

**EIGHTH INTERNATIONAL CONFERENCE ON
LANGUAGE RESOURCES AND EVALUATION**

*Held under the Patronage of Ms Neelie Kroes, Vice-President of the
European Commission, Digital Agenda Commissioner*

MAY 21 – 27, 2012

CONFERENCE PROGRAMME

For more information about ELRA:

ELRA-ELDA

55/57, rue Brillat Savarin

75013 Paris – FRANCE

Tel. : +33 1 43 13 33 33

Fax : + 33 1 43 13 33 30

info@elda.org

© ELRA – European Language Resources Association. All rights reserved.

LREC 2012 Committees

Conference Programme Committee

Nicoletta Calzolari	ILC/CNR, Pisa, Italy (Conference chair)
Khalid Choukri	ELRA, Paris, France
Thierry Declerck	DFKI GmbH, Saarbrücken, Germany
Mehmet Uğur Doğan	Tübitak-BILGEM, Istanbul, Turkey
Bente Maegaard	CST, University of Copenhagen, Denmark
Joseph Mariani	LIMSI-CNRS, Orsay, France
Jan Odijk	UIL-OTS, Utrecht, The Netherlands
Asuncion Moreno	Universitat Politècnica de Catalunya, Barcelona, Spain
Stelios Piperidis	ILSP, Athens, Greece

International Advisory Committee

Georges Carayannis	National Technical University of Athens, Athens, Greece
Roberto Cencioni	Head of Language Technologies and Machine Translation Unit, Directorate General for Information Society and Media, European Commission Luxembourg
Charles J. Fillmore	University of California at Berkeley, USA
Hiroya Fujisaki	Science University of Tokyo, Japan
Sadaoki Furui	Tokyo Institute of Technology, Tokyo, Japan
Tatiana Korelsky	Human Language and Communication, Robust Intelligence Cluster, Division of Information and Intelligent Systems, Directorate of Computer and Information Science and Engineering, National Science Foundation, Arlington, USA
Terence Langendoen	Professor Emeritus, Department of Linguistics, University of Arizona, and Expert, Division of Information and Intelligent Systems, National Science Foundation, USA
Makoto Nagao	President of the National Diet Library, Tokyo, Japan
Bernard Quemada	Conseil Supérieur de la Langue Française, Paris, France
Piet G.J. Van Sterkenburg	International Permanent Committee of Linguists - Leiden, The Netherlands
Shiwen Yu	School of Electronics Engineering & Computer Science, Beijing University, China

Scientific Committee

The Programme Committee is very grateful to Scientific Committee members who reviewed the submissions and contributed to designing the conference programme. The list of the members of Scientific Committee is published on the LREC 2012 web site.

Local Committee

Yücel Bicil	TÜBİTAK - BİLGEM
Assist. Prof. Barış Bozkurt	İzmir Institute of Technology
Dr. Ruken Çakıcı	Middle East Technical University
Assist. Prof. Gülşen Cebiroğlu Eryiğit	İstanbul Technical University
Assoc. Prof. Tolga Çiloğlu	Middle East Technical University
Dr. Şeniz Demir	TÜBİTAK - BİLGEM
Assist. Prof. Cenk Demiroğlu	Özyeğin University
Assist. Prof. Banu Diri	Yıldız Technical University
Dr. İlknur Durgar El-Kahlout	TÜBİTAK - BİLGEM
Assist. Prof. Hakan Erdoğan	Sabancı University
Assist. Prof. Ümit Güz	Işık University
Alper Kanak	TÜBİTAK - BİLGEM
Dr. Oğuzhan Külekçi	TÜBİTAK - BİLGEM
Coşkun Mermer	TÜBİTAK - BİLGEM
Assist. Prof. Zeynep Orhan	Fatih University
Dr. Arzucan Özgür	Boğaziçi University
Prof. Dr. Mehmed Özkan	Boğaziçi University (Chair)
Assist. Prof. Murat Saraçlar	Boğaziçi University
Assist. Prof. Cüneyd Tantuğ	İstanbul Technical University
Assist. Prof. Deniz Yüret	Koç University
Prof. Dr. Deniz Zeyrek	Middle East Technical University

LREC Organising Committee

Cécile Barbier	ELDA/ELRA, Paris, France
Roberto Bartolini	ILC/CNR, Pisa, Italy
Sara Goggi	ILC/CNR, Pisa, Italy (Co-chair)
Olivier Hamon	ELDA/ELRA, Paris, France
Hélène Mazo	ELDA/ELRA, Paris, France (Co-chair)
Vincenzo Parrinelli	ILC/CNR, Pisa, Italy
Caroline Rannaud	ELDA/ELRA, Paris, France
Irene Russo	ILC/CNR, Pisa, Italy

Acknowledgements

The European Language Resources Association, ELRA, and the LREC Committees acknowledge with gratitude the support and sponsoring of the following institutions.

Sponsors and supporters

CELI (<i>Silver</i>)
EML European Media Laboratory (<i>Bronze</i>)
K-Dictionaries (<i>Bronze</i>)
IMMI (<i>Bronze</i>)
MetaNet (<i>Bronze</i>)
Nuance (<i>Silver</i>)
Quaero (<i>Bronze</i>)

Supporting Institutions

Evaluations and Language resources Distribution Agency (ELDA), Paris (France)
Istituto di Linguistica Computazionale (ILC) of the Italian National Research Council (CNR), Pisa (Italy)
Bogazici University, Istanbul (Turkey)
Tubitak-Bilgem, Istanbul (Turkey)

Pre-conference Workshops and Tutorials

Monday, May 21st

Topkapi A	Topkapi B	Dolmabahçe A	Dolmabahçe B	Dolmabahçe C	Sultan 3	Sultan 1
Morning						
T17 Rich Prior Knowledge in Learning for Reducing Annotation Cost	T113 NooJ	T105 ISocat	T81 Bootstrapping Ontology Evolution	W112 WILDRE Workshop on Indian Language Data: Resources and Evaluation	W100 -100A IPR Issues	T134 Computer assisted terminology processing
<i>Lunch break</i>						
Afternoon						
T115 META-SHARE	W18 Best Practices for Speech Corpora in Linguistic Research	T123 Rich Prior Knowledge in Learning for Reducing Annotation Cost	T98 Enriching the Web with Ontology-Lexica	W114 Workshop on Language Resources and Technologies for Turkic Languages	W100 -100A IPR Issues	T127 Using games to model and recognize interactions

Tuesday, May 22nd

Topkapi A	Topkapi B	Dolmabahçe A	Dolmabahçe B	Dolmabahçe C	Sultan 3	Sultan 1
Morning						
W83 Semantic relations-II	W110 Multimodal Corpora: How should multimodal corpora deal with the situation?	W51 SALT MIL & AFlaT2012: Language technology for normalisation of less-resourced languages	W87 Describing Language Resources with Metadata	W89 META-RESEARCH Workshop on Advanced Treebanking	W57 LRE-Rel: Language Resources and Evaluation for Religious Texts	T25 CHILDES
<i>Lunch break</i>						
Afternoon						
W83 Semantic relations-II	W110 Multimodal Corpora: How should multimodal corpora deal with the situation?	W51 SALT MIL & AFlaT2012: Language technology for normalisation of less-resourced languages	W87 Describing Language Resources with Metadata	W121 Challenges in the management of large corpora	W129 Merging of Language Resources	W54 ColabTKR 2012 Collaboration in Terminology & Knowledge Representation

Post-conference Workshops and Tutorials

Saturday, May 26th

Topkapi A	Topkapi B	Dolmabahçe A	Dolmabahçe B	Dolmabahçe C	Sultan 3	Sultan 1
Morning						
W49 Building and Using Comparable Corpora	W93 ES ³ 2012 Corpora for Research on Emotion, Sentiment and Social Signals	W80 CoCo-FLaRE: Reinforcing International Collaboration on LRE	W111 Third Workshop on Building and Evaluating Resources for Biomedical Text Mining	W102 Adaptation of Language Resources and Tools for Processing Cultural Heritage	W120 Language Engineering for Online Reputation Management	W6 Joint ISA-7, SRSL-3, and I2MRT
<i>Lunch break</i>						
Afternoon						
W49 Building and Using Comparable Corpora	W93 ES ³ 2012 Corpora for Research on Emotion, Sentiment and Social Signals	W80 CoCo-FLaRE: Reinforcing International Collaboration on LRE	W111 Third Workshop on Building and Evaluating Resources for Biomedical Text Mining	W46 Computational Models of Narrative	W103 @NLP can u tag #user_generated content?! via lrec-conf.org	W6 Joint ISA-7, SRSL-3, and I2MRT

Sunday, May 27th

Topkapi A	Topkapi B	Dolmabahçe A	Dolmabahçe A	Dolmabahçe C	Sultan 3	Sultan 1
Morning						
W77 Representation and Processing of Sign Languages	W134 Collaborative Resource Development and Delivery	W84 NLP4ITA Natural Language Processing for Improving Textual Accessibility	W88 EEO2012 Exploring and Exploiting Official Publications	W46 Computational Models of Narrative	W97 SPLeT-2012 Semantic Processing of Legal Texts	
<i>Lunch break</i>						
Afternoon						
W77 Representation and Processing of Sign Languages	W134 Collaborative Resource Development and Delivery	W47 Language Resources for Public Security Applications	W86 CREDISLAS Creating Cross-Language Resources for Disconnected Languages and Styles	W46 Computational Models of Narrative	W97 SPLeT-2012 Semantic Processing of Legal Texts	

Conference Programme at a Glance

Morning

23 rd May 2012	24 th May 2012	25 th May 2012
	09:00-09:40 Keynote Speech Sören Auer: The Web of Data: Decentralized, collaborative, interlinked and interoperable <i>Anadolu Auditorium</i>	09:00-09:40 Keynote Speech Catherine Pelachaud: Socio- emotional virtual agent <i>Anadolu Auditorium</i>
09:45-10:30	09:45-11:25 4 Oral and 4 Poster Sessions in Parallel ▪ Infrastructures and Strategies for LRs (2) <i>Anadolu Auditorium</i> ▪ Dialogue <i>Topkapi A</i> ▪ Resource Creation and Acquisition <i>Topkapi B</i> ▪ Corpus and Annotation <i>Dolmabahçe A</i>	09:45-11:25 4 Oral and 4 Poster Sessions in Parallel ▪ Subjectivity and Emotions <i>Anadolu Auditorium</i> ▪ Names Entities <i>Topkapi A</i> ▪ Treebanks and Syntax <i>Topkapi B</i> ▪ Semantic Lexica and Semantic Annotation <i>Dolmabahçe A</i>
10-30-11:15 Introductory Session: Language Resources and Evaluation: Present and Future <i>Anadolu Auditorium</i>	Posters in Poster Area 1 ▪ Semantic Annotation ▪ Document Classification, Text Categorisation ▪ Grammar and Syntax ▪ Digital Libraries	Posters in Poster Area 1 ▪ Corpus Creation, Processing, Usage (3) ▪ Language Resource Infrastructures (1) ▪ Speech Synthesis ▪ Speech Resources
11:35-13:15 4 Oral and 3 Poster Sessions in Parallel ▪ Corpora for Machine Translation <i>Anadolu Auditorium</i> ▪ Infrastructures and Strategies for LRs (1) <i>Topkapi A</i> ▪ Semantics <i>Topkapi B</i> ▪ Speech Corpora <i>Dolmabahçe A</i> Posters in Poster Area 1 and 2 ▪ Anaphora and Coreference ▪ Tools, Systems and Evaluation ▪ Lexical Resources	11:45-13:05 4 Oral and 3 Poster Sessions in Parallel ▪ Speech Corpora and Tools <i>Anadolu Auditorium</i> ▪ Machine Translation and Evaluation (2) <i>Topkapi A</i> ▪ Semantic Resources <i>Topkapi B</i> ▪ Trends in Corpora <i>Dolmabahçe A</i> Posters in Poster Area 2 ▪ Treebanks ▪ Parsing ▪ Information Extraction (2)	11:45-13:25 4 Oral and 6 Poster Sessions in Parallel ▪ Machine Translation and Language Resources <i>Anadolu Auditorium</i> ▪ WordNets <i>Topkapi A</i> ▪ Text Mining <i>Topkapi B</i> ▪ Evaluation of Systems and Applications <i>Dolmabahçe A</i> Posters in Poster Area 2 ▪ Subjectivity: Sentiments, Emotions, Opinions (2) ▪ Language Resource Infrastructures (3) ▪ Knowledge and Ontologies ▪ Semantics ▪ Temporal Information ▪ Sign Language
	13:10-13:30 Invited Talk Kemal Oflazer: The Turkish Language and its Challenges for Language Processing <i>Anadolu Auditorium</i>	

Conference Programme at a Glance Afternoon

23 rd May 2012	24 th May 2012	25 th May 2012
14:45-16:25	14:55-16:35	14:55-16:35
<p>4 Oral and 4 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Crowdsourcing (Special Session) <i>Anadolu Auditorium</i> ▪ Dialogue and Multimodality <i>Topkapi A</i> ▪ Machine Translation and Language Resources <i>Topkapi B</i> ▪ Corpus Processing and Infrastructure <i>Dolmabahçe A</i> 	<p>4 Oral and 4 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Multimodal Corpora (2) <i>Anadolu Auditorium</i> ▪ Child Language Corpus <i>Topkapi A</i> ▪ MultiWord Expressions <i>Topkapi B</i> ▪ Sign Language <i>Dolmabahçe A</i> 	<p>4 Oral and 5 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ New Media (Special Session) <i>Anadolu Auditorium</i> ▪ Semantics, Knowledge and Ontologies <i>Topkapi A</i> ▪ Segmentation, Tagging, Parsing <i>Topkapi B</i> ▪ Named Entities and Subjectivity <i>Dolmabahçe A</i>
<p><i>Posters in Poster Area 2</i></p> <ul style="list-style-type: none"> ▪ Information Extraction (1) ▪ Word Sense Disambiguation and Evaluation ▪ Multiword Expressions and Term Extraction ▪ Authoring Tools, Proofing 	<p><i>Posters in Poster Area 1</i></p> <ul style="list-style-type: none"> ▪ Part of Speech Tagging ▪ Machine Translation (1) ▪ Corpus Creation, Processing, Usage (1) ▪ Evaluation Methodologies 	<p><i>Posters in Poster Area 1</i></p> <ul style="list-style-type: none"> ▪ Machine Translation (2) ▪ Natural Language Generation ▪ Crowdsourcing ▪ Text Mining and Text Entailment ▪ Speech/Multimodal Tools, Systems, Applications
16:45-18:05	16:55-18:15	16:55-17:55
<p>4 Oral & 3 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Endangered Languages <i>Anadolu Auditorium</i> ▪ Document Classification, Text Categorisation <i>Topkapi A</i> ▪ Discourse (1) <i>Topkapi B</i> ▪ Word Sense Disambiguation <i>Dolmabahçe A</i> 	<p>4 Oral and 4 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Language Generation and Paraphrasing <i>Anadolu Auditorium</i> ▪ Computer-aided Language Learning <i>Topkapi A</i> ▪ Discourse (2) <i>Topkapi B</i> ▪ Syntax & Parsing <i>Dolmabahçe A</i> 	<p>Antonio Zampolli Prize Talk <i>Anadolu Auditorium</i></p>
<p><i>Posters in Poster Area 1</i></p> <ul style="list-style-type: none"> ▪ Morphology ▪ Prosody and Phonetics ▪ Language Resource Infrastructures (1) 	<p><i>Posters in Poster Area 2</i></p> <ul style="list-style-type: none"> ▪ Multilinguality ▪ Question Answering and Summarization ▪ Multimodal Corpus for Interaction ▪ Ontologies 	
18:10-19:30	18:20-19:40	17:55-18:15
<p>4 Oral and 3 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Multimodal Corpora (1) <i>Anadolu Auditorium</i> ▪ Machine Translation and Evaluation (1) <i>Topkapi A</i> ▪ Information Extraction and Question Answering <i>Topkapi B</i> ▪ Web Services <i>Dolmabahçe A</i> 	<p>4 Oral and 4 Poster Parallel Sessions</p> <ul style="list-style-type: none"> ▪ Semantics from Corpora <i>Anadolu Auditorium</i> ▪ Authoring and Related Tools <i>Topkapi A</i> ▪ Word Sense Annotation and Disambiguation <i>Topkapi B</i> ▪ Time and Space <i>Dolmabahçe A</i> 	<p>Closing Session by Programme Committee <i>Anadolu Auditorium</i></p>
<p><i>Posters in Poster Area 2</i></p> <ul style="list-style-type: none"> ▪ Subjectivity: Sentiments, Emotions, Opinions (1) ▪ Named Entity Recognition ▪ Dialogue 	<p><i>Posters in Poster Area 2</i></p> <ul style="list-style-type: none"> ▪ Discourse ▪ Lexical Acquisition ▪ Corpus Creation, Processing, Usage (2) ▪ Web Services 	
21:00: Welcome Reception		21:00: Gala Dinner

Special Session: EC Projects, a broad overview
Thursday 24 May 2012 – 14:55 – 19:30p.m.

