both in the Google and Apple stores.
- The **OrphaMine** platform[1], developed as part of the ANR Hybrid project, enables visualization, data integration and in-depth analytics. The data at the heart of the platform is about orphan diseases and is extracted from the OrphaData ontology. The aim is to build a true collaborative portal that will serve the different actors of the Hybrid project: (i) A general visualization of OrphaData data for physicians working, maintaining and developing this knowledge database about orphan diseases. (ii) The integration of analytics (data mining) algorithms developed by the different academic actors. (iii) The use of these algorithms to improve our general knowledge of rare diseases.

The academic reputation and appeal

7 Prizes and Distinctions

Best papers

- Best paper in International Conference on Formal Concept Analysis (ICFCA) in 2014: *A Proposition for Combining Pattern Structures and Relational Concept Analysis* [464]
- The paper *Relation graphs and partial clones on a 2-element set* [479] of ISMVL 2014 was awarded the "Outstanding Contributed Paper Award" at the conference ISMVL 2015
- Best paper award at the 6th International Conference on Information Systems and Economic Intelligence (SIIE) in 2015 for the paper *Neural Networks for Proper Name Retrieval in the Framework of Automatic Speech Recognition* [119]

[1] `http://webloria.loria.fr/~mosmuk/orphamine/`

- In 2012, a best paper award was granted to the paper *Adapting Spatial and Temporal Cases* [497] published in the international conference on case-based reasoning (ICCBR)
- Best paper award in 27th ACM Symposium On Applied Computing in 2012: *Affine Invariant Shape Matching Using Radon Transform and Dynamic Time Warping Distance* [709]
- Best paper award in The 13th ACM Symposium on Document Engineering in 2013 for the paper *Document Noise Removal using Sparse Representations over Learned Dictionary* [697]

Distinctions

- Ekaterina Lebedava was awarded the *E. W. Beth Dissertation Prize* of the Association for Logic, Language and Information (FoLLI) for her thesis [807].

Participation to challenges

- We ranked 2nd among 9 teams for the ”Professionally produced music recordings” task of the 2015 Signal Separation Evaluation Campaign (SiSEC) [58].
- We ranked 4th among 25 teams and as the best European team for the 3rd CHiME Speech Separation and Recognition Challenge [206].

Invited Talks

- Invited keynote by Abdel Belaïd at the International Conference on Intelligent Systems Design and Applications, 2015, Marrakech.
- Philippe de Groote gave an invited talk at the Center for Logic and Philosophy of Science of the Tilburg University, on the occasion of Reinhard Muskens' 60th birthday.
- Samuel Cruz-Lara was invited for the International Conference on Engineering and Computer Education, New Trends in Engineering Education Workshop in 2011, Guimarães, Portugal.
- Claire Gardent gave a Keynote Talk at the 2nd International Conference on Statistical Language and Speech Processing, SLSP 2014, Grenoble (France).
- Emmanuel Vincent was invited in the 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics [277].

8 Editorial and organizational activities

Conference organization & chairing.

- Amedeo Napoli was the president of the organization committee and the chairman of “CLA 2011” (International Conference on Concept Lattices and their Applications).
- Marie-Dominique Devignes was the scientific chair of the conference and the scientific editor of the proceedings of “ECCB 2014” (European Conference on Computational Biology, 1100 participants).
- Amedeo Napoli and Chedy Raïssi were the Conference chairs of “ECML-PKDD 2014” (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 500 participants).
- Slim Ouni was the General Chair of “AVSP 2013” (International Conference on Auditory-Visual Speech Processing).
- Emmanuel Vincent was the General Chair of “CHiME 2013” (International Workshop on Machine Listening in Multisource Environments).
- Emmanuel Vincent was the General Chair of “HSCMA 2014” (Joint Workshop on Hands-free Speech Communication and Microphone Arrays).
- Bart Lamiroy was the General Chair of “GREC 2013” (IAPR International Workshop on Graphics Recognition, Bethlehem, PA, USA).
- Bart Lamiroy was program co-chair and organizing chair of “GREC’2015” (11th International IAPR Workshop on Graphics Recognition, Nancy)
- Abdel Belaïd was invited to serve as PC-Chair of ICDAR’13 (Washington, USA) and ICDAR’15 (Nancy).
- Sylvain Pogodalla co-chaired “LACL” (Logical Aspects of Computational Linguistics) in 2011.
- Kamel Smaïli has been a joint organizer and was the PC chair of “ICNLSP” (International Conference on Natural Language and Speech Processing) in 2015.
- Abdel Belaïd is co-editor of the International Journal on Document Analysis and Recognition since 2008.
- Claire Gardent was PC Chair for: ESSLLI 2016, *SEM 2016, SigDIAL 2012, ENLG 2011.

- Claire Gardent was Editor in Chief for Language and Linguistics Compass, Computational and Mathematical section.

Other

- Kamel Smaïli is the joint director of the LIA (Laboratoire International Associé) DATANET. This LIA concerns Big Data and includes two labs from France: Loria and CRAN and 5 universities from Morocco.
- Emmanuel Vincent is the elected president of ISCA Special Interest Group on Robust Speech Processing
- Claire Gardent is Chair of SIGGEN (ACL Special Interest Group on Natural Language Generation, 2015-2019)
- Philippe de Groote was elected in 2016 to be the future president of SIGMOL (ACL Special Interest Group for Mathematics of Language)
- Karl Tombre is vice-president of *Université de Lorraine*, in charge of partnerships and international affairs.
- Salvatore Tabbone is president of the GRCE (Groupe de Recherche en Communication Ecrite) since December 2010.

9 Services as expert or evaluator

Members of the department are regularly invited to national committees and to PhD committees. We highlight a few important facts, leaving a more complete description in team reports.

- 1 member of the CNU section 27
- 2 members of the CoCNRS (section 7 and 34)
- reviewer for the EU (FP7 project), reviewer of projects in the European FET projects
- Samuel Cruz-Lara is the Project Leader of the ISO Normalization committee for the MultiLingual Information Framework (*MLIF ISO 2012:24616*) TC37 / SC4[2]

10 Collaborations

Teams of the department are involved in many collaborations. Please refer to detailed reports of each team for a full list.

Collaborations trough European funding: H2020 CrossCult project, Chist-Era project AMIs with AGH University of Science and Technology Krakow (Poland) and University of DEUSTO Facultad de Ingenieria Bilbao (Spain).

Among international collaborations, we are involved in 3 collaborations funded by PHC: Zenon project with University of Nicosia, CMCU project with University of Tunis and Laboratoire d'Informatique de Grenoble, Van Gogh with Utrech University. We have several relations with SRI (Stanford): Inria Associate Team Snowflake and WebNLG.

Co-supervised PhD theses are also a fruitful source of collaboration in the department: we have 10 such co-supervision with, for instance, ENIT in Tunis, Algeria, Luxembourg, Inria teams (PANAMA, ALPAGE), Ircam, Télécom Paristech.

We also have close collaborations (other kind of projects or joint publications) with: UPC Barcelona, UFPE Recife, Brazil and several other University from South america,UQAM Montréal (CNRS PICS), National Institute for Informatics (NII, Tokyo, Japan); Netherlands; New-York University, USA; Chicago University, USA; Düsseldorf University; Saarland University (Saarbrucken); Bosphorus University (Istanbul); University of Sheffield (UK); Cork Institute of Technology; Northwestern University; CENATAV (Habana, Cuba); University of Bolzano; University of West Bohemia; ENSIT (Ecole Natioanle Supérieure des Ingénieurs de Tunis);

11 External support and funding

International projects

[2] http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37330

- **ERC EPS** 313360
- **DOVSA, Marie Curie IEF** (2010-2012);
- **H2020: CrossCult**;
- **ITEA Metaverse** (2009-2011);
- **ITEA Empathics Products** (2012-2015);
- **ITEA ModelWriter** (2014-2017);
- **AMIS** CHIST-ERA (2015-2018); **SMART** is leader of the project
- **ALLEGRO** Interreg Project (2010-2012);
- **EMOSPEECH** Eurostar project (2010-2013);
- **IFCASL**; *Programme franco-allemand en SHS*; (2013-2016);
- **Eureka label SCANPLAN** (2009-2012).

Notable national projects

- **DOD** (Oséo Consortium with two companies: Itesoft and Sagecom)

The department was also involved in 21 ANR projects (9 as leader), in 8 PEPS projects (4 as leader) and 3 PHC.

Several teams of the department participate to two PIA: **Equipex Ortolang** and **Equipex Istex**

Involvement with social, economic and cultural environment

Start-ups

- **Aetheris**[3] won the national I-lab competition for the most innovative start-ups. The goal of this company is to provide holistic travel planning methods that use the notion of data mining and preferences to provide good results in less time for the traveler. The engine used to answer the different travelers query is the result of technological transfers of our mining approaches.
- The **Harmonic Pharma** (`http://www.harmonicpharma.com`) start-up company (emerged from the **ORPAILLEUR** team in 2009) is still active and current members of **ORPAILLEUR** collaborate regularly with it.

Industrial contracts

- 4 PhD thesis with CIFRE grant (with XILOPIX, OCE, XRCE, eNovalys)
- **Studio MAIA**; 2014 - 2015; speech processing tools.
- **Venathec SAS**; 2014-2017; real-time control of wind farms.
- OCE Canon (2011-2013)
- Xilopix (2014-2017)
- Exameca (2014)
- ITESOFT (DOD Oséo project)
- **ECLEIR** project with the company eNovalys
- "BioIntelligence" project with "Dassault Systèmes" (2010-2014).
- Quaero project (Inria, Oséo, Technicolor, `http://www.quaero.org`, 2010-2013).
- The Orpailleur team was a member of the BioProLor consortium, which is composed of 5 enterprises and 7 academic research teams.

Mediation Members of the department were involved in mediation activities. We highlight a few facts here, more details are available in each team report.

- Several teams participate to Science & You, Nancy (June 2015). The exposition were opened during 4 days for children and large audience;
- The movie "*Je peux voir les mots que tu dis !*" (I can see the words you say!) has received the "best documentary" award at the "*Festival du film universitaire pédagogique de Lyon*";
- Maxime Amblard is member of the InterStice)i(editorial board; he was curator of the *Fascination ou aversion pour le numérique : encodage/décodage* exhibition (2011).

[3] `http://www.loria.fr/~raissi/Aetheris/`

The involvement in training through research

- We have strong links with the local master programs. Members of the department were at the origin of the implication of Lorraine University as a member of the Erasmus Mundus Master program in Natural Language Processing called “Language and Communication Technology” (LCT); most of the courses in this Master program are given by members of the NLPKD department. Some members gave master courses in other universities (MRPI in Paris, Tunis).
- We have a strong implication in PhD supervision: 45 PhD theses defended during evaluation period and 22 on-going PhD. 20 of the 67 theses are co-supervised either with industrial or with academic partners.

Computational Epistemic Logic in Lorraine

Synopsis

1 Team Composition

Permanents

Hans van Ditmarsch (DR CNRS, arrived Dec 2012)

	PR	MCF	DR	CR	Total
2011					
2016			1		

Post-docs, and engineers

Pere Pardo (CRNS, post-doc Dec 2014- Dec 2016) (ERC)
Bastien Maubert (CNRS, post-doc Sept 2014- Feb 2015) (ERC)
Petar Iliev (CNRS, post-doc, Sept 2013- Sept 2015) (ERC)
Sophia Knight (CNRS, post-doc Sept 2013- Jan 2016) (ERC)

Doctoral students

Aybüke Özgün (UL, 2013-...), van Ditmarsch (ERC, University of Amsterdam, 2013-2017) Zeinab Bakhtiari (UL, 2014-...), van Ditmarsch (ERC, 2014-2017)

Phd's defended	0	On-going PhD's	2

Team evolution

The CELLO team was born in 2013, shortly after the arrival of Hans van Ditmarsch at LORIA. It is the vehicle for the execution of the European Research Council Starting Grant EPS (Epistemic Protocol Synthesis) 313360. By the start of the academic year 2013/2014 three people had joined the project: Sophia Knight, postdoc, Aybüke Özgün, PhD student, and Petar Iliev, postdoc, and also a (12 months)

visiting PhD student Jie Fan, Peking University. Around the start of the academic year 2014/2015 more people joined the project: Bastien Maubert, postdoc, Zeinab Bakhtiari, PhD, Pere Pardo, postdoc, and, by a part-time form of assignment (20% during the academic year 2014/2015), François Schwarzentruber from IRISA Rennes. A 2 months visiting PhD student in 2014 was James Hales, University of Western Australia, and a long-term (8 months) visiting PhD student in 2015 was Rahim Ramezanian, Sharif University of Technology, Iran. Currently, a new arrival is envisaged by May 2016: Bouke Kuijer, postdoc. Currently, all CELLO members are EPS project members. It is conceivable that the research group CELLO may attract in the coming years yet other members but who are not involved in the project, thus establishing the group as a permanent feature at LORIA.

2 Life of the team

The CELLO team is the vehicle for the execution of the European Research Council Starting Grant EPS (Epistemic Protocol Synthesis) 313360. Hans van Ditmarsch, also the principal investigator of the ERC project, is head of the team.

3 Research topics

Keywords

Modal logic, knowledge and belief, epistemic protocols, multi-agent systems, information-theoretic cryptography

Research area and main goals

Given my current state of knowledge, and a desirable state of knowledge, how do I get from one to the other? Is it possible in principle to reach the desirable state of knowledge, i.e., does it make sense at all to start trying to obtain the desirable state? If I know it is impossible to obtain, there is no use trying. But even if I know that it is possible in principle, is there a way to approach the desirable state in steps or phases, i.e., can I iteratively construct an epistemic protocol to achieve the desirable state? And can this be done with some or with full assurance that I am getting closer to the goal? Such problems become more complex if they involve more agents. The knowledge states of agents may be in terms of knowledge properties of other agents. Such assumed knowledge properties may be incorrect, or the agents may act at unpredictable or unknown moments, or with delayed or faulty communication channels, as typically in asynchronous systems.

The focus of much research in dynamic epistemic logic, and more generally in epistemic and temporal modal logics, is analysis: given a well-specified input epistemic state, and some well-specified dynamic process, compute the output epistemic state. We focus on synthesis: given a well-specified input epistemic state, and desirable output (typically less well specified), find the process transforming the input into the output. The process found is the epistemic protocol. We are aided by recent advances in logics for propositional quantification. Areas of specific interest are protocols for secure communication, protocol languages, and agency. The general goal of this research group is epistemic protocol synthesis for synchronous and asynchronous multi-agent systems, by way of using and developing dynamic epistemic logics. More specifically, the following themes are in our scope:

- Epistemic protocol synthesis from initial state to goal state
- Knowledge progression
- Protocol synthesis in asynchronous systems

- Security protocol synthesis
- Heuristics for epistemic planning
- Logics for quantifying over information change
- Planning with unawareness
- Protocol languages
- Agency in logics for epistemic protocol synthesis

4 Main Achievements

The CELLO team has only existed for a few years and there are no important facts to report up to date.

5 Research activities

Propositional quantification in modal logic

Description This research is involved with adding propositional quantifiers to modal logics for information change. Examples are arbitrary public announcement logic and refinement modal logic. Let us consider refinement modal logic. A refinement is like a bisimulation, except that from the three relational requirements only 'atoms' and 'back' need to be satisfied. Refinement modal logic contains a new operator 'all' in addition to the standard modalities 'box' for each agent. The operator 'all' acts as a quantifier over the set of all refinements of a given model. As a variation on a bisimulation quantifier, this refinement operator or refinement quantifier 'all' can be seen as quantifying over a variable not occurring in the formula bound by it. The logic combines the simplicity of multi-agent modal logic with some powers of monadic second-order quantification. The logic can be axiomatized, and has high complexity. However, it is equally expressive as base modal logic. Whereas the arbitrary public announcement logic (APAL), also mentioned above, is more expressive. Various developments on both logics are ongoing.

Main results Together with Philippe Balbiani (IRIT) we made a novel, simplified, completeness proof of the logic APAL (with quantification over announcements) [16]. The works [2, 1, 17] describe progress in refinement modal logic consisting of axiomatizations, semantical results, and complexility results. In [11, 27] we review undecidability results for various logics with quantification over announcements (and some progress on related logics). Obviously, decidable versions of such logics are more suitable for practical applications.

Cards cryptography

Description Information-based cryptography employing card deals allows for agents to share secret keys as a result of protocols for public communication. In the generalized Russian cards problem, Alice, Bob and Cath draw a, b and c cards, respectively, from a deck of size a+b+c. Alice and Bob must then communicate their entire hand to each other, without Cath learning the owner of a single card she does not hold. Unlike many traditional problems in cryptography, however, they are not allowed to encode or hide the messages they exchange from Cath. The problem is then to find methods through which they can achieve this.

Main results The works [3, 4] focus on geometric techniques to ensure such information-based secure communication, where the main results are protocols consisting of three or four steps (instead of the two-announcement protocols, one by the sender and a response by the receiver, that are better known in this area), and protocols wherein the eavesdropper (spy) can hold more than one card.

Gossip protocols

Description A gossip protocol is a procedure for spreading secrets among a group of agents, using a connection graph. We investigate epistemic (or knowledge-based) gossip protocols, where the agents themselves instead of a global scheduler determine whom to call. They are also distributed protocols. Investigations involve protocols for synchronous systems, but also for asynchronous (more fully distributed) systems. Yes another various is to give such gossip protocols a dynamic twist by assuming that when a call is established not only secrets are exchanged but also telephone numbers. In that setting, distributed dynamic gossip protocols can be characterized them in terms of the class of graphs where they terminate successfully.

Main results This area is currently actively investigated by members of the CELLO team and the results up to date are few, mainly work in progress or under submission to journals, and include [14, 15] (author Attamah was the Liverpool PhD student of van Ditmarsch and graduated last December). These two publications describe some knowledge-based gossip protocols such as the 'Learn New Secrets' protocol wherein agents call another agent if they do not know the secret of that agent.

Scientific production and quality

6 Synthesis of publications

[19, 16, 26, 29, 28, 21, 13, 7, 8, 2, 5, 3, 4, 20, 14, 11, 27, 24, 30, 25, 6, 12, 10, 15, 9, 1, 17, 18, 23, 22]

	2011	2012	2013	2014	2015	2016
PhD Thesis						
H.D.R						
Journal			3	5	2	
Conference proceedings			6	10	4	
Conference proceedings (non selective)						
Book chapter						
Book (written)						
Book or special issue (edited)						
Patent						
General audience papers						

7 List of top journals in which we have published

Artificial Intelligence Journal (1) [7]
Theoretical Computer Science (2) [3, 2]
Designs, Codes, and Cryptography (1) [4]
Information and Computation (1) [1]

8 List of top conferences in which we have published

AAMAS (2) [22, 11]
AiML (2) [24, 21]
TARK (1) [25]
IJCAI (1) [17]

9 Software

Not applicable.

The academic reputation and appeal

- **ANR:** Hans van Ditmarsch is a participant in ANR DynRes, a project coordinated by Didier Galmiche. This project ran until 2015.
- **MEALS:** Hans van Ditmarsch is a participant in the project MEALS, `http://www.meals-project.eu/`, the local LORIA branch of this project is coordinated by Stephan Merz. This project ran until 2015.
- **ERC:** The activities of all CELLO team members coincide with the activities of the ERC project EPS 313360 (EPS = Epistemic Protocol Synthesis). This 5-year project runs from 1 February 2012 – 1 February 2018.

10 Collaborations

A substantial part of progress in the CELLO team is due to continued international collaborations, often with members of the external support group of the ERC Starting Grant EPS 313360. There are also collaborations with people who are not in the external support group. Among long-term visits of the kind are: Hans van Ditmarsch' regular visits to Institute of Mathematical Sciences, Chennai (where he holds an unpaid associate position), (host Ramanujam), visit to Peking University (host Yanjing Wang), the stay of Tim French and his PhD student James Hales (Australia) at LORIA (two months in 2014), the 8 month visit of PhD student Rahim Ramezanian (Iran) (2015). Short-term visitors to LORIA include: Jerome Lang, Abdallah Saffidine, Wiebe van der Hoek, Maduka Attamah, Andreas Herzig, Philippe Balbiani, Thomas Bolander, Mikkel Birkegaard Andersen, Martin Holm Jensen, Rohit Parikh, Thomas Agotnes, Sabine Frittella, Alessandra Palmigiano, Yongmei Liu, Liangda Fang, Yi Wang, Katsuhiko Sano, Minghui Ma, Mehrnoosh Sadrzadeh, Raul Fervari, Jan van Eijck. Hans van Ditmarsch acted at the opening workshop for the logic group at Andalucia Tech, a collaboration between the Universities of Sevilla and Malaga, November 2014. Pere Pardo, currently postdoc at LORIA, held a prior postdoc position in the logic group at the University of Sevilla.

11 External support and funding

ERC: Hans van Ditmarsch is Principal Investigator of ERC project EPS 313360 (EPS = Epistemic Protocol Synthesis). This 5-year project runs from 1 February 2012 – 1 February 2018.

Speech Modeling for Facilitating Oral-Based Communication

This reports concerns the PAROLE team (from 2011 until June 2014) and the MULTISPEECH team (since July 2014). It is limited to the perimeter of the MULTISPEECH research topics. Thus, as Christophe Cerisara has left the PAROLE team in December 2011 to create the SYNALP team, his activities, as well as those of Christian Gillot are described in the SYNALP report. Also, as Kamel Smaïli and David Langlois left the PAROLE team in December 2014 to create the SMART team, their activities as well as those of Sylvain Raybaud and Motaz Saad are presented in the SMART report.

Synopsis

1 Team Composition

Permanents

Denis Jouvet (HDR, DR Inria), Yves Laprie (HDR, DR CNRS), Anne Bonneau (CR CNRS), Dominique Fohr (CR CNRS), Antoine Liutkus (CR Inria, arrived Jan. 2014), Emmanuel Vincent (HDR, CR Inria, arrived Jan. 2013), Martine Cadot (MCF, UL, arrived Feb. 14), Vincent Colotte (MCF UL), Joseph di Martino (MCF, UL, left Dec. 15), Irina Illina (MCF, UL), Odile Mella (MCF, UL), Slim Ouni (HDR, MCF, UL), Agnès Piquard-Kipffer (MCF, UL).

	PR	MCF	DR	CR	Total
2011		6	2	2	10
2016		6	2	4	12

Post-docs, and engineers

Benjamin Elie (post-doc, Oct. 2013 - Feb. 2016), Juan Andrés Morales Cordovilla (post-doc, Mar. 2015 - Feb. 2016), Camille Fauth (post-doc, Mar. 2013 - July 2014), Thibaut Fux (post-doc, Sept. 2014 - Aug. 2015), Emad Girgis (post-doc, Nov. 2014 - June 2015), Sucheta Ghosh (post-doc, Sept. 2015 - Dec. 2016), Larbi Mesbahi (post-doc, until Oct. 2011), Ingmar Steiner (post-doc, Jan. 2011 - April 2012), Asterios Toutios (post-doc, until April 2011).
Ilef Ben Farhat (engineer, Nov. 2013 - Oct. 2015), Julie Busset (engineer, June 2014 - May 2015), Antoine Chemardin (engineer, Oct. 2014 - Dec. 2015), Sara Dahmani (engineer, Dec. 2014 - Nov. 2016), Sebastien Demange (engineer, 2012), Célie Depardieu (engineer, June - July 2014) Valérian Girard

(engineer, May 2015 - May 2016), Jean-François Grand (engineer, Jan. 2011 - Dec. 2012), Caroline Lavecchia (engineer, Nov. 2011 - Dec. 2012), Jeremy Miranda (engineer, Sept. 2013 - Mar. 2014), Luiza Orosanu (engineer, Nov. 2011 - Dec. 2012), Yann Salaün (engineer, Dec. 2012 - Nov. 2014), Aghilas Sini (engineer, Nov. 2014 - Dec. 2016), Sunit Sivasankaran (engineer, Mar. 2015 - Feb. 2017).

Doctoral students

Uptala Musti (Inria, Nov. 2009 - Feb. 2013, [36]), Julie Busset (ANR ARTIS, 2010 - Mar. 2013, [32]) Alex Mesnil (ENS, Nov. 2012 - has quit in Aug. 2013), Arseniy Gorin (Inria, Nov. 2010 - Nov. 2014, [33]), Dung Tran (Inria, Dec. 2012 - Nov. 2015, [40]), Luiza Orosanu (FUI RAPSODIE, Dec. 2012 - Dec. 2015, [37]), Imran Sheikh (ANR CONTNOMINA, Feb. 2014 - ...), Baldwin Dumortier (Industrial contract VENATHEC, Sept. 2014 - ...), Aditya Arie Nugraha (Inria, Jan. 2015 - ...), Ken Deguernel (ANR DYCI2, Mar. 2015 - ...).

And co-supervised theses: Imen Jemaa (Co-tutelle ENIT/UL, PhD in Feb. 2013, [35]), Fadoua Bahja (COADVISE-FP7, PhD in Jul. 2013, [31]), Xabier Jaureguiberry (Institut Télécom & Institut Carnot, PhD in June 2105, [34]), Nathan Souviràa-Labastie (Univ. Rennes 1, PhD in Nov. 2015, [39]), Othman Lachhab (COADVISE-FP7, to be defended in 2016), Quan V. Nguyen (Inria, LARSEN team, 2014 - ...), Amal Houidhek (Co-tutelle ENIT/UL, 2015 - ...), Imene Zangar (Co-tutelle ENIT/UL, 2015 - ...).

Phd's defended	5	On-going PhD's	4
Co-supervised Phd's defended	4	On-going co-supervised PhD's	4

Team evolution

Compared to the PAROLE project, MULTISPEECH has a stronger focus on speech signals. On the one hand, activities related to syntax have given rise to the LORIA team SYNALP (*Statistical and Symbolic Natural Language Processing*) headed by C. Cerisara, former member of PAROLE until Dec. 2011, and activities related to speech translation have led to the creation of the LORIA team SMART (*Statistical Machine Translation*) by K. Smaïli and D. Langlois, who were members of PAROLE until Dec. 2013. On the other hand, activities related to speech signal processing and to Bayesian speech processing have expanded following the arrival of E. Vincent (CR Inria) in Jan. 2013, who was previously with the METISS project-team in Rennes, and the recruitment of A. Liutkus (CR Inria) in Jan. 2014.

Left the team *C. Cerisara (2011)*, *K. Smaïli (2013)*, *D. Langlois (2013)*, J. Di Martino (2015).

Joined the team E. Vincent (2013), A. Liutkus (2014), M. Cadot (2014).

2 Life of the team

There are regular team meetings/seminars, on average, one every two weeks. Most of them are dedicated to scientific presentations followed by discussions. A few team meetings are limited to permanent members to discuss research orientations, budget, ... Invited speakers make presentations either in the team seminars or in the department seminars. The last full day team seminar took place in Dec. 2015.

Two mailing lists are used for team communication, one is for the whole team, the other one is limited to permanent members. A web site is available. Following the creation of the MULTISPEECH team-project, a new website was set up.

3 Research topics

Keywords

Speech, machine learning, statistical methods, perception, speech recognition, multimodal speech processing, speech synthesis, audio, linguistics, phonetics, natural language, signal processing, modeling, foreign language learning, articulatory modeling.

Research area and main goals

Speech is the scientific object studied. The main objectives are a better understanding of how human produce and perceive speech, and the enhancement of communication channels for efficient vocal interactions via automatic processing (speech enhancement, synthesis and recognition).

The physical nature of the speech is taken into account through an explicit modeling of speech production and perception phenomena. Articulatory modeling exploits the geometrical form of the vocal tract, and makes the link between the shape and position of the articulators and the speech signal. Multimodal speech synthesis considers both the audio and the visual modalities of the speech signal. With respect to the production and the perception of speech, the relations between articulatory features, acoustic cues and human perception are investigated.

Many automatic speech processing aspects are based on statistical modeling. This includes acoustic, lexical and language modeling. Statistical modeling relies on efficient machine learning techniques and takes benefit of large corpora. Statistical models are used for source separation, automatic speech recognition, and speech generation.

Estimating and dealing with uncertainty is a theme introduced with the MULTISPEECH project. The uncertainty due to the variability of the speech production processes and of the transmission channels makes speech processing an arduous task. It is important to estimate and exploit the uncertainty. This is now the case in robust speech recognition where the uncertainty estimated during the source separation process is used in decoding for improved performance. Confidence measures in speech recognition gives an idea of the reliability of the recognized words. We aim at developing similar measures for speech segmentation and prosodic features.

Although large speech-only and text-only corpora are now available, specific acoustic, visual and articulatory data are necessary to design and validate the algorithms and approaches. Hence the collection of specific data and the development of associated tools to process them.

4 Main Achievements

Articulatory modeling has been investigated further, and in collaboration with the IADI laboratory (at Nancy hospital) new Magnetic Resonance Imaging (MRI) acquisition protocols have been developed and articulatory data has been collected.

A first version of audio-visual speech synthesis has been developed relying on bimodal (audio + visual) speech unit concatenation. Then, a new approach was proposed for an efficient animation of 3D models from a limited number of points. More audio-visual data is under collection through various acquisition devices.

Source separation is applied both for extracting speech from a mixture signal and for robust speech recognition. Various approaches have been developed corresponding to the FASST and KAM source separation frameworks. A new multichannel approach relying on deep neural networks has been developed and used in two challenges in 2015, leading to good results.

In parallel a learning-based uncertainty estimation and propagation paradigm was elaborated, leading to a 30% word error rate reduction on the CHiME2 dataset compared to conventional decoding (without

uncertainty).

A French-German bilingual learner speech corpus was designed, collected and manually annotated. It was used for investigating various non-native speech phenomena. This corpus will be made available for research at the end of the IFCASL project.

5 Research activities

Explicit modeling of speech production and perception

Description. Speech signals result from the movements of articulators. A good knowledge of their position with respect to sounds is essential to improve articulatory speech synthesis and the relevance of the diagnosis and feedback in computer assisted language learning. Production and perception processes are interrelated, so a better understanding of how humans perceive speech will lead to more relevant diagnoses in language learning as well as pointing out critical parameters for expressive speech synthesis. The expressivity, which is a long term goal, translates into both visual and acoustic effects that must be considered simultaneously to produce expressive speech synthesis. These research axis often require the acquisition of specific data: articulatory data, multimodal speech and non-native speech.

Main results. ***Articulatory modeling and synthesis.*** Our research on acoustic to articulatory inversion is based on the analysis by synthesis paradigm. An extension of this approach to using cepstral coefficients as input [32, 98, 97] has shown the importance of improving the quality of the direct model; thus articulatory synthesis became an important objective. The determination of the instantaneous area functions of the vocal tract requires the construction of articulatory models from X-ray films [157, 158, 161, 233], or from MRI images and films [156, 279]. The Xarticulators software was developed to construct articulatory models from images of the mid-sagittal slices (MRI) or projections (X-ray) of the vocal tract. The resulting 2D articulatory models [159] are far more complete that other models. Other contributions are about articulatory copy synthesis from X-ray films [160, 56], and acquisition of more precise MRI articulatory data in collaboration with the IADI laboratory [278, 280]. We now have quite a complete environment in the domain of articulatory synthesis, which is seldom found in speech laboratories.

Audio-visual speech synthesis. Based on our text-to-speech synthesis system (SoJA), we have developed a bimodal acoustic-visual synthesis technique that concurrently generates the acoustic speech signal and a 3D animation of the speaker's face, by concatenating bimodal diphone units consisting of both acoustic and visual information [59, 175, 220, 36, 38]. In the visual domain, we had mainly focused on the dynamics of the face rather than on rendering. Thus, we have then developed an algorithm to animate the 3D model of human face from a limited number of markers [259]. We are focusing lately on expressive audiovisual speech synthesis where bimodal signal is augmented by the facial expressions. We are actively conducting experiments in evaluating several acquisition techniques for multimodal speech data [83] using marker-based stereovision techniques and markerless techniques based on kinect-like devices. Our purpose for the future is to acquire different kind of data for different modalities and to combine them.

Categorization of sounds and prosody. We kept investigating on the early predictors of future reading skills. We examined whether reading level at age 8 could be predicted on the basis of phonemic discrimination at age 5, a skill very rarely examined in longitudinal studies [62, 63]. We continued the work concerning the acquisition of language by hard-of-hearing children via cued speech. Three digital books had been elaborated and used by children (3 to 12 years old) [262, 64]. Within the framework of the IFCASL project, a bilingual speech corpus of French and German language learners was designed and recorded from about 100 speakers [114, 51]. A large part was manually annotated. This is the first corpus of this type and size for this language pair. It will be made available for research at the end of the IFCASL project. The inter-annotator agreement was studied [170], and the corpus was used to analyze various

non-native pronunciation phenomena [143, 94, 92, 230] in the French-German pair. The results aim at being used to create individualized training and feedback for foreign language learners.

Statistical modeling of speech

Description. Statistical approaches are common for processing speech and they achieve performance that makes possible their use in actual applications. However, speech recognition systems still have limited capacities (for example, even if large, the vocabulary is limited) and their performance drops significantly when dealing with degraded speech, such as noisy signals and spontaneous speech. Source separation based approaches are investigated both for enhancing the speech signal and for noise robust speech recognition. Dealing with spontaneous speech and handling new proper names are two critical aspects that are tackled, along with the use of statistical models for speech production. Deep Neural Network (DNN) modeling is used in several topics including source separation and acoustic modeling for speech recognition.

Main results. ***Audio source separation and speech enhancement.*** Audio source separation is a key issue for the enhancement of speech in noise or singing voice in a musical accompaniment. Building upon the state-of-the-art Gaussian modeling framework, we showed how to exploit repeated excerpts of the same music or repeated versions of the same sentence uttered by different speakers [68, 39] and prior information regarding the source positions and the room characteristics [43, 50, 53]. Recently, we introduced new modeling frameworks for source separation based on alpha-stable distributions [163, 164], deep neural networks [58], and fusion of multiple separation systems [140, 54, 34]. Our multichannel deep neural network-based system [58] led to good results in 2015 evaluation campaigns [179, 206] (see 11). These results show that we stand among the top teams worldwide in the field of source separation and speech enhancement thanks to the combination of deep learning with our expertise in signal processing. In order to exploit our know-how for industrial applications, we investigated issues such as audio quality [219] and smart user interfaces [197]. We also disseminated these results by means of review papers [71], plenary talks [277, 269], and software [266].

Acoustic modeling. A more refined modeling was investigated for Hidden Markov Models (HMM) with Gaussian Mixture Model (GMM) densities. Class-based modeling was improved through the introduction of a classification tolerance marging [152, 237] and a discriminative criterion [126]; and then used to introduce some structure in the components of GMM densities for better modeling speaker trajectories [127, 128, 130, 33]. Combining forward-based and backward-based decoders was investigated for speech recognition [148, 147] and for unsupervised training [149]. For robust speech recognition, the performance of several types of features has been investigated [124], as well as a new normalization which operates on the log-likelihood scores [229].

Pronunciation modeling. A Conditional Random Field (CRF) based approach was developed for grapheme-to-phoneme conversion, which provides better results than the Joint-Multigram Models (JMM) approach for generating pronunciation variants [132, 131]. It was also shown that combining both approaches improves performance [146]; and a special attention was paid to proper names [236]. We have also investigated the usage of Wiktionary data [145], as well as introducing probabilities of pronunciation variants dependent on the speaking rate [144].

Language modeling. A neural-network based approach was proposed for selecting the most relevant lexicon for a speech transcription task [150]. With respect to language modeling, we investigated the selection of relevant training data [172] and the introduction of semantic information through the Random Indexing paradigm [122]. Syllable units which provide good phonetic decoding performance [182, 181] were merged with word units in hybrid language models [183][184, 37]. A word context similarity was defined for introducing new words in a language model [185]. For dealing with diachronic documents,

the lexical context and the temporal information are used to predict proper-name (PN) words that should be added to the recognition lexicon for improved speech recognition performance [178, 133, 134]. We focused on the continuous space word representation using neural networks. Experimental results suggest that the proposed method that models the semantic and lexical context of proper names (PN) can be useful for PN retrieval [118, 119]. Entity-Topic models have been proposed as extensions of Latent Dirichlet Allocation (LDA) to specifically learn relations between words, entities (PNs) and topics [204, 201]. Similarly to speech, music can be modeled at the symbolic levels using probabilistic models akin to language models. We pursued our pioneering work on this topic, with a particular focus on long-term structures [91], and variations across musical genres [65].

Speech generation by statistical methods. Voice conversion techniques were applied to enhance oesophageal voices. In addition to the statistical aspects of the voice conversion approaches, signal processing is critical for good quality speech output. An algorithm for pitch detection has been proposed, that can operate in real-time [42, 31]. Another practical issue consists in estimating / predicting the phase spectrum for facilitating the re-synthesis process. A comparative study was carried out [47] based on an Inria patent.

Uncertainty estimation and exploitation in speech processing

Description. Our general objective is to quantify the confidence (or uncertainty) in the output of a given speech processing technique and to exploit it for further processing. We are interested in such information regarding for instance speech enhancement and source separation techniques. For what concerns phonetic segment boundaries and prosodic parameters, however, no such information is available yet. Hence the challenge of heading for confidence measures on the speech-text alignment results, and on the prosodic parameters.

Main results. ***Uncertainty and acoustic modeling.*** Speech enhancement is useful to deal with noisy speech data, but some distortion or "uncertainty" remains in the enhanced signal which must be estimated. We proposed to estimate uncertainty using a trained regressor [225, 70, 40] instead of fixed mathematical approximations. Combined with a plain GMM-HMM acoustic model, we achieved 30% relative word error rate reduction with respect to conventional decoding (without uncertainty) on small and medium noise-robust ASR vocabulary tasks, that is about twice as much as the best existing uncertainty estimator and propagator. More recently, we started working on the propagation of uncertainty for deep neural network-based ASR [72] and for speaker recognition [199]. In order to motivate further work by the community, we studied the issue of data collection [76] and created the series of CHiME Speech Separation and Recognition Challenges. This has made it possible to compare various techniques in realistic multichannel noise conditions and contributed to the development of new robust ASR techniques. We organized two editions in 2013 [228] and 2015, the latest of which attracted 26 participating teams worldwide. Finally, we partnered with other teams in Inria to exploit the knowledge gained regarding uncertainty in other fields, namely industrial acoustics and robotics [79].

Uncertainty and speech recognition. In the RAPSODIE project, interviews with hard-of-hearing people were conducted to determine the best way of displaying speech recognition results and to examine which factors the participants consider as being helpful for a better understanding. Hybrid language models were used and confidence measures exploited to adjust the way of displaying the results [193].

Uncertainty and phonetic segmentation. With respect to non-native speech-text alignment, we analyzed the impact of adding pronunciation variants in the lexicon [151, 45] and the reliability of the boundaries with respect to phonetic classes [171]. The COALT software [121] was developed for comparing several sets of phonetic segmentations. An automatic extension of the pronunciation lexicon combining native models from both languages (mother tongue and foreign language) was proposed to better align non-native speech [123]. However, before any automatic speech-text alignment, it is important to check

that the pronounced utterance corresponds to the expected one [238, 180], especially when designing automatic exercises. The speech-text alignment of spontaneous speech was investigated and led to the development of a dedicated web application ASTALI. In the ORFEO project many corpora were aligned, unfortunately having different annotation conventions [120]. The impact of the frame shifts (5 ms vs. 10 ms) on automatic phone segmentation was also investigated [89, 142].

Uncertainty and prosody. The automatic detection of the prosodic structures of speech utterances was investigated [87]; and then, the links between these structures and manually annotated punctuation marks were investigated [88]. A study has begun on the use of prosodic parameters to identify discourse particles in French [102]. The usage of prosodic parameters was also investigated, in addition to linguistic parameters, for the determination of the modality of sentences [187, 186]. As the fundamental frequency is one important prosodic feature, several approaches for its computation have been implemented in the JSNOORI software.

Scientific production and quality

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis			4	1	4	
H.D.R			1			
Journal	2	3	6	8	6	7
Conference proceedings	24	22	23	34	44	9
Conference proceedings (non selective)						
Book chapter	4			1		
Book (written)						
Book or special issue (edited)			1		1	
Patent						1
General audience papers						

7 List of top journals in which we have published

- IEEE Signal Processing Magazine (1) [71]
- IEEE Transactions on Audio, Speech and Language Processing (6) [65, 68, 70, 54, 52, 58]
- IEEE Transactions on Signal Processing (2) [43, 57]
- Journal of the Acoustical Society of America (3) [49, 69, 41]
- IEEE Signal Processing Letters (1) [67]

8 List of top conferences in which we have published

- ASRU (IEEE Auto. Speech Recognition & Understanding Workshop) (4) [84, 135, 206, 228]
- AVSP (Int. Conf. on Auditory-Visual Speech Processing) (4) [173, 174, 175, 215]
- EUSIPCO (European Signal Processing Conference) (6) [110, 124, 153, 158, 208, 218]
- ICASSP (IEEE Int. Conf. on Acoustics, Speech and Signal Processing) (20) [106, 109, 116, 146, 152, 163, 165, 196, 197, 201, 209, 216, 222, 223, 224, 225, 216, 111, 116, 202]

- ICPhS (Int. Congress of Phonetic Sciences) (5) [89, 94, 112, 159, 230]
- INTERSPEECH (Conf. of the Int. Speech Communication Association) (23) [83, 92, 107, 118, 122, 128, 131, 139, 144, 148, 149, 160, 169, 182, 184, 188, 190, 198, 199, 204, 220, 221, 229]
- ISSP (Int. Seminar on Speech Production) (10) [93, 95, 97, 156, 157, 161, 207, 210, 211, 212]
- JEP (Journées d'Etudes sur la Parole) (10) [113, 129, 133, 183, 236, 237, 183, 238, 189, 200]
- MLSP (IEEE Int. Workshop on Machine Learning for Signal Processing) (3) [137, 138, 219]
- WASPAA (IEEEWorkshop on Appli. of Signal Proc. to Audio and Acoustics) (2) [164, 166]

9 Software

Several tools have been improved or developed during the evaluation period. Speech processing tools deal with speech transcription (ANTS), audio sources separation (FASST and KAM), speech-text alignment (ASTALI) and text-to-speech synthesis (SoJA). Another set aims at visualizing various aspects of speech data: speech audio signal (JSNOORI), ElectroMagnetographic Articulography (EMA) data (VisArtico) and speech articulators from X-ray or Magnetic Resonance Imaging (MRI) images (Xarticulators). Platforms were also enhanced or developed for acquiring multichannel speech audio signals (JCorpusRecorder), EMA data and MRI data.

Some tools have been largely used outside the team, as for example FASST (downloaded more than 500 times) and VisArtico (used by more than 130 researchers around the world). The EMA acquisition platform has also been usd by several laboratories.

10 Miscellaneous

Within the project IFCASL, a French-German bilingual learner corpus was designed and recorded [114, 51]. A large part of the corpus has been manually annotated at the word and phoneme levels. This corpus will be made available for research at the end of the project.

After two successful editions in 2011 and 2013, we organized the third edition of the 'CHiME' Speech Separation and Recognition Challenge in 2015 [84]. 25 teams have participated to this third challenge.

In 2015, a hundred full length musical tracks were gathered, whose license permit scientific studies. The tracks were mixed semi-professionally. The resulting Demixing Secrets Dataset 100 is intended to be a major reference in the field of source separation and in the study of singing voice. An international evaluation campaign using this corpus for source separation has been undertaken (SiSEC 2015) and will be pursued in the near future.

The academic reputation and appeal

The team has been, and still is, involved in many collaborative projects, including European and national projects. A full list is given below in the section "External support and funding", and detailed in Annex. Among these collaborative projects, the team is leader in five ANR projects.

11 Prizes and Distinctions

Members of the team received the "best documentary" award at the "*festival du film universitaire pédagogique de Lyon*" (April 2012, www.canal-u.tv.); the best poster prize at EWEA 2015 (European Wind Energy Association 2015 Annual Event) [108], and a best paper award at SIIE 2015 (6th Int. Conference on Information Systems and Economic Intelligence) [119].

We ranked 2nd among 9 teams for the ”Professionally produced music recordings” task of the 2015 Signal Separation Evaluation Campaign (SiSEC) [58]; and 4th among 25 teams, and best European team, for the 3rd CHiME Speech Separation and Recognition Challenge [206].

Members of the team have been invited for delivering keynotes (e.g. at WASPAA 2015) and lectures in various conferences, workshops, spring/summer schools, and thematic days; some recent invited talks being at Dagstuhl spring school 2014, ASA meeting 2015, LVA/ICA 2015, as well as a tutorial at ICASSP’2014.

12 Editorial and organizational activities

A member of the team is the elected president of the ISCA Special Interest Group on Robust Speech Processing; and he is also chairing the Challenges Subcommittee of the IEEE Technical Committee on Audio and Acoustic Signal Processing.

Members of the team were involved as general chair (AVSP’2013; CHiME’2013; HSCMA’2014), program chair (LVA/ICA’2015) and area chairs (ISMIR-2011-2014; ICASSP 2015, 2016; INTERSPEECH 2013, 2015; ...). Some are also involved in the organization of evaluation campaigns (2nd and 3rd CHiME challenges; SiSEC’2015), of workshops and of special sessions.

13 Services as expert or evaluator

Team members frequently review articles for journals (such as IEEE Trans. ASLP, JASA, Journal of Phonetics, Computer Speech and Language, ...) as well as for conferences (ICASSP, INTERSPEECH, JEP, ...). They also report on project proposals (e.g., Europe FET, ANR) and participate in many PhD committees.

Several members belong to journal editorial boards, such as for IEEE Trans. ASLP, Speech Communication, EURASIP ASMP, ...

Some members are involved in local committees related to LORIA, Inria or doctoral school, as well as in external committees: e.g., member of the National Council of Universities (CNU section 61), member of the *Conseil du Pôle Scientifique* AM2I of University of Lorraine, ...

14 Collaborations

During the evaluation period the team had many collaborations, including with:

- Saarland University (Saarbrucken): learner corpus [114, 51], non-native speech analysis [143, 230], co-organization of events (Spring school 2014, workshop 2015).
- Bosphorus University (Istanbul): extension of tensorial decomposition framework to datasets with heterogenous sampling rates [168]. Bayesian methods for alpha-stable models [67]
- National Institute for Informatics (NII, Tokyo, Japan): source separation, including organization of the SiSEC 2015 evaluation campaign [179] among others [137, 53, 223]
- Mitsubishi Electric Research Labs (MERL, Boston, USA): source separation [80]
- University of Sheffield (UK) and Mitsubishi Electric Research Labs (MERL, Boston, USA): dataset collection and organization of the 2nd and 3rd CHiME challenges [76, 228]
- Cork Institute of Technology: source separation. Joint work on statistical models for source separation and music/voice processing [165, 167, 117]
- Northwestern University: joint work on music source separation [167, 117, 245]
- CENATAV (Habana, Cuba): robust speaker recognition [198, 199]
- *Laboratoire Informatique d’Avignon* (LIA, Avignon): Automatic speech transcription for diachronic documents [134, 178, 133, 201, 231]

- Télécom Paristech (Paris): fusion for source separation; co-supervision of PhD [138, 140, 139]
- *Laboratoire de Phonétique et Phonologie* (Paris): articulatory data and improvement of articulatory synthesis.
- *Laboratoire Psychologie de la Perception* (Paris): early assessment of pre-reading skills [62].
- Ircam (Paris): statistical music modeling, co-supervision of K. Deguernel
- Laboratoire IADI (*Imagerie Adaptative Diagnostique et Interventionnelle*, INSERM U947, Nancy): cooperation since 2011 on MRI acquisition (protocols and data) [279].
- ATILF (Nancy): prosody Prosodic structures of speech utterances [87, 88], discourse particles [102], pronunciation variants statistics [89, 142]
- EPC Inria PANAMA: source separation; co-supervision of PhD [208, 68, 209]; FASST toolbox [266]. Finalization of work started by E. Vincent as a former member [43, 50, 218, 219, 239, 71]
- Teams MAGRIT (audio-visual corpus acquisition), CORTEX (acquisition of multimodal data), SYNALP (speech segmentation), SMART (language modeling [172]), TOSCA (co-supervision PhD), LARSEN (co-supervision PhD [79].

15 External support and funding

European projects: ALLEGRO (Interreg 2010-2012), EMOSPEECH (Eurostar 2010-2013), IFCASL (French-German program in SHS, 2013-2016), and AMIS (CHIST-ERA, 2015-2018).

ANR projects, where the team was or is coordinator: ARTIS (2009-2012), VISAC (2009-2012); CONTNOMINA (2013-2016); KAMoulox (2015-2019), and ARTSPEECH (2015-2019).

Other national projects: DOCVACIM (ANR, 2008-2011), ORTOLANG (EQUIPEX, 2013-2016), ORFEO (ANR, 2013-2016), RAPSODIE (FUI + FEDER, 2012-2016), VOICEHOME (FUI, 2015-2017), and DYCI2 (ANR, 2015-2018).

Funding from *région* Lorraine: partial funding of three theses and two post-doctoral positions, and of the COREXp project.

Miscellaneous: Three doctoral positions, four PhD scholarships, and four technology development actions have been / are funded by Inria. One PhD scholarship is funded by University of Lorraine. A CMCU contract partially support two PhD students.

Industrial contracts: Studio MAIA (2014-2015), and Venathec SA (2014-2017).

Involvement with social, economic and cultural environment

Team members have presented demonstrations at several events including the Renaissance festival of Nancy (June 2013), and at Science & You, Nancy (June 2015); as well as demonstrations for lycée, high school and master students. Other general audience dissemination include "*Minute de la science*" by France Bleu Lorraine (Spring 2011); a movie "*Je peux voir les mots que tu dis !*" (I can see the words you say!) (April 2012); podcast "*Quand les sons se séparent*" 'When the sounds get separated), Interstices (May 2012, Oct. 2014); and participation in a short TV report (France 3 Lorraine) about audiovisual speech synthesis (March 2015).

Our participation in FUI projects led to industrial partnerships with eROCCA and OnMobile SA. Also, industrial contracts were signed with Studio MAIA and Venathec SAS.

The involvement in training through research

Team members contribute to the Master of Informatics and to the master of Cognitive Sciences for courses related to speech. One person from the team is a member of the Doctoral school Committee for the Informatics specialty. Every year several team members supervised several master internships, as well as several PhD students.

Knowledge Discovery and Knowldge Engineering

Synopsis

1 Team Composition

Permanents

Miguel Couceiro (Prof. UL, since September 2014), Adrien Coulet (MCF UL), Esther Galbrun (CR Inria, since October 2015), Nicolas Jay (Prof. Faculté de Médecine, UL), Jean Lieber (MCF UL), Jean-François Mari (Prof. UL), Amedeo Napoli (DR CNRS), Emmanuel Nauer (MCF UL), Malika Smaïl-Tabbone (MCF UL), Chedy Raïssi (CR Inria), Jean-Sébastien Sereni (CR CNRS), Yannick Toussaint (CR Inria).

Former members: Mario Valencia (MCF Université de Paris 13, Inria delegation until August 2015) ; Marie-Dominique Devignes (CR CNRS), Bernard Maigret (DR CNRS emeritus) and Dave Ritchie (DR Inria) together moved to constitute the Caspid Inria Team (since January 1st 2015).

	PR	MCF	DR	CR	Total
2011	1	5	3	3	12
2016	3	4	1	4	12

Post-docs, and engineers

Sébastien Da Silva (ATER UL), Dhouha Grisa (postdoc, INRA Clermont-Ferrand/LORIA), Nguyen Le Thi Nhu (Engineer Inria), Mickaël Zehren (engineer Inria).

Former members: Aurélie Bertaux (MCF Université de Dijon), Thomas Bourquard (postdoc Université de Tours), Jérémie Bourseau (industry), Melisachew Chekol (postdoc Universität Mannheim), Sami Ghadfi (industry), Renaud Grisoni (industry), Laura Infante-Blanco (engineer LORIA), Jean-François Kneib (industry), Alice Hermann (industry), Van-Thai Hoang (postdoc IGBMC Strasbourg), Ionna Lykourentzou (researcher LIST Luxembourg), Olfa Makkaoui (industry), Luis-Felipe Melo-Mora (industry), Matthieu Osmuk (M2 student), Nicolas Pépin-Hermann (industry), Violeta Pérez-Nueno (researcher Harmonic Pharma).

Doctoral students

Younès Abid (MAIF Contract since April 2015), Quentin Brabant (MESR, since October 2015), Emmanuelle Gaillard (ATER UL), Mohsen Hassan (ANR Hybride, since May 2013), Justine Reynaud (DGA + Lorraine Region Grant, since October 2015), My Thao Tang (Hybride).

Former students:

Seyed Ziaeddin Alborzi (Inria, Capsid since January 1st), Mehwish Alam (postdoc Paris-Nord University), Yasmine Assess (preparation of CAPES), Sid-Ahmed Benabderrahmane (postdoc at Université de Paris 8), Emmanuel Bresso (postdoc University of Brasilia, Brazil), Alexey Buzmakov (postdoc Perm University, Russia), Victor Codocedo (postdoc LIRIS-INSA Lyon), Julien Cojan (industry), Valmi Dufour-Lussier (postdoc Québec Canada), Elias Egho (researcher Orange Lab Lannion), Anisah Ghoorah (Professor Maurice Island), Mehdi Kaytoue (Assistant Professor LIRIS-INSA Lyon), Thomas Meilender (industry position). Gabin Personeni (MESR, Capsid since January 1st).

Phd's defended	13	On-going PhD's	6

Team evolution

The organization of the Orpailleur team has changed since the last evaluation in January 2012 and four main facts should be mentioned: (i) Jean-Sébastien Sereni joined the team in 2013 as a CNRS CR1, (ii) Miguel Couceiro joined the team as a University Professor in September 2014, (iii) Esther Galbrun joined the team as an Inria CR2 in October 2015, and (iv) the Inria Capsid Team was created on January 1st 2015, whose research activity is on structural systems biology (this was last research theme related to the fourth objective in Orpailleur team in the preceding evaluation).

2 Life of the team

The scientific animation in the Orpailleur team is based on scientific discussions and meetings, and on the Team Seminar which is called the "Malotec" seminar (`http://malotec.loria.fr`) for "Séminaire de MAthématique et LOgique pour l'exTraction et le traitEment de Connaissances". The Malotec seminar is (roughly) held twice a month and is used either for invited presentations of external researchers or for general presentations of members of the team. During the year, there are also special events, i.e. a whole day of presentations, about a chosen topic which is of interest for the team and for the whole LORIA laboratory Besides that, the members of the Orpailleur team are involved in various scientific animation operations such as the organization of conferences and workshops attached to conferences of interest.

3 Research topics

Keywords

Knowledge discovery in databases, pattern mining, formal concept analysis, text mining, preference modeling, skylines, privacy, knowledge discovery in life sciences, knowledge engineering, knowledge systems, web of data.

Research area and main goals

The three main scientific objectives of Orpailleur Team for the evaluation period are: (1) Knowledge Discovery guided by Domain Knowledge (KDDK), (2) KDDK in Life Sciences, and (3) Knowledge Engineering. The fourth theme, namely (4) Structural Systems Biology, is now the main objective of the

Capsid Team and thus will not be discussed in length later. However, life sciences, i.e. agronomy, biology, chemistry, and medicine, are application domains where the Orpailleur team has a very rich experience, and the team intends to keep and to extend this experience, focusing on subjects which deserve a lot of consideration nowadays (e.g. sustainable development).

Knowledge discovery in databases (KDD) is aimed at discovering patterns in large databases. These patterns can then be interpreted as knowledge units to be reused in knowledge systems. The KDD process is based on three main steps: (i) selection and preparation of the data, (ii) data mining, (iii) interpretation of the discovered patterns. The KDD process –as implemented in the Orpailleur team– is based on data mining methods which are either symbolic or numerical. Symbolic methods are based on pattern mining (e.g. mining frequent itemsets, association rules, sequences...), Formal Concept Analysis (FCA[GW99]) and extensions of FCA such as Pattern Structures [378] and Relational Concept Analysis (RCA [395]). Numerical methods are based on probabilistic approaches such as second-order Hidden Markov Models (HMM[ML06]), which are well adapted to the mining of temporal and spatial data.

Domain knowledge, when available, can improve and guide the KDD process, materializing the idea of *Knowledge Discovery guided by Domain Knowledge* or KDDK. In KDDK, domain knowledge plays a role at each step of KDD: the discovered patterns can be interpreted as knowledge units and reused for problem-solving activities in knowledge systems, implementing the operational sequence “mining, interpreting (modeling), representing, and reasoning”. Hence, knowledge discovery appears as a core task in knowledge engineering, with an impact in various semantic activities, e.g. information retrieval, recommendation and ontology engineering, and in various application domains.

4 Main Achievements

The team has an original and very productive research activity in knowledge discovery and especially in developing symbolic data mining methods. Moreover, the team has a very valuable and visible activity in Formal Concept Analysis –and especially in the extensions of FCA–, and as well in pattern mining, in sequence mining, graph mining and text mining. Moreover, the team combines in an original way the mining of web of data and knowledge engineering activities, and the integration of preferences in knowledge discovery.

5 Research activities

Fundamentals of KDDK

Description. The team has a prominent position in the study of FCA and pattern mining which are at the core of numerous data mining tasks. Accordingly, the team is developing knowledge discovery methods based on pattern mining, FCA and extensions, to be used in real-sized applications on complex and massive data (multi-valued attributes, n-ary relations, sequences, trees, and graphs). One main aspect is to make pattern mining more useful and to take into account domain knowledge, user-preferences and constraints to guide the discovery process.

Main results. Advances in data and knowledge engineering have emphasized the needs for pattern mining tools working on complex data. We have contributed to two main extensions of FCA, namely Pattern Structures and Relational Concept Analysis. Pattern Structures (PS[GK01]) allow to build a concept lattice

[GW99] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis*. Springer, Berlin, 1999.

[ML06] Jean-François Mari and Florence Le Ber. Temporal and spatial data mining with second-order hidden models. *Soft Computing*, 10(5):406–414, 2006.

[GK01] Bernhard Ganter and Sergei O. Kuznetsov. Pattern structures and their projections. In *Proceedings of ICCS 2001*, LNCS 2120, pages 129–142. Springer, 2001.

from complex data, e.g. numbers, sequences, trees and graphs. Relational Concept Analysis (RCA) is able to analyze objects described both by binary and relational attributes and can play an important role in text classification and text mining [395]. Following this way, and regarding itemset and association rule discovery, we improved standard algorithms for building lattices from large data and for completing the algorithm collection of the Coron platform [398].

We designed new information retrieval methods based on FCA where the concept lattice is considered as an index space for answering disjunctive queries [317, 467]. We developed also a whole line of work on pattern structures for the discovery of functional dependencies [302], text classification and heterogeneous pattern structures [464]. FCA can also be used for clustering and recommendation [466]. Projections can be associated with pattern structures for leveraging the volume and the complexity of the computation [453]. We designed also a quasi-polynomial algorithm for mining top patterns w.r.t. measures satisfying special properties in a FCA framework [452].

Considering complex data, FCA and its extensions can be used with success in text mining [530, 537]. We also worked on temporal data and healthcare patient trajectories in the "Trajcan Project", focusing on a multi-dimensional representation of trajectories [354, 506] and the definition of a similarity measure adapted to sequences and trajectories [355]. We also worked on the analysis of molecular structures (or molecular graphs) [428, 386].

In the last four years, we worked on privacy challenges which are becoming a very important issue (regarding reputation for example) in every day life. We introduced different data anonymization methodologies based on different usability scenarios [556, 554]. On the practical side, these algorithms are applied in the PoQeMON project to monitor and analyze mobile phone data while taking into account the privacy issues.

KDDK in Life Sciences

Description. Life Sciences constitute a challenging domain for knowledge discovery. Biological data are complex from many points of views, e.g. huge size, high-dimensionality, deep inter-connections, and quick evolutions. Analyzing such data for discovering interesting patterns is an important issue in domains such as health, environment and agronomy (sustainable development), for tasks such as disease understanding, drug discovery, pharmacovigilance and pharmacogenomics. Moreover, many bio-ontologies are available and can (should) be used to improve knowledge discovery in biological data.

Main results. Data reduction is an important issue in the analysis of biological high-dimensional datasets. IntelliGO[BSTP+10] is a semantic similarity measure based on Gene Ontology (GO) which is used for classifying sets of genes. Moreover we explored the potential complementarity of Inductive Logic Programming (ILP) and FCA for analyzing biological data. ILP can be combined with FCA for interpreting theories w.r.t. domain knowledge, e.g. for understanding drug side-effects [310]. Actually, improving our ability to understand drug side-effects is necessary to reduce costly and long drug development process.

Increasing amounts of biomedical data in Linked Open Data (LOD) offer new opportunities for knowledge discovery in biomedicine[SCH+12]. We proposed some approaches for selecting, integrating, and mining LOD with the objective of discovering genes responsible for a disease [521, 539]. In addition,

---

[BSTP+10] Sidahmed Benabderrahmane, Malika Smaïl-Tabbone, Olivier Poch, Amedeo Napoli, and Marie-Dominique Devignes. Intelligo: a new vector-based semantic similarity measure including annotation origin. *BMC Bioinformatics*, 11(1):588, December 2010.

[SCH+12] Matthias Samwald, Adrien Coulet, Iker Huerga, Robert L. Powers, Joanne S. Luciano, Robert R. Freimuth, Frederick Whipple, Elgar Pichler, Eric Prud'Hommeaux, Michel Dumontier, and M. Scott Marshall. Semantically enabling pharmacogenomic data for the realization of personalized medicine. *Pharmacogenomics*, 13(2):201–212, Jan 2012.

annotating data w.r.t. concepts of an ontology is a common practice. The resulting annotations define links between data and ontologies that are usable for data exchange, data integration and data analysis. We used BioPortal (`http://bioportal.bioontology.org/`) to analyze annotations and We proposed a generic method based on FCA and pattern structures to explore and complete hand-made annotations –based on BioPortal– in biomedical document databases [490].

Finally, in the ANR Hybride project, we focused on extracting relations between named entities using graph mining methods over a collection of dependency graphs. We also applied this idea to learn syntactic patterns relating diseases with symptoms which are used for discovering potential relationships in biomedical texts [518].

Knowledge Engineering

Description. The web of data constitutes a good platform for experimenting ideas on knowledge engineering and knowledge discovery, in relation with the principles of semantic web. Actually, there are many interconnections between concept lattices in FCA and ontologies, e.g. the partial order underlying an ontology can be extracted from a concept lattice. Moreover, a pair of implications within a concept lattice can be adapted for designing concept definitions in ontologies. Accordingly, we are interested here in three main challenges: (i) how the web of data, as a set of potential knowledge sources (e.g. DBpedia, Wikipedia, Yago, Freebase…) can be mined for helping the design of definitions and knowledge bases, (ii) how knowledge discovery techniques can be applied for providing a better usage of the web of data (e.g. LOD classification), and (iii) how to build then efficient knowledge systems.

Main results. A part of the research work in Knowledge Engineering is oriented towards knowledge discovery in the web of data, as, with the increased interest in machine processable data, more and more data is now published in RDF (Resource Description Framework) format. Particularly, we are interested in the completeness of the data and the their potential to provide concept definitions in terms of necessary and sufficient conditions that can also be considered as a search for descriptions [415]. Other important aspects are concerned with data access, data visualization w.r.t. the SPARQL query language [419]. Lattice-Based View Access (LBVA) is a framework based on FCA, providing a classification of answers to SPARQL queries based on a concept lattice, which can be visualized and navigated for retrieving or mining specific patterns.

Regarding knowledge systems, the Taaable system was originally created as a challenger of the Computer Cooking Contest (co-located with ICCBR Conferences). The goal of Taaable (`http://intoweb.loria.fr/taaable3ccc/`) is to solve cooking problems on the basis of a recipe book. The Taaable system aims at federating various research themes, such as case-based reasoning, knowledge engineering, knowledge discovery, text-mining and revision.

Case-based reasoning (CBR) is a reasoning paradigm based on experience reuse (and as such well adapted to cooking). A classical case-based inference consists in selecting in a case base a case similar to the query (retrieval step) and in modifying it in order to meet the query (adaptation step). Theoretical and practical researches conducted in the Orpailleur team have led to the development of tools for building and improving CBR systems [608, 352], such as a case-based inference engine and a library for revision-based adaptation. In addition, some studies have addressed the issue of knowledge acquisition for CBR, in order to feed the knowledge base [514] or to manage knowledge reliability [510].

Structural Systems Biology

Note. This research activity is now carried out in the Capsid team (since January 2015). Below, we just recall some important facts in the evaluation period.

Description. Structural systems biology aims to describe and analyze the many components and interactions within living cells in terms of their three-dimensional (3D) molecular structures. We have been developing advanced computing techniques for molecular shape representation, protein-protein docking, and knowledge discovery in databases dedicated to protein-protein interactions.

Main results One of the current challenges in the life sciences is to understand how complex biological systems and processes can be modeled at the three-dimensional (3D) molecular level. We adapted the Hex protein docking software to use modern graphics processors (GPUs) to carry out the expensive FFT part of a docking calculation (`http://hex.loria.fr`). The docking work has facilitated further developments on modeling the assembly of multi-component molecular structures and on modeling protein flexibility during docking [402]. In order to explore the possibilities of using structural knowledge of protein-protein interactions, the KBDOCK system (`http://kbdock.loria.fr`) combines coordinate data from the Protein Data Bank with the Pfam protein domain family classification in order to describe and analyze protein-protein interactions [365, 366].

Scientific production and quality

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis	4	1	2	3	3	
H.D.R	2		1	1		
Journal	16	17	17	22	12	4
Conference proceedings	26	35	22	26	22	
Conference proceedings (national) (*)	10	4	8	3	2	3
Book chapter	3	13	1	3		
Book (written)						
Book or special issue (edited)	1	1	2	5	1	
Patent						
General audience papers (**)						

(*) and (**): We did not reported here papers published in “non selective conferences” as we do not have any clear definition of such a kind of conference. The same thing applies to “general audience papers”.

7 List of top journals in which we have published

- International journals (A+): Artificial Intelligence, Data Mining and Knowledge Discovery, IEEE Transactions on Data and Knowledge Engineering, Information Science, Journal of Multiple-Valued Logic and Soft Computing…

- Other journals (A): Annals of Mathematics and Artificial Intelligence, Expert Systems with Applications, International Journal on General Systems (IJGS), Journal of Intelligent Information Systems (JIIS)…

8 List of top conferences in which we have published

- International conferences (A+): European Conference on Artificial Intelligence (ECAI), European Conference on Machine-Learning and Practical Applications of Knowledge Discovery in databases (ECML-PKDD), International Conference on Data-Mining (ICDM), Knowledge Discovery in databases (KDD), International Joint Conference on Artificial Intelligence (IJCAI) International Conference on Knowledge Representation and Reasoning (KR), International Conference on Very Large Database (VLDB)…
- Other international conferences (A): Concept Lattices and Applications (CLA), International Conference on Case-Based Reasoning (ICCBR), International Conference on Formal Concept Analysis (ICFCA)…
- National Conferences: BDA (Bases de Données Avancées), Extraction et gestion des connaissances (EGC), Ingénierie des connaissances (IC), Jounrées ouvertes en biologie, informatique et mathématiques (JOBIM), Reconnaissance des formes et intelligence artificielle (RFIA)…

9 Software

Below we list the software which is developed in the team and made available.

