



## CHIST-ERA Conference 2014

### Participate in CHIST-ERA Call 2014 Definition

The CHIST-ERA ERA-NET is a consortium of funding organisations mainly in Europe with programmes supporting ICST. The consortium is itself supported by the European Union's Future & Emerging Technologies scheme (FET). CHIST-ERA promotes multidisciplinary and transnational ICT research

with the potential to lead to significant breakthroughs. The funding organisations jointly support research projects selected in the framework of CHIST-ERA. The **Call 2014**, to be published in October, addresses two new and emerging topics:

**Resilient Trustworthy Cyber-Physical Systems**

**Human Language Understanding**

The **CHIST-ERA Conference 2014** in Istanbul (Turkey), June 17-18, brings together scientists working in these research areas and CHIST-ERA representatives to refine the topics contour and scope of the Call 2014.

The topic keywords illustrate the topics, but will be refined. All attendees will join plenary and facilitated break-out sessions to identify and formulate the promising scientific and technological challenges at the frontier of research. High level keynote talks by internationally renowned scientists and poster presentations will further contribute to the discussions.

June 17-18, 2014  
Point Hotel  
Istanbul, Turkey

Information: <http://conference2014.chistera.eu>



## ***Topics of the Call 2014***

### **Resilient Trustworthy Cyber-Physical Systems**

Cyber-physical systems (CPS) refer to novel hardware and software compositions creating smart, autonomously acting devices, enabling efficient end-to-end workflows and new forms of user-machine interaction. In manifold emerging application domains such as health care, traffic management or energy supply CPS carry a high potential for creating new markets and solutions to societal hazards, but impose highest requirements to quality in terms of resilience, safety, security and privacy. However, the heterogeneous, evolving and distributed nature of CPS bears major challenges to continuously assure these quality requirements employing state of the art methods and technologies. Foundational research efforts are needed to achieve a predictable quality level in an efficient, traceable and measurable way, coping efficiently with external and internal changes, supporting necessary transitions between mechanical, electrical and software engineering, as well as integrating management, design and deployment aspects.

**Keywords:**

- Security and safety requirements engineering, security infrastructures, intrusion detection
- Risk management, quality and risk models
- Model-based software and systems engineering
- Multi-level architectures
- Verification and validation, automated testing
- Security testing, security and privacy protocols, certification processes

### **Human Language Understanding**

Having a machine understand language like a human being can be considered as the epitome of Artificial Intelligence, as exemplified by the design of the Turing test. Even though the domain of automatic language processing has made steady progress over the last decades and some applications are out on the market, the machine is still far from reaching human performance. New approaches are needed, in particular to model high-level, semantic and pragmatic knowledge in a robust fashion. For that purpose, the machine learning approach which has proved very successful to train models from linguistic data should be extended to use more varied data, potentially covering the whole gamut of stimuli a human can learn from, in a much more multidisciplinary approach.

**Keywords:**

- Machine reading, question answering, dialog systems, automatic speech recognition, optical character and handwriting recognition, opinion and sentiment analysis
- Machine learning, semi-supervised learning, reinforcement learning, active learning
- Semantic analysis, context modelling, concept learning, deep learning
- Embodied intelligence, multimodal data processing, information fusion

## Conference Programme

Schedule	17/06/2014 - RTCPS	18/06/2012 - HLU
08:30	<i>Welcome and registration (coffee/pastries available)</i>	<i>Welcome and registration (coffee/pastries available)</i>
09:00	<b>Welcome address</b> Ezgi Bener (TUBITAK, Conference Organiser) Edouard Geoffrois (ANR, CHIST-ERA Coordinator)	<b>Welcome address</b> Ezgi Bener (TUBITAK, Conference Organiser) Edouard Geoffrois (ANR, CHIST-ERA Coordinator)
09:15	<b>Living Security – Managing Security Of Ultra-Large IT Platforms</b> Ruth Breu – University of Innsbruck  <b>Towards Rigorous Design Of Cyber-Physical Systems: A First Step</b> Axel Legay - INRIA  <b>Security for CPS: Moving Towards The Vehicles Of The Future</b> Siraj Shaikh – Coventry University	<b>Understanding Spoken Language Understanding: From Big Data To Real-World Context</b> Roger Moore – University of Sheffield  <b>Natural Languages And Semantics</b> Adam Przepiorkowski – The Polish Academy of Sciences  <b>Language Technology Challenges At The Crossroads Of Data, Language And Evaluation</b> Nicoletta Calzolari – CNR
10:45	<i>Coffee pause / poster networking session</i>	<i>Coffee pause / poster networking session</i>
11:15	<b>Plenary poster session</b> Poster presenters	<b>Plenary poster session</b> Poster presenters
12:00	<b>Information session for Call 2014 applicants</b> Ezgi Bener (TUBITAK, Conference Organiser)	<b>Information session for Call 2014 applicants</b> Ezgi Bener (TUBITAK, Conference Organiser)
12:20	<i>Lunch break / poster networking session</i>	<i>Lunch break / poster networking session</i>
14:00	<b>Parallel break-out sessions: Brainstorming on call content</b> Session attendees	<b>Parallel break-out sessions: Brainstorming on call content</b> Session attendees
16:00	<i>Coffee pause / poster networking session</i>	<i>Coffee pause / poster networking session</i>
16:30	<b>Plenary session: Report on parallel sessions</b> Rapporteurs	<b>Plenary session: Report on parallel sessions</b> Rapporteurs
17:30	<b>Conclusions</b> Georges Klein (SNSF, Conference Organiser)	<b>Conclusions</b> Georges Klein (Conference Organiser)
17:45	<b>Poster networking session</b>	<b>Poster networking session</b>
18:30	<i>End of Conference Day 1</i>	<i>End of Conference Day 2</i>
20:00	<i>Conference Networking dinner</i>	<i>Conference Networking Dinner</i>

