

FETLT-2016 2nd International Workshop

SEVILLE (SPAIN) 30TH NOVEMBER, 1ST - 2ND DECEMBER 2016

Future Emerging Trends in Language Technologies

WORKSHOP PROGRAM

30TH NOVEMBER



Venue: Pabellón de México (Universidad de Sevilla)

ADDRESS: PASEO DE LAS DELICIAS, S/N



08:00 - 09:00

Registration and workshop welcome event



09:00 - 10:00

Opening ceremony

Dr. Manuel García León

DIRECTOR OF RESEARCH AND KNOWLEDGE TRANSFER
SECRETARY OF ECONOMY AND KNOWLEDGE. REGIONAL GOVERNMENT OF ANDALUCIA

Dr. Julián Martínez

VICE-CHANCELLOR FOR RESEARCH, UNIVERSITY OF SEVILLE

Dr. Victor Muñoz

VICE-CHANCELLOR FOR STRATEGIC PROJECTS, UNIVERSITY OF MALAGA

Dr. Juan Carlos Toscano

SECRETARY OF SCIENCE, ORGANIZATION OF IBERO-AMERICAN STATES (OEI)

Dr. Jose F. Quesada

PROGRAM CHAIR FETLT16
DPT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE. UNIVERSITY OF SEVILLE
DIRECTOR ANDALUSIAN SCIENTIFIC INFORMATION SYSTEM. REGIONAL GOVERNMENT OF ANDALUCIA



10:00 - 11:00

Hans Uszkoreit

SAARLAND UNIVERSITY, DFKI LANGUAGE TECHNOLOGY LAB

Opportunities and challenges for language technology in the new era of AI

Biography:

Hans Uszkoreit is Scientific Director at the German Research Center for Artificial Intelligence (DFKI) where he heads the DFKI Language Technology Lab in Berlin. He studied Linguistics and Computer Science at the Technical University of Berlin and the University of Texas at Austin. While he was studying in Austin, he also worked as a research associate in a large machine translation project at the Linguistics Research Center. After he received his Ph.D. in Linguistics from the University of Texas, he worked as a computer scientist at the Artificial Intelligence Center and was affiliated with the Center for the Study of Language and Information at Stanford University.

After a stay at IBM Germany, first as a research fellow and later as a Project Leader he was appointed Full Professor for Computational Linguistics at Saarland University, where he taught for more than 25 years. From 2010 till 2015 Uszkoreit led META-NET, a European Network of Excellence dedicated to Technologies for Multilingual Europe comprising 60 research centers in 34 countries. In several projects he is currently developing language technology applications in the area of data analytics.

Abstract:

Currently language technology is blessed with unprecedented opportunities in means and demands. New means emerge from progress in AI, especially in machine learning and semantic technologies, but also from the availability of huge volumes of language data and very large repositories of structured knowledge. Increased demands and practical chances for applications originate from the digital transformation of industry and society. These applications in turn supply even more data for the acquisition of linguistic skills through machine learning methods.

Whereas progress in the last decades mainly came from raw or hand-labelled linguistic data, we witness a move towards data that are already associated with nonlinguistic information providing references, contexts, needs and actions. For some applications such data are relatively easy to obtain, for others the data acquisition is costly or highly difficult.

In my presentation I will provide examples from our own work and from other research to illustrate this development. I will then discuss the ramifications to research planning and expectation management.



11:00 - 11:30

Coffee break



11:30 - 12:30

Research Unit in Logic, Language and Information (Andalucia Tech)

UILLI - AT

Explora Research Project

NEW DESIGN OF DIALOGUE SYSTEMS



Lekta.ai

A RESEARCH PLATFORM FOR CONVERSATIONAL USER INTERFACES

Research Unit in Logic, Language and Information (Andalucia Tech)

UILLI - AT

Abstract:

The research network UILLI-AT is a coordinated Research Unit that incorporates other labs/research teams within University of Seville and University of Malaga.

Its main objective is to promote collaboration through the realization of great research projects that require the active participation of multiple researchers in various disciplines. These fields include the following:

- Linguistics.
- Logics and Philosophy of Science.
- Computer Science and Artificial Intelligence.
- Mathematics applied to Computer Science.
- Cooperative Information Systems.
- Other convergent fields, like Robotics and Domotics.

The unit is fully operative and is currently involved in the research Project "New Design of Dialogue Systems" with a focus on the study of the techniques for the application of new approaches to the design, implementation and evaluation of Dialog Systems.

Speakers:



Dr. Teresa Lopez-Soto. Associate Professor at the Department of English Language. University of Seville.

She holds a PhD in Computational Linguistics (Applying linguistic information to enrich decision making in Speech Recognition). Present research in Cognitive Linguistics, Semantic Information and Semantic Representation of Language with a special interest on Speech Perception.

She is also part a multidisciplinary team on Clinical Research (Perception, Neurosciences and Cognitive Sciences) in which she works on the design of intensive auditory training (cochlear implants) and in connection with Neurolinguistics: cognitive therapy in neurodegenerative diseases.



Dr. Alfredo Burrieza Muñiz, full profesor of Logic and Philosophy of Science at the University of Málaga.

He is the Director of the Unit of Research in "Lógica, Lenguaje e Información" (Andalucía Tech). He usually works in the development of non-standard logics for Artificial Intelligence in multidisciplinary projects of Applied Mathematics and Engineering.



Dr. Ángel Nepomuceno. Full Professor of Logics. University of Seville.

He is Full Professor of Logic and Philosophy of Science at University of Sevilla. He belongs to the interdisciplinary Research Group on Logic, Language and Information at the University of Seville (GILLIUS). He has worked in the area of logical treatment of abduction and non classical logics. He is Principal Investigator of the projects Dynamics of Information: Reasoning, Interaction and Abduction, and New Designs of Dialogue Systems.



Francisco J. Salguero-Lamillar is Full Professor of Linguistics at the University of Seville.

He usually works in the field of cognitive linguistics and its relationship to formal models of grammar and semantics. Currently he belongs to the Research Group on Logic, Language and Information at the University of Seville (GILLIUS) and works on the projects Dynamics of Information: Reasoning, Interaction and Abduction, and New Designs of Dialogue Systems.

