
29th May 2008

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

**Julia Hirschberg: Detecting Deceptive Speech:
Requirements, Resources and Evaluation**

Chairperson: Daniel Tapias

Ministres

9:40-9:45 *5-minute Break*

9:45-11:25 **4 Oral Sessions and 1 Poster Session in Parallel**

O15 - LRs: Large Programs, Policies, Strategies

Chairperson: Steven Krauwer

Ministres

9:45-10:05 The Dutch-Flemish Comprehensive Approach to HLT Stimulation and Innovation: STEVIN, HLT Agency and beyond
Peter Spyns, Elisabeth D'Halleweyn and Catia Cucchiarini

10:05-10:25 15 Years of Language Resource Creation and Sharing: a Progress Report on LDC Activities
Christopher Cieri and Mark Liberman

10:25-10:45 Estimating the Resource Adaption Cost from a Resource Rich Language to a Similar Resource Poor Language
Anil Kumar Singh, Kiran Pala and Harshit Surana

10:45-11:05 Latest Developments in ELRA's Services
Valérie Mapelli, Victoria Arranz, Hélène Mazo and Khalid Choukri

11:05-11:25 From Research to Application in Multilingual Information Access: the Contribution of Evaluation
Carol Peters, Martin Braschler, Giorgio Di Nunzio, Nicola Ferro, Julio Gonzalo and Mark Sanderson

O16 - Biomedical Resources

Chairperson: Jian Su

Ambassadeurs

9:45-10:05 Clustering Related Terms with Definitions
Scott Piao, John McNaught and Sophia Ananiadou

10:05-10:25 Challenges in Pronoun Resolution System for Biomedical Text
Ngan Nguyen, Jin-Dong Kim and Jun'ichi Tsujii

10:25-10:45 Exploiting Multiply Annotated Corpora in Biomedical Information Extraction Tasks
Barry Haddow and Beatrice Alex

10:45-11:05 GENIA-GR: a Grammatical Relation Corpus for Parser Evaluation in the Biomedical Domain
Yuka Tateisi, Yusuke Miyao, Kenji Sagae and Jun'ichi Tsujii

11:05-11:25 Learning the Species of Biomedical Named Entities from Annotated Corpora
Xinglong Wang and Claire Grover

O17 - Semantics in Lexicons and Corpora

Chairperson: Nancy Ide

Fez 1

9:45-10:05 Acquiring Naturalistic Concept Descriptions from the Web
Tony Veale and Yanfen Hao

10:05-10:25 Tools for Collocation Extraction: Preferences for Active vs. Passive
Ulrich Heid and Marion Weller

10:25-10:45 Boot-Strapping a WordNet Using Multiple Existing WordNets
Francis Bond, Hitoshi Isahara, Kyoko Kanzaki and Kiyotaka Uchimoto

10:45-11:05 Corpus-based Semantic Relatedness for the Construction of Polish WordNet
Bartosz Broda, Magdalena Derwojedowa, Maciej Piasecki and Stanislaw Szpakowicz

11:05-11:25 Developing Verb Frames for Hindi
Rafiya Begum, Samar Husain, Lakshmi Bai and Dipti Misra Sharma

O18 - Affect and Emotion in Speech

Chairperson: Nick Campbell

Fez 2

9:45-10:05 Uncertainty Corpus: Resource to Study User Affect in Complex Spoken Dialogue Systems
Kate Forbes-Riley, Diane Litman, Scott Silliman and Amruta Purandare

10:05-10:25 On the Role of the NIMITEK Corpus in Developing an Emotion Adaptive Spoken Dialogue System
Milan Gnjatovic and Dietmar Roesner

- 10:25-10:45 Emotion Recognition from Speech: Stress Experiment
Stefan Scherer, Hansjörg Hofmann, Malte Lampmann, Martin Pfeil, Steffen Rhinow, Friedhelm Schwenker and Günther Palm
- 10:45-11:05 Automatic Phone Segmentation of Expressive Speech
Laure Charonnat, Gaëlle Vidal and Olivier Boeffard
- 11:05-11:25 Multimodal Spontaneous Expressive Speech Corpus for Hungarian
Márk Fék, Nicolas Audibert, János Szabó, Albert Rilliard, Géza Németh and Véronique Aubergé
- 9:45-11:25 **Poster Sessions Asni Room**

P13 - Evaluation: Resources, Tools, Methodologies, Campaigns

Chairperson: Margaret King

Evaluation of several Maximum Likelihood Linear Regression Variants for Language Adaptation.

Miriam Luján Mares, Carlos David Martínez Hinarejos and Vicent Alabau Gonzalvo

Evaluation of Lexical Resources and Semantic Networks on a Corpus of Mental Associations

Laurianne Sibon, Patrice Bellot and Philippe Blache

Measures for Term and Sentence Relevances: an Evaluation for German

Heike Bieler and Stefanie Dipper

Annotation of Information Structure: an Evaluation across different Types of Texts

Julia Ritz, Stefanie Dipper and Michael Götze

Word Segmentation of Vietnamese Texts: a Comparison of Approaches

ĐINH Quang Thắng, LÊ Hồng Phương, NGUYỄN Thị Minh Huyền, NGUYỄN Cẩm Tú, Mathias ROSSIGNOL and VŨ Xuân Lương