The European Commission has launched some 60 language-technology projects during the last few years. Some of these projects will be briefly presented highlighting their main innovations and the challenges that should be addressed by follow-on actions.

14:55	Opening and introduction	<i>Roberto Cencioni</i>	EC representative
--------------	--------------------------	-------------------------	-------------------

15:00	Making LR fit for online consumption and reuse	<i>Stelios Piperidis, Tamas Varadi, Andrej Vasiljevs, Antonio Branco</i>	META-SHARE, CESAR, METANORD and METANET4YOU
--------------	--	--	--

16:00	Advancing methods and techniques for compiling LR	<i>Nuria Bel, Marko Tadić</i>	PANACEA & ACCURAT
--------------	---	-------------------------------	----------------------

16:35 break

16:55	Information analysis: text and social media mining	<i>Allan Hanbury, Marko Grobelnik, Thierry Declerck, Paul Buitelaar</i>	KHRESMOI, X-LIKE, TRENDMINER and MONNET
--------------	--	---	--

17:50	Advanced text and speech translation services	<i>Raivis Skadiņš, Volker Steinbiss</i>	LETS MT! and EU-BRIDGE
--------------	---	---	---------------------------

18:25	Evaluation and standards as key R&D enablers	<i>Nicola Ferro, Felix Sasaki</i>	PROMISE and LT-WEB
--------------	--	---------------------------------------	-----------------------

19:00	Federating research efforts: towards a shared vision	<i>Hans Uszkoreit</i>	META-NET
--------------	--	-----------------------	----------

19:40	Closing remarks, upcoming opportunities	<i>Roberto Cencioni</i>	
--------------	---	-------------------------	--

Detailed Conference Programme

23 May 2012

9:45-10:30 **Opening Ceremony in *Anadolu Auditorium***

10:30-11:15 **Introductory Session:**

Language Resources and Evaluation: Present and Future

- Language Technologies and Resources: progress and prospects, Roberto Cencioni
- Meta-Share is here! Present use and future developments, Stelios Piperidis
- LRE Map and Language Library: a service to the community, Nicoletta Calzolari
- ELRA-Forum: a world of services for the LRE friends, Khalid Choukri

▪ ***Anadolu Auditorium***

11:15-11:35 ***Coffee break***

O1 - Corpora for Machine Translation

Chairperson: Josef van Genabith

Anadolu Auditorium

11:35-11:55 PaCo2: A Fully Automated tool for gathering Parallel Corpora from the Web
Iñaki San Vicente and Iker Manterola

11:55-12:15 Terra: a Collection of Translation Error-Annotated Corpora
Mark Fishel, Ondřej Bojar and Maja Popović

12:15-12:35 A light way to collect comparable corpora from the Web
Ahmet Aker, Evangelos Kanoulas and Robert Gaizauskas

12:35-12:55 SUMAT: Data Collection and Parallel Corpus Compilation for Machine Translation of Subtitles
Volha Petukhova, Rodrigo Agerri, Mark Fishel, Sergio Penkale, Arantza del Pozo, Mirjam Sepesy Maucec, Andy Way, Panayota Georgakopoulou and Martin Volk

12:55-13:15 The FAUST Corpus of Adequacy Assessments for Real-World Machine Translation Output
Daniele Pighin, Lluís Màrquez and Lluís Formiga

O2 - Infrastructures and Strategies for LRs (1)

Chairperson: Hans Uszkoreit

Topkapı A

11:35-11:55 The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions
Stelios Piperidis

- 11:55-12:15 The Language Library: supporting community effort for collective resource production
Riccardo Del Gratta, Francesca Frontini, Francesco Rubino, Irene Russo and Nicoletta Calzolari
- 12:15-12:35 Using the International Standard Language Resource Number: Practical and Technical Aspects
Jungyeul Park, Victoria Arranz, Olivier Hamon and Khalid Choukri
- 12:35-12:55 ELRA in the heart of a cooperative HLT world
Valérie Mapelli, Victoria Arranz, Matthieu Carré, Hélène Mazo, Djamel Mostefa and Khalid Choukri
- 12:55-13:15 Twenty Years of Language Resource Development and Distribution: A Progress Report on LDC Activities
Christopher Cieri, Marian Reed, Denise DiPersio and Mark Liberman

O3 - Semantics

Chairperson: Bolette Pedersen

Topkapı B

- 11:35-11:55 Polaris: Lymba's Semantic Parser
Dan Moldovan and Eduardo Blanco
- 11:55-12:15 Automatic classification of German "an" particle verbs
Sylvia Springorum, Sabine Schulte im Walde and Antje Roßdeutscher
- 12:15-12:35 Pragmatic identification of the witness sets
Livio Robaldo and Jakub Szymanik
- 12:35-12:55 Evaluating automatic cross-domain Dutch semantic role annotation
Orphée De Clercq, Veronique Hoste and Paola Monachesi
- 12:55-13:15 Logic Based Methods for Terminological Assessment
Benoît Robichaud

O4 - Speech corpora

Chairperson: Sophie Rosset

Dolmabahçe A

- 11:35-11:55 KALAKA-2: a TV Broadcast Speech Database for the Recognition of Iberian Languages in Clean and Noisy Environments
Luis Javier Rodriguez-Fuentes, Mikel Penagarikano, Amparo Varona, Mireia Diez and German Bordel
- 11:55-12:15 The C-ORAL-BRASIL I: Reference Corpus for Spoken Brazilian Portuguese
Tommaso Raso, Heliana Mello and Maryualê Malvessi Mittmann
- 12:15-12:35 The ETAPE corpus for the evaluation of speech-based TV content processing in the French language
Guillaume Gravier, Gilles Adda, Niklas Paulsson, Matthieu Carré, Aude Giraudel and Olivier Galibert
- 12:35-12:55 Automatic Speech Recognition on a Firefighter TETRA Broadcast Channel
Daniel Stein and Bela Usabaev
- 12:55-13:15 TED-LIUM: an Automatic Speech Recognition dedicated corpus
Anthony Rousseau, Paul Deléglise and Yannick Estève

P1 - Anaphora and Coreference

Chairperson: Ineke Schuurman

QurAna: Corpus of the Quran annotated with Pronominal Anaphora
Abdul-Baqee Sharaf and Eric Atwell

The Use of Parallel and Comparable Data for Analysis of Abstract Anaphora in German and English
Stefanie Dipper, Melanie Seiss and Heike Zinsmeister

Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component
Lucie Poláková, Pavlína Jinová and Jiří Mirovský

A Portuguese-Spanish Corpus Annotated for Subject Realization and Referentiality
Luz Rello and Iria Gayo

Coreference in Spoken vs. Written Texts: a Corpus-based Analysis
Marilisa Amoia, Kerstin Kunz and Ekaterina Lapshinova-Koltunski

Annotating Near-Identity from Coreference Disagreements
Marta Recasens, M. Antònia Martí and Constantin Orasan

This also affects the context - Errors in extraction based summaries
Thomas Kaspersson, Christian Smith, Henrik Danielsson and Arne Jönsson

Annotation of anaphoric relations and topic continuity in Japanese conversation
Natsuko Nakagawa and Yasuharu Den

Domain-specific vs. Uniform Modeling for Coreference Resolution
Olga Uryupina and Massimo Poesio

Creating a Coreference Resolution System for Polish
Mateusz Kopeć and Maciej Ogrodniczuk

P2 - Tools, Systems and Evaluation

Chairperson: Michael Kipp

Fast Labeling and Transcription with the Speechalyzer Toolkit
Felix Burkhardt

Automatic annotation of head velocity and acceleration in Anvil
Bart Jongejan

AVATech – automated annotation through audio and video analysis
Przemysław Lenkiewicz, Binyam Gebrekidan Gebre, Oliver Schreer, Stefano Masneri, Daniel Schneider and Sebastian Tschöpel

An Oral History Annotation Tool for INTER-VIEWS
Henk van den Heuvel, Eric Sanders, Robin Rutten, Stef Scagliola and Paula Witkamp

ELAN development, keeping pace with communities' needs
Han Sloetjes and Aarthy Somasundaram

Inforex -- a web-based tool for text corpus management and semantic annotation
Michał Marcińczuk, Jan Kocoń and Bartosz Broda

Towards Automatic Gesture Stroke Detection
Binyam Gebrekidan Gebre, Peter Wittenburg and Przemysław Lenkiewicz

EXMARaLDA and the FOLK tools – two toolsets for transcribing and annotating spoken language
Thomas Schmidt

Designing a search interface for a Spanish learner spoken corpus: the end-user's evaluation
Leonardo Campillos Llanos

P3 - Lexical Resources

Chairperson: Adam Kilgarriff

Dictionary Look-up with Katakana Variant Recognition
Satoshi Sato

The Rocky Road towards a Swedish FrameNet - Creating SweFN
Karin Friberg Heppin and Maria Toporowska Gronostaj

Capturing syntactico-semantic regularities among terms: An application of the FrameNet methodology to terminology
Marie-Claude L'Homme and Janine Pimentel

Developing LMF-XML Bilingual Dictionaries for Colloquial Arabic Dialects
David Graff and Mohamed Maamouri

UBY-LMF -- A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF
Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek and Christian M. Meyer

Legal electronic dictionary for Czech
František Cvrček, Karel Pala and Pavel Rychlý

Adaptive Dictionary for Bilingual Lexicon Extraction from Comparable Corpora
Amir Hazem and Emmanuel Morin

A New Twitter Verb Lexicon for Natural Language Processing
Jennifer Williams and Graham Katz

P4 - Annotation and Corpora

Chairperson: Andreas Witt

Challenges in the development of annotated corpora of computer-mediated communication in Indian Languages: A Case of Hindi
Ritesh Kumar

Ontologies of Linguistic Annotation: Survey and perspectives
Christian Chiarcos

A High-Quality Web Corpus of Czech
Johanka Spoustová and Miroslav Spousta

WebAnnotator, an Annotation Tool for Web Pages
Xavier Tannier

Development of a Web-Scale Chinese Word N-gram Corpus with Parts of Speech Information
Chi-Hsin Yu, Yi-jie Tang and Hsin-Hsi Chen

CoALT: A Software for Comparing Automatic Labelling Tools
Dominique Fohr and Odile Mella

CAT: the CELCT Annotation Tool
Valentina Bartalesi Lenzi, Giovanni Moretti and Rachele Sprugnoli

ROMBAC: The Romanian Balanced Annotated Corpus
Radu Ion, Elena Irimia, Dan Ștefănescu and Dan Tufis

A French Fairy Tale Corpus syntactically and semantically annotated
Ismail El Maarouf and Jeanne Villaneau

Iula2Standoff: a tool for creating standoff documents for the IULACT
Carlos Morell, Jorge Vivaldi and Núria Bel

ANALEC: a New Tool for the Dynamic Annotation of Textual Data
Frederic Landragin, Thierry Poibeau and Bernard Victorri

The SYNC3 Collaborative Annotation Tool
Georgios Petasis

Simplified guidelines for the creation of Large Scale Dialectal Arabic Annotations
Heba Elfardy and Mona Diab

13:15-14:45 Lunch break

O5 - Crowdsourcing (Special Session)

Chairperson: Karen Fort and Iryna Gurevych

Anadolu Auditorium

- 14:45-15:05 Leveraging the Wisdom of the Crowds for the Acquisition of Multilingual Language Resources
Arno Scharl, Marta Sabou, Stefan Gindl, Walter Rafelsberger and Albert Weichselbraun
- 15:05-15:25 Experiences in Resource Generation for Machine Translation through Crowdsourcing
Anoop Kunchukuttan, Shourya Roy, Pratik Patel, Kushal Ladha, Somya Gupta, Mitesh M. Khapra and Pushpak Bhattacharyya
- 15:25-15:45 Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing
Elena Filatova
- 15:45-16:05 Supervised Topical Key Phrase Extraction of News Stories using Crowdsourcing, Light Filtering and Co-reference Normalization
Luis Marujo, Anatole Gershman, Jaime Carbonell, Robert Frederking and João P. Neto
- 16:05-16:25 Discussion

O6 - Dialogue and Multimodality

Chairperson: Jimmy Kunzmann

Topkapı A

- 14:45-15:05 Constructive Interaction for Talking about Interesting Topics
Kristiina Jokinen and Graham Wilcock
- 15:05-15:25 Using multimodal resources for explanation approaches in intelligent systems
Florian Nothdurft and Wolfgang Minker
- 15:25-15:45 Multimodal Corpus of Multi-party Conversations in Second Language
Shota Yamasaki, Hirohisa Furukawa, Masafumi Nishida, Kristiina Jokinen and Seiichi Yamamoto

- 15:45-16:05 The REX corpora: A collection of multimodal corpora of referring expressions in collaborative problem solving dialogues
Takenobu Tokunaga, Ryu Iida, Asuka Terai and Naoko Kuriyama
- 16:05-16:25 ISO 24617-2: A semantically-based standard for dialogue annotation
Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis and David Traum

O7 - Machine Translation and Language Resources (1)

Chairperson: Gregor Thurmair

Topkapı B

- 14:45-15:05 Collecting and Using Comparable Corpora for Statistical Machine Translation
Inguna Skadiņa, Ahmet Aker, Nikos Mastropavlos, Fangzhong Su, Dan Tufis, Mateja Verlic, Andrejs Vasiljevs, Bogdan Babych, Paul Clough, Robert Gaizauskas, Nikos Glaros, Monica Lestari Paramita and Mārcis Pinnis
- 15:05-15:25 Suffix Trees as Language Models
Casey Redd Kennington, Martin Kay and Annemarie Friedrich
- 15:25-15:45 DGT-TM: A freely available Translation Memory in 22 languages
Ralf Steinberger, Andreas Eisele, Szymon Kłoczek, Spyridon Pilos and Patrick Schlüter
- 15:45-16:05 Identifying Word Translations from Comparable Documents Without a Seed Lexicon
Reinhard Rapp, Serge Sharoff and Bogdan Babych
- 16:05-16:25 Large aligned treebanks for syntax-based machine translation
Gideon Kotzé, Vincent Vandeghinste, Scott Martens and Jörg Tiedemann

O8 - Corpus Processing and Infrastructure

Chairperson: Bente Maegaard

Dolmabahçe A

- 14:45-15:05 Korp – the corpus infrastructure of Språkbanken
Lars Borin, Markus Forsberg and Johan Roxendal
- 15:05-15:25 Annotation Trees: LDC's customizable, extensible, scalable, annotation infrastructure
Jonathan Wright, Kira Griffitt, Joe Ellis, Stephanie Strassel and Brendan Callahan
- 15:25-15:45 Building Large Corpora from the Web Using a New Efficient Tool Chain
Roland Schäfer and Felix Bildhauer
- 15:45-16:05 Annotated Bibliographical Reference Corpora in Digital Humanities
Young-Min Kim, Patrice Bellot, Elodie Faath and Marin Dacos
- 16:05-16:25 Building a 70 billion word corpus of English from ClueWeb
Jan Pomikálek, Miloš Jakubiček and Pavel Rychlý

14:45-16:25

Poster Sessions - Poster Area (2)

P5 - Information Extraction (1)

Chairperson: Günter Neumann

A Gold Standard for Relation Extraction in the Food Domain

Michael Wiegand, Benjamin Roth, Eva Lasarczyk, Stephanie Köser and Dietrich Klakow

Textual Characteristics for Language Engineering
Mathias Bank, Robert Remus and Martin Schierle

Automatically Extracting Procedural Knowledge from Instructional Texts using Natural Language Processing

Ziqi Zhang, Philip Webster, Victoria Uren, Andrea Varga and Fabio Ciravegna

Evolution of Event Designation in Media: Preliminary Study

Xavier Tannier, Véronique Moriceau, Béatrice Arnulphy and Ruixin He

CLTC: A Chinese-English Cross-lingual Topic Corpus

Yunqing Xia, Guoyu Tang, Peng Jin and Xia Yang

A Resource-light Approach to Phrase Extraction for English and German Documents from the Patent Domain and User Generated Content

Julia Maria Schulz, Daniela Becks, Christa Womser-Hacker and Thomas Mandl

An Evaluation of the Effect of Automatic Preprocessing on Syntactic Parsing for Biomedical Relation Extraction