## Book of Abstracts

<i>Resilient Trustworthy Cyber-Physical Systems</i> .....	5
1. <b>Keynote Talks</b> .....	5
Living Security - Managing Security Of Ultra-Large IT Platforms.....	5
Towards Rigorous Design Of Cyber-Physical Systems: A First Step.....	5
Security For Cyber-Physical Systems: Moving Towards The Vehicles Of The Future .....	6
2. <b>Posters</b> .....	7
Resilient Mobile Sensor Platforms - ReSens.....	7
Predictive Techniques For Managing Changes/Failures In Cyber-Physical Systems .....	7
AUGMENT: AUTomated Generation Of Mission ENvironments For Testing .....	8
Software-Defined IoT Units For Cyber-Physical Systems .....	8
Cybernetic And Teleonomic Cooperation In Living Matter .....	9
Cyber-Physical Systems Lab: A Research Centre In CPS Engineering .....	9
Control Of Cyber-Physical Systems By On-Line Simulation .....	9
The Challenge Of Characterizing Quality Properties .....	10
ViaBots: Adaptive And Viable Multi-Robot Systems .....	10
Risk Aware Cyber Security Systems.....	11
Use Of Provenance For Trustworthy Cyber-Physical Systems.....	11
Self-Health Awareness Of Distributed Systems By Decentralized Adaptive Agents .....	12
Formal Methods For Trustworthy Cyber-Physical Systems.....	13
Trustworthy Remote Sensing And Communication .....	13
<i>Human Language Understanding</i> .....	14
1. <b>Keynote Talks</b> .....	14
Understanding Spoken Language Understanding: From Big Data To Real-World Context.....	14
Natural Languages And Semantics .....	15
Language Technology Challenges At The Crossroads Of Data, Language And Evaluation .....	16
2. <b>Posters</b> .....	18
Automatic Classification Of Customer Complaints.....	18
AREA: Autonomous REsearch Assistants.....	18
PARSE4REAL .....	19
Actual Problems Of The Machine-Operated Semantic Analysis.....	19
Robust Auditory Scene Analysis In Interaction With Sensory-Motor Modalities For Human Language Understanding .....	20
Affective And Social Dimensions In Spoken Interaction.....	21
Presentation Of The Laboratory Of Informatics Of Grenoble .....	21
Sentiment Analysis For Free – Can You Detect Positive Texts In A Language That You Don't Understand? 22	



## *Resilient Trustworthy Cyber-Physical Systems*

### 1. Keynote Talks

**Ruth Breu**

#### **Living Security - Managing Security Of Ultra-Large IT Platforms**

Professor Breu is Scientific Head of the Quality Engineering Laura Bassi Lab (QE LaB) in Austria (Innsbruck). The goal of QE LaB is to develop methods and tools to foster the quality management of a new generation of dynamically evolving cooperative systems and to bring high-level research results to the market. Her research interests include Requirements Engineering, Model-Driven Software Development, Software Processes, Security Engineering, Quality Management.

##### **Abstract**

Probably the most severe vulnerability of organizations today is their inability to efficiently coordinate security relevant actions and data. Challenges for assuring security of the emerging generation of “smart” platforms (smart cities, smart grids, smart vehicles, ...) arise from the sheer amount of highly interconnected assets, the necessity to balance security, safety, cost and performance requirements and inter-organizational environments. The presented “Living Security” concept claims the necessity to increase the degree of automation, adaptability and analysability within the security management process.

**Axel Legay**

#### **Towards Rigorous Design Of Cyber-Physical Systems: A First Step**

Dr. Axel Legay (2008 IBM Belgian Prize Winner) held positions at University of Liège and CMU (under the supervision of Ed Clarke). He is now full-time researcher at INRIA where he leads the ESTASE team (8 researchers), and a part-time Reader at the Royal Holloway University of London. His main research interests are in developing formal specification and verification techniques for Software Engineering. Axel Legay is a major contributor in the areas of CPS, embedded systems, and Systems of Systems. Axel Legay is a founder and major contributor of statistical model checking (a statistical variant of model checking effectively used in industry). He also provided the first generic techniques for the verification of infinite-state systems and for dealing with variability. Recently, he has introduced new theories to deal with quantitative problems and reason on energy-centric systems. Axel Legay supervised 3 PhD theses and authored more than 150 peer-reviewed publications. He is a referee for top journals and conferences in formal verification and simulation, and program co-chair of INFINITY'09, FIT'10, Runtime Verification 2013, Splat 2014, and FORMATS 2014. He is also workshop chair at ETAPS'14. He is principal investigator on 3 national and 5 European research projects.



#### **Abstract**

In this short presentation we will first introduce the basic ingredients of Cyber-Physical Systems (CPS) and outline the main challenges and actions needed to make smart CPS happen. This includes new social aspects (adoption by humans), and new challenges in terms of safety/security/privacy. We will also insist on the green CPS challenge. In the second part of the presentation, we will present our vision on the rigorous design of CPS. This vision extends the one from Joseph Sifakis with new simulation-based techniques coupled with smart models and tools used to learn and anticipate systems dynamic. In conclusion, we will outline the main research priorities for the coming years.