Comparing Italian parsers on a common Treebank: the EVALITA experience

Cristina Bosco, Alessandro Mazzei, Vincenzo Lombardo, Giuseppe Attardi, Anna Corazza, Alberto Lavelli, Leonardo Lesmo, Giorgio Satta and Maria Simi

Evaluation of Natural Language Tools for Italian: EVALITA 2007

Bernardo Magnini, Amedeo Cappelli, Fabio Tamburini, Cristina Bosco, Alessandro Mazzei, Vincenzo Lombardo, Francesca Bertagna, Nicoletta Calzolari, Antonio Toral, Valentina Bartalesi Lenzi, Rachele Sprugnoli and Manuela Speranza

A Bottom-up Comparative Study of EuroWordNet and WordNet 3.0 Lexical and Semantic Relations

Maria Teresa Pazienza, Armando Stellato and Alexandra Tudorache

Evaluating the Ontology underlying sMail - the Conceptual Framework for Semantic Email Communication

Simon Scerri, Myriam Mencke, Brian Davis and Siegfried Handschuh

Inter-sentential Coreferences in Semantic Networks: An Evaluation of Manual Annotation

Václav Novák and Keith Hall

Diacritic Annotation in the Arabic Treebank and its Impact on Parser Evaluation

Mohamed Maamouri, Seth Kulick and Ann Bies

Evaluation of Virtual Keyboards for West-African Languages

Chantal Enguehard and Harouna Naroua

Anaphora Resolution Exercise: an Overview

Constantin Orăsan, Dan Cristea, Ruslan Mitkov and António Branco

Portuguese-English Word Alignment: some Experiments

Diana Santos and Alberto Simões

System Evaluation on a Named Entity Corpus from Clinical Notes

Karin Schuler, Vinod Kaggal, James Masanz, Philip Ogren and Guergana Savova

Constructing Evaluation Corpora for Automated Clinical Named Entity Recognition

Philip Ogren, Guergana Savova and Christopher Chute

Assessing the Costs of Machine-Assisted Corpus Annotation through a User Study

Eric Ringger, Marc Carmen, Robbie Haertel, Kevin Seppi, Deryle Lonsdale, Peter McClanahan, James Carroll and Noel Ellison

Training and Evaluation of POS Taggers on the French MULTITAG Corpus

Alexandre Allauzen and Hélène Bonneau-Maynard

Cleaneval: a Competition for Cleaning Web Pages

Marco Baroni, Francis Chantree, Adam Kilgarriff and Serge Sharoff

A Ground Truth Dataset for Matching Culturally Diverse Romanized Person Names

Mark Arehart and Keith J Miller

Producing a Test Collection for Patent Machine Translation in the Seventh NTCIR Workshop

Atsushi Fujii, Masao Utiyama, Mikio Yamamoto and Takehito Utsuro

A Test Suite for Inference Involving Adjectives

Marilisa Amoia and Claire Gardent

Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments: newest Part of the CENSREC Series -

Takanobu Nishiura, Masato Nakayama, Yuki Denda, Norihide Kitaoka, Kazumasa Yamamoto, Takeshi Yamada, Satoru Tsuge, Chiyomi Miyajima, Masakiyo Fujimoto, Tetsuya Takiguchi, Satoshi Tamura, Shingo Kuroiwa, Kazuya Takeda and Satoshi Nakamura

An Experimental Methodology for an End-to-End Evaluation in Speech-to-Speech Translation

Olivier Hamon and Djamel Mostefa

11:25-11:45 *Coffee break*

11:45-13:25 **4 Oral Sessions and 1 Poster Session in Parallel**

O19 - Opinion Mining and Summarization

Chairperson: Sophia Ananiadou

Ministres

- 11:45-12:05 Vox Populi Annotation: Measuring Intensity of Ideological Perspectives by Aggregating Group Judgments
Wei-Hao Lin and Alexander Hauptmann
- 12:05-12:25 A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources
Carmen Banea, Rada Mihalcea and Janyce Wiebe
- 12:25-12:45 Finding the Sources and Targets of Subjective Expressions
Josef Ruppenhofer, Swapna Somasundaran and Janyce Wiebe
- 12:45-13:05 Annotating Topics of Opinions
Veselin Stoyanov and Claire Cardie
- 13:05-13:25 From Extracting to Abstracting: Generating Quasi-abstractive Summaries
Zhuli Xie, Barbara Di Eugenio and Peter C. Nelson

O20 - Coreference and Discourse

Chairperson: Antonio Branco

Ambassadeurs

- 11:45-12:05 Controlling Redundancy in Referring Expressions
Jette Viethen, Robert Dale, Emiel Kraemer, Mariet Theune and Pascal Touset
- 12:05-12:25 Anaphoric Annotation in the ARRAU Corpus
Massimo Poesio and Ron Artstein
- 12:25-12:45 Knowledge Sources for Bridging Resolution in Multi-Party Dialog
Mark-Christoph Mueller, Margot Mieskes and Michael Strube
- 12:45-13:05 The Penn Discourse TreeBank 2.0.
Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi and Bonnie Webber
- 13:05-13:25 A Coreference Corpus and Resolution System for Dutch
Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Veronique Hoste, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet and Jean-Luc Verschelde

