
Detailed Conference Programme

28 May 2014

9:30-10:30

Opening Ceremony in Eldborg

10:30-11:15

Introductory Session

Resources and Evaluation at the confluence of Data and Language

Chairperson: Joseph Mariani

- Language Resources in the Data Economy, *Márta Nagy-Rothengass*
- A World of Language Resources, *Nicoletta Calzolari and Khalid Choukri*
- LREC 15th Anniversary Celebration, *Joseph Mariani*

Eldborg

11:15-11:35

Coffee break

O1 • NLP Workflows

Chairperson: Stephanie Strassel

Nordurljos

11:35-11:55

The Alveo Virtual Laboratory: a Web based Repository API

Steve Cassidy, Dominique Estival, Timothy Jones, Denis Burnham and Jared Burghold

11:55-12:15

A Stream Computing Approach Towards Scalable NLP

Xabier Artola, Zuhaitz Beloki and Aitor Soroa

12:15-12:35

ILLINOISCLOUDNLP: Text Analytics Services in the Cloud

Hao Wu, Zhiye Fei, Aaron Dai, Mark Sammons, Dan Roth and Stephen Mayhew

12:35-12:55

The Language Application Grid

Nancy Ide, James Pustejovsky, Christopher Cieri, Eric Nyberg, Di Wang, Keith Suderman, Marc Verhagen and Jonathan Wright

12:55-13:15

Praaline: Integrating Tools for Speech Corpus Research

George Christodoulides

O2 • Machine Translation and Evaluation (1)

Chairperson: Bente Maegaard

Silfurberg A

11:35-11:55

Linguistic Evaluation of Support Verb Constructions by OpenLogos and Google Translate

Anabela Barreiro, Johanna Monti, Brigitte Orliac, Susanne Preuß, Kutz Arrieta, Wang Ling, Fernando Batista and Isabel Trancoso

11:55-12:15

MTWatch: a Tool for the Analysis of Noisy Parallel Data

Sandipan Dandapat and Declan Groves

12:15-12:35

Machine Translation for Subtitling: a Large-Scale Evaluation

Thierry Etchegoyhen, Lindsay Bywood, Mark Fishel, Panayota Georgakopoulou, Jie Jiang, Gerard van Loenhout, Arantza del Pozo, Mirjam Sepesy Maucec, Anja Turner and Martin Volk

- 12:35-12:55 Machine Translationness: Machine-Likeness in Machine Translation Evaluation
Joaquim Moré and Salvador Climent
- 12:55-13:15 On the Origin of Errors: a Fine-Grained Analysis of MT and PE Errors and their Relationship
Joke Daems, Lieve Macken and Sonia Vandepitte

O3 • Grammar and Parsing (1)

Chairperson: Emily M. Bender

Silfurberg B

- 11:35-11:55 Parsing Chinese Synthetic Words with a Character-based Dependency Model
Fei Cheng, Kevin Duh and Yuji Matsumoto
- 11:55-12:15 Improvements to Dependency Parsing using Automatic Simplification of Data
Tomáš Jelinek
- 12:15-12:35 All Fragments Count in Parser Evaluation
Joost Bastings and Khalil Sima'an
- 12:35-12:55 Less is More? Towards a Reduced Inventory of Categories for Training a Parser for the Italian Stanford Dependencies
Maria Simi, Cristina Bosco and Simonetta Montemagni
- 12:55-13:15 Rapid Deployment of Phrase Structure Parsing for Related Languages: a Case Study of Insular Scandinavian
Anton Karl Ingason, Hrafn Loftsson, Eiríkur Rögnvaldsson, Einar Freyr Sigurðsson and Joel C. Wallenberg

O4 • Information Extraction and Knowledge Discovery

Chairperson: Ricardo Baeza-Yates

Kaldalon

- 11:35-11:55 Boosting Open Information Extraction with Noun-based Relations
Clarissa Xavier and Vera Lima
- 11:55-12:15 Clinical Data-driven Probabilistic Graph Processing
Travis Goodwin and Sanda Harabagiu
- 12:15-12:35 A Framework for Compiling High Quality Knowledge Resources from Raw Corpora
Gongye Jin, Daisuke Kawahara and Sadao Kurohashi
- 12:35-12:55 Thematic Cohesion: Measuring Terms Discriminatory Power Toward Themes
Clément de Groc, Xavier Tannier and Claude de Loupy
- 12:55-13:15 Extracting Information for Context-aware Meeting Preparation
Simon Scerri, Behrang Q. Zadeh, Maciej Dabrowski and Ismael Rivera

11:35 - 13:15

Poster Sessions - Poster Area (1)

P1 • Corpora and Annotation

Chairperson: Marko Tadić

TaLAPI - a Thai Linguistically Annotated Corpus for Language Processing
AiTi Aw, Sharifah Mahani Aljunied, Nattadaporn Lertcheva and Sasiwimon Kalunsima