Md. Faisal Mahbub Chowdhury and Alberto Lavelli

Evaluation of Unsupervised Information Extraction

Wei Wang, Romaric Besançon, Olivier Ferret and Brigitte Grau

Extraction of unmarked quotations in Newspapers

Stéphanie Weiser and Patrick Watrin

NgramQuery - Smart Information Extraction from Google N-gram using External Resources

Martin Aleksandrov and Carlo Strapparava

P6 - Word Sense Disambiguation and Evaluation

Chairperson: Sanni Nimb

A voting scheme to detect semantic underspecification

Héctor Martínez Alonso, Núria Bel and Bolette Sandford Pedersen

A Comparative Evaluation of Word Sense Disambiguation Algorithms for German

Verena Henrich and Erhard Hinrichs

DutchSemCor: Targeting the ideal sense-tagged corpus

Piek Vossen, Attila Görög, Rubén Izquierdo and Antal Van den Bosch

Mapping WordNet synsets to Wikipedia articles

Samuel Fernando and Mark Stevenson

A new semantically annotated corpus with syntactic-semantic and cross-lingual senses

Myriam Rakho, Éric Laporte and Matthieu Constant

Detection of Peculiar Word Sense by Distance Metric Learning with Labeled Examples

Minoru Sasaki and Hiroyuki Shinnou

Using semi-experts to derive judgments on word sense alignment: a pilot study

Soojeong Eom, Markus Dickinson and Graham Katz

ATLIS: Identifying Locational Information in Text Automatically

John Vogel, Marc Verhagen and James Pustejovsky

P7 - Multiword Expressions and Term Extraction

Chairperson: Karel Pala

Semi-Supervised Technical Term Tagging With Minimal User Feedback
Behrang QasemiZadeh, Paul Buitelaar, Tianqi Chen and Georgeta Bordea

Linguistic knowledge for specialized text production
Miriam Buendía-Castro and Beatriz Sánchez-Cárdenas

In the same boat and other idiomatic seafaring expressions
Rita Marinelli and Laura Cignoni

Association Norms of German Noun Compounds
Sabine Schulte im Walde, Susanne Borgwaldt and Ronny Jauch

Medical Term Extraction in an Arabic Medical Corpus
Doaa Samy, Antonio Moreno-Sandoval, Conchi Bueno-Díaz, Marta Garrote-Salazar and José M. Guirao

Evaluating the Impact of External Lexical Resources into a CRF-based Multiword Segmenter and Part-of-Speech Tagger
Mathieu Constant and Isabelle Tellier

Adapting and evaluating a generic term extraction tool
Anita Gojun, Ulrich Heid, Bernd Weißbach, Carola Loth and Insa Mingers

Evaluation of Classification Algorithms and Features for Collocation Extraction in Croatian
Mladen Karan, Jan Šnajder and Bojana Dalbelo Bašić

The Quaero Evaluation Initiative on Term Extraction
Thibault Mondary, Adeline Nazarenko, Haïfa Zargayouna and Sabine Barreaux

Using Noun Similarity to Adapt an Acceptability Measure for Persian Light Verb Constructions
Shiva Taslimipoor, Afsaneh Fazly and Ali Hamzeh

Identifying bilingual Multi-Word Expressions for Statistical Machine Translation
Dhouha Bouamor, Nasredine Semmar and Pierre Zweigenbaum

Detecting Japanese Compound Functional Expressions using Canonical/Derivational Relation
Takafumi Suzuki, Yusuke Abe, Itsuki Toyota, Takehito Utsuro, Suguru Matsuyoshi and Masatoshi Tsuchiya

Building a database of French frozen adverbial phrases
Aude Grezka and Céline Poudat

German Verb Patterns and Their Implementation in an Electronic Dictionary
Marc Luder

P8 - Authoring Tools, Proofing

Chairperson: Catia Cucchiari

Risk Analysis and Prevention: LELIE, a Tool dedicated to Procedure and Requirement Authoring

Flore Barcellini, Camille Albert, Corinne Grosse and Patrick Saint-Dizier

A Framework for Spelling Correction in Persian Language Using Noisy Channel Model
Mohammad Hoseyn Sheykholeslam, Behrouz Minaei-Bidgoli and Hossein Juzi

Conventional Orthography for Dialectal Arabic

Nizar Habash, Mona Diab and Owen Rambow

Arabic Word Generation and Modelling for Spell Checking

Khaled Shaalan, Mohammed Attia, Pavel Pecina, Younes Samih and Josef van Genabith

Similarity Ranking as Attribute for Machine Learning Approach to Authorship Identification

Jan Rygl and Aleš Horák

Spell Checking for Chinese

Shaohua Yang, Hai Zhao, Xiaolin Wang and Bao-liang Lu

Spell Checking in Spanish: The Case of Diacritic Accents

Jordi Atserias, Maria Fuentes, Rogelio Nazar and Irene Renau

Incorporating an Error Corpus into a Spellchecker for Maltese

Michael Rosner, Albert Gatt, Andrew Attard and Jan Joachimsen

16:25-16:45 **Coffee break**

O9 - Endangered Languages

Chairperson: Dafydd Gibbon

Anadolu Auditorium

16:45-17:05 A Rule-based Morphological Analyzer for Murrinh-Patha

Melanie Seiss

17:05-17:25 Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages

Dirk Goldhahn, Thomas Eckart and Uwe Quasthoff

17:25-17:45 “Rendering Endangered Lexicons Interoperable through Standards Harmonization”¹ : the RELISH project

Helen Aristar-Dry, Sebastian Drude, Menzo Windhouwer, Jost Gippert and Irina Nevskaya

17:45-18:05 Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms

Ryan Georgi, Fei Xia and William Lewis

O10 - Document Classification, Text Categorisation

Chairperson: Luca Dini

Topkapı A

16:45-17:05 Measuring Interlanguage: Native Language Identification with L1-influence Metrics

Julian Brooke and Graeme Hirst

17:05-17:25 Distractorless Authorship Verification

John Noecker Jr and Michael Ryan

17:25-17:45 Correlation between Similarity Measures for Inter-Language Linked Wikipedia Articles

Monica Lestari Paramita, Paul Clough, Ahmet Aker and Robert Gaizauskas

17:45-18:05 JRC Eurovoc Indexer JEX - A freely available multi-label categorisation tool

Ralf Steinberger, Mohamed Ebrahim and Marco Turchi

O11 - Discourse (1)

Chairperson: Haïfa Zargayouna

Topkapı B

- 16:45-17:05 Annotations for Power Relations on Email Threads
Vinodkumar Prabhakaran, Huzaifa Neralwala, Owen Rambow and Mona Diab
- 17:05-17:25 A Corpus for Research on Deliberation and Debate
Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott and Joseph King
- 17:25-17:45 Annotating Agreement and Disagreement in Threaded Discussion
Jacob Andreas, Sara Rosenthal and Kathleen McKeown
- 17:45-18:05 Evaluation of Discourse Relation Annotation in the Hindi Discourse Relation Bank
Sudheer Kolachina, Rashmi Prasad, Dipti Misra Sharma and Aravind Joshi

O12 - Word Sense Disambiguation

Chairperson: Kiril Simov

Dolmabahçe A

- 16:45-17:05 Using Verb Subcategorization for Word Sense Disambiguation
Will Roberts and Valia Kordoni
- 17:05-17:25 Applying cross-lingual WSD to wordnet development
Marianna Apidianaki and Benoit Sagot
- 17:25-17:45 Discovering Missing Wikipedia Inter-language Links by means of Cross-lingual Word Sense Disambiguation
Els Lefever, Veronique Hoste and Martine De Cock
- 17:45-18:05 Unsupervised Word Sense Disambiguation with Multilingual Representations
Erwin Fernandez-Ordonez, Rada Mihalcea and Samer Hassan

16:45-18:05

Poster Sessions - Poster Area (1)

P9 - Morphology

Chairperson: Pushpak Bhattacharyya

First Steps towards the Semi-automatic Development of a Wordformation-based Lexicon of Latin

Marco Passarotti and Francesco Mambrini

PoliMorf: a (not so) new open morphological dictionary for Polish

Marcin Woliński, Marcin Milkowski, Maciej Ogrodniczuk and Adam Przepiórkowski

Unsupervised acquisition of concatenative morphology

Lionel Nicolas, Jacques Farré and Cécile Darne

Annotating and Learning Morphological Segmentation of Egyptian Colloquial Arabic

Emad Mohamed, Behrang Mohit and Kemal Oflazer

First Results in a Study Evaluating Pre-annotation and Correction Propagation for Machine-Assisted Syriac Morphological Analysis

Paul Felt, Eric Ringger, Kevin Seppi, Kristian Heal, Robbie Haertel and Deryle Lonsdale

Evaluating Hebbian Self-Organizing Memories for Lexical Representation and Access
Claudia Marzi, Marcello Ferro, Claudia Caudai and Vito Pirrelli

A Morphological Analyzer For Wolof Using Finite-State Techniques
Cheikh M. Bamba Dione

IDENTIC Corpus: Morphologically Enriched Indonesian-English Parallel Corpus
Septina Dian Larasati

The Romanian Neuter Examined Through A Two-Gender N-Gram Classification System
Liviu P. Dimu, Vlad Nicolae and Octavia-Maria Şulea

UniDic for Early Middle Japanese: a Dictionary for Morphological Analysis of Classical Japanese
Toshinobu Ogiso, Mamoru Komachi, Yasuharu Den and Yuji Matsumoto

Recognition of Polish Derivational Relations Based on Supervised Learning Scheme
Maciej Piasecki, Radosław Ramocki and Marek Maziarz

Reconstructing the Diachronic Morphology of Romanian from Dictionary Citations
Dan Cristea, Radu Simionescu and Gabriela Haja

Generation of Verbal Stems in Derivationally Rich Language
Krešimir Šojat, Nives Mikelić Preradović and Marko Tadić

A finite-state morphological transducer for Kyrgyz
Jonathan Washington, Mirlan Ipasov and Francis Tyers

AnIta: a powerful morphological analyser for Italian
Fabio Tamburini and Matias Melandri

P10 - Prosody and Phonetics

Chairperson: Laurence Devillers

A topologic view of Topic and Focus marking in Italian
Gloria Gagliardi, Edoardo Lombardi Vallauri and Fabio Tamburini

Comparison between two models of language for the automatic phonetic labeling of an undocumented language of the South-Asia: the case of Mo Piu
Geneviève Caelen-Haumont and Sethserey Sam

MISTRAL+: A Melody Intonation Speaker Tonal Range semi-automatic Analysis using variable Levels
Benoît Weber, Geneviève Caelen-Haumont, Binh Hai Pham and Do-Dat Tran

Comparing performance of different set-covering strategies for linguistic content optimization in speech corpora
Nelly Barbot, Olivier Boeffard and Arnaud Delhay

Towards Fully Automatic Annotation of Audio Books for TTS
Olivier Boeffard, Laure Charonnat, Sébastien Le Maguer and Damien Lolive

Statistical Evaluation of Pronunciation Encoding
Iris Merkus and Florian Schiel

Annotating a corpus of human interaction with prosodic profiles - focusing on Mandarin repair/disfluency
Helen Kaiyun Chen

Prediction of Non-Linguistic Information of Spontaneous Speech from the Prosodic Annotation: Evaluation of the X-JToBI system
Kikuo Maekawa

Prosomarker: a prosodic analysis tool based on optimal pitch stylization and automatic syllabification

Antonio Origlia and Iolanda Alfano

Designing French Tale Corpora for Entertaining Text To Speech Synthesis

David Doukhan, Sophie Rosset, Albert Rilliard, Christophe d'Alessandro and Martine Adda-Decker

Open-Source Boundary-Annotated Corpus for Arabic Speech and Language Processing

Claire Brierley, Majdi Sawalha and Eric Arwell

A Phonemic Corpus of Polish Child-Directed Speech

Luc Boruta and Justyna Jastrzebska

P11 - Language Resource Infrastructures (1)

Chairperson: Daan Broeder

Smooth Sailing for STEVIN

Peter Spyns and Elisabeth D'Halleweyn

Semantic metadata mapping in practice: the Virtual Language Observatory

Dieter van Uytvanck, Herman Stehouwer and Lari Lampen

Aspects of a Legal Framework for Language Resource Management

Aditi Sharma Grover, Annamart Nieman, Gerhard Van Huyssteen and Justus Roux

Introducing the Swedish Kelly-list, a new lexical e-resource for Swedish

Elena Volodina and Softe Johansson Kokkinakis

Texto4Science: a Quebec French Database of Annotated Short Text Messages

Philippe Langlais, Patrick Drouin, Amélie Paulus, Eugénie Rompré Brodeur and Florent Cottin

Recent Developments in CLARIN-NL

Jan Odijk

A Metadata Editor to Support the Description of Linguistic Resources

Emanuel Dima, Christina Hoppermann, Erhard Hinrichs, Thorsten Trippel and Claus Zinn

Fivehundredmillionandone Tokens. Loading the AAC Container with Text Resources for Text Studies.

Hanno Biber and Evelyn Breiteneder

The Common Orthographic Vocabulary of the Portuguese Language: a set of open lexical resources for a pluricentric language

José Pedro Ferreira, Maarten Janssen, Gladis Barcellos de Oliveira, Margarita Correia and Gilvan Müller de Oliveira

Creation of an Open Shared Language Resource Repository in the Nordic and Baltic Countries

Andrejs Vasiljevs, Markus Forsberg, Tatiana Gornostay, Dorte Haltrup Hansen, Kristín Jóhannsdóttir, Gunn Lyse, Krister Lindén, Lene Offersgaard, Sussi Olsen, Bolette Pedersen, Eiríkur Rögnvaldsson, Inguna Skadiņa, Koenraad De Smedt, Ville Oksanen and Roberts Rozis

The LRE Map. Harmonising Community Descriptions of Resources

Nicoletta Calzolari, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo and Claudia Soria

The META-SHARE Metadata Schema for the Description of Language Resources

Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papageorgiou, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz and Valérie Mapelli

Towards automation in using multi-modal language resources: compatibility and interoperability for multi-modal features in Kachako

Yoshinobu Kano

5-minute break

O13 - Multimodal Corpora (1)

Chairperson: Christopher Cieri

Anadolu Auditorium

- 18:10-18:30 The REPERE Corpus : a multimodal corpus for person recognition
Aude Giraudel, Matthieu Carré, Valérie Mapelli, Juliette Kahn, Olivier Galibert and Ludovic Quintard
- 18:30-18:50 Polish Multimodal Corpus - a collection of referential gestures
Magdalena Lis
- 18:50-19:10 An audiovisual political speech analysis incorporating eye-tracking and perception data
Stefan Scherer, Georg Layher, John Kane, Heiko Neumann and Nick Campbell

O14 - Machine Translation and Evaluation (1)

Chairperson: Robert Frederking

Topkapı A

- 18:10-18:30 Eye Tracking as a Tool for Machine Translation Error Analysis
Sara Stymne, Henrik Danielsson, Sofia Bremin, Hongzhan Hu, Johanna Karlsson, Anna Prytz Lillkull and Martin Wester
- 18:30-18:50 Involving Language Professionals in the Evaluation of Machine Translation
Eleftherios Avramidis, Aljoscha Burchardt, Christian Federmann, Maja Popović, Cindy Tschewinka and David Vilar
- 18:50-19:10 An Analysis (and an Annotated Corpus) of User Responses to Machine Translation Output
Daniele Pighin, Lluís Màrquez and Jonathan May

O15 - Information Extraction and Question Answering

Chairperson: Allan Hanbury

Topkapı B

- 18:10-18:30 Challenges in the Knowledge Base Population Slot Filling Task
Bonan Min and Ralph Grishman
- 18:30-18:50 Evaluating Machine Reading Systems through Comprehension Tests
Anselmo Peñas, Eduard Hovy, Pamela Forner, Álvaro Rodrigo, Richard Sutcliffe, Corina Forascu and Caroline Sporleder
- 18:50-19:10 Biomedical Chinese-English CLIR Using an Extended CMeSH Resource to Expand Queries
Xinkai Wang, Paul Thompson, Jun'ichi Tsujii and Sophia Ananiadou

O16 - Web Services

Chairperson: Marc Kemps-Snijders

Dolmabahçe A

- 18:10-18:30 Towards a User-Friendly Platform for Building Language Resources based on Web Services
Marc Poch, Antonio Toral, Olivier Hamon, Valeria Quochi and N ria Bel
- 18:30-18:50 Web Service integration platform for Polish linguistic resources
Maciej Ogrodniczuk and Michal Lenart
- 18:50-19:10 Classifying Standard Linguistic Processing Functionalities based on Fundamental Data Operation Types
Yoshihiko Hayashi and Chiharu Narawa

18:10-19:30

Poster Sessions - Poster Area (2)

P12 - Subjectivity: Sentiments, Emotions, Opinions (1)