O21 - Semantic Resources and Acquisition

Chairperson: Peter Anick

Fez 1

- 11:45-12:05 Statistical Identification of English Loanwords in Korean Using Automatically Generated Training Data
Kirk Baker and Chris Brew
- 12:05-12:25 Romanian Semantic Role Resource
Diana Trandabăţ and Maria Husarciu
- 12:25-12:45 Unsupervised Acquisition of Verb Subcategorization Frames from Shallow-Parsed Corpora
Alessandro Lenci, Barbara McGillivray, Simonetta Montemagni and Vito Pirrelli
- 12:45-13:05 A Method for Automatically Constructing Case Frames for English
Daisuke Kawahara and Kiyotaka Uchimoto
- 13:05-13:25 Automatic Acquisition for low frequency lexical items
Núria Bel, Sergio Espeja and Montserrat Marimon

O22 - Speaker and Dialect Identification

Chairperson: Andrea Paoloni

Fez 2

- 11:45-12:05 BioSec Multimodal Biometric Database in Text-Dependent Speaker Recognition
Doroteo Toledano, Daniel Hernandez-Lopez, Cristina Esteve-Elizalde, Julian Fierrez, Javier Ortega-Garcia, Daniel Ramos and Joaquin Gonzalez-Rodriguez
- 12:05-12:25 Text Independent Speaker Identification in Multilingual Environments
Iker Luengo, Eva Navas, Iñaki Sainz, Ibon Saratxaga, Jon Sanchez, Igor Odriozola and Inma Hernaez
- 12:25-12:45 NineOneOne: Recognizing and Classifying Speech for Handling Minority Language Emergency Calls
Udhyakumar Nallasamy, Alan Black, Tanja Schultz and Robert Frederking
- 12:45-13:05 Bridging the Gap between Linguists and Technology Developers: Large-Scale, Sociolinguistic Annotation for Dialect and Speaker Recognition
Christopher Cieri, Stephanie Strassel, Meghan Glenn, Reva Schwartz, Wade Shen and Joseph Campbell
- 13:05-13:25 Speaker Recognition: Building the Mixer 4 and 5 Corpora
Linda Brandschain, Christopher Cieri, David Graff, Abby Neely and Kevin Walker

P14 - Evaluation: Resources, Tools, Systems, Methodologies

Chairperson: Diana Santos

Evaluation of Different Segmentation Techniques for Dialogue Turns

Carlos D. Martínez-Hinarejos and Vicent Tamarit

Acquisition and Evaluation of a Dialog Corpus through WOZ and Dialog Simulation Techniques

David Griol, Lluís F. Hurtado, Encarna Segarra and Emilio Sanchis

What would you Ask a conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting

Susan Robinson, David Traum, Midhun Ittycheriah and Joe Henderer

An Evaluation of Spoken and Textual Interaction in the RITEL Interactive Question Answering System

Dave Toney, Sophie Rosset, Aurélien Max, Olivier Galibert and Eric Bilinski

Classification Procedures for Software Evaluation

Muriel Amar, Sophie David, Rachel Panckhurst and Lisa Whistlecroft

Cross-Corpus Evaluation of Word Alignment

Sylwia Ozdowska

Evaluating Evaluation Metrics for Ontology-Based Applications: Infinite Reflection

Diana Maynard, Wim Peters and Yaoyong Li

Lexical Substitution as a Framework for Multiword Evaluation

Diana McCarthy

Tree Distance and Some Other Variants of Evalb

Martin Emms

BLEU+: a Tool for Fine-Grained BLEU Computation

A. Cuneyd Tantug, Kemal Oflazer and Ilknur Durgar El-Kahlout

Elicited Imitation as an Oral Proficiency Measure with ASR Scoring

C. Ray Graham, Deryle Lonsdale, Casey Kennington, Aaron Johnson and Jeremiah McGhee

Methodology for Evaluating the Usability of User Interfaces in Mobile Services

Pedro Concejero, Daniel Tapias, Juan José Rodríguez, Juan Carlos Luengo and Sebastián Sánchez

An Economic View on Human Language Technology Evaluation

Edouard Geoffrois

Comparing Corpus-based to Web-based Lookup Techniques for Automatic English Inclusion Detection

Beatrice Alex

Centering Theory for Evaluation of Coherence in Computer-Aided Summaries

Laura Hasler

Linguistic Resources and Evaluation Techniques for Evaluation of Cross-Document Automatic Content Extraction

Stephanie Strassel, Mark Przybocki, Kay Peterson, Zhiyi Song and Kazuaki Maeda

Let's not Argue about Semantics

Johan Bos

Can we Evaluate the Quality of Generated Text?

David Harcastle and Donia Scott

An Infrastructure, Tools and Methodology for Evaluation of Multicultural Name Matching Systems

Keith J. Miller, Mark Arehart, Catherine Ball, John Polk, Alan Rubenstein, Kenneth Samuel, Elizabeth Schroeder, Eva Vecchi and Chris Wolf

Evaluating Robustness Of A QA System Through A Corpus Of Real-Life Questions

Laurianne Sibon, Patrice Bellot and Philippe Blache

Sentiment Analysis and the Use of Extrinsic Datasets in Evaluation

Ann Devitt and Khurshid Ahmad

Certification and Cleaning up of a Text Corpus: Towards an Evaluation of the "Grammatical" Quality of a Corpus

Cyril Grouin

WEB-Based Listening Test System for Speech Synthesis and Speech Conversion Evaluation