Explora Research Project

New Design of Dialogue Systems

Abstract:

Dialog Systems (DS) that are oriented both to spoken and written interaction represent one area of major economic impact and scientific interest in the field of Language Technologies. To design a DS is a highly complex process that must integrate operative functional requirements (what to do and how), linguistic issues (natural language) and computational elements (response times must be extremely short).

The main research lines have been lately focused around the study of the techniques that allow for the quasi-automatic creation of such systems from training bases (annotated corpora), while addressing open domains (one of the present-day objectives is precisely the idea of universal domain). Having into account the main goals of the Explora program (exploration and investigation), researchers from different teams raise the following question: How the analysis of human language learning can inspire the creation of a model applicable to the design of dialog systems.

The first observation to be made in our proposal is the abductive nature of language learning, the temporal and epistemic characteristics of such process and the existence of cognitive and metacognitive phenomena that come along and could be spread to modulate DSs. This Project must therefore be considered an exploratory one that can allow for the development of a formal framework that may produce a dynamic design of a DS. This project is endorsed by the collaboration of researchers from different teams in all related fields: Advanced Dialog Systems (UGr), Logic, Language and Information (US), Cognitive Sciences, Applied Mathematics and Computation and InDoMiTo (UMA).

Lekta.ai

A Research Platform for Conversational User Interfaces

Abstract:

Innovative research on hot topics related to New Design of Dialogue Systems requires the use of Advanced platforms for the implementation, use and evaluation of Conversational User Interfaces. In order to optimize the available resources of the research project and strengthen the collaboration between the academic and industrial sectors, the aforementioned Explora project will use the [Lekta.ai](#) technology as a reference working platform.

Lekta version 2 is the result of several years of intensive research on Spoken Dialogue Systems and Conversational User Interfaces. It follows a multidimensional approach integrating the computational linguistic dimension, with the operative, communicative and business factors involved.

In addition, Lekta follows a hybrid approach that facilitates the integration of knowledge, business-oriented and strategic-driven operational and communicative strategies. This hybrid approach also guarantees the optimization of the whole architecture by means of machine learning, and data science-based techniques.

[Lekta.ai](#) is currently working on version 3 of the technological platform, integrating some of the key research trends in the field, such as incremental understanding and dialogue management, an open-domain approach able to improve the reusability of language resources, application of deep learning to dialogue management and error handling, etc.

Speaker:



Jose F Quesada is the director of the Andalusian Scientific Information System (SICA), Associate professor of the Department of Computer Science and Artificial Intelligence of the University of Seville, coordinator of the Club Scratch Iberciencia (Secretary of Science - Organization of the Iberoamerican States - OEI), and has served the European Commission as well as the COST organization as an Expert Evaluator since 2013.

He got his Doctor degree in Computer Science degree from University of Seville in 1997, BA in Philosophy from the University of Granada in 1992, and BA in Art History from UNED in 2013. After a PostDoc at SRI (Computer Science Lab, Menlo Park, CA), he participated in several Spanish and European Research Projects. From 2005 till 2012 he was the Research Director of the Natural Language Division of Vocalcom. In 2016, he has co-founded the company [Lekta.ai](#), where he is leading the research and innovation programs as the CTO of the company.

His present research interests focus on the integration of Language Technologies, Machine Learning and Data Science. He has published more than 80 papers in different journals and conferences, and has participated in more than 100 industrial events related with language technologies.



12:30 - 13:30

Nicoletta Calzolari Zamorani

ILC-CNR, PISA (ITALY)

An excursus through policy issues (At the crossroads of data, language resources and infrastructure)

Biography:

Research Associate and former Director of Research and Director of the CNR Institute of Computational Linguistics, Pisa, Italy. Received an Honorary Doctorate in Philosophy from the University of Copenhagen “for her significant contribution to the field of Computational Linguistics”. Awarded the title “ACL Fellow” for “significant contributions to computational lexicography, and for the creation and dissemination of language resources” in the founding group of the ACL Fellows program of the Association for Computational Linguistics. Coordinated many international/European/national projects/strategic initiatives.

President of ELRA (European Language Resources Association), permanent member of ICCL, chair of ISO/TC 37/SC 4, vice-president of META-TRUST, member of the Board of UNDL Foundation (Universal Networking Digital Language Foundation), member of the Advisory Board of LIDER (Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe), committee member of ISO/TC 37/AG 0, president of the PAROLE Association, former convenor of the ISO Lexicon WG, 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).

General Chair of LREC (since 2004), of COLING 2016 and COLING-ACL-2006. Invited speaker, member of program committees, organiser of many international conferences/workshops. Co-editor-in-chief of the Journal Language Resources and Evaluation, Springer. Member of journal editorial/advisory boards. More than 400 publications.

Abstract:

I will highlight the importance of policy issues for the future of Language Technology (LT): issues such as standardisation, sharing resources, services and tools, adopting the paradigm of accumulation of knowledge and promoting replicability of research results. The challenges ahead depend on a coherent strategy involving not only the best methods and technologies but also policy dimensions.

LT is a “data-intensive” field and major breakthroughs stemmed from the use of large Language Resources (LRs). It must become also a “knowledge-intensive” field. 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 modalities (social media, and visual and multimodal data) and the integration of text analytics with methods for capturing the full potential of the combination of various modalities and different semantic/pragmatic contexts. A critical point will be the adoption of consolidated methodologies of the LT/LR field (e.g. appropriate evaluation and interoperability) also when working on different types of data.

In the paradigm of open language infrastructures based on sharing LRs, services and tools, a way for LT to achieve the status of a mature science lies in initiatives enabling to join forces both in the creation of large LR pools and in big collaborative experiments using these LRs. This will enable building on each other achievements, integrating results (also with Linked Data). This cannot be achieved without standardisation efforts. I will point at current initiatives within ISO with respect to standardising LRs.

This requires also an effort towards a culture of “service to the community” where everyone has to contribute. This “cultural change” is not a minor issue. I will mention how initiatives like the LRE Map, Share your LRs, ISLRN, are steps towards promoting the concept of Open Science, highlighting the role of ELRA and LREC in pushing towards this vision.