Variations on Quantitative Comparability Measures and their Evaluations on Synthetic French-English Comparable Corpora

Guiyao Ke, Pierre-Francois Marteau and Gildas Menier

Using Transfer Learning to Assist Exploratory Corpus Annotation

Paul Felt, Eric Ringger, Kevin Seppi and Kristian Heal

Priberam Compressive Summarization Corpus: a New Multi-Document Summarization Corpus for European Portuguese

Miguel B. Almeida, Mariana S. C. Almeida, André F. T. Martins, Helena Figueira, Pedro Mendes and Cláudia Pinto

Corpus and Evaluation of Handwriting Recognition of Historical Genealogical Records

Patrick Schone, Heath Nielson and Mark Ward

The SYN-series Corpora of Written Czech

Milena Hnátková, Michal Křen, Pavel Procházka and Hana Skoumalová

Corpus of 19th-Century Czech Texts: Problems and Solutions

Karel Kučera and Martin Stluka

Extending Standoff Annotation

Maik Stührenberg

Constructing and Exploiting an Automatically Annotated Resource of Legislative Texts

Stefan Höfler and Kyoko Sugisaki

P2 • Crowdsourcing

Chairperson: Alain Couillault

A Study on Expert Sourcing Enterprise Question Collection and Classification

Yuan Luo, Thomas Boucher, Tolga Oral, David Osofsky and Sara Weber

Can the Crowd Be Controlled?: a Case Study on Crowd Sourcing and Automatic Validation of Completed Tasks Based on User Modeling

Balamurali A.R

When Transliteration Met Crowdsourcing: an Empirical Study of Transliteration Via Crowdsourcing using Efficient, non-Redundant and Fair Quality Control

Mitesh M. Khapra, Ananthakrishnan Ramanathan, Anoop Kunchukuttan, Karthik Visweswariah and Pushpak Bhattacharyya

Design and Development of an Online Computational Framework to Facilitate Language Comprehension Research on Indian Languages

Manjira Sinha, Tirthankar Dasgupta and Anupam Basu

Collaboration in the Production of a Massively Multilingual Lexicon

Martin Benjamin

A SICK Cure for the Evaluation of Compositional Distributional Semantic Models

Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi and Roberto Zamparelli

Can Crowdsourcing Be Used for Effective Annotation of Arabic?

Wajdi Zaghouni and Kais Dukes

Crowdsourcing as a Preprocessing for Complex Semantic Annotation Tasks

Héctor Martínez Alonso and Lauren Romeo

Online Experiments with the Percy Software Framework - Experiences and some Early Results

Christoph Draxler

P3 • Dialogue

Chairperson: Dan Cristea

First Insight into Quality-Adaptive Dialogue

Stefan Ultes, Hüseyin Dikme and Wolfgang Minker

The DBOX Corpus Collection of Spoken Human-Human and Human-Machine Dialogues

Volha Petukhova, Martin Gropp, Dietrich Klakow, Gregor Eigner, Mario Topf, Stefan Srb, Petr Motlicek, Blaise Potard, John Dines, Olivier Deroo, Ronny Egeler, Uwe Meinz, Steffen Liersch and Anna Schmidt

Modeling and Evaluating Dialog Success in the LAST MINUTE Corpus

Dietmar Rösner, Rafael Friesen, Stephan Günther and Rico Andrich

NASTIA: Negotiating Appointment Setting Interface

Layla El Asri, Rémi Lemonnier, Romain Laroche, Olivier Pietquin and Hatim Khouzaimi

DINASTI: Dialogues with a Negotiating Appointment Setting Interface

Layla El Asri, Romain Laroche and Olivier Pietquin

EL-WOZ: a Client-Server Wizard-of-Oz Interface

Thomas Pellegrini, Vahid Hedayati and Angela Costa

P4 • Phonetic Databases and Prosody

Chairperson: Philippe Martin

Tools for Arabic Natural Language Processing: a Case Study in Qalqalah Prosody

Claire Brierley, Majdi Sawalha and Eric Atwell

A Benchmark Database of Phonetic Alignments in Historical Linguistics and Dialectology

Johann-Mattis List and Jelena Prokić

Rhapsodie: a Prosodic-Syntactic Treebank for Spoken French

Anne Lacheret, Sylvain Kahane, Julie Beliao, Anne Dister, Kim Gerdes, Jean-Philippe Goldman, Nicolas Obin, Paola Pietrandrea and Atanas Tchobanov

C-PhonoGenre: a 7-hour Corpus of 7 Speaking Styles in French: Relations between Situational Features and Prosodic Properties