Laurent Blin, Olivier Boeffard and Vincent Barreaud

Semiotic-based Ontology Evaluation Tool (S-OntoEval)

Renata Dividino, Massimo Romanelli and Daniel Sonntag

ANNALIST - ANNotation ALIgnment and Scoring Tool

George Demetriou, Robert Gaizauskas, Haotian Sun and Angus Roberts

Task-Based Evaluation of Meeting Browsers: from Task Elicitation to User Behavior Analysis

Andrei Popescu-Belis, Mike Flynn, Pierre Wellner and Philippe Baudrion

Improving Contextual Quality Models for MT Evaluation Based on Evaluators' Feedback

Paula Estrella, Andrei Popescu-Belis and Maghi King

Performance Evaluation of Speech Translation Systems

Brian Weiss, Craig Schlenoff, Greg Sanders, Michelle Steves, Sherri Condon, Jon Phillips and Dan Parvaz

Automatic Evaluation Measures for Statistical Machine Translation System Optimization

Arne Mauser, Sasa Hasan and Hermann Ney

14:40-16:20 **4 Oral Sessions and 3 Poster Sessions in Parallel**

O23 - Corpus Annotation and Classification

Chairperson: Christopher Cieri

Ministres

- 14:40-15:00 MASC: the Manually Annotated Sub-Corpus of American English
Nancy Ide, Collin Baker, Christiane Fellbaum, Charles Fillmore and Rebecca Passonneau
- 15:00-15:20 Quality Assurance of Automatic Annotation of Very Large Corpora: a Study based on heterogeneous Tagging System
Chu-Ren Huang, Lung-Hao Lee, Jia-Fei Hog, Wei-guang Qu and Shiwen Yu
- 15:20-15:40 An eRulemaking Corpus: Identifying Substantive Issues in Public Comments
Claire Cardie, Cynthia Farina, Matt Rawding and Adil Aijaz
- 15:40-16:00 Navigating through Dense Annotation Spaces
Branimir Boguraev and Mary Neff
- 16:00-16:20 An Unsupervised Probabilistic Approach for the Detection of Outliers in Corpora
David Guthrie, Louise Guthrie and Yorick Wilks

O24 - Machine Translation and Multilinguality

Chairperson: Gregor Thurmair

Ambassadeurs

- 14:40-15:00 Using Log-linear Models for Tuning Machine Translation Output
Michael Carl
- 15:00-15:20 Generalising Lexical Translation Strategies for MT Using Comparable Corpora
Bogdan Babych, Serge Sharoff and Anthony Hartley
- 15:20-15:40 Post-MT Term Swapper: Supplementing a Statistical Machine Translation System with a User Dictionary
Masaki Itagaki and Takako Aikawa
- 15:40-16:00 Using Parsed Corpora for Estimating Stochastic Inversion Transduction Grammars
Germán Sanchis and Joan Andreu Sánchez
- 16:00-16:20 Experiments on Processing Overlapping Parallel Corpora
Mark Fishel and Heiki-Jaan Kaalep

O25 - Evaluation

Chairperson: Robert Frederking

Fez 1

- 14:40-15:00 Parser Evaluation and the BNC: Evaluating 4 constituency parsers with 3 metrics
Jennifer Foster and Josef van Genabith
- 15:00-15:20 EASY, Evaluation of Parsers of French: what are the Results?
Patrick Paroubek, Isabelle Robba, Anne Vilnat and Christelle Ayache
- 15:20-15:40 Evaluation Metrics for Automatic Temporal Annotation of Texts
Xavier Tannier and Philippe Muller
- 15:40-16:00 A Comparative Study on Language Identification Methods
Lena Grothe, Ernesto William De Luca and Andreas Nürnberg
- 16:00-16:20 PASSAGE: from French Parser Evaluation to Large Sized Treebank
Eric Villemonte de la Clergerie, Olivier Hamon, Djamel Mostefa, Christelle Ayache, Patrick Paroubek and Anne Vilnat

O26 - Broadcast News Processing

Chairperson: Florian Schiel

Fez 2

- 14:40-15:00 Structural Metadata Annotation of Speech Corpora: Comparing Broadcast News and Broadcast Conversations
Jáchym Kolář and Jan Švec
- 15:00-15:20 Thai Broadcast News Corpus Construction and Evaluation
Markpong Jongtaveesataporn, Chai Wutiwivatchai, Koji Iwano and Sadaoki Furui
- 15:20-15:40 RUNDKAST: an Annotated Norwegian Broadcast News Speech Corpus
Ingunn Amdal, Ole Morten Strand, Jørn Almberg and Torbjørn Svendsen
- 15:40-16:00 First Broadcast News Transcription System for Khmer Language
Sopheap Seng, Sethserey Sam, Laurent Besacier, Brigitte Bigi and Eric Castelli
- 16:00-16:20 Quick Rich Transcriptions of Arabic Broadcast News Speech Data
Chomicha Bendahman, Meghan Glenn, Djamel Mostefa, Niklas Paulsson and Stephanie Strassel
- 14:40-16:20 **Poster Sessions Asni Room**

P15 - LR Infrastructures and Architectures

Chairperson: Mike Rosner

RACAI's Linguistic Web Services

Dan Tufiş, Radu Ion, Alexandru Ceaşu and Dan Ştefănescu

Words in Contexts: Digital Editions of Literary Journals in the “AAC - Austrian Academy Corpus”