13:30 - 15:00

Lunch



15:00 - 16:30

Andalusian agency of knowledge

Opportunities for cooperation and funding in research projects and programmes

Miguel Torres

DIRECTOR OF KNOWLEDGE TRANSFER AND ENTREPRENEURSHIP. UNIVERSITY OF SEVILLE

Jose F. Quesada

PROGRAM CHAIR FETLT16

Moderator

Jaime Durán Díaz

Sectorial TIC para Andalucía en Agencia Andaluza del Conocimiento (Transferencia y H2020)



Jaime Durán has a Degree on Business Administration from University of Seville. He completed his studies with a Master's Degree as an Expert on Technology Transfer (University of Seville and IPTS).

Since 2002, he has been involved in issues concerning Technology Transfer and RDi management as a technology consultant in Agencia Andaluza del Conocimiento (formerly CITAndalucía).

His role as ICT sectorial head in Agencia Andaluza del Conocimiento deals with the provision of services to SMEs, Universities, Entrepreneurs, Sectorial association: Knowledge Transfer, Support for international networking events, support for H2020 funding as Regional Contact Point for ICT and Security, IPR, Technology Watch.... focused in the ICT sector. He is a project evaluator at ANEP (National Agency of Evaluation and Prospective). As ICT sectorial, he supports the ICT Sector Group of Enterprise Europe Network (European Commission).

Confirmed Panelists

Aníbal Ollero (University of Seville)

Margaretha Mazura (EMF)

Antonio Jesus Nebro Urbaneja (University of Málaga)

Aníbal Figueiras (Charles III University of Madrid)

Speakers:



Aníbal Ollero is Full professor, head of GRVC (75 members), University of Seville, and Scientific Advisor of the Center for Advanced Aerospace Technologies in Seville (Spain). He has been full professor at the Universities of Santiago and Malaga (Spain), researcher at the Robotics Institute of Carnegie Mellon University (Pittsburgh, USA) and LAAS-CNRS (Toulouse, France). He authored more than 635 publications, including 9 books and 135 SCI journal papers and led about 140 projects, transferring results to many companies.

He has participated in 22 European Projects being coordinator of 6, including the recently concluded FP7-EC integrated projects ARCAS and EC-SAFEMOBIL and the on-going H2020 AEROARMS.

He is recipient of 15 awards, has supervised 32 PhD Thesis and is currently co-chair of the IEEE Technical Committee on Aerial Robotics and Unmanned Aerial Vehicles, member of the Board of Directors and coordinator of the Aerial Robotics Topic Group of euRobotics and president of the Spanish Society for Research and Development in Robotics.



Margaretha Mazura has been working for EU affairs since 1993 with specialisation in Information Society policy issues. In 1997, she was appointed Deputy Secretary General of the EMF – the Forum of e-Excellence, www.emfs.eu and Secretary General in 2007.

In this capacity, she represents EMF at events worldwide and initiates EU-funded projects (e.g. LT Compass and LT Observatory). Since 1999, Margaretha is an external expert to different EC units and programmes, amongst them DG CNECT (Language Technologies). Margaretha was elected member of the Board of TermNet (the Network for Terminology) and the Steering Committee of the "Business Platform for Multilingualism" of DG EAC (2008-2012).

2013/14 She acted as an external expert for DG REGIO in drafting synergy guidelines for combined funding opportunities between Horizon 2020 and ESIF (European Structural and Investment Funds; [synergies_en.pdf](#)).

Currently, she is a partner in the LT-Observatory project where she identified funding opportunities for LT and liaises with EU institutions such as the European Parliament and the Committee of the Regions to "make the Digital Single Market multilingual". Margaretha Mazura holds a doctorate in Law of the University of Vienna, and a diploma of Advanced European Studies from the College of Europe.

www.linkedin.com/in/margarethamazura



Professor Antonio Jesus Nebro Urbaneja holds a PhD in Computer Science from the University of Málaga since 1999, university in which he won the degrees of Graduate in Computer Science (1991) and Diploma in Computer Science (1988). He is currently Associate Professor at the University of Malaga. He has been recognized with three six-year periods of researching, four five-year periods in teaching and four periods of autonomous complements.

He started his postgraduate at the University of Malaga in 1991, where he enjoyed a personal grant as "Junta de Andalucía" professor and researcher for three years in the Department of Languages and Computing Sciences at the University of Malaga into the GISUM research group where he still belongs in the present. He was "Profesor asociado" from 1994 to 1996, "Profesor Titular de Escuela Universitaria" until 2003, and since then he has been Associate Professor ("Profesor Titular"). Yet inside GISUM group, he

has participated in 8 national research projects, 2 regional projects, 2 european projects and 7 projects with companies.

His research activity is related to parallelism issues, Big Data and multi-objective optimization techniques, resulting in 28 articles published in international journals, 26 of which are indexed in JCR, 15 book chapters and 37 articles in international congresses. His H index is 30, and his jobs have more than 3100 citations. He has participated in numerous international conferences, being part of the program committee in many of them, being a reviewer in more than 15 international magazines.



Prof. Dr. Aníbal R. Figueiras-Vidal got his Doctor in Telecom Engineering degree from UPC in 1978, and he is now a Full Professor in Signal Theory and Communications at UC3M. His academic experience includes serving at UPC, UPM and USC.

His present research interest is focused on signal and data processing and machine learning. He has published some 80 papers in international refereed journals and more than 200 at technical conferences, supervised more than 30 doctoral dissertations, and he has been the principal investigator for almost 100 research projects and contracts. He is a member of the Spain Royal Academy of Engineering, and he was its President from 2007 to 2011. He received the "Honoris Causa" Doctor degree from Universities of Vigo (Spain) and San Pablo (Arequipa, Peru).



16:30 - 17:00

Coffee break



LT-OBSERVE

17:00 - 19:00

LT-Observatory

Observatory for language resources and machine translation in Europe

Abstract:

The Observatory for Language Resources (LR) and Machine Translation (MT) in Europe <http://www.lt-innovate.org/lt-observe> derives from the need expressed by multilingual solution providers (enterprises and researchers) to have better access to more usable language resources (LRs), together with the more general need to make available large amounts of content and data that are currently untapped due to language barriers, leading to a loss of economic opportunities. The LT-Observatory aims at making LR more accessible to and usable by users of machine translation in public and commercial services. In parallel, LT strategies at national and regional level were investigated and funding sources identified that provide windows of opportunity for LT projects.