Chairperson: Carlo Strapparava

A Multilingual Natural Stress Emotion Database

Xin Zuo, Tian Li and Pascale Fung

Method for Collection of Acted Speech Using Various Situation Scripts

Takahiro Miyajima, Hideaki Kikuchi, Katsuhiko Shirai and Shigeki Okawa

Annotating Opinions in German Political News

Hong Li, Xiwen Cheng, Kristina Adson, Tal Kirshboim and Feiyu Xu

Hindi Subjective Lexicon: A Lexical Resource for Hindi Adjective Polarity Classification

Akshat Bakliwal, Piyush Arora and Vasudeva Varma

The I3MEDIA speech database: a trilingual annotated corpus for the analysis and synthesis of emotional speech

Juan Mar a Garrido, Yesika Laplaza, Montse Marquina, Andrea Pearman, Jos  Gregorio Escalada, Miguel  ngel Rodr guez and Ana Armenta

A hierarchical approach with feature selection for emotion recognition from speech

Panagiotis Giannoulis and Gerasimos Potamianos

Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text

Alexandra Balahur and Jes s M. Hermida

Fine-grained German Sentiment Analysis on Social Media

Saeedeh Montazi

“You Seem Aggressive!” Monitoring Anger in a Practical Application

Felix Burkhardt

Mining Sentiment Words from Microblogs for Predicting Writer-Reader Emotion Transition

Yi-jie Tang and Hsin-Hsi Chen

Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank

Christian Scheible and Hinrich Sch tze

P13 - Named Entity Recognition

Chairperson: Antonio Branco

Learning Categories and their Instances by Contextual Features

Antje Schlaf and Robert Remus

Rembrandt - a named-entity recognition framework

Nuno Cardoso

An Adaptive Framework for Named Entity Combination

Bogdan Sacaleanu and Günter Neumann

Rule-based Entity Recognition and Coverage of SNOMED CT in Swedish Clinical Text

Maria Skeppstedt, Maria Kvist and Hercules Dalianis

Latvian and Lithuanian Named Entity Recognition with TildeNER

Mārcis Pinnis

Tree-Structured Named Entity Recognition on OCR Data: Analysis, Processing and Results

Marco Dinarelli and Sophie Rosset

Aleda, a free large-scale entity database for French

Benoît Sagot and Rosa Stern

Evaluating the Impact of Phrase Recognition on Concept Tagging

Pablo Mendes, Joachim Daiber, Rohana Rajapakse, Felix Sasaki and Christian Bizer

P14 - Dialogue

Chairperson: Ron Artstein

Adaptive Speech Understanding for Intuitive Model-based Spoken Dialogues

Tobias Heinroth, Maximilian Grotz, Florian Nothdurft and Wolfgang Minker

Relating Dominance of Dialogue Participants with their Verbal Intelligence Scores

Kseniya Zablotskaya, Umair Rahim, Fernando Fernández Martínez and Wolfgang Minker

The coding and annotation of multimodal dialogue acts

Volha Petukhova and Harry Bunt

Using DiAML and ANVIL for multimodal dialogue annotations

Harry Bunt, Michael Kipp and Volha Petukhova

A Scalable Architecture For Web Deployment of Spoken Dialogue Systems

Matthew Fuchs, Nikos Tsourakis and Manny Rayner

A Corpus for a Gesture-Controlled Mobile Spoken Dialogue System

Nikos Tsourakis and Manny Rayner

A Corpus of Spontaneous Multi-party Conversation in Bosnian Serbo-Croatian and British English

Emina Kurtic, Bill Wells, Guy J. Brown, Timothy Kempton and Ahmet Aker

The Herme Database of Spontaneous Multimodal Human-Robot Dialogues

Jing Guang Han, Emer Gilmartin, Celine DeLooze, Brian Vaughan and Nick Campbell

Annotation of response tokens and their triggering expressions in Japanese multi-party conversations

Yasuharu Den, Hanae Koiso, Katsuya Takanashi and Nao Yoshida

Syntactic annotation of spontaneous speech: application to call-center conversation data

Thierry Bazillon, Melanie Deplano, Frederic Bechet, Alexis Nasr and Benoit Favre

DECODA: a call-centre human-human spoken conversation corpus
Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori and Eric Arbillot

Resource Evaluation for Usable Speech Interfaces: Utilizing Human-Human Dialogue
Pepi Stavropoulou, Dimitris Spiliotopoulos and Georgios Kouroupetroglou

3rd party observer gaze as a continuous measure of dialogue flow
Jens Edlund, Simon Alexandersson, Jonas Beskow, Lisa Gustavsson, Mattias Heldner, Anna Hjalmarsson, Petter Kallionen and Ellen Marklund

Pursing power in Arabic on-line discussion forums
Marc Tomlinson, David Bracewell, Mary Draper, Zewar Almissour, Ying Shi and Jeremy Bensley

Causal analysis of task completion errors in spoken music retrieval interactions
Sunao Hara, Norihide Kitaoka and Kazuya Takeda

An Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style
Marilyn Walker, Grace Lin and Jennifer Sawyer

24 May 2012

9:00-9:40 **Keynote Speech**
**Sören Auer: The Web of Data: Decentralized, collaborative, interlinked
and interoperable**

Chairperson: Jan Odijk
Anadolu Auditorium

9:40-9:45 5-minute break

O17 - Infrastructures and Strategies for LRs (2)

Chairperson: Andrejs Vasiljevs
Anadolu Auditorium

9:45-10:05 The FLReNet Strategic Language Resource Agenda
*Claudia Soria, Núria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan
Odijk, Stelios Piperidis, Valeria Quochi and Nicoletta Calzolari*

10:05-10:25 Standardizing a Component Metadata Infrastructure
*Daan Broeder, Dieter van Uytvanck, Maria Gavrilidou, Thorsten Trippel and Menzo
Windhouwer*

10:25-10:45 Citing on-line Language Resources
Daan Broeder, Dieter van Uytvanck and Gunter Senft

10:45-11:05 An Analytical Model of Language Resource Sustainability
Khalid Choukri and Victoria Arranz

11:05-11:25 On Using Linked Data for Language Resource Sharing in the Long Tail of the
Localisation Market
*David Lewis, Alexander O'Connor, Andrzej Zydrón, Gerd Sjögren and Rahzeb
Choudhury*

O18 - Dialogue

Chairperson: Linne Ha
Topkapı A

9:45-10:05 Evaluation of Online Dialogue Policy Learning Techniques
Alexandros Papangelis, Vangelis Karkaletsis and Fillia Makedon

10:05-10:25 The acquisition and dialog act labeling of the EDECAN-SPORTS corpus
Lluis-F. Hurtado, Fernando Garcia, Emilio Sanchis and Encarna Segarra

10:25-10:45 Developing and evaluating an emergency scenario dialogue corpus
Jolanta Bachan

- 10:45-11:05 Building and Exploiting a Corpus of Dialog Interactions between French Speaking Virtual and Human Agents
Lina M. Rojas-Barahona, Alejandra Lorenzo and Claire Gardent
- 11:05-11:25 Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the PORTMEDIA corpora
Fabrice Lefèvre, Djamel Mostefa, Laurent Besacier, Yannick Estève, Matthieu Quignard, Nathalie Camelin, Benoit Favre, Bassam Jabaian and Lina M. Rojas-Barahona

O19 - Resource Creation and Acquisition

Chairperson: James Pustejovsky

Topkapı B

- 9:45-10:05 Building a Basque-Chinese Dictionary by Using English as Pivot
Xabier Saralegi, Iker Manterola and Iñaki San Vicente
- 10:05-10:25 Automatic lexical semantic classification of nouns
Núria Bel, Lauren Romeo and Muntsa Padró
- 10:25-10:45 Assessing Crowdsourcing Quality through Objective Tasks
Ahmet Aker, Mahmoud El-Haj, M-Dyaa Albakour and Udo Kruschwitz
- 10:45-11:05 Rapid creation of large-scale corpora and frequency dictionaries
Attila Zséder, Gábor Recski, Dániel Varga and András Kornai
- 11:05-11:25 Boosting the Coverage of a Semantic Lexicon by Automatically Extracted Event Nominalizations
Kata Gábor, Marianna Apidianaki, Benoît Sagot and Éric Villemonte de la Clergerie

O20 - Corpus and Annotation

Chairperson: Nancy Ide

Dolmabahçe A

- 9:45-10:05 Analyzing the Impact of Prevalence on the Evaluation of a Manual Annotation Campaign
Karën Fort, Claire François, Olivier Galibert and Maha Ghribi
- 10:05-10:25 Corpus Annotation as a Scientific Task
Donia Scott, Rossano Barone and Rob Koeling
- 10:25-10:45 Document Attrition in Web Corpora: an Exploration
Stephen Wattam, Paul Rayson and Damon Berridge
- 10:45-11:05 A Concise Query Language with Search and Transform Operations for Corpora with Multiple Levels of Annotation
Anil Kumar Singh
- 11:05-11:25 A New Method for Evaluating Automatically Learned Terminological Taxonomies
Paola Velardi, Roberto Navigli, Stefano Faralli and Juana Maria Ruiz-Martinez

9:45-11:25

Poster Sessions - Poster Area (1)

P15 - Semantic Annotation

Chairperson: Aline Villavicencio

Event Nominals: Annotation Guidelines and a Manually Annotated Corpus in French
Béatrice Arnulphy, Xavier Tannier and Anne Vilnat

Building a Corpus of Indefinite Uses Annotated with Fine-grained Semantic Functions
Maria Aloni, Andreas van Cranenburgh, Raquel Fernandez and Marta Sznajder

A PropBank for Portuguese: the CINTIL-PropBank
António Branco, Catarina Carvalheiro, Silvia Pereira, Sara Silveira, João Silva, Sérgio Castro and João Graça

Empty Argument Insertion in the Hindi PropBank
Ashwini Vaidya, Jinho D. Choi, Martha Palmer and Bhuvana Narasimhan

Annotating Qualia Relations in Italian and French Complex Nominals
Pierrette Bouillon, Elisabetta Jezek, Chiara Melloni and Aurélie Picton

Semantic annotation of French corpora: animacy and verb semantic classes
Juliette Thuilier and Laurence Danlos

Yes we can!? Annotating English modal verbs
Josef Ruppenhofer and Ines Rehbein

An Annotation Scheme for Quantifier Scope Disambiguation
Mehdi Manshadi, James Allen and Mary Swift

Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure
Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa

Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English
Kyoko Ohara

ConanDoyle-neg: Annotation of negation cues and their scope in Conan Doyle stories
Roser Morante and Walter Daelemans

P16 - Document Classification, Text Categorisation

Chairperson: Serge Sharoff

The Netlog Corpus. A Resource for the Study of Flemish Dutch Internet Language
Mike Kestemont, Claudia Peersman, Benny De Decker, Guy De Pauw, Kim Luyckx, Roser Morante, Frederik Vaassen, Janneke van de Loo and Walter Daelemans

Investigating Verbal Intelligence Using the TF-IDF Approach
Kseniya Zablotskaya, Fernando Fernández Martínez and Wolfgang Minker

Diachronic Changes in Text Complexity in 20th Century English Language: An NLP Approach
Sanja Štajner and Ruslan Mitkov

DeCour: a corpus of DEceptive statements in Italian COURts
Tommaso Fornaciari and Massimo Poesio

French and German Corpora for Audience-based Text Type Classification
Amalia Todirascu, Sebastian Pado, Jennifer Krisch, Max Kisselew and Ulrich Heid

Irregularity Detection in Categorized Document Corpora
Borut Sluban, Senja Pollak, Roel Coesemans and Nada Lavrac

Rhetorical Move Detection in English Abstracts: Multi-label Sentence Classifiers and their Annotated Corpora
Carmen Dayrell, Arnaldo Candido Jr., Gabriel Lima, Danilo Machado Jr., Ann Copestake, Valéria Feltrim, Stella Tagnin and Sandra Aluisio

Unsupervised document zone identification using probabilistic graphical models
Andrea Varga, Daniel Preotiuc-Pietro and Fabio Ciravegna

Improving K-Nearest Neighbor Efficacy for Farsi Text Classification
Mohammad Hossein Elahimanesh, Behrouz Minaei and Hossein Malekinezhad

P17 - Grammar and Syntax

Chairperson: Eleni Efthimiou

Automatic Annotation and Manual Evaluation of the Diachronic German Corpus TüBa-D/DC

Erhard Hinrichs and Thomas Zastrow

Grammatical Error Annotation for Korean Learners of Spoken English

Hongsuck Seo, Kyusong Lee, Gary Geunbae Lee, Soo-Ok Kweon and Hae-Ri Kim

Robust clause boundary identification for corpus annotation

Heiki-Jaan Kaalep and Kadri Muischnek

A Corpus-based Study of the German Recipient Passive

Patrick Ziering, Sina Zarrieß and Jonas Kuhn

Wordnet Based Lexicon Grammar for Polish

Zygmunt Vetulani

A Galician Syntactic Corpus with Application to Intonation Modeling

Montserrat Arza, José M. García-Miguel, Francisco Campillo and Miguel Cuevas – Alonso

A Search Tool for FrameNet Constructicon

Hiroaki Sato

Annotating Errors in a Hungarian Learner Corpus

Markus Dickinson and Scott Ledbetter

Text Simplification Tools for Spanish

Stefan Bott, Horacio Saggion and Simon Mille

CLIMB grammars: three projects using metagrammar engineering

Antske Fokkens, Tania Avgustinova and Yi Zhang

An implementation of a Latvian resource grammar in Grammatical Framework

Peteris Paikens and Normunds Gruzitis

An Open Source Persian Computational Grammar

Shafqat Mumtaz Virk and Elnaz Abolahrar

Reclassifying subcategorization frames for experimental analysis and stimulus generation

Paula Buttery and Andrew Caines

Annotating progressive aspect constructions in the spoken section of the British National Corpus

Andrew Caines and Paula Buttery

P18 - Digital Libraries

Chairperson: Monica Monachini

BUCEADOR, a multi-language search engine for digital libraries

Jordi Adell, Antonio Bonafonte, Antonio Cardenal, Marta R. Costa-Jussà, José A. R. Fonollosa, Asunción Moreno, Eva Navas and Eduardo R. Banga

A tool for enhanced search of multilingual digital libraries of e-journals

Ranka Stanković, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić

A Graphical Citation Browser for the ACL Anthology
Benjamin Weitz and Ulrich Schäfer

LDC Language Resource Database: Building a Bibliographic Database
Eleftheria Ahtaridis, Christopher Cieri and Denise DiPersio

Matching Cultural Heritage items to Wikipedia
Eneko Agirre, Ander Barrena, Oier Lopez de Lacalle, Aitor Soroa, Samuel Fernando and Mark Stevenson

11:25-11:45 **Coffee break**

O21 - Speech Corpora and Tools

Chairperson: Robrecht Comeyne

Anadolu Auditorium

11:45-12:05 Creating a Data Collection for Evaluating Rich Speech Retrieval
Maria Eskevich, Gareth J.F. Jones, Martha Larson and Roeland Ordelman

12:05-12:25 The Political Speech Corpus of Bulgarian
Petya Osenova and Kiril Simov

12:25-12:45 SPPAS: a tool for the phonetic segmentation of speech
Brigitte Bigi

12:45-13:05 Orthographic Transcription: which enrichment is required for phonetization?
Brigitte Bigi, Pauline Péri and Roxane Bertrand

O22 - Machine Translation and Evaluation (2)

Chairperson: François Yvon

Topkapı A

11:45-12:05 Error profiling for evaluation of machine-translated text: a Polish-English case study
Sandra Weiss and Lars Ahrenberg

12:05-12:25 Two Phase Evaluation for Selecting Machine Translation Services
Chunqi Shi, Donghui Lin, Masahiko Shimada and Toru Ishida

12:25-12:45 Italian and Spanish Null Subjects. A Case Study Evaluation in an MT Perspective.
Lorenza Russo, Sharid Loáiciga and Asheesh Gulati

12:45-13:05 On the practice of error analysis for machine translation evaluation
Sara Stymne and Lars Ahrenberg

O23 - Semantic Resources

Chairperson: Zygmunt Vetulani

Topkapı B

11:45-12:05 Identifying equivalents of specialized verbs in a bilingual comparable corpus of judgments: A frame-based methodology
Janine Pimentel

12:05-12:25 Logical metonymies and qualia structures: an annotated database of logical metonymies for German
Alessandra Zarcone and Stefan Rued

12:25-12:45 Modality in Text: a Proposal for Corpus Annotation
Iris Hendrickx, Amália Mendes and Silvia Mencarelli

12:45-13:05 DBpedia: A Multilingual Cross-domain Knowledge Base
Pablo Mendes, Max Jakob and Christian Bizer