Jean-Philippe Goldman, Tea Prsirr and Antoine Auchlin

A Corpus and Phonetic Dictionary for Tunisian Arabic Speech Recognition

Abir Masmoudi, Mariem Ellouze Khmekhem, Yannick Esteve, Lamia Hadrich Belguith and Nizar Habash

Towards Automatic Transformation Between Different Transcription Conventions: Prediction of Intonation Markers from Linguistic and Acoustic Features

Yuichi Ishimoto, Tomoyuki Tsuchiya, Hanae Koiso and Yasuharu Den

RSS-TOBI - a Prosodically Enhanced Romanian Speech Corpus

Tiberiu Boroş, Adriana Stan, Oliver Watts and Stefan Daniel Dumitrescu

Segmentation Evaluation Metrics, a Comparison Grounded on Prosodic and Discourse Units

Klim Peshkov and Laurent Prévot

A Cross-Language Corpus for Studying the Phonetics and Phonology of Prominence
Bistra Andreeva, William Barry and Jacques Koreman

Using a Machine Learning Model to Assess the Complexity of Stress Systems
Liviu Dinu, Alina Maria Ciobanu, Ioana Chitoran and Vlad Niculae

GlobalPhone: Pronunciation Dictionaries in 20 Languages
Tanja Schultz and Tim Schlippe

P5 • Speech Resources

Chairperson: Martine Adda-Decker

New Spanish Speech Corpus Database for the Analysis of People Suffering from Parkinson's Disease

Juan Rafael Orozco-Arroyave, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, María Claudia Gonzalez-Rátiva and Elmar Nöth

An Effortless Way to Create Large-Scale Datasets for Famous Speakers
François Salmon and Félicien Vallet

German Alcohol Language Corpus - the Question of Dialect
Florian Schiel and Thomas Kislser

Vulnerability in Acquisition, Language Impairments in Dutch: Creating a VALID Data Archive

Jetske Klatter, Roeland Van Hout, Henk van den Heuvel, Paula Fikkert, Anne Baker, Jan De Jong, Frank Wijnen, Eric Sanders and Paul Trilsbeek

The Nijmegen Corpus of Casual Czech

Mirjam Ernestus, Lucie Kočková-Amortová and Petr Pollak

CIEMPIESS: a New Open-Sourced Mexican Spanish Radio Corpus

Carlos Daniel Hernandez Mena and Abel Herrera Camacho

Mapping Diatopic and Diachronic Variation in Spoken Czech: the Ortofon and Dialekt Corpora

Marie Kopřivová, Hana Goláňová, Petra Klimešová and David Lukeš

The Research and Teaching Corpus of Spoken German - FOLK

Thomas Schmidt

Free Acoustic and Language Models for Large Vocabulary Continuous Speech Recognition in Swedish

Niklas Vanhainen and Giampiero Salvi

13:15-14:45

Lunch break

O5 • Linked Data (Special Session)

Chairperson: Asunción Gómez-Pérez

Nordurljos

14:45-15:05 Metadata as Linked Open Data: Mapping Disparate XML Metadata Registries into one RdDF/OWL Registry.

Marta Villegas, Maite Melero and Núria Bel

15:05-15:25 Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0

Maud Ehrmann, Francesco Cecconi, Daniele Vannella, John Philip McCrae, Philipp Cimiano and Roberto Navigli

- 15:25-15:45 Enabling Language Resources to Expose Translations as Linked Data on the Web
Jorge Gracia, Elena Montiel-Ponsoda, Daniel Vila-Suero and Guadalupe Aguado-de-Cea
- 15:45-16:05 A SKOS-based Schema for TEI-encoded Dictionaries at ICLTT
Thierry Declerck, Karlheinz Mörth and Eveline Wandl-Vogt
- 16:05-16:25 Discussion

O6 • Audiovisual

Chairperson: Gilles Adda

Silfurberg A

- 14:45-15:05 TVD: a Reproducible and Multiply Aligned TV Series Dataset
Anindya Roy, Camille Guinaudeau, Herve Bredin and Claude Barras
- 15:05-15:25 Image Annotation with ISO-Space: Distinguishing Content from Structure
James Pustejovsky and Zachary Yocum
- 15:25-15:45 SAVAS: Collecting, Annotating and Sharing Audiovisual Language Resources for Automatic Subtitling
Arantza del Pozo, Carlo Aliprandi, Aitor Alvarez, Carlos Mendes, Joao P. Neto, Sérgio Paulo, Nicola Piccinini and Matteo Raffaelli
- 15:45-16:05 Phoneme Similarity Matrices to Improve Long Audio Alignment for Automatic Subtitling
Pablo Ruiz, Aitor Álvarez and Haritz Arzelus
- 16:05-16:25 KALAKA-3: a Database for the Recognition of Spoken European Languages on YouTube Audios
Luis Javier Rodríguez-Fuentes, Mikel Penagarikano, Amparo Varona, Mireia Diez and German Bordel