Hanno Biber, Evelyn Breiteneder and Karlheinz Mörh

ASV Toolbox: a Modular Collection of Language Exploration Tools

Chris Biemann, Uwe Quasthoff, Gerhard Heyer and Florian Holz

LX-Service: Web Services of Language Technology for Portuguese

António Branco, Francisco Costa, Pedro Martins, Filipe Nunes, João Silva and Sara Silveira

The TextPro Tool Suite

Emanuele Pianta, Christian Girardi and Roberto Zanoli

An AI-inspired intelligent agent/student architecture to combine Language Resources research and teaching

Bayan Abu Shawar and Eric Atwell

Language Resources and Tools for Swedish: A Survey

Kjell Elenius, Eva Forsbom and Beáta Megyesi

Glossa: a Multilingual, Multimodal, Configurable User Interface

Lars Nygaard, Joel Priestley, Anders Nøklestad and Janne Bondi Johannessen

Ontology-Based Interface Specifications for a NLP Pipeline Architecture

Ekaterina Buyko, Christian Chiarcos and Antonio Pareja-Lora

Foundation of a Component-based Flexible Registry for Language Resources and Technology

Daan Broeder, Thierry Declerck, Erhard Hinrichs, Stelios Piperidis, Laurent Romary, Nicoletta Calzolari and Peter Wittenburg

Building a Federation of Language Resource Repositories: the DAM-LR Project and its Continuation within CLARIN.

Daan Broeder, David Nathan, Sven Strömqvist and Remco van Veenendaal

A Grid of Regional Language Archives

Paul Trilsbeek, Daan Broeder, Tobias Valkenhoef and Peter Wittenburg

Adapting International Standard for Asian Language Technologies

Tokunaga Takenobu, Dain Kaplan, Chu-Ren Huang, Shu-Kai Hsieh, Nicoletta Calzolari, Monica Monachini, Claudia Soria, Shirai Kiyooki, Virach Sornlertlamvanich, Thatsanee Charoenporn and Xia YingJu

A Large-Scale Web Data Collection as a Natural Language Processing Infrastructure

Keiji Shinzato, Daisuke Kawahara, Chikara Hashimoto and Sadao Kurohashi

UFRA: a UIMA-based Approach to Federated Language Resource Architecture

Riccardo Del Gratta, Roberto Bartolini, Tommaso Caselli, Monica Monachini, Claudia Soria and Nicoletta Calzolari

The Metadata-Database of a Next Generation Sustainability Web-Platform for Language Resources

Georg Rehm, Oliver Schonefeld, Andreas Witt, Timm Lehmborg, Christian Chiarcos, Hanan Bechara, Florian Eishold, Kilian Evang, Magdalena Leshitanska, Aleksandar Savkov and Matthias Stark

From Field Notes towards a Knowledge Base

Piroska Lendvai and Steve Hunt

Construction of a Metadata Database for Efficient Development and Use of Language Resources

Hitomi Tohyama, Shunsuke Kozawa, Kiyotaka Uchimoto, Shigeki Matsubara and Hitoshi Isahara

A Taxonomy of Lexical Metadata Categories

Bodil Nistrup Madsen and Hanne Erdman Thomsen

P16 - LR National/International Projects, Organizational/Policy Issues

Chairperson: Tamas Varadi

The 2008 Oriental COCODA Book Project: in Commemoration of the First Decade of Sustained Activities in Asia

Shuichi Itahashi and Chiu-yu Tseng

Towards the National Corpus of Polish

Adam Przepiórkowski, Rafał L. Górski, Barbara Lewandowska-Tomaszyk and Marek Łaziński

Strengthening the Estonian Language Technology

Einar Meister and Jaak Vilo

MEDAR: Collaboration between European and Mediterranean Arabic Partners to Support the Development of Language Technology for Arabic

Bente Maegaard, Mohammed Atiyya, Khalid Choukri, Steven Krauwer, Chafic Mokbel and Mustafa Yaseen

Slovene Terminology Web Portal and the TBX-Compatible Simplified DTD/schema

Simon Krek, Vojko Gorjanc and Špela Arhar

P17 - Standards and Best Practices for LRs

Chairperson: Gil Francopoulo

LIRICS Semantic Role Annotation: Design and Evaluation of a Set of Data Categories

Volha Petukhova and Harry Bunt

Reusable Tagset Conversion Using Tagset Drivers

Daniel Zeman

Presentation of the New ISO-Standard for the Representation of Entries in Dictionaries: ISO 1951

Marie-Jeanne Derouin and André Le Meur

ISOcat: Corraling Data Categories in the Wild

Marc Kemps-Snijders, Menzo Windhouwer, Peter Wittenburg and Sue Ellen Wright

Standardising Bilingual Lexical Resources According to the Lexicon Markup Framework

Isa Maks, Carole Tiberius and Remco van Veenendaal

A Framework for Standardized Syntactic Annotation

Thierry Declerck

A Guide for the Production of Reusable Language Resources

Victoria Arranz, Franck Gandcher, Valérie Mapelli and Khalid Choukri

16:20-16:40 Coffee break

16:40-18:20 **4 Oral Sessions and 2 Poster Sessions in Parallel**

O27 - Ontologies

Chairperson: German Rigau

Ministres

16:40-17:00 A General Methodology for Mapping EuroWordNets to the Suggested Upper Merged Ontology
Dennis Spohr