- Coron includes a collection of pattern mining algorithms (`http://coron.loria.fr`).
- Orion is a "skycube computation software" (`https://github.com/leander256/Orion`).
- Carottage is based on Hidden Markov Models of second order and used for mining spatio-temporal data (`http://www.loria.fr/~jfmari/App/`). Arpentage extends Carottage for space-time clustering.
- The OrphaMine platform, developed as part of the ANR Hybrid project, enables visualization, data integration and in-depth analytics (`http://webloria.loria.fr/~mosmuk/orphamine/`). The data of interest is related to orphan diseases and knowledge discovery in OrphaMine is carried out w.r.t. the OrphaData ontology (`http://www.orpha.net`).
- IntelliGO implements a semantic similarity measure between genes based on Gene Ontology (`http://plateforme-mbi.loria.fr/intelligo/`).
- MODIM for "MOdel-driven Data Integration for Mining" is used for biological data integration (`https://gforge.inria.fr/projects/modim/`).
- Kasimir is aimed at decision support and knowledge management for the treatment of cancer (`http://katexowl.loria.fr`).
- Taaable implements a cased-based reasoning system for managing cooking recipes (`http://intoweb.loria.fr/taaable3ccc/`), related to a specific inference engine and a revision library (`http://tuuurbine.loria.fr/` and `http://revisor.loria.fr/`).
- We recall here Hex and KBDOCK which are now developed within the Capsid Team. Hex is a protein docking program (`http://hex.loria.fr`) while KBDOCK is a database containing the coordinates and annotations of all known protein-protein interfaces (`http://kbdock.loria.fr`).

The academic reputation and appeal

The research activities and results obtained by the Orpailleur team w.r.t. publications and development of software provide the team an original position. The team is requested to participate in many research projects. In addition, the members of the Orpailleur team are all involved, as members or as head persons, in national and international research groups, and evaluation structures, in the organization of conferences, as members of conference program committees and as members of editorial boards, and in the organization of journal special issues. All details are given in the appendix.

10 Prizes and Distinctions

The team has regularly best paper awards in specialized conferences: ICFCA 2015 (best promising student paper), ISMVL 2014, ICFCA 2014, Best Application Award at the 2014 NCBO Hackathon, cover issue of Journal Chemical Information in 2014, and ICCBR 2012. The members of the team are also invited speakers in specialized conferences (details are given in the complete bibliography).

Finally, the Taaable system wins regularly prizes at the "Computer Cooking Contest" (in 2015, it won 3 of the 5 prizes).

11 Editorial and organizational activities

Below we list the main event organization.

- Amedeo Napoli is a member of the steering committee of the International Conference on Concept Lattices and Applications (CLA) while Chedy Raïssi is a member of the steering committee of ECML-PKDD Conference.
- Amedeo Napoli was the general and scientific chair of the CLA 2011 Conference (`http://cla2011.loria.fr/`).
- European Conference on Computational Biology (ECCB 2014), Strasbourg, September 2014. Marie-Dominique Devignes was the scientific chair of the conference (`http://www.ebi.ac.uk/eccb/2014/`).
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2014), Nancy, September 2014 (`http://www.ecmlpkdd2014.org/`). Amedeo Napoli and Chedy Raïssi were the Conference chairs of ECML-PKDD which was joint with ILP 2014 (`http://dtai.cs.kuleuven.be/events/ilp2014/`).
- Workshop Series "FCA4AI 2012–2015" or "What can do FCA for Artificial Intelligence". Amedeo Napoli co-organized with Sergei O. Kuznetsov (HSE Moscow) and Sebastian Rudolph (TU Dresden) the four workshops FCA4AI co-located with ECAI or IJCAI Conferences since 2012 (`http://www.fca4ai.hse.ru/`, and CEUR Proceedings 1430, 1257, 1058, 939 `http://ceur-ws.org/`).
- Emmanuel Nauer was co-organizing the "Computer Cooking Contest" at ICCBR 2014 and ICCBR 2015 (see `http://liris.cnrs.fr/ccc/`), and he was co-organizing CwC (Cooking with Computer) at ECAI 2012 and IJCAI 2013.
- "SFCI". "Journées de la Société Francophone de ChemoInformatique" (SFCI), Nancy, October 2013. Amedeo Napoli was the general chair of the event (`http://sfci2013.loria.fr/`).

12 Services as expert or evaluator

- The members of the Orpailleur team are all involved, as members or as head persons, in various national research groups (GDR CNRS). Moreover, they take part in various evaluation structures such as “comité national des universités (CNU)”, ANR, and AERES committees. Amedeo Napoli was a member of the so-called “comité national du CNRS, section 7 (CoCNRS)” from 2010 until 2012.

- The members of the Orpailleur team are involved in the program and steering committees of main Conferences (ECAI, ECML-PKDD, ICDM, IJCAI, KDD, VLDB), as members of editorial boards, and in the organization of journal special issues.

13 Collaborations

- Agaur project (with UPC Barcelona 2011-2012). In this project, we study the discovery of functional dependencies and biclustering using pattern structures and FCA [302, 522, 523].

- STIC AmSud AKD (2015-2016). In this project, we study the design of an autonomous system based on FCA for preventing vulnerabilities in computer networks [434].

- Facepe and “Ciência Sem Fronteiras” (UFPE Recife, Brazil). The Facepe – Inria research project (2011-2013) aims at developing and comparing classification and clustering algorithms for interval and multi-valued data. The “French-Brazilian Workshop on Numerical and Symbolic Methods of Data Analysis (WFB 2013)” (`http://www.cin.ufpe.br/~wfb2013/`) was organized in this framework.

  This joint project is completed by the program “Ciência Sem Fronteiras” (2014-2016), a Brazilian research fellowship for three years (funding for a stay of one month of a visiting French researcher).

- PICS CNRS (2011-2013).

  The CNRS PICS collaboration (2011-2013) involving UQAM Montréal and LIRMM Montpellier was interested in the design of algorithms for Relational Concept Analysis and construction of AOC-posets [395, 303].

- Snowflake is an Inria Associate Team whose main objective is to improve biomedical knowledge discovery by connecting Electronic Health Records (EHRs) with LOD (Linked Open Data) (`http://snowflake.loria.fr/`).

- In a Zenon project with University of Nicosia (Cyprus, 2011-2013), we investigated classification of Linked Open Data using FCA and pattern structures for the completion of document annotations in biology [490].

14 External support and funding

- European Projects: CrossCult (2016-2020), ERC DOVSA (Marie Curie IEF, 2010-2012).

- ANR projects: Hybride (2012-2016, `http://hybride.loria.fr/`, headed by Yannick Toussaint), Kolflow (2011-2014, `http://kolflow.univ-nantes.fr/mediawiki/index.php/Main_Page`). PractiKPharma (2016-2020, `http://practikpharma.loria.fr/`, headed by Adrien Coulet), PEPSI (2011-2015, `http://pepsi.gforge.inria.fr/`). TermITH (2014-2016, `http://www.atilf.fr/ressources/termith/`).

- ISTEX: Excellence Initiative of Scientific and Technical Information (CNRS + DIST, 2014-2016, http://www.istex.fr).
- PoQeMON (FUI Project, 2014-2016, https://members.loria.fr/poqemon/).
- Trajcan (INCa 2010-2013).
- PEPS Projects: PEPS Mirabelle EXPLOD-Biomed (2014-2015), PEPS Approppre (2015, headed by Miguel Couceiro), Confocal (2015), Prefute (2015).
- Projects with industry: BioIntelligence (2010-2014, Dassault Systèmes), Quaero (2010-2013, Inria, Oseo, Technicolor, http://www.quaero.org).

Involvement with social, economic and cultural environment

The research activities within the team always have a double preoccupation: theory and practice. Thus, there is a strong activity in the development of systems supporting the research activities of the team. Below we mention some technology transfers.

- Aetheris is a start-up working on travel planning methods based on data mining and preferences (http://www.loria.fr/~raissi/Aetheris/). This start-up won the national I-lab competition[4].
- The Harmonic Pharma start-up company emerged from the Orpailleur team in 2009 (http://www.harmonicpharma.com) and is working in drug design. In particular, Harmonic Pharma was the coordinator of the project "LBS (Le Bois Santé)" for operating wood products in the pharmaceutical and nutriment domains in the framework of the so-called BioProLor consortium. Concerned researchers in the Orpailleur team were working on data management and knowledge discovery about new therapeutic applications. Harmonic Pharma is now more closely related to the Capsid Team.
- The research activities within the industrial BioIntelligence project were aimed at designing software modules for knowledge discovery in biology and bio-medicine. The project was led by Dassault Systèmes and was involving Sanofi Aventis, Laboratoires Pierre Fabre, Ipsen, Servier, Bayer Crops, together with INSERM and Inria.

The involvement in training through research

All members of the Orpailleur team are involved in teaching at all university levels of Université de Lorraine. In addition, some members of the Orpailleur team are also involved in seminars and teaching abroad (North and South America, Africa, Asia...). The specialties of the team are ranging from Artificial Intelligence to Database Theory including knowledge discovery, knowledge engineering, learning, preferences, and discrete maths.

The members of the Orpailleur team are also involved in student supervision, from undergraduate until postgraduate levels. Finally, the members of the Orpailleur team are involved in HDR and PhD thesis defenses, being referees or committee members.

[4] http://cache.media.enseignementsup-recherche.gouv.fr/file/2015/29/2/Palmares_2015_web2_445292.pdf

Querying Graphics through Analysis and Recognition

Synopsis

1 Team Composition

Permanents

Salvatore-Antoine Tabbone (Professor, *Université de Lorraine*), Philippe Dosch (Associate Professor, *Université de Lorraine*), Bart Lamiroy (Associate Professor, *Université de Lorraine*), Gérald Masini (Research Fellow, CNRS), Jonathan Weber (Associate Professor, *Université de Lorraine*, arrived in september 2012),

Associate members

Stéfane Paris (Associate Professor, *Université de Lorraine*), Karl Tombre (Professor, Vice-President at *Université de Lorraine* for partnerships and international affairs)

	PR	MCF	DR	CR	Total
2011	1	3			4
2016	1	3		1	5

Visiting scientist

Makoto Hasegawa (Professor, Tokyo Denki University April 2011–Dec 2012), Daniel Lopresti (Professor, Lehigh University, December 2011, July 2012, June 2013), Ricardo Torres da Silva (Professor, University Campinas, Brazil, December 2012–February 2013), Elisa Barney-Smith (Associate Professor, Boise State University, USA, October–Dec 2014 and March–July 2015)

Doctoral students

Abdessalem Bouzaieni (CIFRE XILOPIX, 2014–), Rachid Hafiane (UL, 2012–), Do Than Ha (MOET Vietnam and Europe, 2010–2013), Mehdi Felhi (CIFRE OCE, 2011–2014), Amani Boumaiza (Europe, 2010–2013), Thai Hoang (BDI PED CNRS, 2008–2012), Salim Jouili, (ANR, 2007–2011), Santosh KC (INRIA CORDI, 2008–2011)

Phd's defended	6	On-going PhD's	2

Team evolution

The team QGAR stops its research activity at the end of this year.

2 Life of the team

Several seminars are organized each year. These seminars are given by the members of the team (permanent members or PhD students) or by invited scientists.

3 Research topics

Keywords

Document analysis and recognition, image analysis, pattern recognition.

Research area and main goals

The Qgar team works on pattern recognition methods to compute useful features for image and document indexing. Our research belongs to image analysis field, and mainly to the document recognition community. The main contributions of the team are in the area of algorithms and methods for image segmentation and recognition, shape recognition with a specific focus on images of graphics-rich documents.

4 Main Achievements

Main works on pattern recognition methods have been published in leading international journals.

5 Research activities

Feature extraction and segmentation

Description As conversion from pixels to features raises a great deal of problems, our project-team have to design several algorithms and methods for noise removal, image segmentation and text detection.

Main results

- Edge noise removal: We propose different methods for denoising bilevel graphical document images by using learning dictionary based on sparse representation. Learning method starts by building a training database from corrupted images and constructing an empirically learned dictionary by using sparse representation. This dictionary can be used as a fixed dictionary to find the solution of the basis pursuit denoising problem. In addition, we provide an interesting energy noise model (the best value of the reconstruction error) which allows us to easier set the threshold required for noise removal when the model of noise is known[665] or unknown[698].

- Image segmentation: We propose a filtering method for the quasi-flat zones [683] which are image simplification operators also used as segmentation preprocessing. Our proposition achieves better filtering and requires less computation time than existing methods. Additionaly, we propose an interactive segmentation method [662] which has been compared to state-of-the-art approach on leaf segmentation and achieves good performance. Moreover, we also introduce new spatio-temporal definition of quasi-flat zones to efficiently analyze satellite time series [732]. Such definition combined with Dynamic Time Warping achieve good performance. Furthermore, we work on

interactive segmentation on tactile tablets [720]. The idea is to combine component-tree based segmentation and meaningful scales to obtain a fast and non parametric segmentation method adapted to tablet specifications.

- Text detection: We investigate two research directions; one to extract text from graphical documents and the other from real scene images. For graphical documents, we propose an original sparse representation method based on multi-learned dictionaries[696] where two sequences of learned dictionaries for the text and graphical parts respectively are defined following different sizes and non-overlapped document patches. Based on these representations, each patch in a document can be classified into the text or graphic layer by comparing its reconstruction errors. Same-sized patches in one category are then merged together to define the corresponding text or graphic parts. For natural scenes, we propose a method without learning phase. The approach is based on present a skeleton based descriptor to describe the strokes of the text candidates that compose a spatial relation graph. Then we apply the graph cuts algorithm to label the nodes of the graph as text or non-text. Finally the image is segmented after a classification step that aims to refine the results[702, 738]. In this perspective, we present a new hybrid page segmentation approach based on connected component and region analysis. Our approach is able to segment and identify lines, background(s), photo regions and multiscale text [739, 703].

- Typeface features: We have initiated a collaboration with the ANRT (*Atelier National de Recherche en Typographie*) and are developing new feature extractors that are able to describe typefaces, and extract the parameters defining an observed font from noisy scanned documents [721]. We address problems like inferring the most likely original shape from multiple noisy instances, robust curvature detection, automatic learning of human perception linked features, *etc.*

Shape descriptors

Description Shape (or symbol) recognition remains an open question when dealing with complex shapes having large variations, distorted by affine transformations and when they are embedded into the rest of the document. In any shape recognition process, before the recognition step, we need to to extract features, called usually *signatures*. Most works are concentrated on text-based or image-based signatures (popular methods are based on Bag of Features) and there is still room for graphic specific signatures to achieve efficient localization and recognition of shapes. Broadly speaking, we can group the descriptors according to their structural or statistical properties. Structural descriptors consider the structure of the shape in their definition. Graphs are data structures that best represent this type of information. In statistical descriptors the relationship between primitives is not made explicit. Therefore, statistical descriptors are feature directly computed from raw images on the pixels (or only the contours) composing the shape.

Main results

- Invariant image transformation descriptors: Polar harmonic transforms were recently proposed and have shown nice properties for image representation and pattern recognition. We proposed generic polar harmonic transforms [670] and a fast and efficient method based on the inherent recurrence relations among harmonic functions that are used in the definition of the radial and angular kernels of the transforms. The employment of these relations leads to recursive strategies for fast computation of harmonic function-based kernels [669].

- Symbol recognition is still an active field when dealing with complex shapes having large variations and distortions, structural representations or when they are embedded into the images [755]. We

intensified our works on the Radon transform with a study on its robustness to white noise [677] and new descriptors robust to affine transformations and deformations [663, 710]. Radon-based invariant shape descriptors are different from the others in the sense that Radon transform is used as an intermediate representation upon which invariant features are extracted from for the purpose of indexing/matching.

We have been focusing on extending work on spatial relation models and have used it as a basis for complex symbol description and recognition [672, 674, 750]. The method based on the spatio-structural description of a ”vocabulary” of extracted visual elementary parts. The method consists of first identifying vocabulary elements and then computing spatial relations between the possible pairs of labelled vocabulary types. These are further used as a basis for building an Attributed Relational Graph that fully describes the symbol. The experiments show that this approach, used for recognition, significantly outperforms both structural and signal-based state-of-the-art methods. It has also been extended to indexing of spatial relations [718].

- Structural description: We propose to adapt the Bag of Words model into the context of graphs. In particular, we define two BoW-based methods: The Bag of Singleton Graphs and the Bag of Visual Graphs[692, 691], which create vector representations for graphs and images, respectively. The first approach generates a bag representation for objects modeled as graphs with attributes associated with their vertices and edges. The second approach creates BoW-based descriptors using graphs to model the spatial relationships between the visual words found within an image. Both methods are validated in classification tasks obtaining significant results in terms of both accuracy and execution time.

Benchmarking and performance evaluation

Description The QGar team is involved in a long-term international collaboration with Lehigh University to develop and promote a new approach to document image analysis benchmarking and performance evaluation. The Document Analysis and Exploitation platform (DAE) is a sophisticated technical environment that consists of a repository containing document images, implementations of document analysis algorithms, and the results of these algorithms when applied to data in the repository. The use of a web-services model makes it possible to set up document analysis pipelines that form the basis for reproducible protocols. Since the platform keeps track of all intermediate results, it becomes an information resource for the analysis of experimental data. The features of the platform have been extended in order to support RDF and SparQL queries [742].

Main results The current adopted methods for experimental validation essentially consist of confronting new approaches to established and commonly accepted annotated benchmark data (usually referred to as Ground Truth of Golden Standard). Results and rankings of participating approaches are generally obtained through precision and recall-like metrics. These metrics have the drawback of solely reflecting the adequacy of the participating methods to the exact interpretation context for which the benchmark was conceived. Very often, slight variations in this context can lead to significantly different results. We have already shown that in most cases, it is theoretically impossible to be sure that two compared methods actually share the same context and that using precision/recall-like metrics are fundamentally biased to that avail. We have therefore started investigating new statistical and formal metrics that consist in measuring the various degrees of agreement and disagreement of sets of methods on benchmarking data, especially when this data itself contains errors.

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis	3		1	2		
H.D.R			1			
Journal	1	7	4	9	1	2
Conference proceedings	14	17	6	7	7	
Conference proceedings (non selective)	5		1	1		
Book chapter			3	2		
Book (written)						
Book or special issue (edited)				3	2	
Patent						
General audience papers						

7 List of top journals in which we have published

IEEE Transactions on Image Processing[669, 662] (IF 3.625), Pattern Recognition[667, 666, 671, 663, 668, 678] (5 years Impact Factor 3.613), Neurocomputing[664] (5 years IF 2.292), Signal Processing[660] (5 years IF 2.36), Pattern Recognition Letters[672] (5 years IF 1.896), Image and VIsion Computing[670, 682] (5 years IF 1.68), International Journal on Document Recognition[681, 665, 674] (IF 1).

8 List of top conferences in which we have published

International Conference on Pattern Recognition ICPR, International Conference on Document Analysis and Recognition (ICDAR) [724, 723, 686, 707, 719, 718, 689, 685], Document Analysis System (DAS)[725, 688, 703, 698], Graph-Based Representations in Pattern Recognition (GBR)[715, 671, 706], Structural, Syntactic, and Statistical Pattern Recognition (SSPR)[704, 708], IEEE International Conference on Image Processing (ICIP)[712, 711, 691], International Conference Computer Analysis of Images and Patterns (CAIP)[705, 690], IEEE International Geoscience and Remote Sensing Symposium (IGARSS)[732], International Symposium on Mathematical Morphology (ISMM)[684].

9 Software

Hasegawa Makoto and Salvatore Tabbone transferred their Radon-based shape descriptor to Universal Robot, a Japanese company developing Intelligent Software based on signal processing. Bart Lamiroy transferred his circle detection technology [751] to Exameca, a French company focusing on metrology and precision engineering.

We developped a symbol spotting method [733]. The software is available on the website http://syssy.loria.fr.

The academic reputation and appeal

10 Prizes and Distinctions

Best paper award in ACM SAC in 2012 and ACM DocEng in 2013. Nominated for the best paper: ICIP 2013 (paper[691]), ICPRAM 2014 (paper[710]), DAS 2014 (paper[703]), CIFED 2014 (paper[739]).

11 Editorial and organizational activities

Bart Lamiroy is member of the editorial board of IJDAR (International Journal of Document Analysis and Recognition), he is associate editor for the Journal of Imaging Science and Technology, published by IS&T, and review editorial board member for Frontiers in Digital Humanities, Cultural Heritage Digitization speciality section. He is co-chair of DRR'2016 and DRR'2015 (Document Recognition and Retrieval XXIII and XXII, San Francisco, USA), area chair for Graphics Analysis and Recognition at ICDAR'2015 (13th International Conference on Document Analysis and Recognition, Gammarth, Tunisia - relocated to Nancy), and program co-chair and organizing chair of GREC'2015 (11th International IAPR Workshop on Graphics Recognition, Sousse, Tunisia - relocated to Nancy) He was also General Chair of GREC 2013 (10th IAPR International Workshop on Graphics Recognition, Bethlehem, PA, USA). Bart Lamiroy was PC member of 22 international and national conferences, member of the RFIA steering committee in 2016 and 2014 and organized GREC 2013 and GREC 2015.

Philippe Dosch was PC member of 4 international and national conferences. Philippe Dosch is also the integrator of the whole current LORIA activity report.

Salvatore Tabbone is member of the editorial board of JUCS (Journal of Universal Computer Science). Salvatore Tabbone was/is program co-chair of ICPRAM'2014 (3rd International Conference on Pattern Recognition Applications and Methods, Angers, France), general chair of the organizing committee of CIFED'2014 (Colloque International Francophone sur l'Ecrit et le Document, Nancy, France). Salvatore Tabbone was PC member of 32 international and national conferences.

Karl Tombre is member of the advisory board of ELCVIA (Electronic Letters on Computer Vision and Image Analysis), and of the editorial board of Machine Graphics and Vision and of *Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées (ARIMA)*. He also co-edited a handbook on Document Image Analysis in 2014. Karl Tombre was the editor in chief of the International Journal on Document Analysis and Recognition (IJDAR). Karl Tombre was/is member of program committee for 9 international conferences.

Jonathan Weber was member of the organizing comittee of CIFED'2014 (Colloque International Francophone sur l'Ecrit et le Document, Nancy, France). Jonathan Weber was PC member of 7 international and national conferences.

12 Services as expert or evaluator

Bart Lamiroy is IAPR TC-10 Vice Chair since 2014 and Chair of the IAPR standing committee on Publications and Publicity. He is treasurer of the French IAPR chapter AFRIF since 2012. He served as a referee for the ANR and is a scientific expert to the French Ministry of Research and Higher Education for the CIR (Crédit Impôt Recherche). He served on 5 PhD committees (2 abroad).

Salvatore Tabbone is president of the GRCE (Groupe de Recherche en Communication Ecrite) since December 2010. Until August 31, 2012, Karl Tombre was the director of the Inria Nancy–Grand Est Research Center. Since September 1, 2012, he is vice-president of *Université de Lorraine*, in charge of partnerships and international affairs.Salvatore Tabbone participated as reviewer to 7 (3 abroad) PhD

thesis committees and 2 HDR committees and as examinator for 13 PhD thesis (1 abroad) and 1 HDR. He served also as expert for funding agencies (AERES, ANR, ANRT, FNRS-Semaphore Belgium).

Karl Tombre was/is member of the HCERES evaluation committee for Université Claude Bernard Lyon 1 (2015) and for ENSTA (2016). He served on one HDR committee.

13 Collaborations

We intensified our long-lasting scientific cooperation with the Computer Vision Center at Universitat Autònoma de Barcelona (Oriol Ramos-Terrades co-advised the PhD thesis of Do Thanh Han[660]). A collaboration has been initiated with University of Campinas (Sao Paulo, Brazil) on graph matching based on bag of graphs[692, 691]. We have a collaboration with Luc Brun (co-advisor of the PhD thesis of Rachid Hafiane) from GREYC-ENSI Caen on graph indexing based on theorical aspects on graph embedding[706].

Bart Lamiroy was a visiting scientist at Lehigh University from 2010–2011, and developed subsequent strong relations with professor Daniel Lopresti who visited the QGar team in 2011, 2012 and 2013. Bart Lamiroy returned the visit in 2015. This collaboration is related to reproducible research and performance analysis in Document Image Analysis and gave rise to the DAE evaluation platform[724, 723, 725, 693, 728, 722].

A strong collaboration with Boise State University was initiated in 2014 and professor Elisa Barney Smith[665] stayed 7 months in Nancy (October–December 2014, March-July 2015). This collaboration concerns image analysis techniques for the extraction of typographic features in historical documents [721, 685]. This work is also supported by ongoing exchanges with the Atelier National de Recherche Typographique (ANRT) in Nancy.

A collaboration has been initiated with Erchan Aptoula (Okan University, Turkey) on some theoretical aspects of mathematical morphology for color or multiband images [754]. We also work with François Petitjean (Monash University, Australia) on simplification, segmentation and classification of Satellite Image Time Series [680]. Moreover, we work on segmentation issues on various types of image with Manuel Grand-Brochier (Université d'Auvergne) [662].

14 External support and funding

- European project with the Eureka label SCANPLAN (2009-12)
- European project with the CHIST-ERA project AMIS (2016-19)
- PEPS CNRS SIGBaM (2012-13, leader), PICS DIA-TRIBE(2015-2017, leader), PHC Volubilis (2010-12), Chercheur d'excellence program by the Région Lorraine (2015), MSH Lorraine preoperation funding (2015), ANR DIA-Tribe (2015) submitted and passed phase 1, received excellent reviews in phase 2 but remained unfunded.
- Industrial contracts: OCE Canon (2011-2013), Xilopix (2014-2017), Exameca (2014)

Involvement with social, economic and cultural environment

Bart Lamiroy recurrently participated in *Fête de la Science* events in 2013, 2014 and 2015.

The involvement in training through research

Bart Lamiroy is chair of the Computer Science department of the École des Mines de Nancy.

Philippe Dosch heads one of the professional bachelor in computer science of *Université de Lorraine*.

Salvatore Tabbone is the director of UFR *Mathématiques et Informatique* of *Université de Lorraine* (see `http://www.univ-lorraine.fr/content/ufr-mathematiques-et-informatique`) and heads one of the computer science masters (M2 Miage-ACSI).

Jonathan Weber is head of the MMI department (Web and Multimedia) of *IUT de Saint-Dié*. He is also Vice-President of the MMI department Head Association, in charge of communication on social networks.

Reconnaissance de l'Ecriture et Analyse de Documents

Synopsis

1 Team Composition

Permanents

Abdel Belaïd (CR1 CNRS from 1984 to 2002, then Pr UL from 2002), Yolande Belaïd (MCF UL from 1984).

Post-docs and engineers

Yves Rangoni (Engineer, DOD-Oséo, UHP, Feb. 2011 - Feb. 2013), Santosh KC (DOD-Oséo grant, Postoc Dec. 2011 - May 2013), Locteau Hervé (DOD-Oséo grant, Postoc Feb. 2012 - June 2012), Montaser Awal (DOD-Oséo grant, Postoc Sept. 2013 - Nov. 2014), Thotreingam Kasar (DOD-Oséo grant, Postoc Jan. 2014 - Apr. 2015), Tapan K. Bhowmik (DOD-Oséo grant, Postoc Jan. 2014 - Jan. 2015), Hani Daher (DOD-Oséo grant, Postoc from Jan. 2013 - Dec. 2015).

Doctoral students

Emilie Philippot (PhD, Cifre Actimage grant, sept. 2008 - sept. 2011), Jean Marc Vauthier (PhD Student, DOD-Oséo grant, Octob. 2011 - 2013), Mohamed Rafik Bouguelia (PhD, DOD-Oséo grant, ATER UL on 2015, 2011 - 2015, Defense on 2015), Nihel Kooli (PhD student, DOD-Oséo grant, UL, Started on Oct. 2012, ATER UL sept. 2015, Aug. 2016, defense planned on 2016), Nabil Ghanmi (PhD, Cifre grant, UL, Started on September 1st 2012, defense planned on 2016, ATER UL sept. 2015, Aug. 2016).

Phd's defended: 2, On-going PhD's: 2.

Team evolution

The team has not changed during this period concerning the hiring of permanent staff, but remained active through its projects: 1 big Oséo five-year project and many postdocs and PhD students engaged.

2 Life of the team

Abdel Belaid is the scientific leader of the team. Yolande Belaid helps in the hardware and software management. Both jointly accompany doctorate students and postdocs by permanent meetings.

3 Research topics

In the READ team, we tackle several problems related to the incremental classification, segmentation and scanned document content analysis. The challenge is the ability to understand, exploit the information content as well as to index document images in the appropriate forms that are guided by the applications. Throughout the period between 2011 and 2016, we managed two projects. A major project Oséo, DOD, over five years with 5 thesis topics and partnership covering most active French teams in the field, and Cifre opening an interesting collaboration with the Alsace region on the recognition of chemistry books.

4 Main Achievements

- A framework for administrative document analysis related to the document flow segmentation, to the table and entity extraction, to the script separation and to the incremental classification.
- A framework for the segmentation and classification of chemistry document for line and region extraction.

5 Research activities

Handwriting recognition

Description: We proposed generative and discriminative models for handwriting recognition. For generative models, we used Dynamic Bayesian Networks which generalize better HMM by allowing the state space to be represented in factored form, instead of as a single discrete random variable. DBNs have also the advantage to represent and solve decision problems under uncertainly. For discriminative models, we addressed the convolutional Neural Networks, and specially Transparent Neural Networks suggested by McClelland for reading. These models replace learning by activation-inhibition processes.

Main results: Our investment has resulted in one book chapter in the domain [805] and the development of several powerful systems such as: Arabic/Latin and machine-printed/handwritten word discrimination using HOG-based shape descriptor and co-occurrence matrix of oriented gradients [786, 801], Probabilistic Graphical Models for Arabic handwritten word recognition [793], Collaborative combination of neuron-linguistic classifiers for large arabic word vocabulary recognition [761], Structural features extraction for handwritten Arabic personal names recognition [785], Fuzzy based preprocessing using fusion of online and offline trait for online Urdu script recognition [759], Signature verification [796].

Document image analysis and indexing

Description: In the Oséo project, we dealt with four subjects related to this theme for incoming mail. The first relates to the segmentation of document streams, through observation of continuity or ruptures in the successive pairs of pages. Structural and factual descriptors are used for this purpose. The second deals with the extraction of tables from patterns proposed by the customer. Graph matching techniques are used to find similar patterns. The third is related to the separation of handwritten/printed scripts. Pseudo-words classification techniques are investigated for this purpose. The last subject corresponds to the extraction of named entities. The idea is to exploit the knowledge in a database to find real entities in

the documents. A resolution entity is first operated on the database, then the model obtained is used for the matching in the document images.

Main results: Our investment has resulted in one book chapter in the domain [803] and the development of several powerful systems: Labelling logical structures of document images using a dynamic perceptive neural network [764, 799], Administrative document flow segmentation [779, 778], Semantic Label and Structure Model based approach for Entity Recognition in Database Context [794, 795], Table information extraction and structure recognition using query patterns [791, 789, 790], Document segmentation [803, 804, 762, 771, 802, 784, 770, 769], Bayesian network for form processing [797, 798, 763].

Incremental classification

Description: This third theme is a transversal theme investigated by M. R. Bouguélia in his thesis, which focuses on classification and learning from evolving data stream in the presence of uncertain labeling knowledges. The main objectives can be summarized in the 3 following points: minimizing the labelling cost using active learning strategies, learning with the possibility of appearence of novel classes that have never been seen before. This implies a mechanism for novel class detection, detection of possible labeling errors, to mitigate their effect on the learning performance.

Main results: We have developped an adaptive streaming active learning strategy based on instance weighting, used now by the ITESOFT company for mail sorting [777, 806, 758, 773, 776, 774, 775, 772]

Scientific production and quality

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016	Total
PhD Thesis	1				1		2
H.D.R							
Journal	1	2	1	5		1	10
Conference proceedings	3	5	8	11	8	3	38
Conference proceedings (non selective)							
Book chapter	2			1	1		4
Book (written)							
Book or special issue (edited)							
Patent							
General audience papers					1		1

7 List of top journals in which we have published

Pattern Recognition Letters (1) [757], Electronic Letters on Computer Vision and Image Analysis(3) [765, 760, 766], International Journal of Pattern Recognition and Artificial Intelligence (1) [761], International Journal on Document Analysis and Recognition (2) [764, 762], International Journal of Innovative Computing, Information and Control (1) [759].

8 List of top conferences in which we have published

ICDAR (13) [800, 767, 792, 782, 795, 788, 786, 798, 796, 772, 790, 768, 783], ICPR (2) [776, 780], ICFHR (4) [769, 770, 793, 782], MVA (2) [771, 789], DAS (1) [797].

9 Software

As we work in the context of industrial projects, several softwares have been developed: smearing for document image cleaning, line-region and pseudo-word segmentation, numeric extraction, M_EROCS: entity extraction in document blocks, GELSE: entity extraction by graph matching, entity resolution in databases, document descriptor extraction by regular expressions, unsupervised incremental learning algorithm, document flow segmentation, kfill adaptation for pepper/salt noise reduction in document images, printed/handwritten script separation.

The academic reputation and appeal

- Participation in scientific networks: involvement in national or international projects :

  At the National level, the team participates to: ISTEX-DATA: scientific resource acquisition program to create a digital library, initiated by the Ministry of Higher Education and Research, where A. Belaïd is expert for OCR evaluation. It participates also to RIMES: Recognition and Data Indexing and handwritten Data and facsimiles. Project of Techno-Vision program of the Ministries of Research and Defence. The READ group participates to this competition.

  At the International level, we particpate to two Technical Committees (TC10 & TC11) of the International Association of Pattern Recognition (IAPR), in relation with Document Analysis activities.

- Prizes and awards received by unit members

  M. R. Bouguélia: best paper in CIFED-CORIA 2014, A. Belaïd, awards for his co-chairing PCs of ICDAR'13 and ICDAR'15.

- National and international appeal (recruitment, guest researchers, etc.) The team welcomes Marianela Parodi for 6 months (Jan. 1st - June. 30th, 2011) from the National University of Rosario, Santa Fe, Argentina.

10 Prizes and Distinctions

Invited keynote by A. Belaïd, Generative-Discriminative Based Methods for Arabic Recognition, ISA, WICT, ISDA 2015, Marrakech.

11 Editorial and organizational activities

- Abdel Belaid was invited to serve as PC-Chair of ICDAR'13 (Washington, USA) and ICDAR'15 (Nancy). Because of the sad events in Tunisia, he relocated the conference from Tunis to Nancy (300 attendies).
- Abdel Belaid is co-editor of the International Journal on Document Analysis and Recognition since 2008.
- Theses and Masters: Abdel Belaid has supervised 5 PhD thesis from 2011 to 2016.

12 Collaborations

- The READ team has constant industrial collaborations with ITESOFT. They are partners in the DOD project and collaborating with ten french teams working in the field of document recognition.
- The team was selected on other projects: Project pole with Alsace BioValley for the treatment of chemistry. Scientifically, the team participates in international activities through journal editing, book chapter writing, co-chairing of program committees of international conferences, etc.
- Collaborations with other laboratories: The team maintains a scientific collaboration with LaTICE laboratory in ENSIT (Ecole Natioanle Supérieure des Ingénieurs de Tunis). This collaboration takes various forms: co-supervising of PhD students, invitations to various conferences, summer schools, seminars, and juries, courses in doctoral formations and animation of team meetings, co-organization of international conferences. Currently, Abdel Belaid co-supervises the HDR of Afef Kacem and PhD students.

13 External support and funding

Currently, the READ team manages two major contracts: oséo DOD, related to flow document analysis, programmed over five years and lasts until 2015 (900 k euros of budget) and ECLEIR, on chemical specifications, starting in July 2011 and lasts until June 2014.

The involvement in training through research

The team is involved in three Masters: the computer science Master of the FST, the Miage Master where A. Belaïd intervenes in L3, in M1 and in M2 , the Master Cognitive Science that he created in 2006. Today, he provides courses in L2 and L3. Yolande Belaïd gives her classes in IUT informatique, in first and second year. She is responsible of the constitution of the program scheduling of the department.

Abdel Belaïd had different responsabilities in the UFR-MI and the University:

- Elected member of the UFR-MI council, 2002-2012;
- Elected member of the specialist Committee of Nancy 2, Section 27, 2002-2012; President of the committee, 2008-2012;
- Elected member of the CEVU, 2006-2011;

Since December 2012, following a serious illness, A. Belaid is handicaped over 80%. He was discharged from teaching in favor of a full-time research.

Semantic Analysis of Natural Language

Synopsis

1 Team Composition

Permanents

Maxime Amblard (Assistant professor, Université de Lorraine), Philippe de Groote (DR INRIA), Bruno Guillaume (CR INRIA), Guy Perrier (Professor, Université de Lorraine, Professor Emeritus since 09/2014), Sylvain Pogodalla (CR INRIA).

	PR	MCF	DR	CR	Total
2011	1	1	1	2	5
2016	1	1	1	2	5

Post-docs, and engineers

Can Baskent (post-doc, INRIA, arrived 10/2013, left 09/2014), Karën Fort (post-doc, Université de Lorraine, arrived 09/2012, left 08/2014), Nicolas Lefebvre (engineer, INRIA, arrived 02/2015), Paul Masson (engineer, INRIA, left 30/09/2011).

Doctoral students

Clément Beysson (Université de Lorraine, arrived 09/2015), Ekaterina Lebedeva (INRIA, left 04/2012), Jiří Maršík (INRIA, arrived 10/2013), Aleksandre Maskharashvili (ANR/INRIA, arrived 01/11/2012), Mathieu Morey (Université de Lorraine, left 30/09/2011), Florent Pompigne (Université de Lorraine, left 08/2013), Sai Qian (Université de Lorraine, left 11/2014), Shohreh Tabatabayi Seifi (INRIA, left 04/2012).

Phd's defended	4	On-going PhD's	3

Team evolution

Maxime Amblard had an INRIA *délégation* (September 2013–September 2015). Guy Perrier is professor Emeritus since 09/2014).

2 Life of the team

After the Calligramme project reached the limit of 12 years, the Sémagramme team was launched first as INRIA and LORIA project on January 1st 2011 and became INRIA team-project on July 1st 2013.

3 Research topics

Keywords

Computational Linguistics, Formal Semantics, Syntax-Semantics Interface, Discourse Dynamics.

Research area and main goals

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

Syntax-semantics interface The Sémagramme project focuses on the semantics of natural languages (in a wider sense than usual, including some pragmatics). Nevertheless, the semantic construction process is syntactically guided, that is, the constructions of logical representations of meaning is based on the analysis of the syntactic structures. Consequently, in order not to commit ourselves to such or such specific theory of syntax, our approach is based on abstract generic models of the syntax-semantic interface.

At the foundational level, our objective is to define and study the formal properties of our models (expressive power, complexity…). At the applicative level, we aim at developing large scale grammars still allowing for fine-grained semantic account of linguistic phenomena.

Discourse dynamics The interpretation of a discourse is a dynamic process. On the one hand, a sentence occurring in a discourse 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.

Our objective is 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[SW74,Bar02,Bar04,cS04] and the related theories of func-

[SW74] Christopher Strachey and Christopher F. Wadsworth. Continuations: a mathematical semantics for handling jumps. Technical Report PRG-11, Oxford University Computing Laboratory, 1974.

[Bar02] Chris Barker. Continuations and the nature of quantification. *Natural Language Semantics*, 10(3):211–242, 2002. http://semanticsarchive.net/Archive/902ad5f7/. DOI: 10.1023/A:1022183511876.

[Bar04] Chris Barker. Continuations in natural language. In Hayo Thielecke, editor, *Proceedings of the Fourth ACM-SIGPLAN Continuations Workshop (CW'04)*, 2004. http://www.cs.bham.ac.uk/~hxt/cw04/barker.pdf.

[cS04] Chung chieh Shan. Delimited continuations in natural language. In Hayo Thielecke, editor, *Proceedings of the Fourth ACM-SIGPLAN Continuations Workshop (CW'04)*, 2004. http://www.cs.bham.ac.uk/~hxt/cw04/shan.pdf.

tional control operators.[FF89,FH92] We have indeed successfully applied such techniques in order to model the way quantifiers in natural languages may dynamically extend their scope.[de 06]

Common basic resources Even if our research primarily focuses on semantics and pragmatics, we nevertheless need syntactic resources. More precisely, we need syntactic trees to start with. We consequently contribute to the production of linguistics resources (mainly, for French) such as syntactic lexicons, grammars, tree banks, annotated corpora, etc.

4 Main Achievements

2011 Maxime Amblard (with Manuel Rebuschi and Michel Musiol) delivered an invited talk at the *Le langage comme Action / l'action par le langage* conference [818], 2011.

2013 Ekaterina Lebedava was awarded the *E. W. Beth Dissertation Prize* of the Association for Logic, Language and Information (FoLLI) for here thesis [807].

2013 The ACG development toolkit was released and distributed as an OPAM (OCaml Package Manager) package.

2014 Sémagramme organized the workshop on the occasion of the award of a Doctor Honoris Causa degree from the Université de Lorraine to Hans Kamp.

5 Research activities

Syntax-semantics interface

Description In our work on abstract generic models of the syntax-semantics interface, we focus on two models. One uses modular graph rewriting. Such systems offer a modular description of the interpretation process of natural language expressions. They also provide a natural account of natural language ambiguity with non-confluent rewriting systems. On the other hand, we focus on a model that follows Montague's idea that semantics must appear as a homomorphic image of syntax. In the field of categorial grammars, van Benthem showed how Montague's requirement can be realized using the Curry-Howard isomorphism.[vB86] This motivated our definition of an Abstract Categorial Grammars.[de 01]

Technically, an Abstract Categorial Grammar consists simply of a (linear) homomorphism between two higher-order signatures. Extensive studies have shown that this simple model allows several grammatical formalisms to be expressed, providing them with a syntax-semantics interface for free.[dGP04,KS07,

[FF89] Matthias Felleisen and Daniel P. Friedman. A syntactic theory of sequential state. *Theoretical Computer Science*, 69(3):243–287, 1989. DOI: 10.1016/0304-3975(89)90069-8.

[FH92] Matthias Felleisen and Robert Hieb. The revised report on the syntactic theories of sequential control and state. *Theoretical Computer Science*, 103(2):235 – 271, 1992. DOI: 10.1016/0304-3975(92)90014-7.

[de 06] Philippe de Groote. Towards a montagovian account of dynamics. In Masayuki Gibson and Jonathan Howell, editors, *Proceedings of Semantics and Linguistic Theory (SALT) 16*, 2006. DOI: 10.3765/salt.v16i0.2952.

[vB86] Johan van Benthem. *Essays in Logical Semantics*. Reidel, Dordrecht, 1986.

[de 01] Philippe de Groote. Towards Abstract Categorial Grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, pages 148–155, 2001. ACL: P01-1033.

[dGP04] Philippe de Groote and Sylvain Pogodalla. On the expressive power of Abstract Categorial Grammars: Representing context-free formalisms. *Journal of Logic, Language and Information*, 13(4):421–438, 2004. HAL: inria-00112956. DOI: 10.1007/s10849-004-2114-x.

[KS07] Makoto Kanazawa and Sylvain Salvati. Generating control languages with abstract categorial grammars. In Gerald Penn, editor, *Proceedings of The 12th conference on Formal Grammar FG 2007*. CSLI Publications, 2007. `http://research.nii.ac.jp/~kanazawa/publications/control.pdf`.

RS10]

Main Results The abstract model of graph rewriting has been studied from a theoretical perspective, in order to ensure termination properties [828, 892]. An dedicated tool has been developed (see Section 9) and applied to various tasks including parsing into syntactic dependencies [846], translating to semantic dependencies [808, 827], and corpus annotation (see Section 5).

We also studied extensively possible type-theoretic extensions of the ACG framework, and we applied them to express various grammatical constraints [809, 869]. We also used ACG to give fully compositional accounts of linguistic phenomena [831, 836]. A special focus on the reversibility property of ACG, that allows for using it in generation tasks was made, in particular within the Polymnie ANR project [832, 833, 834, 851]

Finally, in order to provide a modular approach to the semantic modeling of several phenomena (quantification, intensionality, and discourse dynamics—see Section 5), we defined a general intensionalization procedure that turns an extensional semantics for a language into an intensionalized one that is capable of accommodating truly intensional lexical items without changing the compositional semantic rules [814]. We proved some formal properties of this procedure and clarified its relation to the procedure implicit in Montague's PTQ. We then generalized this work and showed how several interpretation transformations obey a same scheme [837].

Discourse dynamics

Description When dealing with discourse dynamics, the widely used approach is Kamp's Discourse Representation Theory (DRT),[KR93] and its variants such as segmented DRT.[AL03] These theories act at a supra-sentential level and were not built, technically, as extensions of Montague semantics. As a consequence, several proposals have been made in order to accommodate DRT and Montague semantics.[dMS90, Mus96] Most of these proposals consist in lifting up some aspects of Montague's framework at the level of DRT. We attack the problem the other way around, that is to express discourse structure in the same framework as Montague semantics, namely, higher-order logic.

Main Results One of the main structural problems one faces when trying to model discourse interpretation is that the indefinite descriptions introduce existential quantifiers that may extend dynamically their scope on the complete discourse. In a previous work, we showed how a continuation semantics can provide a compositional interpretation to these dynamic phenomena.[de 06] We extracted from the resulting theory a type-theoretic dynamic logic that subsumes Groenendijk and Stokhof's Dynamic Predicate Logic. We then refined it in order to model more precisely the accessibility constraints that exist between a referring expression and its possible discourse referents [810, 859], in particular in the case of

[RS10] Christian Retoré and Sylvain Salvati. A faithful representation of non-associative lambek grammars in abstract categorial grammars. *Journal of Logic, Language and Information*, 19(2):185–200, 2010. DOI: 10.1007/s10849-009-9111-z.

[KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, 1993.

[AL03] Nicholas Asher and Alex Lascarides. *Logics of conversation*. Cambridge University Press, 2003.

[dMS90] Jeroen Groenendijk dand Martin Stokhof. Dynamic montague grammar. In László Kalmand and László Pólos, editors, *Papers from the Second Symposium on Logic and Language*, pages 3–48, Budapest, 1990. Akademiai Kiadoo.

[Mus96] Reinhard Muskens. Combining montague semantics and discourse representation. *Linguistics and Philosophy*, 19(2):143–186, 1996. DOI: 10.1007/BF00635836.

[de 06] Philippe de Groote. Towards a montagovian account of dynamics. In Masayuki Gibson and Jonathan Howell, editors, *Proceedings of Semantics and Linguistic Theory (SALT) 16*, 2006. DOI: 10.3765/salt.v16i0.2952.

presupposition induced by the use of proper names [807]. The latter work was awarded the *E. W. Beth Dissertation Prize* of the Association for Logic, Language and Information (FoLLI).

We also studied discourse dynamics by looking at morbid discourses. In collaboration with a psycholinguistist (Michel Musiol, Analyse et Traitement Informatique de la Langue Française, ATILF) and an epistemologist (Manuel Rebuschi, Laboratoire d'Histoire des Sciences et de Philosophie - Archives Henri-Poincaré, LHSP-AP), we performed a formal analysis of pathological conversations involving a schizophrenic patient and a psychiatrist. Such dialogues are characterized by so-called discourse ruptures (on the patient side), which appear as non-sensical to any "normal" speaker. Our analysis, which is based on SDRT, relies both on semantic and pragmatic features [811, 817, 820, 853, 887].

We also developed a method to interface a sentential grammar and a discourse grammar. It integrates smoothly the two grammars without using any intermediate processing step, and offers the possibility of modelling discourse structures as direct acyclic graphs (rather than mere trees). This method is based on an encoding of D-STAGs (Discourse Synchronous Tree Adjoining Grammars) as ACGs [835].

Common basic resources

Description Many NLP applications require digital linguistic resources such as tree banks, annotated corpora, or real size lexicons. Our team is active in producing such resources, mainly for the French language. This activity is usually carried on in collaboration with other French teams.

Main Results We have used our theoretical and practical tools in order to develop several resources. We develop a wide coverage grammar for French [816] that can be used by the Leopar parser (see Section 9). In collaboration with the Alpage INRIA project-team, we hav provided the Sequoia French Treebank[CS12] (3,100 sentences treebank covering several domains: news, medical, europarl and fr-wikipedia) with deep syntax dependencies [829, 854] (it was previously annotated with surface syntax dependencies).

We also developed tools for pattern searching of dependencies within an annoted corpus [847]. And in order to build new corpora annotated with syntactic dependencies, we proposed a serious game (namely, ZombiLingo[5]) with the purpose of producing dependency syntax annotations [864, 865]. A prototype has been developed and first experiments were performed (see Section 9).

[5] `http://zombilingo.org/`

[CS12] Marie Candito and Djamé Seddah. Le corpus sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical (the sequoia corpus : Syntactic annotation and use for a parser lexical domain adaptation method) [in french]. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2: TALN*, pages 321–334, Grenoble, France, June 2012. ATALA/AFCP. ACL: F12-2024.

Scientific production and quality

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis	1	1	1	1		
H.D.R						
Journal	1	2	1	3		
Conference proceedings	8	5	6	16	6	
Conference proceedings (non selective)	3	2	5	5	3	
Book chapter	1			3	1	
Book (written)						
Book or special issue (edited)	3	1				
Patent						
General audience papers						

7 List of top journals in which we have published

Journal of Logic, Language and Information (JoLLI) [814], Fundamenta Informaticae (1) [815], Traitement Automatique de Langues (TAL) (2) [813, 811], Journal of Language Modeling (1) [816], L'Évolution Psychiatrique (1) [817].

8 List of top conferences in which we have published

International Conference in Computational Linguistics (Coling) (1) [852], Traitement Automatique des Langues Naturelles (TALN) (11) [820, 821, 826, 834, 843, 845, 847, 854, 855, 856, 865], Semantics and Linguistic Theory (SALT) (1), Amsterdam Colloquium (1) [825], International Conference on Computational Semantics (IWCS) (1) [827], International conference on Language Resources and Evaluation (LREC) (4) [829, 840, 844, 849], Workshop on Logic, Language, Information and Computation (WoLLIC) (1) [838], International Conference on Parsing Technologies (IWPT) (1) [846], Logic and Engineering of Natural Language Semantics (LENLS) (4) [823, 831, 836, 859], International Natural Language Generation Conference (INLG) (1) [832].

9 Software

The ACG development toolkit (ACGtk) is a prototype for developing and manipulating Abstract Categorial Grammars. The type system currently implemented is the linear core system plus the (non-linear) intuitionistic implication. The current version is available is available as an open-source software under a CeCILL license. It has also been made available as an OPAM (OCaml Package Manager) package.

Grew is a Graph Rewriting tools dedicated to applications in NLP. It allows for confluent and non-confluent graph rewriting and includes several mechanisms that are useful in the context of NLP applications (built-in notion of feature structures, parametrization of rules with lexical information, ...). Grew is freely-available.

Leopar is a parser for natural languages based on the formalism of Interaction Grammars [GP09]. It is available as an open-source software under a CeCILL license.

ZombiLingo [865, 864] is a prototype of a GWAP (Game With A Purpose), where gamers have to give linguistic information about the syntax of French natural language sentences. It allows dependency syntax annotations of French sentences to be collected by crowdsourcing.

The academic reputation and appeal

10 Prizes and Distinctions

2013 Ekaterina Lebedava was awarded the *E. W. Beth Dissertation Prize* of the Association for Logic, Language and Information (FoLLI) for here thesis [807].

2013 Philippe de Groote gave an invited talk at the Center for Logic and Philosophy of Science of the Tilburg University, on the occasion of Reinhard Muskens' 60th birthday.

2014 Jiří Maršík has participated to the regional final of the competition *Ma thèse en 180 secondes*.

11 Editorial and organizational activities

The Sémagramme team members are involved in the editorial boards of several journals and series (FoLLI LNCS Publications on Logic, Language and Information, Higher-Order and Symbolic Computation, Traitement Automatique des langues, Cahiers du Centre de Logique InterStice)i(). They also belonged or are belonging to several standing committees (Formal Grammar, Logical Aspects of Computational Linguistics, European Summer School in Logic, Language and Information). They also chaired conferences (FG, LACL) and organized several workshops.

They have served as reviewers for several journals: Journal of Language Modeling, Computational Linguistics, Traitement Automatique de Langues (TAL), Journal of Logic, Language and Information (JoLLI), Linguistic Issues in Language Technology (LiLT) and as program and scientific committee members in several international conferences and workshop (TALN, LACL, Coling, MoL, FG, CID, IWCS, Linguistic Annotation Workshop (LAW) 2013, International Conference on Software Language Engineering (SLE 2013), PolTAL 2014, JFLA, LENLS, LICS, Journées de Phonétique Clinique (JPC), LREC).

12 Services as expert or evaluator

Sémagramme team members participated in 2 AERES evaluation, 13 PhD committees, 1 HDR committees, 8 hiring committees. They are members at several levels of the scientific organization of UL. Bruno Guillaume is head of the LORIA "Natural Language Processing and Knowledge Discovery" department since 2013.

[GP09] Bruno Guillaume and Guy Perrier. Interaction grammars. *Research on Language and Computation*, 7(2):171–208, 2009. HAL: inria-00568888. DOI: 10.1007/s11168-010-9066-x.

13 Collaborations

The Sémagramme team has collaboration both at the national level: Collaboration with Alpage (Paris 7 university & INRIA Paris-Rocquencourt), MELODI (IRIT, CNRS, Toulouse), LaBRI (CNRS, Université de Bordeaux), LIRMM) (Université de Montpellier, CNRS) via the Polymnie (see section 14) ANR project (ccordinated by Sémagramme); with ATILF and LHSP-APP (Nancy), via the participation to the SLAM project (ccordinated by Sémagramme); with LHSP-APP (Nancy) via the participation to the *Dialogues, rationalités et formalismes* MSH-Lorraine project; with Paris 4 – Sorbonne via the ZombiLingo project. A at the international level: NII, Tokyo, Japan [814]; Utrecht University (Van Gogh PHC), Netherlands [825, 836];New-York University, USA; Chicago University, USA; Düsseldorf University [867, 848].

14 External support and funding

- Van Gogh Partenariat Hubert Curien (PHC), 2012-2013.
- Polymnie ANR project (ANR-12-CORD-0004), Sep. 2012–Feb. 2016. Sémagramme is project leader.
- SLAM project, supported by the MSH Lorraine and HuMaIn PEPS CNRS. Sémagramme is project leader.
- ZombiLingo project is supported by DGLFLF (Direction Générale de la Langue Française et des Langues de France, ministère de la culture et de la communication) since 2015 and by the INRIA ADT program.

Involvement with social, economic and cultural environment

- Guy Perrier and Bruno Guillaume were interviewed in the scientific TV program “C à Savoir” on France 3 Lorraine channel in 2013.
- Maxime Amblard has participate to a television report about the SLAM project for France 3 Lorraine, may 2014. He wrote an article about the SLAM project for the Journal du CNRS, January 2014, and gave an interview about *Real Humans* (a television drama series), May 2014. He delivered an invited talk “Le langage, logique !” at the “Conférence Curieuse” of Univ. Lorraine, at the Nancy musée aquarium (September 17th 2015) and at Luxembourg University (October 15th 2015) and at the *Journées Hubert Curien*, 2012 (now part of the Science&You conference).
- Bruno Guillaume presented the game ZombiLingo during the conference Science&You in Nancy (June 2015).

The involvement in training through research

- Maxime Amblard was head of the master *Sciences Cognitives et Applications (SCA)* of UL (until July 2012).

- Maxime Amblard was head (2010–2012) and is head (since Sep. 2015) of the master 2 *Traitement Automatique des Langues* of the SCA master of UL.
- Guy Perrier was the local coordinator of the Erasmus Mundus Master program Language and Communication Technologies for the University of Nancy 2 until August 2011.
- Sylvain Pogodalla was the local coordinator (Sep. 2011–Sep. 2013) of the Erasmus Mundus Master program Language and Communication Technologies for the UL. This program were renewed in Sep. 2014.
- All the members of the Sémagramme team have taught in the master 2 *Traitement Automatique des Langues* of the SCA master of UL.
- Philippe de Groote has taught in the *Master Parisien de Recherche Informatique*

All the team members have been involved in L3, M1 (10) and M2 (11) thesis supervisions.

Speech Modelization and Text

Synopsis

1 Team Composition

Permanents

Kamel Smaïli (Pr UL), David Langlois (MCF UL), Joseph Di-Martino (MCF UL, arrived 15/12/15), Chiraz Latiri (Associated member Assistant Professor HDR, University of Tunis), Mourad Abbas (Associated member Researcher, Centre de recherche Scientifique et Technique pour le Développement de la Langue Arabe - Algeria), Karima Meftouh (Associated member Assistant Professor, Université de Annaba, Algeria).

	PR	MCF	DR	CR	Total
2011					
2012					
2013					
2014	1	1			2
2015	1	2			3
2016	1	2			3

Post-docs, and engineers

Doctoral students

Sylvain Raybaud (2009-2012), ENS (2008 - 2012) Motaz Saad (2012-2015), Campus France (2012-2015) Ameur Douib (UL, 2014-), Bourse ministérielle Menacer Mohamed-Amine (UL, 2016-), Bourse projet Chist-Era Salima Harrat (2012-), Algérie

Phd's defended	2	On-going PhD's	3

Team evolution

SMa^rT (Speech, Model and Text - Statistical Machine Translation) created on 2014. The members of the research group were in MULTISPEECH research team. For a long time, in Multispeech K. Smaïli managed a particular research axis concerning language modeling and machine translation. In 2014, he decided to create with D. Langlois a new research team.
In December 2015, Joseph Di-Martino joined SMa^rT. His research interest concerns mainly the recognition of pathological speech. In fact, SMa^rT would like to work on the recognition of this kind of issue. The idea is to use machine translation to transform a source signal into a target one. The source one in this case is the pathological speech and the target one is an improved speech

2 Life of the team

Members who worked for a long time with K. Smaïli have been associated to SMa^rT . This concerns K. Meftouh, C. Latiri and M. Abbas. K. Smaïli supervised the PhD thesis of Mrs Meftouh and Mr Abbas and the HDR of Mrs Latiri. Few projects have been conducted with C. Latiri such as CMCU and PNR with K. Meftouh.

3 Research topics

Keywords

Statistical Machine Translation, Language Modelling, Cross-lingual Sentiment analysis and opinion mining, Quality estimation of Machine Translation, Modelization and machine translation of under-resourced language (Arabic dialects)

Research area and main goals

The objective of SMa^rT team is to develop statistical approaches for applications related to Natural Language Processing. Our researches are focused on multi-lingual aspects: speech-to-speech translation, cross-lingual sentiment analysis and opinion mining, quality estimation of machine translation and machine translation of under-resourced languages such as Arabic dialects. Our approach is corpus-based: we extract from different types of corpora, statistical information which allow to modelize the language for a particular application. The corpora are mono-lingual or multi-lingual. Concerning multi-lingual corpora, they could be parallel or comparable (the topic of documents is the same, but documents are not strictly translations of each others). In our researches, we focus on French, English and Arabic. We have paid special attention on machine translation of Arabic dialects for which several scientific challenges have to be overcomed due to the fact that these dialects are not written which makes their statistical modelization complicated.

4 Research activities

Corpora

Our research activity allows us to develop several corpora since our models are based on statistical methods and machine learning techniques.

- PADIC: A Parallel Arabic DIalect Corpus. This corpus is a parallel corpus composed of five dialects: Algiers, Annaba (East of Algeria), Tunisian, Palestinian and Syrian [915]. All these corpora

are aligned with a MSA (Modern Standard Arabic). This corpus is very important for the international community which pays more and more attention to this research topic.

- Multilingual comparable corpora have been collected from different sources: Wikipedia (English 91 MW, French 57MW and Arabic 22MW), Euronews (English 7MW, French 7MW and Arabic 5MW) and several millions of words in the three languages form parallel corpora (AFP, ANN, MEDAR, NIST, TED, ...) [923]

- Multilingual annotated sentiment corpus of few million of words. This is one of the most interesting achievement in terms of collected corpus because this kind of corpora is extremely rare [921].

Machine Translation

One of the main research area of SMa^rT is Machine Translation and Speech-To-Speech machine Translation. Speech translation is a more challenging problem than text translation for the following reasons:

- Oral language and written language are very different. The syntax, for example is more permissive for oral. Classical written parallel corpora such as Europarl are therefore not adapted to speech translation. Moreover the speech production contains hesitations, repetitions. Must we translate that? Or must we clean the transcription before translating?

- In the scope of statistical machine translation, speech parallel corpus are less available than written ones. Then, we need to build new parallel corpus or we need to adapt existing corpora to the speech modality.

- The speech translation involves speech transcription, and machine translation, it is to say two independent systems. Can we simply plug the both systems? Or must we build a new system merging truly transcription and translation processes? Another problem is that a speech recognition system makes errors. This is an important problem because text translation systems assume that the text input contains no error. Then, how to deal with an erroneous input? A last question is what to translate? The 1-best, the n-best, or a word-trellis ? In this last case, how to translate a word-trellis?

We developed in the last five years, a system which achieves results comparable to those obtained by combining state of the art speech recognition systems and state of the art translation systems and implement an original method for resegmenting the hypothesis generated by the recognition system in a way that should make them easier to translate for the translation system [920]. This issue is not completely solved since the results are not sufficiently good to hope to propose a product based on this kind of model. A lot of effort has to be done to achieve this objective. The Chist-Era project on which we are involved has been structured in order to achieve this goal [903].

SMa^rT developed several new models to deal with the issue of phrase-based machine translation [900][919][898][924]. We proposed, for instance a new phrase-based translation model based on inter-lingual triggers. The originality of this method is double. First we identify common source phrases separately of the target phrases. Then we use inter-lingual triggers in order to retrieve their translations. Furthermore, we consider the way of extracting phrase translations as an optimization issue. For that we use simulated annealing algorithm to find out the best phrase translations among all those determined by inter-lingual triggers. Other variants have been proposed based on Conditional Mutual Information to deal directly with the extraction of long phrases on the target and the source language without composing them by concatenation [902][898].
Another aspect of the research we conduct concerns the evolution of the present decoder of MT from A^*

algorithm to an evolutionary algorithm . The main argument for adopting this approach is to use from the beginning step of decoding a complete solution and not a partially one which evolves as in MOSES (the baseline system). Another reason concerns the fact that this kind of algorithms impose no constraints about the underlying structure of the solution.

Machine Translation for Arabic Dialects

The translation of Arabic dialect constitutes a real challenge since it is an under-resourced language. In fact, Modern Standard Arabic is as any other language. That means it could be processed by the available tools. Unfortunately, in Arabic countries people speak an Arabic language which is different from the official one and differs from one region to another. Our objective is then to propose a speech-to-speech system converting modern standard Arabic to different Arabic dialects. The challenge is that for theses dialects, there is no available corpora which makes the statistical processing impossible while all our methods are based on statistical approach [904][899].

Cross-Lingual Sentiment Analysis and Opinion Mining

Sentiment analysis (or opinion mining) consists in identifying the subjectivity or the polarity of a text. Our objective is to compare documents dealing with the same subject, in terms of opinions and sentiments. In order to retrieve such couple of documents, the lexical content of each document is compared. We developed methods for collecting comparable corpora from different sources: Aljazeera, Euronews, AFP, ANN, MEDAR, NIST, UN, etc. The developed method extracts and aligns comparable corpora at the article level from Wikipedia based on interlingual links. To evaluate the closeness of corpora we proposed several comparability measures. Our evaluations show that the proposed comparability measures are able to capture the comparability degree of any comparable corpora. Then we developed algorithms to compare multi-lingual corpora not in terms of their degree of comparability but in terms of their underlying sentiments [921]. The final objective is to propose a multilingual press review concerning a given topic. This review should use several multilingual resources (electronic newspapers), and should class resources according to the included sentiments (fear, joy...) about the subject, polarity (against or not to the subject). This is one of the objective of Chist-Era project in which we are involved. This work should be extended to social networks for which we worked on to classify Like-minded people [908][925]

Quality Estimation for Machine Translation

In the scope of Machine Translation, Quality Estimation consists in automatically evaluating the quality (correctness, fluency) of a translated sentence or the correctness of each word in this sentence. This automatic evaluation can be useful to help an expert to post-edit and correct the sentence (or the document), or to decide if the translated document can be disseminated at it is.

Several confidence measures have been proposed and machine learning techniques have been used to deal with this issue [910][912]. In the scope of Confidence Measures, we participated to the World Machine Translation evaluation campaign in 2012, 2013 and in 2015. The objective of these campaigns is that reserachers evaluate their approaches in Machine Translation using the common data sets created for the shared tasks. All participants who submit entries will have their translations evaluated. Translation performance will be evaluated by human judgment. In 2013 and 2015, the systems we proposed have been ranked respectively 5^{th} and 1^{st}.

Vocabulary and Data selection for speech recognition

We investigated the data selection process in this context of building interpolated language models for speech transcription [916]. The objective is to automatically extract from huge out-domain corpora the most interesting part for the target task (in-domain). Results show that, in the selection process, the choice of the language models for representing in-domain and non-domain-specific data is critical. Moreover, it is better to apply the data selection only on some selected data sources.

Vocabulary selection consists in automatically selecting the most interesting words for the target task. For that we proposed a machine learning approach based on neural networks [909].

Scientific production and quality

5 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis		1			1	
H.D.R			1			
Journal	2					2
Conference proceedings	3	3	5	7	5	1
Conference proceedings (non selective)						
Book chapter						
Book (written)						
Book or special issue (edited)						3
Patent						
General audience papers						

6 List of top journals in which we have published

- Machine Translation - MT [900]
- International Journal of Computational Linguistics and Applications [898]
- International Journal of Advanced Computer Science and Applications [899]

7 List of top conferences in which we have published

- Annual Conference of the International Communication Association Interspeech [906],[905].
- International Conference on Intelligent Text Processing and Computational Linguistics - CICLING [913], [904]
- Machine Translation Summit [920]
- Workshop on Machine Translation [911], [910],[912]
- International Conference on Language Resources and Evaluation - LREC [922]
- Workshop on Building and Using Comparable Corpora - BUCC [921]

8 Software

SUBWEB

We published in 2007 a method which allows to align sub-titles comparable corpora [LSL+07]. In 2009, we proposed SUBWEB, an alignment web tool based on the developed algorithm. It allows to: upload a source and a target files, obtain an alignment at a sub-title level with a verbose option, and a graphical representation of the course of the algorithm. This work has been supported by CPER/TALC/SUBWEB[6]

QUesT

QUesT[7] is a quality estimation tool. It allows to automatically describe a sentence and its translation with numerical features. We contributed to the development of QUesT. For that, David Langlois was invited by Lucia Specia at University of Sheffield, Computer Sciences department, Natural Language Processing group. We added our own features into QUesT. This tool is dedicated to be available for the research community.