#### **Siraj Shaikh**

### **Security For Cyber-Physical Systems: Moving Towards The Vehicles Of The Future**

Dr. Siraj Ahmed Shaikh is a Reader in Cyber Security at the Faculty of Engineering and Computing at Coventry University. He leads the Digital Security and Forensics (SaFe) Group. He has been involved in research of large-scale distributed secure systems for over twelve years. His doctoral and post-doctoral research involved design and verification of security and safety-critical systems. He has over fifty peer-reviewed research publications. He holds a PhD in Computer Security, MSc in Computer Networking, MA in Higher and Professional Education, and BSc (Honours) in Computing. He is a Chartered Fellow of BCS and also a Chartered Scientist.

#### **Abstract**

Cyber-Physical Systems (CPS) pose a remarkable challenge to security and resilience. One example of such systems are modern vehicles, which are increasingly sophisticated in terms of communication, sensors and actuators all interconnected. Due to the advances in electronic sensors and digital platforms for communication and control, physical security of today provides enhanced capability on the one hand, yet opens up doors for new types of threats on the other. In the case of vehicles, this opens up new attack vectors, which when coordinated across the physical and cyber space serve to be a real challenge for the detection and prevention of attacks. Added to this, are challenges of non-technical factors such as economics, national policy and human factors that serve as an influence. The need to respond to this problem in a manner that is converged across both physical and cyber space hence is greater than ever. This keynote will highlight some of these issues and emphasise on relevant future research directions.



## 2. Posters

### Agris Nikitenko

#### Resilient Mobile Sensor Platforms - ReSens

Agris Nikitenko is assistant professor at Riga Technical University (Latvia). The research interests of M. Nikitenko include Intelligent Agents, Multi-Agent Systems, Agent-Oriented Software Engineering, Intelligent Tutoring Systems, Ontology-Based Software Systems, Autonomous Software and Robotic Systems, Machine Learning.

##### Abstract

The project idea is to develop a set of methods enabling development of adaptive software and hardware for applications in mobile sensor platforms. The main scientific challenge is to create software and hardware frameworks enabling tolerance to potential damages, malfunctioning and performance losses caused by changing environment that might be both highly dynamic and potentially dangerous. The proposer sees that necessary behaviour of such systems can be reached by introduction of high level agent-based software layer, where software and hardware entities are represented by intelligent agents. Thereby the possible behaviour alternatives can be modeled ahead of the unwanted accident actual occurrence. The envisioned application area of the proposed methods is outdoor mobile sensor platforms with long-term autonomous operation capabilities.

### Alberto Ferrante

#### Predictive Techniques For Managing Changes/Failures In Cyber-Physical Systems

Dr. Alberto Ferrante is a researcher at the ALaRI institute of the University of Lugano, Switzerland. His research interests are in Communication Security and Security for Embedded Systems.

##### Abstract

Change management is fundamental in guaranteeing resilience and trustworthiness of cyber-physical systems. The capability of the system to cope with changes in its conditions, determine, in fact, its ability to survive faults or security attacks. Short-term prediction may be used to dramatically improve resilience; In particular, predictive techniques can be used for prediction of failures and early detection of security attacks. Thus, the system, can timely put in place suitable countermeasures. We propose the development of two-level attack detection system based on predictive techniques.

**Hélène Waeselynck**

## **AUGMENT: AUTomated Generation Of Mission ENvironments For Testing**

Dr. Waeselynck is Researcher in the Laboratory for Analysis and Architecture of Systems (LAAS) of the CNRS in France. Her research interests concern software testing, including Probabilistic Methods for Test Generation, Search-based Testing, Testing and Formal Methods, Mutation Analysis, Test Languages, Robustness Testing, Testing of Safety-Critical Control Systems, Testing of Mobile Computing Systems.

### **Abstract**

Autonomous cyber-physical systems (robots, UAVs, intelligent cars) must operate in diverse and previously unseen environments. The principled and rigorous simulation-based testing of such systems remains a major research challenge. We aim to harness insight, experience and science from software and system testing, and from meta-heuristic search, for the rigorous testing of cyber-physical systems in an economically efficient manner. In particular, we will develop and evaluate automated stochastic strategies to find safety-relevant corner cases for systems and their environments that are beyond the reach of current techniques. Success will greatly accelerate the wide-scale deployment of justifiably trustworthy and resilient cyber-physical systems.

**Hong-Linh Truong**

## **Software-Defined IoT Units For Cyber-Physical Systems**

Dr. Hong-Linh Truong is an Assistant Professor for Service Engineering Analytics at the Distributed Systems Group, Institute of Information Systems, Vienna University of Technology.

His research interests are various fields pertained to distributed and parallel systems with an applied, systems-oriented focus. The main research interest of Dr. Truong focuses Service Engineering Analytics by obtaining an understanding of the behaviour and quality of distributed and parallel applications and systems through monitoring and analysis.

### **Abstract**

We introduce the concept of software-defined IoT units for cloud-based cyber-physical systems that encapsulates fine-grained IoT resources and IoT capabilities in a well-defined API in order to provide a unified view on accessing, configuring and operating IoT systems. Our software-defined IoT units are the fundamental building blocks of software-defined IoT systems. We present our framework for dynamic, on-demand provisioning and deploying such software-defined IoT systems in the cloud. By automating main provisioning aspects and providing support for managed configurations, our framework simplifies provisioning of software-defined IoT systems and enables flexible runtime customization.