The LTO session will present the practical results of the project, i.e. tools and on-line information for the benefits of all involved in LT/MT and that will be maintained after the project end. These include:

- A language resources catalogue, annotated LR for commercial and other purpose.
- On-line guide to national and regional funding opportunities.
- The MT EcoGuide: Practical steps for LT developers and users, and recommendations to decision-makers in easy-to-consume modules.

The LT_Observatory project was a two year CSA funded by H2020 ICT 17-2014 c) "Cracking the language barrier".

Draft Agenda:

LT_Observatory - New ways towards Language Resources - an introduction

ZABALA, Luz Esparza

The challenge of collecting and evaluating LRs for commercial use

CLARIN, Bente Maegaard

Access is king - Demo of LTO LR Catalogue on-line

LT-Innovate, Philippe Wacker

The Future of R&I in LT - the Strategic Research & Innovation Agenda

CRACKER project, Georg Rehm

A path through the funding maze: LT National & Regional Funding Opportunities on-line

EMF, Margaretha Mazura

MT EcoGuide: "What MT professionals need to know and what decision makers should know" - a practical vademecum

University of Vienna, Vesna Lusicky

Final take-away message

EMF, Margaretha Mazura

Speakers:



Luz Esparza, Telecommunication Engineer and Senior Consultant at Zabala Innovation Consulting, S.A. since September 2010.

Coordinator of the LT_Observatory project funded by the H2020 programme.

From 2007 to 2010, consultant at the Spanish Ministry of Industry, Tourism and Trade (MITYC) under the framework of Plan Avanza; the global Plan for the development of the Information Society in Spain. She worked as Technical Assistant for the State Secretary Cabinet and in the Avanza's Follow-up office (OTSA)



Bente Maegaard is Vice Executive Director of CLARIN ERIC. Bente Maegaard is employed at the University of Copenhagen, Centre for Language Technology. Her areas of expertise are in particular research infrastructure, language resources and tools, and machine translation. Bente Maegaard has been project coordinator of many national and European research projects. She is a member of the ESFRI Strategic Working Group for Social and Cultural Innovation.



Philippe Wacker is the Secretary General of LT-Innovate, the Association of the Language Technology Industry (www.lt-innovate.org).

A multilingual and cross-cultural manager with a strategic outlook and broad international experience, he has managed leading international ICT and high technology networks for more than 25 years.

His skills set encompasses association and interest group management, international project management, strategic project planning and development, international project finance, event management, international networking and business expansion, European public affairs and lobbying.

www.linkedin.com/in/philippewacker



Georg Rehm works in the [Language Technology Lab](#) at the German Research Center for Artificial Intelligence (DFKI), in Berlin. He is the General Secretary of [META-NET](#), an EU/EC-funded Network of Excellence consisting of 60 research centers from 34 countries, dedicated to building the technological foundations of a multilingual European information society. Furthermore, he is the Coordinator of the EU/EC-funded project [CRACKER](#) which initiated, among others, the emerging European federation [Cracking the Language Barrier](#). Georg is also the coordinator of the BMBF-funded project [Digitale Kuratierungstechnologien](#).

Additionally, Georg Rehm is the Manager of the [German/Austrian Office](#) of the [World Wide Web Consortium \(W3C\)](#), hosted at DFKI in Berlin. In that capacity, Georg is involved in several bridge building activities between W3C work on the Web of Things on the one hand and Industrie 4.0-related activities in Germany on the other.

Georg Rehm holds an M.A. in Computational Linguistics and Artificial Intelligence, Linguistics and Computer Science from the University of Osnabrück. After completing his PhD in Computational Linguistics at the University of Gießen, he worked at the University of Tübingen, leading projects on the sustainability of language resources and technologies. After being responsible for the language technology development at an award-winning internet startup in Berlin, he joined DFKI in early 2010.

Georg Rehm has authored, co-authored or edited more than 130 [research publications](#) and co-edited, together with Hans Uszkoreit, the [META-NET White Paper Series Europe's Languages in the Digital Age](#) as well as the [META-NET Strategic Research Agenda for Multilingual 2020](#). He is also one of the editors of [Language as a Data Type and Key Challenge for Big Data - Strategic Research and Innovation Agenda for the Multilingual Digital Single Market](#).



Vesna Lušicky is a researcher and lecturer at the University of Vienna, Austria. Her work has centered around language technology, language resources and translation. She is the (co-)author of a number of publications in the area of translation, language technology and language resources. She has participated in several national and international R&D projects (LISE, SFB German in Austria, CLARIN-AT, TransCert). Vesna has consulted for companies, governmental agencies and NGOs on language technology and translation services.

www.linkedin.com/in/vesnalusicky



Margaretha Mazura has been working for EU affairs since 1993 with specialisation in Information Society policy issues. In 1997, she was appointed Deputy Secretary General of the EMF - the Forum of e-Excellence, www.emfs.eu and Secretary General in 2007.

In this capacity, she represents EMF at events worldwide and initiates EU-funded projects (e.g. LT Compass and LT Observatory). Since 1999, Margaretha is an external expert to different EC units and programmes, amongst them DG CNECT (Language Technologies). Margaretha was elected member of the Board of TermNet (the Network for Terminology) and the Steering Committee of the "Business Platform for Multilingualism" of DG EAC (2008-2012).

2013/14 She acted as an external expert for DG REGIO in drafting synergy guidelines for combined funding opportunities between Horizon 2020 and ESIF (European Structural and Investment Funds; [synergies_en.pdf](#)).