The academic reputation and appeal

- K. Smaïli is the joint director of the LIA (Laboratoire International Associé). This LIA include two labs from France: Loria and CRAN and 5 universities from Morocco. The scientific topic of this LIA concerns Big Data. The objective is to exchange experiences, skills, students and develop datasets.
- SMa^rT Applied to a Chist-Era call and the proposed project has been selected for funding .
- K. Smaïli has been invited to give a tutorial in Ecole d'Automne en Recherche d'Information : Fondements et Applications, Tunisia, 2014. The tutorial concerned the scientific basis on statistical machine translation"[8]
- K. Smaïli Visited Professor Nizar Habbash at the university of New York at Abu Dhabi. This well known researcher who was in Columbia works today for the university of NYU but at Abu Dhabi. We started discussing about the opportunity to work together on statistical processing of Arabic dialects.

9 Prizes and Distinctions

10 Editorial and organizational activities

K. Smaïli has been a joint organizer and was the PC chair of ICNLSP (International Conference on Natural Language and Speech Processing) in 2015.
He has been a member of PC of:

[6] http://wikitalc.loria.fr/dokuwiki/doku.php?id=operations:subweb
[7] https://github.com/lspecia/quest
[8] http://www.asso-aria.org/presentation-earia

[LSL+07] Caroline Lavecchia, Kamel Smaïli, David Langlois, et al. Building a bilingual dictionary from movie subtitles based on inter-lingual triggers. In *Translating and the Computer*, 2007.

- the International Speech Communication Association - INTERSPEECH in 2011, 2012, 2013, 2014, 2015 and 2016
- Invited to the editorial board of the JNLE (Journal of Natural Language Engineering) special issue.
- TALN: Traitement Automatique des Langues Naturelles in 2011, 2012, 2013, 2014, 2015 and 2016.
- CARI: Colloque Africain sur la Recherche en Informatique et Mathématiques Appliquée in 2014
- SIIE: Systèmes D'Information et Intelligence Économique

He reviewed papers for Machine Translation Journal in 2012 and for The Arabian Journal for Science and Engineering in 2012 and 2015.

David Langlois is member of AFCP (Association Francophone pour la Communication Parlée). He is member of the "Comité d'Administration" and contributed to discussions on scientific supports, student grants, forensics in the speech domain…Moreover, he organized the AFCP PhD Thesis Prize 2012, 2013, 2014-2015. Last, he was member of the scientific program of the JEP 2014 conference (Journées d'Études de la Parole). He reviewed papers for LREC2016, ICNLSP 2015, WMT2015, ACL2014, EAMT2012, JEP2012, Machine Translation Journal (2012)

11 Services as expert or evaluator

As several researchers, the members of SMa^rT are members of CP, reviewers for journals and participate to PhD defence theses. Kamel Smaïli participated to the following research activities:

Expertise

K. Smaïli expertised ANR projects (2009, 2011, 2015 and 2016) and makes a report for ANRT (Association Nationale Recherche Technologie) in 2016 .

PhD Committee

K. Smaïli participated to 18 PhD committes as a reviewer or as an examiner for these last 5 years. David Langlois participated to 2 PhD thesis

12 Collaborations

The researchers of SMa^rT developed several collaborations which conducted to the proposition of international projects. We can name few of them:

- AGH University of Science and Technology Krakow (Poland) and University of DEUSTO Facultad de Ingenieria Bilbao (Spain). With the colleagues of these universities and with colleagues from LIA (University of Avignon) we proposed a Chist-Era project which has been accepted (2016-2019).
- **CMCU PHC-UTIQUE 2014-2016** With colleagues from the university of Tunis (C. Latiri) and LIG (Catherine Berrut) we applied for a CMCU project which have been accepted. The objective of this project is to develop an efficient system for multilingual Information Retrieval.
- The collaboration with colleagues from university of Annaba (K. Meftouh) and with colleagues from CRSTDLA (M. Abbas) allowed us to apply to a PNR, the corresponding ANR Algerian project. The objective of this project which has been accepted is to translate under-resourced languages and more especially those concerning Arabic dialects.

- We welcomed Mr fadi Ghawanmeh a researcher from the university of Amman (Jordan). He contacted us in order to work on arab music improvisation. The issue is to propose an automatic system of improvisation to accompany a musician. In fact, the issue is that "Question & Answer" technique of accompaniment is common in music: either alone are together with other techniques. SMa^rT proposes to consider this issue as a machine translation problem. The question (the song part by the singer) is the source language and the answer to provide by a musician (the corresponding instrumental part) is the target language. Mr Ghawanmeh spent 4 mounths is our lab, we started collecting music corpora and we contacted experts in music processing from the university of Huddersfield (Great Britain) in order to apply to a H2020 project.

13 External support and funding

AMIS: Access Multilingual Information opinionS is a Chist-Era project involving colleagues from AGH (Poland), DEUSTO (Spain), LIA (France) and MULTISPEECH and QGAR resarch teams from Loria. K. Smaïli from SMa^rT is the leader of this project [903]. The global asked funding is around 600K euros with 203K euros for Loria.

Involvement with social, economic and cultural environment

The involvement in training through research

- K. Smaïli has been the director of UFR Mathematiques et Informatique of University of Lorraine from 2010 until 2014.
- K. Smaïli is the head of Bachelor and MIAGE outsourced in Morocco.
- K. Smaïli was in charge from 2003 until 2015 of ERASMUS exchanges at UFR Mathématiques et Informatiques especially with the university of Kuopio (Finland) and the university polytechnic of Valencia (Spain)
- K. Smaïli gives several courses: Big data, Business Intelligence, Statistical language Modelling for Master students (ERASMUS MUNDUS LCT), MIAGE in France, Morocco and Luxembourg. Statistical language modelling for speech recognition and machine translation, Course done in English
- D. Langlois has given courses on speech analysis and recognition. He has given a course on Machine Translation at University of Tunis.

Statistic and Symbolic Natural Language Processing

Synopsis

1 Team Composition

Permanents

Nadia Bellalem (MCF UL), Lotfi Bellalem (Praag UL), Christophe Cerisara (CR CNRS, arrived 1/1/12), Samuel Cruz-Lara (MCF UL), Christine Fay-Varnier (MCF UL), Claire Gardent (DR CNRS), Jean-Charles Lamirel (MCF UL, HDR)

	PR	MCF	DR	CR	Total
2011	0	5	1	0	6
2016	0	5	1	1	7

Post-docs, and engineers

Marilisa Amoia (post-doc 2011), Corinna Anderson (engineer 2011-2012), Nouha Boujelben (engineer 2012), Treveur Bretaudiere (engineer 2011-2012), Émilie Colin (engineer 2016-2017), Alexandre Denis (post-doc 2011-2015), Emmanuel Didiot (engineer 2011-2012), Nicolas Dugué (post-doc 2015-2016), Ingrid Falk (post-doc 2012-2013), Anass El Haddadi (engineer 2013-2014), German Kruszewski (engineer 2011-2012), Mariem Mahfoudh (post-doc 2015-2016), Lina María Rojas Barahona (post-doc 2011-2015), Céline Morot (engineer 2012-2013), Laura Perez-Beltrachini (post-doc 2014-2016), Sayed Rania (engineer 2015), Anselme Revuz (engineer 2015), Guillaume Serrière (engineer 2016), Anastasia Shimora (engineer 2016), Aghilas Sini (engineer 2014), Ali Tebbakh (engineer 2013-2016)

Doctoral students

Ingrid Falk (Région & Europe, 2008-2012, defended June 2012), Christian Gillot (UL, 2009-2012, defended December 2012), Bikash Gyawali (UL, 2013-2016, defended January 2016), Alejandra Lorenzo (Europe, 2011-..., cancelled for medical reasons), Shashi Narayan (UL, 2011-2014, defended November 2014), Laura Perez-Beltrachini (UL, 2009-2013, defended April 2013), Chunyang Xiao (CIFRE, 2014-...)

Phd's defended	5	On-going PhD's	1

Team evolution

The team mainly evolved at the end of 2011, by changing its name (from Talaris to Synalp) and its administrative authority (from INRIA/CNRS/UL to CNRS/UL). Several researchers have at that time left the Talaris team and Nancy: Patrick Blackburn (DR INRIA), Carlos Aceres (CR INRIA), Fabienne Venant (MCF) and Matthieu Quignard (CR CNRS). Starting from 2012, the new Synalp team is still working on Natural Language Processing, but with less focus on logic and an increasing machine learning component.

2 Life of the team

All formal and informal team meetings are conducted in English. In addition to classical team seminars, new forms of meetings and activities are regularly proposed to foster dynamism and creativity. Hence, in 2013, a *paper reading group* was formed and active all along the year. Starting in 2015, a new type of *lunch working meeting* was organized, which allowed for more free exchanges on work in progress.

We also favor as much as possible interactions with other teams in the laboratory. Hence, every year, we have invited in at least one of our team meetings all members of the Multispeech team. We are also at the origin of a working group on deep learning that groups together about 80 researchers from all laboratory departments.

3 Research topics

Keywords

Natural Language Processing, Natural Language Generation, Syntactic and semantic parsing, weakly supervised training, text clustering, dialogue processing.

Research area and main goals

Synalp stands for *Symbolic and Statistical Natural Language Processing*. As this name suggests, the aim of the Synalp team is to investigate hybrid, symbolic and statistical approaches to natural language processing. More concretely, Synalp's goal is to develop computational grammars with a semantic dimension; to investigate the interplay between symbolic and stochastic processing; to explore supervised, semi-supervised and unsupervised approaches; and to explore the linguistic and computational issues involved in such areas as syntactic and semantic parsing, natural language generation, dialog modeling and automatic resource acquisition.

The major long term computational goals of the Synalp team are:

- The design and implementation of powerful clustering techniques which support both the incremental classification of large amount of heterogeneous textual data and a detailed, supervised and unsupervised, evaluation of the output clusters.
- The development of robust methods for text generation.
- The integration of language technology and semantic resources into multimedia applications.
- The adaptation of deep and weakly supervised machine learning approaches for NLP.

These computational goals are pursued in the context of theoretical investigations that rigorously map out the required linguistic, scientific and mathematical context.

4 Main Achievements

- Generating Instructions in a Virtual Environment: the team participated in the preparation of the international GIVE 2.5 challenge on the Generation of Instructions in a Virtual Environment. This challenge brought together researchers from six universities and evaluated the participating systems on their ability to generate instructions in a dynamic 3D setting. One of our systems won the second place both in terms of objective and of subjective metrics.
- Involvment in standardization activities, with the release of MLIF as a new ISO/TC37/SC4 standard, published on 1st September 2012 `http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=37330`. MLIF is an XML-based standard to represent multimodal and multilingual dialogue corpora.
- Licencing of the GenI software developed in the team and diffusion under an open-source licence of the SATI mobile application plus two other software (JTrans & JSafran)
- Participation in 17 projects
- 5 PhDs successfully defended, 21 journal publications and 15 papers in major conferences
- Participation in the organization of Interspeech 2013, a major and large (about 1400 participants) conference in speech processing
- Organisation of the KBGen shared task, an international evaluation campaign for generation from ontologies
- Good performances (9th over 50 participants) at the SemEval2014 challenge on sentiment analysis

5 Research activities

Natural Language Generation

Description Natural Language Generation (NLG) consists in mapping a communicative goal and a dataset to a text. It involves selecting the content to be verbalised (content selection), structuring this content (document planning) and planning the verbalisation of the structured content (microplanning).

Main results Synalp has been working for several years now on different aspects of surface realisation i.e., the subtask of microplanning which maps a sentence size semantic input to a sentence. A distinctive feature of the approach being pursued is that it aims to combine linguistically informed models (e.g., grammar or lexicons) with statistical models. Thus [927, 941, 1004, 948, 980] proposes statistical and symbolic methods for automatically creating syntax/semantic lexicon that could be used for generation while [935] introduces an abstract specification language for grammar writing and shows that this language can be used to develop syntactic/semantic grammars for generation. Because hand-written grammars inevitably contains errors and missing information, we developed both symbolic [984] and statistical [986, 1029] error mining approaches which permit automatically detecting not only these errors but also mismatches between the input and the input format expected by the grammar. Applying this method to a generation input corpus derived from the Penn Treebank by the international shared task on surface realisation, we showed that it permits increasing coverage (the ratio of input for which the generator produces an output) from 38.5% before error mining to 81% afterwards [943, 930]. To handle the high combinatorial complexity of generating with symbolic grammars (the surface realisation task is known to be NP complete in the length of the input), we proposed to use regular tree grammars to prune the initial search space [942] and we developed an efficient surface realisation algorithm which combines

top-down and bottom-up filter with parallel search over the input tree and allows for generation of the PTB sentences in an average of 2 seconds per input [1030].

While this work assumed either a logical representation or an unordered dependency tree as input for surface realisation, we moved on more recently to work on surface realisation from data issued from the semantic web typically, OWL or RDF data.

In 2013, we collaborated with Stanford SRI to create of an international shared task for surface realisation from knowledge bases [962, 961]. Using the data released by this shared task, we defined a linguistically driven grammar induction method and showed that the induced grammar outperformed a competing statistical system and was competitive with an existing manual approach [992, 993, 995].

Based on a collaboration with the University of Bolzano, we proposed a hybrid approach to surface realisation which permits querying arbitrary knowledge bases in natural language. The approach combines a small generic grammar with a conditional random field model used for filtering the initial search space and select a sequence of grammar rules yielding a good quality output sentence [982, 1032].

We also worked on text simplification using statistical machine translation techniques and weakly supervised learning [1031]; on the special case of ellipsis [987]; on exploiting surface realisation to support computer aided language learning applications; [958, 957, 959, 932, 958, 977, 989, 1033, 931] and on entity linking.

Syntax and semantic

Description Syntactic parsing is the task of automatically deriving the syntactic dependency structure of a given sentence. Such a structure is represented as a labeled directed graph, where each node corresponds to a token in the sentence and has a link (dependency) to another token (head), except for the root token that has no head. The task of Semantic Role Labeling (SRL) is to automatically identify and label the semantic relations that hold between predicates (typically verbs), and their associated arguments [MCLS08]. We address the automatic parsing of both the independent and joint syntactic and semantic structures, both in the supervised and weakly supervised learning paradigms.

Main results At the beginning of 2012, our works about dependency parsing were focused on supervised methods [967, 970] Then, in 2012 and 2013, we investigated weakly supervised learning via generative models, their combination with user-defined rules and their training with Markov Chain Monte Carlo methods [1024, 1026, 969]. In 2014, we rather focused on unsupervised and weakly-supervised learning of discriminative models. We thus designed a semisupervised model that combines a Bayesian model with a maximum entropy model for Semantic Role Labelling [1025]. This model has been validated on both English and French. In 2015, we further investigated unsupervised learning of linear models, which constitute the core of many succesful models, such as SVM, CRF and deep neural networks. When a labeled corpus is given, the training objective is approximated with the empirical risk computed on this specific corpus. But without supervision, empirical risk minimization is not possible, and optimization requires alternative criteria. Based on a novel approach proposed in [BDL11] that approximates the true risk minimization objective even without supervision, we have derived a new method to train a linear classifier in an unsupervised way and applied it to predicate identification and Named Entity Recognition [1037]. Our main contributions are (a) the derivation of a formal solution to the integration of the risk in the binary case, which alleviates the need for expensive numerical integration; (b) features selection methods to focus the stochastic gradient onto the most interesting part of the search space; (c) new

[MCLS08] Lluís Màrquez, Xavier Carreras, Kenneth C. Litkowski, and Suzanne Stevenson. Semantic role labeling: an introduction to the special issue. *Comput. Linguist.*, 34(2):145–159, June 2008.

[BDL11] Krishnakumar Balasubramanian, Pinar Donmez, and Guy Lebanon. Unsupervised supervised learning II: Margin-based classification without labels. *Journal of Machine Learning Research*, 12:3119–3145, 2011.

and faster Gaussian Mixture Model training algorithms that are adapted to the specificities of this model. Recently, in parallel to this unsupervised track, we also pursued our investigations of recent supervised training algorithms, especially with tree kernels [1036] and deep neural networks, such as recursive autoencoders.

Dialogue and interaction

Description There has been much work over the last two decades on developing supervised dialog systems which support Man-Machine dialogic interactions. More recently, an emerging strand of work in supervised dialog research targets the rapid prototyping of virtual humans capable of conducting a dialog with a human user in the context of a virtual world. Much of this work however focuses on English interactions and few supervised dialog systems exists for French. Within the EU funded EMOSPEECH projects, Synalp has started to work on developing virtual dialog agents using supervised learning techniques.

Sentiment analysis and emotion detection consists in retrieving the sentiment or emotion carried by a sentence or a document. These tasks raise many difficult questions relative to the coverage of the underlying linguistic models with regards to all levels (lexical, syntactic, semantics or pragmatics). Our approach relies on emotionally-annotated lexicons, with lexical emotions propagated and combined upwards following the syntactic structure of sentences.

Main results Our work is two-fold: first the exploration of different approaches and their evaluation on standard resources; second the representation of sentiments and emotions. The first aspect deals with the specification and implementation of both symbolic approaches (mostly based on lexicons and valence shifting modelling) and machine learning approaches (based on annotated corpora). The second aspect refers to our participation in the EmotionML specification, a W3C upcoming standard for the representation of sentiments and emotions. While we are not among the authors of the specification, we contributed to its clarification and dissemination thanks to its first Java implementation.

In 2012, the Synalp team developed several dialog systems (symbolic, supervised and hybrid symbolic/supervised) within the framework of the Emospeech project [1038, 1040, 1039].

In 2013, we explored to what extent lemmatisation, lexical resources, distributional semantics and paraphrases could be used to increase the accuracy of our supervised models [990]. We further extended part of these works to the unsupervised paradigm applied to dialog management, first using a generative Bayesian model to automatically annotate the manual transcriptions in the corpus with semantic relations.

In 2014, we explored Bayesian Inverse Reinforcement Learning (BIRL) to infer human behavior in the context of the Emospeech serious game, given evidence in the form of stored dialogues provided by experts (The Wizard of Oz) who play the role of several conversational agents in the game [1035].

Text clustering and mining

Descriptions From 2011 to 2015, we focused our research activity on the development and the exploitation of new metrics adapted to the treatment of large textual data collections that can be applied in a unified way in a variety of contexts, whether supervised or unsupervised. We have especially developed for this the feature maximization metric that can be substituted for the usual distances typical in the big data context while providing explanatory capabilities to the learning methods applied to these data, conversely to the kernel methods usually exploited in this context. Our metric is further free of parameters, unlike the vast majority of alternative methods. We privileged applications of this new metric for analysis of large textual data collections, and in particular for their diachronic analysis.

Main results To tackle the challenge of incremental clustering that is a complex problem for which few convincing solutions have been proposed until now, we have notably developed an adaptation of the IGNG algorithm, called IGNGF, substituting the usual distances by feature maximization and performing learning with a principle similar to that of the expectation maximization (EM) algorithm. We have shown that this new algorithm provides superior results than alternatives from the state-of-the-art in the case of heterogeneous data clustering [998] [1007]. In NLP, we have also shown that this approach, although incremental, allowed to surpass the current best performing approaches such as formal concepts (AFC) analysis or spectral clustering in the field of automatic classification of French verbs, both for generalization and semantic role labelling tasks [980] [1004] [948].

In the supervised context, we also discussed a common issue with large corpus size, which concerns highly unbalanced classes. This problem may be exacerbated by a high degree of similarity between classes, as it might be the case for example in text classification using a hierarchical category repository. We have addressed this issue in an original way by exploiting feature selection combined with data contrasting relying on a new adaptation of the feature maximization metric. We have shown that this approach allowed both to compensate for the imbalance and to discriminate between similar classes, by improving the performance of the classifiers to over 90%, while other approaches (feature selection and/or resampling) prove completely ineffective, or even, degrade performance particularly in the case of the processing of multidimensional, unbalanced, noisy and sparse data [950] [999] [1012] [945] [1014] [1059] [1013] [1051] [1049] [947].

We are currently working on an adaptation of the feature maximization metric of maximization for the determination of optimal model in the clustering approaches, which is an open research problem. It appears that this new approach is the most effective in this context, in particular to address the clustering problems involving highly multidimensional data [1019] [1055] [1022] [1003] [1022].

Scientific production and quality

6 Synthesis of publications

	2011	2012	2013	2014	2015	2016
PhD Thesis		2	1	1		1
H.D.R						
Journal	6	2	6	5	2	0
Conference proceedings	24	26	14	11	15	1
Conference proceedings (non selective)	4	5	3	3	1	0
Book chapter	6			2		1
Book (written)						
Book or special issue (edited)			1			
Patent						
General audience papers	1					

7 List of top journals in which we have published

Computational Linguistics [935, 943], Scientometrics [950, 939], Neurocomputing [948], International Journal of Computer Applications [952]

8 List of top conferences in which we have published

ACL [995, 980, 1031], AAAI [1028], COLING [1029, 1030], EACL [1032], EMNLP [990], IJCNN [1012, 1007], Interspeech [970, 969, 967, 968, 1060, 991]

9 Software

Many software are developed in the team. But only a few of them can be considered as visible pieces of software that are licensed and are distributed and regularly used outside of our laboratory by users who are not collaborators in any of our project: In particular, we would like to highlight GeNi (commercial licensed to SRI), JTrans (open-source Cecill-C licence, distributed via github), JSafran (open-source Cecill-C licence, distributed via github), and SATI (free licence, distributed on Google and Apple playstore).

The academic reputation and appeal

10 Prizes and Distinctions

- The ITEA *METAVERSE* project received the Silver 2011 Outstanding Achievement Award.
- The team members have given 25 invited talks during the reporting period: Christophe Cerisara(1), Samuel Cruz-Lara(3), Claire Gardent(11), Jean-Charles Lamirel(8) and Laura Perez(2).

11 Editorial and organizational activities

- Christophe Cerisara: Publication chair of Interspeech 2013. Primary reviewer for ACL, SIGDIAL, Interspeech, TSD. Member of the editorial board for the special issue on *automatic processing of spoken language* of the TAL journal. Reviewer for Elsevier: Signal Processing & Computer Speech and Language; for IEEE: Trans. on Acoustics Speech and Language Processing & Signal Processing Letters; for Springer: Multimedia Tools and Applications. Reviewer for the AFIA Ph.D. thesis prize. Head of the CPER LCHN Ingénierie des Langues project.
- Claire Gardent: Chair of SIGGEN (ACL Special Interest Group on Natural Language Generation, 2015-2019); PC Chair for: ESSLLI 2016, *SEM 2016, SigDIAL 2012, ENLG 2011; Area Chair for: TALN 2011 and 2012; Co-Organiser of the first international shared task on generating from ontologies (KBGEN 2012); Local organizer of NaTAL (2011) and of ENLG 2011 (the *13th European Workshop on Natural Language Generation*); Editor in Chief for Language and Linguistics Compass, Computational and Mathematical section (with Roxanna Girju); Member of the Editorial Board for the online journal Journal of Language Modelling; Elite committee reviewer for TACL (Transaction of the ACL). PC member for EMNLP, COLING, ACL, EACL, IJCNLP, NAACL, COLING, TALN, IWCS, ENLG, INLG. Head of the CPER TALC project (2007-2013).
- Jean-Charles Lamirel: Program Chair of COLLNET and organizer of COLLNET'2016. Organizer of special session at IJCNN'2015. PC member of IEEE ICDM, ICTAI, IEA-AIE, WSOM, BDAS, ICWIS, IJCNN, COLLNET, ISKO-Maghreb, PAKDD-QIMIE, SFC. External reviewer for "Discovery Grants" of Natural Sciences and Engineering Research Council of Canada (NSERC). Reviewer for the *Scientometrics*, *PLOS-One*, *Journal of Applied Intelligence*, *Neurocomputing*, *Neural Computing Letters*, *Neural Computing and Applications*, *Journal of Intelligent Information*

Systems, *Information sciences*, *Computational Intelligence*, *Data Mining and Knowledge Discovery*, *International Journal of Artificial Intelligence Tools Advances in Knowledge Discovery and Management* journals.

- Samuel Cruz-Lara: Project leader of MLIF ISO 2012:24616, member of the WWW SYMM consortium, co-editor of the journal of Virtual Worlds Research, member of the Scientific Board of the IEEE-RITA journal.

12 Services as expert or evaluator

- Claire Gardent has reviewed for the ANR, for the EU (FP7 project) and is a member of the Comité National de la Recherche Scientifique (Section 34).
- Members of the team have participated to 11 different Ph.D. Committees.

13 Collaborations

Stanford Research International (1 project), Xerox Grenoble (1 CIFRE Ph.D.), Univ. of Bolzano (2 joint publications, 1 project), IIMAS UNAM Mexico (1 project), LIPN Paris (1 project), Univ. of West Bohemia (4 joint publications)

14 External support and funding

(FR) Involved in 2 CPER (MISN and LCHN): for both, Synalp was leader of a research axis.

(FR) Involved in 1 “Investments for the Future”: ISTEX-R

(FR) Involved in 1 Equipex : ORTOLANG

(FR) 6 ANR projects (Synalp is coordinator of 1 of them)

(FR) 2 PEPS projects (Synalp is coordinator of 1 of them)

(FR) 1 CIFRE Ph.D. thesis with Xerox

(EU) 3 ITEA projects

(EU) 1 Eurostars project

(EU) 1 Interreg project

Involvement with social, economic and cultural environment

- Design of the mobile application SATI that showcases our researches on sentiment analysis to a wider public. SATI is distributed in the Google and Apple stores.
- Design of the Allegro serious game that showcased our researches on language learning in a virtual 3D world from Second Life.

- The Emosim serious game for investigating the use of emotions in games is deployed as a demonstration game on our web site `http://talc1.loria.fr/empathic/emosim`.
- Interviews about the ALLEGRO project (radio.Saarländischer Rundfunk, “la minute science” on radio France Bleu Sud Lorraine in 2011; “C’est à savoir” on FR3)

The involvement in training through research

Members of the team are involved in the *Master 2 Sciences cognitives spécialité TAL de l’Université de Lorraine*:

- 3 members (Claire Gardent, Jean-Charles Lamirel and Christophe Cerisara) are yearly reviewers of the applicants to the Erasmus Mundus LCT
- Claire Gardent supervises the software project of a course in the Erasmus Mundus Master “Language and Communication Technology”.

1 References for Cello

Articles in International Peer-Reviewed Journal

[1] L. Bozzelli, H. P. v. Ditmarsch, T. French, J. Hales, S. Pinchinat, "Refinement modal logic", *Journal of Information and Computation 239*, December 2014, p. 37, https://hal.inria.fr/hal-01098737.

[2] L. Bozzelli, H. Van Ditmarsch, S. Pinchinat, "The complexity of one-agent refinement modal logic", *Theor. Comput. Sci. 603*, 2015, p. 58–83, https://hal.archives-ouvertes.fr/hal-01273557.

[3] A. Cordón-Franco, H. Van Ditmarsch, D. F. Duque, F. Soler-Toscano, "A colouring protocol for the generalized Russian cards problem", *Theor. Comput. Sci. 495*, 2013, p. 81–95, https://hal.archives-ouvertes.fr/hal-01273573.

[4] A. Cordón-Franco, H. Van Ditmarsch, D. F. Duque, F. Soler-Toscano, "A geometric protocol for cryptography with cards", *Des. Codes Cryptography 74*, 1, 2015, p. 113–125, https://hal.archives-ouvertes.fr/hal-01273551.

[5] H. Van Ditmarsch, D. F. Duque, W. v. d. Hoek, "On the definability of simulation and bisimulation in epistemic logic", *J. Log. Comput. 24*, 6, 2014, p. 1209–1227, https://hal.archives-ouvertes.fr/hal-01273566.

[6] H. Van Ditmarsch, T. French, "Semantics for Knowledge and Change of Awareness", *Journal of Logic, Language and Information 23*, 2, 2014, p. 169–195, https://hal.archives-ouvertes.fr/hal-01273565.

[7] H. Van Ditmarsch, S. Ghosh, R. Verbrugge, Y. Wang, "Hidden protocols: Modifying our expectations in an evolving world", *Artif. Intell. 208*, 2014, p. 18–40, https://hal.archives-ouvertes.fr/hal-01273564.

[8] H. Van Ditmarsch, W. v. d. Hoek, J. Ruan, "Connecting dynamic epistemic and temporal epistemic logics", *Logic Journal of the IGPL 21*, 3, 2013, p. 380–403, https://hal.archives-ouvertes.fr/hal-01273575.

[9] H. Van Ditmarsch, "Revocable Belief Revision", *Studia Logica 101*, 6, 2013, p. 1185–1214, https://hal.archives-ouvertes.fr/hal-01273574.

[10] H. Van Ditmarsch, "Dynamics of lying", *Synthese 191*, 5, 2014, p. 745–777, https://hal.archives-ouvertes.fr/hal-01273567.

Major International Conferences

[11] T. Ågotnes, H. Van Ditmarsch, T. French, "The undecidability of group announcements", *in : International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, p. 893–900, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273571.

[12] M. B. Andersen, T. Bolander, H. Van Ditmarsch, M. H. Jensen, "Bisimulation for Single-Agent Plausibility Models", *in : AI 2013: Advances in Artificial Intelligence - 26th Australasian Joint Conference, Dunedin, New Zealand, December 1-6, 2013. Proceedings*, p. 277–288, Unknown, Unknown or Invalid Region, 2013, https://hal.archives-ouvertes.fr/hal-01273576.

[13] C. Areces, H. Van Ditmarsch, R. Fervari, F. Schwarzentruber, "Logics with Copy and Remove", *in : Logic, Language, Information, and Computation - 21st International Workshop, WoLLIC 2014, Valparaíso, Chile, September 1-4, 2014. Proceedings*, p. 51–65, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273560.

[14] M. Attamah, H. Van Ditmarsch, D. Grossi, W. v. d. Hoek, "A Framework for Epistemic Gossip Protocols", *in : Multi-Agent Systems - 12th European Conference, EUMAS 2014, Prague, Czech Republic, December 18-19, 2014, Revised Selected Papers*, p. 193–209, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273558.

[15] M. Attamah, H. Van Ditmarsch, D. Grossi, W. v. d. Hoek, "Knowledge and Gossip", *in : ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, p. 21–26, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273563.

[16] P. Balbiani, H. Van Ditmarsch, A. Kudinov, "Subset space logic with arbitrary announcements", *in : 5th Indian Conference on Logics and its Applications (ICLA)*, p. pp. 233–244, Chennai, India, January 2013, https://hal.archives-ouvertes.fr/hal-01202518.

[17] L. Bozzelli, H. P. v. Ditmarsch, S. Pinchinat, "The Complexity of One-Agent Refinement Modal Logic", *in : 23rd International Joint Conference on Artificial Intelligence, AAAI*, Beijing, China, August 2013, https://hal.inria.fr/hal-01098744.

[18] L. Bozzelli, B. Maubert, S. Pinchinat, "Unifying Hyper and Epistemic Temporal Logics", *in : Foundations of Software Science and Computation Structures - 18th International Conference, FoSSaCS 2015, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2015, London, UK, April 11-18, 2015. Proceedings*, p. 167–182, Unknown, Unknown or Invalid Region, 2015, https://hal.archives-ouvertes.fr/hal-01273553.

[19] J.-R. Courtault, H. van Ditmarsch, D. Galmiche, "An Epistemic Separation Logic", *in : WoLLIC 2015 - 22nd Int. Workshop on Logic, Language, Information, and ComputationWoLLIC 2015, Lecture Notes in Computer Science, 9160*, Springer Verlag, p. 156–173, Bloomington, IN, United States, 2015, https://hal.archives-ouvertes.fr/hal-01259768.

[20] J. Fan, H. Van Ditmarsch, "Neighborhood Contingency Logic", *in : Logic and Its Applications - 6th Indian Conference, ICLA 2015, Mumbai, India, January 8-10, 2015. Proceedings*, p. 88–99, Unknown, Unknown or Invalid Region, 2015, https://hal.archives-ouvertes.fr/hal-01273556.

[21] J. Fan, Y. Wang, H. Van Ditmarsch, "Almost Necessary", *in : Advances in Modal Logic 10, invited and contributed papers from the tenth conference on "Advances in Modal Logic," held in Groningen, The Netherlands, August 5-8, 2014*, p. 178–196, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273569.

[22] W. v. d. Hoek, P. Iliev, "On the relative succinctness of modal logics with union, intersection and quantification", *in : International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, p. 341–348, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273561.

[23] S. Knight, B. Maubert, F. Schwarzentruber, “Asynchronous Announcements in a Public Channel”, *in : Theoretical Aspects of Computing - ICTAC 2015 - 12th International Colloquium Cali, Colombia, October 29-31, 2015, Proceedings*, p. 272–289, Unknown, Unknown or Invalid Region, 2015, https://hal.archives-ouvertes.fr/hal-01273555.

[24] H. Van Ditmarsch, J. Fan, W. v. d. Hoek, P. Iliev, “Some Exponential Lower Bounds on Formula-size in Modal Logic”, *in : Advances in Modal Logic 10, invited and contributed papers from the tenth conference on ”Advances in Modal Logic,” held in Groningen, The Netherlands, August 5-8, 2014*, p. 139–157, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273568.

[25] H. Van Ditmarsch, T. French, F. R. Velázquez-Quesada, Y. N. Wáng, “Knowledge, awareness, and bisimulation”, *in : Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013), Chennai, India, January 7-9, 2013*, Unknown, Unknown or Invalid Region, 2013, https://hal.archives-ouvertes.fr/hal-01273577.

[26] H. Van Ditmarsch, A. Herzig, E. Lorini, F. Schwarzentruber, “Listen to me! Public announcements to agents that pay attention - or not”, *in : 4th International Workshop on Logic, Rationality and Interaction (LORI IV), LNCS*, 8196, p. pp. 96–109, Hangzhou, China, October 2013, https://hal.archives-ouvertes.fr/hal-01187760.

[27] H. Van Ditmarsch, S. Knight, A. Özgün, “Arbitrary Announcements on Topological Subset Spaces”, *in : Multi-Agent Systems - 12th European Conference, EUMAS 2014, Prague, Czech Republic, December 18-19, 2014, Revised Selected Papers*, p. 252–266, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273562.

[28] H. Van Ditmarsch, S. Knight, “Partial Information and Uniform Strategies”, *in : Computational Logic in Multi-Agent Systems - 15th International Workshop, CLIMA XV, Prague, Czech Republic, August 18-19, 2014. Proceedings*, p. 183–198, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273570.

[29] H. Van Ditmarsch, J. LANG, A. Saffidine, “Strategic voting and the logic of knowledge”, *in : Proceedings of the 14th Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013), Chennai, India, January 7-9, 2013*, Unknown, Unknown or Invalid Region, 2013, https://hal.archives-ouvertes.fr/hal-01273578.

[30] H. Van Ditmarsch, “The Ditmarsch Tale of Wonders”, *in : KI 2014: Advances in Artificial Intelligence - 37th Annual German Conference on AI, Stuttgart, Germany, September 22-26, 2014. Proceedings*, p. 1–12, Unknown, Unknown or Invalid Region, 2014, https://hal.archives-ouvertes.fr/hal-01273559.

2 References for Multispeech

Doctoral Dissertations

[31] F. Bahja, *Détection du fondamental de la parole en temps réel : application aux voix pathologiques*, Theses, Université Mohammed V-Agdal UFR Informatique et Télécommunications Laboratoire LRIT Unité associée au CNRST, URAC 29, Faculté des sciences, June 2013, https://tel.archives-ouvertes.fr/tel-00927147.

[32] J. Busset, *Acoustic-to-articulatory mapping from cepstral coefficients*, Theses, Université de Lorraine, March 2013, https://tel.archives-ouvertes.fr/tel-00838913.

[33] A. Gorin, *Acoustic Model Structuring for Improving Automatic Speech Recognition Performance*, Theses, University of Lorraine, November 2014, https://hal.inria.fr/tel-01102029.

[34] X. Jaureguiberry, *Fusion for audio source separation*, Theses, TELECOM ParisTech ; INRIA Nancy, équipe Multispeech, June 2015, https://hal.archives-ouvertes.fr/tel-01189560.

[35] I. Jemaa, *Formant tracking via a multiresolution analysis*, Theses, Université de Lorraine ; Faculté des Sciences de Tunis, February 2013, https://tel.archives-ouvertes.fr/tel-00836717.

[36] U. Musti, *Acoustic-Visual Speech Synthesis by Bimodal Unit Selection*, Theses, Université de Lorraine, February 2013, https://tel.archives-ouvertes.fr/tel-00927121.

[37] L. Orosanu, *Speech recognition as a communication aid for deaf and hearing impaired people*, Theses, Université de Lorraine, December 2015, https://hal.inria.fr/tel-01251128.

[38] S. Ouni, *Multimodal Speech: from articulatory speech to audiovisual speech*, Habilitation à diriger des recherches, Université de Lorraine, November 2013, https://tel.archives-ouvertes.fr/tel-00927119.

[39] N. Souviraà-Labastie, *Audio motif spotting for guided source separation. Application to movie soundtracks.*, Theses, Université de Rennes 1, November 2015, https://hal.inria.fr/tel-01245318.

[40] D. Tran, *Uncertainty learning for noise robust ASR*, Theses, Universite de Lorraine, November 2015, https://hal.inria.fr/tel-01246481.

Articles in International Peer-Reviewed Journal

[41] M. Aron, M.-O. Berger, E. Kerrien, B. Wrobel-Dautcourt, B. Potard, Y. Laprie, "Multimodal acquisition of articulatory data: Geometrical and temporal registration", *Journal of the Acoustical Society of America 139*, 2, 2016, p. 13, https://hal.inria.fr/hal-01269578.

[42] F. Bahja, J. Di Martino, E. H. Ibn Elhaj, D. Aboutajdine, "An overview of the CATE algorithms for real-time pitch determination", *Signal, Image and Video Processing*, 2013, https://hal.inria.fr/hal-00831660.

[43] A. Benichoux, L. S. R. Simon, E. Vincent, R. Gribonval, "Convex regularizations for the simultaneous recording of room impulse responses", *IEEE Transactions on Signal Processing*, January 2014, https://hal.inria.fr/hal-00934941.

[44] F. Bimbot, E. Deruty, G. Sargent, E. Vincent, "System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces", *Music Perception*, 2016, p. 41, To appear - http://mp.ucpress.edu/, https://hal.inria.fr/hal-01188244.

[45] A. Bonneau, D. Fohr, I. Illina, D. Jouvet, O. Mella, L. Mesbahi, L. Orosanu, "Gestion d'erreurs pour la fiabilisation des retours automatiques en apprentissage de la prosodie d'une langue seconde", *Traitement Automatique des Langues 53*, 3, 2013, https://hal.inria.fr/hal-00834278.

[46] G. Bouselmi, D. Fohr, I. Illina, "Multilingual Recognition of Non-Native Speech using Acoustic Model Transformation and Pronunciation Modeling", *International Journal of Speech Technology 15*, 2, June 2012, p. 203 – 213, https://hal.archives-ouvertes.fr/hal-00764626.

[47] M. Chami, M. Immassi, J. Di Martino, "An architectural comparison of signal reconstruction algorithms from short-time Fourier transform magnitude spectra", *International Journal of Speech Technology 18*, 3, September 2015, p. 9, https://hal.inria.fr/hal-01184625.

[48] M. Dargnat, V. Colotte, K. Bartkova, A. Bonneau, "Continuations intra- et interphrastiques du français : premiers résultats expérimentaux", *SHS Web of Conferences 1*, July 2012, p. 1471–1485, Cette revue reprend les articles sélectionnés par le comité de lecture du 3e Congrès Mondial de Linguistique Française., https://hal.inria.fr/hal-00764639.

[49] S. Demange, S. Ouni, "An episodic memory-based solution for the acoustic-to-articulatory inversion problem", *Journal of the Acoustical Society of America 133*, 5, May 2013, p. 2921–2930, https://hal.inria.fr/hal-00834556.

[50] N. Duong, E. Vincent, R. Gribonval, "Spatial location priors for Gaussian model based reverberant audio source separation", *EURASIP Journal on Advances in Signal Processing 2013*, 1, 2013, p. 149, https://hal.inria.fr/hal-00870191.

[51] C. Fauth, A. Bonneau, O. Mella, V. Colotte, D. Fohr, D. Jouvet, Y. Laprie, J. Trouvain, "Constitution d'un Corpus de Français Langue Etrangère destiné aux Apprenants Allemands", *SHS Web of Conferences 8*, July 2014, p. 14, https://hal.inria.fr/hal-01080630.

[52] D. Fitzgerald, A. Liutkus, R. Badeau, "Projection-based demixing of spatial audio", *IEEE Transactions on Audio, Speech and Language Processing*, May 2016, https://hal.inria.fr/hal-01260588.

[53] N. Ito, E. Vincent, T. Nakatani, N. Ono, S. Araki, S. Sagayama, "Blind suppression of nonstationary diffuse noise based on spatial covariance matrix decomposition", *Journal of Signal Processing Systems 79*, 2, May 2015, p. 145–157, https://hal.inria.fr/hal-01020255.

[54] X. Jaureguiberry, E. Vincent, G. Richard, "Fusion methods for speech enhancement and audio source separation", *IEEE Transactions on Audio, Speech and Language Processing*, April 2016, https://hal.archives-ouvertes.fr/hal-01120685.

[55] O. Lachhab, J. Di Martino, E. H. Ibn Elhaj, A. Hammouch, "A preliminary study on improving the recognition of esophageal speech using a hybrid system based on statistical voice conversion", *SpringerPlus*, October 2015, https://hal.inria.fr/hal-01221503.

[56] Y. Laprie, R. Sock, B. Vaxelaire, B. Elie, "Comment faire parler les images aux rayons X du conduit vocal ?", *SHS Web of Conferences 8*, July 2014, p. 14, https://hal.inria.fr/hal-01059887.

[57] A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, L. Daudet, "Kernel Additive Models for Source Separation", *IEEE Transactions on Signal Processing*, June 2014, p. 14, https://hal.inria.fr/hal-01011044.

[58] A. A. Nugraha, A. Liutkus, E. Vincent, "Multichannel audio source separation with deep neural networks", *IEEE/ACM Transactions on Audio, Speech, and Language Processing 24*, 10, June 2016, https://hal.inria.fr/hal-01163369.

[59] S. Ouni, V. Colotte, U. Musti, A. Toutios, B. Wrobel-Dautcourt, M.-O. Berger, C. Lavecchia, "Acoustic-visual synthesis technique using bimodal unit-selection", *EURASIP Journal on Audio, Speech, and Music Processing*, 2013:16, June 2013, https://hal.inria.fr/hal-00835854.

[60] S. Ouni, S. Dahmani, "Is markerless acquisition of speech production accurate ?", *JASA Express Letters*, May 2016, (accepté et en cours de publication) - Date prévue : avant fin mai 2016., https://hal.inria.fr/hal-01315579.

[61] S. Ouni, "Tongue control and its implication in pronunciation training", *Computer Assisted Language Learning 27*, 5, 2014, p. 439–453, https://hal.inria.fr/hal-00834554.

[62] A. Piquard-Kipffer, L. Sprenger-Charolles, "Early predictors of future reading skills: A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8)", *Annee Psychologique 113*, 4, 2013, p. 491–521, https://hal.inria.fr/hal-00925411.

[63] A. Piquard-Kipffer, "Prédire dès l'âge de 5 ans le niveau de lecture de fin de cycle 2. Suivi de 85 enfants de langue maternelle française de 4 à 8 ans.", *L'information grammaticale*, 133, March 2012, p. 20–26, https://hal.inria.fr/hal-00681684.

[64] A. Piquard-Kipffer, "Faire voir une histoire : Louis et son incroyable chien Noisette", *Les Cahiers Pédagogiques Hors série numérique N°42*, February 2016, p. 7, https://hal.inria.fr/hal-01191878.

[65] S. Raczynski, E. Vincent, "Genre-based music language modelling with latent hierarchical Pitman-Yor process allocation", *IEEE/ACM Transactions on Audio, Speech, and Language Processing 22*, 3, January 2014, p. 672–681, https://hal.inria.fr/hal-00804567.

[66] J. Razik, O. Mella, D. Fohr, J.-P. Haton, "Frame-Synchronous and Local Confidence Measures for Automatic Speech recognition", *International Journal of Pattern Recognition and Artificial Intelligence 25*, 2, 2011, p. 1–26, https://hal.archives-ouvertes.fr/hal-00579092.

[67] U. Simsekli, A. Liutkus, T. Cemgil, "Alpha-Stable Matrix Factorization", *IEEE Signal Processing Letters*, September 2015, p. 5, https://hal.inria.fr/hal-01194354.

[68] N. Souviraà-Labastie, A. Olivero, E. Vincent, F. Bimbot, "Multi-channel audio source separation using multiple deformed references", *IEEE Transactions on Audio, Speech and Language Processing 23*, 11, June 2015, p. 1775–1787, https://hal.inria.fr/hal-01070298.

[69] A. Toutios, S. Ouni, Y. Laprie, "Estimating the control parameters of an articulatory model from electromagnetic articulograph data", *Journal of the Acoustical Society of America 129*, 5, May 2011, p. 3245–3257, https://hal.inria.fr/inria-00578733.

[70] D. T. Tran, E. Vincent, D. Jouvet, "Nonparametric uncertainty estimation and propagation for noise robust ASR", *IEEE/ACM Transactions on Audio, Speech and Language Processing 23*, 11, November 2015, p. 1835–1846, https://hal.inria.fr/hal-01114329.

[71] E. Vincent, N. Bertin, R. Gribonval, F. Bimbot, "From blind to guided audio source separation: How models and side information can improve the separation of sound", *IEEE Signal Processing Magazine 31*, 3, May 2014, p. 107–115, https://hal.inria.fr/hal-00922378.

Invited Conferences

[72] A. H. Abdelaziz, S. Watanabe, J. R. Hershey, E. Vincent, D. Kolossa, "Uncertainty propagation through deep neural networks", *in : Interspeech 2015*, Dresden, Germany, September 2015, https://hal.inria.fr/hal-01162550.

[73] A. Bonneau, "Troubles du traitement temporel du langage oral : vers des outils en recherche clinique", *in : 23eme SFNP congres de la société française de neurologie pédiatrique*, SFNP société française de neurologie pédiatrique, Nancy, France, January 2013, https://hal.inria.fr/hal-00925640.

[74] Y. Laprie, S. Ouni, "Articulatory Data Acquisition and Processing", *in : Colloque Corpus et Outils en Linguistique Langue et Parole : Statuts, Usages et Mésusages*, Strasbourg, France, July 2013, https://hal.inria.fr/hal-00921142.

[75] Y. Laprie, "Contribution of the acoustic cues to the non-native accent", *in : 169th meeting: Acoustical Society of America*, Pittsburgh, United States, May 2015, https://hal.inria.fr/hal-01188770.

[76] J. Le Roux, E. Vincent, J. R. Hershey, D. P. Ellis, "MICbots: collecting large realistic datasets for speech and audio research using mobile robots", *in : IEEE 2015 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, https://hal.inria.fr/hal-01116822.

[77] S. Ouni, "Acquisition de données articulatoires par un articulographe", *in : Typologie des rhétiques : manifestations phonétiques et enjeux phonologiques*, Paris, France, June 2011, https://hal.inria.fr/hal-00934377.

[78] S. Ouni, "Toward Realistic Expressive Audiovisual Speech Synthesis", *in : Expressive Virtual Actors workshop*, G. B. et Remi Ronfard (editor), Gipsa-Lab, Grenoble, France, November 2015. This event is a special session in the series supported by the ERC Expressive grant, https://hal.inria.fr/hal-01243644.

[79] E. Vincent, A. Sini, F. Charpillet, "Audio source localization by optimal control of a mobile robot", *in : IEEE 2015 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, https://hal.inria.fr/hal-01103949.

[80] F. Weninger, H. Erdogan, S. Watanabe, E. Vincent, J. Le Roux, J. R. Hershey, B. Schuller, "Speech enhancement with LSTM recurrent neural networks and its application to noise-robust ASR", *in : 12th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Liberec, Czech Republic, August 2015, https://hal.inria.fr/hal-01163493.

Major International Conferences

[81] F. Bahja, J. Di Martino, E. H. Ibn Elhaj, D. Aboutajdine, "Prediction of Cepstral Excitation Pulses for Voice Conversion", *in : 5th. International Conference on Information Systems and Economic Intelligence - SIIE 2012*, Djerba, Tunisia, February 2012, https://hal.inria.fr/hal-00761776.

[82] F. Bahja, E. H. Ibn Elhaj, J. Di Martino, "On the Use of Wavelets and Cepstrum Excitation for Pitch Determination in Real-Time", *in : 3rd International Conference on Multimedia Computing and Systems - ICMCS'12*, Tangier, Morocco, May 2012, https://hal.inria.fr/hal-00761819.

[83] A. Bandini, S. Ouni, P. Cosi, S. Orlandi, C. Manfredi, "Accuracy of a markerless acquisition technique for studying speech articulators. In Interspeech 2015", *in : Interspeech 2015*, ISCA (editor), Dresden, Germany, September 2015, https://hal.inria.fr/hal-01189000.

[84] J. Barker, R. Marxer, E. Vincent, S. Watanabe, "The third 'CHiME' Speech Separation and Recognition Challenge: Dataset, task and baselines", *in : 2015 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2015)*, Scottsdale, AZ, United States, December 2015, https://hal.inria.fr/hal-01211376.

[85] K. Bartkova, A. Bonneau, V. Colotte, M. Dargnat, "Productions of "continuation contours" by French speakers in L1 (French) and L2 (English)", *in : Speech Prosody*, *1*, p. 426–429, Shangai, China, May 2012, https://hal.inria.fr/hal-00763919.

[86] K. Bartkova, D. Jouvet, E. Delais-Roussarie, "Prosodic Parameters and Prosodic Structures of French Emotional Data", *in : Speech Prosody 2016*, *Speech Prosody 2016*, Boston, United States, May 2016, https://hal.inria.fr/hal-01293516.

[87] K. Bartkova, D. Jouvet, "Automatic Detection of the Prosodic Structures of Speech Utterances", *in : SPECOM - 15th International Conference on Speech and Computer - 2013*, M. Železný, I. Habernal, A. Ronzhin (editors), *Lecture Notes in Artificial Intelligence*, *8113*, Springer Verlag, p. 1–8, Pilsen, Czech Republic, September 2013, https://hal.inria.fr/hal-00834318.

[88] K. Bartkova, D. Jouvet, "Links between Manual Punctuation Marks and Automatically Detected Prosodic Structures", *in : Speech Prosody 2014*, Dublin, Ireland, May 2014, https://hal.archives-ouvertes.fr/hal-00998031.

[89] K. Bartkova, D. Jouvet, "Impact of frame rate on automatic speech-text alignment for corpus-based phonetic studies", *in : ICPhS'2015 - 18th International Congress of Phonetic Sciences*, Glasgow, United Kingdom, August 2015, https://hal.inria.fr/hal-01183637.

[90] J. Beliao, A. Liutkus, "OOPS: une approche orientée objet pour l'interrogation et l'analyse linguistique de l'interface prosodie/syntaxe/discours", *in : 4e Congrès Mondial de Linguistique Française*, *8*, p. 2565–2581, Berlin, Germany, July 2014, https://hal.archives-ouvertes.fr/hal-01053422.

[91] F. Bimbot, G. Sargent, E. Deruty, C. Guichaoua, E. Vincent, "Semiotic Description of Music Structure: an Introduction to the Quaero/Metiss Structural Annotations", *in : AES 53rd International Conference on Semantic Audio*, p. P1–1, London, United Kingdom, January 2014. 12 pages, https://hal.archives-ouvertes.fr/hal-00931859.

[92] A. Bonneau, M. Cadot, "German non-native realizations of French voiced fricatives in final position of a group of words", *in : Interspeech 2015*, Möller, S., Ney, H., Moebius, B., Nöth, E., Dresde, Germany, September 2015, https://hal.inria.fr/hal-01186068.

[93] A. Bonneau, B. Wrobel-Dautcourt, "Efficiency of five labial correlates for /i/ and /y/ in adverse contexts", *in : The ninth International Seminar on Speech Production - ISSP'11*, Montreal, Canada, June 2011, https://hal.inria.fr/inria-00579160.

[94] A. Bonneau, "Realizations of French voiced fricatives by German learners as a function of speaker level and prosodic boundaries", *in : 18th International Congress of Phonetic Sciences, ICPhS 2015*, University of Glasgow, p. 5, Glasgow, United Kingdom, August 2015, https://hal.inria.fr/hal-01186062.

[95] F. Bouarourou, B. Vaxelaire, Y. Laprie, R. Ridouane, M. Bechet, R. Sock, "PLOSIVE AND FRICATIVE GEMINATES IN TARIFIT AN ARTICULATORY AND ACOUSTIC STUDY", *in : International, Proceedings of ISSP 2014*, 9, ISSP 2014, Köln, Germany, May 2014, https://halshs.archives-ouvertes.fr/halshs-01151411.

[96] J. Busset, M. Cadot, “Démêler les actions des articulateurs en jeu lors de la production de parole avec le logiciel C.H.I.C. : Analyse de séquences de radiographies de la tête.”, *in : 6th International Conference Implicative Statistic Analysis - A.S.I. 6 - 2012*, p. 291–305, Caen, France, November 2012, https://hal.archives-ouvertes.fr/hal-00759054.

[97] J. Busset, Y. Laprie, “Adaptation of cepstral coefficients for acoustic-to-articulatory inversion”, *in : International Seminar on Speech Production 2011 - ISSP’11*, Montréal, Canada, June 2011, https://hal.inria.fr/inria-00599108.

[98] J. Busset, Y. Laprie, “Acoustic-to-articulatory inversion by analysis-by-synthesis using cepstral coefficients”, *in : ICA - 21st International Congress on Acoustics - 2013*, Montréal, Canada, June 2013, https://hal.inria.fr/hal-00836808.

[99] M. Cadot, A. Bonneau, “Pourquoi et comment transformer des variables quantitatives en catégorielles ? Application à l’intonation de la langue française.”, *in : ASI-8*, Radès, France, November 2015, https://hal.archives-ouvertes.fr/hal-01188477.

[100] M. Cadot, A. Bonneau, “Du fichier audio à l’intonation en Français :Graphes pour l’apprentissage de 3 classes intonatives”, *in : Fouille de données complexes (FDC@EGC2016), Proceedings of FDC@EGC2016*, Reims, France, January 2016, https://hal.archives-ouvertes.fr/hal-01292121.

[101] M. Chami, J. Di Martino, L. Pierron, E. H. Ibn Elhaj, “Real-Time Signal Reconstruction from Short-Time Fourier Transform Magnitude Spectra Using FPGAs”, *in : 5th. International Conference on Information Systems and Economic Intelligence - SIIE 2012*, Djerba, Tunisia, February 2012, https://hal.inria.fr/hal-00761783.

[102] M. Dargnat, K. Bartkova, D. Jouvet, “Discourse Particles In French: Prosodic Parameters Extraction and Analysis”, *in : International Conference on Statistical Language and Speech Processing*, Budapest, Hungary, November 2015, https://hal.inria.fr/hal-01184197.

[103] M. Dargnat, A. Bonneau, K. Bartkova, V. Colotte, “”Non-conclusive” Slopes in French: First Results”, *in : Interface Discours et prosodie 2011*, University of Salford, Manchester, United Kingdom, September 2011, https://hal.inria.fr/hal-00642721.

[104] M. Dargnat, A. Bonneau, V. Colotte, K. Bartkova, “Intra- and Inter-clausal Continuation Slopes in French: First Results”, *in : Experimental and Theoretical Advances in Prosody 2*, Montréal, Canada, September 2011, https://hal.inria.fr/hal-00642735.

[105] M. Dargnat, V. Colotte, K. Bartkova, A. Bonneau, “Continuations intra et interphrastiques du français : premiers résultats expérimentaux”, *in : CMLF 2012*, p. 1471–1485, Lyon, France, 2012, https://halshs.archives-ouvertes.fr/halshs-00958196.

[106] S. Demange, S. Ouni, “Acoustic-to-articulatory inversion using an episodic memory”, *in : International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE (editor), IEEE - Signal Processing Society, Prague, Czech Republic, May 2011, https://hal.inria.fr/inria-00578740.

[107] S. Demange, S. Ouni, “Continuous episodic memory based speech recognition using articulatory dynamics”, *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, ISCA (editor), Florence, Italy, August 2011, https://hal.inria.fr/inria-00602414.

[108] B. Dumortier, E. Vincent, M. Deaconu, "Acoustic Control of Wind Farms", *in : EWEA 2015 - European Wind Energy Association*, Paris, France, November 2015, https://hal.inria.fr/hal-01233730.

[109] B. Dumortier, E. Vincent, "Blind RT60 estimation robust across room sizes and source distances", *in : 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Firenze, Italy, May 2014, https://hal.inria.fr/hal-00941061.

[110] B. Elie, Y. Laprie, "Audiovisual to area and length functions inversion of human tract", *in : Eusipco 2014*, Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01096547.

[111] B. Elie, Y. Laprie, "A GLOTTAL CHINK MODEL FOR THE SYNTHESIS OF VOICED FRICATIVES", *in : International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01314308.

[112] M. Embarki, S. Ouni, F. Salam, "Speech clarity and coarticulation in Modern standard Arabic and Dialectal Arabic", *in : 17th International Congress of Phonetic Sciences - ICPhS XVII*, p. 635–638, Hong Kong, China, August 2011, https://hal.archives-ouvertes.fr/hal-00671225.

[113] M. Embarki, S. Ouni, F. Salam, "Clarté de la parole et effets coarticulatoires en arabe standard et dialectal", *in : Actes de la conférence conjointe JEP-TALN-RECITAL 2012*, A. . AFCP (editor), *1:JEP*, p. 209–216, Grnoble, France, June 2012, https://hal.archives-ouvertes.fr/hal-00762581.

[114] C. Fauth, A. Bonneau, F. Zimmerer, J. Trouvain, B. Andreeva, V. Colotte, D. Fohr, D. Jouvet, J. Jügler, Y. Laprie, O. Mella, B. Möbius, "Designing a Bilingual Speech Corpus for French and German Language Learners: a Two-Step Process", *in : LREC - 9th Language Resources and Evaluation Conference*, The European Language Resources Association, Reykjavik, Iceland, May 2014, https://hal.inria.fr/hal-00979026.

[115] C. Fauth, A. Bonneau, "L1-L2 interference: the case of devoicing of French voiced obstruents in final position by German learners - Pilot study", *in : International Workshop on Multilinguality in Speech Research: Data, Methods and Models.*, Bernd Möbius et Jürgen Trouvain, Université de la Sarre, Allemagne, Dagstuhl, Germany, April 2014, https://hal.inria.fr/hal-01095183.

[116] D. Fitzgerald, A. Liutkus, R. Badeau, "PROJET - Spatial Audio Separation Using Projections", *in : 41st International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Shanghai, China, 2016, https://hal.archives-ouvertes.fr/hal-01248014.

[117] D. Fitzgerald, A. Liutkus, Z. Rafii, B. Pardo, L. Daudet, "Harmonic/Percussive Separation Using Kernel Additive Modelling", *in : IET Irish Signals & Systems Conference 2014*, Limerick, Ireland, June 2014, https://hal.inria.fr/hal-01000001.

[118] D. Fohr, I. Illina, "Continuous Word Representation using Neural Networks for Proper Name Retrieval from Diachronic Documents", *in : Interspeech 2015*, Dresden, Germany, September 2015, https://hal.archives-ouvertes.fr/hal-01184951.

[119] D. Fohr, I. Illina, "Neural Networks for Proper Name Retrieval in the Framework of Automatic Speech Recognition", *in : IEEE International Conference on Information Systems and Economic Intelligence*, hammamet, Tunisia, 2015, https://hal.archives-ouvertes.fr/hal-01184957.

[120] D. Fohr, O. Mella, D. Jouvet, "De l'importance de l'homogénéisation des conventions de transcription pour l'alignement automatique de corpus oraux de parole spontanée", *in : 8es Journées*

Internationales de Linguistique de Corpus (JLC2015), Orléans, France, September 2015, https://hal.inria.fr/hal-01183352.

[121] D. Fohr, O. Mella, "CoALT: A Software for Comparing Automatic Labelling Tools", *in : Language Resources and Evaluation LREC 2012*, p. 325–328, Istanbul, Turkey, May 2012, https://hal.archives-ouvertes.fr/hal-00761781.

[122] D. Fohr, O. Mella, "Combination of Random Indexing based Language Model and N-gram Language Model for Speech Recognition", *in : INTERSPEECH - 14th Annual Conference of the International Speech Communication Association - 2013*, p. –, Lyon, France, August 2013, https://hal.archives-ouvertes.fr/hal-00833898.

[123] D. Fohr, O. Mella, "Detection of Phone Boundaries for Non-Native Speech using French-German Models", *in : Workshop on Speech and Language Technology in Education*, Leipzig, Germany, September 2015, https://hal.archives-ouvertes.fr/hal-01185195.

[124] T. Fux, D. Jouvet, "Evaluation of PNCC and extended spectral subtraction methods for robust speech recognition", *in : EUSIPCO 2015 - 23rd European Signal Processing Conference*, Nice, France, August 2015, https://hal.inria.fr/hal-01183645.

[125] A. Gorin, D. Jouvet, E. Vincent, D. Tran, "Investigating Stranded GMM for Improving Automatic Speech Recognition", *in : 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2014)*, Nancy, France, May 2014, https://hal.inria.fr/hal-01003054.

[126] A. Gorin, D. Jouvet, "Class-based speech recognition using a maximum dissimilarity criterion and a tolerance classification margin", *in : SLT 2012 - 4th IEEE Workshop on Spoken Language Technology*, Miami, United States, December 2012, https://hal.inria.fr/hal-00753454.

[127] A. Gorin, D. Jouvet, "Efficient constrained parametrization of GMM with class-based mixture weights for Automatic Speech Recognition", *in : LTC'13 - 6th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, Poznań, Poland, December 2013, https://hal.inria.fr/hal-00923202.

[128] A. Gorin, D. Jouvet, "Component Structuring and Trajectory Modeling for Speech Recognition", *in : Interspeech*, Singapoore, Singapore, September 2014, https://hal.inria.fr/hal-01063653.

[129] A. Gorin, D. Jouvet, "Explicit trajectories and speaker class modeling for child and adult speech recognition", *in : XXXème édition des Journées d'Etudes sur la Parole*, Le Mans, France, June 2014, https://hal.inria.fr/hal-01080343.

[130] A. Gorin, D. Jouvet, "Structured GMM Based on Unsupervised Clustering for Recognizing Adult and Child Speech", *in : SLSP 2014, 2nd International Conference on Statistical Language and Speech Processing*, p. 108 – 119, Grenoble, France, October 2014, https://hal.inria.fr/hal-01090472.

[131] I. Illina, D. Fohr, D. Jouvet, "Grapheme-to-Phoneme Conversion using Conditional Random Fields", *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, International Speech Communication Association (ISCA) et The Italian Regional SIG - AISV (Italian Speech Communication Association), Florence, Italy, August 2011, https://hal.inria.fr/inria-00614981.

[132] I. Illina, D. Fohr, D. Jouvet, "Multiple Pronunciation Generation using Grapheme-to-Phoneme Conversion based on Conditional Random Fields", *in : XIV International Conference "Speech and Computer" (SPECOM'2011)*, Kazan, Russia, September 2011, https://hal.inria.fr/inria-00616325.

[133] I. Illina, D. Fohr, G. Linares, "Extension du vocabulaire d'un système de transcription avec de nouveaux noms propres en utilisant un corpus diachronique", *in : Journées d'Etude sur la parole*, Le Mans, France, June 2014, https://hal.inria.fr/hal-01092214.

[134] I. Illina, D. Fohr, G. Linares, "Proper Name Retrieval from Diachronic Documents for Automatic Speech Transcription using Lexical and Temporal Context", *in : Workshop on Speech, Language and Audio in Multimedia*, Penang, Malaysia, September 2014, https://hal.inria.fr/hal-01092224.

[135] I. Illina, D. Fohr, "Different word representations and their combination for proper name retrieval from diachronic documents", *in : IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2015)*, Scottsdale, United States, December 2015, https://hal.inria.fr/hal-01201533.

[136] I. Illina, D. Fohr, "Neural Networks Revisited for Proper Name Retrieval from Diachronic Documents", *in : LTC Language & Technology Conference*, p. 120–124, Poznan, Poland, November 2015, https://hal.archives-ouvertes.fr/hal-01240480.

[137] N. Ito, E. Vincent, N. Ono, S. Sagayama, "General algorithms for estimating spectrogram and transfer functions of target signal for blind suppression of diffuse noise", *in : 2013 IEEE International Workshop on Machine Learning for Signal Processing*, Southampton, United Kingdom, September 2013, https://hal.inria.fr/hal-00849791.

[138] X. Jaureguiberry, G. Richard, P. Leveau, R. Hennequin, E. Vincent, "Introducing a simple fusion framework for audio source separation", *in : 2013 IEEE International Workshop on Machine Learning for Signal Processing*, p. 6, Southampton, United Kingdom, September 2013, https://hal.archives-ouvertes.fr/hal-00846834.

[139] X. Jaureguiberry, E. Vincent, G. Richard, "Multiple-order non-negative matrix factorization for speech enhancement", *in : Interspeech*, p. 4, Singapore, June 2014, https://hal.archives-ouvertes.fr/hal-01023399.

[140] X. Jaureguiberry, E. Vincent, G. Richard, "Variational Bayesian model averaging for audio source separation", *in : SSP (IEEE Workshop on Statistical Signal Processing)*, p. 4, Australia, June 2014, https://hal.archives-ouvertes.fr/hal-00986909.

[141] I. Jemaa, K. Ouni, Y. Laprie, S. Ouni, J.-P. Haton, "A new Automatic Formant Tracking approach based on scalogram maxima detection using complex wavelets", *in : CEIT - International Conference on Control, Engineering & Information Technology - 2013*, Sousse, Tunisia, June 2013, https://hal.inria.fr/hal-00836854.

[142] D. Jouvet, K. Bartkova, "Acoustical Frame Rate and Pronunciation Variant Statistics", *in : International Conference on Statistical Language and Speech Processing*, Budapest, Hungary, November 2015, https://hal.inria.fr/hal-01184195.

[143] D. Jouvet, A. Bonneau, J. Trouvain, F. Zimmerer, Y. Laprie, B. Möbius, "Analysis of phone confusion matrices in a manually annotated French-German learner corpus", *in : Workshop on*

Speech and Language Technology in Education, Leipzig, Germany, September 2015, https://hal.inria.fr/hal-01184186.

[144] D. Jouvet, D. Fohr, I. Illina, "About handling boundary uncertainty in a speaking rate dependent modeling approach", *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, International Speech Communication Association (ISCA) et The Italian Regional SIG - AISV (Italian Speech Communication Association), Florence, Italy, August 2011, https://hal.inria.fr/inria-00614781.

[145] D. Jouvet, D. Fohr, I. Illina, "Building a Pronunciation Lexicon for a Speech Transcription System from Wiktionary Pronunciations only", *in : XIV International Conference "Speech and Computer" (SPECOM'2011)*, Kazan, Russia, September 2011, https://hal.inria.fr/inria-00616330.

[146] D. Jouvet, D. Fohr, I. Illina, "Evaluating grapheme-to-phoneme converters in automatic speech recognition context", *in : ICASSP - 2012 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4821 – 4824, Kyoto, Japan, March 2012, https://hal.inria.fr/hal-00753364.