## Jean-Yves Tallet

### Cybernetic And Teleonomic Cooperation In Living Matter

The research interest of Mr. Tallet include Consciousness and Mathematical Philosophy.

#### Abstract

Hypothesis: evolution of life or Cyber-Physical Systems is a result of cooperation of their parts. The only logical mean of this cooperation seen as aggregation of elements up to perform operations guided by goals seems to me as feedbacks through different levels of the organisation of the system experienced as such. Also cooperation is defined in this hypothesis as binding of elements and parts able to determine and bound each other under pressure of a guidance by "perception" of one or more dividable goals to share. The numerous feedbacks encountered in life would be able to produce this guidance.

## Jeremy Bryans

### Cyber-Physical Systems Lab: A Research Centre In CPS Engineering

Dr. Jeremy Bryans is a Senior Research Associate at Newcastle University in United Kingdom. His research interests include Cyber Physical Systems, Formal Methods, Semantics, Contract-Based Design, Resilience, Dependability and Provenance.

#### Abstract

In response to the engineering challenges of CPS, Newcastle University is founding a Cyber-Physical Laboratory (CPLab) to create a learning, research and innovation community in CPSs. As part of a £50m investment from Newcastle University in the "Science Central" initiative, CPLab will be co-located with laboratories on smart grids, transportation, digital interaction and cloud computing, as well as a decision theatre and an urban observatory; all sharing a common theme of digitally enabled urban sustainability via integrated research.

## John Brooke

### Control Of Cyber-Physical Systems By On-Line Simulation

Dr. John Brooke is an honorary lecturer in the School of Computer Science of Manchester University in the United Kingdom. His main research interests include Computational Science and Engineering, Distributed Computing, Applied Dynamical Systems.

#### Abstract

We present a software architecture and implementation to integrate simulation of a cyber-physical system with monitoring and control of the system. The simulation is continually steered by sensor nets embedded in the cyber-physical system and allows prediction of future state of the system. We present an example of the concept implementing control of a water distribution network.

Laura M. Castro

## The Challenge Of Characterizing Quality Properties

Dr. Laura M. Castro is an Assistant Professor at the University of A Coruña (Spain), member of the Models and Applications of Distributed Systems research group. Her research focuses on software testing (automated, model and property-based testing), applied to software in general, and distributed, concurrent, functional systems in particular.

### Abstract

Software quality can only be assessed through efficient, effective testing. The most promising automated testing techniques and tools are steaming from model-based and property-based testing areas. However, to date, there has been no successful characterization of software quality non-functional properties such as security, that can be used in such a way. We know these quality properties will make a difference in the next generation of highly critical, autonomous software, so we need to ask ourselves: what do we need a risk model to include, in order to be able to automatically derive tests from it?

Mara Pudane

## ViaBots: Adaptive And Viable Multi-Robot Systems

Mara Pudane is Researcher at the Riga Technical University. The research interests of Mara Pudane include Viable Systems Model, Multi-Robot Systems, Intelligent Autonomous Systems, Emotional Agents, Human Behaviour Simulation.

### Abstract

Currently autonomous robotic systems are designed for operation in environments defined by designers therefore suffering from limited knowledge of the designers which do not reflect the full spectrum of environments features and dynamics. The objective of the project is to create a framework for highly adaptive and long-term operation capable robotic system (including multi-robot) design using the Viable Systems Model as a core concept to achieve system's adaptation. The framework will be based on a model which specifies the functions that are needed to implement adaptation to the changes in the environment and the system itself. The fundamental breakthrough provided by the project will be the new level of adaptation and autonomy of robotic systems.

**Mariam Kiran**

## Risk Aware Cyber Security Systems

Dr. Kiran is a Research Fellow in Cloud Computing at the Department of Computer Science, University of Sheffield in United Kingdom. Her research interests include Cloud Computing, Agent-Based Modelling, Risk Assessment, Trust and Risk on SLAs in Clouds, Security Issues, Market Behaviour Algorithms, Game Theory, Parallelisation, Multi-Objective Optimisation Techniques, Verification and Testing, Evolutionary Computation, Economic Systems, Socioeconomic and Biology Modelling and Social Networks.

### Abstract

The ever changing world for fast connectivity and information at our finger tips is driving the movement towards Cloud computing and Big Data mechanisms to make all information open information. Although, information is power, in some cases such as the NHS pulling out a month before, has raised concerns in the security to protect individual identities. There is a need to have a risk aware guidance of what can be made open and how with mechanisms put in place to make sure data is not compromised across the dynamic nature of networks and connectivity mechanisms.

**Mehmet Aktas**

## Use Of Provenance For Trustworthy Cyber-Physical Systems

Dr. Mehmet Aktas is Assistant Professor at the Yildiz Technical University in Turkey. His research interests include Systems Science (distributed systems, web-based systems and grid computing); Data Science (data/web mining and information retrieval).

### Abstract

The convergence between the cyber and physical worlds generates new challenges for handling the information flowing between the cyber and physical worlds. Huge amount of information will be generated by several sources both in the physical (e.g., by environmental phenomena) and in the virtual world (by sensors, RFIDs, etc.), which need to be stored for future processing. In order to use this huge amount of heterogeneous data coming from a variety of sources, the quality of such data needs to be evaluated and taken into consideration before using it.