O24 - Trends in Corpora

Chairperson: Victoria Arranz

Dolmabahçe A

- 11:45-12:05 A corpus of general and specific sentences from news
Annie Louis and Ani Nenkova
- 12:05-12:25 Brand Pitt: A Corpus to Explore the Art of Naming
Gozde Ozbal, Carlo Strapparava and Marco Guerini
- 12:25-12:45 The WeSearch Corpus, Treebank, and Treecache - A Comprehensive Sample of User-Generated Content
Jonathon Read, Dan Flickinger, Rebecca Dridan, Stephan Oepen and Lilja Øvrelid
- 12:45-13:05 Collecting humorous expressions from a community-based question-answering-service corpus
Masashi Inoue and Toshiki Akagi

11:45-13:05 **Poster Sessions - Poster Area (2)**

P19 - Treebanks

Chairperson: Menno van Zaanen

Further Developments in Treebank Error Detection Using Derivation Trees
Seth Kulick, Ann Bies and Justin Mott

Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures
Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies and Nianwen Xue

Expanding Arabic Treebank to Speech: Results from Broadcast News
Mohamed Maamouri, Ann Bies and Seth Kulick

Propbank-Br: a Brazilian Treebank annotated with semantic role labels
Magali Sanches Duran and Sandra Maria Aluisio

Joint Grammar and Treebank Development for Mandarin Chinese with HPSG
Yi Zhang, Rui Wang and Yu Chen

A tree is a Baum is an árbol is a sach'a: Creating a trilingual treebank
Annette Rios and Anne Göhring

German and English Treebanks and Lexica for Tree-Adjoining Grammars
Miriam Kaeshammer and Vera Demberg

Prague Dependency Style Treebank for Tamil
Loganathan Ramasamy and Zdeněk Žabokrtský

Treebanking by Sentence and Tree Transformation: Building a Treebank to support Question Answering in Portuguese
Patricia Gonçalves, Rita Santos and António Branco

Croatian Dependency Treebank: Recent Development and Initial Experiments
Dasa Berovic, Zeljko Agic and Marko Tadić

A GUI to Detect and Correct Errors in Hindi Dependency Treebank

Rahul Agarwal, Bharat Ram Ambati and Anil Kumar Singh

From Grammar Rule Extraction to Treebanking: A Bootstrapping Approach

Masood Ghayoomi

The IULA Treebank

Montserrat Marimon, Beatriz Fisas, Núria Bel, Marta Villegas, Jorge Vivaldi, Sergi Torner, Mercè Lorente, Silvia Vázquez and Marta Villegas

Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3

Astro Voutilainen, Kristiina Muhonen, Tanja Purtonen and Krister Lindén

The Parallel-TUT: a multilingual and multiformat treebank

Cristina Bosco, Manuela Sanguinetti and Leonardo Lesmo

Irish Treebanking and Parsing: A Preliminary Evaluation

Teresa Lynn, Ozlem Cetinoglu, Jennifer Foster, Elaine Uí Dhonnchadha, Mark Dras and Josef van Genabith

P20 - Parsing

Chairperson: Antonio Moreno-Sandoval

Automatic Extraction and Evaluation of Arabic LFG Resources

Mohammed Attia, Khaled Shaalan, Lamia Tounsi and Josef van Genabith

Rule-Based Detection of Clausal Coordinate Ellipsis

Kristiina Muhonen and Tanja Purtonen

The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish

Gülşen Eryiğit

Task-Driven Linguistic Analysis based on an Underspecified Features Representation

Stasinos Konstantopoulos, Valia Kordoni, Nicola Cancedda, Vangelis Karkaletsis, Dietrich Klakow and Jean-Michel Renders

“Combining Language Resources Into A Grammar-Driven Swedish Parser”

Malin Ahlberg and Ramona Enache

The Icelandic Parsed Historical Corpus (IcePaHC)

Eiríkur Rögnvaldsson, Anton Karl Ingason, Einar Freyr Sigurðsson and Joel Wallenberg

A treebank-based study on the influence of Italian word order on parsing performance

Anita Alicante, Cristina Bosco, Anna Corazza and Alberto Lavelli

Effort of Genre Variation and Prediction of System Performance

Dong Wang and Fei Xia

P21 - Information Extraction (2)

Chairperson: Paul Buitelaar

Statistical Section Segmentation in Free-Text Clinical Records

Michael Tepper, Daniel Capurro, Fei Xia, Lucy Vanderwende and Meliha Yetisgen-Yildiz

A Corpus of Scientific Biomedical Texts Spanning over 168 Years Annotated for Uncertainty

Ramona Bongelli, Carla Canestrari, Ilaria Riccioni, Andrzej Zuczkowski, Cinzia Buldorini, Ricardo Pietrobon, Alberto Lavelli and Bernardo Magnini

Págico : Evaluating Wikipedia-based information retrieval in Portuguese
Cristina Mota, Alberto Simões, Cláudia Freitas, Luis Costa and Diana Santos

Applying Random Indexing to Structured Data to Find Contextually Similar Words
Danica Damljanovic, Udo Kruschwitz, M-Dyaa Albakour, Johann Petrak and Mihai Lupu

The CONCUSUS Corpus of Event Summaries
Horacio Saggion and Sandra Szasz

Building and Exploring Semantic Equivalences Resources
Gracinda Carvalho, David Martins de Matos and Vitor Rocio

The TARSQI Toolkit
Marc Verhagen and James Pustejovsky

From medical language processing to BioNLP domain
Gabriella Pardelli, Manuela Sassi, Sara Goggi and Stefania Biagioni

Evaluation of a Complex Information Extraction Application in Specific Domain
Romarc Besançon, Olivier Ferret and Ludovic Jean-Louis

A methodology for the extraction of information about the usage of formulaic expressions in scientific texts
Hannah Kermes

Structural alignment of plain text books
André Santos, José João Almeida and Nuno Carvalho

Dependency parsing for interaction detection in pharmacogenomics
Gerold Schneider, Fabio Rinaldi and Simon Clematide

A data and analysis resource for an experiment in text mining a collection of micro-blogs on a political topic.
William Black, Rob Procter, Steven Gray and Sophia Ananiadou

13:05-13:10 5-minute break

13:10-13:30

Invited Talk:

Kemal Oflazer: The Turkish Language and its Challenges for Language Processing

Chairperson: Mehmed Özkan
Anadolu Auditorium

13:30-14:55 Lunch break

O25 - Multimodal Corpora (2)

Chairperson: Nick Campbell
Anadolu Auditorium

14:55-15:15 SUTAV: A Turkish Audio-Visual Database
Ibrahim Saygin Topkaya and Hakan Erdogan

15:15-15:35 Multimodal Behaviour and Feedback in Different Types of Interaction
Costanza Navarretta and Patrizia Paggio

15:35-15:55 A Parallel Corpus of Music and Lyrics Annotated with Emotions
Carlo Strapparava, Rada Mihalcea and Alberto Battocchi

15:55-16:15 Building a Multimodal Laughter Database for Emotion Recognition
Merlin Teodosia Suarez, Jocelynn Cu and Madelene Sta. Maria

16:15-16:35 A Speech and Gesture Spatial Corpus in Assisted Living
Dimitra Anastasiou

O26 - Child Language Corpus

Chairperson: Massimo Poesio

Topkapı A

- 14:55-15:15 The Twins Corpus of Museum Visitor Questions
Priti Aggarwal, Ron Artstein, Jillian Gerten, Athanasios Katsamanis, Shrikanth Narayanan, Angela Nazarian and David Traum
- 15:15-15:35 Korean Children's Spoken English Corpus and an Analysis of its Pronunciation Variability
Hyejin Hong, Sunhee Kim and Minhwa Chung
- 15:35-15:55 Corpus of Children Voices for Mid-level Markers and Affect Bursts Analysis
Marie Tahon, Agnes Delaborde and Laurence Devillers
- 15:55-16:15 A large scale annotated child language construction database
Aline Villavicencio, Beracah Yankama, Marco Idiart and Robert Berwick
- 16:15-16:35 Morphosyntactic Analysis of the CHILDES and TalkBank Corpora
Brian MacWhinney

O27 - MultiWord Expressions

Chairperson: Emanuele Pianta

Topkapı B

- 14:55-15:15 Light Verb Constructions in the SzegedParalellFX English--Hungarian Parallel Corpus
Veronika Vincze
- 15:15-15:35 Measuring the compositionality of NV expressions in Basque by means of distributional similarity techniques
Antton Gurrutxaga and Iñaki Alegria
- 15:35-15:55 Analyzing and Aligning German compound nouns
Marion Weller and Ulrich Heid
- 15:55-16:15 Automatic Term Recognition Needs Multiple Evidence
Natalia Loukachevitch
- 16:15-16:35 Constraint Based Description of Polish Multiword Expressions
Roman Kurc, Maciej Piasecki and Bartosz Broda

O28 - Sign Language

Chairperson: Javier Caminero

Dolmabahçe A

- 14:55-15:15 Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking
Dimitris Metaxas, Bo Liu, Fei Yang, Peng Yang, Nicholas Michael and Carol Neidle
- 15:15-15:35 Comparing computer vision analysis of signed language video with motion capture recordings
Matti Karppa, Tommi Jantunen, Ville Viitaniemi, Jorma Laaksonen, Birgitta Burger and Danny De Weerd

- 15:35-15:55 DEGELS1: A comparable corpus of French Sign Language and co-speech gestures
Annelies Braffort and Leïla Boutora
- 15:55-16:15 Semi-Automatic Sign Language Corpora Annotation using Lexical Representations of Signs
Matilde Gonzalez, Michael Filhol and Christophe Collet
- 16:15-16:35 A platform-independent user-friendly dictionary from Italian to LIS
Umar Shoaib, Nadeem Ahmad, Paolo Prinetto and Gabriele Tiotto

14:55-16:35

Poster Sessions - Poster Area (1)

P22 - Part-of-Speech Tagging

Chairperson: Reinhard Rapp

A Universal Part-of-Speech Tagset

Slav Petrov, Dipanjan Das and Ryan McDonald

Improving corpus annotation productivity: a method and experiment with interactive tagging

Atro Voutilainen

Lemmatizing Serbian as Category Tagging with Bidirectional Sequence Classification

Andrea Gesmundo and Tanja Samardzic

Joint Segmentation and POS Tagging for Arabic Using a CRF-based Classifier

Souhir Gahbiche-Braham, H el ene Bonneau-Maynard, Thomas Lavergne and Fran ois Yvon

Boosting statistical tagger accuracy with simple rule-based grammars

Mans Hulden and Jerid Francom

NeoTag: a POS Tagger for Grammatical Neologism Detection

Maarten Janssen

Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser

Francesco Rubino, Francesca Frontini and Valeria Quochi

P23 - Machine Translation (1)

Chairperson: Philippe Langlais

Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies

Bruno Cartoni and Thomas Meyer

Can Statistical Post-Editing with a Small Parallel Corpus Save a Weak MT Engine?

Marianna J. Martindale

BLEU Evaluation of Machine-Translated English-Croatian Legislation

Sanja Seljan, Marija Brki  and Tomislav Vi i 

Chinese Characters Mapping Table of Japanese, Traditional Chinese and Simplified Chinese

Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi

Free/Open Source Shallow-Transfer Based Machine Translation for Spanish and Aragonese

Juan Pablo Mart nez Cort s, Jim O'Regan and Francis Tyers

Automatic MT Error Analysis: Hjerson Helping Addicter
Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović and Daniel Zeman

Re-ordering Source Sentences for SMT
Amit Sangodkar and Om Damani

An English-Portuguese parallel corpus of questions: translation guidelines and application in SMT
Ángela Costa, Tiago Luís, Joana Ribeiro, Ana Cristina Mendes and Luísa Coheur

Word Alignment for English-Turkish Language Pair
Mehmet Talha Çakmak, Süleyman Acar and Gülşen Eryiğit

PEXACC: A Parallel Sentence Mining Algorithm from Comparable Corpora
Radu Ion

A Richly Annotated, Multilingual Parallel Corpus for Hybrid Machine Translation
Eleftherios Avramidis, Marta R. Costa-Jussà, Christian Federmann, Josef van Genabith, Maitte Melero and Pavel Pecina

Automatic word alignment tools to scale production of manually aligned parallel texts
Stephen Grimes, Katherine Peterson and Xuansong Li

Design and compilation of a specialized Spanish-German parallel corpus
Carla Parra Escartin

A Distributed Resource Repository for Cloud-Based Machine Translation
Jörg Tiedemann, Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen and Matthias Zumpfe

P24 - Corpus Creation, Processing, Usage (1)

Chairperson: Takenobu Tokunaga

Parallel Data, Tools and Interfaces in OPUS
Jörg Tiedemann

The Polish Sejm Corpus
Maciej Ogrodniczuk

From keystrokes to annotated process data: Enriching the output of Inputlog with linguistic information
Lieve Macken, Veronique Hoste, Marielle Leijten and Luuk Van Waes

A Curated Database for Linguistic Research: The Test Case of Cimbrian Varieties
Maristella Agosti, Birgit Alber, Giorgio Maria Di Nunzio, Marco Dussin, Stefan Rabanus and Alessandra Tomaselli

Introducing the Reference Corpus of Contemporary Portuguese Online
Michel Génèreux, Iris Hendrickx and Amália Mendes

A Basic Language Resource Kit for Persian
Mojgan Seraji, Beáta Megyesi and Joakim Nivre

Collecting and Analysing Chats and Tweets in SoNaR
Eric Sanders

The goo300k corpus of historical Slovene
Tomaž Erjavec

Kitten: a tool for normalizing HTML and extracting its textual content
Mathieu-Henri Falco, Véronique Moriceau and Anne Vilnat

Collection of a corpus of Dutch SMS
Maaske Treurniet, Orphée De Clercq, Henk van den Heuvel and Nelleke Oostdijk

RIDIRE-CPI: an Open Source Crawling and Processing Infrastructure for Supervised Web-Corpora Building

Alessandro Panunzi, Marco Fabbri, Massimo Moneglia, Lorenzo Gregori and Samuele Paladini

The Minho Quotation Resource

Brett Drury and José João Almeida

Evaluating Query Languages for a Corpus Processing System

Elena Frick, Carsten Schnober and Piotr Bański

P25 - Evaluation Methodologies

Chairperson: Mathieu Lafourcade

QurSim: A corpus for evaluation of relatedness in short texts

Abdul-Baqee Sharaf and Eric Atwell

EVALIEX - A Proposal for an Extended Evaluation Methodology for Information Extraction Systems

Christina Feilmayr, Birgit Pröll and Elisabeth Linsmayr

A Rough Set Formalization of Quantitative Evaluation with Ambiguity

Patrick Paroubek and Xavier Tannier

The Influence of Corpus Quality on Statistical Measurements on Language Resources

Thomas Eckart, Uwe Quasthoff and Dirk Goldhahn

Identifying Nuggets of Information in GALE Distillation Evaluation

Olga Babko-Malaya, Greg Milette, Michael Schneider and Sarah Scogin

NTUSocialRec: An Evaluation Dataset Constructed from Microblogs for Recommendation Applications in Social Networks

Chieh-Jen Wang, Shuk-Man Cheng, Lung-Hao Lee, Hsin-Hsi Chen, Wen-shen Liu, Pei-Wen Huang and Shih-Peng Lin

16:35-16:55 **Coffee break**

O29 - Language Generation and Paraphrasing

Chairperson: Robert Dale

Anadolu Auditorium

16:55-17:15 A contrastive review of paraphrase acquisition techniques

Houda Bouamor, Aurélien Max, Gabriel Illouz and Anne Vilnat

17:15-17:35 Chinese Whispers: Cooperative Paraphrase Acquisition

Matteo Negri, Yashar Mehdad, Alessandro Marchetti, Danilo Giampiccolo and Luisa Bentivogli

17:35-17:55 Diversifiable Bootstrapping for Acquiring High-Coverage Paraphrase Resource

Hideki Shima and Teruko Mitamura

17:55-18:15 SemScribe: Natural Language Generation for Medical Reports

Sebastian Varges, Heike Bieler, Manfred Stede, Lukas C. Faulstich, Kristin Irsig and Malik Atalla

O30 - Computer Aided Language Learning

Chairperson: Justus Roux

Topkapı A

- 16:55-17:15 Item Development and Scoring for Japanese Oral Proficiency Testing
Hitokazu Matsushita and Deryle Lonsdale
- 17:15-17:35 Evaluating Appropriateness Of System Responses In A Spoken CALL Game
Manny Rayner, Pierrette Bouillon and Johanna Gerlach
- 17:35-17:55 Spontaneous Speech Corpora for language learners of Spanish, Chinese and Japanese
Antonio Moreno-Sandoval, Leonardo Campillos Llanos, Yang Dong, Emi Takamori, José M. Guirao, Paula Gozalo, Chieko Kimura, Kengo Matsui and Marta Garrote-Salazar
- 17:55-18:15 The DISCO ASR-based CALL system: practicing L2 oral skills and beyond
Helmer Strik, Jozef Colpaert, Joost Van Doremalen and Catia Cucchiarini

O31 - Discourse (2)

Chairperson: Aravind Joshi

Topkapı B

- 16:55-17:15 A Tool for Extracting Conversational Implicatures
Marta Tatu and Dan Moldovan
- 17:15-17:35 Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns
Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni and Sandrine Zufferey
- 17:35-17:55 Annotating Story Timelines as Temporal Dependency Structures
Steven Bethard, Oleksandr Kolomyets and Marie-Francine Moens
- 17:55-18:15 An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus
Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cecile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paul Pery-Woodley, Laurent Prevot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret and Laure Vieu

O32 - Syntax and Parsing

Chairperson: Adam Przepiórkowski

Dolmabahçe A

- 16:55-17:15 HamleDT: To Parse or Not to Parse?
Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský and Jan Hajič
- 17:15-17:35 Evaluating and improving syntactic lexica by plugging them within a parser
Elsa Tolone, Benoît Sagot and Éric Villemonte de la Clergerie
- 17:35-17:55 Efficient Dependency Graph Matching with the IMS Open Corpus Workbench
Thomas Proisl and Peter Uhrig
- 17:55-18:15 MaltOptimizer: A System for MaltParser Optimization
Miguel Ballesteros and Joakim Nivre

Poster Sessions - Poster Area (2)
P26 - Multilinguality
Chairperson: Gil Francopoulo

Representing the Translation Relation in a Bilingual Wordnet

Jyrki Niemi and Krister Lindén

Building a multilingual parallel corpus for human users

Alexandr Rosen and Martin Vavřín

HunOr: A Hungarian-Russian Parallel Corpus

Martina Katalin Szabó, Veronika Vincze and István Nagy T.

Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics

Kanika Gupta, Monojit Choudhury and Kalika Bali

Dbnary: Wiktionary as a LMF based Multilingual RDF network

Gilles Sérasset

FreeLing 3.0: Towards Wider Multilinguality

Lluís Padró and Evgeny Stanilovsky

Bulgarian X-language Parallel Corpus

Svetla Koeva, Ivelina Stoyanova, Rositsa Dekova, Borislav Rizov and Angel Genov

Automatically Generated Online Dictionaries

Enikő Héja and Dávid Takács

Feedback in Nordic First-Encounters: a Comparative Study

Costanza Navarretta, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen and Patrizia Paggio

MultiUN v2: UN Documents with Multilingual Alignments

Yu Chen and Andreas Eisele

Customization of the Europarl Corpus for Translation Studies

Zahurul Islam and Alexander Mehler

Accessing and standardizing Wiktionary lexical entries for the translation of labels in Cultural Heritage taxonomies

Thierry Declerck, Karlheinz Mörth and Piroska Lendvai

A Mandarin-English Code-Switching Corpus

Ying Li, Yue Yu and Pascale Fung

A Fast, Memory Efficient, Scalable and Multilingual Dictionary Retriever

Paulo Fernandes, Lucelene Lopes, Carlos A. Prolo, Afonso Sales and Renata Vieira

Multilingual Central Repository version 3.0

Aitor Gonzalez-Agirre, Egoitz Laparra and German Rigau

P27 - Question Answering and Summarisation

Chairperson: Horacio Saggion

A good space: Lexical predictors in word space evaluation

Christian Smith, Henrik Danielsson and Arne Jönsson

Creation and use of Language Resources in a Question-Answering eHealth System

Ulrich Andersen, Anna Braasch, Lina Henriksen, Csaba Huszka, Anders Johannsen, Lars Kayser, Bente Maegaard, Ole Norgaard, Stefan Schulz and Jürgen Wedekind

Effects of Document Clustering in Modeling Wikipedia-style Term Descriptions

Atsushi Fujii, Yuya Fujii and Takenobu Tokunaga

Evaluating Multi-focus Natural Language Queries over Data Services
Silvia Quarteroni, Vincenzo Guerrisi and Pietro La Torre

Summarizing a multimodal set of documents in a Smart Room
Maria Fuentes, Horacio Rodriguez and Jordi Turmo

P28 - Multimodal Corpus for Interaction

Chairperson: Vangelis Karkaletsis

LAST MINUTE: a Multimodal Corpus of Speech-based User-Companion Interactions
Dietmar Rösner, Jörg Frommer, Rafael Friesen, Matthias Haase, Julia Lange and Mirko Otto

Annotating Football Matches: Influence of the Source Medium on Manual Annotation
Karën Fort and Vincent Claveau

Creating HAVIC: Heterogeneous Audio Visual Internet Collection
Stephanie Strassel, Amanda Morris, Jonathan Fiscus, Christopher Caruso, Haejoong Lee, Paul Over, James Fiumara, Barbara Shaw, Brian Antonishek and Martial Michel

MULTIPHONIA: a MULTImodal database of PHONetics teaching methods in classroom InterActions.
Charlotte Alazard, Corine Astésano and Michel Billières

P29 - Ontologies

Chairperson: Paola Velardi

Mapping WordNet to the Kyoto ontology
Egoitz Laparra, German Rigau and Piek Vossen

Constructing a Class-Based Lexical Dictionary using Interactive Topic Models
Kugatsu Sadamitsu, Kuniko Saito, Kenji Imamura and Yoshihiro Matsuo

Adding Morpho-semantic Relations to the Romanian Wordnet
Verginica Barbu Mititelu

An ontological approach to model and query multimodal concurrent linguistic annotations
Julien Seinturier, Elisabeth Muriasco, Emmanuel Bruno and Philippe Blache

The IMAGACT Cross-linguistic Ontology of Action. A new infrastructure for natural language disambiguation
Massimo Moneglia, Monica Monachini, Omar Calabrese, Alessandro Panunzi, Francesca Frontini, Gloria Gagliardi and Irene Russo

Towards a methodology for automatic identification of hypernyms in the definitions of large-scale dictionary
Inga Gheorghita and Jean-Marie Pierrel

Collaborative semantic editing of linked data lexica
John McCrae, Elena Montiel-Ponsoda and Philipp Cimiano

Ontoterminology: How to unify terminology and ontology into a single paradigm
Christophe Roche

Representation of linguistic and domain knowledge for second language learning in virtual worlds
Alexandre Denis, Ingrid Falk, Claire Gardent and Laura Perez-Beltrachini

A Treebank-driven Creation of an OntoValence Verb lexicon for Bulgarian
Petya Osenova, Kiril Simov, Laska Laskova and Stanislava Kancheva

Creation of a bottom-up corpus-based ontology for Italian Linguistics

Elisa Bianchi, Mirko Tamosanis and Emiliano Giovannetti

Visualizing word senses in WordNet Atlas

Matteo Abrate and Clara Bacciu

5-minute break

O33 - Semantics from Corpora

Chairperson: Maria Teresa Pazienza

Anadolu Auditorium

- 18:20-18:40 Concept-based Selectional Preferences and Distributional Representations from Wikipedia Articles
Alex Judea, Vivi Nastase and Michael Strube
- 18:40-19:00 Associative and Semantic Features Extracted From Web-Harvested Corpora
Elias Iosif, Maria Giannoudaki, Eric Fosler-Lussier and Alexandros Potamianos
- 19:00-19:20 Building a Resource of Patterns Using Semantic Types
Octavian Popescu

O34 - Authoring and Related Tools

Chairperson: Bernardo Magnini

Topkapı A

- 18:20-18:40 CLCM - A Linguistic Resource for Effective Simplification of Instructions in the Crisis Management Domain and its Evaluations
Irina Temnikova, Constantin Orasan and Ruslan Mitkov
- 18:40-19:00 A Framework for Evaluating Text Correction
Robert Dale and George Narroway
- 19:00-19:20 Typing Race Games as a Method to Create Spelling Error Corpora
Paul Rodrigues and C. Anton Rytting

O35 - Word Sense Annotation and Disambiguation

Chairperson: Yoshihiko Hayashi

Topkapı B

- 18:20-18:40 The MASC Word Sense Corpus
Rebecca J. Passonneau, Collin F. Baker, Christiane Fellbaum and Nancy Ide
- 18:40-19:00 Addressing polysemy in bilingual lexicon extraction from comparable corpora
Darja Fišer, Nikola Ljubešić and Ozren Kubelka
- 19:00-19:20 Empirical Comparisons of MASC Word Sense Annotations
Gerard de Melo, Collin F. Baker, Nancy Ide, Rebecca J. Passonneau and Christiane Fellbaum

O36 - Time and Space

Chairperson: Piek Vossen

Dolmabahçe A

- 18:20-18:40 TIMEN: An Open Temporal Expression Normalisation Resource
Hector Llorens, Leon Derczynski, Robert Gaizauskas and Estela Saquete
- 18:40-19:00 Annotating Spatial Containment Relations Between Events
Kirk Roberts, Travis Goodwin and Sanda M. Harabagiu
- 19:00-19:20 The Role of Model Testing in Standards Development: The Case of ISO-Space
James Pustejovsky and Jessica Moszkowicz

18:20-19:40

Poster Sessions - Poster Area (1)

P30 - Discourse

Chairperson: David Traum

Investigating Engagement - intercultural and technological aspects of the collection, analysis, and use of the Estonian Multiparty Conversational video data
Kristiina Jokinen and Silvi Tenjes

DISLOG: A logic-based language for processing discourse structures
Patrick Saint-Dizier

A Repository of Rules and Lexical Resources for Discourse Structure Analysis: the Case of Explanation Structures
Sarah Bourse and Patrick Saint-Dizier

Feature Discovery for Diachronic Register Analysis: a Semi-Automatic Approach
Siefania Degaetano-Ortlieb, Ekaterina Lapshinova-Koltunski and Elke Teich

Improving the Recall of a Discourse Parser by Constraint-based Postprocessing
Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli

Annotating dropped pronouns in Chinese newswire text
Elizabeth Baran, Yaqin Yang and Nianwen Xue

Alternative Lexicalizations of Discourse Connectives in Czech
Magdalena Rysova

METU Turkish Discourse Bank Browser
Utku Şirin, Ruket Çakıcı and Deniz Zeyrek

DramaBank: Annotating Agency in Narrative Discourse
David Elson

Multi-Layer Discourse Annotation of a Dutch Text Corpus
Gisela Redeker, Ildikó Berzlánovich, Nynke van der Vliet, Gosse Bouma and Markus Egg

Clause-based Discourse Segmentation of Arabic Texts
Iskandar Keskes, Farah Benamara and Lamia Hadrich Belguith

Project FLY: a multidisciplinary project within Linguistics
Mariana Gomes, Ana Guilherme, Leonor Tavares and Rita Marquilhas

Revealing Contentious Concepts Across Social Groups
Ching-Sheng Lin, Zumrut Akcam, Samira Shaikh, Sharon Small, Ken Stahl, Tomek Strzalkowski and Nick Webb

P31 - Lexical Acquisition

Chairperson: Pierre Zweigenbaum

Customizable SCF Acquisition in Italian

Tommaso Caselli, Francesco Rubino, Francesca Frontini, Irene Russo and Valeria Quochi

Large Scale Lexical Analysis

Gregor Thurmair, Vera Aleksic and Christoph Schwarz

Extending the adverbial coverage of a French morphological lexicon

Elsa Tolone, Stavroula Voyatzi, Claude Martineau and Matthieu Constant

Corpus based Semi-Automatic Extraction of Persian Compound Verbs and their Relations

Somayeh Bagherbeygi and Mehrnoush Shamsfard

P32 - Corpus Creation, Processing, Usage (2)

Chairperson: Shyam Agrawal

Extending the MPC corpus to Chinese and Urdu - A Multiparty Multi-Lingual Chat Corpus for Modeling Social Phenomena in Language

Ting Liu, Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stromer-Galley, Sarah Taylor, Umit Boz, Xiaoi Ren and Jingsi Wu

Multimedia database of the cultural heritage of the Balkans

Ivana Tanasijević, Biljana Sikimić and Gordana Pavlović-Lažetić

YADAC: Yet another Dialectal Arabic Corpus

Rania Al-Sabbagh and Roxana Girju

The Trilingual ALLEGRA Corpus: Presentation and Possible Use for Lexicon Induction

Yves Scherrer and Bruno Cartoni

Beyond SoNaR: towards the facilitation of large corpus building efforts

Martin Reynaert, Ineke Schuurman, Veronique Hoste, Nelleke Oostdijk and Maarten van Gompel

The New IDS Corpus Analysis Platform: Challenges and Prospects

Piotr Bański, Peter M. Fischer, Elena Frick, Erik Ketzan, Marc Kupietz, Carsten Schnober, Oliver Schonefeld and Andreas Witt

A Tool/Database Interface for Multi-Level Analyses

Kurt Eberle, Kerstin Eckart, Ulrich Heid and Boris Haselbach

New language resources for the Pashto language

Djamel Mostefa

CALBC: Releasing the Final Corpora

Şenay Kafkas, Ian Lewin, David Milward, Erik van Mulligen, Jan Kors, Udo Hahn and Dietrich Rebholz-Schuhmann

Language Richness of the Web

Martin Majliš and Zdeněk Žabokrtský

P33 - Web Services

Chairperson: Nuria Bel

Cloud Logic Programming for Integrating Language Technology Resources
Markus Forsberg and Torbjörn Lager

Dynamic web service deployment in a cloud environment
Marc Kemps-Snijders, Matthijs Brouwer, Jan Pieter Kunst and Tom Visser

Word Sketches for Turkish
Bharat Ram Ambati, Siva Reddy and Adam Kilgarriff

Service Composition Scenarios for Task-Oriented Translation
Chunqi Shi, Donghui Lin and Toru Ishida

Linguistic Analysis Processing Line for Bulgarian
Aleksandar Savkov, Laska Laskova, Stanislava Kancheva, Petya Osenova and Kiril Simov

On the Way to a Legal Sharing of Web Applications in NLP
Victoria Arranz and Olivier Hamon

Collaborative Development and Evaluation of Text-processing Workflows in a UIMA-supported Web-based Workbench
Rafal Rak, Andrew Rowley and Sophia Ananiadou

The SERENOA Project: Multidimensional Context-Aware Adaptation of Service Front-Ends
Javier Caminero, Mari Carmen Rodríguez, Jean Vanderdonckt, Fabio Paternò, Joerg Rett, Dave Raggett, Jean-Loup Comeliau and Ignacio Marín

25 May 2012

9:00-9:40 **Keynote Speech**

Catherine Pelachaud: Socio-emotional virtual agent

Chairperson: Asuncion Moreno

Anadolu Auditorium

9:40-9:45 5-minute break

O37 - Subjectivity and Emotions

Chairperson: Julio Gonzalo

Anadolu Auditorium

- 9:45-10:05 Towards Emotion and Affect Detection in the Multimodal LAST MINUTE Corpus
Jörg Frommer, Bernd Michaelis, Dietmar Rösner, Andreas Wendemuth, Rafael Friesen, Matthias Haase, Manuela Kunze, Rico Andrich, Julia Lange, Axel Panning and Ingo Siegert
- 10:05-10:25 Building a fine-grained subjectivity lexicon from a web corpus
Isa Maks and Piek Vossen
- 10:25-10:45 Learning Sentiment Lexicons in Spanish
Veronica Perez-Rosas, Carmen Banea and Rada Mihalcea
- 10:45-11:05 Assigning Connotation Values to Events
Tommaso Caselli, Irene Russo and Francesco Rubino
- 11:05-11:25 Cost and Benefit of Using WordNet Senses for Sentiment Analysis
Balamuraliar, Aditya Joshi and Pushpak Bhattacharyya

O38 - Named Entities

Chairperson: Satoshi Sato

Topkapı A

- 9:45-10:05 Linguistic Resources for Entity Linking Evaluation: from Monolingual to Cross-lingual
Xuansong Li, Stephanie Strassel, Heng Ji, Kira Griffitt and Joe Ellis
- 10:05-10:25 Creating and Curating a Cross-Language Person-Entity Linking Collection
Dawn Lawrie, James Mayfield, Paul McNamee and Douglas Oard
- 10:25-10:45 International Multicultural Name Matching Competition: Design, Execution, Results, and Lessons Learned
Keith J. Miller, Elizabeth Schroeder Richerson, Sarah McLeod, James Finley and Aaron Schein