[147] D. Jouvet, D. Fohr, "Analysis and Combination of Forward and Backward based Decoders for Improved Speech Transcription", *in : TSD - 16th International Conference on Text, Speech and Dialogue - 2013*, I. Habernal, V. Matoušek (editors), *Lecture Notes in Artificial Intelligence, 8082*, Springer Verlag, p. 84–91, Pilsen, Czech Republic, September 2013, https://hal.inria.fr/hal-00834296.

[148] D. Jouvet, D. Fohr, "Combining Forward-based and Backward-based Decoders for Improved Speech Recognition Performance", *in : InterSpeech - 14th Annual Conference of the International Speech Communication Association - 2013*, Lyon, France, August 2013, https://hal.inria.fr/hal-00834282.

[149] D. Jouvet, D. Fohr, "About Combining Forward and Backward-Based Decoders for Selecting Data for Unsupervised Training of Acoustic Models", *in : INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association*, Singapour, Singapore, September 2014, https://hal.inria.fr/hal-01090483.

[150] D. Jouvet, D. Langlois, "A Machine Learning Based Approach for Vocabulary Selection for Speech Transcription", *in : TSD - 16th International Conference on Text, Speech and Dialogue - 2013*, I. Habernal, V. Matoušek (editors), *Lecture Notes in Artificial Intelligence, 8082*, Springer Verlag, p. 60–67, Pilsen, Czech Republic, September 2013, https://hal.inria.fr/hal-00834302.

[151] D. Jouvet, L. Mesbahi, A. Bonneau, D. Fohr, I. Illina, Y. Laprie, "Impact of Pronunciation Variant Frequency on Automatic Non-Native Speech Segmentation", *in : 5th Language & Technology Conference - LTC'11*, p. 145–148, Poznan, Poland, November 2011, https://hal.archives-ouvertes.fr/hal-00639118.

[152] D. Jouvet, N. Vinuesa, "Classification margin for improved class-based speech recognition performance", *in : ICASSP - 2012 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4285 – 4288, Kyoto, Japan, March 2012, https://hal.inria.fr/hal-00753345.

[153] S. Kırbız, A. Ozerov, A. Liutkus, L. Girin, "Perceptual coding-based informed source separation", *in : 22nd European Signal Processing Conference (EUSIPCO-2014)*, Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01016314.

[154] O. Lachhab, J. Di Martino, E. H. Ibn Elhaj, A. Hammouch, "Real Time Context-Independent Phone Recognition Using a Simplified Statistical Training Algorithm", *in : 3rd International Conference on Multimedia Computing and Systems - ICMCS'12*, Tangier, Morocco, May 2012, https://hal.inria.fr/hal-00761816.

[155] O. Lachhab, J. Di Martino, E. H. Ibn Elhaj, A. Hammouch, "Improving the recognition of pathological voice using the discriminant HLDA transformation", *in : 3rd International IEEE Colloquium on Information Science and Technology*, Tetuan-Chefchaouen, Morocco, October 2014, https://hal.inria.fr/hal-01093309.

[156] Y. Laprie, M. Aron, M.-O. Berger, B. Wrobel-Dautcourt, "Studying MRI acquisition protocols of sustained sounds with a multimodal acquisition system", *in : 10th International Seminar on Speech Production (ISSP)*, Köln, Germany, May 2014, https://hal.inria.fr/hal-01002121.

[157] Y. Laprie, J. Busset, "A curvilinear tongue articulatory model", *in : International Seminar on Speech Production 2011 - ISSP'11*, Montréal, Canada, June 2011, https://hal.inria.fr/inria-00599109.

[158] Y. Laprie, J. Busset, "Construction and evaluation of an articulatory model of the vocal tract", *in : 19th European Signal Processing Conference - EUSIPCO-2011*, Barcelona, Spain, August 2011, https://hal.inria.fr/inria-00599130.

[159] Y. Laprie, B. Elie, A. Tsukanova, "2D Articulatory Velum Modeling Applied to Copy Synthesis of Sentences Containing Nasal Phonemes", *in : International Congress of Phonetic Sciences*, Glasgow, United Kingdom, August 2015, https://hal.inria.fr/hal-01188738.

[160] Y. Laprie, M. Loosvelt, S. Maeda, R. Sock, F. Hirsch, "Articulatory copy synthesis from cine X-ray films", *in : InterSpeech - 14th Annual Conference of the International Speech Communication Association - 2013*, Lyon, France, August 2013, https://hal.inria.fr/hal-00836838.

[161] Y. Laprie, B. Vaxelaire, M. Cadot, "Geometric articulatory model adapted to the production of consonants", *in : 10th International Seminar on Speech Production (ISSP)*, Köln, Germany, May 2014, https://hal.inria.fr/hal-01002125.

[162] Y. Laprie, "An articulatory model of the velum developed from cineradiographic data", *in : 169th Meeting: Acoustical Society of America*, Pittsburgh, United States, May 2015, https://hal.inria.fr/hal-01188760.

[163] A. Liutkus, R. Badeau, "Generalized Wiener filtering with fractional power spectrograms", *in : 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Brisbane, Australia, April 2015, https://hal.archives-ouvertes.fr/hal-01110028.

[164] A. Liutkus, D. Fitzgerald, R. Badeau, "Cauchy Nonnegative Matrix Factorization", *in : IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, United States, October 2015, https://hal.inria.fr/hal-01170924.

[165] A. Liutkus, D. Fitzgerald, Z. Rafii, "Scalable audio separation with light kernel additive modelling", *in : IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Brisbane, Australia, April 2015, https://hal.inria.fr/hal-01114890.

[166] A. Liutkus, T. Olubanjo, E. Moore, M. Ghovanloo, "Source Separation for Target Enhancement of Food Intake Acoustics from Noisy Recordings", *in : IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, United States, October 2015, https://hal.inria.fr/hal-01174886.

[167] A. Liutkus, Z. Rafii, B. Pardo, D. Fitzgerald, L. Daudet, "Kernel Spectrogram models for source separation", *in : HSCMA*, Nancy, France, May 2014, https://hal.inria.fr/hal-00959384.

[168] A. Liutkus, U. Şimşekli, T. Cemgil, "Extraction of Temporal Patterns in Multi-rate and Multi-modal Datasets", *in : International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, Liberec, Czech Republic, August 2015, https://hal.inria.fr/hal-01170932.

[169] S. Maeda, Y. Laprie, "Vowel and prosodic factor dependent variations of vocal-tract length", *in : InterSpeech - 14th Annual Conference of the International Speech Communication Association - 2013*, Lyon, France, August 2013, https://hal.inria.fr/hal-00836829.

[170] O. Mella, D. Fohr, A. Bonneau, "Inter-annotator agreement for a speech corpus pronounced by French and German language learners", *in : Workshop on Speech and Language Technology in Education*, ISCA Special Interest Group (SIG) on Speech and Language Technology in Education, Leipzig, Germany, September 2015, https://hal.archives-ouvertes.fr/hal-01185194.

[171] L. Mesbahi, D. Jouvet, A. Bonneau, D. Fohr, I. Illina, Y. Laprie, "Reliability of non-native speech automatic segmentation for prosodic feedback", *in : Workshop on Speech and Language Technology in Education - SLaTE 2011*, ISCA, Venise, Italy, August 2011, https://hal.inria.fr/inria-00614930.

[172] F. Mezzoudj, D. Langlois, D. Jouvet, A. Benyettou, "Textual Data Selection for Language Modelling in the Scope of Automatic Speech Recognition", *in : International Conference on Natural Language and Speech Processing*, Alger, Algeria, October 2015, https://hal.inria.fr/hal-01184192.

[173] J. Miranda, S. Ouni, "Mixing faces and voices: a study of the influence of faces and voices on audiovisual intelligibility", *in : AVSP - 12th International Conference on Auditory-Visual Speech Processing - 2013*, Annecy, France, August 2013, https://hal.inria.fr/hal-00835855.

[174] U. Musti, V. Colotte, S. Ouni, C. Lavecchia, B. Wrobel-Dautcourt, M.-O. Berger, "Automatic Feature Selection for Acoustic-Visual Concatenative Speech Synthesis: Towards a Perceptual Objective Measure", *in : AVSP - Audio Visual Speech Processing*, Annecy, France, September 2013, https://hal.inria.fr/hal-00925115.

[175] U. Musti, V. Colotte, A. Toutios, S. Ouni, "Introducing Visual Target Cost within an Acoustic-Visual Unit-Selection Speech Synthesizer", *in : International Conference on Auditory-Visual Speech Processing - AVSP2011*, Volterra, Italy, August 2011, https://hal.inria.fr/inria-00602403.

[176] U. Musti, C. Lavecchia, V. Colotte, S. Ouni, B. Wrobel-Dautcourt, M.-O. Berger, "ViSAC : Acoustic-Visual Speech Synthesis: The system and its evaluation", *in : FAA: The ACM 3rd International Symposium on Facial Analysis and Animation*, p. –, Vienne, Austria, September 2012, https://hal.archives-ouvertes.fr/hal-00762568.

[177] U. Musti, S. Ouni, Z. Ziheng, "3D Visual Speech Animation from Image Sequences", *in : Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, ACM, Bangalore, India, December 2014, https://hal.archives-ouvertes.fr/hal-01086073.

[178] I. Nkairi, I. Illina, G. Linarès, D. Fohr, "Exploring temporal context in diachronic text documents for automatic OOV proper name retrieval", *in : Language & Technology Conference*, p. 540–544, Poznań, Poland, December 2013, https://hal.archives-ouvertes.fr/hal-00924696.

[179] N. Ono, Z. Rafii, D. Kitamura, N. Ito, A. Liutkus, "The 2015 Signal Separation Evaluation Campaign", *in : International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, *Latent Variable Analysis and Signal Separation*, *9237*, p. 387–395, Liberec, France, August 2015, https://hal.inria.fr/hal-01188725.

[180] L. Orosanu, D. Jouvet, D. Fohr, I. Illina, A. Bonneau, "Combining criteria for the detection of incorrect entries of non-native speech in the context of foreign language learning", *in : SLT 2012 - 4th IEEE Workshop on Spoken Language Technology*, Miami, United States, February 2012, https://hal.inria.fr/hal-00753458.

[181] L. Orosanu, D. Jouvet, "Comparison and Analysis of Several Phonetic Decoding Approaches", *in : TSD - 16th International Conference on Text, Speech and Dialogue - 2013*, I. Habernal, V. Matoušek (editors), *Lecture Notes in Artificial Intelligence*, *8082*, Springer Verlag, p. 161–168, Pilsen, Czech Republic, September 2013, https://hal.inria.fr/hal-00834313.

[182] L. Orosanu, D. Jouvet, "Comparison of approaches for an efficient phonetic decoding", *in : InterSpeech - 14th Annual Conference of the International Speech Communication Association - 2013*, Lyon, France, August 2013, https://hal.inria.fr/hal-00834284.

[183] L. Orosanu, D. Jouvet, "Combining words and syllables for speech transcription", *in : XXXème édition des Journées d'Etudes sur la Parole*, Le Mans, France, June 2014, https://hal.inria.fr/hal-01080351.

[184] L. Orosanu, D. Jouvet, "Hybrid language models for speech transcription", *in : INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association*, Singapour, Singapore, September 2014, https://hal.inria.fr/hal-01090478.

[185] L. Orosanu, D. Jouvet, "Adding new words into a language model using parameters of known words with similar behavior", *in : International Conference on Natural Language and Speech Processing*, Alger, Algeria, October 2015, https://hal.inria.fr/hal-01184194.

[186] L. Orosanu, D. Jouvet, "Combining lexical and prosodic features for automatic detection of sentence modality in French", *in : International Conference on Statistical Language and Speech Processing*, Budapest, Hungary, November 2015, https://hal.inria.fr/hal-01184196.

[187] L. Orosanu, D. Jouvet, "Detection of sentence modality on French automatic speech-to-text transcriptions", *in : International Conference on Natural Language and Speech Processing*, Alger, Algeria, October 2015, https://hal.inria.fr/hal-01184193.

[188] S. Ouni, L. Mangeonjean, I. Steiner, "VisArtico: a visualization tool for articulatory data", *in : 13th Annual Conference of the International Speech Communication Association - InterSpeech 2012*, Portland, OR, United States, September 2012, https://hal.inria.fr/hal-00730733.

[189] S. Ouni, L. Mangeonjean, "VisArtico : visualiser les données articulatoires obtenues par un articulographe", *in : Actes de la conférence conjointe JEP-TALN-RECITAL 2012*, *1:JEP*, p. 129–135, Grenoble, France, June 2012, https://hal.archives-ouvertes.fr/hal-00762575.

[190] S. Ouni, "Tongue Gestures Awareness and Pronunciation Training", *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, ISCA (editor), Florence, Italy, August 2011. (accepted), https://hal.inria.fr/inria-00602418.

[191] A. Piquard-Kipffer, H. Adam-Piquard, S. Nussbaum, "Création d'une collection de livres numériques pour enfants présentant des difficultés de langage (enfants sourds, enfants avec troubles spécifiques ou retards de langage)", *in : L'apprentissage de la lecture : convergences, innovations, perspectives.*, Université de Cergy-Pontoise & IUFM, Gennevilliers, France, December 2011, https://hal.inria.fr/hal-00681701.

[192] A. Piquard-Kipffer, B. Christian, "Je peux voir les mots que tu dis ! Histoire d'un projet", *in : 13ème édition du Festival du film de chercheur CNRS 2012*, Nancy, France, June 2012, https://hal.inria.fr/hal-01263907.

[193] A. Piquard-Kipffer, O. Mella, J. Miranda, D. Jouvet, L. Orosanu, "Qualitative investigation of the display of speech recognition results for communication with deaf people", *in : 6th Workshop on Speech and Language Processing for Assistive Technologies*, SIG-SLPAT, p. 7, Dresden, Germany, September 2015, https://hal.inria.fr/hal-01183349.

[194] A. Piquard-Kipffer, "Reconnaissance de la parole, application aux personnes sourdes et malentendantes", *in : Journées scientifiques d'Inria*, Inria, Villers-Les-nancy, France, June 2015, https://hal.inria.fr/hal-01263928.

[195] A. Piquard-Kipffer, "Terminal portable de communication et affichage de la reconnaissance vocale. Enjeux et rapports à l'écrit. Etude préliminaire auprès d'adultes déficients auditifs.", *in : 3ème COLLOQUE INTERNATIONAL IDEKI 2015. Didactiques, Métiers de l'humain et Intelligence collective. Construction de savoirs et de dispositifs.*, IDEKI, COLMAR, France, December 2015, https://hal.inria.fr/hal-01243142.

[196] T. Prätzlich, R. Bittner, A. Liutkus, M. Müller, "Kernel additive modeling for interference reduction in multi-channel music recordings", *in : IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, April 2015, https://hal.inria.fr/hal-01116686.

[197] Z. Rafii, A. Liutkus, B. Pardo, "A simple user interface system for recovering patterns repeating in time and frequency in mixtures of sounds", *in : IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, France, April 2015, https://hal.inria.fr/hal-01116689.

[198] D. Ribas, E. Vincent, J. R. Calvo, "Full multicondition training for robust i-vector based speaker recognition", *in : Interspeech 2015*, Dresden, Germany, September 2015, https://hal.inria.fr/hal-01158774.

[199] D. Ribas, E. Vincent, J. R. Calvo, "Uncertainty propagation for noise robust speaker recognition: the case of NIST-SRE", *in : Interspeech 2015*, p. 5, Dresden, Germany, September 2015, https://hal.inria.fr/hal-01158775.

[200] C. Savariaux, P. Badin, S. Ouni, B. Wrobel-Dautcourt, "Étude comparée de la précision de mesure des systèmes d'articulographie électromagnétique 3D : Wave et AG500", *in : 29e Journées d'Études sur la Parole (JEP-TALN-RECITAL'2012)*, ATALA-AFCP (editor), p. 513–520, Grenoble, France, June 2012, https://hal.archives-ouvertes.fr/hal-00724682.

[201] I. Sheikh, I. Illina, D. Fohr, G. Linarès, "OOV Proper Name Retrieval using Topic and Lexical Context Model", *in : IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Australia, 2015, https://hal.archives-ouvertes.fr/hal-01184963.

[202] i. sheikh, I. Illina, D. Fohr, G. Linares, "Document Level Semantic Context for Retrieving OOV Proper Names", *in : 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, *Proceeding of IEEE ICASSP 2016*, IEEE, p. 6050–6054, shanghai, China, March 2016, https://hal.archives-ouvertes.fr/hal-01331716.

[203] I. Sheikh, I. Illina, D. Fohr, "Recognition of OOV Proper Names in Diachronic Audio News", *in : IEEE International Conference on Information Systems and Economic Intelligence*, Hammamet, Tunisia, 2015, https://hal.archives-ouvertes.fr/hal-01184958.

[204] i. sheikh, I. Illina, D. Fohr, "Study of Entity-Topic Models for OOV Proper Name Retrieval", *in : Interspeech 2015*, Dresden, Germany, September 2015, https://hal.archives-ouvertes.fr/hal-01184955.

[205] i. sheikh, I. Illina, D. Fohr, "How Diachronic Text Corpora Affect Context based Retrieval of OOV Proper Names for Audio News", *in : LREC 2016, proceedings of LREC 2016*, Portoroz, Slovenia, May 2016, https://hal.archives-ouvertes.fr/hal-01331714.

[206] S. Sivasankaran, A. A. Nugraha, E. Vincent, J. A. Morales Cordovilla, S. Dalmia, I. Illina, A. Liutkus, "Robust ASR using neural network based speech enhancement and feature simulation", *in : 2015 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2015)*, Arizona, United States, December 2015, https://hal.inria.fr/hal-01204553.

[207] R. Sock, F. Hirsch, Y. Laprie, P. Perrier, B. Vaxelaire, G. Brock, F. Bouarourou, C. Fauth, V. Ferbach-Hecker, L. Ma, J. Busset, J. Sturm, "An X-ray database, tools and procedures for the study of speech production", *in : 9th International Seminar on Speech Production (ISSP 2011)*, L. Ménard, S. Baum, V. Gracco, D. Ostry (editors), p. 41–48, Montréal, Canada, June 2011, https://hal.archives-ouvertes.fr/hal-00610297.

[208] N. Souviraà-Labastie, A. Olivero, E. Vincent, F. Bimbot, "Audio source separation using multiple deformed references", *in : Eusipco*, Lisboa, Portugal, September 2014, https://hal.inria.fr/hal-01017571.

[209] N. Souviraà-Labastie, E. Vincent, F. Bimbot, "Music separation guided by cover tracks: designing the joint NMF model", *in : 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015*, Brisbane, Australia, April 2015, https://hal.archives-ouvertes.fr/hal-01108675.

[210] I. Steiner, P. Knopp, S. Musche, A. Schmiedel, A. Braun, S. Ouni, "Investigating the effects of posture and noise on speech production", *in : 10th International Seminar on Speech Production (ISSP)*, Susanne Fuchs, Martine Grice, Anne Hermes, Leonardo Lancia, Doris Mücke, Cologne, Germany, May 2014, https://hal.archives-ouvertes.fr/hal-01086066.

[211] I. Steiner, S. Ouni, "Investigating articulatory differences between upright and supine posture using 3D EMA", *in : 9th International Seminar on Speech Production - ISSP'11*, UQAM, Montreal, Canada, June 2011, https://hal.inria.fr/inria-00602427.

[212] I. Steiner, S. Ouni, "Towards an articulatory tongue model using 3D EMA", *in : 9th International Seminar on Speech Production - ISSP'11*, UQAM, p. 147–154, Montreal, Canada, June 2011, https://hal.inria.fr/inria-00602423.

[213] I. Steiner, S. Ouni, "Artimate: an articulatory animation framework for audiovisual speech synthesis", *in : Workshop on Innovation and Applications in Speech Technology*, J. Kane (editor), UCD, TCD, Dublin, Ireland, March 2012, https://hal.inria.fr/hal-00678964.

[214] I. Steiner, K. Richmond, S. Ouni, "Using multimodal speech production data to evaluate articulatory animation for audiovisual speech synthesis", *in : 3rd International Symposium on Facial Analysis and Animation - FAA 2012*, Vienna, Austria, September 2012, https://hal.inria.fr/hal-00734464.

[215] I. Steiner, K. Richmond, S. Ouni, "Speech animation using electromagnetic articulography as motion capture data", *in : AVSP - 12th International Conference on Auditory-Visual Speech Processing - 2013*, p. 55–60, Annecy, France, August 2013, https://hal.inria.fr/hal-00835856.

[216] F.-R. Stöter, A. Liutkus, R. Badeau, B. Edler, P. Magron, "Common Fate Model for Unison source Separation", *in : 41st International Conference on Acoustics, Speech and Signal Processing (ICASSP), Proceedings of the 41st International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Shanghai, China, 2016, https://hal.archives-ouvertes.fr/hal-01248012.

[217] F. Stouten, I. Illina, D. Fohr, "Clustering repeated Out-Of-Vocabulary word tokens in order to model them for broadcast news transcription", *in : The XIVth International Conference Speech and Computer - SPECOM'2011*, p. 73–80, Kazan, Russia, September 2011, https://hal.archives-ouvertes.fr/hal-00639123.

[218] J. Thiemann, E. Vincent, "A fast EM algorithm for Gaussian model-based source separation", *in : EUSIPCO - 21st European Signal Processing Conference - 2013*, Marrakech, Morocco, September 2013, https://hal.inria.fr/hal-00840366.

[219] J. Thiemann, E. Vincent, "An experimental comparison of source separation and beamforming techniques for microphone array signal enhancement", *in : MLSP - 23rd IEEE International Workshop on Machine Learning for Signal Processing - 2013*, Southampton, United Kingdom, September 2013, https://hal.inria.fr/hal-00850173.

[220] A. Toutios, U. Musti, S. Ouni, V. Colotte, "Weight Optimization for Bimodal Unit-Selection Talking Head Synthesis", *in : 12thAnnual Conference of the International Speech Communication Association - Interspeech 2011*, ISCA (editor), Florence, Italy, August 2011. (accepted), https://hal.inria.fr/inria-00602407.

[221] A. Toutios, S. Ouni, "Predicting Tongue Positions from Acoustics and Facial Features", *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, ISCA (editor), Florence, Italy, August 2011, https://hal.inria.fr/inria-00602412.

[222] D. Tran, D. Jouvet, E. Vincent, "Fusion of Multiple Uncertainty Estimators and Propagators for Noise Robust ASR", *in : 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, https://hal.inria.fr/hal-00955185.

[223] D. Tran, N. Ono, E. Vincent, "Fast DNN training based on auxiliary function technique", *in : ICASSP 2015 - 40th IEEE International Conference on Acoustics, Speech and Signal Processing*, Brisbane, Queensland, Australia, April 2015, https://hal.inria.fr/hal-01107809.

[224] D. Tran, E. Vincent, D. Jouvet, "Extension of uncertainty propagation to dynamic MFCCs for noise robust ASR", *in : 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, Italy, May 2014, https://hal.inria.fr/hal-00954654.

[225] D. Tran, E. Vincent, D. Jouvet, "Discriminative uncertainty estimation for noise robust ASR", *in : 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015*, Brisbane, Queensland, Australia, April 2015, https://hal.inria.fr/hal-01103969.

[226] D. T. Tran, E. Vincent, D. Jouvet, K. Adiloglu, "Using full-rank spatial covariance models for noise-robust ASR", *in : CHiME - 2nd International Workshop on Machine Listening in Multisource Environments - 2013*, p. 31–32, Vancouver, Canada, June 2013, https://hal.inria.fr/hal-00801162.

[227] J. Trouvain, A. Bonneau, V. Colotte, C. Fauth, D. Fohr, D. Jouvet, J. Jügler, Y. Laprie, O. Mella, B. Möbius, F. Zimmerer, "The IFCASL Corpus of French and German Non-native and Native Read Speech", *in : LREC'2016, 10th edition of the Language Resources and Evaluation Conference, Proceedings LREC'2016*, Portorož, Slovenia, May 2016, https://hal.inria.fr/hal-01293935.

[228] E. Vincent, J. Barker, S. Watanabe, J. Le Roux, F. Nesta, M. Matassoni, "The Second 'CHiME' Speech Separation and Recognition Challenge: An overview of challenge systems and outcomes", *in : 2013 IEEE Automatic Speech Recognition and Understanding Workshop*, Olomouc, Czech Republic, December 2013, https://hal.inria.fr/hal-00862750.

[229] E. Vincent, A. Gkiokas, D. Schnitzer, A. Flexer, "An investigation of likelihood normalization for robust ASR", *in : Interspeech*, Singapore, Singapore, September 2014, https://hal.inria.fr/hal-01006142.

[230] F. Zimmerer, J. Trouvain, A. Bonneau, "One corpus, one research question, three methods "German vowels produced by French speakers"", *in : Worshop on Phonetic learner corpora. Satellite meeting of ICPhS 2015.*, Trouvain, J., Zimmerer, F., Gosy, M., Bonneau, A., Glasgow, United Kingdom, August 2015, https://hal.inria.fr/hal-01186078.

Articles in National Peer-Reviewed Journal

[231] I. Illina, D. Fohr, G. Linarès, "Ajout de nouveaux noms propres au vocabulaire d'un système de transcription en utilisant un corpus diachronique", *Traitement Automatique des Langues 55*, 2, 2014, p. 47–72, https://hal.archives-ouvertes.fr/hal-01184950.

National Peer-Reviewed Conferences

[232] J. Busset, M. Cadot, "Fouille d'images animées : cinéradiographies d'un locuteur", *in : FOSTA 2013, atelier de EGC 2013*, p. 1–12, Toulouse, France, January 2013, https://hal.archives-ouvertes.fr/hal-00773448.

[233] M. Cadot, Y. Laprie, "Méthodologie 3-way d'extraction d'un modèle articulatoire de la parole à partir des données d'un locuteur", *in : Atelier Fouille de Données Complexes des 14èmes Journées Francophones "Extraction et Gestion des Connaissances"*, p. 1–12, Rennes, France, January 2014, https://hal.archives-ouvertes.fr/hal-00934436.

[234] B. Elie, Y. Laprie, P.-A. Vuissoz, "Acquisition temps- réel de données articulatoires par IRM : application a la synthèse par copie", *in : 13ème Congrès Français d'Acoustique (CFA 2016)*, SFA, Le Mans, France, April 2016, https://hal.archives-ouvertes.fr/hal-01314313.

[235] B. Elie, Y. Laprie, "Estimation de la longueur du conduit vocal pour l'inversion acoustique-articulatoire", *in : Congrès Français d'Acoustique*, SFA, Poitiers, France, April 2014, https://hal.archives-ouvertes.fr/hal-01096571.

[236] I. Illina, D. Fohr, D. Jouvet, "Génération des prononciations de noms propres à l'aide des champs aéatoires conditionnels", *in : JEP-TALN-RECITAL 2012*, Grenoble, France, June 2012, https://hal.inria.fr/hal-00753381.

[237] D. Jouvet, A. Gorin, N. Vinuesa, "Exploitation d'une marge de tolérance de classification pour améliorer l'apprentissage de modèles acoustiques de classes en reconnaissance de la parole", *in : JEP-TALN-RECITAL 2012*, p. 763–770, Grenoble, France, June 2012, https://hal.inria.fr/hal-00753394.

[238] L. Orosanu, D. Jouvet, D. Fohr, I. Illina, A. Bonneau, "Détection de transcriptions incorrectes de parole non-native dans le cadre de l'apprentissage de langues étrangères", *in : JEP-TALN-RECITAL 2012*, Grenoble, France, June 2012, https://hal.inria.fr/hal-00753387.

[239] J. Thiemann, E. Vincent, S. Van De Par, "Spatial properties of the DEMAND noise recordings", *in : 40th Annual German Congress on Acoustics (DAGA 2014)*, Oldenburg, Germany, March 2014, https://hal.inria.fr/hal-00985979.

Books or Proceedings Editing

[240] *LNCS 9237 - Proceedings of the 12th International Conference on Latent Variable Analysis and Signal Separation*, Liberec, Czech Republic, Springer, August 2015, https://hal.inria.fr/hal-01183618.

[241] Inria (editor), *The 12th International Conference on Auditory-Visual Speech Processing*, Slim Ouni and Frédéric Berthomier and Alexandra Jesse, Annecy, France, August 2013, 247p. Proceedings are online: http://avsp2013.loria.fr/proceedings, https://hal.inria.fr/hal-01188878.

Book chapters

[242] A. Bonneau, V. Colotte, "Automatic Feedback for L2 Prosody Learning", *in : Speech and Language Technologies*, I. Ipsic (editor), Intech, June 2011, p. 55–70, https://hal.inria.fr/inria-00579255.

[243] M. Embarki, S. Ouni, C. Guilleminot, M. Yeou, S. Al Maqtari, "Acoustic and electromagnetic articulographic study of pharyngealisation", *in : Instrumental Studies in Arabic Phonetics*, Z. M. Hassan and B. Heselwood (editors), *Current Issues in Linguistic Theory*, Benjamins, December 2011, p. 193–215, https://hal.archives-ouvertes.fr/hal-00671220.

[244] M. Embarki, S. Ouni, M. Yeou, C. Guilleminot, S. Al Maqtari, "Acoustic and EMA study of pharyngealization : Coarticulatory effects as index of stylistic and regional distinction", *in : Instrumental Studies in Arabic Phonetics*, M. Hassan and B. Heselwood (editors), *Current Issues in Linguistic Theory*, Benjamins, 2011, p. 1–56, https://hal.archives-ouvertes.fr/hal-00348775.

[245] Z. Rafii, A. Liutkus, B. Pardo, "REPET for Background/Foreground Separation in Audio", *in : Blind Source Separation*, G. Naik and W. Wang (editors), Springer Berlin Heidelberg, 2014, p. 395–411, https://hal.inria.fr/hal-01025563.

[246] I. Steiner, S. Ouni, "Progress in animation of an EMA-controlled tongue model for acoustic-visual speech synthesis", *in : Elektronische Sprachsignalverarbeitung 2011*, B. J. Kröger and P. Birkholz (editors), TUDpress, 2011, p. 245–252, https://hal.inria.fr/hal-00661045.

Other Publications

[247] R. Badeau, A. Liutkus, "Proof of Wiener-like linear regression of isotropic complex symmetric alpha-stable random variables", *research report*, September 2014, https://hal.archives-ouvertes.fr/hal-01069612.

[248] A. Bonneau, "Phonetic variation in non-native speech", Spring School : "Individual-centered Approaches to Speech Processing", April 2014, https://hal.inria.fr/hal-01095804.

[249] P. Carroll, J. Trouvain, F. Zimmerer, Y. Laprie, O. Mella, D. Fohr, "Improvements for a German Vowel Trainer CAPT Tool", Individualized Feedback for Computer-Assisted Spoken Language Learning, November 2015, Poster, https://hal.archives-ouvertes.fr/hal-01243043.

[250] L. Catanese, N. Souviraà-Labastie, B. Qu, S. Campion, G. Gravier, E. Vincent, F. Bimbot, "MODIS: an audio motif discovery software", Show & Tell - Interspeech 2013, August 2013, Poster, https://hal.inria.fr/hal-00931227.

[251] B. Christian, L. Denis, P.-K. Agnès, "Je peux voir les mots que tu dis !", 2011, Film de recherche réalisé par Christian Blonz, Inria-Rocquencourt. Co-scénaristes : Blonz Christian, Lelarge Denis, Piquard-Kipffer Agnès, https://hal.inria.fr/hal-00681724.

[252] V. Colotte, C. Emilien, "JCorpusRecorder", *Technical report*, Université de Lorraine, December 2015, https://hal.inria.fr/hal-01244553.

[253] B. Elie, Y. Laprie, "Extension of the single-matrix formulation of the vocal tract: consideration of bilateral channels and connection of self-oscillating models of vocal folds with glottal chink", working paper or preprint, September 2015, https://hal.archives-ouvertes.fr/hal-01199792.

[254] B. Elie, Y. Laprie, "Copy synthesis of phrase-level utterances", working paper or preprint, February 2016, https://hal.archives-ouvertes.fr/hal-01278462.

[255] J. Le Roux, E. Vincent, "A categorization of robust speech processing datasets", *Technical Report number Mitsubishi Electric Research Labs TR2014-116*, September 2014, https://hal.inria.fr/hal-01063805.

[256] T. Léonova, P.-K. Agnès, "Parcours des enfants présentant des Troubles Spécifiques du Langage (TSL) en situation de handicap. Région Lorraine Enfance de 4 à 20 ans", *Research report*, ARS, 2011, https://hal.inria.fr/hal-01263893.

[257] A. Liutkus, "Scale-Space Peak Picking", *Research report*, Inria Nancy - Grand Est (Villers-lès-Nancy, France), January 2015, https://hal.inria.fr/hal-01103123.

[258] M. Moussallam, A. Liutkus, L. Daudet, "Listening to features", *Research report*, Institut Langevin, ESPCI - CNRS - Paris Diderot University - UPMC, January 2015, https://hal.inria.fr/hal-01118307.

[259] S. Ouni, G. Gris, "Dynamic realistic lip animation using a limited number of control points", SIGGRAPH 2015, ACM, *Proceeding SIGGRAPH '15 ACM SIGGRAPH 2015 Posters*, August 2015, Poster, https://hal.inria.fr/hal-01188997.

[260] A. Piquard-Kipffer, O. Mella, J. Miranda, D. Jouvet, L. Orosanu, "Terminal portable de communication et affichage de la reconnaissance vocale. Enjeux et rapports à l'écrit. Étude préliminaire auprès d'adultes déficients auditifs.", March 2016, In M.Frisch (Eds) Le réseau Idéki : Didactiques, métiers de l'humain et Intelligence collective. Nouveaux espaces et dispositifs en question. Nouveaux horizons en éducation, formation et en recherche. L'harmattan, Collection I.D., https://hal.inria.fr/hal-01239910.

[261] A. Piquard-Kipffer, “Développement et apprentissages : enjeux, démarches, perspectives”, May 2012, Conférence invitée in Séminaire enfants, famille et société : droits, développement et apprentissage. EHESP.Paris., https://hal.inria.fr/hal-01263910.

[262] A. Piquard-Kipffer, “Critères d’évaluation d’un album numérique pour des enfants en difficulté de langage”, December 2014, In M. Frisch (Eds) Le réseau Idéki : objets de recherche, d’éducation et de formation émergents, problématisés, mis en tension, réélaborés. Préface de Joël Lebeaume. Paris : L’harmattan, Collection I.D, 287-309, https://hal.inria.fr/hal-01097278.

[263] A. Piquard-Kipffer, “Poursuivre une scolarité avec une langue déficiente”, January 2014, Conférence invitée dans un séminaire ”Troubles du langage et des apprentissages”, https://hal.inria.fr/hal-01263922.

[264] A. Piquard-Kipffer, “Les troubles Dys : la dyslexie-dysorthographie”, January 2015, Conférence invitée dans le cadre des Conf’curieuses, Musée et aquarium de Nancy, Grand Nancy & Université de Lorraine., https://hal.inria.fr/hal-01263931.

[265] A. Rousseau, A. Darnaud, B. Goglin, C. Acharian, C. Leininger, C. Godin, C. Holik, C. Kirchner, D. Rives, E. Darquie, E. Kerrien, F. Neyret, F. Masseglia, F. Dufour, G. Berry, G. Dowek, H. Robak, H. Xypas, I. Illina, I. Gnaedig, J. Jongwane, J. Ehrel, L. Viennot, L. Guion, L. Calderan, L. Kovacic, M. Collin, M.-A. Enard, M.-H. Comte, M. Quinson, M. Olivi, M. Giraud, M. Dorémus, M. Ogouchi, M. Droin, N. Lacaux, N. P. Rougier, N. Roussel, P. Guitton, P. Peterlongo, R.-M. Cornus, S. Vandermeersch, S. Maheo, S. Lefebvre, S. Boldo, T. Viéville, V. Poirel, A. Chabreuil, A. Fischer, C. Farge, C. Vadel, I. Astic, J.-P. Dumont, L. Féjoz, P. Rambert, P. Paradinas, S. De Quatrebarbes, S. Laurent, “Médiation Scientifique : une facette de nos métiers de la recherche”, *Interne*, none, March 2013, https://hal.inria.fr/hal-00804915.

[266] Y. Salaün, E. Vincent, N. Bertin, N. Souviraà-Labastie, X. Jaureguiberry, D. T. Tran, F. Bimbot, “The Flexible Audio Source Separation Toolbox Version 2.0”, ICASSP, May 2014, Poster, https://hal.inria.fr/hal-00957412.

[267] L. S. R. Simon, E. Vincent, “Combining blockwise and multi-coefficient stepwise approches in a general framework for online audio source separation”, *Research Report number RR-8766*, Inria, August 2015, https://hal.inria.fr/hal-01186948.

[268] E. Vincent, J. Barker, S. Watanabe, J. Le Roux, F. Nesta, M. Matassoni, “Overview of the 2nd ’CHiME’ Speech Separation and Recognition Challenge”, CHiME - 2nd International Workshop on Machine Listening in Multisource Environments - 2013, June 2013, CHiME - 2nd International Workshop on Machine Listening in Multisource Environments - 2013, https://hal.inria.fr/hal-00833252.

[269] E. Vincent, E. A. P. Habets, “Advanced spatial speech and audio processing”, August 2015, LVA/ICA 2015 Summer School on Latent Variable Analysis and Signal Separation, https://hal.inria.fr/hal-01183505.

[270] E. Vincent, “Incertitudes en traitement de la parole et de l’audio”, Journée scientifique de la Fédération Charles Hermite ” Incertitudes: approches et défis ”, December 2013, Journée scientifique de la Fédération Charles Hermite ” Incertitudes: approches et défis ”, https://hal.inria.fr/hal-00923070.

[271] E. Vincent, “Introduction to sound scene analysis - Source localization”, Journées Thématiques ’Localisation, séparation et suivi de sources audio’, February 2013, Journées Thématiques ’Localisation, séparation et suivi de sources audio’, https://hal.inria.fr/hal-00833535.

[272] E. Vincent, "Leveraging online sound exposure and big data", 2013 Hearing Aid Developers Forum, June 2013, 2013 Hearing Aid Developers Forum, https://hal.inria.fr/hal-00833239.

[273] E. Vincent, "Source classification", Journées Thématiques 'Localisation, séparation et suivi de sources audio', February 2013, Journées Thématiques 'Localisation, séparation et suivi de sources audio', https://hal.inria.fr/hal-00833537.

[274] E. Vincent, "Source separation", Journées Thématiques 'Localisation, séparation et suivi de sources audio', February 2013, Journées Thématiques 'Localisation, séparation et suivi de sources audio', https://hal.inria.fr/hal-00833536.

[275] E. Vincent, "Evaluation campaigns and reproducibility", Journée GdR ISIS "reproductibilité en traitement du signal et des images", January 2014, Journée GdR ISIS "reproductibilité en traitement du signal et des images", https://hal.inria.fr/hal-00927741.

[276] E. Vincent, "Les sons à domicile", Séminaire SAILOR "Imaginer des nouveaux lieux de vie", April 2014, Séminaire SAILOR "Imaginer des nouveaux lieux de vie", https://hal.inria.fr/hal-00977674.

[277] E. Vincent, "Is audio signal processing still useful in the era of machine learning?", October 2015, 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), https://hal.inria.fr/hal-01183506.

[278] P.-A. Vuissoz, F. Odille, Y. Laprie, E. Vincent, J. Felblinger, "Sound synchronization and motion compensated reconstruction for speech Cine MRI", ISMRM 2015 Annual Meeting, May 2015, Poster, https://hal.inria.fr/hal-01183504.

[279] P.-A. Vuissoz, F. Odille, Y. Laprie, E. Vincent, G. Hossu, J. Felblinger, "Speech Cine SSFP with optical microphone synchronization and motion compensated reconstruction", ISMRM Workshop on Motion Correction in MRI, July 2014, Poster, https://hal.inria.fr/hal-00994526.

[280] P.-A. Vuissoz, F. Odille, E. Vincent, J. Felblinger, Y. Laprie, "Synchronisation vocale et mouvement compensé en reconstruction pour une ciné IRM de la parole", 2e Congrès de la SFRMBM (Société Française de Résonance Magnétique en Biologie et Médecine), March 2015, Poster, https://hal.inria.fr/hal-01104230.

Patents

[281] S. Ouni, G. Gris, *Dispositif de traitement d'image*, Patent, France, 15 52058, January 2016, Le rapport de recherche reconnait la brevetabilité., https://hal.inria.fr/hal-01294028.

3 References for Orpailleur

Doctoral Dissertations

[282] M. Alam, *Interactive Knowledge Discovery over Web of Data.*, Theses, Loria & Inria Grand Est, December 2015, https://hal.inria.fr/tel-01245458.

[283] Y. Asses, *c-Met receptor inhibitors design by molecular modeling and in silico screening*, Theses, Université Henri Poincaré - Nancy I, October 2011, https://tel.archives-ouvertes.fr/tel-00653609.

[284] S. Benabderrahmane, *Using domain knowledge in the transcriptome analysis: semantic similarity, functional classification and fuzzy profiles. Application to colorectal cancer.*, Theses, Université Henri Poincaré - Nancy I, December 2011, https://tel.archives-ouvertes.fr/tel-00653169.

[285] E. Bresso, *Organization and exploitation of biological molecular networks for studying the etiology of genetic diseases and for characterizing drug side effects*, Theses, Université de Lorraine, September 2013, https://tel.archives-ouvertes.fr/tel-00917934.

[286] A. Buzmakov, *Formal Concept Analysis and Pattern Structures for mining Structured Data*, Theses, Universite de Lorraine, October 2015, https://hal.inria.fr/tel-01229062.

[287] V. Codocedo-Henriquez, *Contributions to indexing and retrieval using Formal Concept Analysis*, Theses, Université de Lorraine, September 2015, https://hal.inria.fr/tel-01241474.

[288] J. Cojan, *Application de la théorie de la révision des connaissances au raisonnement à partir de cas*, Theses, Université Henri Poincaré - Nancy I, October 2011, https://tel.archives-ouvertes.fr/tel-00646841.

[289] S. Da Silva, *Spatial data mining and modelling of hedgrows in agricultural landscapes*, Theses, Université de Lorraine, September 2014, https://hal.inria.fr/tel-01101424.

[290] V. Dufour-Lussier, *Reasoning with Qualitative Spatial and Temporal Textual Cases*, Theses, Université de Lorraine, October 2014, https://hal.inria.fr/tel-01098087.

[291] E. Egho, *Mining Heterogeneous Multidimensional Sequential Data*, Theses, Université de Lorraine, July 2014, https://hal.inria.fr/tel-01094400.

[292] A. Ghoorah, *Knowledge-based Approaches for Modelling the 3D Structural Interactome*, Theses, Université de Lorraine, November 2012, https://tel.archives-ouvertes.fr/tel-00762444.

[293] M. Kaytoue, *Mining numerical data with formal concept analysis and pattern structures*, Theses, Université Henri Poincaré - Nancy I, April 2011, https://tel.archives-ouvertes.fr/tel-00599168.

[294] E. G. Lazrak, *Stochastic data-mining for understanding time and space dynamics of agricultural landscapes. Contribution to numerical methods for landscape agronomy.*, Theses, Université de Lorraine, September 2012, https://tel.archives-ouvertes.fr/tel-00782768.

[295] T. Meilender, *A Semantic Wiki for Decision Knowledge Management - Application in Oncology*, Theses, Université de Lorraine, June 2013, https://tel.archives-ouvertes.fr/tel-00919997.

[296] D. Ritchie, *High Performance Algorithms for Molecular Shape Recognition*, Habilitation à diriger des recherches, Université Henri Poincaré - Nancy I, April 2011, https://tel.archives-ouvertes.fr/tel-00587962.

[297] M. Smaïl-Tabbone, *Contribution to knowledge extraction from biological data*, Habilitation à diriger des recherches, Université de Lorraine, November 2014, https://hal.inria.fr/tel-01093943.

[298] Y. Toussaint, *Text Mining: Symbolic methods to build ontologies and to semantically annotate texts*, Habilitation à diriger des recherches, Université Henri Poincaré - Nancy I, November 2011, https://tel.archives-ouvertes.fr/tel-00764162.

Articles in International Peer-Reviewed Journal

[299] A. K. R. Abadio, E. S. Kioshima, M. M. Teixeira, N. F. Martins, B. Maigret, M. S. S. Felipe, "Comparative genomics allowed the identification of drug targets against human fungal pathogens.", *BMC Genomics 12*, January 2011, p. 75, https://hal.inria.fr/inria-00610630.

[300] N. Agrinier, C. Altieri, F. Alla, N. Jay, D. Dobre, N. Thilly, F. Zannad, "Effectiveness of a multidimensional home nurse led heart failure disease management program-A French nationwide time-series comparison", *International Journal of Cardiology 168*, 4, October 2013, p. 3652–8, https://hal.inria.fr/hal-00903263.

[301] J. Almeida, M. Couceiro, T. Waldhauser, "On the topological semigroup of equational classes of finite functions under composition", *Journal of Multiple-Valued Logic and Soft Computing*, 2016, p. 24, https://hal.archives-ouvertes.fr/hal-01090645.

[302] J. Baixeries, M. Kaytoue, A. Napoli, "Characterizing functional dependencies in formal concept analysis with pattern structures", *Annals of Mathematics and Artificial Intelligence 72*, 2014, p. 129 – 149, https://hal.inria.fr/hal-01101107.

[303] A. Berry, A. Gutierrez, M. Huchard, A. Napoli, A. Sigayret, "Hermes: a simple and efficient algorithm for building the AOC-poset of a binary relation", *Annals of Mathematics and Artificial Intelligence ISSN*, 72, 2014, p. 45 – 71, https://hal.inria.fr/hal-01101144.

[304] S. Bessy, D. Gonçalves, J.-S. Sereni, "Two floor building needing eight colors", *Journal of Graph Algorithms and Applications 19*, 1, January 2015, p. 1–9, https://hal.archives-ouvertes.fr/hal-00996709.

[305] F. Bonomo, G. Duran, A. Napoli, M. Valencia-Pabon, "A one-to-one correspondence between potential solutions of the cluster deletion problem and the minimum sum coloring problem, and its application to P4 -sparse graphs", *Information Processing Letters (IPL) 115*, 6-8, 2015, p. 600–603, Paper submitted to a journal, 11/05/2014, https://hal.archives-ouvertes.fr/hal-01102515.

[306] F. Bonomo, G. Duran, M. Valencia-Pabon, "Complexity of the cluster deletion problem on some subclasses of chordal graphs", *Journal of Theoretical Computer Science (TCS) 600*, 2015, p. 59–69, Paper submitted to a Journal, 28/10/2014, https://hal.archives-ouvertes.fr/hal-01102512.

[307] F. Bonomo, O. Schaudt, M. Stein, M. Valencia-Pabon, "b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs", *Algorithmica*, 2014, p. 17, An extended abstract of this paper appears in the proceedings of ISCO symposium, LNCS 8596, pp. 100-11, 2014, https://hal.archives-ouvertes.fr/hal-01102516.

[308] G. Bosc, P. Tan, J.-F. Boulicaut, C. Raïssi, M. Kaytoue, "A Pattern Mining Approach to Study Strategy Balance in RTS Games", *IEEE Transactions on Computational Intelligence and AI in Games (T-CIAIG)*, December 2015, https://hal.archives-ouvertes.fr/hal-01252728.

[309] T. Bourquard, J. Bernauer, J. Azé, A. Poupon, "A collaborative filtering approach for protein-protein docking scoring functions.", *PLoS ONE 6*, 4, 2011, p. e18541, https://hal.inria.fr/inria-00625000.

[310] E. Bresso, R. Grisoni, G. Marchetti, A.-S. Karaboga, M. Souchet, M.-D. Devignes, M. Smaïl-Tabbone, "Integrative relational machine-learning for understanding drug side-effect profiles", *BMC Bioinformatics 14*, 1, 2013, p. 207, https://hal.inria.fr/hal-00843914.

[311] A. Buzmakov, E. Egho, N. Jay, S. O. Kuznetsov, A. Napoli, C. Raïssi, "On Mining Complex Sequential Data by Means of FCA and Pattern Structures", *Int. J. Gen. Syst.*, 2015, p. 1–25, https://hal.archives-ouvertes.fr/hal-01186715.

[312] A. Buzmakov, S. O. Kuznetsov, A. Napoli, "Is Concept Stability a Measure for Pattern Selection?", *Procedia Computer Science 31*, 2014, p. 918 – 927, https://hal.inria.fr/hal-01095914.

[313] T. Caradec, M. Pupin, A. Vanvlassenbroeck, M.-D. Devignes, M. Smaïl-Tabbone, P. Jacques, V. Leclère, "Prediction of Monomer Isomery in Florine: A Workflow Dedicated to Non-ribosomal Peptide Discovery", *PLoS ONE 9*, 1, January 2014, p. e85667, https://hal.archives-ouvertes.fr/hal-01090619.

[314] A. Carrieri, V. I. Pérez-Nueno, G. Lentini, D. Ritchie, "Recent trends and future prospects in computational GPCR drug discovery: from virtual screening to polypharmacology", *Current Topics in Medicinal Chemistry 13*, 9, 2013, p. 1069–1097, https://hal.inria.fr/hal-00880351.

[315] M. Chavent, B. Lévy, M. Krone, K. Bidmon, J.-P. Nominé, T. Ertl, M. Baaden, "GPU-powered tools boost molecular visualization.", *Briefings in Bioinformatics 12*, 6, November 2011, p. 689–701, https://hal.archives-ouvertes.fr/hal-00645161.

[316] M. Chavent, A. Vanel, A. Tek, B. Levy, S. Robert, B. Raffin, M. Baaden, "GPU-accelerated atom and dynamic bond visualization using hyperballs: a unified algorithm for balls, sticks, and hyperboloids.", *Journal of Computational Chemistry 32*, 13, October 2011, p. 2924–35, https://hal.archives-ouvertes.fr/hal-00645162.

[317] V. Codocedo, I. Lykourentzou, A. Napoli, "A semantic approach to concept lattice-based information retrieval", *Annals of Mathematics and Artificial Intelligence 72*, October 2014, p. 169 – 195, https://hal.inria.fr/hal-01095859.

[318] V. Coevoet, J. Fresson, R. Vieux, N. Jay, "Socioeconomic deprivation and hospital length of stay: a new approach using area-based socioeconomic indicators in multilevel models", *Medical Care 51*, 6, June 2013, p. 548–54, https://hal.inria.fr/hal-00903258.

[319] M. Couceiro, M. Behrisch, E. Lehtonen, K. A. Kearnes, Á. Szendrei, "Commuting polynomial operations of distributive lattices", *Order 29*, 2, 2012, p. 25, https://hal.archives-ouvertes.fr/hal-01090597.

[320] M. Couceiro, D. Dubois, H. Prade, T. Waldhauser, "Decision-making with sugeno integrals", *Order*, December 2015, https://hal.archives-ouvertes.fr/hal-01184340.

[321] M. Couceiro, L. Haddad, I. G. Rosenberg, "Partial clones containing all Boolean monotone self-dual partial functions", *Journal of Multiple-Valued Logic and Soft Computing*, 2015, p. 10, https://hal.archives-ouvertes.fr/hal-01093942.

[322] M. Couceiro, L. Haddad, K. Schölzel, T. Waldhauser, "A Solution to a Problem of D. Lau: Complete Classification of Intervals in the Lattice of Partial Boolean Clones", *Journal of Multiple-Valued Logic and Soft Computing*, 2016, p. 8, https://hal.inria.fr/hal-01183004.

[323] M. Couceiro, L. Haddad, "Intersections of finitely generated maximal partial clones", *Journal of Multiple-Valued Logic and Soft Computing 19*, 1-3, 2012, p. 10, https://hal.archives-ouvertes.fr/hal-01090584.

[324] M. Couceiro, E. Lehtonen, K. Schölzel, "A complete classification of equational classes of threshold functions included in clones", *RAIRO - Operations Research 49*, 1, 2015, p. 39–66, To appear: RAIRO - Operations Research, https://hal.archives-ouvertes.fr/hal-01090621.

[325] M. Couceiro, E. Lehtonen, K. Schölzel, "Hypomorphic Sperner Systems and Non-Reconstructible Functions", *Order 32*, 2, July 2015, p. 255–292, https://hal.archives-ouvertes.fr/hal-01090540.

[326] M. Couceiro, E. Lehtonen, K. Schölzel, "Set-reconstructibility of Post classes", *Discrete Applied Mathematics 187*, May 2015, p. 12–18, 8 pages. arXiv admin note: text overlap with arXiv:1306.5578, https://hal.archives-ouvertes.fr/hal-01090618.

[327] M. Couceiro, E. Lehtonen, T. Waldhauser, "Decompositions of functions based on arity gap", *Discrete Mathematics 312*, 2, 2012, p. 10, https://hal.archives-ouvertes.fr/hal-01090606.

[328] M. Couceiro, E. Lehtonen, T. Waldhauser, "The arity gap of order-preserving functions and extensions of pseudo-Boolean functions", *Discrete Applied Mathematics 160*, 2012, p. 8, https://hal.archives-ouvertes.fr/hal-01090604.

[329] M. Couceiro, E. Lehtonen, T. Waldhauser, "Additive decomposability of functions over Abelian groups", *International Journal of Algebra and Computation 23*, 3, 2013, p. 20, https://hal.archives-ouvertes.fr/hal-01090576.

[330] M. Couceiro, E. Lehtonen, T. Waldhauser, "Parametrized arity gap", *Order 30*, 2, 2013, p. 16, https://hal.archives-ouvertes.fr/hal-01090574.

[331] M. Couceiro, E. Lehtonen, T. Waldhauser, "Additive decomposition schemes for polynomial functions over fields", *Novi Sad J. Math. 44*, 2, 2014, p. 16, https://hal.archives-ouvertes.fr/hal-01090554.

[332] M. Couceiro, E. Lehtonen, T. Waldhauser, "On equational definability of function classes", *Journal of Multiple-Valued Logic and Soft Computing 24*, 1–4, 2015, p. 203–222, https://hal.archives-ouvertes.fr/hal-01093668.

[333] M. Couceiro, E. Lehtonen, "Galois theory for sets of operations closed under permutation, cylindrification and composition", *Algebra Universalis 67*, 3, 2012, p. 25, https://hal.archives-ouvertes.fr/hal-01090609.

[334] M. Couceiro, E. Lehtonen, "A survey on the arity gap", *Journal of Multiple-Valued Logic and Soft Computing 24*, 1–4, 2015, p. 223–249, https://hal.archives-ouvertes.fr/hal-01093666.

[335] M. Couceiro, E. Lehtonen, "On the arity gap of finite functions : results and applications.", *Journal of Multiple-Valued Logic and Soft Computing 27*, 1-2, 2016, p. 15, https://hal.inria.fr/hal-01175695.

[336] M. Couceiro, J.-L. Marichal, B. Teheux, "Conservative Median Algebras and Semilattices", *Order*, May 2015, p. 11, https://hal.inria.fr/hal-01175696.

[337] M. Couceiro, J.-L. Marichal, B. Teheux, "Relaxations of associativity and preassociativity for variadic functions", *Fuzzy Sets and Systems*, November 2015, https://hal.archives-ouvertes.fr/hal-01184334.

[338] M. Couceiro, J.-L. Marichal, T. Waldhauser, "Locally monotone Boolean and pseudo-Boolean functions", *Discrete Applied Mathematics 160*, 12, 2012, p. 10, https://hal.archives-ouvertes.fr/hal-01090591.

[339] M. Couceiro, J.-L. Marichal, "Aczelian n-ary semigroups", *Semigroup Forum 85*, 1, 2012, p. 10, https://hal.archives-ouvertes.fr/hal-01090588.

[340] M. Couceiro, J.-L. Marichal, "Polynomial functions over bounded distributive lattices", *Journal of Multiple-Valued Logic and Soft Computing 18*, 2012, p. 10, https://hal.archives-ouvertes.fr/hal-01090600.

[341] M. Couceiro, J.-L. Marichal, "Discrete integrals based on comonotonic modularity", *Axioms 2*, 3, 2013, p. 14, https://hal.archives-ouvertes.fr/hal-01090567.

[342] M. Couceiro, T. Waldhauser, "Interpolation by polynomial functions of distributive lattices : a generalization of a theorem of R. L. Goodstein", *Algebra Universalis 69*, 3, 2013, p. 13, https://hal.archives-ouvertes.fr/hal-01090572.

[343] M. Couceiro, T. Waldhauser, "Pseudo-polynomial functions over finite distributive lattices", *Fuzzy Sets and Systems 239*, 2014, p. 21–34, https://hal.archives-ouvertes.fr/hal-01090562.

[344] A. Coulet, K. B. Cohen, R. B. Altman, "Guest Editorial: The state of the art in text mining and natural language processing for pharmacogenomics", *Journal of Biomedical Informatics 45*, 5, October 2012, p. 825–826, https://hal.inria.fr/hal-00752106.

[345] A. Coulet, Y. Garten, M. Dumontier, R. B. Altman, M. A. Musen, N. H. Shah, "Integration and publication of heterogeneous text-mined relationships on the Semantic Web", *Journal of Biomedical Semantics 2*, S2, May 2011, p. S10, https://hal.archives-ouvertes.fr/hal-00585215.

[346] S. Da Silva, F. Le Ber, C. Lavigne, "Structure Analysis of Hedgerows with Respect to Perennial Landscape Lines in Two Contrasting French Agricultural Landscapes", *International Journal of Agricultural and Environmental Information Systems (IJAEIS) 5*, 1, 2014, p. 19, https://hal.archives-ouvertes.fr/hal-01057108.

[347] M. D'Aquin, J. Lieber, A. Napoli, "Decentralized case-based reasoning and Semantic Web technologies applied to decision support in oncology", *Knowledge Engineering Review 28*, 4, March 2013, p. 425–449, https://hal.inria.fr/hal-00922080.

[348] H. De Almeida, I. Bastos, B. Ribeiro, B. Maigret, J. Santana, "New Binding Site Conformations of the Dengue Virus NS3 Protease Accessed by Molecular Dynamics Simulation", *PLoS ONE 8*, 8, August 2013, p. 1–15, https://hal.inria.fr/hal-00922971.

[349] M.-D. Devignes, Y. Moreau, "ECCB 2014: The 13th European Conference on Computational Biology", *Bioinformatics 30*, 17, September 2014, p. 4, https://hal.inria.fr/hal-01097339.

[350] M.-D. Devignes, B. Sidahmed, M. Smail-Tabbone, N. Amedeo, P. Olivier, "Functional classification of genes using semantic distance and fuzzy clustering approach: Evaluation with reference sets and overlap analysis", *international Journal of Computational Biology and Drug Design. Special Issue on: "Systems Biology Approaches in Biological and Biomedical Research" 5*, 3/4, 2012, p. 245–260, https://hal.inria.fr/hal-00734329.

[351] M.-D. Devignes, M. Smaïl-Tabbone, D. Ritchie, "Kbdock - Searching and organising the structural space of protein-protein interactions", *ERCIM News*, 104, January 2016, p. 24–25, https://hal.inria.fr/hal-01258117.

[352] V. Dufour-Lussier, F. Le Ber, J. Lieber, E. Nauer, "Automatic case acquisition from texts for process-oriented case-based reasoning", *Information Systems*, December 2012, Date de publication : 2014, https://hal.inria.fr/hal-00812704.

[353] Z. Dvořák, J.-S. Sereni, J. Volec, "Fractional coloring of triangle-free planar graphs", *Electronic Journal of Combinatorics 22*, 4, 2015, p. #P4.11, https://hal.archives-ouvertes.fr/hal-00950493.

[354] E. Egho, N. Jay, C. Raïssi, D. Ienco, P. Poncelet, M. Teisseire, A. Napoli, "A contribution to the discovery of multidimensional patterns in healthcare trajectories", *Journal of Intelligent Information Systems 42*, April 2014, p. 283 – 305, https://hal.inria.fr/hal-01094377.

[355] E. Egho, C. Raïssi, T. Calders, N. Jay, A. Napoli, "On measuring similarity for sequences of itemsets", *Data Mining and Knowledge Discovery 29*, 3, May 2015, p. 33, https://hal.inria.fr/hal-01094383.

[356] C. Eng, A. Thibessard, M. Danielsen, T. B. Rasmussen, J.-F. Mari, P. Leblond, "In silico prediction of horizontal gene transfer in Streptococcus thermophilus", *Archives of Microbiology 193*, 4, January 2011, p. 287–297, https://hal.inria.fr/inria-00569081.

[357] S. J. Fleishman, T. A. Whitehead, E.-M. Strauch, J. E. Corn, S. Qin, H.-X. Zhou, J. C. Mitchell, O. N. A. Demerdash, M. Takeda-Shitaka, G. Terashi, I. H. Moal, X. Li, P. A. Bates, M. Zacharias, H. Park, J.-S. Ko, H. Lee, C. Seok, T. Bourquard, J. Bernauer, A. Poupon, J. Azé, S. Soner, S. K. Ovalı, P. Ozbek, N. B. Tal, T. Haliloglu, H. Hwang, T. Vreven, B. G. Pierce, Z. Weng, L. Pérez-Cano, C. Pons, J. Fernández-Recio, F. Jiang, F. Yang, X. Gong, L. Cao, X. Xu, B. Liu, P. Wang, C. Li, C. Wang, C. H. Robert, M. Guharoy, S. Liu, Y. Huang, L. Li, D. Guo, Y. Chen, Y. Xiao, N. London, Z. Itzhaki, O. Schueler-Furman, Y. Inbar, V. Patapov, M. Cohen, G. Schreiber, Y. Tsuchiya, E. Kanamori, D. M. Standley, H. Nakamura, K. Kinoshita, C. M. Driggers, R. G. Hall, J. L. Morgan, V. L. Hsu, J. Zhan, Y. Yang, Y. Zhou, P. L. Kastritis, A. M. J. J. Bonvin, W. Zhang, C. J. Camacho, K. P. Kilambi, A. Sircar, J. J. Gray, M. Ohue, N. Uchikoga, Y. Matsuzaki, T. Ishida, Y. Akiyama, R. Khashan, S. Bush, D. Fouches, A. Tropsha, J. Esquivel-Rodríguez, D. Kihara, P. B. Stranges, R. Jacak, B. Kuhlman, S.-Y. Huang, X. Zou, S. J. Wodak, J. Janin, D. Baker, "Community-Wide Assessment of Protein-Interface Modeling Suggests Improvements to Design Methodology.", *Journal of Molecular Biology*, September 2011, https://hal.inria.fr/inria-00637848.

[358] B. Fuchs, J. Lieber, A. Mille, A. Napoli, "Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR", *Knowledge-Based Systems 68*, 2014, p. 103 – 114, https://hal.inria.fr/hal-01101145.

[359] A. Furlan, F. Colombo, A. Kover, N. Issaly, C. Tintori, L. Angeli, V. Leroux, S. Letard, M. Amat, Y. Asses, B. Maigret, P. Dubreuil, M. Botta, R. Dono, J. Bosch, O. Piccolo, D. Passarella, F. Maina, "Identification of new aminoacid amides containing the imidazo[2,1-b]benzothiazol-2-ylphenyl moiety as inhibitors of tumorigenesis by oncogenic Met signaling", *European Journal of Medicinal Chemistry 47*, 1, 2012, p. 239 – 254, https://hal.inria.fr/hal-00763849.

[360] L. Ghemtio, V. I. Pérez-Nueno, V. Leroux, Y. Asses, M. Souchet, L. Mavridis, B. Maigret, D. Ritchie, "Recent Trends and Applications in 3D Virtual Screening", *Combinatorial Chemistry and High Throughput Screening 15*, 9, August 2012, p. 749–769, https://hal.inria.fr/hal-00756800.

[361] L. Ghemtio, M. Smaïl-Tabbone, A. Djikeng, M.-D. Devignes, L. Keminse, P. Kelbert, J. Fokam, B. Maigret, O. Ouwe-Missi-Oukem-Boyer, "HIV-PDI: A Protein-Drug Interaction Resource for Structural Analyses of HIV Drug Resistance: 1. Concepts and Associated Database", *Journal of Health & Medical Informatics 2*, 1, 2011, p. 1000104, https://hal.archives-ouvertes.fr/hal-00642539.

[362] L. Ghemtio, M. Souchet, A. Djikeng, L. Keminse, P. Kelbert, D. Ritchie, B. Maigret, O. Ouwe-Missi-Oukem-Boyer, "HIV-PDI: A Protein Drug Interaction Resource for Structural Analyses of HIV Drug Resistance: 2. Examples of Use and Proof-of-Concept", *Journal of Health & Medical Informatics 2*, 1, 2011, p. 1000105, https://hal.archives-ouvertes.fr/hal-00642567.

[363] A. Ghoorah, M.-D. Devignes, S. Z. Alborzi, M. Smaïl-Tabbone, D. Ritchie, "A Structure-Based Classification and Analysis of Protein Domain Family Binding Sites and Their Interactions", *Biology 4*, 2, April 2015, p. 327–343, https://hal.inria.fr/hal-01216748.

[364] A. Ghoorah, M.-D. Devignes, M. Smaïl-Tabbone, D. Ritchie, "Spatial clustering of protein binding sites for template based protein docking", *Bioinformatics 27*, 20, August 2011, p. 2820–2827, https://hal.inria.fr/inria-00617921.

[365] A. Ghoorah, M.-D. Devignes, M. Smaïl-Tabbone, D. Ritchie, "Protein Docking Using Case-Based Reasoning", *Proteins 81*, 12, October 2013, p. 2150–2158, https://hal.inria.fr/hal-00880341.

[366] A. Ghoorah, M.-D. Devignes, M. Smaïl-Tabbone, D. Ritchie, "KBDOCK 2013: A spatial classification of 3D protein domain family interactions", *Nucleic Acids Research 42*, D1, January 2014, p. 389–395, https://hal.inria.fr/hal-00920612.

[367] C. Goetz, A. Zang, N. Jay, "Apports d'une méthode de fouille de données pour la détection des cancers du sein incidents dans les données du programme de médicalisation des systèmes d'information", *INFORMATIQUE ET SANTE 1*, 2012, p. 189–199, https://hal.archives-ouvertes.fr/hal-00764959.

[368] S. Guessoum, M. T. Laskri, J. Lieber, "RespiDiag: a Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease", *Expert Systems with Applications 41*, 2, February 2014, p. pp. 267–273, https://hal.inria.fr/hal-00912641.

[369] D. Hervé, J.-H. Ramaroson, A. Randrianarison, F. Le Ber, "Comment les paysans du corridor forestier de Fianarantsoa (Madagascar) dessinent-ils leur territoire ? Des cartes individuelles pour confronter les points de vue", *Cybergeo : Revue européenne de géographie / European journal of geography*, 2014, p. 681, https://hal.archives-ouvertes.fr/hal-01057112.

[370] T. Hoang, X. Cavin, P. Schultz, D. Ritchie, "gEMpicker: a highly parallel GPU-accelerated particle picking tool for cryo-electron microscopy", *BMC Structural Biology 13*, 1, 2013, p. 25, https://hal.inria.fr/hal-00955580.