Currently, she is a partner in the LT-Observatory project where she identified funding opportunities for LT and liaises with EU institutions such as the European Parliament and the Committee of the Regions to "make the Digital Single Market multilingual". Margaretha Mazura holds a doctorate in Law of the University of Vienna, and a diploma of Advanced European Studies from the College of Europe.

www.linkedin.com/in/margarethamazura

1ST DECEMBER



Venue: Pabellón de México (Universidad de Sevilla)

ADDRESS: PASEO DE LAS DELICIAS, S/N

Paper session #1 09:00 - 10:00

Session Chair: Blanca Rodríguez

ZABALA INNOVATION CONSULTING



09:00 - 09:20

**Asier López Zorrilla¹, Ahmed Ratni², Trung Ngo Trong³,
Javier Mikel Olaso¹, Seth Montenegro⁴, Minha Lee⁵,
Fasih Haider⁶, Stephan Schlögl⁷, Gérard Chollet⁸,
Kristina Jokinen³, Dijana Petrovska², Hugues Sansen⁹ and M. Inés Torres¹**

¹UNIVERSIDAD DEL PAÍS VASCO (UPV/EHU), SPAIN; ²TÉLÉCOM SUDPARIS, CNRS SAMOVAR UMR 5157, PARIS-SACLAY UNIVERSITY, FRANCE;
³UNIVERSITY OF HELSINKI, FINLAND; ⁴HYX, UK; ⁵EINDHOVEN INSTITUTE OF TECHNOLOGY, THE NETHERLANDS; ⁶TRINITY COLLEGE, IRELAND;
⁷MANAGEMENT CENTER INNSBRUCK, AUSTRIA; ⁸INTELLIGENT VOICE, UK; ⁹SHANKAA, FRANCE

LifeLine dialogues with Roberta

Abstract:

This paper describes work on dialogue data collection and dialogue system design for personal assistant humanoid robots undertaken at eINTERFACE 2016. The emphasis has been on the system's speech capabilities and dialogue modeling of what we call LifeLine Dialogues, i.e. dialogues that help people tell stories about their lives. The main goal behind this type of application is to help elderly people exercise their speech and memory capabilities. The system further aims at acquiring a good level of knowledge about the person's interests and thus is expected to feature open-domain conversations, presenting useful and interesting information to the user. The novel contributions of this work are: (1) a flexible spoken dialogue system that extends the Ravenclaw-type agent-based dialogue management model with topic management and multi-modal capabilities, especially with face recognition technologies, (2) a collection of WOZ-data related to initial encounters and presentation of information to the user, and (3) the establishment of a closer conversational relationship with the user by utilizing additional data (e.g. context, dialogue history, emotions, user goals, etc.).

09:20 - 09:40



Zoraida Callejas¹ and David Griol²

¹DEPT. OF LANGUAGES AND COMPUTER SYSTEMS, UNIVERSITY OF GRANADA (GRANADA, SPAIN),

²COMPUTER SCIENCE DEPARTMENT, CARLOS III UNIVERSITY OF MADRID (LEGANÉS, SPAIN)

An affective utility model of user motivation for counselling dialogue systems

Abstract:

Counselling dialogue systems are designed to help users to change and monitor their behaviours in order to achieve beneficial goals, such as the acquisition of healthy habits. To be effective, it is important that these systems include a model that accounts for the effort that users are investing to achieve the goals. However, most of the systems available nowadays carry out a naïve calculation based on the attained results, rather than on the reasons behind the successes or failures and their consequences for future user behaviour. In contrast to this, in this paper we propose a model that characterizes user motivation considering various aspects of psychological theories on subjective expected utility and attribution. Moreover, we provide a specification that allows carrying out calculations that replicate the users' decision-making process considering its emotional implications. The model is general-purpose and can be employed in standard architectures to make interpretations that adapt to each user, thus fostering more flexible and personalized interactions.

09:40 - 10:00



Marta Vicente and Elena Lloret

UNIVERSITY OF ALICANTE, DEPARTMENT OF SOFTWARE AND COMPUTING SYSTEMS

Exploring flexibility in natural language generation throughout discursive analysis of new textual genres

Abstract:

Since automatic language generation is a task able to enrich applications rooted in most of the language-related areas, from machine translation to interactive dialogue, it seems worthwhile to undertake a strategy focused on enhancing generation system's adaptability and flexibility. It is our first objective to understand the relation between the factors that contribute to discourse articulation in order to devise the techniques that will generate it. From that point, we want to determine the appropriate methods to automatically learn those factors. The role of genre on this approach remains essential as provider of the stable forms that are required in the discourse to meet certain communicative goals. The arising of new web-based genres and the accessibility of the data due to its digital nature, has prompted us to use reviews in our first attempt to learn the characteristics of their singular non-rigid structure. The process and the preliminary results are explained in the present paper.

Paper session #2 10:10 - 11:10

Session Chair: Francisco J Salguero

UNIVERSITY OF SEVILLE

10:10 - 10:30



Lea Canales¹, Carlo Strapparava²,
Ester Boldrini¹ and Patricio Martinez-Barco¹

¹UNIVERSITY OF ALICANTE, ALICANTE, SPAIN, ²FBK-IRST, TRENTO, ITALY

Bootstrapping technique + embeddings = Emotional corpus annotated automatically

Abstract:

Detecting depression or personality traits, tutoring and student behaviour systems, or identifying cases of cyber-bullying are a few of the wide range of the applications applications, in which the automatic detection of emotion is crucial. This task can contribute to the benefit of business, society, politics or education. The main objective of our research is focused on the improvement of the supervised emotion detection systems developed so far, through the definition and implementation of a technique to annotate large scale English emotional corpora automatically and with high standards of reliability. Our proposal is based on a bootstrapping process made up two main steps: the creation of the seed using NRC Emotion Lexicon and its extension employing the distributional semantic similarity through *words embeddings*. The results obtained are promising and allow us to confirm the soundness of the bootstrapping technique combined with the *word embedding* to label emotional corpora automatically.

10:30 - 10:50



Farhia Ahmed, Pierrette Bouillon, Chelle Destefano,
Johanna Gerlach, Angela Hooper, Manny Rayner,
Irene Strasly, Nikolaos Tsourakis and Catherine Weiss

UNIVERSITY OF GENEVA, FTI/TIM (SWITZERLAND),
GENEVA SOCIETY FOR THE DEAF, GENEVA (SWITZERLAND),
GYPSYSNAIL ARTS (SOUTH AUSTRALIA),
NABS INTERPRETING SERVICES (SOUTH AUSTRALIA),
SCHOOL OF GLOBAL, URBAN AND SOCIAL STUDIES, RMIT UNIVERSITY, MELBOURNE (VICTORIA)

Rapid construction of a web-enabled medical speech to sign language translator using recorded video

Abstract:

We describe an experiment in which sign-language output in Swiss French Sign Language (LSF-CH) and Australian Sign Language (Auslan) was added to a limited-domain medical speech translation system using a recorded video method. By constructing a suitable web tool to manage the recording procedure, the overhead involved in creating and manipulating the large set of files involved could be made easily manageable, allowing us to focus on the interesting and non-trivial problems which arise at the translation level. Initial experiences with the system suggest that the recorded videos, despite their unprofessional appearance, are readily comprehensible to Deaf informants, and that the method is promising as a simple short-term solution for this type of application.