O7 • Processing of Social Media

Chairperson: Paul Rayson

Silfurberg B

- 14:45-15:05 Named Entity Recognition on Turkish Tweets
Dilek Kucuk, Guillaume Jacquet and Ralf Steinberger
- 15:05-15:25 Comprehensive Annotation of MultiWord Expressions in a Social Web Corpus
Nathan Schneider, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad and Noah A. Smith
- 15:25-15:45 Re-using an Argument Corpus to Aid in the Curation of Social Media Collections
Clare Llewellyn, Claire Grover, Jon Oberlander and Ewan Klein
- 15:45-16:05 #mygoal: Finding Motivations on Twitter
Marc Tomlinson, David Bracewell, Wayne Krug and David Hinote
- 16:05-16:25 A Framework for Public Health Surveillance
Andrew Yates, Jon Parker, Nazli Goharian and Ophir Frieder

O8 • Acquisition

Chairperson: Xavier Tannier

Kaldalon

- 14:45-15:05 Bootstrapping Term Extractors for Multiple Languages
Ahmet Aker, Monica Paramita, Emma Barker and Robert Gaizauskas
- 15:05-15:25 Evaluation of Automatic Hypernym Extraction from Technical Corpora in English and Dutch
Els Lefever, Marjan Van de Kauter and Véronique Hoste
- 15:25-15:45 Resources for the Detection of Conventionalized Metaphors in Four Languages
Lori Levin, Teruko Mitamura, Brian MacWhinney, Davida Fromm, Jaime Carbonell, Weston Feely, Robert Frederking, Anatole Gershman and Carlos Ramirez
- 15:45-16:05 A Language-independent and fully Unsupervised Approach to Lexicon Induction and Part-of-Speech Tagging for Closely Related Languages
Yves Scherrer and Benoît Sagot
- 16:05-16:25 Optimizing a Distributional Semantic Model for the Prediction of German Particle Verb Compositionality
Stefan Bott and Sabine Schulte im Walde

14:45 - 16:25

Poster Sessions - Poster Area (2)

P6 • Endangered Languages

Chairperson: Laurette Pretorius

Open-Domain Interaction and Online Content in the Sami Language

Kristiina Jokinen

The Gulf of Guinea Creole Corpora

Tjerk Hagemeijer, Michel Génèreux, Iris Hendrickx, Amália Mendes, Abigail Tiny and Armando Zamora

Languagesindanger.eu - Including Multimedia Language Resources to disseminate Knowledge and Create Educational Material on less-Resourced Languages

Dagmar Jung, Katarzyna Klessa, Zsuzsa Duray, Beatrix Oszkó, Mária Sipos, Sándor Szeverényi, Zsuzsa Várnai, Trilsbeek Paul and Tamás Váradi

Casa De La Lhéngua: a Set of Language Resources and Natural Language Processing Tools for Mirandese

José Pedro Ferreira, Cristiano Chesi, Daan Baldewijns, Fernando Miguel Pinto, Margarita Correia, Daniela Braga, Hyongsil Cho, Amadeu Ferreira and Miguel Dias

A Finite-State Morphological Analyzer for a Lakota HPSG Grammar

Christian Curtis

P7 • Evaluation Methodologies

Chairperson: Violeta Seretan

Extrinsic Corpus Evaluation with a Collocation Dictionary Task

Adam Kilgarriff, Pavel Rychlý, Milos Jakubicek, Vojtěch Kovář, Vit Baisa and Lucia Kocincová

Evaluating the Effects of Interactivity in a Post-Editing Workbench

Nancy Underwood, Bartolomé Mesa-Lao, Mercedes García Martínez, Michael Carl, Vicent Alabau, Jesús González-Rubio, Luis A. Leiva, Germán Sanchis-Trilles, Daniel Ortíz-Martínez and Francisco Casacuberta

Bridging the Gap between Speech Technology and Natural Language Processing: An Evaluation Toolbox for Term Discovery Systems

Bogdan Ludusan, Maarten Versteegh, Aren Jansen, Guillaume Gravier, Xuan-Nga Cao, Mark Johnson and Emmanuel Dupoux

Introducing a Framework for the Evaluation of Music Detection Tools

Paula Lopez-Otero, Laura Docio-Fernandez and Carmen Garcia-Mateo

Measuring Readability of Polish Texts: Baseline Experiments

Bartosz Broda, Bartłomiej Nitoń, Włodzimierz Gruszczyński and Maciej Ogrodniczuk

Fuzzy V-Measure - an Evaluation Method for Cluster Analyses of Ambiguous Data

Jason Utt, Sylvia Springorum, Maximilian Köper and Sabine Schulte im Walde

Finding a Tradeoff Between Accuracy and Rater's Workload in Grading Clustered Short Answers