[371] T. V. Hoang, X. Cavin, D. Ritchie, "gEMfitter: A highly parallel FFT-based 3D density fitting tool with GPU texture memory acceleration", *Journal of Structural Biology*, September 2013, https://hal.inria.fr/hal-00866871.

[372] T. V. Hoang, S. Tabbone, "Errata and comments on "Generic orthogonal moments: Jacobi–Fourier moments for invariant image description"", *Pattern Recognition 46*, 11, April 2013, p. 3148–3155, https://hal.inria.fr/hal-00820279.

[373] N. Jay, G. Nuemi, M. Gadreau, C. Quantin, "A data mining approach for grouping and analysing trajectories of care using claim data : the exemple of breast cancer", *BCM Medical Informatics and Decisions Making*, 130, 2013, https://halshs.archives-ouvertes.fr/halshs-01227973.

[374] N. Jay, G. Nuemi, M. Gadreau, C. Quantin, "A data mining approach for grouping and analyzing trajectories of care using claim data: the example of breast cancer", *BMC Medical Informatics and Decision Making 13*, 1, November 2013, p. 130, http://www.hal.inserm.fr/inserm-00917359.

[375] C. Jonquet, P. LePendu, S. Falconer, A. Coulet, N. F. Noy, M. A. Musen, N. H. Shah, "NCBO Resource Index: Ontology-Based Search and Mining of Biomedical Resources", *Journal of Web Semantics 9*, 3, September 2011, p. 316–324, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00622155.

[376] A. S. Karaboga, F. Petronin, G. Marchetti, M. Souchet, B. Maigret, "Benchmarking of HPCC: A novel 3D molecular representation combining shape and pharmacophoric descriptors for efficient molecular similarity assessments", *Journal of Molecular Graphics and Modelling 41*, April 2013, p. 20–30, https://hal.inria.fr/hal-01101863.

[377] M. Kaytoue, S. O. Kuznetsov, J. Macko, A. Napoli, "Biclustering meets triadic concept analysis", *Annals of Mathematics and Artificial Intelligence 70*, 2014, p. 55 – 79, https://hal.inria.fr/hal-01101143.

[378] M. Kaytoue, S. O. Kuznetsov, A. Napoli, S. Duplessis, "Mining gene expression data with pattern structures in formal concept analysis", *Information Sciences 181*, 10, 2011, p. 1989–2001, https://hal.archives-ouvertes.fr/hal-00541100.

[379] D. Král', L. Mach, J.-S. Sereni, "A New Lower Bound Based on Gromov's Method of Selecting Heavily Covered Points", *Discrete and Computational Geometry 48*, 2, September 2012, p. 487–498, https://hal.archives-ouvertes.fr/hal-00912562.

[380] T. Leviandier, A. Alber, F. Le Ber, H. Piégay, "Comparison of statistical algorithms for detecting homogeneous river reaches along a longitudinal continuum", *Geomorphology 138*, 1, 2012, p. 130–144, https://hal.archives-ouvertes.fr/hal-00640698.

[381] Y. Liu, A. Coulet, P. LePendu, N. H. Shah, "Using ontology-based annotation to profile disease research", *Journal of the American Medical Informatics Association 19*, e1, June 2012, p. e177–e186, https://hal.inria.fr/hal-00752101.

[382] J.-F. Mari, E.-G. Lazrak, M. Benoît, "Time space stochastic modelling of agricultural landscapes for environmental issues", *Environmental Modelling and Software 46*, March 2013, p. 219–227, https://hal.inria.fr/hal-00807178.

[383] L. Martin, J. Wohlfahrt, F. Le Ber, M. Benoît, "L'insertion territoriale des cultures biomasses pérennes. Etude de cas sur le miscanthus en Côte d'Or (bourgogne, France)", *Espace Geographique 2*, 41, 2012, p. 138–153, https://hal.archives-ouvertes.fr/hal-00733672.

[384] S. Martin, A. Bertaux, F. Le Ber, E. Maillard, G. Imfeld, "Seasonal Changes of Macroinvertebrate Communities in a Stormwater Wetland Collecting Pesticide Runoff From a Vineyard Catchment (Alsace, France)", *Archives of Environmental Contamination and Toxicology 62*, 1, 2012, p. 29–41, https://hal.archives-ouvertes.fr/hal-00607741.

[385] L. Mavridis, A. Ghoorah, V. Venkatraman, D. Ritchie, "Representing and comparing protein folds and fold families using 3D shape-density representations", *Proteins 80*, 2, November 2011, p. 530–545, https://hal.inria.fr/hal-00641815.

[386] J.-P. Metivier, A. Lepailleur, A. Buzmakov, G. Poezevara, B. Crémilleux, S. O. Kuznetsov, J. Le Goff, A. Napoli, R. Bureau, B. Cuissart, "Discovering structural alerts for mutagenicity using stable emerging molecular patterns", *Journal of Chemical Information and Modeling 55*, 5, 2015, p. 925–940, https://hal.archives-ouvertes.fr/hal-01186716.

[387] A. Napoli, "Preface to the special issue on "Concept Lattice and their Applications" (CLA-2011)", *Annals of Mathematics and Artificial Intelligence 70*, 2014, p. 1 – 3, https://hal.inria.fr/hal-01101104.

[388] V. Pérez-Nueno, A. S. Karaboga, M. Souchet, D. Ritchie, "GES Polypharmacology Fingerprints: A Novel Approach for Drug Repositioning", *Journal of Chemical Information and Modeling 54*, 3, February 2014, p. 720–734, https://hal.inria.fr/hal-01092928.

[389] V. I. Pérez-Nueno, D. Ritchie, "Identifying and characterizing promiscuous targets: Implications for virtual screening", *Expert Opinion on Drug Discovery*, November 2011, https://hal.inria.fr/hal-00641835.

[390] V. I. Pérez-Nueno, D. Ritchie, "Using Consensus-Shape Clustering To Identify Promiscuous Ligands and Protein Targets and To Choose the Right Query for Shape-Based Virtual Screening", *Journal of Chemical Information and Modeling 51*, 6, May 2011, p. 1233–1248, https://hal.inria.fr/inria-00617922.

[391] V. I. Pérez-Nueno, V. Venkatraman, L. Mavridis, D. Ritchie, "Detecting Drug Promiscuity Using Gaussian Ensemble Screening", *Journal of Chemical Information and Modeling 52*, 8, July 2012, p. 1948–1961, https://hal.inria.fr/hal-00756804.

[392] P. Popov, D. Ritchie, S. Grudinin, "DockTrina: Docking triangular protein trimers", *Proteins: Structure, Function, and Genetics 82*, 1, January 2014, p. 34–44, https://hal.inria.fr/hal-00880359.

[393] J.-H. Ramaroson, F. Le Ber, B. Ramamonjisoa, D. Hervé, "Treillis de Galois pour la fusion de connaissances spatiales sur des territoires villageois malgaches", *Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle 2013*, 4-5, 2013, p. 595–617, https://hal.archives-ouvertes.fr/hal-00862250.

[394] D. Ritchie, A. Ghoorah, L. Mavridis, V. Venkatraman, "Fast Protein Structure Alignment using Gaussian Overlap Scoring of Backbone Peptide Fragment Similarity", *Bioinformatics 28*, 24, October 2012, p. 3274–3281, https://hal.inria.fr/hal-00756813.

[395] M. Rouane-Hacene, M. Huchard, A. Napoli, P. Valtchev, "Relational Concept Analysis: Mining Concept Lattices From Multi-Relational Data", *Annals of Mathematics and Artificial Intelligence 67*, 1, January 2013, p. 81–108, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00816300.

[396] N. Schaller, E. G. Lazrak, P. Martin, J.-F. Mari, C. Aubry, M. Benoît, "Combining farmers' decision rules and landscape stochastic regularities for landscape modelling", *Landscape Ecology 27*, 3, March 2012, p. 433–446, https://hal.inria.fr/hal-00656407.

[397] J.-S. Sereni, Z. Yilma, "A Tight Bound on the Set Chromatic Number", *Discussiones Mathematicae Graph Theory 33*, 2, 2013, p. 461–465, https://hal.archives-ouvertes.fr/hal-00624323.

[398] L. Szathmary, P. Valtchev, A. Napoli, R. Godin, A. Boc, V. Makarenkov, "A fast compound algorithm for mining generators, closed itemsets, and computing links between equivalence classes",

Annals of Mathematics and Artificial Intelligence 70, 2014, p. 81 – 105, https://hal.inria.fr/hal-01101140.

[399] P. Torres, M. Valencia-Pabon, "On the packing chromatic number of hypercubes", *Electronic Notes in Discrete Mathematics 44*, 5, November 2013, p. 263–268, https://hal.archives-ouvertes.fr/hal-00926875.

[400] W. Ugarte Rojas, P. Boizumault, B. Crémilleux, A. Lepailleur, S. Loudni, M. Plantevit, C. Raïssi, A. Soulet, "Skypattern mining: From pattern condensed representations to dynamic constraint satisfaction problems", *Artificial Intelligence*, April 2015, p. 22, https://hal.inria.fr/hal-01188928.

[401] M. Van Leeuwen, E. Galbrun, "Association Discovery in Two-View Data", *IEEE Transactions on Knowledge and Data Engineering 27*, 12, December 2015, p. 3190 – 3202, https://hal.archives-ouvertes.fr/hal-01242988.

[402] V. Venkatraman, D. Ritchie, "Flexible protein docking refinement using pose-dependent normal mode analysis", *Proteins 80*, 9, June 2012, p. 2262–2274, https://hal.inria.fr/hal-00756809.

[403] V. Venkatraman, D. Ritchie, "Predicting Multi-component Protein Assemblies Using an Ant Colony Approach", *International Journal of Swarm Intelligence Research 3*, September 2012, p. 19–31, https://hal.inria.fr/hal-00756807.

[404] Y. Xiao, C. Mignolet, M. Benoît, J.-F. Mari, " Characterizing historical (1992–2010) transitions between grassland and cropland in mainland France through mining land-cover survey data ", *Journal of Integrative Agriculture 14*, 8, August 2015, p. "1511 – 1523", https://hal.inria.fr/hal-01202510.

[405] Y. Xiao, C. Mignolet, J.-F. Mari, M. Benoît, "Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France", *Computers and Electronics in Agriculture 012*, March 2014, p. 51–63, https://hal.inria.fr/hal-00967890.

[406] A. Yasmine, V. Venkatraman, V. Leroux, D. Ritchie, B. Maigret, "Exploring c-Met kinase flexibility by sampling and clustering its conformational space", *Proteins 80*, 4, January 2012, p. 1227–1238, https://hal.inria.fr/hal-00756791.

Invited Conferences

[407] M. Couceiro, T. Waldhauser, "Lattice-Theoretic Approach to Version Spaces in Qualitative Decision Making", *in : Statistical Learning and Data Sciences, Lecture Notes in Computer Science, 9047*, p. 234–238, London, United Kingdom, April 2015, https://hal.inria.fr/hal-01175691.

[408] X. Goaoc, A. Hubard, R. De Joannis De Verclos, J.-S. Sereni, J. Volec, "Limits of order types", *in : Symposium on Computational Geometry 2015*, L. A. Janos Pach (editor), *Symposium on Computational Geometry 2015: Eindhoven, The Netherlands, 34*, p. 876, Eindhoven, Netherlands, June 2015, https://hal.inria.fr/hal-01172466.

[409] A. Napoli, "Concept Lattices for Knowledge Discovery and Knowledge Engineering", *in : Brazilian Conference on Artificial Intelligence, Natal, Brazil (BRACIS 2015)*, G. L. Pappa, K. C. Revoredo (editors), *Brazilian Conference on Artificial Intelligence, Natal, Brazil (BRACIS 2015)*, Anne Magaly de Paula Canuto (UFRN), Natal, Brazil, November 2015, https://hal.inria.fr/hal-01254131.

[410] A. Napoli, "Exploratory Knowledge Discovery with Formal Concept Analysis", *in : Advanced Information Technology, Services and Systems (AIT2S-15)*, M. Bahaj (editor), *Advanced Information Technology, Services and Systems (AIT2S-15)*, Settat, Morocco, December 2015, https://hal.inria.fr/hal-01254141.

[411] A. Napoli, "Exploring Complex and Large Data with Formal Concept Analysis", *in : Colloque sur l'Optimisation et les Systèmes d'Information (COSI 2015)*, J.-M. Petit (editor), *Colloque sur l'Optimisation et les Systèmes d'Information (COSI 2015)*, Rachid Nourine, Oran, Algeria, June 2015, https://hal.inria.fr/hal-01254129.

[412] C. Raïssi, "Multidimensional skylines", *in : Actes des 8èmes journées francophones sur les Entrepôts de Données et l'Analyse en ligne, EDA 2012*, S. Maabout (editor), Bordeaux, France, June 2012, https://hal.inria.fr/hal-00768448.

Major International Conferences

[413] P. Agarwal, M. Kaytoue, S. O. Kuznetsov, A. Napoli, G. Polaillon, "Symbolic Galois Lattices with Pattern Structures", *in : Thirteenth International Conference on Rough Sets, Fuzzy Sets and Granular Computing - RSFDGrC-2011*, S. O. Kuznetsov, D. Slezak, D. H. Hepting, B. G. Mirkin (editors), *Lecture Notes in Computer Science, 6743*, Springer-Verlag, p. 191–198, Moscou, Russia, June 2011, https://hal-supelec.archives-ouvertes.fr/hal-00631473.

[414] M. Alam, A. Buzmakov, V. Codocedo, A. Napoli, "Bridging DBpedia Categories and DL-Concept Definitions using Formal Concept Analysis", *in : Proceedings of the 4th International Workshop "What can FCA do for Artificial Intelligence?", FCA4AI 2015, co-located with the International Joint Conference on Artificial Intelligence (IJCAI 2015), Buenos Aires, Argentina.*, *CEUR Workshop Proceedings, 1430*, Buenos Aires, Argentina, July 2015, https://hal.inria.fr/hal-01186330.

[415] M. Alam, A. Buzmakov, V. Codocedo, A. Napoli, "Mining Definitions from RDF Annotations Using Formal Concept Analysis", *in : International Joint Conference in Artificial Intelligence, Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, July 2015, https://hal.archives-ouvertes.fr/hal-01186204.

[416] M. Alam, A. Buzmakov, A. Napoli, A. Sailanbayev, "Revisiting Pattern Structures for Structured Attribute Sets", *in : Proceedings of the Twelfth International Conference on Concept Lattices and Their Applications Clermont-Ferrand, France, October 13-16, 2015.*, *CEUR Workshop Proceedings, 1466*, p. 241–252, Clermont-Ferrand, France, August 2015, https://hal.inria.fr/hal-01186339.

[417] M. Alam, A. Coulet, A. Napoli, M. Smaïl-Tabbone, "Formal Concept Analysis Applied to Transcriptomic Data", *in : The Ninth International Conference on Concept Lattices and Their Applications - CLA 2012*, Fuengirola (Málaga), Spain, October 2012, https://hal.inria.fr/hal-00761003.

[418] M. Alam, A. Coulet, A. Napoli, M. Smaïl-Tabbone, "Formal Concept Analysis Applied to Transcriptomic Data", *in : What can FCA do for Artificial Intelligence (FCA4AI) (ECAI 2012)*, Montpellier, France, August 2012, https://hal.inria.fr/hal-00760993.

[419] M. Alam, A. Napoli, M. Osmuk, "RV-Xplorer: A Way to Navigate Lattice-Based Views over RDF Graphs", *in : Proceedings of the Twelfth International Conference on Concept Lattices and*

Their Applications, *1466*, Clermont-Ferrand, France, October 2015, https://hal.inria.fr/hal-01186344.

[420] M. Alam, A. Napoli, "Defining Views with Formal Concept Analysis for Understanding SPARQL Query Results", *in : Proceedings of the Eleventh International Conference on Concept Lattices and Their Applications*, Košice, Slovakia, October 2014, https://hal.inria.fr/hal-01089770.

[421] M. Alam, A. Napoli, "Lattice-Based View Access: A way to Create Views over SPARQL Query for Knowledge Discovery", *in : What can FCA do for Artificial Intelligence? (Third FCA4AI Workshop)*, Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01089777.

[422] M. Alam, A. Napoli, "Lattice-Based Views over SPARQL Query Results", *in : Proceedings of the 1st Workshop on Linked Data for Knowledge Discovery co-located with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2014)*, Nancy, France, September 2014, https://hal.inria.fr/hal-01089774.

[423] M. Alam, A. Napoli, "An Approach Towards Classifying and Navigating RDF data based on Pattern Structures", *in : Proceedings of the International Workshop on Formal Concept Analysis and Applications 2015 co-located with 13th International Conference on Formal Concept Analysis.*, *Formal Concept Analysis and Applications*, *1434*, p. 33–48, Nerja, Spain, June 2015, https://hal.inria.fr/hal-01186327.

[424] M. Alam, A. Napoli, "Interactive Exploration over RDF Data using Formal Concept Analysis", *in : Proceedings of IEEE International Conference on Data Science and Advanced Analytics*, Paris, France, August 2015, https://hal.inria.fr/hal-01186335.

[425] M. Alam, M. Wudage Chekol, A. Coulet, A. Napoli, M. Smaïl-Tabbone, "Lattice Based Data Access (LBDA): An Approach for Organizing and Accessing Linked Open Data in Biology", *in : DMoLD'13 - Data Mining on Linked Data Workshop (ECML/PKDD, 2013)*, Springer, Prague, Czech Republic, September 2013, https://hal.inria.fr/hal-00903450.

[426] Z. Assaghir, M. Kaytoue, W. Meira, J. Villerd, "Extracting Decision Trees from Interval Pattern Concept Lattices", *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, A. Napoli, V. Vychodil (editors), INRIA Nancy Grand Est - LORIA, Nancy, France, October 2011, https://hal.inria.fr/hal-00640938.

[427] Z. Assaghir, A. Napoli, M. Kaytoue, D. Dubois, H. Prade, "Numerical information fusion: Lattice of answers with supporting arguments", *in : 23rd IEEE International Conference on Tools with Artificial Intelligence - ICTAI 2011*, p. 621–628, Boca-Raton, United States, November 2011, https://hal.archives-ouvertes.fr/hal-00646484.

[428] Y. Asses, A. Buzmakov, T. Bourquard, S. O. Kuznetsov, A. Napoli, "A Hybrid Classification Approach based on FCA and Emerging Patterns - An application for the classification of biological inhibitors", *in : CLA'12: The Ninth International Conference on Concept Lattices and Their Applications - 2012*, Fuengirola, Spain, October 2012, https://hal.inria.fr/hal-00761586.

[429] J. Azé, T. Bourquard, S. Hamel, A. Poupon, D. Ritchie, "Using Kendall-Tau Meta-Bagging to Improve Protein-Protein Docking Predictions", *in : PRIB 2011*, M. Loog (editor), Marcel Reinders and Dick de Ridder, p. 284–295, DELFT, Netherlands, November 2011, https://hal.inria.fr/inria-00628038.

[430] Z. Azmeh, M. Huchard, A. Napoli, M. Rouane Hacene, P. Valtchev, "Querying Relational Concept Lattices", *in : CLA'2011: 8th International Conference on Concept Lattices and their Applications*, p. 377–392, France, October 2011, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00646409.

[431] J. Baixeries, M. Kaytoue, A. Napoli, "Computing Functional Dependencies with Pattern Structures", *in : The 9th International Conference on Concept Lattices and Their Applications - CLA 2012*, U. Priss, L. Szathmary (editors), Malaga, Spain, October 2012, https://hal.inria.fr/hal-00763748.

[432] J. Baixeries, M. Kaytoue, A. Napoli, "Computing Similarity Dependencies with Pattern Structures", *in : The Tenth International Conference on Concept Lattices and their Applications - CLA 2013*,, M. Ojeda-Aciego, J. Outrata (editors), Karell Bertet, p. 33–44, La Rochelle, France, France, 2013, https://hal.inria.fr/hal-00922592.

[433] J. Baixeries, M. Kaytoue, A. Napoli, "Characterization of Database Dependencies with FCA and Pattern Structures", *in : Third International Conference, Analysis of Images, Social Networks and Texts (AIST 2014), Communications in Computer and Information Science, 436*, Springer, p. 3 – 14, Yekaterinburg, Russia, 2014, https://hal.inria.fr/hal-01101147.

[434] M. Barrère, G. Betarte, V. Codocedo, M. Rodríguez, H. Astudillo, M. Aliquintuy, J. Baliosian, R. Badonnel, O. Festor, C. Raniery Paula dos Santos, J. Campos Nobre, L. Z. Granville, A. Napoli, "Machine-assisted Cyber Threat Analysis using Conceptual Knowledge Discovery", *in : FCA4AI 2015 - Workshop What can FCA do for Artificial Intelligence ?*, p. 75 – 85, Buenos Aires, Argentina, July 2015, https://hal.archives-ouvertes.fr/hal-01186213.

[435] S. Basu Roy, I. Lykourentzou, S. Thirumuruganathan, S. Amer-Yahia, G. Das, "Crowds, not Drones: Modeling Human Factors in Interactive Crowdsourcing", *in : DBCrowd 2013 - VLDB Workshop on Databases and Crowdsourcing*, R. Cheng, A. D. Sarma, S. Maniu, P. Senellart (editors), *CEUR Workshop Proceedings*, CEUR-WS, p. 39–42, Riva del Garda, Trento, Italy, August 2013, https://hal.inria.fr/hal-00923542.

[436] S. Ben Abbès, A. Scheuermann, T. Meilender, M. D'Aquin, "Characterizing Modular Ontologies", *in : 7th International Conference on Formal Ontologies in Information Systems - FOIS 2012*, p. 13–25, Graz, Austria, July 2012, https://hal.archives-ouvertes.fr/hal-00710035.

[437] S. Benabderrahmane, M.-D. Devignes, M. Smaïl-Tabbone, A. Napoli, O. Poch, "Ontology-based functional classification of genes: evaluation with reference sets and overlap analysis", *in : The second workshop on Integrative Data in Systems Biology held in conjunction with the IEEE BIBM 2011 conference*, Z. Zhao (editor), IEEE Computer Society, p. –, Atlanta, United States, November 2011, https://hal.archives-ouvertes.fr/hal-00644438.

[438] S. Benabderrahmane, M.-D. Devignes, M. Smail-Tabbone, O. Poch, A. Napoli, W. Raffelsberger, D. Guenot, N. Hoan, E. Guerin, "Benchmarking a new semantic similarity measure using fuzzy clustering and reference sets: Application to cancer expression data", *in : 11ème Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances - EGC 2011*, Brest, France, January 2011, https://hal.inria.fr/inria-00617692.

[439] A. Berry, M. Huchard, A. Napoli, A. Sigayret, "Hermes: an efficient algorithm for building Galois sub-hierarchies", *in : CLA'2012: 9th International Conference on Concept Lattices and Applications*, U. P. Laszlo Szathmary (editor), Universidad de Malaga, p. 21–32, Fuengirola (Málaga), Spain, October 2012, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00743882.

[440] G. Bosc, M. Kaytoue, C. Raïssi, J.-F. Boulicaut, P. Tan, "Mining Balanced Sequential Patterns in RTS Games 1", *in : ECAI 2014 - 21st European Conference on Artificial Intelligence*, Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01100933.

[441] A. Braud, C. Nica, C. Grac, F. Le Ber, "A lattice-based query system for assessing the quality of hydro-ecosystems", *in : CLA 2011*, V. V. A Napoli (editor), INRIA NGE et LORIA, p. 265–277, Nancy, France, October 2011, https://hal.archives-ouvertes.fr/hal-00640048.

[442] E. Bresso, S. Benabderrahmane, M. Smaïl-Tabbone, G. Marchetti, A. S. Karaboga, M. Souchet, A. Napoli, M.-D. Devignes, "Use of domain knowledge for dimension reduction: application to mining of drug side effects", *in : International Conference on Knowledge Discovery and Information Retrieval - KDIR 2011*, A. Fred (editor), INSTICC, SciTePress Digital Library, p. 8, Paris, France, October 2011, https://hal.archives-ouvertes.fr/hal-00642520.

[443] E. Bresso, R. Grisoni, M.-D. Devignes, A. Napoli, M. Smail-Tabbone, "Formal Concept Analysis for the Interpretation of Relational Learning applied on 3D Protein-Binding Sites", *in : 4th international conference on Knowledge Discovery and Information Retrieval - KDIR 2012*, A. Fred (editor), p. 12 pages, Barcelona, Spain, October 2012, https://hal.inria.fr/hal-00734349.

[444] O. Bruneau, S. Garlatti, M. Guedj, S. Laubé, J. Lieber, "SemanticHPST: Applying Semantic Web Principles and Technologies to the History and Philosophy of Science and Technology", *in : The Semantic Web: ESWC 2015 Satellite Events*, F. Gandon, A. Zimmermann, C. Faron-Zucker, J. Breslin, S. Villata, C. Guéret (editors), *Lecture Notes in Computer Science, 9341*, Springer International Publishing, p. 416–427, Portoroz, Slovenia, May 2015, https://hal.archives-ouvertes.fr/hal-01214698.

[445] A. Buzmakov, E. Egho, N. Jay, S. O. Kuznetsov, A. Napoli, C. Raïssi, "FCA and pattern structures for mining care trajectories", *in : Workshop FCA4AI, "What FCA can do for artificial intelligence?"*, Beijing, China, August 2013, https://hal.inria.fr/hal-00910290.

[446] A. Buzmakov, E. Egho, N. Jay, S. O. Kuznetsov, A. Napoli, C. Raïssi, "On Projections of Sequential Pattern Structures (with an application on care trajectories)", *in : The Tenth International Conference on Concept Lattices and Their Applications - CLA'13*, La Rochelle, France, October 2013, https://hal.inria.fr/hal-00910300.

[447] A. Buzmakov, E. Egho, N. Jay, S. O. Kuznetsov, A. Napoli, C. Raïssi, "The representation of sequential patterns and their projections within Formal Concept Analysis", *in : Workshop Notes for LML (PKDD)*, Prague, Czech Republic, September 2013, https://hal.inria.fr/hal-00910266.

[448] A. Buzmakov, S. O. Kuznetsov, A. Napoli, "A New Approach to Classification by Means of Jumping Emerging Patterns", *in : FCA4AI: International Workshop "What can FCA do for Artificial Intelligence?" - 2012*, Montpellier, France, December 2012, https://hal.inria.fr/hal-00761602.

[449] A. Buzmakov, S. O. Kuznetsov, A. Napoli, "Concept Stability as a Tool for Pattern Selection", *in : FCA4AI 2014. What can FCA do for Artificial Intelligence?*, Praque, Czech Republic, August 2014, https://hal.inria.fr/hal-01095903.

[450] A. Buzmakov, S. O. Kuznetsov, A. Napoli, "On Evaluating Interestingness Measures for Closed Itemsets", *in : 7th European Starting AI Researcher Symposium (STAIRS 2014)*, J. L. Ulle Endriss (editor), *Proceedings of the 7th European Starting AI Researcher Symposium (STAIRS 2014), 264*, p. 71 – 80, Prague, Czech Republic, 2014, https://hal.inria.fr/hal-01095927.

[451] A. Buzmakov, S. O. Kuznetsov, A. Napoli, “Scalable Estimates of Concept Stability”, *in : 12th International Conference on Formal Concept Analysis (ICFCA 2014)*, C. S. Cynthia Vera Glodeanu, Mehdi Kaytoue (editor), *Formal Concept Analysis*, *8478*, Springer, p. 157 – 172, Cluj-Napoca, Romania, 2014, https://hal.inria.fr/hal-01095920.

[452] A. Buzmakov, S. O. Kuznetsov, A. Napoli, “Fast Generation of Best Interval Patterns for Nonmonotonic Constraints”, *in : Machine Learning and Knowledge Discovery in Databases*, *Lecture Notes in Computer Science*, *9285*, p. 157–172, Porto, Portugal, September 2015, https://hal.archives-ouvertes.fr/hal-01186718.

[453] A. Buzmakov, S. O. Kuznetsov, A. Napoli, “Revisiting Pattern Structure Projections”, *in : International Conference in Formal Concept Analysis - ICFCA 2015*, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), *Lecture Notes in Computer Science*, *9113*, Springer International Publishing, p. 200–215, Nerja, Spain, June 2015, https://hal.archives-ouvertes.fr/hal-01186719.

[454] A. Buzmakov, A. Neznanov, “Practical Computing with Pattern Structures in FCART Environment”, *in : Workshop FCA4AI, ”What FCA can do for artificial intelligence?”*, Beijing, China, August 2013, https://hal.inria.fr/hal-00910296.

[455] G. Canals, A. Cordier, E. Desmontils, L. Infante-Blanco, E. Nauer, “Collaborative Knowledge Acquisition under Control of a Non-Regression Test System”, *in : Workshop on Semantic Web Collaborative Spaces*, Montpellier, France, May 2013. 14 pages, https://hal.archives-ouvertes.fr/hal-00880347.

[456] C. Chamard, E. Bresso, T. Boukhobza, M.-D. Devignes, M. Smaïl-Tabbone, A. Chesnel, H. Dumond, “Long chain alkylphenol mixture promotes mammary epithelial cell metaplastic phenotype through an estrogen receptor alpha 36 mediated mechanism”, *in : 23rd Biennial Congress of the European Association for Cancer Research, EACR-23, European Journal of cancer*, *50*, Supplement 5, p. S118, Munich, Germany, July 2014. Présentation Poster, https://hal.archives-ouvertes.fr/hal-01095429.

[457] M. W. Chekol, M. Alam, A. Napoli, “A Study on the Correspondence between FCA and ELI Ontologies”, *in : CLA - The Tenth International Conference on Concept Lattices and Their Applications - 2013*, La Rochelle, France, October 2013, https://hal.inria.fr/hal-00906757.

[458] M. W. Chekol, J. Euzenat, P. Genevès, N. Layaïda, “Evaluating and benchmarking SPARQL query containment solvers”, *in : Proc. 12th International semantic web conference (ISWC)*, H. Alani, L. Kagal, A. Fokoue, P. Groth, C. Biemann, J. X. Parreira, L. Aroyo, N. Noy, C. Welty, K. Janowicz (editors), *8219*, Springer Verlag, p. 408–423, Sydney, Australia, October 2013. wudagechekol2013a, https://hal.inria.fr/hal-00917911.

[459] M. W. Chekol, A. Napoli, “A Study on the Correspondence between FCA and ELI Ontologies”, *in : ISWC - The 12th International Semantic Web Conference - 2013*, Sydney, Australia, October 2013, https://hal.inria.fr/hal-00906736.

[460] M. W. Chekol, A. Napoli, “An FCA Framework for Knowledge Discovery in SPARQL Query Answers”, *in : The 12th International Semantic Web Conference*, Sydney, Australia, October 2013, https://hal.inria.fr/hal-00881080.

[461] V. Codocedo, I. Lykourentzou, H. Astudillo, A. Napoli, “Using pattern structures to support information retrieval with Formal Concept Analysis”, *in : International Workshop ”What can FCA do for Artificial Intelligence?”*, S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), p. 15–24, Beijing, China, August 2013, https://hal.inria.fr/hal-00880020.

[462] V. Codocedo, I. Lykourentzou, A. Napoli, “A contribution to semantic indexing and retrieval based on FCA - An application to song datasets”, *in : Proceedings of the conference on concept lattices and their applications (CLA)*, Malaga, Spain, October 2012, https://hal.archives-ouvertes.fr/hal-00760764.

[463] V. Codocedo, I. Lykourentzou, A. Napoli, “Semantic querying of data guided by Formal Concept Analysis”, *in : Formal Concept Analysis for Artificial Intelligence*, Montpellier, France, December 2012, https://hal.archives-ouvertes.fr/hal-00760757.

[464] V. Codocedo, A. Napoli, “A Proposition for Combining Pattern Structures and Relational Concept Analysis”, *in : Formal Concept Analysis - 12th International Conference, ICFCA 2014, Cluj-Napoca, Romania, June 10-13, 2014. Proceedings*, p. 96 – 111, Cluj-Napoca, Romania, June 2014, https://hal.inria.fr/hal-01095870.

[465] V. Codocedo, A. Napoli, “Bicluster enumeration using Formal Concept Analysis”, *in : What formal concept analysis can do for artificial intelligence? (FCA4AI 2014) Workshop at ECAI 2014*, Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01095884.

[466] V. Codocedo, A. Napoli, “Lattice-based biclustering using Partition Pattern Structures”, *in : ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS) 2014*, Prague, Czech Republic, August 2014, https://hal.inria.fr/hal-01095865.

[467] V. Codocedo, A. Napoli, “Formal Concept Analysis and Information Retrieval – A Survey”, *in : International Conference in Formal Concept Analysis - ICFCA 2015*, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), *Formal Concept Analysis - Lecture Notes in Computer Science, 9113*, Springer, p. 61–77, Nerja, Spain, June 2015, https://hal.archives-ouvertes.fr/hal-01186196.

[468] V. Codocedo, C. Taramasco, H. Astudillo, “Cheating to achieve Formal Concept Analysis over a large formal context”, *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, p. 349–362, Nancy, France, October 2011, https://hal.archives-ouvertes.fr/hal-00654576.

[469] J. Cojan, V. Dufour-Lussier, E. Gaillard, J. Lieber, E. Nauer, Y. Toussaint, “Knowledge extraction for improving case retrieval and recipe adaptation”, *in : Computer Cooking Contest Workshop*, London, United Kingdom, September 2011, https://hal.inria.fr/hal-00646717.

[470] J. Cojan, J. Lieber, “An Algorithm for Adapting Cases Represented in ALC”, *in : 22nd International Joint Conference on Artificial Intelligence - IJCAI 2011*, Barcelone, Spain, July 2011, https://hal.inria.fr/inria-00584103.

[471] J. Cojan, J. Lieber, “Belief revision-based case-based reasoning”, *in : ECAI-2012 Workshop SAMAI*, G. Richard (editor), *Proceedings of the ECAI-2012 Workshop SAMAI: Similarity and Analogy-based Methods in AI*, Gilles Richard, p. 33–39, Montpellier, France, August 2012, https://hal.inria.fr/hal-00763220.

[472] A. Cordier, E. Gaillard, E. Nauer, “Man-Machine Collaboration to Acquire Cooking Adaptation Knowledge for the TAAABLE Case-Based Reasoning System”, *in : SWCS Semantic Web Collaborative Spaces*, ACM, p. 1113–1120, Lyon, France, April 2012, https://hal.inria.fr/hal-00696013.

[473] A. Cordier, J. Lieber, J. Stevenot, "Towards an operator for merging taxonomies", *in : ECAI-2012 Workshop BNC: Belief change, Non-monotonic reasoning and Conflict resolution*, S. Konieczny, T. Meyer (editors), *Proceedings of the ECAI-2012 Workshop BNC: Belief change, Non-monotonic reasoning and Conflict resolution*, S. Konieczny and T. Meyer, Montpellier, France, August 2012, https://hal.inria.fr/hal-00763228.

[474] M. Couceiro, Q. Brabant, "Axiomatisation des intégrales de Sugeno k-maxitives", *in : 24ème Conférence Rencontres francophones sur la Logique Floue et ses Applications (LFA 2015)*, Poitiers, France, November 2015, https://hal.inria.fr/hal-01247510.

[475] M. Couceiro, D. Dubois, H. Prade, A. Rico, T. Waldhauser, "General Interpolation by Polynomial Functions of Distributive Lattices", *in : 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012, Communications in Computer and Information Science*, Catania, Italy, July 2012, https://hal.archives-ouvertes.fr/hal-01093655.

[476] M. Couceiro, D. Dubois, H. Prade, T. Waldhauser, "Decision-making with Sugeno integrals: DMU vs. MCDM", *in : ECAI 2012 - 20th European Conference on Artificial Intelligence., Frontiers in Artificial Intelligence and Applications*, Montpellier, France, August 2012, https://hal.archives-ouvertes.fr/hal-01093646.

[477] M. Couceiro, L. Haddad, M. Pouzet, K. Schölzel, "Hereditary Rigid Relations.", *in : 45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015), Proceedings of 45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015), IEEE Computer Society.*, Waterloo, Canada, May 2015, https://hal.inria.fr/hal-01175699.

[478] M. Couceiro, L. Haddad, K. Schölzel, T. Waldhauser, "A Solution to a Problem of D. Lau: Complete Classification of Intervals in the Lattice of Partial Boolean Clones", *in : IEEE 43rd International Symposium on Multiple-Valued Logic (ISMVL), 2013*, Toyama, Japan, May 2013, https://hal.archives-ouvertes.fr/hal-01090640.

[479] M. Couceiro, L. Haddad, K. Schölzel, T. Waldhauser, "Relation Graphs and Partial Clones on a 2-Element Set", *in : IEEE 44th International Symposium on Multiple-Valued Logic (ISMVL), 2014*, IEEE (editor), IEEE Computer Society, p. 161–166, Bremen, Germany, May 2014, https://hal.archives-ouvertes.fr/hal-01090638.

[480] M. Couceiro, L. Haddad, "A Survey on Intersections of Maximal Partial Clones of Boolean Partial Functions", *in : 42nd IEEE International Symposium on Multiple-Valued Logic, ISMVL 2012*, Victoria, BC, Canada, May 2012, https://hal.archives-ouvertes.fr/hal-01093663.

[481] M. Couceiro, E. Lehtonen, K. Schölzel, "Sur des classes de fonctions à seuil caractérisables par des contraintes relationnelles", *in : Rencontres francophones sur la Logique Floue et ses Applications (LFA 2013)*, Reims, France, October 2013, https://hal.archives-ouvertes.fr/hal-01093637.

[482] M. Couceiro, E. Lehtonen, T. Waldhauser, "GAP vs. PAG", *in : 42nd IEEE International Symposium on Multiple-Valued Logic, ISMVL 2012*, Victoria, BC, Canada, May 2012, https://hal.archives-ouvertes.fr/hal-01093660.

[483] M. Couceiro, J.-L. Marichal, B. Teheux, "Median preserving aggregation functions", *in : 8th International Summer School on Aggregation Operators and their Applications (AGOP 2015), Proc. 8th Int. Summer School on Aggregation Operators and their Applications (AGOP 2015)*, Michal Baczyński, Bernard De Baets, Radko Mesiar, p. 85–89, Katowice, Poland, July 2015, https://hal.archives-ouvertes.fr/hal-01184116.

[484] M. Couceiro, J.-L. Marichal, T. Waldhauser, "Hierarchies of Local Monotonicities and Lattice Derivatives for Boolean and Pseudo-Boolean Functions", *in : 42nd IEEE International Symposium on Multiple-Valued Logic, ISMVL 2012*, Victoria, BC, Canada, May 2012, https://hal.archives-ouvertes.fr/hal-01093657.

[485] M. Couceiro, J.-L. Marichal, "Quasi-extensions de Lovász et leur version symétrique", *in : Rencontres francophones sur la Logique Floue et ses Applications (LFA 2012)*, Compiègne, France, November 2012, https://hal.archives-ouvertes.fr/hal-01093642.

[486] M. Couceiro, J.-L. Marichal, "Quasi-Lovász Extensions and Their Symmetric Counterparts", *in : 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2012, Communications in Computer and Information Science*, Catania, Italy, July 2012, https://hal.archives-ouvertes.fr/hal-01093650.

[487] M. Couceiro, J.-L. Marichal, "Quasi-Lovász extensions on bounded chains", *in : 15th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2014, Communications in Computer and Information Science, 442*, Springer, Montpellier, France, July 2014, https://hal.archives-ouvertes.fr/hal-01090634.

[488] M. Couceiro, B. Teheux, "Clones of pivotally decomposable functions.", *in : 45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015), 45th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2015), IEEE Computer Society.*, Waterloo, Canada, May 2015, https://hal.inria.fr/hal-01175698.

[489] M. Couceiro, T. Waldhauser, J. Almeida, "On the Semigroup of Equational Classes of Finite Functions", *in : IEEE 43rd International Symposium on Multiple-Valued Logic (ISMVL), 2013*, Toyama, Japan, May 2013, https://hal.archives-ouvertes.fr/hal-01090642.

[490] A. Coulet, F. Domenach, M. Kaytoue, A. Napoli, "Using Pattern Structures for Analyzing Ontology-Based Annotations of Biomedical Data", *in : International Conference on Formal Concept Analysis, LNCS/LNAI series*, Springer, Dresden, Germany, May 2013, https://hal.inria.fr/hal-00880643.

[491] S. Da Silva, C. Lavigne, F. Le Ber, "Analyse des structures des haies et des linéaires pérennes dans deux paysages agricoles contrastés", *in : SAGEO - International Conference on Spatial Analysis and GEOmatics*, p. 16, Liège, Belgium, 2012, https://hal.archives-ouvertes.fr/hal-00742681.

[492] S. Da Silva, F. Le Ber, C. Lavigne, "Structures de haies dans un paysage agricole : une étude par chemin de Hilbert adaptatif et chaînes de Markov ", *in : EGC 2016 – 16èmes Journées Francophones "Extraction et Gestion des Connaissances", Revue des Nouvelles Technologies de l'Information, RNTI-E-30*, p. 279–290, Reims, France, January 2016, https://hal.archives-ouvertes.fr/hal-01266344.

[493] K. Dalleau, N. Coumba Ndiaye, A. Coulet, "Suggesting valid pharmacogenes by mining linked data", *in : Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015, Proceedings of the Semantic Web Applications and Tools for Life Sciences (SWAT4LS) 2015*, Cambridge, United Kingdom, December 2015, https://hal.inria.fr/hal-01239568.

[494] M. D'Aquin, N. Jay, "Interpreting data mining results with linked data for learning analytics: motivation, case study and directions", *in : LAK - Third International Conference on Learning Analytics and Knowledge - 2013*, ACM, p. 155–164, Leuven, Belgium, April 2013, https://hal.inria.fr/hal-00903259.

[495] V. Dufour-Lussier, B. Guillaume, G. Perrier, "Parsing Coordination Extragrammatically", *in : 5th Language & Technology Conference - LTC'11*, Poznan, Poland, November 2011, https://hal.inria.fr/hal-00639929.

[496] V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber, "Belief revision in the propositional closure of a qualitative algebra ", *in : 14th International Conference on Principles of Knowledge Representation and Reasoning*, AAAI Press, p. 4, Vienne, Austria, July 2014, https://hal.inria.fr/hal-01094264.

[497] V. Dufour-Lussier, F. Le Ber, J. Lieber, L. Martin, "Adapting Spatial and Temporal Cases", *in : International Conference for Case-Based Reasoning*, B. D. A. Ian Watson (editor), *7466*, Amélie Cordier, Marie Lefevre, Springer, p. 77–91, Lyon, France, September 2012, https://hal.inria.fr/hal-00735231.

[498] V. Dufour-Lussier, F. Le Ber, J. Lieber, L. Martin, "Case Adaptation with Qualitative Algebras", *in : International Joint Conferences on Artificial Intelligence*, F. Rossi (editor), *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, AAAI Press, p. 3002–3006, Pékin, China, August 2013, https://hal.inria.fr/hal-00871703.

[499] V. Dufour-Lussier, F. Le Ber, J. Lieber, T. Meilender, E. Nauer, "Semi-automatic annotation process for procedural texts: An application on cooking recipes", *in : Cooking with Computers workshop - ECAI 2012*, E. N. Amélie Cordier (editor), Montpellier, France, August 2012, https://hal.inria.fr/hal-00735262.

[500] V. Dufour-Lussier, J. Lieber, E. Nauer, Y. Toussaint, "Improving case retrieval by enrichment of the domain ontology", *in : 19th International Conference on Case Based Reasoning - ICCBR'2011*, London, United Kingdom, September 2011, https://hal.inria.fr/inria-00617621.

[501] V. Dufour-Lussier, J. Lieber, "Evaluating a textual adaptation system", *in : International Conference on Case-Based Reasoning*, Frankfurt, Germany, September 2015, https://hal.inria.fr/hal-01178331.

[502] E. Egho, D. Ienco, N. Jay, A. Napoli, P. Poncelet, C. Quantin, C. Raïssi, M. Teisseire, "Healtcare Trajectory Mining by Combining Multi-dimensional Component and Itemsets", *in : NFMCP'2012: New Frontiers in Mining Complex Patterns, Workshop in conjunction with European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2012), LNCS*, Springer, p. 12, United Kingdom, September 2012, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00732661.

[503] E. Egho, N. Jay, C. Raïssi, A. Napoli, "A FCA-based analysis of sequential care trajectories", *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, A. Napoli, V. Vychodil (editors), INRIA Nancy Grand Est - LORIA, Nancy, France, October 2011, https://hal.inria.fr/hal-00641649.

[504] E. Egho, N. Jay, C. Raïssi, G. Nuemi, C. Quantin, A. Napoli, "An approach for mining care trajectories for chronic diseases", *in : 14th Conference on Artificial Intelligence in Medicine, 7885*, Springer, Murcia, Spain, May 2013, https://hal.inria.fr/hal-00883117.

[505] E. Egho, C. Raïssi, D. Ienco, N. Jay, A. Napoli, P. Poncelet, C. Quantin, M. Teisseire, "Healthcare trajectory mining by combining multidimensional component and itemsets", *in : ECML-PKDD 2012*, p. p. 116 – p. 127, Bristol, United Kingdom, September 2012, https://hal.archives-ouvertes.fr/hal-00801813.

[506] E. Egho, C. Raïssi, N. Jay, A. Napoli, "Mining Heterogeneous Multidimensional Sequential Patterns", *in : European Conference on Artificial Intelligence*, *263*, p. 6, Prague, Czech Republic, France, August 2014, https://hal.inria.fr/hal-01094365.

[507] M. Fabrègue, A. Braud, S. Bringay, F. Le Ber, M. Teisseire, "Including spatial relations and scales within sequential pattern extraction", *in : DS'2012: 15th International Conference on Discovery Science*, *LNCS/LNAI*, 7569, p. N/A, Lyon, France, October 2012, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00735617.

[508] E. Gaillard, L. Infante-Blanco, J. Lieber, E. Nauer, "Tuuurbine: A Generic CBR Engine over RDFS", *in : Case-Based Reasoning Research and Development*, *8765*, p. 140 – 154, Cork, Ireland, September 2014, https://hal.inria.fr/hal-01082372.

[509] E. Gaillard, J. Lieber, Y. Naudet, E. Nauer, "Case-Based Reasoning on E-Community Knowledge", *in : 21st International Conference on Case-Based Reasoning*, *7969*, Springer, p. 104–118, Saratoga Springs, NY, United States, 2013, https://hal.inria.fr/hal-00918518.

[510] E. Gaillard, J. Lieber, E. Nauer, A. Cordier, "How Case-Based Reasoning on e-Community Knowledge Can Be Improved Thanks to Knowledge Reliability", *in : Case-Based Reasoning Research and Development*, *8765*, Luc Lamontagne and Enric Plaza, p. 155 – 169, Cork, Ireland, Ireland, September 2014, https://hal.inria.fr/hal-01082369.

[511] E. Gaillard, J. Lieber, E. Nauer, "Adaptation knowledge discovery for cooking using closed itemset extraction", *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, Nancy, France, October 2011, https://hal.inria.fr/hal-00646732.

[512] E. Gaillard, J. Lieber, E. Nauer, "Case-Based Cooking with Generic Computer Utensils: Taaable Next Generation", *in : Proceedings of the ICCBR 2014 Workshops*, pp 89-100, David B. Leake and Jean Lieber, p. 254, Cork, Ireland, September 2014, https://hal.archives-ouvertes.fr/hal-01082683.

[513] E. Gaillard, J. Lieber, E. Nauer, "How Managing the Knowledge Reliability Improves the Results of a Reasoning Process", *in : 16th European Conference on Knowledge Management - ECKM 2015*, p. 10, Udine, Italy, September 2015, https://hal.inria.fr/hal-01178315.

[514] E. Gaillard, J. Lieber, E. Nauer, "Improving Case Retrieval Using Typicality", *in : 23rd International Conference on Case-Based Reasoning (ICCBR 2015)*, Frankfurt am Main, Germany, September 2015, https://hal.inria.fr/hal-01178317.

[515] E. Gaillard, J. Lieber, E. Nauer, "Improving Ingredient Substitution using Formal Concept Analysis and Adaptation of Ingredient Quantities with Mixed Linear Optimization", *in : Computer Cooking Contest Workshop*, Frankfort, Germany, September 2015, https://hal.inria.fr/hal-01240383.

[516] E. Gaillard, E. Nauer, M. Lefevre, A. Cordier, "Extracting Generic Cooking Adaptation Knowledge for the TAAABLE Case-Based Reasoning System", *in : Cooking with Computers workshop @ ECAI 2012*, Montpellier, France, August 2012, https://hal.inria.fr/hal-00720481.

[517] M. Hassan, A. Coulet, Y. Toussaint, "Learning Subgraph Patterns from text for Extracting Disease–Symptom Relationships", *in : 1st International Workshop on Interactions between Data Mining and Natural Language Processing*, P. Cellier, T. Charnois, A. Hotho, S. Matwin, M.-F. Moens, Y. Toussaint (editors), *1202*, ceur-ws, Nancy, France, September 2014, https://hal.inria.fr/hal-01095595.

[518] M. Hassan, O. Makkaoui, A. Coulet, Y. Toussaint, "Extracting Disease-Symptom Relationships by Learning Syntactic Patterns from Dependency Graphs", *in : BioNLP 15, Proceedings of BioNLP 15*, Association for Computational Linguistics, p. 184, Beijing, China, July 2015, https://hal.inria.fr/hal-01184655.

[519] T. V. Hoang, S. Tabbone, "Fast Computation of Orthogonal Polar Harmonic Transforms", *in : ICPR 2012 - The 21st International Conference on Pattern Recognition*, Tsukuba Science City, Japan, November 2012, https://hal.inria.fr/hal-00734307.

[520] P. Holat, M. Plantevit, C. Raïssi, N. Tomeh, T. Charnois, B. Crémilleux, "Sequence Classification Based on Delta-Free Sequential Pattern", *in : IEEE International Conference on Data Mining*, Shenzhen, China, December 2014, https://hal.inria.fr/hal-01100929.

[521] N. Jay, M. D'Aquin, "Linked Data and Online Classifications to Organise Mined Patterns in Patient Data", *in : Proceedings of the AMIA 2013 Annual Symposium*, Washington, United States, November 2013, https://hal.inria.fr/hal-00903261.

[522] M. Kaytoue, V. Codocedo, J. Baixeries, A. Napoli, "Three Related FCA Methods for Mining Biclusters of Similar Values on Columns", *in : Proceedings of the Eleventh International Conference on Concept Lattices and Their Applications, Kosice, Slovakia, October 7-10, 2014*, Kosice, Slovakia, October 2014, https://hal.inria.fr/hal-01095877.

[523] M. Kaytoue, V. Codocedo, A. Buzmakov, J. Baixeries, S. O. Kuznetsov, A. Napoli, "Pattern Structures and Concept Lattices for Data Mining and Knowledge Processing", *in : Machine Learning and Knowledge Discovery in Databases*, A. Bifet, M. May, B. Zadrozny, R. Gavalda, D. Pedreschi, F. Bonchi, J. Cardoso, M. Spiliopoulou (editors), *Lecture Notes in Computer Science*, *9286*, Springer International Publishing, p. 227–231, Porto, Portugal, 2015, https://hal.archives-ouvertes.fr/hal-01188637.

[524] M. Kaytoue, S. O. Kuznetsov, J. Macko, W. Meira, A. Napoli, "Mining Biclusters of Similar Values with Triadic Concept Analysis", *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, A. Napoli, V. Vychodil (editors), INRIA Nancy Grand Est - LORIA, Nancy, France, October 2011, https://hal.inria.fr/hal-00640873.

[525] M. Kaytoue, S. O. Kuznetsov, A. Napoli, "Biclustering Numerical Data in Formal Concept Analysis", *in : 9th International Conference on Formal Concept Analysis - ICFCA 2011*, P. Valtchev, R. Jäschke (editors), *6628*, Springer, p. 135–150, Nicosia, Cyprus, May 2011, https://hal.inria.fr/inria-00600203.

[526] M. Kaytoue, S. O. Kuznetsov, A. Napoli, "Revisiting Numerical Pattern Mining with Formal Concept Analysis", *in : Twenty second International Joint Conference on Artificial Intelligence - IJCAI 2011*, Barcelona, Spain, July 2011, https://hal.inria.fr/inria-00584371.

[527] M. Kaytoue, A. Silva, L. Cerf, W. Meira, C. Raïssi, "Watch me playing, I am a professional: a first study on video game live streaming", *in : MSND@WWW - International Workshop on Mining Social Network Dynamics - 2012 (in conjunction with WWW - World Wild Web - 2012)*, H. Hacid, S. Guo, J. Velcin (editors), *WWW '12 Companion - Proceedings of the 21st international conference companion on World Wide Web*, ACM, p. 1181–1188, Lyon, France, May 2012, https://hal.inria.fr/hal-00697150.

[528] B. Lamiroy, D. Lopresti, H. Korth, J. Heflin, "How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research", *in : Document Recognition and Retrieval*

XVIII - DRR 2011, C. V.-G. Gady Agam (editor), *Document Recognition and Retrieval XVIII*, *Part of the IST/SPIE 23rd Annual Symposium on Electronic Imaging*, *7874*, SPIE, SPIE, San Francisco, United States, January 2011. ISBN : 9780819484116, https://hal.inria.fr/inria-00537035.

[529] F. Le Ber, C. Lavigne, S. Da Silva, "Structure analysis of hedgerows and other perennial landscape lines in two French agricultural landscapes", *in : AGILE 2012*, p. 6, Avignon, France, April 2012, https://hal.archives-ouvertes.fr/hal-00685777.

[530] A. Leeuwenberg, A. Buzmakov, Y. Toussaint, A. Napoli, "Exploring Pattern Structures of Syntactic Trees for Relation Extraction", *in : International Conference in Formal Concept Analysis - ICFCA 2015*, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), *Lecture Notes in Computer Science*, *9113*, Springer, p. 153–168, Nerja, Spain, June 2015, https://hal.archives-ouvertes.fr/hal-01186717.

[531] D. Lopresti, B. Lamiroy, "Document Analysis Research in the Year 2021", *in : Twenty-fourth International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems - IEA/AIE 2011*, K. G. Mehrotra, C. K. Mohan, J. C. Oh, P. K. Varshney, M. Ali (editors), *6703*, Syracuse University, Springer, p. 264–274, Syracuse, NY, United States, June 2011. The original publication is available at www.springerlink.com, https://hal.inria.fr/inria-00570000.

[532] C. Low-Kam, C. Raïssi, M. Kaytoue, J. Pei, "Mining Statistically Significant Sequential Patterns", *in : IEEE International Conference on Data Mining*, Dallas, United States, December 2013, https://hal.inria.fr/hal-00922255.

[533] L. Martin, F. Le Ber, J. Wohlfahrt, G. Bocquého, M. Benoît, "Modelling farmers' choice of miscanthus allocation in farmland: a case-based reasoning model", *in : 2012 International Congress on Environmental Modelling and Software, Managing Resources of a Limited Planet, Sixth Biennial Meeting*, R. Seppelt, A. Voinov, S. Lange, D. Bankamp (editors), International Environmental Modelling and Software Society (iEMSs), p. 8, Leipzig, Germany, July 2012, https://hal.archives-ouvertes.fr/hal-00724085.

[534] T. Meilender, J. Lieber, G. Herengt, F. Palomares, N. Jay, "A Semantic Wiki for Editing and Sharing Decision Guidelines in Oncology", *in : The 24th European Medical Informatics Conference - MIE 2012*, J. Mantas (editor), M.Cristina Mazzoleni, Pise, Italy, August 2012, https://hal.inria.fr/hal-00736711.

[535] T. Meilender, J. Lieber, F. Palomares, N. Jay, "From Web 1.0 to Social Semantic Web: Lessons Learnt from a Migration to a Medical Semantic Wiki", *in : 9th Extended Semantic Web Conference - ESWC 2012*, Heraklion, Greece, May 2012, https://hal.inria.fr/hal-00736706.

[536] T. Meilender, J. Lieber, F. Palomares, N. Jay, "Semantic decision trees editing for decision support with KcatoS - Application in oncology", *in : SIMI 2012: Semantic Interoperability in Medical Informatics*, Heraklion, Greece, May 2012, https://hal.inria.fr/hal-00736710.

[537] L. F. Melo Mora, Y. Toussaint, "Automatic Validation of Terminology by Means of Formal Concept Analysis", *in : International Conference in Formal Concept Analysis - ICFCA 2015*, J. Baixeries, C. Sacarea, M. Ojeda-Aciego (editors), *Formal Concept Analysis - Lecture Notes in Computer Science*, *9113*, Springer, p. 236–251, Nerja , Spain, June 2015, https://hal.inria.fr/hal-01176422.

[538] G. Personeni, S. Daget, C. Bonnet, P. Jonveaux, M.-D. Devignes, M. Smaïl-Tabbone, A. Coulet, “ILP for Mining Linked Open Data: a biomedical Case Study”, *in : The 24th International Conference on Inductive Logic Programming (ILP 2014)*, Nancy, France, September 2014, https://hal.inria.fr/hal-01095597.

[539] G. Personeni, S. Daget, C. Bonnet, P. Jonveaux, M.-D. Devignes, M. Smaïl-Tabbone, A. Coulet, “Mining Linked Open Data: A Case Study with Genes Responsible for Intellectual Disability”, *in : Data Integration in the Life Sciences - 10th International Conference, DILS 2014*, E. R. Helena Galhardas (editor), *Lecture Notes in Computer Science*, *8574*, Springer, p. 16 – 31, Lisbon, Portugal, July 2014, https://hal.inria.fr/hal-01095591.

[540] G. Personeni, A. Hermann, J. Lieber, “Adapting Propositional Cases Based on Tableaux Repairs Using Adaptation Knowledge”, *in : International Conference on Case-Based Reasoning*, L. Lamontagne, E. Plaza (editors), *Lecture Notes in Computer Science*, *Case-Based Reasoning Research and Development*, 8765, Luc Lamontagne and Enric Plaza, Springer International Publishing, p. 390–404, Cork, Ireland, September 2014, https://hal.inria.fr/hal-01092210.

[541] C. Raïssi, J. Pei, “Towards Bounding Sequential Patterns”, *in : 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD-2011*, C. Apté, J. Ghosh, P. Smyth (editors), ACM, San Diego, United States, August 2011. ISBN : 978-1-4503-0813-7, https://hal.inria.fr/inria-00623550.

[542] J.-H. Ramaroson, D. Hervé, B. Ramamonjisoa, F. Le Ber, “Organisation spatiale, dessinée par les paysans, du corridor forestier de Fianarantsoa”, *in : 4ème édition Forum de la recherche: ” Innovations scientifiques et technologiques - Valorisation de la recherche ”*, Antananarivo, Madagascar, July 2012, https://hal.archives-ouvertes.fr/hal-00764152.

[543] D. Rizzo, J.-F. Mari, E. Marraccini, E.-G. Lazrak, “Agricultural landscape segmentation: a stochastic method to map heterogeneous variables”, *in : Advances in Spatial Typologies: How to move from concepts to practice?*, IALE-Europe Thematic Workshop 2014:, Lisbon, Portugal, July 2014, https://hal.inria.fr/hal-01098402.

[544] M. Rouane-Hacene, M. Huchard, A. Napoli, P. Valtchev, “Soundness and Completeness of Relational Concept Analysis”, *in : ICFCA’2013: 11th International Conference on Formal Concept Analysis*, P. Cellier, F. Distel, B. Ganter (editors), *Lecture Notes in Computer Science*, *7880*, Springer Netherlands, p. 228–243, Dresden, Germany, May 2013, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00833506.

[545] L. Shi, Y. Toussaint, A. Napoli, A. Blansché, “Mining for Reengineering: an Application to Semantic Wikis using Formal and Relational Concept Analysis”, *in : The Semanic Web: Research and Applications - 8th Extended Semantic Web Conference, ESWC 2011*, G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, J. Pan, P. D. Leenhee (editors), *Lecture Notes in Computer Science*, *6644*, Springer, p. 421–435, Heraklion, Crete, Greece, May 2011, https://hal.inria.fr/hal-00646450.

[546] H. Skaf-Molli, E. Desmontils, E. Nauer, G. Canals, A. Cordier, M. Lefevre, P. Molli, Y. Toussaint, “Knowledge Continuous Integration Process (K-CIP)”, *in : WWW 2012 - SWCS’12 Workshop - 21st World Wide Web Conference - Semantic Web Collaborative Spaces workshop, WWW 2012 Companion*, p. 1075–1082, Lyon, France, April 2012, https://hal.inria.fr/hal-00765596.

[547] A. Soulet, C. Raïssi, M. Plantevit, B. Crémilleux, “Mining Dominant Patterns in the Sky”, *in : The 11th IEEE International Conference on Data Mining - ICDM 2011*, IEEE (editor), Vancouver,

B.C, Canada, December 2011. This is still a ”to appear” publication., https://hal.inria.fr/inria-00623566.

[548] R. Stanley, H. Astudillo, V. Codocedo, A. Napoli, “A Conceptual-KDD approach and its application to cultural heritage”, *in : Concept Lattices and their Applications*, M. Ojeda-Aciego, J. Outrata (editors), L3i laboratory, University of La Rochelle, p. 163–174, La Rochelle, France, October 2013, https://hal.inria.fr/hal-00880002.

[549] L. Szathmary, P. Valtchev, A. Napoli, R. Godin, A. Boc, V. Makarenkov, “Fast Mining of Iceberg Lattices: A Modular Approach Using Generators”, *in : The Eighth International Conference on Concept Lattices and their Applications - CLA 2011*, A. Napoli, V. Vychodil (editors), INRIA Nancy Grand Est - LORIA, Nancy, France, October 2011, https://hal.inria.fr/hal-00640898.

[550] L. Szathmary, P. Valtchev, A. Napoli, R. Godin, “Efficient Vertical Mining of Minimal Rare Itemsets”, *in : CLA - The Ninth International Conference on Concept Lattices and Their Applications - 2012*, U. Priss, L. Szathmary (editors), University of Malaga, Spain, Fuengirola, Spain, 2012, https://hal.inria.fr/hal-00769031.

[551] L. Szathmary, P. Valtchev, A. Napoli, R. Godin, “Finding minimal rare itemsets in a depth-first manner”, *in : ECAI Workshop on Formal Concept Analysis for Artificial Intelligence (FCA4AI), Montpellier*, S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), *ECAI Workshop on Formal Concept Analysis for Artificial Intelligence (FCA4AI), Montpellier*, *939*, CEUR Proceedings, p. 71–78, Montpellier, France, 2012, https://hal.inria.fr/hal-00769028.

[552] M. T. Tang, Y. Toussaint, “A Collaborative Approach for FCA-Based Knowledge Extraction”, *in : CLA - The Tenth International Conference on Concept Lattices and Their Applications - 2013*, La Rochelle, France, October 2013, https://hal.inria.fr/hal-00906814.

[553] J. Wiederkehr, B. Fontan, C. Grac, F. Labat, F. Le Ber, M. Trémolières, “Stream multi-index assessment and associated uncertainties : application to macroinvertebrate and macrophytes”, *in : Journées Internationales de Limnologie et d’Océanographie - JILO 2012*, Clermont-Ferrand, France, October 2012, https://hal.archives-ouvertes.fr/hal-00764287.

[554] M. Xue, P. Karras, C. Raïssi, P. Kalnis, H. K. Pung, “Delineating social network data anonymization via random edge perturbation”, *in : CIKM - 21st ACM International Conference on Information and Knowledge Management - 2012*, X. wen Chen, G. Lebanon, H. Wang, M. J. Zaki (editors), Maui, United States, October 2012, https://hal.inria.fr/hal-00768441.

[555] M. Xue, P. Karras, C. Raïssi, H. K. Pung, “Utility-Driven Anonymization in Data Publishing”, *in : 20th ACM Conference on Information and Knowledge Management - CIKM 2011*, ACM, ACM, Glasgow, United Kingdom, October 2011. This is a ”to appear” publication, https://hal.inria.fr/inria-00623578.

[556] M. Xue, P. Karras, C. Raïssi, J. Vaidya, K.-L. Tan, “Anonymizing set-valued data by nonreciprocal recoding”, *in : KDD - The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - 2012*, Q. Yang, D. Agarwal, J. Pei (editors), Beijing, China, August 2012, https://hal.inria.fr/hal-00768428.

[557] M. Xue, P. Papadimitriou, C. Raïssi, P. Kalnis, H. K. Pung, “Distributed Privacy Preserving Data Collection”, *in : 16th International Conference on Database Systems for Advanced Applications - DASFAA 2011*, J. X. Yu, M. H. Kim, R. Unland (editors), *Lecture Notes in Computer Science*, *6587*, springer, p. 93–107, Hong Kong, China, April 2011, https://hal.inria.fr/inria-00610951.

Articles in National Peer-Reviewed Journal

[558] Z. Dvořák, J.-S. Sereni, J. Volec, "Subcubic triangle-free graphs have fractional chromatic number at most $14/5$", *Journal of the London Mathematical Society 89*, 3, June 2014, p. 641–662, https://hal.archives-ouvertes.fr/hal-00779634.

[559] C. Grac, A. Braud, F. Le Ber, M. Trémolières, "Un système d'information pour le suivi et l'évaluation de la qualité des cours d'eau – Application à l'hydro-écorégion de la plaine d'Alsace", *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information 16*, 2011, p. 9–30, https://hal.archives-ouvertes.fr/hal-00601991.

[560] F. Kohler, N. Jay, F. Ducreau, G. Casanova, C. Kohler, A.-C. Benhamou, "COURLIS (COURs en LIgne de Statistiques appliquées) : un MOOC francophone innovant", *HEGEL 3*, 1, 2013, p. 27–32, https://hal.inria.fr/hal-00903262.

[561] F. Le Ber, S. Nogry, C. Brassac, M. Benoît, "Capitalisation d'expériences pour la mise en place d'observatoires de pratiques agricoles", *International Journal of Geomatics and Spatial Analysis / Revue Internationale de Géomatique 21*, 2011, p. 99–118, https://hal.archives-ouvertes.fr/hal-00576238.

National Peer-Reviewed Conferences

[562] T. Bourquard, J. Azé, A. Poupon, D. Ritchie, "Protein-protein docking based on shape complementarity and Voronoi fingerprint", *in : Journées Ouvertes Biologie Informatique Mathématiques*, E. P. R. Emmanuel BARILLOT, Christine F ROIDEVAUX (editor), *JOBIM 2011*, Institut Pasteur, p. 9–16, Paris, France, June 2011, https://hal.inria.fr/inria-00613186.

[563] C. Chamard, E. Bresso, T. Boukobza, M.-D. Devignes, M. Smaïl-Tabbone, A. Chesnel, H. Dumond, "Long chain alkylphenol mixture promotes mammary epithelial cell metaplastic phenotype through an ERalpha36-mediated mechanism.", *in : 8e Forum du Cancéropôle Grand Est*, Strasbourg, France, November 2014. Présentation Poster, https://hal.archives-ouvertes.fr/hal-01095763.

[564] J. Cojan, V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber, E. Nauer, G. Personeni, "Révisor : un ensemble de moteurs d'adaptation de cas par révision des croyances", *in : JIAF - Septièmes Journées de l'Intelligence Artificielle Fondamentale - 2013*, Aix-en-Provence, France, June 2013, https://hal.inria.fr/hal-00856487.