Provenance collection is an important component of enhancing the long-term preservation of digital data in the physical sciences and beyond. For the collections to have lasting value, however, and not all collections of data were intended to have lasting value, the data must have undergone some level of preservation. Usually the data pre-processing, such as data calibration, is fulfilled by external systems. Therefore, it is necessary to take into account the provenance information from both outside and inside the datasets. Besides, simply tracking what happened in history is far from enough. Tasks including maintaining lineage of data and the ability to use various versions of them should be involved as well.

Our provenance research has a major focus on instrumentation that minimizes perturbation on an application and minimizes the burden on the application programmer. The preservation step we focus on is automating

the process of gathering of provenance about data. Data provenance is the lineage of a data product or collection of data and applies to observational and imagery data arriving in real time from cyber-physical systems, particularly from sensors, networks of sensors, and instruments; results from computational models and data mining; field studies such as documenting human use of a plot of land over time; regional positional data; scholarly reports in journals, etc. The larger context in which our research is carried out is enhancing the preservation record for the digital data coming from cyber-physical world. The drawback to minimized instrumentation and human burden costs is loss of provenance completeness or quality. This research focusses on trustworthy cyber-physical systems by putting emphasis on the Quality of Information by analyzing the Provenance Aspect.

**Suresh Perinpanayagam**

## **Self-Health Awareness Of Distributed Systems By Decentralized Adaptive Agents**

Dr. Suresh Perinpanayagam is Lecturer at the University of Cranfield in the United Kingdom. His research interests include Aerospace, Autonomous Systems, Automotive, Mechatronics and Advanced Controls, Electric & Hybrid Vehicles, Integrated Vehicle Health Management, Energy, Electric Power Machines and Grid Systems for Energy, Environment, Monitoring and Environmental Informatics, Manufacturing, Through-life Engineering Services, Sensor Technologies.

### **Abstract**

In large-scale engineering and vehicular systems (aircraft, rail/road) as well, the trend is toward the use of network embedded for measurement and control components in order to enable autonomous behavior. Autonomous systems need to know their own current internal state and predict their potentially future state, so that their own capabilities and limitations are understood. In consequence, achieving adaptive, self-health awareness is crucially important for autonomous systems when operating in complex environments. The focus of this proposal is a mission-centric function of self-health awareness and its interpretation as an autonomous, intelligent, multi-agent system. In particular, an aircraft requires constant assessment of its health status in order to determine the future course of its actions. The health assessment is supervised by and assists the human factor in taking adequate decisions.

The goal of an autonomous, intelligent, multi-agent, self-health aware system is to reliably inform the platform about its current capabilities, thus allowing it to make a rational and informed decision about mission feasibility in the light of platform health information. This change of focus from advising a human decision maker to informing an autonomous on-board decision of the platform will require a shift in the design paradigm with emphasis on efficient and reliable operation of the platform as a self-contained unit. The autonomous, intelligent, multi-agent, self-health aware system will enable not only detection and understanding of critical failures, but will also enable autonomous response to malfunctions, mitigation of the impacts of failures on the mission, and more efficient planning and scheduling systems. The agents can adaptively focus on the platform's regions that warrant more attention in order to recommend a maintenance activity, a system reconfiguration

or a system action. In particular, the network of aircrafts will autonomously take decisions by quickly reacting to its state changes, thus replicating the coordination and decision making of the human factor.

The main aim of this proposal is to produce a rigorous, yet practical, design framework for a self-health awareness system interpreted as a decentralised, autonomous multi-agent system which will: provide intelligent monitoring, enable reconfiguration and recovery operations, and allow efficient adaptive reasoning in complex environments. An agent-based system will extend the current notions of health management, especially at the higher levels: fusion, decision-aiding, reconfiguration, and will address the aspects of self-health awareness more effectively than current systems.

[Sylvain Hallé](#)

## Formal Methods For Trustworthy Cyber-Physical Systems

Sylvain Hallé is Assistant Professor at Université du Québec at Chicoutimi in Canada. His research interests include Web Services, Formal Methods, Model Checking, Computer Networks, Logic, AI, Formal Aspects of XML Query Languages.

### Abstract

In recent years, formal methods have proved to be important concepts for the testing, verification and validation of various kinds of computer systems. It is expected that important gains in terms of quality, safety and efficiency can be achieved through the formalization and automation of techniques such as trace analysis, runtime monitoring, behavioural specification and model-based testing.

[Tobias Oechtering](#)

## Trustworthy Remote Sensing And Communication

Dr. Tobias Oechtering is an Associate Professor in the Communication Theory Lab of KTH in Sweden. His research interests include Information Theory (in particular network information theory and security), Communications (in particular in wireless communication), Statistical Inference (in particular distributed detection and privacy), Signal Processing & Algorithms, Networked control.

### Abstract

The ever changing world for fast connectivity and information at our finger tips is driving the movement towards Cloud computing and Big Data mechanisms to make all information open information. Although, information is power, in some cases such as the NHS pulling out a month before, has raised concerns in the security to protect individual identities. There is a need to have a risk aware guidance of what can be made open and how with mechanisms put in place to make sure data is not compromised across the dynamic nature of networks and connectivity mechanisms.