Andrea Horbach, Alexis Palmer and Magdalena Wolska

Improving Evaluation of English-Czech MT through Paraphrasing

Petra Barancikova, Rudolf Rosa and Ales Tamchyna

On the Reliability and Inter-Annotator Agreement of Human Semantic MT Evaluation via HMEANT

Chi-kiu Lo and Dekai Wu

P8 • Language Resource Infrastructures

Chairperson: Georg Rehm

The Evolving Infrastructure for Language Resources and the Role for Data Scientists

Nelleke Oostdijk and Henk van den Heuvel

Using TEI, CMDI and ISOcat in CLARIN-DK

Dorte Haltrup Hansen, Lene Offersgaard and Sussi Olsen

ROOTS: a Toolkit for Easy, Fast and Consistent Processing of Large Sequential Annotated Data Collections

Jonathan Chevelu, Gwéólé Lecorvé and Damien Lolive

Sharing Cultural Heritage: the Clavius on the Web Project

Matteo Abrate, Angelo Mario Del Grosso, Emiliano Giovannetti, Angelica Lo Duca, Damiana Luzzi, Lorenzo Mancini, Andrea Marchetti, Irene Pedretti and Silvia Piccini

'interHist' - an Interactive Visual Interface for Corpus Exploration

Verena Lyding, Lionel Nicolas and Egon Stemle

P9 • Machine Translation

Chairperson: Jordi Atserias

Constructing a Chinese-Japanese Parallel Corpus from Wikipedia

Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi

An Iterative Approach for Mining Parallel Sentences in a Comparable Corpus

Lise Rebout and Philippe Langlais

Large SMT Data-Sets Extracted from Wikipedia

Dan Tufiş

Production of Phrase Tables in 11 European Languages using an Improved Sub-Sentential Aligner

Juan Luo and Yves Lepage

Collection of a Simultaneous Translation Corpus for Comparative Analysis
Hiroaki Shimizu, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura

English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling
Sharid Loaigca, Thomas Meyer and Andrei Popescu-Belis

Two-Step Machine Translation with Lattices
Bushra Jawaid and Ondrej Bojar

P10 • Metadata

Chairperson: Victoria Arranz

The CMD Cloud

Matej Durco and Menzo Windhouwer

The e-Identity Exploration Workbench

Fritz Kliche, Andre Blessing, Jonathan Sonntag and Dr. Ulrich Heid

Visualization of Language Relations and Families: MultiTree

Damir Cavar and Malgorzata Cavar

P11 • MultiWord Expressions and Terms

Chairperson: Francesca Frontini

Linked Open Data and Web Corpus Data for Noun Compound Bracketing

Pierre André Ménard and Caroline Barrière

4FX: Light Verb Constructions in a Multilingual Parallel Corpus

Anita Rácz, István Nagy T. and Veronika Vincze

Identifying Idioms in Chinese Translations

Wan Yu Ho, Christine Kng, Shan Wang and Francis Bond

Narrowing the Gap Between Termbases and Corpora in Commercial Environments

Kara Warburton

Identification of Multiword Expressions in the brWAC

Rodrigo Boos, Kassius Prestes and Aline Villavicencio

Collocation or Free Combination? - Applying Machine Translation Techniques to

Identify Collocations in Japanese

Lis Pereira, Elga Strafella and Yuji Matsumoto

Building a Crisis Management Term Resource for Social Media: the Case of Floods and Protests

Irina Temnikova, Andrea Varga and Dogan Biyikli

P12 • Treebanks

Chairperson: Béatrice Daille

Towards Building a Kashmiri Treebank: Setting Up the Annotation Pipeline
Riyaz Ahmad Bhat, Shahid Musjtaq Bhat and Dipti Misra Sharma

A Japanese Word Dependency Corpus
Shinsuke Mori, Hideki Ogura and Tetsuro Sasada

A Compact Interactive Visualization of Dependency Treebank Query Results
Chris Culy, Marco Passarotti and Ulla König-Cardanobile

Thomas Aquinas in the TüNDRA: Integrating the Index Thomisticus Treebank into CLARIN-D

Scott Martens and Marco Passarotti

Boosting the Creation of a Treebank

Blanca Arias, Nuria Bel, Mercè Lorente, Montserrat Marimón, Alba Milà, Jorge Vivaldi, Muntsa Padró, Marina Fomicheva and Imanol Larrea

The IULA Spanish LSP Treebank

Montserrat Marimon, Núria Bel, Beatriz Fisas, Blanca Arias, Silvia Vázquez, Jorge Vivaldi, Carlos Morell and Mercè Lorente

The Norwegian Dependency Treebank

Per Erik Solberg, Arne Skjærholt, Lilja Øvrelid, Kristin Hagen and Janne Bondi Johannessen