[565] J. Cojan, J. Lieber, "Adaptation par révision et adaptation différentielle : comparaison de deux approches de l'adaptation", *in : 19ème atelier Français de Raisonnement à Partir de Cas - Rapc2011*, Fadi Badra and Amélie Cordier, Chambéry, France, May 2011, https://hal.inria.fr/inria-00595400.

[566] M. Couceiro, B. Teheux, J.-L. Marichal, "Agrégation des valeurs médianes et fonctions compatibles pour la comparaison", *in : 24ème Conférence Rencontres francophones sur la Logique Floue et ses Applications (LFA 2015)*, Poitiers, France, November 2015, https://hal.inria.fr/hal-01247525.

[567] A. Coulet, F. Domenach, M. Kaytoue, A. Napoli, "Using pattern structures for analyzing ontology-based annotations of biomedical data", *in : Septièmes Journées d'Intelligence Artificielle Fondamentale*, S. Konieczny, N. Maudet, D. Habet, V. Risch (editors), LSIS and LIF Marseille, p. 97–106, Aix-en-Provence, France, June 2013, https://hal.inria.fr/hal-00922392.

[568] S. Da Silva, C. Lavigne, F. Le Ber, "Analyse de la structure des haies dans les vergers pour la définition de paysages mieux adaptés contre les bioagresseurs", *in : SAGEO 2011 - International Conference on Spatial Analysis and GEOmatics Conférence internationale de Géomatique et d'Analyse Spatiale Au sein de la 25e Conférence Internationale de Cartographie*, p. 1–4, Paris, France, July 2011, https://hal.archives-ouvertes.fr/hal-00615301.

[569] V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber, "Révision des croyances dans la clôture propositionnelle d'une algèbre qualitative", *in : JIAF-2014 – Huitièmes Journées de l'Intelligence Artificielle Fondamentale, JIAF - Huitièmes Journées de l'Intelligence Artificielle Fondamentale*, p. 10, Angers, France, June 2014, https://hal.inria.fr/hal-01093977.

[570] V. Dufour-Lussier, F. Le Ber, J. Lieber, L. Martin, "Adaptation de cas spatiaux et temporels", *in : 20ème atelier Français de Raisonnement à Partir de Cas*, Laura Martin and Zied Yakoubi, Paris, France, June 2012, https://hal.inria.fr/hal-00712982.

[571] V. Dufour-Lussier, F. Le Ber, J. Lieber, "Quels formalismes temporels pour représenter des connaissances extraites de textes de recettes de cuisine ?", *in : Représentation et raisonnement sur le temps et l'espace*, S. Laborie, F. L. Ber (editors), Chambéry, France, May 2011, https://hal.inria.fr/inria-00634735.

[572] V. Dufour-Lussier, F. Le Ber, J. Lieber, "Extension du formalisme des flux opérationnels par une algèbre temporelle", *in : Sixièmes Journées de l'Intelligence Artificielle Fondamentale (JIAF), Sixièmes Journées de l'Intelligence Artificielle Fondamentale*, Sébastien Konieczny, p. 133–142, Toulouse, France, May 2012, https://hal.inria.fr/hal-00712978.

[573] E. Egho, C. Raïssi, T. Calders, N. Jay, A. Napoli, "Vers une mesure de similarité pour les séquences complexes", *in : Extraction et gestion des connaissances (EGC'2013), Extraction et gestion des connaissances*, Cépaduès, p. 335–340, Toulouse, France, January 2013, https://hal.inria.fr/hal-00885965.

[574] M. Fabrègue, A. Braud, S. Bringay, F. Le Ber, M. Teisseire, "Extraction de motifs spatio-temporels à différentes échelles avec gestion de relations spatiales qualitatives.", *in : 30ème édition, Inforsid 2012*, p. 123–138, France, May 2012, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00735616.

[575] E. Gaillard, J. Lieber, Y. Naudet, E. Nauer, "Raisonner sur des connaissances provenant d'une e-communauté", *in : IC - 24èmes Journées francophones d'Ingénierie des Connaissances*, Lille, France, 2013, https://hal.inria.fr/hal-00918540.

[576] E. Gaillard, J. Lieber, E. Nauer, "Taaable : un système de raisonnement à partir de cas qui adapte des recettes de cuisine", *in : 1ère Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA 2015)*, Rennes, France, July 2015, https://hal.inria.fr/hal-01178316.

[577] T. Guyet, F. Le Ber, S. Da Silva, C. Lavigne, "Comparaison des chemins de Hilbert adaptatif et des graphes de voisinage pour la caractérisation d'un parcellaire agricole", *in : Conférence Extraction et Gestion de Connaissances*, Rennes, France, January 2014, https://hal.inria.fr/hal-00916964.

[578] R. Hafiane, M. Smaïl-Tabbone, M.-D. Devignes, S. Tabbone, "Clustering optimal de gènes fondé sur une mesure de similarité sémantique", *in : 10ème édition de la COnférence en Recherche d'Information et Applications - CORIA 2013*, C. Berrut (editor), p. 15 p, Neufchâtel, Switzerland, April 2013, https://hal.inria.fr/hal-00920700.

[579] A. Hermann, M. Ducassé, S. Ferré, J. Lieber, "Une approche fondée sur le raisonnement à partir de cas pour la mise à jour interactive d'objets du Web sémantique", *in : 21ème atelier Français de Raisonnement à Partir de Cas (RàPC)*, Lille, France, 2013, https://hal.inria.fr/hal-00910294.

[580] M. Kaytoue, S. O. Kuznetsov, A. Napoli, "Revisiting Numerical Pattern Mining with Formal Concept Analysis", *in : JIAF - Cinquièmes Journées de l'Intelligence Artificielle Fondamentale - 2011*, Lyon, France, June 2011, https://hal.inria.fr/inria-00600222.

[581] M. Kaytoue, S. O. Kuznetsov, A. Napoli, G. Polaillon, "Symbolic Data Analysis and Formal Concept Analysis", *in : XVIIIeme Rencontres de la Société Francophone de Classification - SFC 2011*, R. E. G. Cleuziou (editor), MAPMO - LIFO Orléans, Orléans, France, September 2011, https://hal.inria.fr/hal-00646457.

[582] E. G. Lazrak, N. Schaller, J.-F. Mari, "Extraction de connaissances agronomiques par fouille des voisinages entre occupations du sol", *in : Atelier en marge d'EGC 2011*, Brest, France, January 2011, https://hal.inria.fr/inria-00560098.

[583] J. Lieber, "Révision des croyances dans une clôture propositionnelle de contraintes linéaires", *in : Journées d'intelligence artificielle fondamentale, plate-forme intelligence artificielle*, N. Maudet, B. Zanuttini (editors), p. 10, Rennes, France, June 2015, https://hal.inria.fr/hal-01178267.

[584] T. Meilender, N. Jay, J. Lieber, F. Palomares, "Édition sémantique d'arbres de décision pour l'oncologie avec KcatoS", *in : 1ère édition du Symposium sur l'Ingénierie de l'Information Médicale - SIIM 2011*, Toulouse, France, June 2011, https://hal.inria.fr/hal-00646843.

[585] T. Meilender, N. Jay, J. Lieber, F. Palomares, "Les moteurs de wikis sémantiques : un état de l'art", *in : Extraction et gestion des connaissances (EGC'2011)*, p. 575–580, Brest, France, January 2011, https://hal.archives-ouvertes.fr/hal-00573821.

[586] G. Personeni, A. Hermann, J. Lieber, "Adaptation de cas propositionnels par réparations fondées sur des connaissances d'adaptation", *in : RàPC - 21ème atelier Français de Raisonnement à Partir de Cas - 2013*, V. Dufour-Lussier, E. Gaillard, A. Hermann (editors), Lille, France, July 2013, https://hal.inria.fr/hal-00910276.

[587] M. Plantevit, C. Raïssi, B. Crémilleux, "Motifs séquentiels δ-libres", *in : Extraction et gestion des connaissances (EGC'2011)*, Ali Khenchaf, Pascal Poncelet, Hermann-Éditions, Brest, France, January 2011, https://hal.inria.fr/hal-00653579.

[588] M. Pupin, M. Smaïl-Tabbone, P. Jacques, M.-D. Devignes, V. Leclère, "NRPS toolbox for the discovery of new nonribosomal peptides and synthetases", *in : Journées Ouvertes en Biologie, l'Informatique et les Mathématiques - JOBIM 2012*, F. C. et Denis Tagu (editor), p. 89–93, Rennes, France, July 2012, https://hal.inria.fr/hal-00734312.

[589] M. T. Tang, Y. Toussaint, "Vers un processus continu d'extraction de connaissances à partir de textes", *in : IC - 24èmes Journées francophones d'Ingénierie des Connaissances, Ingénierie des Connaissances*, Lille, France, July 2013, https://hal.inria.fr/hal-00861865.

[590] V. Venkatraman, D. Ritchie, "Predicting Multicomponent Protein Assemblies Using an Ant Colony Approach", *in : International Conference on Swarm Intelligence*, Cergy, France, June 2011, https://hal.inria.fr/inria-00619204.

Books

[591] B. Bucher, F. Le Ber, *Développements logiciels en géomatique*, Hermes Lavoisier, 2012, https://hal.archives-ouvertes.fr/hal-00724084.

[592] B. Bucher, F. Le Ber, *Innovative Software Development in GIS*, ISTE - WILEY, 2012, https://hal.archives-ouvertes.fr/hal-00724080.

Books or Proceedings Editing

[593] *Proceedings of ECCB 2014: The 13th European Conference on Computational Biology*, *Bioinformatics*, *30*, 17, Strasbourg, France, Oxford University Press, September 2014, https://hal.inria.fr/hal-01097346.

[594] C. Carpineto, S. O. Kuznetsov, A. Napoli (editors), *FCAIR 2012 Formal Concept Analysis Meets Information Retrieval Workshop co-located with the 35th European Conference on Information Retrieval (ECIR 2013) March 24, 2013, Moscow, Russia*, *977*, CEUR Proceedings, 2013, 157p., https://hal.inria.fr/hal-00922617.

[595] N. Jay, I. Tiddi, M. D'Aquin (editors), *Proceedings of the Workshop on Linked Data for Knowledge Discovery (LD4KD 2014)*, *Linked Data for Knowledge Discovery, Proceedings of the 1st Workshop on Linked Data for Knowledge Discovery*, *1232*, 3, Nicolas Jay and Illaria Tiddi and Mathieu D'Aquin, Nancy, France, ceur workshop proceedings, September 2014, https://hal.archives-ouvertes.fr/hal-01101149.

[596] S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), *Proceedings of the ECAI Workshop on Formal Concept Analysis for Artificial Intelligence (FCA4AI)*, *CEUR Proceedings*, *939*, CEUR Proceedings (http://ceur-ws.org/Vol-939/), 2012, 88p., https://hal.inria.fr/hal-00768961.

[597] S. O. Kuznetsov, A. Napoli, S. Rudolph, *International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI at IJCAI 2013, Beijing, China, August 4 2013)*, *1058*, CEUR Proceedings, 2013, https://hal.inria.fr/hal-00922616.

[598] S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), *Proceedings of the International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2014)*, *CEUR Proceedings*, *1257*, Prague, Czech Republic, August 2014, 100p., https://hal.inria.fr/hal-01101150.

[599] S. O. Kuznetsov, A. Napoli, S. Rudolph (editors), *Workshop NotesInternational Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI 2015)*, *CEUR Workshop Proceedings 1430*, *CEUR Workshop Proceedings 1430*, France, July 2015, https://hal.inria.fr/hal-01252624.

[600] D. B. Leake, J. Lieber (editors), *Proceedings of the Workshops of the 22nd International Conference on Case-Based Reasoning*, Cork, Ireland, September 2014, 260p., https://hal.inria.fr/hal-01097373.

[601] M. Minor, E. Nauer (editors), *Proceedings of the Computer Cooking Contest Workshop at ICCBR 2014*, *Proceedings of the Workshops of the 22nd International Conference on Case-Based Reasoning*, Cork, Ireland, September 2014, 48p., https://hal.inria.fr/hal-01095840.

[602] A. Napoli, V. Vychodil (editors), *The 8th International Conference on Concept Lattices and Their Applications CLA 2011*, INRIA Nancy-Grand-Est and LORIA, 2011, 416p., https://hal.inria.fr/hal-00758850.

Book chapters

[603] B. Bucher, J. Gaffuri, F. Le Ber, T. Libourel, "Challenges and Proposals for Software Development Pooling in Geomatics", *in : Innovative Software Development in GIS*, B. B. et F. Le Ber (editor), *GIS Series*, ISTE - WILEY, 2012, p. 293–316, ISBN : 978-1848213647, https://hal.archives-ouvertes.fr/hal-00724082.

[604] B. Bucher, J. Gaffuri, F. Le Ber, T. Libourel, "Défis et propositions pour la mutualisation de développements logiciels en géomatique", *in : Développements logiciels en géomatique – innovations et mutualisation*, B. B. et F. Le Ber (editor), *IGAT*, Hermes Lavoisier, 2012, p. 271–290, https://hal.archives-ouvertes.fr/hal-00718768.

[605] B. Bucher, F. Le Ber, "Introduction", *in : Développements logiciels en géomatique – innovations et mutualisation*, B. B. et F. Le Ber (editor), *IGAT*, Hermes Lavoisier, 2012, p. 17–34, https://hal.archives-ouvertes.fr/hal-00718748.

[606] B. Bucher, F. Le Ber, "Introduction", *in : Innovative Software Development in GIS*, B. B. et F. Le Ber (editor), *GIS Series*, ISTE - WILEY, 2012, p. 1–21, https://hal.archives-ouvertes.fr/hal-00724078.

[607] J. Cojan, J. Lieber, "Applying Belief Revision to Case-Based Reasoning", *in : Computational Approaches to Analogical Reasoning: Current Trends*, *Studies in Computational Intelligence*, *548*, Springer, 2014, p. 133 – 161, https://hal.inria.fr/hal-01095344.

[608] A. Cordier, V. Dufour-Lussier, J. Lieber, E. Nauer, F. Badra, J. Cojan, E. Gaillard, L. Infante-Blanco, P. Molli, A. Napoli, H. Skaf-Molli, "Taaable: a Case-Based System for personalized Cooking", *in : Successful Case-based Reasoning Applications-2*, S. Montani and L. C. Jain (editors), *Studies in Computational Intelligence*, *494*, Springer, January 2014, p. 121–162, https://hal.inria.fr/hal-00912767.

[609] A. Coulet, M. Smaïl-Tabbone, A. Napoli, M.-D. Devignes, "Ontology-based knowledge discovery in pharmacogenomics.", *in : Software Tools and Algorithms for Biological Systems*, H. R. Arabnia and Q.-N. Tran (editors), *696*, Springer, 2011, p. 357–66, https://hal.inria.fr/inria-00585072.

[610] V. Dufour-Lussier, B. Guillaume, G. Perrier, "Parsing Coordination Extragrammatically", *in : Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers*, Z. Vetulani and J. Mariani (editors), *Lecture Notes in Computer Science (LNCS) series / Lecture Notes in Artificial Intelligence (LNAI) subseries*, *8387*, Springer International Publishing, July 2014, p. 12, https://hal.inria.fr/hal-00921033.

[611] E. Egho, C. Raïssi, D. Ienco, N. Jay, A. Napoli, P. Poncelet, C. Quantin, M. Teisseire, "Healthcare Trajectory Mining by Combining Multidimensional Component and Itemsets", *in : New Frontiers in Mining Complex Patterns - First International Workshop, NFMCP 2012, Held in Conjunction with ECML/PKDD 2012*, A. Appice (editor), Springer, September 2012, https://hal.inria.fr/hal-00922512.

[612] M. Kaytoue, S. Duplessis, A. Napoli, "Toward the Discovery of Itemsets with Significant Variations in Gene Expression Matrices", *in : Classification and Multivariate Analysis for Complex Data Structures*, B. Fichet, D. Piccolo, R. Verde, and M. Vichi (editors), *Studies in Classification, Data Analysis, and Knowledge Organization*, Springer Berlin Heidelberg, 2011, p. 463–471, https://hal.inria.fr/inria-00600233.

[613] F. Le Ber, C. Brassac, "Modéliser " l'entre deux " dans l'organisation spatiale des exploitations agricoles – Mise en évidence de quelques problématiques", *in : Géoagronomie, paysage et projets de territoire*, S. Lardon (editor), *Indisciplines*, QUAE, October 2012, p. 63–72, https://hal.archives-ouvertes.fr/hal-00742683.

[614] F. Le Ber, B. Bucher, "Analyse des spécificités des développements logiciels en géomatique", *in : Développements logiciels en géomatique – innovations et mutualisation*, B. B. et F. Le Ber (editor), *IGAT*, Hermes Lavoisier, 2012, p. 264–269, https://hal.archives-ouvertes.fr/hal-00718762.

[615] F. Le Ber, B. Bucher, "Analysis of the specificities of Software Development in Geomatics Research", *in : Innovative Software Development in GIS*, B. B. et F. Le Ber (editor), *GIS Series*, ISTE - WILEY, 2012, p. 285–292, ISBN : 978-1848213647, https://hal.archives-ouvertes.fr/hal-00724081.

[616] F. Le Ber, J.-F. Mari, "GenExp-LandSiTes : un logiciel générateur de paysages agricoles 2D", *in : Développements logiciels en géomatique*, B. B. et Florence Le Ber (editor), *Information Géographique et Aménagement du Territoire*, HERMES Lavoisier, July 2012, p. 181 – 202, https://hal.inria.fr/hal-00718266.

[617] F. Le Ber, J.-F. Mari, "GenExp-LandSiTes: a 2D Agricultural Generating Piece of Software", *in : Innovative Software Development in GIS*, B. Bucher and F. L. Ber (editors), *GIS*, ISTE WILEY, July 2012, p. 189 – 214, https://hal.inria.fr/hal-00718259.

[618] J.-F. Mari, F. Le Ber, E.-G. Lazrak, M. Benoît, C. Eng, A. Thibessard, P. Leblond, "Using Markov Models to Mine Temporal and Spatial Data", *in : New Fundamental Technologies in Data Mining*, K. Funatsu and K. Hasegawa (editors), Intech, 2011, p. 561–584, https://hal.inria.fr/inria-00566801.

[619] D. Ritchie, V. I. Pérez-Nueno, "ParaFit", *in : Scaffold Hopping in Medicinal Chemistry*, N. Brown (editor), *Methods and Principles in Medicinal Chemistry*, *58*, Wiley, 2013, https://hal.inria.fr/hal-00880352.

[620] D. Ritchie, "Modeling Protein-Protein Interactions by Rigid-Body Docking", *in : Drug Design Strategies: Computational Techniques and Applications*, T. Clark and L. Banting (editors), *RSC Drug Discovery Series*, RSC Publishing, 2012, p. 56–86, https://hal.inria.fr/hal-00666809.

Other Publications

[621] Y. Abid, A. Imine, A. Napoli, C. Raïssi, M. Rigolot, M. Rusinowitch, "Analyse d'activité et exposition de la vie privée sur les médias sociaux", 16ème conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2016), January 2016, Poster, https://hal.inria.fr/hal-01241619.

[622] M. Alam, A. Napoli, "Lattice-Based View Access: A way to Create Views over SPARQL Query for Knowledge Discovery.", *research report*, INRIA, September 2014, https://hal.archives-ouvertes.fr/hal-01059528.

[623] F. Bonomo, I. Koch, P. Torres, M. Valencia-Pabon, "k-tuple chromatic number of the cartesian product of graphs", working paper or preprint, November 2014, https://hal.archives-ouvertes.fr/hal-01103534.

[624] F. Bonomo, O. Schaudt, M. Stein, M. Valencia-Pabon, "b-coloring is NP-hard on co-bipartite graphs and polytime solvable on tree-cographs", working paper or preprint, October 2013, https://hal.archives-ouvertes.fr/hal-00926924.

[625] M. W. Chekol, "Schema Query Containment", *Research Report number RR-8484*, Inria Nancy - Grand Est (Villers-lès-Nancy, France), February 2014, https://hal.inria.fr/hal-00951201.

[626] V. Codocedo, I. Lykourentzou, A. Napoli, "Semantic Indexing and Retrieval based on Formal Concept Analysis", *Research report*, Inria, 2012, https://hal.archives-ouvertes.fr/hal-00713202.

[627] M. Couceiro, S. Foldes, G. Meletiou, "Arrow Type Impossibility Theorems over Median Algebras", working paper or preprint, August 2015, https://hal.archives-ouvertes.fr/hal-01185977.

[628] M. Couceiro, L. Haddad, K. Schölzel, T. Waldhauser, "On the interval of strong partial clones of Boolean functions containing Pol((0,0),(0,1),(1,0))", working paper or preprint, August 2015, https://hal.archives-ouvertes.fr/hal-01184404.

[629] M. Couceiro, J.-L. Marichal, B. Teheux, "Conservative median algebras", working paper or preprint, 2014, https://hal.archives-ouvertes.fr/hal-01090614.

[630] M. Couceiro, G. Meletiou, "On a special class of median algebras", working paper or preprint, August 2015, https://hal.inria.fr/hal-01248142.

[631] A. P. Dove, J. R. Griggs, R. J. Kang, J.-S. Sereni, "Supersaturation in the Boolean lattice", *Research report*, Loria & Inria Grand Est, 2013, https://hal.archives-ouvertes.fr/hal-00802000.

[632] V. Dufour-Lussier, A. Hermann, F. Le Ber, J. Lieber, "Belief revision in the propositional closure of a qualitative algebra (extended version)", *Technical report*, INRIA Nancy, July 2014, This is the extended version of an article originally presented at the 14th International Conference on Principles of Knowledge Representation and Reasoning, https://hal.inria.fr/hal-00954512.

[633] K. Edwards, J. Van Den Heuvel, R. J. Kang, J.-S. Sereni, "Extension from Precoloured Sets of Edges", *Research report*, Loria & Inria Grand Est, 2014, https://hal.archives-ouvertes.fr/hal-01024843.

[634] E. Egho, C. Raïssi, T. Calders, N. Jay, A. Napoli, "On Measuring Similarity for Sequences of Itemsets", *Research Report number RR-8086*, INRIA, October 2012, https://hal.inria.fr/hal-00740231.

[635] E. Egho, C. Raïssi, N. Jay, A. Napoli, "Mining Heterogeneous Multidimensional Sequential Patterns", *Research Report number RR-8521*, INRIA, April 2014, https://hal.inria.fr/hal-00979804.

[636] E. Gaillard, "Les systèmes informatiques fondés sur la confiance : un état de l'art.", *Research report*, Loria & Inria Grand Est, November 2011, https://hal.inria.fr/hal-00662479.

[637] T. Guyet, S. Da Silva, C. Lavigne, F. Le Ber, "Caractérisation d'un parcellaire agricole : comparaison des sacs de noeuds obtenus par un chemin de Hilbert adaptatif et un graphe de voisinage", January 2014, Atelier FST "fouille de données spatiales et temporelles" associé à la conférence EGC, Rennes, 28 janvier 2014, https://hal.inria.fr/hal-01100583.

[638] M. Huchard, A. Napoli, M. Rouane-Hacene, P. Valtchev, "A gentle introduction to Relational Concept Analysis, Tutorial ICFCA 2011", May 2011, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00616275.

[639] M. Huchard, A. Napoli, "Analyse Relationnelle de Concepts: Une approche pour fouiller des ensembles de données multi-relationnels", October 2012, http://hal-lirmm.ccsd.cnrs.fr/lirmm-00808769.

[640] D. Král', C.-H. Liu, J.-S. Sereni, P. Whalen, Z. Yilma, "A new bound for the 2/3 conjecture", *Research report*, Loria & Inria Grand Est, 2012, https://hal.archives-ouvertes.fr/hal-00686989.

[641] S. Laborie, F. Le Ber, "SIXIÈME ATELIER : Représentation et raisonnement sur le temps et l'espace (RTE 2011)", Atelier Représentation et raisonnement sur le temps et l'espace (RTE 2011), 2011, https://hal.inria.fr/hal-00640820.

[642] J. Lieber, "Révision des croyances dans une clôture propositionnelle de contraintes linéaires (version étendue)", *Research report*, LORIA, UMR 7503, Université de Lorraine, CNRS, Vandoeuvre-lès-Nancy ; Inria Nancy - Grand Est (Villers-lès-Nancy, France), May 2015, https://hal.inria.fr/hal-01155235.

[643] M. Loebl, J.-S. Sereni, "Potts partition function and isomorphisms of trees", *Research report*, Loria & Inria Grand Est, 2014, https://hal.archives-ouvertes.fr/hal-00992104.

[644] J.-F. Mari, "CarottAge Windows pour les données Teruti : manuel d'utilisation", *Technical report*, Loria & Inria Grand Est, February 2014, https://hal.inria.fr/hal-00951102.

[645] A. Muller-Gueudin, M. Buee, A. Deveau, A. Gégout-Petit, S. Martin, I.-C. Morarescu, C. Raïssi, ""Truffinet": Inférence de réseaux d'interactions microbiennes dans la truffe. Rapport scientique du PEPS Mirabelles 2014", *Research report*, IECL ; INRIA BIGS, September 2015, https://hal.archives-ouvertes.fr/hal-01236087.

[646] G. Personeni, S. Daget, C. Bonnet, P. Jonveaux, M.-D. Devignes, M. Smaïl-Tabbone, A. Coulet, "Mining Linked Open Data: a Case Study with Genes Responsible for Intellectual Disability", ECCB'14 (European Conference on Computational Biology 2014), September 2014, Poster, https://hal.inria.fr/hal-01092800.

[647] G. Personeni, A. Hermann, J. Lieber, "Adapting propositional cases based on tableaux repairs using adaptation knowledge – extended report", *Research report*, Loria & Inria Grand Est, 2014, https://hal.archives-ouvertes.fr/hal-01011751.

[648] D. Rautenbach, J.-S. Sereni, "Transversals of Longest Paths and Cycles", *Research report*, Loria & Inria Grand Est, 2013, https://hal.archives-ouvertes.fr/hal-00793271.

[649] J.-S. Sereni, J. Volec, "A note on acyclic vertex-colorings", *Research report*, Loria & Inria Grand Est, 2013, https://hal.archives-ouvertes.fr/hal-00921122.

[650] P. Torres, M. Valencia-Pabon, "Stable Kneser Graphs are almost all not weakly Hom-Idempotent ", working paper or preprint, February 2015, https://hal.archives-ouvertes.fr/hal-01119741.

[651] J. Van Den Heuvel, D. Král', M. Kupec, J.-S. Sereni, J. Volec, "Extensions of Fractional Precolorings show Discontinuous Behavior", *Research report*, Loria & Inria Grand Est, 2012, https://hal.archives-ouvertes.fr/hal-00932924.

4 References for QGAR

Doctoral Dissertations

[652] A. Boumaiza, *Reconnaissance et localisation de symboles dans les documents graphiques: approches basées sur le treillis de concepts*, Theses, Université de Lorraine, May 2013.

[653] T. H. Do, *Sparse Representation over Learned Dictionary for Document Analysis*, Theses, Université de Lorraine, April 2014.

[654] M. Felhi, *Segmentation d'images de documents : catégorisation du contenu*, Theses, Université de Lorraine, June 2014.

[655] T. V. Hoang, *Image Representations for Pattern Recognition*, Theses, Université Nancy II, December 2011, https://tel.archives-ouvertes.fr/tel-00714651.

[656] S. Jouili, *Indexation de masses de documents graphiques : approches structurelles*, Theses, Université Nancy II, March 2011, https://tel.archives-ouvertes.fr/tel-00597711.

[657] B. Lamiroy, *On the Limits of Machine Perception and Interpretation*, Habilitation à diriger des recherches, Université de Lorraine, December 2013, https://tel.archives-ouvertes.fr/tel-00940209.

[658] K. Santosh, *Graphics Recognition using Spatial Relations and Shape Analysis*, Theses, Institut Polytechnique de Lorraine, November 2011.

[659] J. Weber, *Interactive morphological segmentation for video mining*, Theses, Université de Strasbourg, September 2011, https://tel.archives-ouvertes.fr/tel-00643585.

Articles in International Peer-Reviewed Journal

[660] T. S. Do Thanh Ha, Ramos-Terrades, "Sparse Representation over Learned Dictionary for Symbol Recognition (to appear)", *Signal Processing*, 2016.

[661] M. Felhi, S. Tabbone, M. V. Ortiz-Segovia, "Approche hybride de segmentation de pages à base d'un descripteur de traits", *Revue des Sciences et Technologies de l'Information (Série IDocument Numérique) 17*, 3, 2014, https://hal.archives-ouvertes.fr/hal-01254452.

[662] M. Grand-Brochier, A. Vacavant, G. Cerutti, C. Kurtz, J. Weber, L. Tougne, "Tree leaves extraction in natural images: Comparative study of pre-processing tools and segmentation methods", *IEEE Transactions on Image Processing 24*, 5, 2015, p. 1549–1560, https://hal.archives-ouvertes.fr/hal-01176699.

[663] M. Hasegawa, S. Tabbone, "Amplitude-only log Radon transform for geometric invariant shape descriptor", *Pattern Recognition 47*, 2, 2014, p. 643–658, https://hal.archives-ouvertes.fr/hal-01098678.

[664] M. Hasegawa, S. Tabbone, "Histogram of Radon transform with angle correlation matrix for distortion invariant shape descriptor", *Neurocomputing 173*, part 1, 2016, https://hal.archives-ouvertes.fr/hal-01254923.

[665] T. V. Hoang, E. H. Barney Smith, S. Tabbone, "Sparsity-based edge noise removal from bilevel graphical document images", *International Journal on Document Analysis and Recognition 17*, 2, June 2014, p. 161–179, https://hal.inria.fr/hal-00852418.

[666] T. V. Hoang, S. Tabbone, "Invariant pattern recognition using the RFM descriptor", *Pattern Recognition 45*, 1, January 2012, p. 271–284, https://hal.inria.fr/inria-00614835.

[667] T. V. Hoang, S. Tabbone, "The generalization of the R-transform for invariant pattern representation", *Pattern Recognition*, January 2012, https://hal.inria.fr/inria-00637631.

[668] T. V. Hoang, S. Tabbone, "Errata and comments on "Generic orthogonal moments: Jacobi–Fourier moments for invariant image description"", *Pattern Recognition 46*, 11, April 2013, p. 3148–3155, https://hal.inria.fr/hal-00820279.

[669] T. V. Hoang, S. Tabbone, "Fast generic polar harmonic transforms", *IEEE Transactions on Image Processing 23*, 7, May 2014, p. 2961 – 2971, https://hal.inria.fr/hal-01083716.

[670] T. V. Hoang, S. Tabbone, "Generic polar harmonic transforms for invariant image representation", *Image and Vision Computing 32*, 8, 2014, p. 497 – 509, https://hal.inria.fr/hal-01083722.

[671] S. Jouili, S. Tabbone, "Hypergraph-based image retrieval for graph-based representation", *Pattern Recognition 45*, 11, 2012, p. 4054–4068, https://hal.archives-ouvertes.fr/hal-00764616.

[672] S. K.C., B. Lamiroy, L. Wendling, "Symbol Recognition using Spatial Relations", *Pattern Recognition Letters 33*, 3, February 2012, p. 331–341, https://hal.inria.fr/inria-00626996.

[673] S. K.C., B. Lamiroy, L. Wendling, "DTW-Radon-based Shape Descriptor for Pattern Recognition", *International Journal of Pattern Recognition and Artificial Intelligence 27*, 3, March 2013, https://hal.inria.fr/hal-00823961.

[674] S. K.C., B. Lamiroy, L. Wendling, "Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition", *International Journal on Document Analysis and Recognition*, May 2013, https://hal.inria.fr/hal-00824521.

[675] S. K.C., C. Nattee, B. Lamiroy, "Relative Positioning of Stroke Based Clustering: A New Approach to On-line Handwritten Devanagari Character Recognition", *International Journal of Image and Graphics*, January 2012, https://hal.inria.fr/hal-00661237.

[676] S. K.C., L. Wendling, B. Lamiroy, "BoR: Bag-of-Relations for Symbol Retrieval", *International Journal of Pattern Recognition and Artificial Intelligence 28*, 6, 2014, https://hal.inria.fr/hal-01025765.

[677] N. Nacereddine, S. Tabbone, D. Ziou, "Robustness of Radon transform to white additive noise: ggeneral case study", *Electronics Letters 50*, 15, 2014, p. 1063 – 1065, https://hal.archives-ouvertes.fr/hal-01111246.

[678] N. Nacereddine, S. Tabbone, D. Ziou, "Similarity transformation parameters recovery based on Radon transform. Application in image registration and object recognition", *Pattern Recognition 48*, 7, 2015, https://hal.archives-ouvertes.fr/hal-01254921.

[679] N.-V. Nguyen, A. Boucher, J.-M. Ogier, S. Tabbone, "Cluster-based relevance feedback for CBIR: a combination of query point movement and query expansion", *Journal of ambient intelligence and humanized computing 3*, 4, 2012, p. 281–292, https://hal.archives-ouvertes.fr/hal-00764644.

[680] F. Petitjean, J. Weber, "Efficient Satellite Image Time Series Analysis Under Time Warping", *IEEE Geoscience and Remote Sensing Letters 11*, 6, June 2014, p. 1143 – 1147, https://hal.inria.fr/hal-00940767.

[681] M. Visani, O. Ramos-Terrades, S. Tabbone, "A protocol to characterize the descriptive power and the complementarity of shape descriptors", *International Journal on Document Analysis and Recognition 14*, 1, 2011, p. 87–100, https://hal.inria.fr/inria-00580311.

[682] J. Weber, S. Lefèvre, "Spatial and Spectral Morphological Template Matching", *Image and Vision Computing 30*, 12, December 2012, p. 934–945, https://hal.archives-ouvertes.fr/hal-00761281.

[683] J. Weber, S. Lefèvre, "Fast Quasi-Flat Zones Filtering Using Area Threshold and Region Merging", *Journal of Visual Communication and Image Representation 24*, 3, 2013, p. 397–409, https://hal.archives-ouvertes.fr/hal-00763484.

Major International Conferences

[684] E. Aptoula, J. Weber, S. Lefèvre, "Vectorial Quasi-flat Zones for Color Image Simplification", *in : International Symposium on Mathematical Morphology, ISMM 2013*, *Lecture Notes in Computer Science*, *7883*, Springer Berlin Heidelberg, p. 231–242, Uppsala, Sweden, 2013, https://hal.archives-ouvertes.fr/hal-00905178.

[685] E. H. Barney Smith, B. Lamiroy, "Effects of Clustering Algorithms on Typographic Reconstruction", *in : 13th International Conference on Document Analysis and Recognition*, Nancy, France, August 2015, https://hal.archives-ouvertes.fr/hal-01154603.

[686] A. Boumaiza, S. Tabbone, "A Novel Approach for Graphics Recognition based on Galois lattice and Bag of Words representation", *in : 10th International Conference on Document Analysis and Recognition - ICDAR 2011*, ICDAR (editor), *11th International Conference on Document Analysis and Recognition - ICDAR 2011 - Proceedings*, IEEE, p. 829–833, Pekin, China, September 2011, https://hal.inria.fr/hal-00658235.

[687] A. Boumaiza, S. Tabbone, "Impact of a Codebook Filtering Step on a Galois Lattice Structure for Graphics Recognition", *in : 21st International Conference on Pattern Recognition*, p. 4, Tsukuba International Congress Center, Japan, November 2012. 4, https://hal.archives-ouvertes.fr/hal-00759516.

[688] A. Boumaiza, S. Tabbone, "Symbol Recognition using a Galois Lattice of Frequent Graphical Patterns", *in : 10th IAPR International Workshop on Document Analysis Systems (DAS)*, *IEEE*, Gold Coast, Queensland, Australia, March 2012, https://hal.inria.fr/hal-00658255.

[689] A. Bouzaieni, S. Barrat, S. Tabbone, "Automatic annotation extension and classification of documents using a probabilistic graphical model", *in : 13th International Conference on Document Analysis and Recognition (ICDAR 2015)*, Nancy, France, August 2015, https://hal.archives-ouvertes.fr/hal-01254933.

[690] A. Bouzaieni, S. Tabbone, S. Barrat, "Automatic Images Annotation Extension Using a Probabilistic Graphical Model", *in : Computer Analysis of Images and Patterns (CAIP)- 16th International Conference*, *lecture Notes in Computer Science*, *9257*, Valetta, Malta, September 2015, https://hal.archives-ouvertes.fr/hal-01254926.

[691] F. Brandao Da Silva, S. Goldenstein, S. Tabbone, R. Da Silva Torres, "Image classification based on bag of visual graphs", *in : IEEE International Conference on Image Processing (ICIP)*, p. 4312–4316, Melbourne, Australia, September 2013, https://hal.archives-ouvertes.fr/hal-00939183.

[692] F. Brandao Da Silva, S. Tabbone, R. Da Silva Torres, "BoG: A New Approach for Graph Matching", *in : international conference on pattern recognition (ICPR 2014)*, p. 82 – 87, Stockholm, Sweden, August 2014, https://hal.archives-ouvertes.fr/hal-01098679.

[693] J. Chen, D. Lopresti, B. Lamiroy, "A Real-World Noisy Unstructured Handwritten Notebook Corpus for Document Image Analysis Research", *in : Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data - (J-MOCR-AND 2011)*, *Proceedings of the 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*, IAPR, ACM, Beijing, China, September 2011. ISBN: 978-1-4503-0685-0, https://hal.inria.fr/inria-00627844.

[694] R. Da Silva Torres, M. Hasegawa, S. Tabbone, J. Almeida, J. A. Dos Santos, B. Alberton, P. C. Morellato, "Shape-based Time Series Analysis for Remote Phenology Studies", *in : IEEE International Geoscience And Remote Sensing Symposium*, p. ;, Melbourne, Australia, July 2013, https://hal.archives-ouvertes.fr/hal-00939207.

[695] T. H. Do, S. Tabbone, O. Ramos-Terrades, "Noise suppression over bi-level graphical documents by sparse representation", *in : Colloque International Francophone sur l'Écrit et le Document - CIFED 2012*, *Actes du douzième Colloque International Francophone sur l'écrit et le Document*, Bordeaux, France, March 2012, https://hal.inria.fr/hal-00759555.

[696] T. H. Do, S. Tabbone, O. Ramos Terrades, "Text/graphic separation using a sparse representation with multi-learned dictionaries", *in : 21st International Conference on Pattern Recognition - ICPR 2012*, Tsukuba, Japan, November 2012, https://hal.inria.fr/hal-00759554.

[697] T. H. Do, S. Tabbone, O. Ramos Terrades, "Document Noise Removal using Sparse Representations over Learned Dictionary", *in : The 13th ACM Symposium on Document Engineering*, ACM, Florence, Italy, September 2013, https://hal.inria.fr/hal-00939174.

[698] T. H. Do, S. Tabbone, O. Ramos Terrades, "Spotting Symbol Using Sparsity over Learned Dictionary of Local Descriptors", *in : 11th IAPR International Workshop on Document Analysis Systems (DAS)*, Tours, France, April 2014, https://hal.archives-ouvertes.fr/hal-01098684.

[699] G. Drusch, J. M. C. Bastien, S. Paris, "Analysing eye-tracking data: From scanpaths and heatmaps to the dynamic visualisation of areas of interest", *in : International Conference on Applied Human Factors and Ergonomics*, Krakow, Poland, 2014, https://hal.archives-ouvertes.fr/hal-01223743.

[700] K. Falkenstern, N. Bonnier, F. Viénot, M. Felhi, H. Brettel, "Adaptively Selecting a Printer Color Workflow", *in : Nineteeth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, p. 205–210, San Jose, CA, United States, November 2011, https://hal.archives-ouvertes.fr/hal-00664244.

[701] M. Felhi, N. Bonnier, S. Tabbone, "A robust skew detection method based on maximum gradient difference and R-signature", *in : IEEE 18th International Conference on Image Processing*, p. 2665, Bruxelles, Belgium, September 2011, https://hal.archives-ouvertes.fr/hal-00657661.

[702] M. Felhi, N. Bonnier, S. Tabbone, "A Skeleton Based Descriptor for Detecting Text in Real Scene Images", *in : 21st International Conference on Pattern Recognition - ICPR 2012*, IEEE, p. 282–285, Tsukuba, Japan, November 2012, https://hal.inria.fr/hal-00764633.

[703] M. Felhi, S. Tabbone, M.-V. Ortiz-Segovia, "Multiscale Stroke-Based Page Segmentation Approach", *in : 11th IAPR International Workshop on Document Analysis Systems (DAS)*, Tours, France, April 2014, https://hal.archives-ouvertes.fr/hal-01098683.

[704] B. Gaüzère, H. Makoto, L. Brun, S. Tabbone, "Implicit and Explicit Graph Embedding: Comparison of both Approaches on Chemoinformatics Applications", *in : Joint IAPR International Workshop, SSPR & SPR 2012*, G. Gimel'farb (editor), *Lecture Notes in Computer Science*, *7626*, Springer, p. 510–518, Hiroshima, Japan, November 2012, https://hal.archives-ouvertes.fr/hal-00768654.

[705] A.-B. Goumeidane, A. Bouzaieni, N. Nacereddine, S. Tabbone, "Bayesian Networks-Based Defects Classes Discrimination in Weld Radiographic Images", *in : Computer Analysis of Images and Patterns (CAIP)- 16th International Conference*, *Lecture Notes in Computer Science*, *9257*, Valetta, Malta, September 2015, https://hal.archives-ouvertes.fr/hal-01254925.

[706] R. Hafiane, L. Brun, S. Tabbone, "Incremental Embedding Within a Dissimilarity-Based Framework", *in : Graph-Based Representations in Pattern Recognition - 10th International Workshop, GbRPR 2015*, Beijing, China, May 2015, https://hal.archives-ouvertes.fr/hal-01254942.

[707] M. Hasegawa, S. Tabbone, "A Shape Descriptor combining Logarithmic-scale Histogram of Radon Transform and Phase-only Correlation Function", *in : 10th International Conference on Document Analysis and Recognition - ICDAR 2011*, IEEE, p. 182–186, Pekin, China, September 2011. ISBN: 978-0-7695-4520-2, https://hal.archives-ouvertes.fr/hal-00657641.

[708] M. Hasegawa, S. Tabbone, "A Local Adaptation of the Histogram Radon Transform Descriptor: An Application to a Shoe Print Dataset", *in : Structural, Syntactic, and Statistical Pattern Recognition: Lecture Notes in Computer Science*, G. Gimel'farb (editor), *Lecture Notes in Computer Science*, *7626*, Springer, p. 675–683, December 2012, https://hal.inria.fr/hal-00764901.

[709] M. Hasegawa, S. Tabbone, "Affine Invariant Shape Matching Using Radon Transform and Dynamic Time Warping Distance", *in : 27th Symposium On Applied Computing - SAC 2012*, ACM, p. 777–781, Trento, Italy, March 2012. ISBN: 978-1-4503-0857-1, https://hal.archives-ouvertes.fr/hal-00657642.

[710] M. Hasegawa, S. Tabbone, "Affine Invariant Shape Matching using Histogram of Radon Transform and Angle Correlation Matrix", *in : International Conference on Pattern Recognition Applications and Methods (ICPRAM 2014)*, Angers, France, 2014, https://hal.archives-ouvertes.fr/hal-01098680.

[711] T. V. Hoang, E. H. Barney Smith, S. Tabbone, "Edge noise removal in bilevel graphical document images using sparse representation", *in : IEEE International Conference on Image Processing - ICIP'2011*, IEEE, p. 3549 – 3552, Brussels, Belgium, September 2011, https://hal.inria.fr/inria-00593510.

[712] T. V. Hoang, S. Tabbone, "Generic polar harmonic transforms for invariant image description", *in : IEEE International Conference on Image Processing - ICIP'2011*, Brussels, Belgium, September 2011, https://hal.inria.fr/inria-00593509.

[713] T. V. Hoang, S. Tabbone, "Generic R-transform for invariant pattern representation", *in : International Workshop on Content-Based Multimedia Indexing - CBMI'2011*, IEEE Computer Society, p. 157–162, Madrid, Spain, June 2011. ISBN: 978-1-61284-432-9, https://hal.inria.fr/inria-00593469.

[714] T. V. Hoang, S. Tabbone, "Fast Computation of Orthogonal Polar Harmonic Transforms", *in : ICPR 2012 - The 21st International Conference on Pattern Recognition*, Tsukuba Science City, Japan, November 2012, https://hal.inria.fr/hal-00734307.

[715] S. Jouili, S. Tabbone, "Towards Performance evaluation of Graph-based Representation", *in : 8th IAPR - TC-15 Workshop on Graph-based Representations in Pattern Recognition - 2011*, *LNCS - Lecture Notes in Computer Science*, *6658*, Springer-Heidelberg, p. 72–81, Munster, Germany, May 2011, https://hal.inria.fr/inria-00580312.

[716] S. K.C., B. Lamiroy, L. Wendling, "DTW for Matching Radon Features: A Pattern Recognition and Retrieval Method", *in : Advanced Concepts for Intelligent Vision Systems*, J. Blanc-Talon, R. P. Kleihorst, W. Philips, D. C. Popescu, P. Scheunders (editors), *13th International Conference on Advanced Concepts for Intelligent Vision Systems - ACIVS 2011*, *6915*, Springer, p. 249–260, Ghent, Belgium, August 2011. The original publication is available at www.springerlink.com, https://hal.inria.fr/inria-00617287.

[717] S. K.C., B. Lamiroy, L. Wendling, "Spatio-structural Symbol Description with Statistical Feature Add-on", *in : Workshop on Graphics RECognition (GREC)*, Seoul, South Korea, September 2011, https://hal.inria.fr/inria-00617296.

[718] S. K.C., L. Wendling, B. Lamiroy, "Relation Bag-of-Features for Symbol Retrieval", *in : ICDAR - International Conference on Document Analysis and Recognition - 2013*, Washington DC, United States, August 2013, https://hal.inria.fr/hal-00823960.

[719] S. K.C., "Character Recognition based on DTW-Radon", *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, p. 264–268, Beijing, China, September 2011, https://hal.inria.fr/inria-00617298.

[720] M. Kowalczyk, B. Kerautret, B. Naegel, J. Weber, "Revisiting Component Tree Based Segmentation Using Meaningful Photometric Informations", *in : International Conference on Computer Vision and Graphics (ICCVG*, *7594*, p. 475–482, Varsovie, Poland, September 2012, https://hal.archives-ouvertes.fr/hal-00761303.

[721] B. Lamiroy, T. Bouville, J. Blégean, H. Cao, S. Ghamizi, R. Houpin, M. Lloyd, "Re-Typograph Phase I: a Proof-of-Concept for Typeface Parameter Extraction from Historical Documents", *in : Document Recognition and Retrieval XXII*, B. Lamiroy, E. Ringger (editors), *Electronic Imaging*, 9402, IS&T/SPIE, SPIE, San Francisco, United States, February 2015, https://hal.inria.fr/hal-01086803.

[722] B. Lamiroy, D. Lopresti, H. Korth, J. Heflin, "How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research", *in : Document Recognition and Retrieval XVIII - DRR 2011*, C. V.-G. Gady Agam (editor), *Document Recognition and Retrieval XVIII*, *Part of the IS and T/SPIE 23rd Annual Symposium on Electronic Imaging*, *7874*, SPIE, SPIE, San Francisco, United States, January 2011. ISBN : 9780819484116, https://hal.inria.fr/inria-00537035.

[723] B. Lamiroy, D. Lopresti, T. Sun, "Document Analysis Algorithm Contributions in End-to-End Applications: Report on the ICDAR 2011 Contest", *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, Beijing, China, September 2011. ISBN: 978-1-4577-1350-7, https://hal.inria.fr/inria-00608371.

[724] B. Lamiroy, D. Lopresti, “An Open Architecture for End-to-End Document Analysis Benchmarking”, *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, p. 42–47, Beijing, China, September 2011. ISBN: 978-1-4577-1350-7, https://hal.inria.fr/inria-00598907.

[725] B. Lamiroy, D. Lopresti, “The Non-Geek’s Guide to the DAE Platform”, *in : DAS - 10th IAPR International Workshop on Document Analysis Systems*, International Association for Pattern Recognition, IEEE, p. 27–32, Gold Coast, Queensland, Australia, March 2012, https://hal.inria.fr/hal-00658281.

[726] B. Lamiroy, T. Sun, “Precision and Recall Without Ground Truth”, *in : Ninth IAPR International Workshop on Graphics RECognition - GREC 2011*, IAPR, Seoul, South Korea, September 2011, https://hal.inria.fr/inria-00617314.

[727] B. Lamiroy, “Interpretation, Evaluation and the Semantic Gap ... What if we Were on a Side-Track?”, *in : 10th IAPR International Workshop on Graphics Recognition, GREC 2013*, B. Lamiroy, J.-M. Ogier (editors), *8746*, Springer, p. 213–226, Bethlehem, PA, United States, August 2013, https://hal.inria.fr/hal-01057362.

[728] D. Lopresti, B. Lamiroy, “Document Analysis Research in the Year 2021”, *in : Twenty-fourth International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems - IEA/AIE 2011*, K. G. Mehrotra, C. K. Mohan, J. C. Oh, P. K. Varshney, M. Ali (editors), *6703*, Syracuse University, Springer, p. 264–274, Syracuse, NY, United States, June 2011. The original publication is available at www.springerlink.com, https://hal.inria.fr/inria-00570000.

[729] N. Nacereddine, S. Tabbone, D. Ziou, “Object recognition using Radon transform based RST parameter estimation”, *in : 14th International Conference on Advanced Concepts for Intelligent Vision Systems - ACIVS 2012, Lecture notes in computer science, 7517*, p. 515–526, Brno, Czech Republic, 2012, https://hal.archives-ouvertes.fr/hal-00764767.

[730] S. Paris, “A proposal project for a blind image quality assessment by learning distortions from the full reference image quality assessments”, *in : International Workshop on Quality of Multimedia Experience*, Melbourne, Australia, 2012, https://hal.archives-ouvertes.fr/hal-01223739.

[731] E. Valveny, M. Delalandre, R. Raveaux, B. Lamiroy, “Report on the Symbol Recognition and Spotting Contest”, *in : 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16, 2011, Revised Selected Papers, Graphics Recognition. New Trends and Challenges, 7423*, Springer, p. 198–207, Seoul, North Korea, September 2011, https://hal.archives-ouvertes.fr/hal-01178076.

[732] J. Weber, F. Petitjean, P. Gancarski, “Towards efficient satellite time series analysis: combination of Dynamic Time Warping and Quasi-Flat Zones”, *in : IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, p. 4387–4390, Munich, Germany, July 2012, https://hal.archives-ouvertes.fr/hal-00761291.

[733] J. Weber, S. Tabbone, “Symbol spotting for technical documents : An efficient template-Matching approach”, *in : International Conference on Pattern Recognition (ICPR)*, p. 12, Tsukuba, Japan, November 2012, https://hal.archives-ouvertes.fr/hal-00761324.

[734] D. Ziou, D. Rizo-Rodriguez, A. Tabbone, N. Nacereddine, “Head Pose Classification Using a Bidimensional Correlation Filter”, *in : Image Analysis and Recognition - 12th International Conference (ICIAR)*, Springer (editor), *Lecture Notes in Computer Science, 9164*, Niagara Falls, Canada, July 2015, https://hal.archives-ouvertes.fr/hal-01254945.

National Peer-Reviewed Conferences

[735] A. Boumaiza, S. Tabbone, “Classification de symboles avec un treillis de Galois et une représentation par sac de mots”, *in : ORASIS - Congrès des jeunes chercheurs en vision par ordinateur*, INRIA Grenoble Rhône-Alpes, Praz-sur-Arly, France, June 2011, https://hal.inria.fr/inria-00595291.

[736] A. Boumaiza, S. Tabbone, “Vers l’utilisation d’un Treillis de Galois Fréquent pour la Reconnaissance de Symboles Graphiques”, *in : Colloque International Francophone sur l’Écrit et le Document 2012 (CIFED)*, ARIA (Association francophone de Recherche d’Information et Applications) et le GRCE (Groupement de Recherche en Communication écrite), BORDEAUX, Philippines, March 2012, https://hal.inria.fr/hal-00662559.

[737] H. Chouaib, F. Cloppet, S. Tabbone, N. Vincent, “Combinaison de classificateurs simples pour une sélection rapide de caractéristiques”, *in : 12e Conférence Internationale Francophone sur l’Extraction et la Gestion des Connaissances*, Bordeaux, France, 2012, https://hal.archives-ouvertes.fr/hal-00766545.

[738] M. Felhi, S. Tabbone, N. Bonnier, “Un nouveau descripteur de texte pour la détection des lignes de texte multi-orientées dans les scènes réelles”, *in : 7ème Colloque International Francophone sur l’Écrit et le Document - CIFED 2012*, Bordeaux, France, March 2012, https://hal.inria.fr/hal-00764708.

[739] M. Felhi, S. Tabbone, M.-V. Ortiz-Segovia, “Approche hybride de segmentation de page à base d’un descripteur de traits”, *in : Colloque International Francophone sur l’Écrit et le Document (CIFED)*, Nancy, France, March 2014, https://hal.archives-ouvertes.fr/hal-01098686.

[740] R. Hafiane, M. Smaïl-Tabbone, M.-D. Devignes, S. Tabbone, “Clustering optimal de gènes fondé sur une mesure de similarité sémantique”, *in : 10ème édition de la COnférence en Recherche d’Information et Applications - CORIA 2013*, C. Berrut (editor), p. 15 p, Neufchâtel, Switzerland, April 2013, https://hal.inria.fr/hal-00920700.

[741] R. Hafiane, S. Tabbone, L. Brun, “Plongement incrémental dans un contexte de dissimilarité”, *in : Colloque International Francophone sur l’Écrit et le Document (CIFED)*, Nancy, France, March 2014, https://hal.archives-ouvertes.fr/hal-01098685.

[742] B. Lamiroy, S. Zheng, “An Attempt to Use Ontologies for Document Image Analysis”, *in : RISE 2014 6ème Atelier Recherche d’Information SEmantique*, ARIA, Nancy, France, March 2014, https://hal.inria.fr/hal-00959487.

[743] H. Locteau, “Décomposition en Polygones de forme Étoile - Application à la Détection de Pièces”, *in : CIFED - Colloque International Francophone sur l’Écrit et le Document - 2012*, J.-M. Ogier, T. Paquet, J.-Y. Ramel, A. Tabbone, N. Vinvent (editors), Bordeaux, France, March 2012, https://hal.inria.fr/hal-00762857.

[744] S. Paris, “Étude expérimentale d’une nouvelle mesure de qualité d’image fondée sur les propriétés de l’analyse en ondelettes”, *in : Ateliers de Traitement et Analyse de l’Information - Méthodes et Applications*, Hammamet, Tunisia, 2011, https://hal.archives-ouvertes.fr/hal-01223730.

Books

[745] D. Doermann, K. Tombre, *Handbook of Document Image Processing and Recognition*, Springer, 2014, https://hal.inria.fr/hal-00995696.

Books or Proceedings Editing

[746] M. De Marsico, A. Tabbone, A. L. N. Fred (editors), *ICPRAM 2014 - Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*, Angers, France, SciTePress, 2014, https://hal.archives-ouvertes.fr/hal-01254481.

[747] A. L. N. Fred, M. De Marsico, A. Tabbone (editors), *Pattern Recognition Applications and Methods - Third International Conference, ICPRAM 2014, Angers, France, March 6-8, 2014, Revised Selected Papers*, *Lecture Notes in Computer Science*, *9443*, France, Springer, 2015, https://hal.archives-ouvertes.fr/hal-01254470.

[748] B. Lamiroy, J.-M. Ogier (editors), *Graphics Recognition. Current Trends and Challenges*, *Lecture Notes in Computer Science*, *8746*, Bethlehem, PA, United States, Springer, November 2014, 267p., https://hal.inria.fr/hal-01055238.

[749] B. Lamiroy, E. Ringger (editors), *Document Recognition and Retrieval XXII*, *PROCEEDINGS OF SPIE*, *4902*, IS&T/SPIE, San Francisco, United States, SPIE, February 2015, 254p., https://hal.archives-ouvertes.fr/hal-01117730.

Book chapters

[750] S. K.C., B. Lamiroy, L. Wendling, "Spatio-structural Symbol Description with Statistical Feature Add-on", *in : Graphics Recognition. New Trends and Challenges*, J.-M. O. Young-Bin Kwon (editor), *Lecture Notes in Computer Science*, *7423*, Springer-Verlag Berlin Heidelberg, February 2013, p. 228–237, The original publication is available at www.springerlink.com, https://hal.inria.fr/hal-00789268.

[751] B. Lamiroy, Y. Guebbas, "Robust and Precise Circular Arc Detection", *in : Graphics Recognition. Achievements*, *Challenges*, *and Evolution*, J. L. Jean-Marc Ogier, Wenyin Liu (editor), *6020*, Springer-Verlag, 2010, p. 49–60, The original publication is available at www.springerlink.com, https://hal.inria.fr/inria-00516712.

[752] B. Lamiroy, J.-M. Ogier, "Analysis and Interpretation of Graphical Documents", *in : Handbook of Document Image Processing and Recognition*, D. Doermann and K. Tombre (editors), Springer, May 2014, https://hal.inria.fr/hal-00943058.

[753] B. Lamiroy, T. Sun, "Computing Precision and Recall with Missing or Uncertain Ground Truth", *in : Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011*, *Seoul, Korea, September 15-16, 2011, Revised Selected Papers*, Y.-B. Kwon and J.-M. Ogier (editors), *Lecture Notes in Computer Science*, *7423*, Springer, February 2013, p. 149–162, https://hal.inria.fr/hal-00778188.

[754] S. Lefèvre, E. Aptoula, B. Perret, J. Weber, "Morphological Template Matching in Color Images", *in : Advances in Low-Level Color Image Processing*, *Lecture Notes in Computational Vision and Biomechanics*, *11*, Springer Netherlands, 2014, p. 241–277, https://hal.archives-ouvertes.fr/hal-01111954.

[755] S. Tabbone, O. Ramos-Terrades, "An Overview of Symbol Recognition", *in : Handbook of Document Image Processing and Recognition*, K. T. David Doermann (editor), Springer London, 2014, https://hal.archives-ouvertes.fr/hal-01098681.

[756] E. Valveny, M. Delalandre, R. Raveaux, B. Lamiroy, "Report on the Symbol Recognition and Spotting Contest", *in : Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16, 2011, Revised Selected Papers*, Y.-B. Kwon and J.-M. Ogier (editors), *Lecture Notes in Computer Science, 7423*, Springer, February 2013, p. 198–207, https://hal.inria.fr/hal-00788250.

5 References for Read

Articles in International Peer-Reviewed Journal

[757] M.-R. Bouguelia, Y. Belaïd, B. Abdel, "An adaptive streaming active learning strategy based on instance weighting", *Pattern Recognition Letters*, 70, November 2015, p. 38–44, https://hal.inria.fr/hal-01254510.

[758] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "Classification de flux de documents évolutifs avec apprentissage de classes inconnues", *Revue des Sciences et Technologies de l'Information - Série Document Numérique 17*, 3, 2014, p. 21, https://hal.inria.fr/hal-01131453.

[759] M. Imran Razzak, A. Syed, A. A. Mirza, B. Abdel, "FUZZY BASED PREPROCESSING USING FUSION OF ONLINE AND OFFLINE TRAIT FOR ONLINE URDU SCRIPT BASEDLANGUAGES CHARACTER RECOGNITION", *International Journal of Innovative Computing, Information and Control 8*, 5, May 2012, p. 21, https://hal.inria.fr/hal-01254627.

[760] A. Kacem, K. Akram, B. Abdel, "A PGM-based System for Arabic Handwritten Word Recognition", *Electronic Letters on Computer Vision and Image Analysis 13*, 3, September 2014, p. 22, https://hal.inria.fr/hal-01254553.

[761] A. Kacem, I. Ben Cheikh, B. Abdel, "Collaborative combination of neuron-linguistic classifiers for large arabic word vocabulary recognition", *International Journal of Pattern Recognition and Artificial Intelligence 28*, 1, January 2014, p. 39, https://hal.inria.fr/hal-01254577.

[762] N. Ouwayed, A. Belaïd, "A general approach for multi-oriented text line extraction of handwritten document", *International Journal on Document Analysis and Recognition 14*, 4, September 2011, https://hal.inria.fr/inria-00635363.

[763] E. Philippot, S. K.C., A. Belaïd, Y. Belaïd, "Bayesian networks for incomplete data analysis in form processing", *International journal of machine learning and cybernetics*, February 2014, p. 25, https://hal.inria.fr/hal-01099727.

[764] Y. Rangoni, A. Belaïd, S. Vajda, "Labelling logical structures of document images using a dynamic perceptive neural network", *International Journal on Document Analysis and Recognition 15*, 1, 2011, p. 45–55, The original publication is available at www.springerlink.com, https://hal.inria.fr/inria-00579825.

[765] A. Saïdani, A. Kacem, B. Abdel, "Arabic/Latin and Machine-printed/Handwritten Word Discrimination using HOG-based Shape Descriptor", *Electronic Letters on Computer Vision and Image Analysis 2*, 14, August 2015, p. 24, https://hal.inria.fr/hal-01254539.

[766] A. Saïdani, A. Kacem, B. Abdel, "Arabic/Latin and Machine-printed/Handwritten Word Discrimination using HOG-based Shape Descriptor", *Electronic Letters on Computer Vision and Image Analysis 14*, 2, August 2015, p. 23, https://hal.inria.fr/hal-01254563.

Major International Conferences

[767] K. Akram, A. Kacem, B. Abdel, M. Elloumi, "Arabic handwritten words off-line recognition based on HMMs and DBNs", *in : 13th International Conference on Document Analysis and Recognition (ICDAR)* , Nancy, France, August 2015, https://hal.inria.fr/hal-01254724.

[768] N. Aouadi, A. Kacem, B. Abdel, "A Recognition based Approach for Segmenting Touching Components in Arabic Manuscripts", *in : International Conference on Document Analysis and Recognition ICDAR*, Nancy, France, August 2015, https://hal.inria.fr/hal-01254711.

[769] N. Aouadi, A. Kacem, A. Belaïd, "Segmentation of Touching Component in Arabic Manuscripts", *in : International Conference on Frontiers in Handwriting Recognition*, Crète, Greece, September 2014, https://hal.inria.fr/hal-01111750.

[770] A.-M. A. Awal, A. Belaïd, V. Poulain d'Andecy, "Handwritten/printed text separation Using pseudo- lines for contextual re-labeling", *in : International Conference on Frontiers on Handwriting Recognition*, Crète, Greece, September 2014, https://hal.inria.fr/hal-01111715.

[771] A. Belaïd, S. K.C., V. Poulain D'Andecy, "Handwritten and Printed Text Separation in Real Document", *in : The Thirteenth IAPR International Conference on Machine Vision Applications - 2013*, Kyoto, Japan, May 2013, https://hal.inria.fr/hal-00799331.

[772] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "A Stream-Based Semi-Supervised Active Learning Approach for Document Classification", *in : 12th International Conference on Document Analysis and Recognition - ICDAR 2013*, IEEE, p. 611–615, Washington, United States, August 2013, https://hal.inria.fr/hal-00855184.

[773] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "An Adaptive Incremental Clustering Method Based on the Growing Neural Gas Algorithm", *in : 2nd International Conference on Pattern Recognition Applications and Methods - ICPRAM 2013*, SciTePress, p. 42–49, Barcelona, Spain, February 2013, https://hal.inria.fr/hal-00794354.

[774] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "DOCUMENT IMAGE AND ZONE CLASSIFICATION THROUGH INCREMENTAL LEARNING", *in : International Conference on Image Processing (ICIP)*, IEEE, p. 4230–4234, Melbourne, Australia, September 2013, https://hal.inria.fr/hal-00865765.

[775] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "Classification active de flux de documents avec identification des nouvelles classes", *in : CIFED - Colloque International Francophone sur l'Écrit et le Document*, p. 75–89, Nancy, France, March 2014, https://hal.inria.fr/hal-00980698.

[776] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "Efficient Active Novel Class Detection for Data Stream Classification", *in : ICPR - International Conference on Pattern Recognition*, IEEE, p. 2826–2831, Stockholm, Sweden, August 2014, https://hal.inria.fr/hal-01065043.

[777] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "Stream-based Active Learning in the Presence of Label Noise", *in : 4th International Conference on Pattern Recognition Applications and Methods - ICPRAM 2015*, SciTePress, p. 25 – 34, Lisbon, Portugal, January 2015, https://hal.inria.fr/hal-01116114.

[778] H. Daher, B. Abdel, "Document Flow Segmentation for Business Applications", *in : Document Recognition and Retrieval XXI*, p. 9201–15, San Francisco, France, February 2014, https://hal.archives-ouvertes.fr/hal-00926615.

[779] H. Daher, A. Belaïd, V. Poulain d'Andecy, "Segmentation de flux de documents Application aux documents administratifs", *in : Conférence Internationale Francophone sur l'Ecrit et le Document*, Nancy, France, March 2014, https://hal.inria.fr/hal-01111746.

[780] H. Daher, M.-R. Bouguelia, B. Abdel, V. Poulain D'Andecy, "Multipage Administrative Document Stream Segmentation", *in : International Conference on Pattern Recognition*, Stokholm, Sweden, August 2014, https://hal.inria.fr/hal-01254785.

[781] N. Ghanmi, B. Abdel, "Extraction de formules chimiques dans des documents manuscrits composites", *in : Colloque International Francophone sur l'Écrit et le Document*, CIFED (editor), p. 185–197, Nancy, France, March 2014, https://hal.archives-ouvertes.fr/hal-01070831.

[782] N. Ghanmi, B. Abdel, "Table detection in handwritten chemistry documents using conditional random fields", *in : ICFHR*, p. p. 146–151, Crete, Greece, September 2014, https://hal.archives-ouvertes.fr/hal-01070743.

[783] N. Ghanmi, B. Abdel, "Separator and content based approach for table extraction in handwritten chemistry documents", *in : International Conference on Document Analysis and Recognition*, Nancy, France, August 2015, https://hal.archives-ouvertes.fr/hal-01255383.

[784] D. Grzejszczak, Y. Rangoni, A. Belaïd, "Séparation manuscrit et imprimé dans des documents administratifs complexes par utilisation de SVM et regroupement", *in : CIFED-CORIA*, Bordeaux, France, March 2012, https://hal.inria.fr/hal-00779237.

[785] A. Kacem, N. Aouïti, A. Belaïd, "Structural Features Extraction for Handwritten Arabic Personal Names Recognition", *in : ICFHR - 13th International Conference on Frontiers in Handwriting Recognition - 2012*, IEEE, p. 268–273, Bari, Italy, September 2012, https://hal.inria.fr/hal-00779260.

[786] A. Kacem, A. Saïdani, A. Belaïd, "A System for an automatic reading of student information sheets", *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, p. 1265–1269, Bejing, China, September 2011. ISBN: 978-1-4577-1350-7, https://hal.inria.fr/inria-00635376.