## *Human Language Understanding*

### 1. Keynote Talks

**Roger K. Moore**

#### **Understanding Spoken Language Understanding: From Big Data To Real-World Context**

Prof. Roger K. Moore is Chair of Spoken Language Processing in the Speech and Hearing Research Group at the University of Sheffield, and Visiting Professor at the Bristol Robotics Laboratory and Psychology/Language Sciences at University College London. He has over forty years experience in speech technology R&D and has authored and co-authored over 150 scientific publications. He is well known for investigating the similarities and differences between human and machine spoken language behaviour, and has championed a unified theory of spoken language processing known as PRESENCE (PREdictive SENSorimotor Control and Emulation) that weaves together accounts from a wide variety of disciplines with a view to breathing new life into a new generation of spoken language processing systems - especially human-robot interaction. His recent work in this area includes a novel mathematical interpretation of the 'Uncanny Valley' effect which was published in Nature.

##### **Abstract**

The release of Siri - Apple's speech-driven automated personal assistant for the iPhone 4s - in November 2011 heralded a new era in public awareness of (and engagement with) spoken language technology. Before Siri, automatic speech recognition tended to be used in specialist dictation applications (such as transcribing medical notes) and text-to-speech synthesis was beginning to appear in our cars (in the more advanced satellite navigation systems). After Siri, users began to appreciate the potential for more general-purpose interaction in everyday productivity applications and the market began to realise the importance of understanding what was said and conversing appropriately rather than just writing it down or speaking it back. Since 2011, a number of competitors to Siri have arisen from major players such as Google and Microsoft, and there is plenty of research back in the labs, but it is generally acknowledged that there is a long way to go before spoken language interfaces can support the demands of many real-world applications in a consistent and reliable manner.

This talk will address the fundamental issues facing spoken language understanding, and will highlight the need to go beyond the current fashion for using machine learning in a more-or-less blind attempt to train static models on ecologically unrealistic amounts of unrepresentative training data. Rather, the talk will focus on critical developments outside the field of speech and language - particularly in the neurosciences and in cognitive robotics - and will show how insights into the behaviour of living systems in general and human beings in particular could have a direct impact on the next generation of spoken language systems. In particular, it will be suggested that future progress in spoken language understanding might require us to

refocus our attention on generative models of spoken language production (derived from models of movement and action), and that traditional learning paradigms using off-line training with static corpora need to be replaced by on-line interactive skill acquisition in real-world situations and environments.

It will be suggested that, not only is such an approach necessary if we are to have technology that can figure out why people are saying what they are saying (and what to do about it), but that progress in this area is essential if we are to move towards a new generation of context-aware intelligent agents that are capable of engaging in genuinely communicative behaviour with their human users.

Adam Przepiorkowski

## Natural Languages And Semantics

Adam Przepiorkowski is an Associate Professor at the Institute of Computer Science, Polish Academy of Sciences, Head of the Linguistic Engineering Group (<http://zil.ipipan.waw.pl/>). He holds MSc in Computer Science, PhD in Linguistics and Habilitation in Natural Language Processing (NLP). During the last 15 years he has been involved in numerous national and European projects; he has led the National Corpus of Polish (<http://nkip.pl/>) and he is currently the vice leader of the PARSEME COST Action (<http://parseme.eu/>). He is the co-founder and chief editor of the Journal of Language Modelling (<http://ilm.ipipan.waw.pl/>). Adam Przepiorkowski is the author or co-author of around 150 publications and an editor or co-editor of some 10 volumes. His publications cover topics ranging from theoretical morphosyntax, syntax and semantics to the use of novel machine learning methods in NLP and further to NLP applications.

### Abstract

In this talk I will relate to two components of the title of the CHIST-ERA Call 2014 on Human Language Understanding: Natural Languages and Semantics.

The paradigm dominant in Natural Language Processing (NLP) since the early 1990s is that of statistical induction and machine learning. Large corpora have been built and annotated in ways that make the automatic induction of linguistic models possible. This approach, championed by the speech community, proved particularly successful in learning lexical models, e.g., in constructing part of speech (POS) taggers, but to varying extents also in machine translation and many other tasks.

Given the initial successes, it is reasonable that NLP research concentrated on “language-independent” approaches: why construct language-specific systems if a single system may be built, which learns from various corpora of particular languages? However, such research has never been really “language-independent”, as it has always relied on annotated corpora and language-specific lexical resources, and the cost of adding such annotation or developing such lexical resources has often been overlooked.

After over 20 years of the domination of the machine learning paradigm, its limits become clear. In particular, while this paradigm proved successful in the development of POS taggers, shallow syntactic parsers or in the so-called phrase-based machine translation, such statistical approaches are not so successful where more complex linguistic levels and true language understanding are involved. (Interestingly, this seems to be recognised by some of the precursors of the statistical paradigm, e.g., by Kenneth Church in his paper “A Pendulum Swung Too Far”.) Correspondingly, the first thesis of this talk is that, if we want to substantially

move forward towards real human language understanding, we need to combine “language-independent” methods with the construction of non-trivial “language-specific” resources representing complex syntactic, semantic and pragmatic information about linguistic constructions.

The second part of the talk will be devoted to some new and exciting research on natural language semantics. During the last 20 years or so, NLP research concentrated on lexical semantics and great progress has been made in tasks such as Word Sense Disambiguation and Semantic Role Labeling, mostly using the so-called distributional approach to semantics. Again, this progress is correlated with the development of language-specific resources such as Wordnets and lexica containing semantic role information (FrameNet, VerbNet, PropBank). While reasonably-sized Wordnets exist for a number of European languages, semantic role resources – which bring us closer to compositional semantics (combining meanings of components into meanings of larger components), leading to the understanding of full sentences and paragraphs – are still very rare.