A Persian Treebank with Stanford Typed Dependencies

Mojgan Seraji, Carina Jahani, Beáta Megyesi and Joakim Nivre

Converting an HPSG-based Treebank into its Parallel Dependency-based Treebank
Masood Ghayoomi and Jonas Kuhn

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

O9 • Sentiment Analysis and Social Media (1)

Chairperson: Stelios Piperidis

Nordurljos

16:45-17:05 On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter
Hassan Saif, Miriam Fernandez, Yulan HE and Harith Alani

17:05-17:25 Investigating the Image of Entities in Social Media: Dataset Design and First Results
Julien Velcin, Young-Min Kim, Caroline Brun, Jean-Yves Dormagen, Eric SanJuan, Leila Khouas, Anne Peradotto, Stéphane Bonnevey, Claude Roux, Julien Boyadjian, Alejandro Molina and Marie Neihouser

17:25-17:45 Benchmarking Twitter Sentiment Analysis Tools
Ahmed Abbasi, Ammar Hassan and Milan Dhar

17:45-18:05 Recognising Suicidal Messages in Dutch Social Media
Bart Desmet and Véronique Hoste

O10 • Conversational (1)

Chairperson: Nick Campbell

Silfurberg A

- 16:45-17:05 Automatic Detection of Other-Repetition Occurrences: Application to French Conversational Speech
Brigitte Bigi, Roxane Bertrand and Mathilde Guardiola
- 17:05-17:25 ANCOR_Centre, a Large Free Spoken French Coreference Corpus: Description of the Resource and Reliability Measures
Judith Muzerelle, Anaïs Lefevvre, Emmanuel Schang, Jean-Yves Antoine, Aurore Pelletier, Denis Maurel, Iris Eshkol and Jeanne Villaneau
- 17:25-17:45 Phone Boundary Annotation in Conversational Speech
Yi-Fen Liu, Shu-Chuan Tseng and J.-S Roger Jang
- 17:45-18:05 Automatically Enriching Spoken Corpora with Syntactic Information for Linguistic Studies
Alexis Nasr, Frédéric Bechet, Benoît Favre, Thierry Bazillon, José Deulofeu and Andre Valli

O11 • Collaborative Resources (1)

Chairperson: Iryna Gurevych

Silfurberg B

- 16:45-17:05 Corpus Annotation through Crowdsourcing: Towards Best Practice Guidelines
Marta Sabou, Kalina Bontcheva, Leon Derczynski and Arno Scharl
- 17:05-17:25 Towards an Environment for the Production and the Validation of Lexical Semantic Resources
Mikaël Morardo and Eric De La Clergerie
- 17:25-17:45 Towards an Encyclopedia of Compositional Semantics: Documenting the Interface of the English Resource Grammar
Dan Flickinger, Emily M. Bender and Stephan Oepen
- 17:45-18:05 Mapping CPA Patterns onto OntoNotes Senses
Octavian Popescu, Martha Palmer and Patrick Hanks

O12 • Semantics (1)

Chairperson: Eva Hajičová

Kaldalon

- 16:45-17:05 T-PAS; a Resource of Typed Predicate Argument Structures for Linguistic Analysis and Semantic Processing
Elisabetta Jezeck, Bernardo Magnini, Anna Feltracco, Alessia Bianchini and Octavian Popescu
- 17:05-17:25 The N2 Corpus: a Semantically Annotated Collection of Islamist Extremist Stories
Mark Finlayson, Jeffrey Halverson and Steven Corman
- 17:25-17:45 Predicate Matrix: Extending SemLink through WordNet Mappings
Maddalen Lopez de Lacalle, Egoitz Laparra and German Rigau
- 17:45-18:05 A Unified Annotation Scheme for the Semantic/Pragmatic Components of Definiteness
Archana Bhatia, Mandy Simons, Lori Levin, Yulia Tsvetkov, Chris Dyer and Jordan Bender

P13 • Discourse Annotation, Representation and Processing

Chairperson: Ann Bies

A Model for Processing Illocutionary Structures and Argumentation in Debates

Kasia Budzynska, Mathilde Janier, Chris Reed, Patrick Saint-Dizier, Manfred Stede and Olena Yakorska

Potsdam Commentary Corpus 2.0: Annotation for Discourse Research

Manfred Stede and Arne Neumann

Verbs of Saying with a Textual Connecting Function in the Prague Discourse Treebank

Magdalena Rysova

Building a Corpus of Manually Revised Texts from Discourse Perspective

Ryu Iida and Takenobu Tokunaga

The CUHK Discourse Treebank for Chinese: Annotating Explicit Discourse Connectives for the Chinese Treebank

Lanjun Zhou, Binyang Li, Zhongyu Wei and Kam-Fai Wong

Computational Narratology: Extracting Tense Clusters from Narrative Texts

Thomas Bögel, Jannik Strötgen and Michael Gertz

Can Numerical Expressions Be Simpler? Implementation and Demonstration of a Numerical Simplification System for Spanish