[787] K. Kaouther, A. Kacem, A. Belaïd, "Recognition of Machine-Printed Arabic Mathematical Formulas", *in : Information and Communication Technologies Innovations and Applications (ICTIA)*, Sousse, Tunisia, March 2014, https://hal.archives-ouvertes.fr/hal-01112674.

[788] T. Kasar, T. K. Bhowmik, B. Abdel, "Table information extraction and structure recognition using query patterns", *in : International Conference on Document Analysis and Recognition ICDAR*, Nancy, France, August 2015, https://hal.inria.fr/hal-01254761.

[789] S. K.C., A. Belaïd, "Client-Driven Content Extraction Associated with Table", *in : IAPR MVA - The Thirteenth IAPR International Conference on Machine Vision Applications - 2013*, Kyoto, Japan, May 2013, https://hal.inria.fr/hal-00808678.

[790] S. K.C., A. Belaïd, "Document Information Extraction and its Evaluation based on Client's Relevance", *in : ICDAR - International Conference on Document Analysis and Recognition - 2013*, IEEE, Washington DC, United States, August 2013, https://hal.inria.fr/hal-00822479.

[791] S. K.C., A. Belaïd, "Pattern-Based Approach to Table Extraction", *in : IbPRIA 2013: 6th Iberian Conference on Pattern Recognition and Image Analysis*, J. S. C. João M. Sanches, Luisa Micó (editor), *Pattern Recognition and Image Analysis*, Springer, Madeira, Portugal, June 2013, https://hal.inria.fr/hal-00788323.

[792] K. Khazri, A. Kacem, B. Abdel, "A syntax directed system for the recognition of printed Arabic mathematical formulas", *in : International Conference on Document Analysis and Recognition (ICDAR)*, Nancy, France, August 2015, https://hal.inria.fr/hal-01254736.

[793] A. Khémiri, A. Kacem, A. Belaïd, "Towards Arabic Handwritten Word Recognition via Probabilistic Graphical Models", *in : International Conference on Frontiers in Handwriting Recognition*, Crete, Greece, September 2014, https://hal.inria.fr/hal-01111751.

[794] N. Kooli, A. Belaïd, "Entity Matching in OCRed Documents with Redundant Databases", *in : International Conference on Pattern Recognition Applications and Methods (ICPRAM-2015)*, Lisbon, Portugal, January 2015, https://hal.archives-ouvertes.fr/hal-01144641.

[795] N. Kooli, A. Belaïd, "Semantic Label and Structure Model based Approach for Entity Recognition in Database Context", *in : 13th International Conference on Document Analysis and Recognition (ICDAR)*, p. 5, Nancy, France, August 2015, https://hal.archives-ouvertes.fr/hal-01191425.

[796] M. Parodi, J. C. Gomez, A. Belaïd, "A Circular Grid-Based Rotation Invariant Feature Extraction Approach for Off-line Signature Verification", *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, p. 1289–1293, Bejing, China, October 2011. ISBN: 978-1-4577-1350-7, https://hal.inria.fr/inria-00635385.

[797] E. Philippot, A. Belaïd, Y. Belaïd, "Use of PGM for Form recognition", *in : 10th IAPR International Workshop on Document Analysis Systems - DAS 2012*, IEEE, p. 374–378, Gold Coast, Australia, March 2012. ISBN : 978-1-4673-0868-7, https://hal.inria.fr/hal-00697794.

[798] E. Philippot, Y. Belaïd, A. Belaïd, "Use of semantic and physical constraints in Bayesian Networks for Form Recognition", *in : 11th International Conference on Document Analysis and Recognition - ICDAR 2011*, International Association for Pattern Recognition, IEEE Computer Society, p. 946–950, Beijing, China, September 2011. ISBN: 978-1-4577-1350-7, https://hal.inria.fr/inria-00629334.

[799] Y. Rangoni, A. Belaid, "Segmentation de documents composites par une technique de recouvrement des espaces blancs", *in : CIFED-CORIA*, Bordeaux, France, March 2012, https://hal.inria.fr/hal-00779235.

[800] A. Saïdani, A. Kacem, B. Abdel, "Co-occurrence Matrix of Oriented Gradients for Word Script and Nature Identification", *in : International Conference on Document Analysis and Recognition ICDAR*, Nancy, France, August 2015, https://hal.inria.fr/hal-01254691.

[801] A. Saïdani, A. Kacem Echi, A. Belaïd, "Proposition to distinguish Machine-Printed from Handwritten Arabic and Latin Words", *in : Information and Communication Technologies Innovations and Applications (ICTIA)*, Information and Communication Technologies Innovations and Applications (ICTIA), Sousse, Tunisia, March 2014, https://hal.archives-ouvertes.fr/hal-01112678.

[802] J.-M. Vauthier, A. Belaïd, "Segmentation et classification des zones d'une page de document", *in : CIFED-CORIA*, Bordeaux, France, March 2012, https://hal.inria.fr/hal-00779232.

Book chapters

[803] A. Belaïd, N. Ouwayed, "Segmentation of ancient Arabic documents", *in : Guide to OCR for Arabic Scripts*, V. Märgner and H. E. Abed (editors), Springer, March 2011, https://hal.inria.fr/inria-00579840.

[804] A. Belaïd, V. Poulain D'Andecy, H. Hamza, Y. Belaïd, "Administrative Document Analysis and Structure", *in : Learning Structure and Schemas from Documents*, M. Biba and F. Xhafa (editors), *Studies in Computational Intelligence*, *375*, Springer Verlag, March 2011, p. 51–72, https://hal.inria.fr/inria-00579833.

[805] A. Belaid, M. I. Razzak, "Middle Eastern Character Recognition", *in : Handbook of Document Image Processing and Recognition*, *13*, University of Maryland, Université de Lorraine, April 2014, p. 31, https://hal.inria.fr/hal-01254510.

[806] M.-R. Bouguelia, Y. Belaïd, A. Belaïd, "Online Unsupervised Neural-Gas Learning Method for Infinite Data Streams", *in : Pattern Recognition Applications and Methods*, *Advances in Intelligent Systems and Computing*, *318*, springer, 2015, p. 57 – 70, https://hal.inria.fr/hal-01116082.

6 References for Sémagramme

Doctoral Dissertations

[807] E. Lebedeva, *Expressing discourse dynamics through continuations*, Theses, Université de Lorraine, 2012, https://hal.inria.fr/tel-00783245.

[808] M. Morey, *Symbolic supertagging and syntax-semantics interface of polarized lexicalized grammatical formalisms*, Theses, Université de Lorraine, 2011, https://hal.inria.fr/tel-00640561.

[809] F. Pompigne, *Logical modelization of language and Abstract Categorial Grammars*, Theses, Université de Lorraine, 2013, https://hal.inria.fr/tel-00921040.

[810] S. Qian, *Accessibility of Referents in Discourse Semantics*, Theses, Université de Lorraine, 2014, https://hal.inria.fr/tel-01104091.

Articles in International Peer-Reviewed Journal

[811] M. Amblard, K. Fort, C. Demily, N. Franck, M. Musiol, "Analyse lexicale outillée de la parole transcrite de patients schizophrènes", *Traitement Automatique des Langues 55*, 3, 2015, p. 91 – 115, https://hal.inria.fr/hal-01188677.

[812] M. Amblard, C. Retoré, "Partially Commutative Linear Logic and Lambek Caculus with Product: Natural Deduction, Normalisation, Subformula Property", *the Journal of Logics and their Applications 1*, 1, 2014, p. 53–94, https://hal.inria.fr/hal-01071642.

[813] M. Amblard, "La non-commutativité comme argument linguistique : modéliser la notion de phase dans un cadre logique", *Traitement Automatique des Langues 56*, 1, 2015, p. 91 – 115, https://hal.inria.fr/hal-01188669.

[814] P. de Groote, M. Kanazawa, "A Note on Intensionalization", *Journal of Logic, Language and Information 22*, 2, 2013, p. 173–194, https://hal.inria.fr/hal-00909207.

[815] P. de Groote, S. Pogodalla, C. Pollard, "About Parallel and Syntactocentric Formalisms: A Perspective from the Encoding of Convergent Grammar into Abstract Categorial Grammar", *Fundamenta Informaticae 106*, 2-4, 2011, p. 211–231, https://hal.inria.fr/inria-00565598.

[816] G. Perrier, B. Guillaume, "Frigram: a French Interaction Grammar", *Journal of Language Modelling 3*, 1, 2015, p. 265–316, https://hal.inria.fr/hal-01188687.

[817] M. Rebuschi, M. Amblard, M. Musiol, "Schizophrénie, logicité et compréhension en première personne", *L'Évolution Psychiatrique 78*, 1, 2013, p. 127–141, https://hal.inria.fr/hal-00869681.

Invited Conferences

[818] M. Amblard, M. Musiol, M. Rebuschi, "Folie, conversation et contradictions", *in : Le langage comme Action / l'action par le langage*, Grenoble, France, 2011, https://hal.inria.fr/halshs-01231986.

[819] F. Lamarche, "A new homotopy-theoretic interpretation of Martin-Löf's identity type.", *in : Models, Logics and Higher-Dimensional Categories: A Tribute to the Work of Mihaly Makkai, CRM Proceedings & Lecture Notes, 53*, Bradd Hart and Thomas G. Kucera and Anand Pillay and Philip J. Scott and Robert A. G. Seely, Montréal, Canada, 2011, https://hal.inria.fr/inria-00625956.

Major International Conferences

[820] M. Amblard, K. Fort, "Étude quantitative des disfluences dans le discours de schizophrènes : automatiser pour limiter les biais", *in : TALN - Traitement Automatique des Langues Naturelles*, p. 292–303, Marseille, France, 2014, https://hal.inria.fr/hal-01054391.

[821] M. Amblard, M. Musiol, M. Rebuschi, "Une analyse basée sur la S-DRT pour la modélisation de dialogues pathologiques", *in : Actes de la 18e conférence sur le Traitement Automatique des Langues Naturelles - TALN 2011*, M. Lafourcade, V. Prince (editors), Laboratoire d'Informatique de Robotique et de Microélectronique, p. 93 – 98, Montpellier, France, 2011, https://hal.inria.fr/hal-00601622.

[822] M. Amblard, "Encoding Phases using Commutativity and Non-commutativity in a Logical Framework", *in : 6th International Conference on Logical Aspect of Computational Linguistic - LACL 2011*, S. Pogodalla, J.-P. Prost (editors), *Lecture Notes in Artificial Intelligence, 6736*, Springer, p. 1–16, Montpellier, France, 2011. ISBN : 978-3-642-22220-7, https://hal.inria.fr/hal-00601621.

[823] N. Asher, S. Pogodalla, "SDRT and Continuation Semantics", *in : New Frontiers in Artificial Intelligence JSAI-isAI 2010 Workshops, LENLS, JURISIN, AMBN, ISS, Tokyo, Japan, November 18-19, 2010, Revised Selected Papers*, T. Onada, D. Bekki, E. McCready (editors), *Lecture Notes in Computer Science, 6797*, Springer Berlin Heidelberg, p. 3–15, 2011, https://hal.inria.fr/inria-00565744.

[824] C. Baskent, "Public Announcements, Topology and Paraconsistency", *in : Informal Proceedings of LOFT Conference 2014*, Bregen, Norway, 2014, https://hal.inria.fr/hal-01094782.

[825] C. Blom, P. de Groote, Y. Winter, J. Zwarts, "Implicit Arguments: Event Modification or Option Type Categories?", *in : 18th Amsterdam Colloquium on Logic, Language and Meaning*, M. Aloni,

V. Kimmelman, F. Roelofsen, G. W. Sassoon, K. Schulz, M. Westera (editors), *7218*, Springer, p. 240–250, Amsterdam, Netherlands, 2011, https://hal.inria.fr/hal-00763102.

[826] G. Bonfante, B. Guillaume, M. Morey, G. Perrier, "Enrichissement de structures en dépendances par réécriture de graphes", *in : Traitement Automatique des Langues Naturelles (TALN)*, Montpellier, France, 2011, https://hal.inria.fr/inria-00579251.

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

[828] G. Bonfante, B. Guillaume, "Non-simplifying Graph Rewriting Termination", *in : TERMGRAPH*, R. Echahed, D. Plump (editors), *7th International Workshop on Computing with Terms and Graphs*, p. 4–16, Rome, Italy, 2013, https://hal.inria.fr/hal-00921053.

[829] M. Candito, G. Perrier, B. Guillaume, C. Ribeyre, K. Fort, D. Seddah, E. Villemonte De La Clergerie, "Deep Syntax Annotation of the Sequoia French Treebank", *in : International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, 2014, https://hal.inria.fr/hal-00969191.

[830] A. Couillault, K. Fort, G. Adda, H. De Mazancourt, "Evaluating Corpora Documentation with regards to the Ethics and Big Data Charter", *in : International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, 2014, https://hal.inria.fr/hal-00969180.

[831] L. Danlos, P. de Groote, S. Pogodalla, "A Type-Theoretic Account of Neg-Raising Predicates in Tree Adjoining Grammars", *in : New Frontiers in Artificial Intelligence. JSAI-isAI 2013 Workshops, LENLS, JURISIN, MiMI, AAA, and DDS, Kanagawa, Japan, October 27-28, 2013, Revised Selected Papers*, Y. Nakano, K. Satoh, D. Bekki (editors), *Lecture Notes in Computer Science, 8417*, Springer International Publishing, p. 3–16, 2014, https://hal.inria.fr/hal-00868382.

[832] L. Danlos, A. Maskharashvili, S. Pogodalla, "An ACG Analysis of the G-TAG Generation Process", *in : INLG 2014 - 8th International Natural Language Generation Conference*, M. Mitchell, K. McCoy, D. McDonald, A. Cahill (editors), *Proceedings of the 8th International Natural Language Generation Conference (INLG)*, Association for Computational Linguistics, p. 35–44, Philadelphia, PA, United States, 2014, https://hal.inria.fr/hal-00999595.

[833] L. Danlos, A. Maskharashvili, S. Pogodalla, "An ACG View on G-TAG and Its g-Derivation", *in : Logical Aspects of Computational Linguistics. 8th International Conference, LACL 2014*, N. Asher, S. Soloviev (editors), *Lecture Notes in Computer Science, 8535*, Springer, p. 70–82, Toulouse, France, 2014, https://hal.inria.fr/hal-00999633.

[834] L. Danlos, A. Maskharashvili, S. Pogodalla, "Génération de textes : G-TAG revisité avec les Grammaires Catégorielles Abstraites", *in : TALN 2014 - 21ème conférence sur le Traitement Automatique des Langues Naturelles, Actes de TALN 2014, 1*, Association pour le Traitement Automatique des Langues, p. 161–172, Marseille, France, 2014, https://hal.inria.fr/hal-00999589.

[835] L. Danlos, A. Maskharashvili, S. Pogodalla, "Grammaires phrastiques et discursives fondées sur les TAG : une approche de D-STAG avec les ACG", *in : TALN 2015 - 22e conférence sur le Traitement Automatique des Langues Naturelles, Actes de TALN 2015*, Association pour le Traitement Automatique des Langues, p. 158–169, Caen, France, 2015, https://hal.inria.fr/hal-01145994.

[836] P. de Groote, Y. Winter, "A type-logical account of quantification in event semantics", *in : Logic and Engineering of Natural Language Semantics 11*, Tokyo, Japan, 2014, https://hal.inria.fr/hal-01102261.

[837] P. de Groote, "On Logical Relations and Conservativity", *in : NLCS'15. Third Workshop on Natural Language and Computer Science*, M. Kanazawa, L. S. Moss, V. de Paiva (editors), *EasyChair Proceedings in Computing*, *32*, p. 1–11, Kyoto, Japan, 2015, https://hal.inria.fr/hal-01188619.

[838] P. de Groote, "Proof-Theoretic Aspects of the Lambek-Grishin Calculus", *in : Logic, Language, Information, and Computation - 22nd International Workshop, WoLLIC 2015*, V. de Paiva, R. J. G. B. de Queiroz, L. S. Moss, D. Leivant, A. G. de Oliveira (editors), *Lecture Notes in Computer Science*, *9160*, p. 109–123, Bloomington, United States, 2015, https://hal.inria.fr/hal-01188612.

[839] V. Dufour-Lussier, B. Guillaume, G. Perrier, "Parsing Coordination Extragrammatically", *in : Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers*, Z. Vetulani, J. Mariani (editors), *Lecture Notes in Computer Science (LNCS) series / Lecture Notes in Artificial Intelligence (LNAI) subseries*, *8387*, Springer International Publishing, p. 12, 2014, https://hal.inria.fr/hal-00921033.

[840] K. Fort, G. Adda, B. Sagot, J. Mariani, A. Couillault, "Crowdsourcing for Language Resource Development: Criticisms About Amazon Mechanical Turk Overpowering Use", *in : Human Language Technology Challenges for Computer Science and Linguistics*, Z. Vetulani, J. Mariani (editors), *Lecture Notes in Computer Science*, *8387*, Springer International Publishing, p. 303–314, 2014, https://hal.inria.fr/hal-01053047.

[841] K. Fort, A. Nazarenko, S. Rosset, "Modeling the Complexity of Manual Annotation Tasks: a Grid of Analysis", *in : International Conference on Computational Linguistics*, p. 895–910, Mumbaï, India, 2012, https://hal.inria.fr/hal-00769631.

[842] C. Gardent, Y. Parmentier, G. Perrier, S. Schmitz, "Lexical Disambiguation in LTAG using Left Context", *in : Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers*, Z. Vetulani, J. Mariani (editors), *Lecture Notes in Computer Science (LNCS) series / Lecture Notes in Artificial Intelligence (LNAI) subseries*, *8387*, Springer, p. 67–79, 2014, https://hal.inria.fr/hal-00921246.

[843] B. Guillaume, G. Bonfante, P. Masson, M. Morey, G. Perrier, "Grew : un outil de réécriture de graphes pour le TAL", *in : 12ième Conférence annuelle sur le Traitement Automatique des Langues (TALN'12)*, G. S. Georges Antoniadis, Hervé Blanchon (editor), ATALA, p. 1–2, Grenoble, France, 2012, https://hal.inria.fr/hal-00760637.

[844] B. Guillaume, K. Fort, G. Perrier, P. Bedaride, "Mapping the Lexique des Verbes du Français (Lexicon of French Verbs) to a NLP Lexicon using Examples", *in : International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, 2014, https://hal.inria.fr/hal-00969184.

[845] B. Guillaume, K. Fort, "Expériences de formalisation d'un guide d'annotation : vers l'annotation agile assistée", *in : 20e conférence sur le Traitement Automatique des Langues Naturelles*, p. 628–635, Les Sables d'Olonne, France, 2013, https://hal.inria.fr/hal-00840895.

[846] B. Guillaume, G. Perrier, "Dependency Parsing with Graph Rewriting", *in : IWPT 2015, 14th International Conference on Parsing Technologies, 14th International Conference on Parsing Technologies - Proceedings of the Conference*, Bilbao, Spain, 2015, https://hal.inria.fr/hal-01188694.

[847] B. Guillaume, "Online Graph Matching", *in : Traitement Automatique des Langues Naturelles (TALN), Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2015), Caen (France)*, p. 648–649, Caen, France, 2015, https://hal.inria.fr/hal-01188682.

[848] L. Kallmeyer, R. Osswald, S. Pogodalla, "Progression and Iteration in Event Semantics - An LTAG Analysis Using Hybrid Logic and Frame Semantics", *in : The 11th Syntax and Semantics Conference in Paris (CSSP 2015)*, Paris, France, 2015, https://hal.inria.fr/hal-01184872.

[849] M. Lafourcade, K. Fort, "Propa-L: a Semantic Filtering Service from a Lexical Network Created using Games With A Purpose", *in : International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, 2014, https://hal.inria.fr/hal-00969161.

[850] F. Lamarche, N. Novakovic, "Frobenius Algebras and Classical Proof Nets", *in : Fifth International Conference on Topology, Algebra and Categories in Logic - TACL 2011*, Achim Jung and Yde Venema, Marseille, France, 2011, https://hal.inria.fr/inria-00620126.

[851] A. Maskharashvili, S. Pogodalla, "Constituency and Dependency Relationship from a Tree Adjoining Grammar and Abstract Categorial Grammar Perspective", *in : 6th International Joint Conference on Natural Language Processing*, The Asian Federation of Natural Language Processing, p. 1257–1263, Nagoya, Japan, 2013, https://hal.inria.fr/hal-00868363.

[852] Y. Mathet, A. Widlöcher, K. Fort, C. François, O. Galibert, C. Grouin, J. Kahn, S. Rosset, P. Zweigenbaum, "Manual Corpus Annotation: Giving Meaning to the Evaluation Metrics", *in : International Conference on Computational Linguistics*, p. 809–818, Mumbaï, India, 2012, https://hal.inria.fr/hal-00769639.

[853] M. Musiol, M. Amblard, M. Rebuschi, "Approche sémantico-formelle des troubles du discours : les conditions de la saisie de leurs aspects pyscholinguistiques", *in : 27ème Congrès International de Linguistique et de Philologie Romanes*, Nancy, France, 2013, https://hal.inria.fr/hal-00910701.

[854] G. Perrier, M. Candito, B. Guillaume, C. Ribeyre, K. Fort, D. Seddah, "Un schéma d'annotation en dépendances syntaxiques profondes pour le français", *in : TALN - Traitement Automatique des Langues Naturelles*, p. 574–579, Marseille, France, 2014, https://hal.inria.fr/hal-01054407.

[855] G. Perrier, B. Guillaume, "Annotation sémantique du French Treebank à l'aide de la réécriture modulaire de graphes", *in : Conférence annuelle sur le Traitement Automatique des Langues - TALN'12*, G. S. Georges Antoniadis, Hervé Blanchon (editor), ATALA, p. 293–306, Grenoble, France, 2012, https://hal.inria.fr/hal-00760629.

[856] G. Perrier, "Analyse statique des interactions entre structures élémentaires d'une grammaire", *in : TALN 2013*, E. Morin, Y. Estève (editors), Université de Nantes, p. 604–611, Les Sables d'Olonne, France, 2013, https://hal.inria.fr/hal-00920757.

[857] S. Pogodalla, F. Pompigne, “Controlling Extraction in Abstract Categorial Grammars”, *in : Formal Grammar. 15th and 16th International Conferences, FG 2010, Copenhagen, Denmark, August 2010, FG 2011, Ljubljana, Slovenia, August 2011, Revised Selected Papers*, P. de Groote, M.-J. Nederhof (editors), *Lecture Notes in Computer Science*, *7395*, Springer Berlin Heidelberg, p. 162–177, 2012, https://hal.inria.fr/inria-00565629.

[858] S. Qian, M. Amblard, “Event in Compositional Dynamic Semantics”, *in : 6th International Conference on Logical Aspect of Computational Linguistic - LACL 2011*, S. Pogodalla, J.-P. Prost (editors), *Lecture Notes in Artificial Intelligence*, *6736*, Springer, p. 219–234, Montpellier, France, 2011. ISBN : 978-3-642-22220-7, https://hal.inria.fr/hal-00601620.

[859] S. Qian, M. Amblard, “Accessibility for Plurals in Continuation Semantics”, *in : New Frontiers in Artificial Intelligence - JSAI-isAI 2012 Workshops, LENLS, JURISIN, MiMI, Miyazaki, Japan, November 30 and December 1, 2012, Revised Selected Papers.*, M. Yoichi, B. Alastair, B. Daisuke (editors), *Lecture Notes in Computer Science*, *7856*, Springer International Publishing, p. 53–68, 2013, https://hal.inria.fr/hal-00762203.

Secondary International Conferences

[860] M. Amblard, M. Musiol, M. Rebuschi, “Schizophrénie et Langage : Analyse et modélisation. De l’utilisation des modèles formels en pragmatique pour la modélisation de discours pathologiques”, *in : Congrès MSH 2012*, Caen, France, 2012, https://hal.inria.fr/hal-00761540.

[861] M. Amblard, “Quantitative study of linguistic phenomenas as indices of Thought and Language Disorders”, *in : (In)Coherence du Discours 2*, Nancy, France, 2014, https://hal.inria.fr/hal-01164738.

[862] P. de Groote, “Abstract Categorial Parsing as Linear Logic Programming”, *in : Proceedings of the 14th Meeting on the Mathematics of Language (MoL 2015)*, Association for Computational Linguistics, p. 15–25, Chicago, United States, 2015, https://hal.inria.fr/hal-01188632.

[863] V. Dufour-Lussier, B. Guillaume, G. Perrier, “Parsing Coordination Extragrammatically”, *in : 5th Language & Technology Conference - LTC’11*, Poznan, Poland, 2011, https://hal.inria.fr/hal-00639929.

[864] K. Fort, B. Guillaume, H. Chastant, “Creating Zombilingo, a Game With A Purpose for dependency syntax annotation”, *in : Gamification for Information Retrieval (GamifIR’14) Workshop*, Amsterdam, Netherlands, 2014, https://hal.inria.fr/hal-00969157.

[865] K. Fort, B. Guillaume, V. Stern, “ZOMBILINGO : manger des têtes pour annoter en syntaxe de dépendances”, *in : TALN - Traitement Automatique des Langues Naturelles*, p. 15–16, Marseille, France, 2014, https://hal.inria.fr/hal-01054395.

[866] C. Gardent, Y. Parmentier, G. Perrier, S. Schmitz, “Lexical Disambiguation in LTAG using Left Context”, *in : 5th Language & Technology Conference - LTC’11*, p. 395–399, Poznań, Poland, 2011, https://hal.inria.fr/hal-00629902.

[867] L. Kallmeyer, T. Lichte, R. Osswald, S. Pogodalla, C. Wurm, “Quantification in Frame Semantics with Hybrid Logic”, *in : Type Theory and Lexical Semantics*, R. Cooper, C. Retoré (editors), *ESSLLI 2015*, Barcelona, Spain, 2015, https://hal.inria.fr/hal-01151641.

[868] J. Marsik, M. Amblard, “Pragmatic Side Effects”, *in : Redrawing Pragmasemantic Borders*, Groningen, Netherlands, 2015, https://hal.inria.fr/hal-01164729.

[869] J. Maršík, M. Amblard, "Integration of Multiple Constraints in ACG", *in : Logic and Engineering of Natural Language Semantics 10*, p. 1–14, Kanagawa, Japan, 2013, https://hal.inria.fr/hal-00869748.

[870] J. Maršík, M. Amblard, "Algebraic Effects and Handlers in Natural Language Interpretation", *in : Natural Language and Computer Science*, V. de Paiva, W. Neuper, P. Quaresma, C. Retoré, L. S. Moss, J. Saludes (editors), *Joint Proceedings of the Second Workshop on Natural Language and Computer Science (NLCS'14) & 1st International Workshop on Natural Language Services for Reasoners (NLSR 2014), TR 2014-002*, Valeria de Paiva and Walther Neuper and Pedro Quaresma and Christian Retoré and Lawrence S. Moss and Jordi Saludes, Center for Informatics and Systems of the University of Coimbra, Vienne, Austria, 2014, https://hal.inria.fr/hal-01079206.

[871] G. Perrier, B. Guillaume, "Semantic Annotation of the French Treebank with Modular Graph Rewriting", *in : META-RESEARCH Workshop on Advanced Treebanking, LREC 2012 Workshop*, J. Hajic (editor), META-NET, Istanbul, Turkey, 2012, https://hal.inria.fr/hal-00760577.

[872] G. Perrier, B. Guillaume, "FRIGRAM: a French Interaction Grammar", *in : ESSLLI 2013 - Workshop on High-level Methodologies for Grammar Engineering*, D. Duchier, Y. Parmentier (editors), Université de Düsseldorf, p. 63–74, Düsseldorf, Germany, 2013, https://hal.inria.fr/hal-00920717.

[873] S. Pogodalla, "Comments to Yao and Zou on: "Hybrid Categorial Type Logics and the Formal Treatment of Chinese"", *in : Tsinghua Logic Conference*, J. van Benthem, F. Liu (editors), *47*, College Publications, p. 269–274, Beijing, China, 2013, https://hal.inria.fr/hal-00868241.

[874] S. Qian, "Workshop Proposal for Young Researchers' Roundtable on Spoken Dialogue Systems 2011", *in : Young Researchers' Roundtable on Spoken Dialogue Systems 2011*, Portland, United States, 2011, https://hal.inria.fr/hal-00646845.

[875] M. Rebuschi, M. Amblard, M. Musiol, "Interpreting conversations in pathological contexts ", *in : Investigating semantics: Empirical and philosophical approaches*, Bochum, Germany, 2013, https://hal.inria.fr/hal-01253359.

National Peer-Reviewed Conferences

[876] G. Adda, L. Besacier, A. Couillault, K. Fort, J. Mariani, H. De Mazancourt, ""Where the data are coming from?" Ethics, crowdsourcing and traceability for Big Data in Human Language Technology", *in : Crowdsourcing and human computation multidisciplinary workshop*, CNRS, Paris, France, 2014, https://hal.inria.fr/hal-01078045.

[877] M. Amblard, K. Fort, M. Musiol, M. Rebuschi, "L'impossibilité de l'anonymat dans le cadre de l'analyse du discours", *in : Journée ATALA éthique et TAL*, Paris, France, 2014, https://hal.inria.fr/hal-01079308.

[878] A. Couillault, K. Fort, "Charte Éthique et Big Data : parce que mon corpus le vaut bien !", *in : Linguistique, Langues et Parole : Statuts, Usages et Mésusages*, p. 4, Strasbourg, France, 2013, https://hal.inria.fr/hal-00820352.

Books or Proceedings Editing

[879] P. de Groote, M. Egg, L. Kallmeyer (editors), *Formal Grammar, 14th International Conference, FG 2009, Revised Selected Papers, Lecture Notes in Artificial Intelligence, 5591*, Springer, 2011, 213p., https://hal.inria.fr/inria-00608514.

[880] P. de Groote, M.-J. Nederhof (editors), *Formal Grammar - 15th and 16th International Conferences, FG 2010, Copenhagen, Denmark, August 2010, FG 2011, Ljubljana, Slovenia, August 2011, Revised Selected Papers*, *Lecture Notes in Computer Science*, *7395*, Springer, 2012, 307p., https://hal.inria.fr/hal-00763203.

[881] S. Pogodalla, J.-P. Prost (editors), *Logical Aspects of Computational Linguistics, 6th International Conference, LACL 2011*, *LNCS/LNAI*, *6736*, Springer, 2011, 283p., https://hal.inria.fr/inria-00607874.

[882] S. Pogodalla, M. Quatrini, C. Retoré, *Logic and Grammar. Essays Dedicated to Alain Lecomte on the Occasion of His 60th Birthday*, *Lecture Notes in Computer Science*, *6700*, Springer, 2011, https://hal.inria.fr/inria-00607880.

Book chapters

[883] M. Amblard, A. Boumaza, "Robots humains, avez-donc une réalité ?", *in : AI:Philosophy, Geoinformatics & Law*, M. Jankowska, M. Pawelczyk, S. Allouche, and M. Kulawiak (editors), IUS PUBLICUM, 2015, p. 14, https://hal.inria.fr/hal-01188684.

[884] M. Amblard, S. Pogodalla, "Modeling the Dynamic Effects of Discourse: Principles and Frameworks", *in : Dialogue, Rationality, and Formalism*, M. Rebuschi, M. Batt, G. Heinzmann, F. Lihoreau, M. Musiol, and A. Trognon (editors), *Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics. Logic, Argumentation & Reasoning*, *3*, Springer, 2014, p. 247–282, https://hal.inria.fr/hal-00737765.

[885] M. Amblard, "Minimalist Grammars and Minimalist Categorial Grammars, definitions toward inclusion of generated languages", *in : Logic and Grammar: Essays Dedicated to Alain Lecomte on the Occasion of His 60th Birthday*, S. Pogodalla, M. Quatrini, and C. Retoré (editors), *LNCS/LNAI*, *6700*, springer, 2011, p. 61–80, ISBN : 978-3-642-21489-9, https://hal.inria.fr/hal-00617040.

[886] G. Bonfante, B. Guillaume, M. Morey, G. Perrier, "Supertagging with Constraints", *in : Constraint and Language*, P. Blache, H. Christiansen, V. Dahl, D. Duchier, and J. Villadsen (editors), Cambridge Scholar Publishing, 2014, p. 253–297, https://hal.inria.fr/hal-01097999.

[887] M. Rebuschi, M. Amblard, M. Musiol, "Using SDRT to analyze pathological conversations. Logicality, rationality and pragmatic deviances", *in : Interdisciplinary Works in Logic, Epistemology, Psychology and Linguistics: Dialogue, Rationality, and Formalism*, M. Rebuschi, M. Batt, G. Heinzmann, F. Lihoreau, M. Musiol, and A. Trognon (editors), *Logic, Argumentation & Reasoning*, *3*, Springer, 2014, p. 343 – 368, https://hal.inria.fr/hal-00910725.

Other Publications

[888] C. Baskent, "Inquiry, Refutations and the Inconsistent", working paper or preprint, 2014, https://hal.inria.fr/hal-01094783.

[889] C. Baskent, "Some Non-Classical Approaches to the Branderburger-Keisler Paradox", working paper or preprint, 2014, https://hal.inria.fr/hal-01094784.

[890] C. Baskent, "Topological Semantics for da Costa Paraconsistent Logics Cω and C*ω", working paper or preprint, 2014, https://hal.inria.fr/hal-01094786.

[891] C. Baskent, “Which Society, Which Software?”, working paper or preprint, 2014, https://hal.inria.fr/hal-01094785.

[892] G. Bonfante, B. Guillaume, “Non-size increasing Graph Rewriting for Natural Language Processing”, Accepted for publication in MSCS (Mathematical Structures for Computer Science), 2013, https://hal.inria.fr/hal-00921038.

[893] M. Candito, G. Perrier, “Guide d’annotation en dépendances profondes pour le français”, 2015, Ce guide décrit le schéma d’annotation en dépendances profondes utilisé pour l’annotation en syntaxe profonde du corpus Sequoia initialement annoté en syntaxe de surface, https://hal.inria.fr/hal-01249907.

[894] G. Perrier, B. Guillaume, “Leopar: an Interaction Grammar Parser”, ESSLLI 2013 - Workshop on High-level Methodologies for Grammar Engineering, 2013, https://hal.inria.fr/hal-00920728.

[895] G. Perrier, “FRIGRAM: a French Interaction Grammar”, *Research Report number RR-8323*, INRIA Nancy ; INRIA, 2014, https://hal.inria.fr/hal-00840254.

[896] S. Pogodalla, “Functional Approach to Tree-Adjoining Grammars and Semantic Interpretation: an Abstract Categorial Grammar Account”, working paper or preprint, 2015, https://hal.inria.fr/hal-01242154.

[897] M. Rebuschi, M. Amblard, M. Musiol, “Interpreting conversations in pathological contexts”, Investigating semantics: Empirical and philosophical approaches, 2013, Poster. hal-00910709, https://hal.inria.fr/hal-00910709.

7 References for SMa^rT

Articles in International Peer-Reviewed Journal

[898] A. Ben Romdhane, S. Jamoussi, A. Ben Hamadou, K. Smaili, “Phrase-Based Language Model in Statistical Machine Translation”, *International Journal of Computational Linguistics and Applications*, December 2016, https://hal.inria.fr/hal-01336485.

[899] S. Harrat, K. Meftouh, M. Abbas, K. Smaïli, “An Algerian dialect: Study and Resources”, *International Journal of Advanced Computer Science and Applications - IJACSA 7*, 3, March 2016, https://hal.archives-ouvertes.fr/hal-01297415.

[900] S. Raybaud, D. Langlois, K. Smaïli, ““This sentence is wrong.” Detecting errors in machine-translated sentences.”, *Machine Translation 25*, 1, August 2011, p. p. 1–34, https://hal.archives-ouvertes.fr/hal-00606350.

[901] M. Saad, D. Langlois, K. Smaili, “Extracting Comparable Articles from Wikipedia and Measuring their Comparabilities”, *Procedia - Social and Behavioral Sciences 95*, 2013, p. 40–47, Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013), https://hal.inria.fr/hal-00907442.

Major International Conferences

[902] A. Ben Romdhane, S. Jamoussi, A. Ben Hamadou, K. Smaïli, "Phrase-based Language Modelling for Statistical Machine Translation", *in : 11th International Workshop on Spoken Language*, Lake Tahoe, United States, December 2014, https://hal.archives-ouvertes.fr/hal-01112760.

[903] J. Derkacz, M. Leszczuk, M. Grega, A. Kozbial, K. Smaïli, "Definition of Requirements for Accessing Multilingual Information and Opinions", *in : The 10th edition of International Conference on Multimedia & Network Information Systems*, WROCLAW, Poland, September 2016, https://hal.inria.fr/hal-01336568.

[904] S. Harrat, K. Meftouh, M. Abbas, S. Jamoussi, M. Saad, K. Smaili, "Cross-Dialectal Arabic Processing", *in : International Conference on Intelligent Text Processing and Computational Linguistics*, *Lecture Notes in Computer Science*, cairo, Egypt, April 2015, https://hal.archives-ouvertes.fr/hal-01261598.

[905] S. Harrat, K. Meftouh, M. Abbas, K. Smaïli, "Diacritics Restoration for Arabic Dialects", *in : INTERSPEECH 2013 - 14th Annual Conference of the International Speech Communication Association*, ISCA, Lyon, France, August 2013, https://hal.inria.fr/hal-00925815.

[906] S. Harrat, K. Meftouh, M. Abbas, K. Smaïli, "Building Resources for Algerian Arabic Dialects", *in : 15th Annual Conference of the International Communication Association Interspeech*, ISCA, Singapour, Singapore, September 2014, https://hal.inria.fr/hal-01066989.

[907] S. Harrat, K. Meftouh, M. Abbas, K. Smaïli, "Grapheme To Phoneme Conversion - An Arabic Dialect Case", *in : Spoken Language Technologies for Under-resourced Languages*, Saint Petesbourg, Russia, May 2014, https://hal.inria.fr/hal-01067022.

[908] S. Jaffali, S. Jamoussi, B. Abdelmajid, K. Smaili, "Clustering and Classification of Like-Minded People from Their Tweets", *in : COOL-SNA Workshop on Connecting Online and Offline Social Network Analysis' of the IEEE International Conference on Data Mining (ICDM'14)*, Shenzen, China, December 2014, https://hal.archives-ouvertes.fr/hal-01112778.

[909] D. Jouvet, D. Langlois, "A Machine Learning Based Approach for Vocabulary Selection for Speech Transcription", *in : TSD - 16th International Conference on Text, Speech and Dialogue - 2013*, I. Habernal, V. Matoušek (editors), *Lecture Notes in Artificial Intelligence, 8082*, Springer Verlag, p. 60–67, Pilsen, Czech Republic, September 2013, https://hal.inria.fr/hal-00834302.

[910] D. Langlois, S. Raybaud, K. Smaïli, "LORIA System for the WMT12 Quality Estimation Shared Task", *in : The Seventh Workshop on Statistical Machine Translation - NAACL 2012*, *Statistical Machine Translation*, Association for Computational Linguistics, p. 114–119, Montréal, Canada, June 2012, https://hal.inria.fr/hal-00726372.

[911] D. Langlois, K. Smaïli, "LORIA System for the WMT13 Quality Estimation Shared Task", *in : ACL 2013 Eighth Workshop on Statistical Machine Translation*, Sofia, Bulgaria, 2013, https://hal.inria.fr/hal-00923623.

[912] D. Langlois, "LORIA System for the WMT15 Quality Estimation Shared Task", *in : Workshop on Statistical Machine Translation, Proceedings of the Tenth Workshop on Statistical Machine Translation*, Association for Computational Linguistics, p. 8, Lisbonne, Portugal, September 2015, https://hal.inria.fr/hal-01262104.

[913] C. Latiri, K. Smaïli, C. Lavecchia, C. Nasri, D. Langlois, "Phrase-based machine translation based on text mining and statistical language modeling techniques", *in : 12th International Conference on Intelligent Text Processing and Computational Linguistics - CICLing2011*, Tokyo, Japan, February 2011, https://hal.inria.fr/inria-00579335.

[914] K. Meftouh, N. Bouchemal, K. Smaïli, "A Study of a Non-Resourced Language: The Case of one of the Algerian Dialects", *in : The third International Workshop on Spoken Languages Technologies for Under-resourced Languages - SLTU'12*, p. –, Cape-town, South Africa, May 2012, https://hal.archives-ouvertes.fr/hal-00727042.

[915] K. Meftouh, S. Harrat, S. Jamoussi, M. Abbas, K. Smaili, "Machine Translation Experiments on PADIC: A Parallel Arabic DIalect Corpus", *in : The 29th Pacific Asia Conference on Language, Information and Computation*, shanghai, China, October 2015, https://hal.archives-ouvertes.fr/hal-01261587.

[916] F. Mezzoudj, D. Langlois, D. Jouvet, A. Benyettou, "Textual Data Selection for Language Modelling in the Scope of Automatic Speech Recognition", *in : International Conference on Natural Language and Speech Processing*, Alger, Algeria, October 2015, https://hal.inria.fr/hal-01184192.

[917] C. Nasri, L. Chiraz, K. Smaili, "STATISTICAL MACHINE TRANSLATION IMPROVEMENT BASED ON PHRASE SELECTION", *in : Recent Advances in Natural Language Processing*, Hissar, Bulgaria, September 2015, https://hal.inria.fr/hal-01261563.

[918] C. Nasri, K. Smaïli, C. Latiri, Y. Slimani, "A new method for learning Phrase Based Machine Translation with Multivariate Mutual Information", *in : The 8th International Conference on Natural Language Processing and Knowledge Engineering - NLP-KE'12*, HuangShan, China, September 2012, https://hal.archives-ouvertes.fr/hal-00727044.

[919] C. Nasri, K. Smaïli, C. Latiri, "Training phrase-based SMT without explicit word aligment", *in : 15th International Conference on Intelligent Text Processing and Computational Linguistics*, Springer, Kathmandu, Nepal, April 2014, https://hal.inria.fr/hal-01067051.

[920] S. Raybaud, D. Langlois, K. Smaïli, "Broadcast news speech-to-text translation experiments", *in : The Thirteenth Machine Translation Summit*, p. 378–381, Xiamen, China, September 2011, https://hal.archives-ouvertes.fr/hal-00628101.

[921] M. Saad, D. Langlois, K. Smaïli, "Comparing Multilingual Comparable Articles Based On Opinions", *in : Proceedings of the 6th Workshop on Building and Using Comparable Corpora*, Association for Computational Linguistics ACL, p. 105–111, Sofia, Bulgaria, August 2013, https://hal.inria.fr/hal-00851959.

[922] M. Saad, D. Langlois, K. Smaili, "Building and Modelling Multilingual Subjective Corpora", *in : Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, Iceland, Iceland, May 2014, https://hal.inria.fr/hal-00995755.

[923] M. Saad, D. Langlois, K. Smaïli, "Cross-Lingual Semantic Similarity Measure for Comparable Articles", *in : Advances in Natural Language Processing - 9th International Conference on NLP, PolTAL 2014, Warsaw, Poland, September 17-19, 2014. Proceedings*, Springer International Publishing, p. 105–115, Warsaw, Poland, September 2014, https://hal.inria.fr/hal-01067687.

Book chapters

[924] A. Douib, D. Langlois, K. Smaili, "Genetic-based decoder for statistical machine translation", *in : Springer LNCS series, Lecture Notes in Computer Science*, December 2016, https://hal.inria.fr/hal-01336546.

[925] S. Jaffali, A. Ben Hamadou, S. Jamoussi, K. Smaili, "Grouping Like-Minded Users for Ratings' Prediction", *in : Intelligent Decision Technologies*, June 2016, https://hal.inria.fr/hal-01336507.

[926] M.-A. Menacer, A. Boumerdas, C. Zakaria, K. Smaili, "A new language model based on possibility theory", *in : Springer LNCS series, Lecture Notes in Computer Science.*, December 2016, https://hal.inria.fr/hal-01336535.

8 References for Synalp

Doctoral Dissertations

[927] I. Falk, *Making Use of Existing Lexical Resources to Build a Verbnet like Classification of French Verbs*, Theses, Université Nancy II, June 2012, https://tel.archives-ouvertes.fr/tel-00714737.

[928] C. Gillot, *Language models exploiting the structural similarity between sequences for automatic speech recognition*, Theses, Université de Lorraine, December 2012, https://hal.archives-ouvertes.fr/tel-01258153.

[929] B. Gyawali, *Surface Realisation from Knowledge Bases*, Theses, Universite de Lorraine, January 2016, https://hal.inria.fr/tel-01276008.

[930] S. Narayan, *Generating and Simplifying Sentences*, Theses, Université de Lorraine, November 2014, https://hal.archives-ouvertes.fr/tel-01256262.

[931] L. H. Perez, *Natural Language Generation for Language Learning*, Theses, Université de Lorraine, April 2013, https://hal.inria.fr/tel-01256196.

Articles in International Peer-Reviewed Journal

[932] M. Amoia, T. Brétaudière, A. Denis, C. Gardent, L. Perez-Beltrachini, "A Serious Game for Second Language Acquisition in a Virtual Environment", *Journal on Systemics, Cybernetics and Informatics (JSCI) 10*, 1, 2012, p. 24–34, https://hal.inria.fr/hal-00766460.

[933] T. Bretaudière, S. Cruz-Lara, L. M. Rojas Barahona, "Associating Automatic Natural Language Processing to Serious Games and Virtual Worlds", *The Journal of Virtual Worlds Research 4*, 3, December 2011, https://hal.inria.fr/hal-00655980.

[934] J. M. Cabello, J. M. Franco, A. Collado, J. Janer, S. Cruz-Lara, D. Oyarzun, A. Armisen, R. Geraerts, "Standards in Virtual Worlds Virtual Travel Use Case Metaverse1 Project", *The Journal of Virtual Worlds Research 4*, 3, December 2011, https://hal.inria.fr/hal-00655974.

[935] B. Crabbé, D. Duchier, C. Gardent, J. Le Roux, Y. Parmentier, "XMG : eXtensible MetaGrammar", *Computational Linguistics 39*, 3, September 2013, p. 591–629, https://hal.archives-ouvertes.fr/hal-00768224.

[936] S. Cruz-Lara, A. Denis, N. Bellalem, "Linguistic and Multilingual Issues in Virtual Worlds and Serious Games: a General Review", *The Journal of Virtual Worlds Research 7*, 1, January 2014, https://hal.inria.fr/hal-00949908.

[937] S. Cruz-Lara, B. Fernández Manjón, C. Vaz De Carvalho, "Enfoques Innovadores en Juegos Serios", *IEEE VAEP RITA 1*, 1, March 2013, p. 19–21, https://hal.inria.fr/hal-00820350.

[938] S. Cruz-Lara, T. Osswald, J. Guinaud, N. Bellalem, L. Bellalem, "A Chat Interface Using Standards for Communication and E-learning in Virtual Worlds", *Lecture Notes in Business Information Processing 73*, Part 6, 2011, p. 541–554, https://hal.inria.fr/inria-00580198.

[939] P. Cuxac, J.-C. Lamirel, V. Bonvallot, "Efficient supervised and semi-supervised approaches for affiliations disambiguation", *Scientometrics 97*, 1, October 2013, p. 47–58, https://hal.archives-ouvertes.fr/hal-00960435.

[940] P. Cuxac, J.-C. Lamirel, "Analyse intelligente de documents et découverte de connaissances", *Bulletin de l'AFIA*, April 2011, p. 36–40, https://hal.archives-ouvertes.fr/hal-00955416.

[941] I. Falk, C. Gardent, "Combining Formal Concept Analysis and Translation to Assign Frames and Semantic Role Sets to French Verbs", *Annals of Mathematics and Artificial Intelligence*, August 2013, https://hal.inria.fr/hal-00852861.

[942] C. Gardent, B. Gottesman, L. Perez-Beltrachini, "Using Regular Tree Grammars to enhance Sentence Realisation", *Natural Language Engineering 17*, 2, January 2011, p. 185–201, https://hal.inria.fr/inria-00537219.

[943] C. Gardent, S. Narayan, "Multiple Adjunction in Feature-Based Tree-Adjoining Grammar", *Computational Linguistics 41*, 1, March 2015, p. 29, https://hal.inria.fr/hal-01249975.

[944] K. Hajlaoui, P. Cuxac, J.-C. Lamirel, C. François, "Aide à l'expertise des brevets par alignement avec les publications scientifiques", *Revue des Sciences et Technologies de l'Information - Série Document Numérique*, April 2013, p. 11–29, https://hal.archives-ouvertes.fr/hal-00959424.

[945] K. Hajlaoui, J.-C. Lamirel, "A promising combination of approaches for solving complex text classification tasks: application to the classification of scientific papers into patents classes", *International Journal of Knowledge and Learning 9*, 1, 2014, p. 22, https://hal.inria.fr/hal-01263648.

[946] P. Kral, C. Cerisara, "Automatic dialogue act recognition with syntactic features", *Language Resources and Evaluation*, February 2014, p. http://link.springer.com/article/10.1007 https://hal.archives-ouvertes.fr/hal-00954623.

[947] J.-C. Lamirel, P. Cuxac, K. Hajlaoui, "Optimizing text classification through efficient feature selection based on quality metric", *Journal of Intelligent Information Systems*, 2014, p. 18, https://hal.inria.fr/hal-01263651.

[948] J.-C. Lamirel, I. Falk, C. Gardent, "Federating clustering and cluster labelling capabilities with a single approach based on feature maximization: French verb classes identification with IGNGF neural clustering.", *Neurocomputing 147*, January 2015, p. 136–146, https://hal.inria.fr/hal-01074277.

[949] J.-C. Lamirel, D. Reymond, "Automatic websites classification and retrieval using websites communication signatures", *Journal of Information Management and Scientometrics*, 2014, https://hal.inria.fr/hal-01263646.

[950] J.-C. Lamirel, "A new approach for automatizing the analysis of research topics dynamics: application to optoelectronics research", *Scientometrics 93*, 1, 2012, p. 151–166, https://hal.inria.fr/hal-00781576.

[951] J.-C. Lamirel, "Multi-View Data Analysis and Concept Extraction Methods for Text", *Knowledge Organization 40*, 5, 2013, p. 305–319, https://hal.inria.fr/hal-00939034.

[952] M. Zennaki, A. Ech-Cherif, J.-C. Lamirel, "Learning Solution Quality of Hard Combinatorial Problems by Kernel Methods", *International Journal of Computer Applications (IJCA)*, August 2011, https://hal.inria.fr/hal-00645400.

Invited Conferences

[953] J.-C. Lamirel, "A new approach for improving clustering quality based on connected maximized cluster features", *in : Classification Society 2012 Annual Meeting*, Pittsburgh, United States, June 2012, https://hal.inria.fr/hal-00781580.

[954] J.-C. Lamirel, "A new incremental clustering approach based on cluster data feature maximization", *in : Classification Society 2012 Annual Meeting*, Pittsburgh, United States, June 2012, https://hal.inria.fr/hal-00781579.

[955] J.-C. Lamirel, "An unbiased symbolico-numeric approach for reliable clustering quality evaluation", *in : Classification Society 2012 Annual Meeting*, Pittsburgh, United States, June 2012, https://hal.inria.fr/hal-00781578.

[956] J.-C. Lamirel, "Opening new perspectives for classifying and mining textual and web data", *in : 2nd international Symposium ISKO-MAGHREB - 2012*, Hammamet, Tunisia, November 2012, https://hal.inria.fr/hal-00781577.

Major International Conferences

[957] M. Amoia, C. Gardent, L. Perez-Beltrachini, "A Serious Game for Second Language Acquisition", *in : Third International Conference on Computer Supported Education - CSEDU 2011*, A. Verbraeck, M. Helfert, J. Cordeiro, B. Shishkov (editors), SciTePress, p. 394–397, Noordwijkerout, Netherlands, May 2011. ISBN : 978-989-8425-49-2, https://hal.inria.fr/hal-00643955.

[958] M. Amoia, C. Gardent, L. Perez-Beltrachini, "Learning a Second Language with a Videogame", *in : ICT for Language Learning*, Pixel Associazione, Firenze, Italy, October 2011, https://hal.inria.fr/hal-00643953.

[959] M. Amoia, "I-FLEG: A 3D-Game for Learning French.", *in : The 9th International Conference on Education and Information Systems, Technologies and Applications - EISTA 2011*, Orlando, United States, July 2011, https://hal.inria.fr/hal-00643959.

[960] C. Areces, P. Fontaine, "Combining theories: the Ackerman and Guarded Fragments", *in : 8th International Symposium Frontiers of Combining Systems - FroCoS 2011*, C. Tinelli, V. Sofronie-Stokkermans (editors), *6989*, Viorica Sofronie-Stokkermans, Springer Verlag, p. 40–54, Saarbrücken, Germany, October 2011, https://hal.inria.fr/hal-00642529.

[961] E. Banik, C. Gardent, E. Kow, "The KBGen Challenge", *in : the 14th European Workshop on Natural Language Generation (ENLG)*, p. 94–97, Sofia, Bulgaria, August 2013, https://hal.archives-ouvertes.fr/hal-00920605.

[962] E. Banik, C. Gardent, D. Scott, N. Dinesh, F. Linag, "KBGen - Text Generation for Knowledge Bases as a New Shared Task", *in : The seventh International Natural Language Generation Conference. Starved Rock, Illinois, USA.*, p. 141–146, Starved Rock, Illinois, United States, May 2012, https://hal.archives-ouvertes.fr/hal-00768616.

[963] P. Bedaride, C. Gardent, "Deep Semantics for Dependency Structures", *in : 12th International Conference on Computational Linguistics and Intelligent Text Processing - CICLing 2011*, A. F. Gelbukh (editor), *Lecture Notes in Computer Science*, *6608*, Springer Verlag, p. 277–288, Tokyo, Japan, February 2011. The original publication is available at www.springerlink.com, https://hal.inria.fr/hal-00639825.

[964] L. Benotti, A. A. J. Denis, "CL system: Giving instructions by corpus based selection", *in : 13th European Workshop on Natural Language Generation*, Nancy, France, September 2011, https://hal.inria.fr/inria-00636474.

[965] L. Benotti, A. A. J. Denis, "Giving instructions in virtual environments by corpus based selection", *in : SIGdial Meeting on Discourse and Dialogue*, Portland, United States, June 2011, https://hal.inria.fr/inria-00636303.

[966] L. Benotti, A. A. J. Denis, "Prototyping virtual instructors from human-human corpora", *in : Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations*, Portland, United States, June 2011, https://hal.inria.fr/inria-00636300.

[967] C. Cerisara, C. Gardent, "The JSafran platform for semi-automatic speech processing", *in : 12th Annual Conference of the International Speech Communication Association - Interspeech 2011*, p. 4, Florence, Italy, August 2011, https://hal.archives-ouvertes.fr/hal-00600520.

[968] C. Cerisara, P. Kral, C. Gardent, "Commas recovery with syntactic features in French and in Czech", *in : 12thAnnual Conference of the International Speech Communication Association - Interspeech 2011*, p. 4, Florence, Italy, August 2011, https://hal.archives-ouvertes.fr/hal-00600528.

[969] C. Cerisara, A. Lorenzo, P. Kral, "Weakly supervised parsing with rules", *in : INTERSPEECH 2013*, p. 2192–2196, Lyon, France, August 2013, https://hal.archives-ouvertes.fr/hal-00850437.

[970] C. Cerisara, A. Lorenzo, "Mixed probabilistic and deterministic dependency parsing", *in : 13th Annual Conference of the International Speech Communication Association - InterSpeech 2012*, p. 4, Portland, United States, September 2012, https://hal.archives-ouvertes.fr/hal-00706547.

[971] P. Cuxac, V. Bonvallot, J.-C. Lamirel, "Efficient supervised and semi-supervised approaches for affliations disambiguation", *in : 13th COLLNET Meeting*, p. 10 p., Seoul, North Korea, October 2012, https://hal.archives-ouvertes.fr/hal-00956386.

[972] P. Cuxac, J.-C. Lamirel, "Efficient supervised and semi- supervised approaches for affiliations disambiguation", *in : WIS - 8th International Conference on Webometrics, Informetrics and Scientometrics - 2012*, Séoul, South Korea, October 2012, https://hal.inria.fr/hal-00781584.

[973] P. Cuxac, J.-C. Lamirel, "Analysis of evolutions and interactions between science fields: the cooperation between feature selection and graph representation", *in : 14th COLLNET Meeting*, Tartu, Estonia, August 2013, https://hal.archives-ouvertes.fr/hal-00959415.

[974] A. Denis, S. Cruz-Lara, N. Bellalem, L. Bellalem, "Synalp-Empathic: A Valence Shifting Hybrid System for Sentiment Analysis", *in : 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 605 – 609, Dublin, Ireland, August 2014, https://hal.archives-ouvertes.fr/hal-01099671.

[975] A. Denis, S. Cruz-Lara, N. Bellalem, L. Bellalem, "Visualization of affect in movie scripts", *in : Empatex, 1st International Workshop on Empathic Television Experiences at TVX 2014*, Newcastle, United Kingdom, June 2014, https://hal.archives-ouvertes.fr/hal-01099668.

[976] A. Denis, S. Cruz-Lara, N. Bellalem, "General Purpose Textual Sentiment Analysis and Emotion Detection Tools", *in : Workshop on Emotion and Computing*, Koblenz, Germany, September 2013, https://hal.inria.fr/hal-00860316.

[977] A. Denis, I. Falk, C. Gardent, L. Perez-Beltrachini, "Representation of Linguistic and Domain Knowledge for Second Language Learning in Virtual Worlds", *in : LREC - The eighth international conference on Language Resources and Evaluation - 2012*, p. 2631–2635, Istanbul, Turkey, May 2012, https://hal.inria.fr/hal-00766418.

[978] A. A. J. Denis, "The Loria Instruction Generation System L in GIVE 2.5", *in : 13th European Workshop on Natural Language Generation*, Nancy, France, September 2011, https://hal.inria.fr/inria-00636479.

[979] N. Dugué, A. Tebbakh, P. Cuxac, J.-C. Lamirel, "Feature selection and complex networks methods for an analysis of collaboration evolution in science: an application to the ISTEX digital library", *in : ISKO-MAGHREB 2015*, Hammamet, Tunisia, November 2015, https://hal.archives-ouvertes.fr/hal-01231791.

[980] I. Falk, C. Gardent, J.-C. Lamirel, "Classifying French Verbs Using French and English Lexical Resources", *in : The 50th annual meeting of the Association for Computational Linguistics - ACL'12, Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 1: Long Papers*, The Association for Computational Linguistics (ACL), Association for Computational Linguistics, p. 854–863, Jeju, South Korea, July 2012, https://hal.inria.fr/hal-00690484.

[981] I. Falk, C. Gardent, "Combining Formal Concept Analysis and Translation to Assign Frames and Thematic Role Sets to French Verbs", *in : Concept Lattices and Their Applications*, A. Napoli, V. Vychodil (editors), *CLA 2011, Proceedings of the Eighth International Conference on Concept Lattices and Their Apllications.*, Nancy, France, October 2011, https://hal.inria.fr/inria-00634270.

[982] E. Franconi, C. Gardent, X. I. Juarez-Castro, L. Perez-Beltrachini, "Quelo Natural Language Interface: Generating queries and answer descriptions", *in : Natural Language Interfaces for Web of Data, Proceedings of NLIWOD 2014*, Riva del Gada - Trention, Italy, October 2014, https://hal.archives-ouvertes.fr/hal-01109607.

[983] D. Fréard, F. Détienne, M. Baker, F. Barcellini, M. Quignard, A. Denis, "Visualising zones of collaboration in online collective activity: a case study in Wikipedia", *in : European Conference on Cognitive Ergonomics*, Edinburgh, United Kingdom, August 2012, https://hal.inria.fr/hal-00766463.

[984] C. Gardent, G. Kruszewski, "Generation for Grammar Engineering", *in : INLG 2012, The seventh International Natural Language Generation Conference.*, p. 31–40, Starved Rock, Illinois, United States, May 2012, https://hal.archives-ouvertes.fr/hal-00768612.

[985] C. Gardent, A. Lorenzo, L. Perez-Beltrachini, L. M. Rojas-Barahona, "Weakly and Strongly Constrained Dialogues for Language Learning", *in : the 14th annual SIGdial Meeting on Discourse and Dialogue SIGDIAL 2013*, p. 357–359, Metz, France, August 2013, https://hal.archives-ouvertes.fr/hal-00920608.

[986] C. Gardent, S. Narayan, "Error Mining on Dependency Trees", *in : 50th Annual Meeting of the Association for Computational Linguistics*, p. 592–600, Jeju Island, South Korea, July 2012, https://hal.archives-ouvertes.fr/hal-00768204.

[987] C. Gardent, S. Narayan, "Generating Elliptic Coordination", *in : the 14th European Workshop on Natural Language Generation (ENLG)*, p. 40–50, Sofia, Bulgaria, August 2013, https://hal.archives-ouvertes.fr/hal-00920606.

[988] C. Gardent, Y. Parmentier, G. Perrier, S. Schmitz, "Lexical Disambiguation in LTAG using Left Context", *in : 5th Language & Technology Conference - LTC'11*, p. 395–399, Poznań, Poland, November 2011, https://hal.archives-ouvertes.fr/hal-00629902.

[989] C. Gardent, L. Perez-Beltrachini, "Using FB-LTAG Derivation Trees to Generate Transformation-Based Grammar Exercices", *in : TAG+11: The 11th International Workshop on Tree Adjoining Grammars and Related Formalisms*, p. 117–125, Paris, France, September 2012, https://hal.archives-ouvertes.fr/hal-00768231.

[990] C. Gardent, L. M. Rojas Barahona, "Using Paraphrases and Lexical Semantics to Improve the Accuracy and the Robustness of Supervised Models in Situated Dialogue Systems", *in : Conference on Empirical Methods in Natural Language Processing*, SIGDAT, the Association for Computational Linguistics special interest group on linguistic data and corpus-based approaches to NLP, p. 808–813, Seattle, United States, October 2013, https://hal.inria.fr/hal-00905405.

[991] C. Gillot, C. Cerisara, "Similarity language model", *in : 12thAnnual Conference of the International Speech Communication Association - Interspeech 2011*, p. 4, Florence, Italy, August 2011, https://hal.archives-ouvertes.fr/hal-00600531.

[992] B. Gyawali, C. Gardent, C. Cerisara, "A Domain Agnostic Approach to Verbalizing n-ary Events without Parallel Corpora", *in : Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)*, p. 18–27, Brighton, United Kingdom, September 2015, https://hal.inria.fr/hal-01207155.

[993] B. Gyawali, C. Gardent, C. Cerisara, "Automatic Verbalisation of Biological Events", *in : International Workshop on Definitions in Ontologies (IWOOD 2015)*, Lisbon, Portugal, July 2015, https://hal.inria.fr/hal-01214569.

[994] B. Gyawali, C. Gardent, "A Hybrid Approach To Generating from the KBGen Knowledge-Base", *in : the 14th European Workshop on Natural Language Generation (ENLG)*, p. 204–205, Sofia, Bulgaria, August 2013, https://hal.archives-ouvertes.fr/hal-00920607.

[995] B. Gyawali, C. Gardent, "Surface Realisation from Knowledge-Bases", *in : the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL (editor), p. 424–434, Baltimore, United States, June 2014, https://hal.archives-ouvertes.fr/hal-01021916.

[996] K. Hajlaoui, P. Cuxac, J.-C. Lamirel, C. François, "Enhancing Patent Expertise through Automatic Matching with Scientific Papers", *in : DS 2012 : The Fifteenth International Conference on Discovery Science*, p. 299–312, Lyon, France, October 2012, https://hal.archives-ouvertes.fr/hal-00962386.

[997] P. Kral, L. Lenc, C. Cerisara, "Semantic Features for Dialogue Act Recognition", *in : 3rd International Conference on Statistical Language and Speech Processing (SLSP)*, Budapest, Hungary, November 2015, https://hal.archives-ouvertes.fr/hal-01256301.

[998] J.-C. Lamirel, S. Al Shehabi, G. Safi, "A new label maximization based incremental neural clustering approach: application to text clustering", *in : 8th Workshop on Self-Organizing Maps (WSOM 2011)*, Espoo, Finland, June 2011, https://hal.inria.fr/hal-00645394.

[999] J.-C. Lamirel, P. Cuxac, K. Hajlaoui, A. S. Chivukula, "A new feature selection and feature contrasting approach based on quality metric: application to efficient classification of complex textual data", *in : International Workshop on Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models (QIMIE)*, Australia, April 2013, https://hal.archives-ouvertes.fr/hal-00960127.

[1000] J.-C. Lamirel, P. Cuxac, R. Mall, "A new efficient and unbiased approach for clustering quality evaluation", *in : QIMIE'11*, p. 209–220, Shenzen, China, May 2011, https://hal.archives-ouvertes.fr/hal-00955498.

[1001] J.-C. Lamirel, P. Cuxac, "Dynamically highlighting influential authors in a topic domain : a tentative approach", *in : 10th International Conference on Webometrics, Informetrics and Scientometrics (WIS), Proceedings of 10th International Conference on Webometrics, Informetrics and Scientometrics (WIS)*, Illmenau, Germany, August 2014, https://hal.inria.fr/hal-01263850.

[1002] J.-C. Lamirel, P. Cuxac, "Improving textual data classification and discrimination using an ad-hoc metric: application to a famous text discrimination challenge", *in : 4th International Symposium ISKO-Maghreb 2014, Proceedngs of 4th International Symposium ISKO-Maghreb 2014*, Alger, Algeria, November 2014, https://hal.inria.fr/hal-01263858.

[1003] J.-C. Lamirel, P. Cuxac, "New quality indexes for optimal clustering model identification with high dimensional data", *in : ICDM-HDM'15 - International Workshop on High Dimensional Data Mining, Proceedings of ICDM-HDM'15 - International Workshop on High Dimensional Data Mining*, Atlantic City, United States, November 2015, https://hal.inria.fr/hal-01263880.

[1004] J.-C. Lamirel, I. Falk, C. Gardent, "Enhancing NLP Tasks by the Use of a Recent Neural Incremental Clustering Approach Based on Cluster Data Feature Maximization", *in : WSOM - 9th Workshop on Self-Organizing Maps - 2012*, Santiago, Chile, December 2012, https://hal.inria.fr/hal-00781589.

[1005] J.-C. Lamirel, R. Mall, M. Ahmad, "Comparative behaviour of recent incremental and non-incremental clustering methods on text: an extended study", *in : The Twenty-fourth International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2011)*, Syracuse, United States, June 2011, https://hal.inria.fr/hal-00645393.

[1006] J.-C. Lamirel, R. Mall, P. Cuxac, G. Safi, "A new efficient and unbiased approach for clustering quality evaluation", *in : PAKDD 2010 2nd International Workshop on Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models (QIMIE)*, Shenzen, China, May 2011, https://hal.inria.fr/hal-00645395.

[1007] J.-C. Lamirel, R. Mall, P. Cuxac, G. Safi, "Variations to incremental growing neural gas algorithm based on label maximization", *in : International Joint Conference on Neural Networks - IJCNN 2011*, San Jose, United States, July 2011, https://hal.inria.fr/hal-00645390.

[1008] J.-C. Lamirel, D. Reymond, "Automatic websites classification and retrieval using websites communication signatures", *in : WIS - 8th International Conference on Webometrics, Informetrics and Scientometrics - 2012*, Séoul, South Korea, October 2012, https://hal.inria.fr/hal-00781585.