The second thesis of this talk is that the coming years should – and hopefully will – see the increased research in compositional semantics. The construction of semantic role resources is one way to go, but in this talk I will instead concentrate on cutting-edge research (e.g., by Marco Baroni, Ann Copestake, Mark Steedman and their colleagues) on combining distributional semantics with the more traditional logical or model-theoretic approaches to semantics. In these approaches, almost all semantic information is learned from large textual corpora which are not human annotated (i.e., which are relatively cheap to construct), and only rather small resources are mutually constructed, e.g., lexica of functional (closed class) words. I will finish the talk by mentioning the possibility of so-called grounded language learning, where grammars and semantic impact of words are learned from multimodal corpora which pair sentences with perceptual contexts (e.g., work by Raymond Mooney).

## Nicoletta Calzolari

### Language Technology Challenges At The Crossroads Of Data, Language And Evaluation

Nicoletta Calzolari Zamorani is Research Associate and former Director of Research and Director (2003-08) of the Institute of Computational Linguistics-CNR, Pisa. She received an Honorary Doctorate in Philosophy from the University of Copenhagen and awarded the title of “ACL Fellow” in the ACL (Association for Computational Linguistics) Fellows founding group for “significant contributions to computational lexicography, and for the creation and dissemination of language resources”. She is coordinating international, European (recently the EC FLaReNet Network), national projects/strategic initiatives. She is President of ELRA (European Language Resources Association), Permanent member of ICCL, vice-president of META-TRUST, convener of the ISO Lexicon WG, president of the PAROLE Association, former chair of the Scientific Board of CLARIN, former member of the ACL Exec, of the META-NET Council, of the ESFRI Social Sciences and Humanities Working Group, and of many International Committees and Advisory Boards (e.g., ELSNET, SENSEVAL, ECOR, SIGLEX). She is General Chair of LREC (since 2004) and of COLING-ACL-2006. She was invited speaker, member of program committees, organiser of many international conferences/workshops. She is co-editor-in-chief of the

Journal Language Resources and Evaluation, Springer. She is member of journal editorial/advisory boards. She has more than 400 publications.

## **Abstract**

Language Technology (LT) is a data-intensive field and major breakthroughs have stemmed from a better use of more and more Language Resources (LRs). The challenges ahead depend on a coherent strategy involving not only the best methods and technologies but also many LR related dimensions. One of these is evaluation. The major target will continue to be the problem of “understanding”: ours must become also a “knowledge-intensive” field. But the next frontier will focus not only on text or multilingualism but also around the concurrent use of different types of data, across sectors and across modalities (social media data, but also visual and multimodal data) and the integration of text analytics methods with methods for dealing with other types of data, to capture the full potential not only of big data but also of the combination of different modalities and different semantic/pragmatic contexts. A critical point will be the adoption of consolidated methodologies and achievements (e.g. appropriate evaluation mechanisms and interoperability) of the LT/LR field also when working on different types of data.

The presentation will also highlight some policy issues that must be considered when making up a strategy for the future of the field: issues such as sharing resources and tools, adopting the paradigm of accumulation of knowledge and allowing replicability of research results.



## 2. Posters

**Bilge Koroglu**

### Automatic Classification Of Customer Complaints

Bilge Koroglu is Senior Software Developer at Yapi Kredi. Her research interests include Natural Language Processing and Machine Learning.

#### **Abstract**

We have been receiving customer complaints from our branches, internet and telephone banking in digital format. These complaints are classified into predefined types of complaints reading by employees from customer services department of our bank. It requires extremely huge amount of human labelling process. We have been starting to work on a decision support system which classifies the complaints according to a statistical model so that human labelling is minimized. By using text categorization and machine learning algorithms, the system can learn a model dynamically and also can answer newly defined complaint class.

**Egons Lavendelis**

### AREA: Autonomous REsearch Assistants

Dr. Lavendelis is Assistant Professor at Riga Technical University in Latvia. The research interests of Dr. Lavendelis include Intelligent Agents, Multi-Agent Systems, Agent-Oriented Software Engineering, Intelligent Tutoring Systems, Ontology-Based Software Systems, Autonomous Software and Robotic Systems, Machine Learning.

#### **Abstract**

The idea is to create a framework for autonomous research assistants (AREAs). We will design an infrastructure and protocols for machine-first (yet human-friendly) research communication, and create several AREA prototypes. While the most creative aspects of the research process must remain a human domain, AREAs will produce, exchange, consume and review research results; discover correlations in data generated by sensor networks and/or robots and pose hypotheses; conduct peer review of published research objects and assess reputation of researchers. In our interdisciplinary work, we will combine insights from the theory of multi-agent systems, scholarly communication studies, and text and data mining.



Gulsen Eryigit

## PARSE4REAL

Dr. Gulsen Eryigit is an Assistant Professor at the Department of Computer Engineering of Computer and Informatics Faculty of Istanbul Technical University in Turkey. Her main research area is Natural Language Processing, and especially the application of Machine Learning techniques into this area.

### Abstract

Despite the importance of structural analysis in properly determining sentence meaning, many commercial Text Analytics solutions do not employ parsing. One reason for this is that it is only very recently that automatic parsers have become efficient enough to be used as part of a content processing workflow operating on web-scale data sets. While the past fifteen years has seen significant advances in statistical parsing, two crucial weaknesses remain:

1. Multilingualism
2. Noisy Web 2.0 Text

The Parse4Real project will explicitly address these two challenges by producing multilingual, adaptive and scalable parsing technology for Web2.0 text.