10:50 - 11:10



Stephen Ash¹ and David Lin²

¹UNIVERSITY OF MEMPHIS, MEMPHIS, TN (USA), ²BAYLOR UNIVERSITY, WACO, TX (USA)

Incorporating syllable phonotactics to improve grapheme to phoneme translation

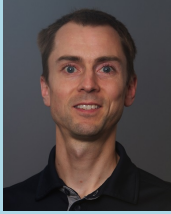
Abstract:

Grapheme to Phoneme (G2P) translation is a critical step in many natural language tasks such as text-to-speech production and automatic speech recognition. Most approaches to the G2P problem ignore phonotactical constraints and syllable structure information, and they rely on simple letter window features to produce pronunciations of words. We present a G2P translator which incorporates syllable structure into the prediction pipeline during structured prediction and re-ranking. In addition, most dictionaries contain only word-to-pronunciation pairs, which is a problem when trying to use these dictionaries as training data in a structured prediction approach to G2P translation. We present a number of improvements to the process of producing high-quality alignments of these pairs for training data. Together these two contributions improve the G2P word error rate (WER) on the CMUDict dataset by ~8%, achieving a new state-of-the-art accuracy level among open-source solutions.



11:10 - 11:30

Coffee break



11:30 - 12:30

Jason D. Williams
MICROSOFT RESEARCH

End-to-end deep learning of task-oriented dialog systems

Biography:

Jason D. Williams is with Microsoft Research in Redmond, Washington, USA. He has published more than 55 peer-reviewed papers on dialog systems and related areas in international conferences and journals, and has received five best paper/presentation awards for work on statistical approaches to dialog systems, including the use of POMDPs (partially observable Markov decision processes), reinforcement learning, turn-taking, and empirical user studies.

In 2012 he initiated the Dialog State Tracking Challenge series, in 2014 he shipped components of the first release of Microsoft Cortana, and in 2015 he launched Microsoft Language Understanding Intelligent Service (www.luis.ai).

He is Vice President of SIGDIAL, and an elected member of the IEEE Speech and Language Technical Committee (SLTC) in the area of spoken dialogue systems. Prior to Microsoft, Jason was with AT&T Labs Research 2006-2012, and also held several positions in industry building spoken dialog systems, including at Tellme Networks (now Microsoft) as Voice Application Development Manager.

Systems he has built over the past 15 years have conducted tens of millions of dialogs with real users.

Abstract:

This talk will present a model for end-to-end learning of task-oriented dialog systems. The main component of the model is a recurrent neural network which maps from a sequence of raw words directly to a distribution over system actions.

As compared to a conventional dialog system, the recurrent neural network takes the place of language understanding, dialog state tracking, and dialog control/policy. In addition, the developer can provide software that expresses business rules and provides access to programmatic APIs, enabling the network to take actions in the real world on behalf of the user. The neural network can be optimized using supervised learning (SL), where a domain expert provides example dialogs which the network should imitate; or using reinforcement learning (RL), where the system improves by interacting directly with end users.



12:30 - 13:30

Michael McTear
UNIVERSITY OF ULSTER

The birth of the conversational interface: a new kid on the block?

Biography:

Michael McTear is Emeritus Professor at the University of Ulster with a special research interest in spoken language technologies. He graduated in German Language and Literature from Queens University Belfast in 1965, was awarded an MA in Linguistics at the University of Essex in 1975, and a PhD at the University of Ulster in 1981. He has been Visiting Professor at the University of Hawaii (1986-87), the University of Koblenz, Germany (1994-95), and the University of Granada, Spain (2006-2010). He has been researching in the field of spoken dialogue systems for more than fifteen years and is the author of the widely used textbook Spoken dialogue technology: toward the conversational user interface (Springer, 2004).

He also is a co-author (with Kristiina Jokinen) of the book Spoken Dialogue Systems, (Morgan and Claypool, 2010), and (with Zoraida Callejas) of the book Voice Application Development for Android (Packt Publishing, 2013). He is co-author (with Zoraida Callejas and David Griol) of a new book entitled The Conversational Interface: Talking to Smart Devices (Springer, May 2016).

Abstract:

Conversational interfaces have become a hot topic. Major tech companies have been making huge investments in research into technologies such as AI, deep neural networks, machine learning, and natural language understanding with the aim of creating intelligent assistants (or bots) that will enable users to interact with information and services in a natural, conversational way. Yet the vision of the conversational interface is not new, and indeed there is a history of research in dialogue systems, chatbots, voice user interfaces, and embodied conversational agents that goes back more than fifty years.

This paper explores what has changed to make the conversational interface relevant today and examines some key issues from earlier work that could inform the next generation of conversational systems.



13:30 - 15:00

Lunch



15:00 - 16:00

German Rigau

UNIVERSITY OF THE BASQUE COUNTRY

Big data for natural language processing

Biography:

German Rigau Ph.D. and B.A. in Computer Science by the Universitat Politècnica de Catalunya (UPC). Formerly member of the Computer Science department at the UPC and member of the TALP research group of the UPC, currently, he is teaching at the Computer Science Faculty of the EHU as an Associate Professor. He has published more than hundred-refereed articles and conference papers in the area of Natural Language Processing, and in particular Acquisition of Lexical Knowledge, Word Sense Disambiguation, Semantic Processing and Inferencing.

He has been involved in several European research projects (ACQUILEX, ACQUILEX II, EuroWordNet, NAMIC, MEANING, KYOTO, PATHS, OpeNER and NewsReader). He coordinated the MEANING project (IST-2001-34460) and the local groups for NAMIC, KYOTO, OpeNER and NewsReader. He has been also involved in several Spanish National research projects (ITEM, HERMES, SENSEM, KNOW, KNOW2, SKaTer and TUNER). Currently, he is coordinating the TUNER project.

He served as PC member and reviewer of the main international conferences and workshops in NLP and AI including ACL, EACL, NAACL, COLING, AAAI, ECAI, IJCAI, EMNLP, IJCNLP, CoNLL, TSD, SENSEVAL/SEMEVAL and IWC. He also served as reviewer of International Journals including: Computers and the Humanities, Journal of Natural Language Engineering, Journal of Artificial Intelligence Research and Artificial Intelligence. He has also participated in all editions of the international competition of SENSEVAL.