17:00-17:20 Extended Named Entity Ontology with Attribute Information
Satoshi Sekine

17:20-17:40 Towards a Glossary of Activities in the Ontology Engineering Field
Mari Carmen Suárez-Figueroa and Asunción Gómez-Pérez

17:40-18:00 Chinese Core Ontology Construction from a Bilingual Term Bank
Yirong Chen, Qin Lu, Wenjie Li and Gaoying Cui

18:00-18:20 The European Thesaurus on International Relations and Area Studies - a Multilingual Resource for Indexing, Retrieval, and Translation
Michael Kluck and Axel Huckstorf

O28 - Machine Translation and Multilinguality

Chairperson: Dan Tufis

Ambassadeurs

16:40-17:00 Building Bilingual Lexicons using Lexical Translation Probabilities via Pivot Languages
Takashi Tsunakawa, Naoaki Okazaki and Jun'ichi Tsujii

17:00-17:20 Improving Statistical Machine Translation Efficiency by Triangulation
Yu Chen, Andreas Eisele and Martin Kay

17:20-17:40 Phrase-Based Machine Translation based on Simulated Annealing
Caroline Lavecchia, David Langlois and Kamel Smaili

17:40-18:00 Evaluation of Context-Dependent Phrasal Translation Lexicons for Statistical Machine Translation
Marine Carpuat and Dekai Wu

18:00-18:20 A Multi-Genre SMT System for Arabic to French
Saša Hasan and Hermann Ney

O29 - Information Extraction and Question Answering

Chairperson: Walter Daelemans

Fez 1

16:40-17:00 Investigating the Structure of Procedural Texts for Answering How-to Questions
Estelle Delpéch and Patrick Saint-Dizier

17:00-17:20 Analysis and Performance of Morphological Query Expansion and Language-Filtering Words on Basque Web Searching
Igor Leturia, Antton Gurrutxaga, Nerea Areta and Eli Pociello

17:20-17:40 Scaling Answer Type Detection to Large Hierarchies
Kirk Roberts and Andrew Hickl

17:40-18:00 Answering List Questions using Co-occurrence and Clustering
Majid Razmara and Leila Kosseim

18:00-18:20 Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary
Torsten Zesch, Christof Müller and Iryna Gurevych

O30 - Evaluation in Speech Processing

Chairperson: Edouard Geoffrois

Fez 2

16:40-17:00 Odds of Successful Transfer of Low-Level Concepts: a Key Metric for Bidirectional Speech-to-Speech Machine Translation in DARPA's TRANSTAC Program
Gregory Sanders, Sebastien Bronsart, Sherri Condon and Craig Schlenoff

17:00-17:20 Question Answering on Speech Transcriptions: the QAST evaluation in CLEF
Lori Lamel, Sophie Rosset, Christelle Ayache, Djamel Mostefa, Jordi Turmo and Pere Comas

17:20-17:40 Evaluation of Spoken Document Retrieval for Historic Speech Collections
Willemijn Heeren, Franciska de Jong, Laurens van der Werff, Marijn Huijbregts and Roeland Ordelman

17:40-18:00 Applying Automatic Metrics to Speech Translation Dialogs
Sherri Condon, Jon Phillips, Christy Doran, John Aberdeen, Dan Parvaz, Beatrice Oshika, Greg Sanders and Craig Schlenoff

18:00-18:20 A Three-stage Disfluency Classifier for Multi-Party Dialogues
Margot Mieskes and Michael Strube

16:40-18:20 **Poster Sessions Menara**

P18 - Lexical Resources and Tools

Chairperson: Francis Bond

Prolexbase: a Multilingual Relational Lexical Database of Proper Names
Denis Maurel

Ontologizing Lexicon Access Functions based on an LMF-based Lexicon Taxonomy
Yoshihiko Hayashi, Chiharu Narawa, Monica Monachini, Claudia Soria and Nicoletta Calzolari

Romanian Lexical Data Bases: Inflected and Syllabic Forms Dictionaries
Ana-Maria Barbu

Producing an Encyclopedic Dictionary using Patent Documents
Atsushi Fujii

Evaluating the Relationship between Linguistic and Geographic Distances using a 3D Visualization
Folkert De Vriend, Jan Pieter Kunst, Louis Ten Bosch, Charlotte Giesbers and Roeland Van Hout

Enhancing an English-Polish Electronic Dictionary for Multiword Expression Research
Piotr Bański and Radosław Moszczyński

ProPOSEL: A Prosody and POS English Lexicon for Language Engineering
Claire Brierley and Eric Atwell

Creating Glossaries Using Pattern-Based and Machine Learning Techniques
Eline Westerhout and Paola Monachesi

Using Similarity Measures to Extend the LinGO Lexicon
Lynne Cahill

Acquiring a Poor Man's Inflectional Lexicon for German
Peter Adolphs

COLDIC, a Lexicographic Platform for LMF compliant lexica
Núria Bel, Sergio Espeja, Montserrat Marimon and Marta Villegas

P19 - Morphology, Syntax and Tools

Chairperson: Patrick Paroubek

The Annotation Guidelines of the Latin Dependency Treebank and Index Thomisticus Treebank: the Treatment of some specific Syntactic Constructions in Latin