Susana Bautista and Horacio Saggion

Cross-Linguistic Annotation of Narrativity for English/French Verb Tense Disambiguation

Cristina Grisot and Thomas Meyer

P14 • Grammar and Syntax

Chairperson: Cristina Bosco

A Database for Measuring Linguistic Information Content

Richard Sproat, Bruno Cartoni, HyunJeong Choe, David Huynh, Linne Ha, Ravindran Rajakumar and Evelyn Wenzel-Grondie

Valency and Word Order in Czech - a Corpus Probe

Katerina Rysova and Jiří Mírovský

Mörkum Njálu. An Annotated Corpus to Analyse and Explain Grammatical Divergences Between 14th-century Manuscripts of Njál's Saga.

Ludger Zeevaert

GenitivDB - a Corpus-generated Database for German Genitive Classification

Roman Schneider

Building a Reference Lexicon for Countability in English

Tibor Kiss, Francis Jeffry Pelletier and Tobias Stadtfeld

P15 • Lexicons

Chairperson: Amália Mendes

Disambiguating Verbs by Collocation: Corpus Lexicography Meets Natural Language Processing

Ismail El Maarouf, Jane Bradbury, Vít Baisa and Patrick Hanks

GLÀFF, a Large Versatile French Lexicon

Nabil Hathout, Franck Sajous and Basilio Calderone

Bilingual Dictionary Construction with Transliteration Filtering

John Richardson, Toshiaki Nakazawa and Sadao Kurohashi

Bootstrapping Open-Source English-Bulgarian Computational Dictionary

Krasimir Angelov

MotàMot Project: Conversion of a French-Khmer Published Dictionary for Building a Multilingual Lexical System

Mathieu Mangeot

RELISH LMF: Unlocking the Full Power of the Lexical Markup Framework

Menzo Windhouwer, Justin Petro and Shakila Shayan

Building a Dataset of Multilingual Cognates for the Romanian Lexicon

Liviu Dinu and Alina Maria Ciobanu

LexTec - a Rich Language Resource for Technical Domains in Portuguese

Palmira Marrafa, Raquel Amaro and Sara Mendes

P16 • Morphology

Chairperson: Benoît Sagot

Amazigh Verb Conjugator

Fadoua Ataa Allah and Siham Boulaknadel

The Development of Dutch and Afrikaans Language Resources for Compound Boundary Analysis.

Menno van Zaanen, Gerhard Van Huyssteen, Suzanne Aussems, Chris Emmery and Roald Eiselen

Zmorge: a German Morphological Lexicon Extracted from Wiktionary

Rico Sennrich and Beat Kunz

A New Form of Humor - Mapping Constraint-based Computational Morphologies to a Finite-State Representation

Attila Novák

Szeged Corpus 2.5: Morphological Modifications in a Manually POS-tagged Hungarian Corpus

Veronika Vincze, Viktor Varga, Katalin Ilona Simkó, János Zsibrita, ágoston Nagy, Richárd Farkas and János Csirik

A Set of Open Source Tools for Turkish Natural Language Processing

Çağrı Çöltekin

Word-Formation Network for Czech

Magda Sevcikova and Zdenek Zabokrtsky

MADAMIRA: a Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic

Arfath Pasha, Mohamed Al-Badrashiny, Mona Diab, Ahmed El Kholly, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow and Ryan Roth

Computer-aided Morphology Expansion for Old Swedish

Yvonne Adesam, Malin Ahlberg, Peter Andersson, Gerlof Bouma, Markus Forsberg and Mans Hulden

Morfeusz Reloaded

Marcin Woliński

P17 • WordNet

Chairperson: Francis Bond

Automatic Creation of WordNets from Parallel Corpora

Antoni Oliver and Salvador Climent

Mapping WordNet Domains, WordNet Topics and Wikipedia Categories to Generate Multilingual Domain Specific Resources

Spandana Gella, Carlo Strapparava and Vivi Nastase

Adapting VerbNet to French using Existing Resources

Quentin Pradet, Laurence Danlos and Gaël de Chalendar

Bootstrapping an Italian VerbNet: Data-driven Analysis of Verb Alternations

Gianluca Lebani, Veronica Viola and Alessandro Lenci

Dense Components in the Structure of WordNet

Ahti Lohk, Kaarel Allik, Heili Orav and Leo Võhandu

The Making of Ancient Greek WordNet

Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini and Gregory Crane

Etymological WordNet: Tracing the History of Words

Gerard de Melo

5-minute break

O13 • Sentiment Analysis (1)