- 10:45-11:05 An Empirical Study of the Occurrence and Co-Occurrence of Named Entities in Natural Language Corpora
K Saravanan, Monojit Choudhury, Raghavendra Udupa and A Kumaran
- 11:05-11:25 Extended Named Entities Annotation on OCR'd Documents: From Corpus Constitution to Evaluation Campaign
Olivier Galibert, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum and Ludovic Quintard

O39 - Treebanks and Syntax

Chairperson: Erhard Hinrichs

Topkapı B

- 9:45-10:05 Making Ellipses Explicit in Dependency Conversion for a German Treebank
Wolfgang Seeker and Jonas Kuhn
- 10:05-10:25 A Grammar-informed Corpus-based Sentence Database for Linguistic and Computational Studies
Hongzhi Xu, Helen Kaiyun Chen, Chu-Ren Huang, Qin Lu, Dingxu Shi and Tin-Shing Chiu
- 10:25-10:45 A Reference Dependency Bank for Analyzing Complex Predicates
Tafseer Ahmed, Miriam Butt, Annette Hautli and Sebastian Sulger
- 10:45-11:05 Announcing Prague Czech-English Dependency Treebank 2.0
Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová and Zdeněk Žabokrtský
- 11:05-11:25 Example-Based Treebank Querying
Liesbeth Augustinus, Vincent Vandeghinste and Frank Van Eynde

O40 - Semantic Lexicons and Semantic Annotation

Chairperson: Dimitrios Kokkinakis

Dolmabahçe A

- 9:45-10:05 A Cross-Lingual Dictionary for English Wikipedia Concepts
Valentin I. Spitzkovsky and Angel X. Chang
- 10:05-10:25 A database of semantic clusters of verb usages
Silvie Cinková, Martin Holub, Adam Rambousek and Lenka Smejkalová
- 10:25-10:45 Is it Useful to Support Users with Lexical Resources? A User Study.
Ernesto William De Luca
- 10:45-11:05 A review corpus annotated for negation, speculation and their scope
Natalia Konstantinova, Sheila C.M. de Sousa, Noa P. Cruz, Manuel J. Maña, Maite Taboada and Ruslan Mitkov
- 11:05-11:25 Developing a large semantically annotated corpus
Valerio Basile, Johan Bos, Kilian Evang and Noortje Venhuizen

Poster Sessions - Poster Area (1)

P34 Corpus Creation, Processing, Usage (3)

Chairperson: German Rigau

Le Petit Prince in UNL

Ronaldo Martins

A generic formalism to represent linguistic corpora in RDF and OWL/DL

Christian Chiarcos

A Database of Attribution Relations

Silvia Pareti

KPW: Towards a Free Corpus of Polish

Bartosz Broda, Michal Marcińczuk, Marek Maziarz, Adam Radziszewski and Adam Wardynski

Construction of the Turkish National Corpus (TNC)

Yeşim Aksan, Mustafa Aksan, Ahmet Koltuksuz, Taner Sezer, Ümit Mersinli, Umut Ufuk Demirhan, Hakan Yılmaz, Gülsüm Atasoy, Seda Öz, İpek Yıldız and Özlem Kurtoğlu

Building a learner corpus

Jirka Hana, Alexandr Rosen, Barbora Štindlová and Petr Jäger

Pedagogical stances and their multimodal signals.

Isabella Poggi, Francesca D'Errico and Giovanna Leone

Annotated Corpora for Word Alignment between Japanese and English and its Evaluation with MAP-based Word Aligner

Tsuyoshi Okita

Ubiquitous Usage of a Broad Coverage French Corpus: Processing the Est Republicain corpus

Djamé Seddah, Marie Candito, Benoit Crabbé and Enrique Henestroza Anguiano

P35 - Language Resource Infrastructures (2)

Chairperson: Claudia Soria

Federated Search: Towards a Common Search Infrastructure

Herman Stehouwer, Matej Durco, Eric Auer and Daan Broeder

Proper Language Resource Centers

Willem Elbers, Daan Broeder and Dieter van Uytvanck

The Language Archive - a new hub for language resources

Sebastian Drude, Daan Broeder, Paul Trilsbeek and Peter Wittenburg

LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis

Kais Dukes and Eric Atwell

An NLP Curator (or: How I Learned to Stop Worrying and Love NLP Pipelines)

James Clarke, Vivek Srikumar, Mark Sammons and Dan Roth

Using Language Resources in Humanities research

Marta Villegas, Núria Bel, Carlos Gonzalo, Amparo Moreno and Nuria Simelio

Glottolog/Langdoc: Increasing the visibility of grey literature for low-density languages

Sebastian Nordhoff and Harald Hammarström

The Australian National Corpus: National Infrastructure for Language Resources

Steve Cassidy, Michael Haugh, Pam Peters and Mark Fallu

META-SHARE v2: An Open Network of Repositories for Language Resources including Data and Tools

Christian Federmann, Ioanna Giannopoulou, Christian Girardi, Olivier Hamon, Dimitris Mavroeidis, Salvatore Minutoli and Marc Schröder

Linguagrid: a network of Linguistic and Semantic Services for the Italian Language.

Alessio Bosca, Luca Dini, Milen Kouylekov and Marco Trevisan

P36 - Speech Synthesis

Chairperson: Martine Garnier-Rizet

Versatile Speech Databases for High Quality Synthesis for Basque

Iñaki Sainz, Daniel Erro, Eva Navas, Inma Hernández, Jon Sánchez, Ibon Saratxaga and Igor Odriozola

Building a synchronous corpus of acoustic and 3D facial marker data for adaptive audio-visual speech synthesis

Dietmar Schabus, Michael Pucher and Gregor Hofer

Building Text-To-Speech Voices in the Cloud

Alistair Conkie, Thomas Okken, Yeon-Jun Kim and Giuseppe Di Fabbrizio

Building Synthetic Voices in the META-NET Framework

Emilia Garcia Casademont, Antonio Bonafonte and Asunción Moreno

Building Text-to-Speech Systems for Resource Poor Languages

Nur-Hana Samsudin and Mark Lee

Evaluating expressive speech synthesis from audiobook corpora for conversational phrases

Eva Szekely, Joao Paulo Cabral, Mohamed Abou-Zleikha, Peter Cahill and Julie Carson-Berndsen

P37 - Speech Resources

Chairperson: Andrea Paoloni

Body-conductive acoustic sensors in human-robot communication

Panikos Heracleous, Carlos Ishi, Takahiro Miyashita and Norihiro Hagita

Balanced data repository of spontaneous spoken Czech

Lucie Válková, Martina Waclawičová and Michal Křen

NKI-CCRT Corpus - Speech Intelligibility Before and After Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy

R.P. Clapham, L. van der Molen, R.J.J.H. van Son, M. van den Brekel and F.J.M. Hilgers

Sense Meets Nonsense - Sense Meets Nonsense - a dual-layer Danish speech corpus for perception studies

Thomas Ulrich Christiansen and Peter Juel Henriksen

SMALLWorlds -- Multilingual Content-Controlled Monologues

Peter Juel Henriksen and Marcus Uneson

A Parameterized and Annotated Spoken Dialog Corpus of the CMU Let's Go Bus Information System

Alexander Schmitt, Stefan Ultes and Wolfgang Minker

Speech and Language Resources for LVCSR of Russian
Sergey Zablotskiy, Alexander Shvets, Maxim Sidorov, Eugene Semenkin and Wolfgang Minker

Dysarthric Speech Database for Development of QoLT Software Technology
Dae-Lim Choi, Bong-Wan Kim, Yeon-Whoa Kim, Yong-Ju Lee, Yongnam Um and Minhwa Chung

The annotation of the C-ORAL-BRASIL oral through the implementation of the Palavras Parser
Eckhard Bick, Heliana Mello, Alessandro Panunzi and Tommaso Raso

The Nordic Dialect Corpus
Janne Bondi Johannessen, Joel Priestley, Kristin Hagen, Anders Nøklestad and André Lynum

ULex: new data models and a mobile environment for corpus enrichment.
Dafydd Gibbon

Developing Partially-Transcribed Speech Corpus from Edited Transcriptions
Kengo Ohta, Masatoshi Tsuchiya and Seiichi Nakagawa

LDC Forced Aligner
Xiaoyi Ma

The KIT Lecture Corpus for Speech Translation
Sebastian Stüker, Florian Kraft, Christian Mohr, Teresa Herrmann, Eunah Cho and Alex Waibel

Development of Text and Speech database for Hindi and Indian English specific to Mobile Communication environment
Shyam Agrawal, Shweta Sinha, Pooja Singh and Jesper Olson

11:25-11:45 **Coffee break**

O41 - Machine Translation and Language Resources (2)

Chairperson: Atsushi Fujii

Anadolu Auditorium

11:45-12:05 Source-Language Dictionaries Help Non-Expert Users to Enlarge Target-Language Dictionaries for Machine Translation
Víctor M. Sánchez-Cartagena, Miquel Esplà-Gomis and Juan Antonio Pérez-Ortiz

12:05-12:25 The ML4HMT Workshop on Optimising the Division of Labour in Hybrid Machine Translation
Christian Federmann, Eleftherios Avramidis, Marta R. Costa-Jussà, Josef van Genabith, Maite Melero and Pavel Pecina

12:25-12:45 Alignment-based reordering for SMT
Maria Holmqvist, Sara Stymne, Lars Ahrenberg and Magnus Merkel

12:45-13:05 Same domain different discourse style - A case study on Language Resources for data-driven Machine Translation
Monica Gavrilă, Walther v. Hahn and Cristina Vertan

13:05-13:25 Automatic Translation of Scholarly Terms into Patent Terms Using Synonym Extraction Techniques
Hidetsugu Nanba, Toshiyuki Takezawa, Kiyoko Uchiyama and Akiko Aizawa

O42 - WordNets

Chairperson: Martha Palmer

Topkapı A

- 11:45-12:05 Towards a richer wordnet representation of properties
Sanni Nimb and Bolette Sandford Pedersen
- 12:05-12:25 A proposal for improving WordNet Domains
Aitor Gonzalez-Agirre, Mauro Castillo and German Rigau
- 12:25-12:45 Corpus+WordNet thesaurus generation for ontology enriching
Fernando Castilho, Roger Granada, Breno Meneghetti, Leonardo Carvalho and Renata Vieira
- 12:45-13:05 Cleaning noisy wordnets
Benoît Sagot and Darja Fišer
- 13:05-13:25 Wordnet extension made simple: A multilingual lexicon-based approach using wiki resources
Valérie Hanoka and Benoît Sagot

O43 - Text Mining

Chairperson: Paul Rayson

Topkapı B

- 11:45-12:05 A Survey of Text Mining Architectures and the UIMA Standard
Mathias Bank and Martin Schierle
- 12:05-12:25 Large Scale Semantic Annotation, Indexing and Search at The National Archives
Diana Maynard and Mark A. Greenwood
- 12:25-12:45 Expertise Mining for Enterprise Content Management
Georgeta Bordea, Sabrina Kirrane, Paul Buitelaar and Bianca Pereira
- 12:45-13:05 SemSim: Resources for Normalized Semantic Similarity Computation Using Lexical Networks
Elias Iosif and Alexandros Potamianos
- 13:05-13:25 Identification of Manner in Bio-Events
Raheel Nawaz, Paul Thompson and Sophia Ananiadou

O44 - Evaluation of Systems and Application

Chairperson: Sabine Schulte im Walde

Dolmabahçe A

- 11:45-12:05 Cross-lingual studies of ASR errors: paradigms for perceptual evaluations
Ioana Vasilescu, Martine Adda-Decker and Lori Lamel
- 12:05-12:25 Practical Evaluation of Human and Synthesized Speech for Virtual Human Dialogue Systems
Kallirroi Georgila, Alan Black, Kenji Sagae and David Traum
- 12:25-12:45 Designing an Evaluation Framework for Spoken Term Detection and Spoken Document Retrieval at the NTCIR-9 SpokenDoc Task
Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyooki Aikawa, Tatsuya Kawahara and Tomoko Matsui

- 12:45-13:05 Evaluation of the KomParse Conversational Non-Player Characters in a Commercial Virtual World
Tina Kluewer, Feiyu Xu, Peter Adolphs and Hans Uszkoreit
- 13:05-13:25 The IWSLT 2011 Evaluation Campaign on Automatic Talk Translation
Marcello Federico, Sebastian Stüker, Luisa Bentivogli, Michael Paul, Mauro Cettolo, Teresa Herrmann, Jan Niehues and Giovanni Moretti

11:45-13:25

Poster Sessions - Poster Area (2)

P38 - Subjectivity: Sentiments, Emotions, Opinions (2)

Chairperson: Paolo Rosso

MLSA - A Multi-layered Reference Corpus for German Sentiment Analysis
Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger and Michael Wiegand

A Classification of Adjectives for Polarity Lexicons Enhancement
Silvia Vázquez and Nùria Bel

SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis
Jorge Carrillo de Albornoz, Laura Plaza and Pablo Gervás

“Vreselijk mooi!” (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives.
Tom De Smedt and Walter Daelemans

Visualizing Sentiment Analysis on a User Forum
Rasmus Sundberg, Anders Eriksson, Johan Bini and Pierre Nugues

Affective Common Sense Knowledge Acquisition for Sentiment Analysis
Erik Cambria, Yunqing Xia and Amir Hussain

P39 - Language Resource Infrastructures (3)

Chairperson: Penny Labropoulou

A Repository for the Sustainable Management of Research Data
Emanuel Dima, Verena Henrich, Erhard Hinrichs, Marie Hinrichs, Christina Hoppermann, Thorsten Trippel, Thomas Zastrow and Claus Zinn

Towards a comprehensive open repository of Polish language resources
Maciej Ogrodniczuk, Piotr Pęzik and Adam Przepiórkowski

The open lexical infrastructure of Språkbanken
Lars Borin, Markus Forsberg, Leif-Jöran Olsson and Jonatan Uppström

The Open Linguistics Working Group
Christian Chiarcos, Sebastian Hellmann, Sebastian Nordhoff, Steven Moran, Richard Littauer, Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek and Christian M. Meyer

GATetoGerManC: A GATE-based Annotation Pipeline for Historical German
Silke Scheible, Richard J. Whitt, Martin Durrell and Paul Bennett

Tackling interoperability issues within UIMA work flows
Nicolas Hernandez

P40 - Knowledge and Ontologies

Chairperson: Robert Gaizauskas

Knowledge-Rich Context Extraction and Ranking with KnowPipe

Anne-Kathrin Schumann

Application of a Semantic Search Algorithm to Semi-Automatic GUI Generation

Maria Teresa Pazienza, Noemi Scarpato and Armando Stellato

The KnowledgeStore: an Entity-Based Storage System

Roldano Cattoni, Francesco Corcoglioniti, Christian Girardi, Bernardo Magnini, Luciano Serafini and Roberto Zanolli

Tools for pIWordNet Development. Presentation and Perspectives

Bartosz Broda, Marek Maziarz and Maciej Piasecki

Combining Formal Concept Analysis and semantic information for building ontological structures from texts : an exploratory study

Silvia Moraes and Vera Lima

RELcat: a Relation Registry for ISocat data categories

Menzo Windhouwer

A disambiguation resource extracted from Wikipedia for semantic annotation

Eric Charton and Michel Gagnon

NLP Challenges for Eunomos a Tool to Build and Manage Legal Knowledge

Guido Boella, Luigi di Caro, Llio Humphreys, Livio Robaldo and Leon van der Torre

Representing General Relational Knowledge in ConceptNet 5

Robert Speer and Catherine Havasi

P41 - Semantics

Chairperson: Marc Verhagen

A new dynamic approach for lexical networks evaluation

Alain Joubert and Mathieu Lafourcade

LIE: Leadership, Influence and Expertise

Roberta Catizone, Louise Guthrie, Arthur Thomas and Yorrick Wilks

Semantic Role Labeling with the Swedish FrameNet

Richard Johansson, Karin Friberg Heppin and Dimitrios Kokkinakis

Extending a wordnet framework for simplicity and scalability

Pedro Fialho, Sérgio Curto, Ana Cristina Mendes and Luisa Coheur

German "nach"-Particle Verbs in Semantic Theory and Corpus Data

Boris Haselbach, Wolfgang Seeker and Kerstin Eckart

LexIt: A Computational Resource on Italian Argument Structure

Alessandro Lenci, Gabriella Lapesa and Giulia Bonansinga

Enriching the ISST-TANL Corpus with Semantic Frames

Alessandro Lenci, Simonetta Montemagni, Giulia Venturi and Maria Grazia Cutrullà

P42 - Temporal Information

Chairperson: Uwe Quasthoff

TimeBankPT: A TimeML Annotated Corpus of Portuguese
Francisco Costa and António Branco

SUtime: A library for recognizing and normalizing time expressions
Angel X. Chang and Christopher Manning