Currently, he is member of the Association for Computational Linguistics (ACL) and the Spanish Society for Natural Language Processing (SEPLN).

Abstract:

Requirements in computational power have grown dramatically in recent years. This is also the case in many language processing tasks, due to the overwhelming and ever increasing amount of textual information that must be processed in a reasonable time frame. This scenario has led to a paradigm shift in the computing architectures and large-scale data processing strategies used in the Natural Language Processing field.

This talk presents new distributed architectures and technology for scaling up text analysis running complete chains of linguistic processors on parallel architectures. The talk also describes a series of experiments carried out with the goal of analyzing the scaling capabilities of the current language processing pipelines on large clusters with many processing nodes.



16:00 - 16:30

Coffee break

Paper session #3

16:30 - 17:50

Session Chair: Bente Maegaard

UNIVERSITY OF COPENHAGEN



16:30 - 16:50

Alejandro Tarafa Guzman and Ademar Ferreira

POLYTECHNIC SCHOOL OF THE UNIVERSITY OF SÃO PAULO, BUTANTÃ, SÃO PAULO (BRAZIL)

Hybrid conceptual and statistical measure for semantic textual similarity evaluation

Abstract:

We propose in this paper an improved methodology to evaluate semantic textual similarity between two sentences. Our model integrates semantic and statistical information by means of a chunking parser in such a way that the combination is inherently internal to the overall system. Evaluation results with SemEval 2016 sentence data sets are encouraging.

16:50 - 17:10



Fernando Giner Martínez and Enrique Amigó Cabrera

DEPARTAMENTO DE LENGUAJES Y SISTEMAS INFORMÁTICOS, UNED, MADRID (SPAIN)

General representation model for text similarity

Abstract:

Text similarity is a central issue in multiple information access tasks. General speaking, most of existing similarity models focus on a particular kind of text features such as words, n-grams, or linguistic features or distributional semantics units. In this paper, we introduce a general theoretical model for integrating multiple sources in the text feature representation called Feature Projection Information model. The proposed model allows us to integrate traditional features such as words with other sources such as the output of classifiers over different categories or distributional semantics information. The theoretical analysis shows that traditional approaches can be seen as particularizations of the model. Our first empirical results support the idea that additional features in the representation step outperform the predictive power of similarity measures.

17:10 - 17:30



Alberto Poncelas, Andy Way and Antonio Toral

ADAPT CENTRE, SCHOOL OF COMPUTING, DUBLIN CITY UNIVERSITY, DUBLIN (IRELAND)

Extending feature decay algorithms using alignment entropy

Abstract:

In machine-learning applications, data selection is of crucial importance if good runtime performance is to be achieved. Feature Decay Algorithms (FDA) have demonstrated excellent performance in a number of tasks. While the decay function is at the heart of the success of FDA, its parameters are initialised with the same weights. In this paper, we investigate the effect on Machine Translation of assigning more appropriate weights to words using word-alignment entropy. In experiments on German to English, we show the effect of calculating these weights using two popular alignment methods, GIZA++ and FastAlign, using both automatic and human evaluations. We demonstrate that our novel FDA model is a promising research direction.

17:30 - 17:50



**Daniel Bruggemann, Yannik Hermeij, Carsten Orth,
Darius Schneider, Stefan Selzer and Gerasimos Spanakis**

DEPARTMENT OF DATA SCIENCE AND KNOWLEDGE ENGINEERING, MAASTRICHT UNIVERSITY, MAASTRICHT (NETHERLANDS)

Towards a topic discovery and tracking system with application to news items

Abstract:

Rapid proliferation of the World Wide Web led to an enormous increase in the availability of textual corpora. In this paper, the problem of topic detection and tracking is considered with application to news items. The proposed approach explores two algorithms (Non-Negative Matrix Factorization and a dynamic version of Latent Dirichlet Allocation (DLDA)) over discrete time steps and makes it possible to identify topics within storylines as they appear and track them through time. Moreover, emphasis is given to the visualization and interaction with the results through the implementation of a graphical tool (regardless the approach). Experimental analysis on Reuters RCV1 corpus and the Reuters 2015 archive reveals that explored approaches can be effectively used as tools for identifying topic appearances and their evolutions while at the same time allowing for an efficient visualization.



18:30 - 19:30

Guided visit to University of Seville main building



20:30

Social dinner

ADDRESS: SALÓN ANDALUCÍA - HOTEL ALFONSO XIII - CALLE SAN FERNANDO, 2

2ND DECEMBER



Venue: Paraninfo (University of Seville main building)

ADDRESS: REAL FÁBRICA DE TABACOS. CALLE SAN FERNANDO S/N



09:30 - 10:30

António Branco
UNIVERSITY OF LISBON

With or without meaning? With or without funding?: Cycles in language technology research and what we may learn from them

Biography:

António Branco is the Director of the Portuguese node of the CLARIN research infrastructure. He is a professor of language science and technology at the University of Lisbon, where he was the founder and is the head of research of the Natural Language and Speech Group (NLX Group) of the Department of Informatics. He is the (co-)author of over 150 publications in the area of language science and technology and has participated and coordinated several national and international R&D projects. He was the coordinator of the European project METANET4U, integrating the R&D network of excellence META-NET. He is a member of the META-NET Executive Board and he is the first author of the White Paper on the Portuguese Language in the Digital Age.

He is coordinating the QTLeap project, an European research project on quality machine translation by deep language engineering approaches.

Abstract:

As language technology has now been an increasingly active research area for long enough time in the last six decades, it offers itself to be the object of a meta-research exercise of seeking to devise successive epistemological and institutional paradigms its progress may have been based upon.

In this talk, I will seek to support and share the view that, in line with the structure of the scientific (r)evolutions that other research areas may go through, also here scientific evolution seems to go through a succession of different epistemological paradigms and institutional supporting schemes. Interestingly, this progression seems to exhibit a somewhat cyclic shape.

There was an initial paradigm where, no matter remote it might be with respect the natural language processing task of interest, the representation of meaning was assumed to be the ultimate cornerstone of the technology to be researched.