[1009] J.-C. Lamirel, "A new diachronic methodology for automatizing the analysis of research topics dynamics : an example of application on optoelectronics research", *in : 7th International Conference on Webometrics, Informetrics and Scientometrics and 12th COLLNET Meeting*, Istanbul, Turkey, September 2011, https://hal.inria.fr/hal-00645389.

[1010] J.-C. Lamirel, "A new method for automatically analyzing the research dynamics: application on optoelectronics research", *in : 13th ISSI 2011 Conference*, Durban, South Africa, July 2011, https://hal.inria.fr/hal-00645391.

[1011] J.-C. Lamirel, "Combining Neural Clustering with Intelligent Labeling and Unsupervised Bayesian Reasoning in a Multiview Context for Efficient Diachronic Analysis", *in : WSOM - 9th Workshop on Self-Organizing Maps - 2012*, Santiago, Chile, December 2012, https://hal.inria.fr/hal-00781588.

[1012] J.-C. Lamirel, "Dealing with highly imbalanced textual data gathered into similar classes", *in : IJCNN - 2013 International Joint Conference on Neural Networks*, Dallas, United States, August 2013, https://hal.inria.fr/hal-00939036.

[1013] J.-C. Lamirel, "Enhancing classification accuracy with the help of feature maximization metric", *in : ICTAI*, Washington, United States, November 2013, https://hal.inria.fr/hal-00939038.

[1014] J.-C. Lamirel, "Enhancing feature selection with feature maximization metric", *in : WSC*, Hong Kong, China, August 2013, https://hal.inria.fr/hal-00939043.

[1015] J.-C. Lamirel, "Supervised and unsupervised multi-view data analysis methods for text", *in : WISSORG'2013*, Postdam, Germany, March 2013, https://hal.inria.fr/hal-00939045.

[1016] J.-C. Lamirel, "Measuring and highlighting changes of influence in authors/topics networks through the use of a new metric: application to a compound dataset of scientific data", *in : DISC 2014, Proceedings of DISC 2014*, Daegu, South Korea, December 2014, https://hal.inria.fr/hal-01263864.

[1017] J.-C. Lamirel, "Exploring the dynamics of scientific collections using a new combination of approaches mixing graph representation and feature selection", *in : DISC 2015, Proceedings of DISC 2015*, Daegu, South Korea, November 2015, https://hal.inria.fr/hal-01263878.

[1018] J.-C. Lamirel, "Feature Maximization: a Highly Accurate Approach for Feature Selection and Meta-Classification", *in : Classification Society 2015 Annual Meeting*, Hamilton, Canada, May 2015, https://hal.inria.fr/hal-01263870.

[1019] J.-C. Lamirel, "Feature maximization based clustering quality evaluation: a promising approach", *in : QIMIE 2015: 4th International PAKDD Workshop on Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models, Proceedings of QIMIE 2015: 4th International PAKDD Workshop on Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models*, Ho Chi Min, Vietnam, April 2015, https://hal.inria.fr/hal-01263867.

[1020] J.-C. Lamirel, "New metrics and related statistical approaches for efficient mining in very large and highly multidimensional databases", *in : 11th IEEE International Conference: Beyond Databases, Architectures and Structures (BDAS'2015), Proceedings of 11th IEEE International Conference: Beyond Databases, Architectures and Structures (BDAS'2015)*, Cracovie, Poland, May 2015, https://hal.inria.fr/hal-01263865.

[1021] J.-C. Lamirel, "New quality indexes for optimal clustering model identification based on cross-domain approach", *in : ICTAI – CIMA workshop - 28th International Conference on Tools with Artificial Intelligence (ICTAI), Proceedings of ICTAI – CIMA workshop - 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, Vietri sul Mare, Italy, November 2015, https://hal.inria.fr/hal-01263872.

[1022] J.-C. Lamirel, "Reliable clustering quality estimation from low to high dimensional data", *in : 11th Workshop on Self-Organizing Maps (WSOM 2016), Proceedings of 11th Workshop on Self-Organizing Maps (WSOM 2016)*, Houston, United States, January 2016, https://hal.inria.fr/hal-01263883.

[1023] F. Lefevre, D. Mostefa, L. Besacier, Y. Estève, M. Quignard, N. Camelin, B. Favre, B. Jabaian, L. M. Rojas Barahona, "Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the PORTMEDIA corpora", *in : The International Conference on Language Resources and Evaluation*, Istanbul, Turkey, May 2012, https://hal.inria.fr/hal-00683433.

[1024] A. Lorenzo, C. Cerisara, "Unsupervised frame based Semantic Role Induction: application to French and English", *in : Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, Association for Computational Linguistics, p. 30–35, Jeju, South Korea, July 2012, https://hal.archives-ouvertes.fr/hal-00758506.

[1025] A. Lorenzo, C. Cerisara, "Semi-supervised SRL system with Bayesian inference", *in : 15th International Conference on Intelligent Text Processing and Computational Linguistics*, p. 433, Nepal, April 2014, https://hal.archives-ouvertes.fr/hal-01015414.

[1026] A. Lorenzo, L. M. Rojas-Barahona, C. Cerisara, "Unsupervised structured semantic inference for spoken dialog reservation tasks", *in : SIGDIAL - 14th annual SIGdial Meeting on Discourse and Dialogue - 2013*, p. 12–20, Metz, France, August 2013, https://hal.archives-ouvertes.fr/hal-00911017.

[1027] J. Mennesson, B. Allaert, I. M. Bilasco, N. van der Aa, A. A. J. Denis, S. Cruz-Lara, "Faces and Thoughts: An Empathic Dairy", *in : IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, Ljubljana, Slovenia, May 2015, https://hal.archives-ouvertes.fr/hal-01150212.

[1028] Y. Mrabet, C. Gardent, M. Foulonneau, S. Elena, R. Eric, "Towards Knowledge-Driven Annotation", *in : Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI 2015., Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI 2015.*, Austin, United States, January 2015, https://hal.inria.fr/hal-01261551.

[1029] S. Narayan, C. Gardent, "Error Mining with Suspicion Trees: Seeing the Forest for the Trees", *in : 24th International Conference in Computational Linguistics (COLING)*, p. 60–73, Mumbai, India, December 2012, https://hal.archives-ouvertes.fr/hal-00768227.

[1030] S. Narayan, C. Gardent, "Structure-Driven Lexicalist Generation", *in : 24th International Conference in Computational Linguistics (COLING)*, p. 100–113, Mumbai, India, December 2012, https://hal.archives-ouvertes.fr/hal-00768228.

[1031] S. Narayan, C. Gardent, "Hybrid Simplification using Deep Semantics and Machine Translation", *in : the 52nd Annual Meeting of the Association for Computational Linguistics*, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL, p. 435 – 445, Baltimore, United States, June 2014, https://hal.archives-ouvertes.fr/hal-01109581.

[1032] L. Perez-Beltrachini, C. Gardent, E. Franconi, "Incremental Query Generatio", *in : the 14th Conference of the European Chapter of the Association for Computational Linguistics.*, p. 183–191, Gothenburg, Sweden, April 2014, https://hal.archives-ouvertes.fr/hal-01021917.

[1033] L. Perez-Beltrachini, C. Gardent, G. Kruszewski, "Generating Grammar Exercises", *in : The 7th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT Worskhop 2012*, p. 147–157, Montreal, Canada, June 2012, https://hal.archives-ouvertes.fr/hal-00768610.

[1034] L. M. Rojas Barahona, T. Bazillon, M. Quignard, F. Lefevre, "Using MMIL for the High Level Semantic Annotation of the French MEDIA Dialogue Corpus", *in : Ninth International Conference on Computational Semantics - IWCS 2011*, J. Bos, S. Pulman (editors), ACL, London, United Kingdom, January 2011, https://hal.inria.fr/inria-00638000.

[1035] L. M. Rojas Barahona, C. Cerisara, "Bayesian Inverse Reinforcement Learning for Modeling Conversational Agents in a Virtual Environment.", *in : Conference on Intelligent Text Processing and Computational Linguistics*, Alexander Gelbukh, Springer, Kathmandu, Nepal, April 2014, https://hal.inria.fr/hal-01002361.

[1036] L. M. Rojas Barahona, C. Cerisara, "Enhanced discriminative models with tree kernels and unsupervised training for entity detection", *in : 6th. International Conference on Information Systems & Economic Intelligence (SIIE), Proc. of the 6th International Conference on Information Systems and Economic Intelligence (SIIE)*, Hammamet, Tunisia, February 2015, https://hal.archives-ouvertes.fr/hal-01184847.

[1037] L. M. Rojas Barahona, C. Cerisara, "Weakly supervised discriminative training of linear models for Natural Language Processing", *in : 3rd International Conference on Statistical Language and Speech Processing (SLSP)*, Budapest, Hungary, November 2015, https://hal.archives-ouvertes.fr/hal-01184849.

[1038] L. M. Rojas Barahona, C. Gardent, "What should I do now? Supporting conversations in a serious game.", *in : SeineDial 2012 - 16th Workshop on the Semantics and Pragmatics of Dialogue*, Jonathan Ginzburg (chair), Anne Abeillé, Margot Colinet, Gregoire Winterstein, Paris, France, September 2012, https://hal.inria.fr/hal-00757892.

[1039] L. M. Rojas Barahona, A. Lorenzo, C. Gardent, "An End-to-End Evaluation of Two Situated Dialog Systems.", *in : Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Association for Computational Linguistics, p. 10–19, Seoul, North Korea, July 2012, https://hal.inria.fr/hal-00726723.

[1040] L. M. Rojas Barahona, A. Lorenzo, C. Gardent, "Building and Exploiting a Corpus of Dialog Interactions between French Speaking Virtual and Human Agents", *in : The eighth international conference on Language Resources and Evaluation (LREC)*, N. C. C. Chair), K. Choukri,

T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis (editors), European Language Resources Association (ELRA), p. 1428–1435, Istanbul, Turkey, May 2012, https://hal.inria.fr/hal-00726721.

[1041] L. M. Rojas Barahona, M. Quignard, "An Incremental Architecture for the Semantic Annotation of Dialogue Corpora with High-Level Structures. A case study for the MEDIA corpus", *in : 12th annual SIGdial Meeting on Discourse and Dialogue - SIGDIAL 2011*, J. Moore, D. Traum (editors), ACL, p. 332–334, Portland Oregon, United States, June 2011, https://hal.inria.fr/inria-00637991.

[1042] K. Striegnitz, A. A. J. Denis, A. Gargett, K. Garoufi, A. Koller, M. Theune, "Report on the Second Second Challenge on Generating Instructions in Virtual Environments (GIVE-2.5)", *in : 13th European Workshop on Natural Language Generation*, Nancy, France, September 2011, https://hal.inria.fr/inria-00636498.

[1043] A. Tebbakh, N. Dugué, J.-C. Lamirel, "Feature selection and complex networks methods for an analysis of collaboration evolution in science: an application to the ISTEX digital library", *in : 5th International Symposium ISKO-MAGHREB 2015*, *Proceedings of 5th International Symposium ISKO-MAGHREB 2015*, Hammamet, Tunisia, November 2015, https://hal.inria.fr/hal-01263876.

Secondary International Conferences

[1044] C. Cerisara, "Semi-supervised experiments at LORIA for the SPMRL 2014 Shared Task", *in : Proc. of the Shared Task on Statistical Parsing of Morphologically Rich Languages*, Dublin, Ireland, August 2014, https://hal.archives-ouvertes.fr/hal-01109886.

[1045] S. Cruz-Lara, "Mundos virtuales: Una infraestructura global para facilitar las interacciones sociales multilingües y el aprendizaje de idiomas", *in : The VII International Conference on Engineering and Computer Education. Workshop "New trends in Engineering Education"*., COPEC Science and Education Research Council, Guimaraes, Portugal, September 2011, https://hal.inria.fr/hal-00694454.

[1046] P. Cuxac, J.-C. Lamirel, V. Bonvallot, "Approche semi-supervisée pour la désambiguïsation des affiliations dans les bases de données bibliographiques", *in : 2nd International Symposium ISKO Maghreb - 2012*, Hammamet, Tunisia, November 2012, https://hal.inria.fr/hal-00781587.

[1047] P. Cuxac, J.-C. Lamirel, "Analyse des évolutions et des interactions entre domaines scientifiques: GRAFSEL, association de la sélection de variables et de la représentation graphique", *in : VSST 2013 - 7e colloque de Veille Stratégique, Scientifique et Technologique*, Nancy, France, October 2013, https://hal.archives-ouvertes.fr/hal-00962312.

[1048] K. Hajlaoui, P. Cuxac, J.-C. Lamirel, C. François, "Aide à l'expertise des brevets par alignement avec les publications scientifiques", *in : 3ème Séminaire de Veille Stratégique, Scientifique et Technologique*, Ajaccio, France, May 2012, https://hal.archives-ouvertes.fr/hal-00959396.

[1049] J.-C. Lamirel, P. Cuxac, K. Hajlaoui, "Une nouvelle approche pour la sélection de variables basée sur une métrique d'estimation de la qualité", *in : EGC 2014*, *Proceedings of EGC 2014*, Rennes, France, January 2014, https://hal.inria.fr/hal-01263843.

[1050] J.-C. Lamirel, P. Cuxac, "Détection des évolutions dans les corpus textuels", *in : EGC 2012 CIDN Workshop*, Bordeaux, France, January 2012, https://hal.inria.fr/hal-00781582.

[1051] J.-C. Lamirel, P. Cuxac, "Amélioration de la qualité des résultats des classifieurs par la métrique de maximisation d'étiquetage", *in : ISKO-MAGHREB*, Marrakech, Morocco, November 2013, https://hal.inria.fr/hal-00939041.

[1052] J.-C. Lamirel, P. Cuxac, "Une nouvelle méthode statistique pour la classification robuste des données textuelles : le cas Mitterand-Chirac", *in : JADT 2014, Proceedings of JADT 2014*, Paris, France, April 2014, https://hal.inria.fr/hal-01263847.

[1053] J.-C. Lamirel, R. Mall, M. Ahmad, "Comportement comparatif des méthodes de clustering incrémentales et non incrémentales sur les données textuelles hétérogènes", *in : 11th International Francophone Conference on Knowledge Extraction and Management (EGC 2011)*, Brest, France, January 2011, https://hal.inria.fr/hal-00645398.

[1054] J.-C. Lamirel, "Une nouvelle méthodologie diachronique pour automatiser l'analyse de la dynamique des domaines de recherche", *in : 1st. International Symposium ISKO-Maghreb*, Hammamet, Tunisia, May 2011, https://hal.inria.fr/hal-00645397.

[1055] J.-C. Lamirel, "Nouveaux indices de qualité de clustering basés sur la maximisation de traits", *in : SFC'15, SFC'15*, Nantes, France, September 2015, https://hal.inria.fr/hal-01263871.

[1056] F. Lefèvre, D. Mostefa, L. Besacier, Y. Estève, M. Quignard, N. Camelin, B. Favre, B. Jabaian, L. Rojas-Barahona, "Robustness and portability of spoken language understanding systems among languages and domains : the PORTMEDIA project", *in : Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 1: JEP*, ATALA/AFCP, p. 779–786, Grenoble, France, 2012, https://hal-amu.archives-ouvertes.fr/hal-01194257.

National Peer-Reviewed Conferences

[1057] C. Anderson, C. Cerisara, C. Gardent, "Vers la détection des dislocations à gauche dans les transcriptions automatiques du Français parlé / Towards automatic recognition of left dislocation in transcriptions of Spoken French", *in : Traitement Automatique des Langues Naturelles - TALN'2011*, p. 6, Montpellier, France, June 2011, https://hal.archives-ouvertes.fr/hal-00600510.

[1058] A. Denis, M. Quignard, D. Fréard, F. Détienne, M. Baker, F. Barcellini, "Détection de conflits dans les communautés épistémiques en ligne", *in : TALN - Actes de la Conférence sur le Traitement Automatique des Langues Naturelles - 2012*, Grenoble, France, June 2012, https://hal.inria.fr/hal-00766397.

[1059] J.-C. Lamirel, P. Cuxac, "Nouvelles méthodes statistiques pour le traitement des données textuelles volumineuses et changeantes", *in : Premières Rencontres Scientifiques du Réseau Mixte LaFEF (Langue Française et Expressions Francophones)*, Strasbourg, France, November 2013, https://hal.inria.fr/hal-00939040.

Books or Proceedings Editing

[1060] F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, P. Perrier (editors), *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech), 25-29 August 2013, Lyon (France)*, France, International Speech Communication Association (ISCA), August 2013, over 3500 pagesp., https://hal.archives-ouvertes.fr/hal-00931864.

Book chapters

[1061] P. Bedaride, C. Gardent, "Non Compositional Semantics Using Rewriting", *in : Human Language Technology. Challenges for Computer Science and Linguistics*, Z. Vetulani (editor), *Lecture Notes in Computer Science*, *6562*, Springer Verlag, February 2011, p. 257–267, https://hal.inria.fr/hal-00639843.

[1062] S. Cruz-Lara, J. M. Cabello, T. Osswald, A. Collado, J. M. Franco, S. Barrera, "Tourism in Virtual Worlds: Means, Goals and Needs", *in : Handbook of Research on Practices and Outcomes in Virtual Worlds*, H. H. Yang and S. C.-Y. Yuen (editors), IGI Globals, 2011, https://hal.inria.fr/inria-00580241.

[1063] S. Cruz-Lara, A. A. J. Denis, N. Bellalem, L. Bellalem, "Handbook on 3D3C Platforms: Applications and Tools for Three Dimensional Systems for Community, Creation and Commerce", *in : Handbook on 3D3C Platforms: Applications and Tools for Three Dimensional Systems for Community, Creation and Commerce*, Y. Sivan (editor), Springer International Publishing, 2016, p. 507, https://hal.inria.fr/hal-01256593.

[1064] S. Cruz-Lara, T. Osswald, J.-P. Camal, N. Bellalem, L. Bellalem, J. Guinaud, "Enabling Multilingual Social Interactions and Fostering Language Learning in Virtual Worlds", *in : Handbook of Research on Practices and Outcomes in Virtual Worlds*, H. H. Yang and S. C.-Y. Yuen (editors), IGI Globals, 2011, https://hal.inria.fr/inria-00580219.

[1065] P. Cuxac, J.-C. Lamirel, "Clustering incrémental et méthodes de détection de nouveauté : application à l'analyse intelligente d'informations évoluant au cours du temps", *in : La recherche d'information en contexte :Outils et usages applicatifs*, L. Grivel Ed, October 2011, p. 00, https://hal.archives-ouvertes.fr/hal-00962376.

[1066] C. Gardent, Y. Parmentier, G. Perrier, S. Schmitz, "Lexical Disambiguation in LTAG using Left Context", *in : Human Language Technology. Challenges for Computer Science and Linguistics. 5th Language and Technology Conference, LTC 2011, Poznan, Poland, November 25-27, 2011, Revised Selected Papers*, Z. Vetulani and J. Mariani (editors), *Lecture Notes in Computer Science (LNCS) series / Lecture Notes in Artificial Intelligence (LNAI) subseries*, *8387*, Springer, July 2014, p. 67–79, https://hal.inria.fr/hal-00921246.

[1067] C. Gardent, "Syntax and Data-to-Text Generation", *in : Lecture Notes in Computer Science*, *8791*, 2014, p. 3 – 20, https://hal.archives-ouvertes.fr/hal-01109617.

[1068] G. Jacquet, J.-L. Manguin, F. Venant, B. Victorri, "Construire le sens d'un verbe dans son cadre prédicatif", *in : Au commencement était le verbe. Syntaxe, Sémantique et Cognition*, N. L. Q. F. Neveu, P. Blumenthal (editor), Peter Lang, 2011, p. 233–252, https://halshs.archives-ouvertes.fr/halshs-00666615.

[1069] J.-C. Lamirel, P. Cuxac, C. François, "Recherche des évolutions technologiques à partir de bases de données bibliographiques : apport de la classification incrémentale", *in : Technologies de la Connaissance et Recherche d'Information en Contexte*, S. Collection (editor), Hermes Science Publishing Ltd, 2011, https://hal.inria.fr/inria-00535972.

Other Publications

[1070] G. Jacquet, J.-L. Manguin, F. Venant, B. Victorri, "Déterminer le sens d'un verbe dans son cadre prédicatif", JACQUET G., MANGUIN J.L., VENANT F., VIC TORRI B., Construire le sens d'un

verbe dans son cadre pré di catif, in F. Neveu, P. Blu menthal, N. Le Querler (éds.), Au com men cement était le verbe. Syntaxe, Séman tique et Cog nition, Peter Lang, 2011, à paraître, July 2011, https://hal.archives-ouvertes.fr/hal-00611665.

[1071] J.-C. Lamirel, "A systemic approach for mining complex data", 2011, Conférence invitée à l'université National Cheng Kung University (NCKU), Taiwan, https://hal.inria.fr/hal-01263769.

[1072] J.-C. Lamirel, "A new approach for improving clustering quality based on connected maximized cluster features", 2012, Conférence invitée au Classification Society 2012 Annual Meeting, Carnegie Mellon University, Pittsburgh, PA, USA, https://hal.inria.fr/hal-01263775.

[1073] J.-C. Lamirel, "A new incremental clustering approach based on cluster data feature maximization", 2012, Conférence invitée au Classification Society 2012 Annual Meeting, Carnegie Mellon University, Pittsburgh, PA, USA, https://hal.inria.fr/hal-01263778.

[1074] J.-C. Lamirel, "An unbiased symbolico-numeric approach for reliable clustering quality evaluation", 2012, Conférence invitée au Classification Society 2012 Annual Meeting, Carnegie Mellon University, Pittsburgh, PA, USA, https://hal.inria.fr/hal-01263781.

[1075] J.-C. Lamirel, "Opening new perspectives for classifying and mining textual and web data", 2012, Conférence invitée au 2nd International Symposium ISKO Maghreb 2012, Hammamet, Tunisia, https://hal.inria.fr/hal-01263783.

[1076] J.-C. Lamirel, "Analysis of evolution and interactions between science fields: the cooperation between feature selection and graph representation", 2013, Conférence invitée au ISSRAS Conference, Moscow, Russia, https://hal.inria.fr/hal-01263795.

[1077] J.-C. Lamirel, "Exploiting feature maximization metric in the framework of webometrics and scientometrics", 2013, Conférence invitée au 9th International Conference on Webometrics, Informetrics and Scientometrics and 14th COLLNET meeting, Tartu, Estonia, https://hal.inria.fr/hal-01263788.

[1078] J.-C. Lamirel, "Nouvelles méthodes statistiques pour le texte", 2013, Conférence invitée au CNPLET Annual colloquium : La Néologie, les corpus informatisés et les processus d'élaboration des langues de moindre diffusion", Ghardaïa, Algeria, https://hal.inria.fr/hal-01263800.

[1079] J.-C. Lamirel, "Feature maximization: a new incremental and flexible approach for the efficient analysis of large and dynamic data collections", 2014, Conférence invitée au Stadius Seminar, ESAT, KU-Leuven, Leuven, https://hal.inria.fr/hal-01263803.

[1080] J.-C. Lamirel, "Feature selection an graph representation for an analysis of science fields evolution: an application to digital library ISTEX", 2015, Conférence invitée au 11th International Conference on Webometrics, Informetrics and Scientometrics and 16th COLLNET meeting, Delhi, India, https://hal.inria.fr/hal-01263830.

[1081] J.-C. Lamirel, "Les mots : tenseurs de sons – vecteurs de sens", 2015, Conférence invitée au Colloque international MUTTUM, MUTMUT, MU, MUSIQUE, Plaisir, France, https://hal.inria.fr/hal-01263822.

[1082] J.-C. Lamirel, "New metrics and related statistical approaches for efficient mining in very large and highly multidimensional databases", 2015, Conférence invitée au 11th IEEE International Conference: Beyond Databases, Architectures and Structures (BDAS'2015), Krakow, Poland, https://hal.inria.fr/hal-01263816.

02

Project

Loria

Department 4

Natural Language Processing & Knowledge Discovery

Department Head: Bruno Guillaume

Department project

The department 4 of the LORIA, named NLPKD (Natural Language Processing & Knowledge Discovery) is composed of 48 permanent members in 7 teams with activities in Natural Language Processing (both text and speech are considered), Knowledge Discovery and Processing, and Document Processing. In this document, we will give perspectives for scientific activities in this department for the next five years. We will first review the important facts about the local environment concerning the scientific fields of the department. Later, we will described what are the main changes that can be expected in terms of methods or of application domains. We then focus on the expected changes in the teams organization and consider what are the strengths and weaknesses of the department and give perspectives concerning organizational evolution of the department. The last part presents the detailed scientific project of each team of the department.

1 Evolutions of the department

Environment

For a long time, the Lorraine is known for its activities in on Natural Language Processing (either in linguistics with the ATILF laboratory or in computer science with the LORIA) and in Knowledge Processing. We give here the most important ongoing projects that are strongly linked to the local environment. We believe that each of these items are examples of environment that will have a positive effects on the activities of the department in terms of local collaborations and applications (either in social sciences or in industrial context).

- ORTOLANG is an EQUIPEX (Équipement d'excellence) that offers a network infrastructure with NLP resources (corpus, lexicons, dictionaries, etc.) and NLP tools. The main goal is to ensure the availability and the quality of the available resources. The long term goal is to place the NLP on the French language to be at the highest international level and to ease the use and the transfer of resources and tools developed within government laboratories to industrial partners. It must also promote French language through sharing knowledge about our language accumulated by public laboratories.

- The Lorraine University and CNRS are strongly implied in the ISTEX (Initiative d'excellence de l'Information Scientifique et Technique) project. The ISTEX project's main objective is to provide the whole higher education and research community online access to retrospective collections of scientific literature in all disciplines by setting up a national document acquisition policy covering journal archives, databases, texts corpora etc. ISTEX will provide a platform with systematic access to full text documents and to data processing services: data extraction, text mining, production of document syntheses and terminological corpora, etc.

- The CPER (Contrat de Plan État/Région) [2015-2020] contains two projects related to computer science. One of these project is called LCHN (French acronym for "Language, Knowledge and Digital Humanities") is leaded by ATILF and LORIA but it implies 85 permanent members of 11 laboratories of Lorraine University. The total funding of the project is 2.46M€ (a large part of it is reserved for materials). This funding will be used to install new research platforms for NLP (with GPU computer adapted to deep learning applications), for Knowledge extraction and for Digital Humanities.

- The project ISITE LUE (Lorraine Université d'Excellence) starts in 2016 for 10 years. One of challenge chosen by Lorraine University is "Natural Language and Knowledge engineering". The goal of this challenge is to stimulate transversal activities implying engineering and Digital Humanities or implying engineering and industries. The project will provide PhD thesis funding (two in 2016) and other funding in the next years.

- Lorraine University is implied in the two ERASMUS Mundus programs: the LCT (Language and Communication Technologies) program and in the EMLex (European Master in Lexicography). Thanks to the two master, each year several good foreign students are in Nancy to attend to one of these master. Many members of department 4 are implied in teaching the LCT Master program.

Scientific evolutions in the department

We want here to highlight evolutions which concern several teams of the department; please refer to detailed team project below for a more precise scientific focus.

The main evolution we can expect for the coming period is the development of deep learning methods. In recent years, deep learning methods were successfully used in several areas of interest for our department. These methods are already used in **MULTISPEECH** activities (for source separation and speech recognition) and there will be given a stronger focus in the next years. In natural language processing, they also become one of the mostly used approach thanks to their very good performances, but also to their capacity to automatically derive high level features from raw textual data. **SYNALP** will explore the use of these methods both for natural language understanding and for natural language generation. The deep learning is also nowadays heavily used in machine translation and in document processing; hence, **SMART** and **READ** teams will also probably be concerned by this evolution.

As described above in the environment of the department, the CPER LCHN [2015-2020] will mainly bring funding for material. In 2016, a new platform dedicated to "Language Engineering" composed of a cluster of GPU (for 60 k€) will be installed in the LORIA. GPU computers are essential for experimentation in deep learning, this new cluster will be a strong help for the new activities around these methods cited above. A new platform for "Digital Humanities" is also planned in 2016 and 2017 in the same CPER; this platform will be installed in the ATILF laboratory but it will be opened to members of the department working in theses topics. Amongst other platforms that will have an impact the department activities, we can cite a "Data Science" platform with a new engineer who will reinforce activities in this domain, mainly in **ORPAILLEUR** with applications to data exploration in Life science and, finally,

the articulatory data acquisition platform (an articulograph in the LORIA and MRI acquisition at our INSERM partner) which is important for articulatory modeling researches in **MULTISPEECH**.

A part of department activities in the next period will be conducted in some new projects; we want to highlight three of them: METAL, CrossCult and AMiS. The e-FRAN METAL project (Modèles Et Traces au service de l'Apprentissage des Langues) [DATE?] is headed by the Kiwi team of department 5 but two teams of department 4 are implied: **SYNALP** will contribute to the project with applications of natural language generation to automatically generation of grammar exercises and **MULTISPEECH** for the development of a 3D talking head for supporting learning of a foreign language. The CrossCult[1] project [2016-2020] is a large European project which aims at gathering many kind of historical data and cultural data from different European countries to build new applications for large audience or for museum visitors for instance. The **ORPAILLEUR** teams is involved in this project to work on the knowledge extraction part of the system. Finally, The Chist-Era[2] project AMIS (Access Multilingual Information opinionS) [2016-2019] is headed by the **SMART** team. The goal of the project is to develop a multilingual help system of understanding without any human being intervention, using techniques from several areas: text, audio and video summarization; automatic speech recognition; machine translation; cross-lingual sentiment analysis. Members of the **QGAR** team are also implied in the project.

Teams evolution

MULTISPEECH is a new EPC Inria since July 2014 which follows the **PAROLE** team. During the last five years, two new Loria teams appeared with former members of **PAROLE** (**SYNALP** in 2012 and **SMART** in 2015). Even if the **MULTISPEECH** team is large (13 permanent members), it is a well structured team which is visible and recognized nationally and internationally. The team is attractive and there are often good applicants for permanent position recruitment. It is highly probable that **MULTISPEECH** will continue to be one of the most active French group in many aspects of Speech Processing of the coming years.

SMART is a small and young team with 3 permanent members and works in a well-identified area: Machine Translation. Being leader of the new ChistEra project AMIS, **SMART**will be able to work in collaboration with other European teams.

SYNALP is a medium team with a good visibility and which is involved in large projects mainly around natural language generation (WebNLG, e-FRAN METAL). No particular evolution is planned for the next period.

SÉMAGRAMME is an EPC Inria since 2011. During fall 2015, the team was evaluated by an Inria committee and has defined scientific objectives for the next years with balance between theoretical works and a focus on implementation of a variety of semantic phenomena able to deal with discourse (through context modeling).

CELLO is a team where Hans Van Ditmarsch is the only permanent member. The team creation starts with the ERC grant of Hans in 2013; and this ERC will stop on February 2018. Hans has a large set of international collaborations and he will focus his research on the promising subject of gossip protocols. Nevertheless, the team will seek for a recruitment in the next two years but, if it fails to recruit, we will have to deal with this situation and find with Hans another team in which he will be able to continue his activity.

ORPAILLEUR is the second large team of the department (12 permanent members). Most of the members of **ORPAILLEUR** have activities focused on knowledge management or discovery. But they are also a few members that are active in more theoretical graph theory. **ORPAILLEUR** is an EPC Inria, and following the Inria process of team management, **ORPAILLEUR** will be stopped in 2019. Of

[1] http://www.crosscult.eu

[2] http://www.chistera.eu/projects/amis

course, this will be an opportunity for some members to create a new team and to define a new scientific project. However, it is too early to predict the new organization of current **ORPAILLEUR** members in the next year.

On the topic of document processing, there were two teams: **QGAR** and **READ**. **QGAR** has stopped its activities in 2016. Former members (4 persons) of **QGAR** will have to choose for their integration in a team of the laboratory in the next months. The second team working on this topic is **READ** is active and gives details about the evolution in the next five years (see below) but this team remain small (2 permanent members). In consequence, there is some uncertainty about the future of this theme in the laboratory.

Department organization

The department organizes a seminar with invited talks every two months both for members of the department and for students of the Erasmus Mundus master program in Natural Language Processing. We have observed during last years that the large area of scientific interest of the different teams makes it difficult to organize invited talks that spread over more then two or three teams of the department. Since 2016, we decide to have several seminars with a more precise scientific focus. Three seminars with three different organizers are planned: one centered on speech processing, one with a focus on NLP and one more centered on knowledge processing and related areas. The three organizers work with the department leader and the department assistant for the practical organization and to make the seminar planning.

Since 2015, a department day is organized each year where all PhD students are invited to present their research works to colleagues of the department. 13 PhD students in March 2015 and 12 in May 2016 had the opportunity to present their thesis work and to exchange with permanent members of the department. We had positive feedbacks from these two days and we plan to organize again this event in the next years.

Several meeting are organized each year for more administrative tasks. For instance, the department is in charge of the ranking of PhD applications and of the profiling of job for permanent positions. The department4 is the biggest one from the laboratory: there are 48 permanent members working in a wide area of topics. As a consequence of this, it is difficult to organize these meetings and to have an fair repartition of PhD funding amongst the teams: the number of PhD funding for the department given by University is between one and two each year. Job profiling is also complicated; due to the small number of permanent positions available it is often difficult to have an agreement about these profiles.

2 Strengths, Weaknesses, Opportunities, Threats.

We will not give teams specific descriptions for all these aspects but a synthesis of these at the department level. Hence, we will focus on elements concerning several teams.

Strengths

Scientific activities of the NLPKD department teams are highly visible both at the national and international level. Members of the department have a high quality publication activity and are involved in many contracts either with academic or industrial partners. The two big teams **MULTISPEECH** and **ORPAILLEUR** are very attractive for recruitment either for Inria or CNRS researchers or for University positions.

Weaknesses

The most frequent weakness identified by the teams of the department is the low number of PhD and post-doc funding. The ratio of number of PhD student by permanent members is much lower than 1 (during the evaluation period (5.5 years), 40 PhD were defended for more than 40 permanent members).

Another recurrent problem also concerns funding: national and international funding is more and more competitive. A lot of time is spent writing proposals, many of them being rejected.

Some specific activities of the department requires manpower: software development, data acquisition, annotation of the needed data either for system training or for systems evaluation. This needs are not satisfactory covered and when they are covered, it is most of the times by non-permanent staff; for software development for instance, engineers are employed for one or two years and they have to leave the laboratory after and then, the long term maintenance and development of software relies on permanent research staff.

Opportunities

For the coming years, we can hope that the actual research potential of the department and its attractiveness will be an opportunity for keeping a high level of scientific production. The emergence of new activities around the deep learning will also be an opportunity for new applications and the development of improved systems in our scientific domains. The audience of the several software tools developed in the department is not as high as expected and we can improve this aspects in the next period. The active environment described above will bring many opportunities of new collaborations at the regional levels either with academic or industrial partners.

Threats

The main threat identified is relative to the weakness cited above concerning funding. If the situation about PhD funding from the University and about the time spent for finding funding does not evolved in a positive way, it will be more and more difficult to conduct high level research. On several topics we are interested in, big companies (Google, Apple, Microsoft, IBM, Nuance, …) make large investments and have much more resources than us. Hence, with respect to complete systems, this is impossible to compete with such companies; we have to set the focus on more specific topics. These big companies are also very attractive for brilliant students or researchers, it is therefore more difficult for us to recruit new members.

	Positive	Negative
Intern	Strengths: • National and international visibility • High quality publications • Contracts with industrial partners • Attractiveness	Weaknesses: • Few PhD and Post-doc funding • Time spent on writing proposals for funding • Lack of manpower for development and data annotation
Extern	Opportunities: • Emergence of deep learning based activities • Active environment described above • Development of new industrial applications of our research activities • Improvement of our software audience	Threats: • Worse funding situation (PhD, Post-doc and other projects) • Competition with big companies with much more resources

3 MULTISPEECH

Team composition: Anne Bonneau (CR CNRS), Martine Cadot (MCF, UL), Vincent Colotte (MCF UL), Dominique Fohr (CR CNRS), Irina Illina (MCF, UL), Denis Jouvet (DR Inria), Yves Laprie (DR CNRS), Antoine Liutkus (CR Inria), Odile Mella (MCF, UL), Slim Ouni (MCF, UL), Agnès Piquard-Kipffer (MCF, UL), Romain Sérizel (MCF, UL), Emmanuel Vincent (DR Inria).

The goal of the **MULTISPEECH** project is the modeling of speech for facilitating oral-based communication, with a particular focus on multisource, multilingual and multimodal aspects. The project is structured along three research axes: explicit modeling of speech, statistical modeling of speech and uncertainty in speech processing.

Explicit modeling of speech production and perception

Speech signals result from the movements of articulators. A good knowledge of their position with respect to sounds is essential to improve articulatory speech synthesis and the relevance of the diagnosis and feedback in computer assisted language learning. Production and perception processes are interrelated, so a better understanding of how humans perceive speech will lead to more relevant diagnoses in language learning as well as pointing out critical parameters for expressive speech synthesis. The expressivity translates into both visual and acoustic effects that must be considered simultaneously to produce expressive speech synthesis.

Articulatory modeling. The main objective is to continue improving the articulatory synthesis by controlling the fine temporal evolution of the vocal tract geometry at supra-segmental and segmental levels and by improving acoustic simulations. This requires the design of better MRI acquisitions protocols to capture the vocal tract shape dynamically together with the corresponding denoised speech signal, and/or with an increased tridimensional precision. This area of research relies on the equipment available in the laboratory to acquire articulatory data (articulograph Carstens AG501) and strong cooperation with IADI lab (INSERM U947, Nancy) specialized in MRI acquisitions and LPP in Paris for acquisition of source parameters.

Expressive acoustic-visual synthesis. In our approach, the acoustic-visual text-to-speech synthesis is performed by considering bimodal signal units comprising both acoustic and visual channels. An important goal is to provide an audiovisual synthesis that is intelligible, acoustically and visually. Thus, we continue working on adding realistic visible components of the head as lips and tongue. To acquire the audio-visual data, we consider using marker-based and marker-less motion capture systems (articulograph, vicon, kinect). Another goal is to add expressivity through the acoustic signal (prosody aspects) and facial expressions (head and eyebrow movements).

Categorization of sounds and prosody for native and non-native speech. Discriminating speech sounds and prosodic patterns is the keystone of language learning whether in the mother tongue or in a second language. This issue is associated with the emergence of phonetic categories (classes of sounds related to phonemes, and prosodic patterns). Studies on the emergence of new categories, in the mother tongue (for children with hearing deficiencies) or in a second language, rely upon studies on native and non-native acoustic realizations of speech sounds and prosody.

Positioning. A strong asset of **MULTISPEECH** is its multidisciplinary expertise covering signal processing, statistical modeling, speech production (and acquisition of articulatory data), speech perception, and phonetics. Such positioning has been adopted by a number of teams abroad (e.g., UCLA and USC), but is unique in France where other teams work typically either only on automatic speech recognition (e.g., LIUM, LIA or LIMSI) or only on phonetics (e.g., LPP, LPL). There are several tracks of research in articulatory modeling: superficial modeling via statistical models linking directly acoustic and vocal tract geometry (GIPSA-Lab, NTT CS-Lab, CSTR), bio-mechanical modeling and finite element methods (TIMC-IMAG) which requires huge calculation resources and medical data hard to collect, and geometrical modeling coupled with acoustical simulations (TUD). Concerning audiovisual synthesis, GIPSA-Lab is using bio-mechanical and articulatory-visual-based approaches. LIMSI and ParisTech are working on close topics but focused on cognitive aspects of human interaction. Internationally, this field is highly competitive with large groups carrying research (Toshiba UK, Microsoft China) and few academic labs (KTH, DSSP, VUB, UEA).

Statistical modeling of speech

Statistical approaches achieve performance that makes possible their use in actual applications. However, speech recognition systems still have limited capacities (e.g., even if large, the vocabulary is limited) and their performance drops significantly when dealing with degraded speech. At the signal level source separation approaches are investigated for enhancing the speech signal and for noise robust speech recognition. At the language level, handling new proper names is a critical aspect that is tackled, along with the use of statistical models for speech production. Deep Neural Network (DNN) modeling is used in several topics including source separation and acoustic modeling for speech recognition.

Source separation. The focus is set on techniques using multiple microphones and/or models of non-speech events. Some of the challenges include getting the most of new modeling frameworks based on alpha-stable distributions and deep neural networks, and combining them with established spatial filtering approaches, modeling more complex properties of speech and audio sources (phase, inter-frame and inter-frequency properties), … Beyond the definition of such models, the difficulty will be to design scalable estimation algorithms robust to overfitting, that will integrate into the recently developed FASST and KAM software frameworks.

Linguistic modeling. A few specific topics are investigated, such as the processing of proper names. The challenge is to dynamically adjust the lexicons and/or the language models through the use of the context of the documents or of some relevant external information possibly collected over the web. We also want to introduce into the models additional relevant information such as linguistic features, semantic relation, topic or user-dependent information.

Speech generation by statistical methods. Our goal is to combine statistical speech synthesis (which offer the possibility to deal with small amounts of speech resources, and flexibility for adapting to new emotions or speakers), with corpus-based speech synthesis (which leads to the highest speech quality). This will be applied to producing expressive audio-visual speech.

Positioning. **MULTISPEECH** is among the few French research groups working on large vocabulary speech transcription (along with LIMSI, LIUM, LIA, …). At the international level, quite many laboratories carry research in speech recognition, among which we can cite university laboratories (RWTH, CU, CMU, …), research institutes (SRI, IDIAP, …), as well as research laboratories from large IT companies (Google, Microsoft, IBM, AT&T, …). On the topic of audio source separation, FASST is one of the few

frameworks together with that of NTT in Japan to jointly exploit spatial and spectral cues, and we aim for FASST to remain one of the most advanced frameworks available. We will keep collaborating with the two other major research teams on source separation in France (PANAMA team of Inria Rennes, and audio team of Telecom ParisTech). At the international level, our expertise in audio source separation is widely acknowledged and we have close relationship with Northwestern University, Cork Institute of Technology, NII (Japan) and MERL (USA). Other teams include Tampere University of Technology, Queen Mary University of London, University of Kyoto, University of Illinois, …

Uncertainty estimation and exploitation in speech processing

Our general objective is to quantify the confidence (or uncertainty) in the output of a given speech processing technique and to exploit it for further processing. We are interested in such information regarding speech enhancement and source separation techniques, and for phonetic segment boundaries and prosodic parameters.

Uncertainty and acoustic modeling. One way of improving speech recognition on noisy data is to estimate the uncertainty on the separated sources in the form of their posterior distribution and to propagate this distribution through the subsequent feature extraction and speech decoding stages. **MULTISPEECH** has recently obtained more accurate estimates by training a non-parametric estimator from data, and used them successfully for acoustic decoding with GMM-HMM models. Next steps are to account for the correlation of distortions over time and frequency, which have not been considered so far, and to exploit them for acoustic decoding with DNN-HMM models.

Uncertainty and phonetic segmentation. Currently phonetic boundaries obtained through forced speech-text alignments are quite correct on good quality speech, but precision degrades significantly on noisy and non-native speech. Phonetic segmentation aspects will be investigated, both in speech recognition (spoken text unknown) and in forced alignment (spoken text known).

Uncertainty and prosody. Prosody parameters (fundamental frequency, duration of the sounds and their energy) provide information for structuring speech data, determining the modality of an utterance, determining accented words, ... Errors in estimating these parameters may lead to a wrong decision. **MULTISPEECH** will investigate estimating the uncertainty on the duration of the phones and on the fundamental frequency, as well as using it in subsequent processing.

Positioning. The idea of estimating the uncertainty in the feature domain after speech enhancement or source separation emerged 10 years ago at Microsoft (Redmond, USA) and has since then been studied by a few teams at the University of Bochum (Germany), the University of Paderborn (Germany), Aalto University (Helsinki, Finland), and NTT (Kyoto, Japan). We are the only French group carrying research on this problem, thanks to our joint expertise in speech enhancement and speech recognition. Our originality has been to recast it as a machine learning problem. Confidence measures associated with word recognition hypotheses have been largely investigated in the past, in many teams, including **PAROLE**. However, there is currently no similar information computed on the prosodic features (F0 estimation and phone and word segmentation).

Dissemination and transfer

We will continue investigating various ways of disseminating and exploiting the research results. Conventional ways include publications in conferences and journals, as well as disseminating the team knowl-

edge around speech through courses at Lorraine University (e.g., at master level), and participating to various public events (exhibition, science fair, presentation to students, …) We will also do our best to continue being involved in bilateral contracts with industrial partners, as well as in collaborative projects that also involve industrial partners. Some visualization tools and speech data will be made available to the community. We plan to use the ORTOLANG platform to boost the visibility of the tools and of data. Other toolkits, such as FASST and KAM will be enriched with new research advances. Another track for transfer of research results is the involvement in evaluation campaigns, as for example the involvement in organizing the CHiME and SiSEC challenges.

4 SMART

Team composition: Kamel Smaïli (Pr UL), David Langlois (MCF UL), Joseph Di-Martino (MCF UL).

SMART(Speech Modelisation and Text) is interested by the issues related to the modelisation of speech and text. The used approaches are based on machine learning techniques for which a training step is necessary. The training necessitates in general huge corpora. The corpora used in **SMART**are of three kinds: monolingual, parallel and comparable. The monolingual corpora are necessary for language modeling, the objective of this model is to propose a well-built sentence from a lattice achieved by a speech recognition or machine translation systems. For the next five years three objectives will be pursued.

Machine translation

The first one is to enhance the research on machine translation, research which we started in Loria in 2005. In fact, **SMART** means also Statistical Machine Translation. One of the objectives in machine translation is to propose another alternative for decoding a lattice. In fact, almost all the decoders today are based on A* algorithm, this approach has the drawback that the final solution is created from a huge number of partial solutions. This system produces generally solutions which are good at a local level but not satisfactory globally. We propose to build a solution by evolving a complete solution by using an evolutionary algorithm.

Under-resourced languages

The second objective is to handle the issue of under-resourced languages by proposing NLP tools for languages which are vernaculars. The languages on which we work are Arabic dialects: Algerian, Moroccan, Tunisian, Syrian and Palestinian. For this purpose, work is under progress and several resources have been developed: a multilingual dialect corpus (PADIC), a morphological analyzer, a diacritization system, a machine translation between all the pairs of languages of PADIC the pairs of language. The issue now is how to adapt the developed methods to other dialects and how to go beyond the limitation of the corpora since they are used to train language and translation models. Social networks will be used to extract more comparable corpora that will be investigated in order to extract some bilingual phrases necessary for the translation. Also these corpora will be used in **SMART** to study cross-lingual sentiments analysis and opinion mining. We will continue working in this area and especially in AMIS (Access Multilingual Information opinionS) a ChisTera project.

Health application

The third objective is how to use the skills we have in **SMART** for improving the recognition of pathological voices, language modeling and machine translation to help people who had a stroke attack in order to recover a better speech and linguistic elocution.

5 SYNALP

Team composition: Lotfi Bellalem (Prag UL), Nadia Bellalem (MCF UL), Christophe Cerisara (CR CNRS), Samuel Cruz-Lara (MCF UL), Christine Fay-Varnier (MCF UL), Claire Gardent (DR CNRS), Jean-Charles Lamirel (MCF UL)

Application domains

In terms of application domains, the **SYNALP** team members will continue to focus their efforts onto Natural Language Processing in general, in particular syntactico-semantic parsing and generation, as well as text clustering and dialog processing. However, these topics of research will be addressed by taking into account larger contextual, paralinguistic information. For instance, the emotion/sentiment/opinion carried on by the words stream influences the semantic content, and reciprocally. We have already proposed emotion-recognition approaches that rely on syntactic parse trees, but we could go further and jointly model emotion with semantic embeddings of sentences.

More generally, current sources of natural language do not generally provide a single stream of words, but the linguistic information is nearly always part of a broader set of related information: hyperlinks and Knowledge-base references on web pages, document structures and metadata in scientific papers, geolocalisation, timestamps, graphs of retweets, at-mentions and users on Twitter, etc. These multiple dimensions of the available information have to be considered jointly. Hence, we plan to investigate the combination of the graph of Twitter users that are related to some Twitter conversations with semantic vectors that represent the raw text.

Furthermore, the form of the linguistic input itself that we manipulate is also evolving, and we have to make our models more robust to such evolutions. For instance, new words and expressions are constantly appearing on the web, ungrammatical sentences are very common in oral dialogs, new abbreviations and emoticons appear on Twitter. We plan to partly address these issues thanks to character-level models that constitute a recent and robust alternative to word-based embeddings.

Fundamental methods

In terms of fundamental approaches, we plan to combine our skills and background on grammatical formalism with recent deep learning models, in particular sequence-to-sequence models, with the objective of automatically extracting and analyzing higher-level features that shall leverage our standard linguistically-motivated designs. Deep neural networks are very efficient to automatically infer hierarchically complex features from raw data, but their success largely depends on (i) a well-defined task; (ii) a linguistically-motivated model design and (iii) a large-enough dataset. Solving the latter issue may require the definition of auxiliary tasks from which generic distributed representations can be computed, and the use of data augmentation strategies. However, no such reliable data augmentation method exists so far in the NLP domain. We hence propose to capitalize on our expertise on text generation to propose new paraphrastic generation approaches that are specific to the target task and that will thus augment the amount of focused training data for deep models.

Transfer

We will emphasize our efforts in submitting and participating to regional, national and European projects. In addition to the international consortium that typically form the main partners in such projects, we also plan, whenever it is possible, to collaborate with regional companies. Hence, we have contacts with 3 start-ups in Lorraine:

- OneMuze in Baccarat, which manage descriptive databases for art;
- Xilopix in Nancy, which have developed a specialized search engine;
- Sesamm in Metz, which analyze tweets for financial predictions.

A contact with Xerox in Grenoble has also led to funding a CIFRE Ph.D. thesis.

In terms of software development, we will continue to favor open-sourcing our software whenever it is possible.

National and international positioning

Well established at the national level as an active NLP team, **SYNALP** is one of the few research units in France (with Grenoble and Paris) to have extensive expertise in the domain of Natural Language Generation. Recently, the team has developed a new focus on deep learning approaches for NLP, focus that we shall develop in the next five years in parallel and in combination with the NLG work.

The **SYNALP** team has good international visibility. Members of the team publish in the best conferences (EMNLP, COLING, ACL, EACL, Interspeech) and journal (Computational Linguistics) of the NLP and Speech Processing domain. They are regularly involved in Program Committees, often as chair or general chair (ESSLI, ENLG, SIGDIAL, StarSem, etc.), give keynotes at international events and participate in international steering committees such as SIGGEN, the Special Interest Group in Natural Language Generation of the Association for Computational Linguistics.

Collaborations at the national and international level are supported and enhanced by a regular participation in medium and large scale funded projects (ANR, ITEA, Interreg, Eurostars, PEPS) as well as by visits abroad and the occasional hosting of national and international researchers (U. of Alicante, Spain; U. of West Bohemia, Czech Republic; U. of Edinburgh, UK).

6 SÉMAGRAMME

Team composition: Maxime Amblard (MCF, UL), Philippe de Groote (DR Inria), Bruno Guillaume (CR Inria), Guy Perrier (Pr emeritus, UL), Sylvain Pogodalla (CR Inria).

The overall objective of the **SÉMAGRAMME** team 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: Syntax-semantics interface, Discourse dynamics and Common basic resources; which correspond to the current objectives of the team.

Syntax-Semantics interface

A system is said to be modular if it is made of relatively independent components. In the case of a semantic system, we will say that it is modular if the ontology on which it is based (including notions such as *truth*, *entities*, *events*, *possible worlds*, *time intervals*, *state of knowledge*, *state of believe*, ...) is obtained by combining relatively independent simple ontologies.

We believe that modularity is a requirement as important as compositionality. A large semantic system that would not be modular will be quite difficult to grasp and present little explanatory power. In addition, modularity plays an essential part in the incremental development of a semantic system. Typically, in the

absence of modularity, when studying a new semantic phenomenon, it will be problematic to decide whether and how the proposed treatment is compatible with the treatments of other phenomena.

In order to tackle the issue of modularity in semantics, we intend to develop a modular semantic interpretation of a large fragment (of English and/or French).

Graph rewriting based techniques are also explored as another robust and modular way to construct syntactic ans semantic representation of natural language utterances.

Positioning Most formal approaches to the semantic of natural language derive from Montague's work. The **SÉMAGRAMME** project-team clearly belongs to this stream. Nevertheless, contrarily to the traditional trend, we are not that much interested in determining model-theoretic truth conditions of utterances. We are more interested in the compositional process that allows a syntactic structure to be transformed into a logical formula of a given formal logic. In this sense, our approach is rather proof-theoretic. In addition, the use of the abstract categorial grammar framework allows us to be as generic as possible by taking into account different syntactic formalisms and different formal logics. In fact, both the syntactic formalism and the target logical language under consideration are parameters of our approach. On this topic, we collaborate with NII (Tokyo), Utrecht University and New-York University.

Discourse dynamics

What characterizes 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.

A context should not only consist of a set of discourse referents. It most also contain information about the accessibility and the salience of referents. In addition, it must record some logical knowledge that allows the discourse referents to be accessed by their properties.

At a foundational level, we intend 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?

Positioning When dealing with discourse dynamics, the widely used approach is Kamp's Discourse Representation Theory (DRT), and its variants such as segmented DRT. These theories act at a supra-sentential level and were not built, technically, as extensions of Montague semantics. As a consequence, several proposals have been made in order to accommodate DRT and Montague semantics. Most of these proposals consist in lifting up some aspects of Montague's framework at the level of DRT. We attack the problem the other way around, that is to express discourse structure in the same framework as Montague semantics, namely, higher-order logic. On discourse dynamics, we have collaboration with NII (Tokyo), New-York University and University of Chicago.

Common basic resources

Our ultimate goal in developing models and methods for the semantic analysis of utterances and discourses is to integrate them into natural language processing systems. To make this effective, we will need further primary resources. In particular, we will need lexicons annotated with semantic and pragmatic information. We intend to contribute to the development of such lexicons in a near future.

We also intend to construct linguistically annotated corpora. Our tools are used to automatically annotate data and these annotation are then validate by a human (using validation by an expert on non-expert validation though games with a purpose). Finally, these corpora are useful in the development of

our methods and tools, first for testing linguistic hypothesis against real data and for evaluation of our approaches.

Positioning Many NLP applications require digital linguistic resources such as treebanks, annotated corpora, or real size lexicons. Our team is active in producing such resources, mainly for the French language. This activity is usually carried on in collaboration with other French teams and with Ohio State University.

7 CELLO

Team composition: Hans van Ditmarsch (DR CNRS).

In the immediate future of the **CELLO** team the activities will remain focused on the European Research Council project Epistemic Protocol Synthesis (EPS 313360). The project will finish 1 February 2018. A rapidly developing theme in this project, initially not foreseen as a possible application of knowledge-based protocols, are gossip protocols. This topic intersects with the network community and dissemination of information in graphs, including graph connectivity transformations. The reason for optimism on further development is based on the appreciation within the logic research community of the results obtained over the past 2 years (2014/2015), as evidenced in citations, discussions, and workshops.

The coming 5 years we will attempt to further develop the topic of gossip protocols, including the obtention of additional funds (for example, an attempt to obtain an advanced ERC). This is anticipated to become the main focus of the team. Other areas of interest are: dynamics of knowledge and belief (in general), modal logics of propositional quantification (and their theoretic properties such as complexity and succinctness and axiomatization), and interactions of concurrency and knowledge.

The future of the **CELLO** team will depend to a large extent on the composition of the team. Currently, team leader Hans van Ditmarsch is the only permanent member.

Links with other national research establishments will be continued and strengthened, in particular with: IRIT, Toulouse and IRISA, Rennes. Concerning international positioning of the group we will attempt to institutionalize links with India (IMSc, Chennai, where Hans van Ditmarsch holds an associate position) and Iran (in view of the country opening up to research contacts with Europe; and given various contacts with PhD students and established Iranian researchers). In contact with Ramanujam, IMSc, and Benedikt Loewe, Amsterdam & Hamburg, a major bi-lateral Indian-European grant proposal is anticipated. Also in view of these goals Hans van Ditmarsch envisages a long-term stay at IMSc in Chennai in 2018, other obligations permitting. Productive links with various Chinese academic institutions may also be further developed (Peking Uni, Tsinghua, Sun Yat-Sen, South-West Uni).

8 ORPAILLEUR

Team composition: Miguel Couceiro (PR UL), Adrien Coulet (MCF UL), Esther Galbrun (CR Inria), Nicolas Jay (PR Faculté de Médecine, UL), Jean Lieber (MCF UL), Jean-François Mari (PR UL), Amedeo Napoli (DR CNRS), Emmanuel Nauer (MCF UL), Malika Smaïl-Tabbone (MCF UL), Chedy Raïssi (CR Inria), Jean-Sébastien Sereni (CR CNRS), Yannick Toussaint (CR Inria).

In the following, we present the research objectives of the **ORPAILLEUR** team for the next evaluation period. They are organized around three main themes in continuation with the past research lines of the team, namely: (1) Exploratory Knowledge Discovery guided by Domain Knowledge, (2) KDDK in Life Sciences, (3) Data and Knowledge Engineering.

Exploratory Knowledge Discovery

In the next years, we will continue our work on pattern mining and FCA, focusing on a "smart approach" to the problem of mining big and complex data for discovering reusable knowledge units. Such a smart approach relies on "data exploration" and is revisiting exploratory data analysis for defining "exploratory knowledge discovery". The main operations that are carried out on large masses of data and supporting knowledge discovery (KD) are: (i) data access and exploration, (ii) data mining, (iii) interpretation of the (discovered) patterns, (iv) data and pattern visualization. The knowledge discovery starts with data and returns patterns that can be then interpreted and represented as knowledge units, to be processed in turn in knowledge engineering.

Knowledge discovery is a flexible process and its results should reflect the nature of knowledge, i.e. discovering procedural or declarative knowledge units, and, by extension, meta-knowledge units when applying the process at the level of knowledge, e.g. mining Linked Open Data. Declarative approaches are more oriented towards pattern mining and conceptual knowledge discovery. The declarative approach involves exploration, interaction and iteration, but also the use of domain knowledge for guiding the mining process. Procedural approaches are more related to algorithmic concerns, e.g. one-pass, anytime and distributed algorithms. Actually, a combination of both declarative and procedural approaches should be carried out.

Moreover, with the proliferation of massive databases and new fields such as computational advertising, search engines and recommender systems, the needs for the construction of user preference models for classification and prediction purposes is parallel to the needs for adapted information retrieval and knowledge discovery processes. In addition, topics related to the management of data should also be revisited in exploratory knowledge discovery, such as data storage, data access, data indexing, querying, visualization and computational aspects. Accordingly, following the idea of putting the analyst at the center of the discovery process, we focus on the notion of preferences to help the analyst to explicitly declare his/her interests. In practice, pattern mining is based on a conjunction of constraints involving thresholds conditions. Setting the thresholds is a well-known issue and a way of minimizing the problem is to consider preferences of the analyst w.r.t. measures and to mine patterns that maximize preferences. This can be likened to computing the Pareto-front or skylines of the pattern space.

Three other directions of investigation should be still mentioned: mining equivalent pattern sets, segmentation of the data space and video game analytics. Data are usually of different natures, as they originate from various and distinct sources. A natural task is to identify the correspondences that exist between these different natures, which is the motivation of "redescription mining". Many potential applications exist in life sciences and social networks. The data space is the basis from which is built the search or pattern space that should be traversed in the most intelligent and efficient ways for discovering the best patterns. The data segmentation relies on internal properties of the data (e.g. equivalences, congruences, dependencies) or on domain knowledge (e.g. feature selection, projections). Accordingly, segmentation techniques for guiding the reduction of the data space in the most efficient and knowledgeable way are under study. Finally, in the last years, we worked on the discovery of strategies in real-time strategy games through pattern mining. Indeed, the video game industry has grown enormously over the last twenty years, bringing new challenges. We are currently studying new sequential pattern mining approaches and measures to understand how a strategy is likely to win.

KDDK in life sciences

In this domain, we are trying to apply knowledge discovery methods for assessing and enriching theoretical (published in literature) and practical (clinical) domain knowledge in pharmacogenomics (PGx). A main objective is to validate theoretical knowledge in PGx on the basis of practice-based or clinical knowledge extracted from electronic health records (EHRs). Actually knowledge units in PGx take

the form of ternary relationships, i.e. (gene variant, drug adverse event), and can be formalized thanks to biomedical ontologies. To reach such an assessment, theoretical knowledge should be extracted from PGx databases and literature. In parallel, clinical knowledge units should be extracted from EHRs. Then, these two kinds of knowledge units should be aligned for assessment. For example, new knowledge units can be confirmed thanks to omics databases, leading to the understanding of molecular mechanisms underlying and explaining drug adverse events (and enabling personalized medicine). Such an analysis of EHRs data is also related to aggregation theory and subgroup discovery. There are two key tasks in the analysis of EHR data: clustering patients into subgroups and designing supervised classification and prediction models for decision support. The use of aggregation theory will be investigated for completing these two tasks.

Finally, the combination of symbolic and numerical data-mining approaches remains a challenge and is not yet correctly solved. Such a combination is particularly interesting in life sciences for mining complex and heterogeneous data. Hybrid methods are based on a combination of supervised numerical methods (SVM, Random Forests) and unsupervised symbolic methods (pattern mining, clustering). The later are used to interpret and visualize the classes built by the former. Extensions to complex data based on graphs are of main importance.

Data and Knowledge Engineering

Here three main problems are considered: Autonomous Knowledge Discovery (AKD), reuse of experiences and recommendation. The objective of the first line of research is to introduce knowledge discovery and knowledge engineering techniques to support security analysis of computer networks and systems. Computer networks are dynamic environments composed by many entities holding thousands of activities, and requiring configuration changes to satisfy new or upgraded services. Such dynamics highly increase the complexity of security management. Even if automated tools help to simplify security tasks there is a need for advanced and flexible solutions able to assist security analysts in better understanding what is happening inside networks. Then, it becomes important to understand to which extent knowledge discovery can enrich and advance the state of the art of vulnerability management techniques, especially for anticipating and remedying security vulnerabilities, by providing specific and multi-dimensional classification of vulnerabilities, in agreement with domain knowledge. Then the links between current security standard languages and security ontologies should be investigated. The integration of such research results will help to develop knowledge discovery techniques able to deal with cyber security threats.

Today, a great deal of decision-making knowledge and know-how is held in large sets of resources on the web which are readily available but hardly reusable. The objective of this line of research is to capture experience, i.e. extract data and discover knowledge from these data, for solving complex human problems such as adaptation and decision-making. We are planning to combine a targeted data extraction/knowledge discovery process and reasoning techniques such as case-based reasoning (CBR) to discover and reuse hidden knowledge units by dynamically integrating, analyzing and adapting (possibly large) sets of resources on the web.

Following an alternative line, Recommender Systems (RSs) can be considered as information retrieval systems whose main goal is predicting which items from a catalog would a user consume or like in the future, given the characteristics of the item and the history of the user interactions with the system. RSs are present now in a very large range of applications and social activities. An important aspect of RSs is to be able to contextualize the output and to explain a user why an item is recommended. For that, we are carrying studies on the discovery of functional dependencies and biclustering which are closely related and directly linked to recommendation. For example, finding biclusters with similar column values within a numerical context is an example of the output expected by a recommender system. Taking into account user profiles and domain knowledge will have an effective impact on this line of research.

Transfer

The research activities within the team are always two-sided with theoretical and practical developments. Thus, there is a strong activity in the development of systems in the framework of industrial, national and international research and development projects (whose funding supports the research activities of the team). Below, we mention some particular technology transfer activities: (i) The industrial "BioIntelligence" project was aimed at designing software modules for helping the biological daily practice and guiding knowledge discovery in biology and bio-medicine (project involving Dassault Systèmes and pharmaceutical industrial partners). (ii) The objective of the Aetheris start-up is to provide travel planning methods based on data mining and preferences techniques (`http://www.loria.fr/~raissi/Aetheris/`). (iii) The Harmonic Pharma start-up is working on the integration of data mining modules in life sciences systems (`http://www.harmonicpharma.com`).

Mediation

The research activities and results obtained by the **ORPAILLEUR** team w.r.t. publications and development of software provide the team an original position. The team is solicited for participating in many new projects and is already contributing to many research projects. In addition, the members of the **ORPAILLEUR** team are all involved, as members or as head persons, in national and international research groups, and evaluation structures, in the organization of conferences, as members of conference program committees and as members of editorial boards, and in the organization of journal special issues.

Positioning

Positioning is made precise in the Inria Activity Report of the team, where the on-going research collaborations are mentioned (and in the HCERES document describing the team). We would like to make precise that the team has very close relations with Inria Teams in France, but also with related French teams in French universities as well. International collaborations are also very important, with America (North and South), Europe, and Russia.

9 READ

Team composition: Abdel Belaïd (Pr UL), Yolande Belaïd (MCF UL).

Transfer

The research in **READ** will evolve in the next five years following three axes: 1) proposal of generic techniques for the evaluation of recognition systems, whatever the script and the layout of documents, accompanied by annotation and ground truth construction techniques, and metrics tailored to the measure of system performance, 2) pursue the active learning in semi-supervised manner and evolve it to deep learning, and 3) evolving the recognition methods of Arabic by taking better account of the morphology and linguistic context. Efforts are still needed for the extraction of descriptors adapted to the morphology of the Arab, and taking them into account in machine learning systems. Dynamic Bayesian Networks techniques are an interesting way to keep exploring. Our collaboration with LATICE team ENSIT (University of Tunis) on the extraction of histogram oriented gradient (HOG) for the identification of the script, and on the use of Dynamic Bayesian Networks will continue with diversification of the scriptures. We also continue to study the problems of segmentation of Arabic documents into lines and words, strengthening our extraction techniques baseline (very difficult in Arabic because of the distortion of the words) and detection and separation of overlapping and attachment lines.

Mediation - Dissemination

READ team adheres to various circuits of research and document industry for mediation and dissemination of his research. First, through the project OSEO DoD, where ITESOFT-Yooz company is in charge of the operation and diffusion techniques developed by **READ**. Then, our collaboration working with the INIST, in the ISTEX-DATA project, will allow us to test our evaluation techniques of OCRs and experiment them on their datasets. Moreover, because of our membership of the international community in handwriting recognition and document analysis and recognition, we will continue to publish our research in prestigious conferences in the field as ICDAR, DAS, ICFHR and ICPR, and we will participate in various competitions in these conferences.

National and international Positioning

At national level, the **READ** team is part of the writing communication group (GRCE) where he participated in theme days and conferences by chairing sessions, proofreading papers, etc. At international level, we have already participated in the organization of three international conferences (ICDAR), we will continue to be part of the editorial boards of these conferences, chairing sessions, to be in the editorial board of several journals of our field. We are part of two Technical Committees of IAPR: TC10 (graphic recognition) and TC 11 (Reading systems).

Report integrators: Éric Domenjoud (Team ADAGIo) and Philippe Dosch (Team QGAR). Report designed under Linux using Emacs, and formated thanks to XɘLATEX.