Temporal Annotation: A Proposal for Guidelines and an Experiment with Inter-annotator Agreement
André Bittar, Caroline Hagège, Véronique Moriceau, Xavier Tannier and Charles Teissèdre

Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards
Jannik Strötgen and Michael Gertz

Massively Increasing TIMEX3 Resources: A Transduction Approach
Leon Derczynski, Hector Llorens and Estela Saquete

Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information
Corina Forascu and Dan Tufis

P43 - Sign Language

Chairperson: Thomas Hanke

Detecting Reduplication in Videos of American Sign Language
Zoya Gavrilov, Stan Sclaroff, Carol Neidle and Sven Dickinson

BiBiKit - A Bilingual Bimodal Reading and Writing Tool for Sign Language Users
Nedelina Ivanova and Olle Eriksen

Resource production of written forms of Sign Languages by a user-centered editor, SWift (SignWriting improved fast transcriber)
Fabrizio Borgia, Claudia S. Bianchini, Patrice Dalle and Maria De Marsico

RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus
Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater and Hermann Ney

13:25-14:55 Lunch break

O45 - New Media (Special Session)

Chairperson: Thierry Declerk

Anadolu Auditorium

14:55-15:15 Two Database Resources for Processing Social Media English Text
Eleanor Clark and Kenji Araki

15:15-15:35 Holaaa!! writin like u talk is kewl but kinda hard 4 NLP
Maite Melero, Marta R. Costa-Jussà, Judith Domingo, Montse Marquina and Martí Quixal

15:35-15:55 Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events
William J. Corvey, Sudha Verma, Sarah Vieweg, Martha Palmer and James H. Martin

- 15:55-16:15 EmpaTweet: Annotating and Detecting Emotions on Twitter
Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie and Sanda M. Harabagiu
- 16:15-16:35 Discussion

O46 - Semantics, Knowledge and Ontologies

Chairperson: Christian Chiarcos

Topkapı A

- 14:55-15:15 Semantic Relations Established by Specialized Processes Expressed by Nouns and Verbs: Identification in a Corpus by means of Syntactico-semantic Annotation
Nava Maroto, Marie-Claude L'Homme and Amparo Alcina
- 15:15-15:35 Using Wikipedia to Validate the Terminology found in a Corpus of Basic Textbooks
Jorge Vivaldi, Luis Adrián Cabrera-Diego, Gerardo Sierra and María Pozzi
- 15:35-15:55 PEARL: ProjECTION of Annotations Rule Language, a Language for Projecting (UIMA) Annotations over RDF Knowledge Bases
Maria Teresa Pazienza, Armando Stellato and Andrea Turbati
- 15:55-16:15 Constructing Large Proposition Databases
Peter Exner and Pierre Nugues
- 16:15-16:35 Highlighting relevant concepts from Topic Signatures
Montse Cuadros, Lluís Padró and German Rigau

O47 - Segmentation, Tagging, Parsing

Chairperson: Valia Kordoni

Topkapı B

- 14:55-15:15 Towards an LFG parser for Polish: An exercise in parasitic grammar development
Agnieszka Patejuk and Adam Przepiórkowski
- 15:15-15:35 Using a Goodness Measurement for Domain Adaptation: A Case Study on Chinese Word Segmentation
Yan Song and Fei Xia
- 15:35-15:55 The Dependency-Parsed FrameNet Corpus
Daniel Bauer, Hagen Fürstenau and Owen Rambow
- 15:55-16:15 Predicting Phrase Breaks in Classical and Modern Standard Arabic Text
Majdi Sawalha, Claire Brierley and Eric Atwell
- 16:15-16:35 Parsing Any Domain English text to CoNLL dependencies
Sudheer Kolachina and Prasanth Kolachina

O48 - Named Entities and Subjectivity

Chairperson: Dan Tufis

Dolmabahçe A

- 14:55-15:15 Iterative Refinement and Quality Checking of Annotation Guidelines - How to Deal Effectively with Semantically Sloppy Named Entity Types, such as Pathological Phenomena
Udo Hahn, Elena Beisswanger, Ekaterina Buyko, Erik Faessler, Jenny Traumüller, Susann Schröder and Kerstin Hornbostel

- 15:15-15:35 GerNED: A German Corpus for Named Entity Disambiguation
Danuta Ploch, Leonhard Hennig, Angelina Duka, Ernesto William De Luca and Sahin Albayrak
- 15:35-15:55 Centroids: Gold standards with distributional variation
Ian Lewin, Şenay Kafkas and Dietrich Rebholz-Schuhmann
- 15:55-16:15 Quantising Opinions for Political Tweets Analysis
Yulan He, Hassan Saif, Zhongyu Wei and Kam-Fai Wong
- 16:15-16:35 AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis
Muhammad Abdul-Mageed and Mona Diab

14:55-16:35 **Poster Sessions - Poster Area (1) 40**

P44 - Machine Translation (2)

Chairperson: Jan Hajic

Arabic-Segmentation Combination Strategies for Statistical Machine Translation
Saab Mansour and Hermann Ney

The Joy of Parallelism with CzEng 1.0
Ondřej Bojar, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel and Aleš Tamchyna

Statistical Machine Translation without Source-side Parallel Corpus Using Word Lattice and Phrase Extension
Takanori Kusumoto and Tomoyosi Akiba

Automatic Translation of Scientific Documents in the HAL Archive
Lambert Patrik, Holger Schwenk and Frédéric Blain

Expanding Parallel Resources for Medium-Density Languages for Free
Georgi Iliev and Angel Genov

VERTa: Linguistic features in MT evaluation
Elisabet Comelles, Jordi Aterias, Victoria Arranz and Irene Castellón

Linguistic Resources for Handwriting Recognition and Translation Evaluation
Zhiyi Song, Safa Ismael, Stephen Grimes, David Doermann and Stephanie Strassel

Development and Application of a Cross-language Document Comparability Metric
Fangzhong Su and Bogdan Babych

Assessing Divergence Measures for Automated Document Routing in an Adaptive MT System
Claire Jaja, Douglas Briesch, Jamal Laoudi and Clare Voss

A Study of Word-Classing for MT Reordering
Ananthakrishnan Ramanathan and Karthik Visweswariah

Dealing with unknown words in statistical machine translation
João Silva, Luísa Coheur, Ângela Costa and Isabel Trancoso

PET: a Tool for Post-editing and Assessing Machine Translation
Wilker Aziz, Sheila Castilho and Lucia Specia

Tajik-Farsi Persian Transliteration Using Statistical Machine Translation
Chris Irwin Davis

Assessing the Comparability of News Texts
Emma Barker and Robert Gaizauskas

P45 - Natural Language Generation

Chairperson: Dan Cristea

Corpus-based Referring Expressions Generation

Hilder Pereira, Eder Novais, Andre Mariotti and Ivandre Paraboni

Portuguese Text Generation from Large Corpora

Eder Novais, Ivandre Paraboni and Douglas Silva

DSim, a Danish Parallel Corpus for Text Simplification

Sigrid Klerke and Anders Søgaard

Acquisition of Syntactic Simplification Rules for French

Violeta Seretan

A Repository of Data and Evaluation Resources for Natural Language Generation

Anja Belz and Albert Gatt

LG-Eval: A Toolkit for Creating Online Language Evaluation Experiments

Eric Kow and Anja Belz

P46 - Crowdsourcing

Chairperson: Collin Baker

Turk Bootstrap Word Sense Inventory 2.0: A Large-Scale Resource for Lexical Substitution

Chris Biemann

Collection of a Large Database of French-English SMT Output Corrections

Marion Potet, Emmanuelle Esperança-Rodier, Laurent Besacier and Hervé Blanchon

Getting more data -- Schoolkids as annotators

Jirka Hana and Barbora Hladka

Word Sense Inventories by Non-Experts.

Anna Rumshisky, Nick Botchan, Sophie Kushkuley and James Pustejovsky

P47 - Text Mining and Text Entailment

Chairperson: Sophia Ananiadou

The BladeMistress Corpus: From Talk to Action in Virtual Worlds

Anton Leuski, Carsten Eickhoff, James Ganis and Victor Lavrenko

Annotating Factive Verbs

Alvin Grissom II and Yusuke Miyao

A Holistic Approach to Bilingual Sentence Fragment Extraction from Comparable Corpora

Mahdi Khademan, Kaveh Taghipour, Saab Mansour and Shahram Khadivi

An Examination of Cross-Cultural Similarities and Differences from Social Media Data with respect to Language Use

Mohammad Fazleh Elahi and Paola Monachesi

Turkish Paraphrase Corpus

Seniz Demir, Ilknur Durgar El-Kahlout, Erdem Unal and Hamza Kaya

Constructing a Question Corpus for Textual Semantic Relations

Rui Wang and Shuguang Li

Evaluating the Similarity Estimator component of the TWIN Personality-based Recommender System

Alexandra Roshchina, John Cardiff and Paolo Rosso

P48 - Speech/Multimodal Tools, Systems, Applications

Chairperson: Tomaž Erjavec

Annotation Facilities for the Reliable Analysis of Human Motion

Michael Kipp

Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research

Michael Carl

Intelligibility assessment in forensic applications

Giovanni Costantini, Andrea Paoloni and Massimiliano Todisco

Strategies to Improve a Speaker Diarisation Tool

David Tavares, Eva Navas, Daniel Erro and Ibon Saratxaga

Using an ASR database to design a pronunciation evaluation system in Basque

Igor Odriozola, Eva Navas, Inma Hernández, Iñaki Sainz, Ibon Saratxaga, Jon Sánchez and Daniel Erro

W-PhAMT: A web tool for phonetic multilevel timeline visualization

Francesco Cutugno, Vincenza Anna Leano and Antonio Origlia

English to Indonesian Transliteration to Support English Pronunciation Practice

Amalia Zahra and Julie Carson-Berndsen

Rapidly Testing the Interaction Model of a Pronunciation Training System via Wizard-of-Oz

Joao Paulo Cabral, Mark Kane, Zeeshan Ahmed, Mohamed Abou-Zleikha, Eva Szekely, Amalia Zahra, Kalu Ogbureke, Peter Cahill, Julie Carson-Berndsen and Stephan Schlogl

PAMOCAT: Automatic retrieval of specified postures

Bernhard Brüning, Christian Schnier, Karola Pitsch and Sven Wasmuth

Coffee break

16:35-16:55

16:55-17:35

Antonio Zampolli Prize

17:35-18:15

Closing Session

21:00

Gala Dinner

Sponsors

SILVER SPONSORS

CELI

CELI is a major provider of software solutions in the field of Natural Language Processing (NLP). Our mission is to improve the process of extracting meaningful intelligence from unstructured multilingual text by developing industry standard linguistic resources, technologies and solutions.

Our monolingual, multilingual, and cross-language technologies apply to language analysis, information extraction, text mining, information retrieval, speech technology, opinion monitoring, machine translation, etc.

CELI is deeply committed to research in order to ensure its position at the forefront of the technological curve. To achieve this result, CELI is involved in important research activities and cooperates with world class institutions on a national and international level.

CELI devotes a large part of its activity to the creation and maintenance of language resources including: morphological databases, monolingual and multilingual lexicons, text corpora, semantic networks. This activity is coupled with language consulting services for over 20 languages in order to help our customers better manage their linguistic assets. Covered languages: Italian, English, French, Spanish, Catalan, Portuguese, German, Dutch, Swedish, Norwegian, Finnish, Danish, Polish, Russian, Estonian, Latvian, Lithuanian, Ukrainian, Greek, Turkish, Arabic, Hebrew, Armenian, Albanian, Croatian, Serbian, Czech, Slovak, Slovenian, Romanian, Bulgarian, Hungarian, Chinese, Japanese.

<http://www.celi.it>

NUANCE

Nuance is the leading provider of speech and imaging solutions for businesses and consumers around the world. Our technologies, applications and services make the user experience more compelling by transforming the way people interact with information and how they create, share and use documents. Every day, millions of users and thousands of businesses experience Nuance by calling directory assistance, requesting account information, dictating patient records, searching the mobile web by voice, telling a navigation system their destination, or digitally reproducing documents that can be shared and searched. Making each of those experiences productive and compelling is what Nuance is all about.

- Nuance has more than 7000 employees in more than 35 offices worldwide.
- Nuance speech and imaging products are used in more the 70 countries and support more than 50 languages
- More than ten billion phone conversations each year are automated with Nuance speech
- Nuance technologies ship in more than 1 billion mobile devices in 2012
- More than 350,000 physicians in 3,000 hospitals and clinics use our dictation solutions
- Nuance has more than 1000 filed and pending speech patents

<http://www.nuance.com>

BRONZE SPONSORS

European Media Laboratory (EML)

The European Media Laboratory GmbH is a private IT enterprise established by Klaus Tschira, one of the founders of the SAP AG software company. In accordance with its motto " Experience IT – Intuitive Technology ", EML is successfully pursuing research and development in the fields of human-computer interaction and automatic speech processing.

For more information see the EML company portrait on our website:

<http://www.eml-development.de/english/>

IMMI-CNRS, A joint Research Unit between LIMSI-CNRS, RWTH and KIT

The Institute for Multilingual and Multimedia Information (IMMI – www.immi-labs.org) brings together the Rheinisch-Westphalian Technical University Aachen (RWTH), the Karlsruhe Institute of Technology (KIT) and LIMSI-CNRS in the field of Multimedia and Multilingual Document Processing within the French Quaero program. The aim is to properly address multilingualism which constitutes a major challenge in Europe and worldwide. The scientific scope focuses on research in Natural Language Processing, Speech and Audio Processing, and Written and Spoken Language Translation. The partners share the same approach based on the use of automatic learning methods, language resources and language technology evaluation, in order to jointly address a problem and compare on the same data, allowing comprehensive scientific directions, measure of progress and estimate of the task/technology adequacy. During the past two years, IMMI has been involved in the FLaReNet and META-NET networks and has co-organized IWSLT 2010 and 2011, Tralogy "Translation and Technology" 2011 and 2013, and will host in Paris IWSDS in November 2012.

<http://www.immi-labs.org>

K Dictionaries

K DICTIONARIES offer world leading content and solutions for dozens of languages and all media, for learning, professional and general purposes. We create smart dictionary databases and applications for monolingual, bilingual and multilingual use, interweaved with each other and covering a wide range of English dictionaries with over forty language versions and a global series with over twenty languages. Our partners include leading publishing houses, technology, information, education and communication firms, universities, associations and editors worldwide.

<http://kdictionaries.com/>

META-NET

META-NET is a **Network of Excellence** of dedicated to fostering the technological foundations of the European multilingual information society by setting up **META**, the Multilingual Europe Technology Alliance to:

- enable communication and collaboration across language barriers
- secure users of any language equal access to the information and knowledge society
- advance the functionalities of networked information technology

META-NET brings together major stakeholders in Language Technology to further these goals by creating a large diverse community with a shared vision and Strategic Research Agenda. Since our launch in 2010, **META-NET** has grown to include 57 research centres in 33 countries and includes the CESAR, METANET4U, and META-NORD consortia.

To find out more and have your say visit our stand at the **EC Projects Village**.

<http://www.meta-net.eu>

Quaero

Quaero is a research and innovation program addressing automatic processing of multimedia and multilingual content aiming at the development of new tools for navigating in large volumes of text and audiovisual content. The research and development undertaken covers automatic information retrieval, analysis, segmentation and classification of text, speech, music, image and video. The program, supported by OSEO, gathers 32 French and German partners – large groups, small and medium size enterprises, research laboratories and public organizations. The program consists of a number of application projects aiming at industrial targets and markets, that are supported by a common shared research structure. Real world data sets (corpora) are used to define the evaluation tasks and to conduct the research challenges between partners. The use of systematic periodic technology evaluation allows to assess progress made and to select the most promising technical and scientific approaches. After four years of existence, Quaero is a very active eco-system that has produced in excess of 580 scientific publications, more than 20 awards, numerous top 3 rankings in technology evaluation campaigns, 20 patent applications and many innovative prototypes

<http://www.quaero.org>

WITH THE SUPPORT OF:

Star Alliance, Official Carrier for LREC 2010

Wherever your conferences take you, Star Alliance can take you.

No matter where you're travelling from, the Star Alliance™ network offers you a wide choice of flights to Istanbul. And with 20,500 flights a day to over 1,293 destinations across 190 countries, our 25 member airlines extend the same choice to any future conferences you're planning to attend. You can also save money when you book your flights. Simply quote the unique conventions code and you and one companion travelling with you will receive a special discount. Better still, no matter which Star Alliance member airline's frequent flyer programme you belong to, you can earn and redeem miles across all 25 airlines.

For more information, or to join the airline network that offers you more choice wherever your conferences take you, simply go to staralliance.com/conventionsplus.

<http://www.staralliance.com>