David Bamman, Marco Passarotti, Roberto Busa and Gregory Crane

Unsupervised Lexical Acquisition for Part of Speech Tagging
Dan Tufiş, Elena Irimia, Radu Ion and Alexandru Ceaşu

A Hybrid Approach to Extracting and Classifying Verb+Noun Constructions
Amalia Todiraşcu, Dan Tufiş, Ulrich Heid, Christopher Gledhill, Dan Ştefanescu, Marion Weller and François Rousselot

Head or Non-head? Semi-automatic Procedures for Extracting and Classifying Subcategorisation Properties of Compounds.
Ekaterina Lapshinova-Koltunski and Ulrich Heid

A LAF/GrAF based Encoding Scheme for underspecified Representations of syntactic Annotations.
Manuel Kountz, Ulrich Heid and Kerstin Eckart

The JOS Morphosyntactically Tagged Corpus of Slovene
Tomaž Erjavec and Simon Krek

Demo: An Open Source Tool for Partial Parsing and Morphosyntactic Disambiguation
Aleksander Buczyński and Adam Przepiórkowski

Annotating Superlatives
Silke Scheible

POS Tagging for German: how important is the Right Context?
Steliana Ivanova and Sandra Kuebler

UnsuParse: unsupervised Parsing with unsupervised Part of Speech Tagging
Christian Hähnig, Stefan Bordag and Uwe Quasthoff

Enriching the Venice Italian Treebank with Dependency and Grammatical Relations
Sara Tonelli, Rodolfo Delmonte and Antonella Bristot

Rule-Based Chunker for Croatian
Kristina Vuckovic, Marko Tadic and Zdravko Dovedan

Learning Properties of Noun Phrases: from Data to Functions
Valeria Quochi and Basilio Calderone

A Study of Parentheticals in Discourse Corpora - Implications for NLG Systems
Eva Banik and Alan Lee

Enhancing the Arabic Treebank: a Collaborative Effort toward New Annotation Guidelines
Mohamed Maamouri, Ann Bies and Seth Kulick

A Pilot Arabic Propbank
Martha Palmer, Olga Babko-Malaya, Ann Bies, Mona Diab, Mohamed Maamouri, Aous Mansouri and Wajdi Zaghouani

18:20-18:25 *5-minute Break*

18:25-19:45 **1 Panel, 3 Oral Sessions and 1 Poster Session in Parallel**

18:25-19:45 **Panel: Key Issues in Building a Common Language Resources and Tools Infrastructure**

Moderator: Erhard Hinrichs

Ministres

O31 - Evaluation and Machine Translation

Chairperson: Malek Boualem

Ambassadeurs

- 18:25-18:45 Towards Heterogeneous Automatic MT Error Analysis
Jesús Giménez and Lluís Màrquez
- 18:45-19:05 Sensitivity of Automated MT Evaluation Metrics on Higher Quality MT Output: BLEU vs Task-Based Evaluation Methods
Bogdan Babych and Anthony Hartley
- 19:05-19:25 Translation Adequacy and Preference Evaluation Tool (TAP-ET)
Mark Przybocki, Kay Peterson and Sebastien Bronsart
- 19:25-19:45 Evaluation of a Cross-lingual Romanian-English Multi-document Summariser
Constantin Orasan and Oana Andreea Chiorean

O32 - Syntactically Annotated Corpora

Chairperson: Philippe Blache

Fez 1

- 18:25-18:45 The BNC Parsed with RASP4UIMA
Øistein E. Andersen, Julien Nioche, Ted Briscoe and John Carroll
- 18:45-19:05 Word-level Dependency-structure Annotation to Corpus of Spontaneous Japanese and its Application
Kiyotaka Uchimoto and Yasuharu Den
- 19:05-19:25 Induction of Treebank-Aligned Lexical Resources
Tejaswini Deoskar and Mats Rooth
- 19:25-19:45 A Unified Database of Dependency Treebanks: Integrating, Quantifying & Evaluating Dependency Data
Olga Pustyl'nikov, Alexander Mehler and Rüdiger Gleim

O33 - Terminology

Chairperson: Christophe Jouis

Fez 2

- 18:25-18:45 Amazigh Language Terminology in Morocco or Management of a "Multidimensional" Variation
Aicha Bouhjar
- 18:45-19:05 Chinese Term Extraction Based on Delimiters
Yuhang Yang, Qin Lu and Tiejun Zhao
- 19:05-19:25 A Multi-Word Term Extraction Program for Arabic Language
Siham Boulaknadel, Beatrice Daille and Driss Aboutajdine
- 19:25-19:45 Using Similarity Metrics for Terminology Recognition
Jonathan Butters and Fabio Ciravegna
- 18:25-19:45 **Poster Sessions Asni Room**

P20 - Multimodal, Multimedia and Subjective Corpus

Chairperson: Doroteo Toledano

Saxon: an Extensible Multimedia Annotator

Mark Greenwood, José Iria and Fabio Ciravegna

Spatiotemporal Coding in ANVIL

Michael Kipp

Integrating Audio and Visual Information for Modelling Communicative Behaviours Perceived as Different

Michelina Savino, Laura Scivetti and Mario Refice

Annotation Tool Development for Large-Scale Corpus Creation Projects at the Linguistic Data Consortium

Kazuaki Maeda, Haejoong Lee, Shawn Medero, Julie Medero, Robert Parker and Stephanie Strassel