From mid nineties onward, given the stalemate of endless new proposals of yet another fragment of some logics to represent natural language meaning with no clearly visible impact on the progress of the area, and having statistical machine translation as one of its major driving application, a new paradigm rapidly spread based initially on the challenging assumption that research could very well be advanced without resorting to semantic representation.

Some twenty years later, given the stalemate of endless new proposals of yet another cocktail of statistical techniques with no clearly perceived cumulative impact on the progress of the area beyond the few positive centesimal points of the delta measured on the specific dataset and domain that is being resorted to, this is the paradigm that is nowadays giving way to a come back of the view that semantic representation is key to leverage the progress of research in language technology.

It is worth noting, though, that this renaissance of the semantic representation primacy is not a mere come back to the logic-based representation of meaning but to a richer view of meaning representation that crucially encompasses also distributional semantics, a vector-based representation of meaning.

Interestingly, on the institutional side, not only a cyclic shaped evolution seems also to emerge in the support that this research area has received at the European level, as its shape seems to be pretty much in phase with the paradigmatic shifts occurred on the epistemological side.

Given the promotion of multilingualism inscribed in its core mission, the interest of the European Commission on language technology was initially focused in machine translation as an advanced support for its translation and multilingual services, which represented around 1/3 of the running costs of the Commission. This interest materialized in the EUROTRA project, a massive, in house research and development project aimed at setting up machine translation software to serve its internal needs and run under the coordination of the services of the Commission, with major components being commissioned to third parties identified by the Commission in the different member states.

By mid nineties, as this machine translation project came to an end, the Commission broadened its interest to the full range of the language technology area. By assuming this was a first rank scientific area – and also a political priority – on a par with any other discipline worth being supported, the Commission supported a series of full-scale, competitive funding schemes and opportunities aiming at the progress of language technology that were aligned with similar schemes used for other research areas.

After being supported during three R&D framework programs, in the current H2020 program, funding of language technology by the European Commission is again focused on machine translation, and again narrowed down to a particular in house project, CEFAT, which is being run under the coordination of the Commission and with components being commissioned to third parties, with the major goal of addressing the specific needs of its services for multilingual technology.

While proceeding with the meta-analysis exercise just sketched above, and in line with the objectives of the present workshop on "Future and Emerging Trends in Language Technologies" (Seville, December 1-2, 2016), in this talk I will ponder and seek to foster the discussion on which long term progression lines may be devised in the scope of such an exercise, and if and how these lines can eventually help us to devise future and emerging trends in our area, which in the epistemological side should thus be pursued, and in the funding side should be fostered.



10:30 - 11:00

Coffee break



11:00 - 12:00

David Pérez Fernández

MINISTRY OF INDUSTRY, ENERGY AND TOURISM

Spanish plan for the impulse of language technologies

Biography:

David Pérez is currently advisor of the Secretary of State for Telecommunications and Information Society inside Ministry of Industry, Energy and Tourism (Spanish Government).

Abstract:

The Plan to promote language technologies (LT-Plan) aims to encourage industrial sector development of natural language processing and machine translation in Spanish and regional languages.

The LT-Plan establishes measures to increase the number, quality and availability of language infrastructures (Spanish and co-official languages), boost language industry by knowledge transfer from research sector and incorporate Public Administration as a driver of natural language processing and machine translation sector.

The Plan aims to improve language technologies in a coordinated manner, seeking synergies and avoiding duplication of efforts, in accordance with the recommendations of the Commission for the Reform of Public Administration (CORA). LT-Plan is structured into four main areas:

Area I: Support for the development of linguistic infrastructure.

This work area aims to facilitate the development of processors (entities recognizers, taggers, semantic similarity measures, etc.) and linguistic resources (parallel corpora for machine translation, dictionaries, taxonomies, etc.) that serve as fuel for the development of the processing industry Spanish natural language and machine translation.

Area II: Promotion of Language Technology Industry.

The second area corresponds to the objective of supporting the transfer of knowledge between research sector and industry, as well as to promote

internationalization of companies and institutions that belong to Language Technology sector.

Area III: Public Administration as a driver of the Language Industry.

It's proposed the creation of open common platforms for language processing and machine translation to accelerate Public Administration and Industry technology adoption. In addition, public policy of open data on public sector information (named RISP in Spain) represents a channel for the development of important linguistic resources (named entities, toponyms, parallel corpora, etc.).

Area IV: Lighthouse projects based on language technologies.

Fourth area is oriented to promotion of the lighthouse projects, based on application of technology of natural language, initially undertaken by the government in strategic sectors (Health, Education, Tourism, etc.). LT-Plan is open to public private initiatives

This projects are intended as a demonstration of language technology capabilities and benefits, generating industry and creating reusable resources for industrial application. They also serve as real scenery for learning and improve future developments.



12:00 - 13:00

Philippe Wacker

LT-INNOVATE

Improving the collaboration between industry and research: offer and demand

Biography:

Philippe Wacker is the Secretary General of LT-Innovate, the Association of the Language Technology Industry (www.lt-innovate.org).

A multilingual and cross-cultural manager with a strategic outlook and broad international experience, he has managed leading international ICT and high technology networks for more than 25 years. His skills set encompasses association and interest group management, international project management, strategic project planning and development, international project finance, event management, international networking and business expansion, European public affairs and lobbying.

www.linkedin.com/in/philippewacker

Abstract:

I will present an overview of LT-Innovate activities, programmes and services of interest to language technology researchers. In addition, I will be reporting on a survey we will be conducting in the autumn about industry's needs in terms of R&D in the field of language technology.



13:00 - 13:30

Closure ceremony

Dr. Teresa López

ORGANIZING COMMITTEE FETLT16

Dr. Jose F. Quesada

PROGRAM CHAIR FETLT16

DPT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE. UNIVERSITY OF SEVILLE
DIRECTOR ANDALUSIAN SCIENTIFIC INFORMATION SYSTEM. REGIONAL GOVERNMENT OF ANDALUCIA

Dr. Francisco José González Ponce

DEAN, SCHOOL OF PHILOLOGY. UNIVERSITY OF SEVILLE

Dr. José Luis Sevillano

DEAN, SCHOOL OF COMPUTER SCIENCE. UNIVERSITY OF SEVILLE

Dr. Pedro Bisbal

ANDALUSIAN AGENCY OF KNOWLEDGE