Iryna Sekret

## Actual Problems Of The Machine-Operated Semantic Analysis

Dr. Iryna Sekret is Professor at the Zirve University in Turkey. Her research interests are in the areas of psychology and language, namely educational psychology, ICT and e-learning, social media in learning, contrastive linguistics, semantics and translation, English language teaching, and teaching English for specific purposes.

### Abstract

The research proposal is within the area of the semantic analysis to improve quality of e-translation and language e-learning. Among urgent problems are:

1. low quality of the machine translation because of the scarcity of efficient language solutions based on the thorough semantic analysis of the language means regarding their derivation, compatibility with other linguistic units, processes of the semantic shift occurring in cases of the language unit borrowing, its dialectal usage, or representation in the jargon speech;
2. lack of e-learning programmes for developing effective vocabulary skills;
3. problems of the semantic analysis of the sign language.

Jean Rouat

## Robust Auditory Scene Analysis In Interaction With Sensory-Motor Modalities For Human Language Understanding

Dr. Jean Rouat is Professor at the Sherbrooke University in Canada. His research interests include Intelligent Systems, Artificial Intelligence, Machine Learning, Signal, Speech and Audible Signals Processing, Visual Processing, Computational Neurosciences, Neurophysiological Signal Analysis, Human-Systems Interfaces, Sensorial Substitution.

### Abstract

Most Human Language Understanding systems are based on statistical and machine learning pattern matching technics either implemented as graphical models (HMM, Language models, etc.) or as formal neural networks encoding the firing rate of neurons (convolutive N.N., deep N.N., Boltzmann N.N., etc.). Impressive practical classification and pattern matching results can now be reach thanks to recent developments in computing power and hardware implementations, notably on GPU (Graphical Processing Units).

Human communication through language is essential to our survival and is in strong interaction with motricity, vision, emotion, etc. Abstract interpretation of the acoustical signal (semantic, emotion, etc.) requires the use of most areas of our brain (motor, visual, planning, ... areas). One of the striking performance of the brain is the auditory scene analysis and the capacity to decompose auditory scenes into auditory streams and objects. The most common known application of the auditory scene is the cocktail party effect, that is in practice only a side effect of a more complex and general process that implies our multisensory brain. In fact, auditory scene analysis is also fundamental to the acquisition of a new language and to the understanding of speech and sounds. Our ability to analyse auditory scenes by integration of visual and motor feedbacks is fundamental to our build up of human language understanding. Taking into account these feedbacks and these multisensory interactions for better human language understanding and acquisition systems cannot be reduced to pattern matching or classification algorithms. Dynamic feedbacks, active cochlea, attentional processes, anticipation, intention, planning,... occurring in the multisensory brain have to be taken into account and implemented for better auditory scene analysis modules that are part of human language understanding systems.

The poster discusses potential research directions and solutions to the design of better human language understanding systems that comprise robust auditory scene analysis modules in interaction with other sensory-motor modalities of our brain. Discussions about software and hardware implementations in relation with state of the art machine learning and NPU (Neural Processing Units) are also presented on the poster.

[Laurence Devillers](#)

## Affective And Social Dimensions In Spoken Interaction

Dr. Laurence Devillers is professor at Paris IV University in France. Her research interests include Human-Human and Human-Machine Interaction, Emotion Detection (audio and multimodal signals), Social Interaction, Social Simulation, Social Signals.

### **Abstract**

In order to better understand spoken language and to design social interaction with machines, experimental grounding is required to study expressions of emotion, attention and intention cues during spoken interaction. Robotics are a relevant framework for designing applications due to the learning and skills of robots. Many research topics linked to spoken language understanding will be presented with some new challenges linked to Multimodal, Multi-Party, Real-World Human-Robot Interaction.

[Laurent Besacier](#)

## Presentation Of The Laboratory Of Informatics Of Grenoble

Dr. Laurent Besacier is Professor at the University Joseph Fourier, Laboratory of Informatics of Grenoble (LIG). His research interests include Speech & Language Understanding, Automatic Speech Transcription and Translation, Computer Aided Translation, Processing of Under-Resourced Languages, Speech Processing / Analysis and Interactions in Ambient Environments, Modeling Social Affects, Automatic and Interactive Meaning Clarification Processes.

### **Abstract**

This poster will present the LIG laboratory as a potential partner for CHIST-ERA project on the topic 'Human Language Understanding' as well as a list of potential projects ideas to which LIG will be interested to participate or coordinate.

**Mark Cieliebak**

## Sentiment Analysis For Free – Can You Detect Positive Texts In A Language That You Don't Understand?

Dr. Mark Cieliebak is researcher and lecturer at Zurich University of Applied Sciences (ZHAW). His expertise includes Efficient Algorithms, Software Engineering and Data Analysis.

### **Abstract**

Sentiment Analysis Systems (SAS) typically rely on human interaction to build them: lexica are assembled, documents need to be tagged manually, POS tagging requires a thorough understanding of the language, etc. Such language resources exist for common languages such as English, German or Chinese. But what if you want to build an SAS for a "new" language, one with poor or no language resources? We want to use large sets of opinionated documents (e.g. Amazon reviews or Tweets) to fully automatically create an SAS for any language, or even a dialect.