Design and Recording of Czech Audio-Visual Database with Impaired Conditions for Continuous Speech Recognition

Jana Trojanová, Marek Hruš, Pavel Campr and Milos Zelezny

The UJIPenchars Database: a Pen-Based Database of Isolated Handwritten Characters

D. Llorens, F. Prat, A. Marzal, J. M. Vilar, M. J. Castro, J. C. Amengual, S. Barrachina, A. Castellanos, S. España, J. A. Gómez, J. Gorbe, A. Gordo, V. Palazón, G. Peris, R. Ramos-Garjón and F. Zamora

Sign Language Corpus Annotation: toward a new Methodology

Emilie Chételat-Pelé and Annelies Braffort

Benchmark Databases for Video-Based Automatic Sign Language Recognition

Philippe Dreuw, Carol Neidle, Vassilis Athitsos, Stan Sclaroff and Hermann Ney

The ATIS Sign Language Corpus
Jan Bungeroth, Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, Andy Way and Lynette van Zijl

Collection and Preprocessing of Czech Sign Language Corpus for Sign Language Recognition
Pavel Campr, Marek Hrzík and Jana Trojanová

A Multimodal Infant Behavior Annotation for Developmental Analysis of Demonstrative Expressions
Shigeyoshi Kitazawa, Shinya Kiriya, Tomohiko Kasami, Shogo Ishikawa, Naofumi Otani, Hiroaki Horiuchi and Yoichi Takebayashi

Automatic Emotional Degree Labeling for Speakers' Anger Utterance during Natural Japanese Dialog
Yoshiko Arimoto, Sumio Ohno and Hitoshi Iida

A Real-World Emotional Speech Corpus for Modern Greek
Theodoros Kostoulas, Todor Ganchev, Iosif Mporas and Nikos Fakotakis

Annotating Subjective Content in Meetings
Theresa Wilson

P21 - Tools and Data for Speech Systems Development

Chairperson: Ryszard Gubrynowicz

The AUTONOMATA Spoken Names Corpus
Henk van den Heuvel, Jean-Pierre Martens, Bart D'hoore, Kristof D'hanens and Nanneke Konings

Acquiring Pronunciation Data for a Placenames Lexicon in a Less-Resourced Language
Briony Williams and Rhys James Jones

Constructing a Database of Non-Japanese Pronunciations of Different Japanese Romanizations
Reiko Kaji and Hajime Mochizuki

Combined Systems for Automatic Phonetic Transcription of Proper Nouns
Antoine Laurent, Téva Merlin, Sylvain Meignier, Yannick Estève and Paul Deléglise

Evaluation of Modules and Tools for Speech Synthesis: the ECESS Framework
Harald Höge, Zdravko Kacic, Bojan Kotnik, Matej Rojc, Nicolas Moreau and Horst-Udo Hain

An Automatic Close Copy Speech Synthesis Tool for Large-Scale Speech Corpus Evaluation
Dafydd Gibbon and Jolanta Bachan

A Flexible Wizard of Oz Environment for Rapid Prototyping
Stefan Scherer and Petra-Maria Strauß

Building of a Speech Corpus Optimised for Unit Selection TTS Synthesis
Jindřich Matoušek, Daniel Tihelka and Jan Romportl

Methodologies for Designing and Recording Speech Databases for Corpus Based Synthesis
Luis Oliveira, Sérgio Paulo, Luís Figueira, Carlos Mendes, Ana Nunes and Joaquim Godinho

MISTRAL: a Statistical Machine Translation Decoder for Speech Recognition Lattices
Alexandre Patry and Philippe Langlais

LC-STAR II: Starring more Lexica
Ute Ziegenhain, Hanne Fersoe, Henk van den Heuvel and Asuncion Moreno

Communicating Unknown Words in Machine Translation
Matthias Eck, Stephan Vogel and Alex Waibel

Developing Non-European Translation Pairs in a Medium-Vocabulary Medical Speech Translation System
Pierrette Bouillon, Sonia Halimi, Yukie Nakao, Kyoko Kanzaki, Hitoshi Isahara, Nikos Tsourakis, Marianne Starlander, Beth Ann Hockey and Manny Rayner

CLIoS: Cross-lingual Induction of Speech Recognition Grammars
Nadine Perera, Michael Pitz and Manfred Pinkal

Construction and Analysis of Word-level Time-aligned Simultaneous Interpretation Corpus
Takahiro Ono, Hitomi Tohyama and Shigeki Matsubara

Semantic Frame Annotation on the French MEDIA corpus
Marie-Jean Meurs, Frédéric Duvert, Frédéric Bechet, Fabrice Lefevre and Renato De Mori

Cross-Domain Dialogue Act Tagging
Nick Webb, Ting Liu, Mark Hepple and Yorick Wilks

Building Mobile Spoken Dialogue Applications Using Regulus
Nikos Tsourakis, Maria Georgescu, Pierrette Bouillon and Manny Rayner

Active Annotation in the LUNA Italian Corpus of Spontaneous Dialogues
Christian Raymond, Kepa Joseba Rodriguez and Giuseppe Riccardi

A Comparison of Various Methods for Concept Tagging for Spoken Language Understanding
Stefan Hahn, Patrick Lehnen, Christian Raymond and Hermann Ney

Morphosyntactic Resources for Automatic Speech Recognition
Stéphane Huet, Guillaume Gravier and Pascale Sébillot