Chairperson: Paul Buitelaar

Nordurljos

- 18:10-18:30 Generating Polarity Lexicons with WordNet Propagation in 5 Languages
Isa Maks, Ruben Izquierdo, Francesca Frontini, Rodrigo Agerrí, Piek Vossen and Andoni Azpeitia
- 18:30-18:50 SANA: a Large Scale Multi-Genre, Multi-Dialect Lexicon for Arabic Subjectivity and Sentiment Analysis
Muhammad Abdul-Mageed and Mona Diab
- 18:50-19:10 On the Importance of Text Analysis for Stock Price Prediction
Heeyoung Lee, Mihai Surdeanu, Bill MacCarney and Dan Jurafsky

O14 • Paralinguistics

Chairperson: Sophie Rosset

Silfurberg A

- 18:10-18:30 An Analysis of Older Users' Interactions with Spoken Dialogue Systems
Jamie Bost and Johanna Moore
- 18:30-18:50 Alert!... Calm Down, There is Nothing to Worry About. Warning and Soothing Speech Synthesis
Milan Rusko, Sakhia Darjaa, Marian Trnka, Marian Ritomsky and Robert Sabo
- 18:50-19:10 Prosodic, Syntactic, Semantic Guidelines for Topic Structures Across Domains and Corpora
Ana Isabel Mata, Helena Moniz, Telmo Mória, Anabela Gonçalves, Fátima Silva, Fernando Batista, Inês Duarte, Fátima Oliveira and Isabel Falé

O15 • Multiword Expressions

Chairperson: Aline Villavicencio

Silfurberg B

- 18:10-18:30 How to Tell a Schneemann from a Milchmann: an Annotation Scheme for Compound-Internal Relations
Corina Dima, Verena Henrich, Erhard Hinrichs and Christina Hoppermann
- 18:30-18:50 Semi-Compositional Method for Synonym Extraction of Multi-Word Terms
Béatrice Daille and Amir Hazem
- 18:50-19:10 Multiword Expressions in Machine Translation
Valia Kordoni and Iliana Simova

O16 • Spelling Normalisation

Chairperson: Hrafn Loftsson

Kaldalon

- 18:10-18:30 A Database of Freely Written Texts of German School Students for the Purpose of Automatic Spelling Error Classification
Kay Berkling, Johanna Fay, Masood Ghayoomi, Katrin Hein, Rémi Lavalley, Ludwig Linhuber and Sebastian Stüker
- 18:30-18:50 Towards Shared Datasets for Normalization Research
Orphée De Clercq, Sarah Schulz, Bart Desmet and Véronique Hoste
- 18:50-19:10 Synergy of Nederlab and @PhilosTEI: Diachronic and Multilingual Text-induced Corpus Clean-up
Martin Reynaert

18:10 - 19:30

Poster Sessions - Poster Area (2)

P18 • Corpora and Annotation

Chairperson: Steve Cassidy

Translation Errors from English to Portuguese: an Annotated Corpus
Angela Costa, Tiago Luís and Luísa Coheur

CoRoLa - the Reference Corpus of Contemporary Romanian Language
Verginica Barbu Mititelu, Elena Irimia and Dan Tufiş

A Multidialectal Parallel Corpus of Arabic
Houda Bouamor, Nizar Habash and Kemal Oflazer

YouDACC: the YouTube Dialectal Arabic Comment Corpus
Ahmed Salama, Houda Bouamor, Behrang Mohit and Kemal Oflazer

Comparing Two Acquisition Systems for Automatically Building an English-Croatian Parallel Corpus from Multilingual Websites
Miquel Esplà-Gomis, Filip Klubička, Nikola Ljubešić, Sergio Ortiz-Rojas, Vassilis Papavassiliou and Prokopis Prokopidis

Towards an Integration of Syntactic and Temporal Annotations in Estonian
Siim Orasmaa

Annotation of Specialized Corpora using a Comprehensive Entity and Relation Scheme
Louise Deleger, Anne-Laure Ligozat, Cyril Grouin, Pierre Zweigenbaum and Aurélie Neveol

Developing Politeness Annotated Corpus of Hindi Blogs
Ritesh Kumar

The MERLIN Corpus: Learner Language and the CEFR
Adriane Boyd, Jirka Hana, Lionel Nicolas, Detmar Meurers, Katrin Wisniewski, Andrea Abel, Karin Schöne, Barbora Šindlová and Chiara Vettori

DysList: an Annotated Resource of Dyslexic Errors
Luz Rello, Ricardo Baeza-Yates and Joaquim Llisterra

Criteria for Identifying and Annotating Caused Motion Constructions in Corpus Data
Jena D. Hwang, Annie Zaenen and Martha Palmer

The American Local News Corpus
Ann Irvine, Joshua Langfus and Chris Callison-Burch

P19 • Document Classification, Text Categorisation

Chairperson: Karën Fort

A LDA-based Topic Classification Approach from highly Imperfect Automatic Transcriptions
Mohamed Morchid, Richard Dufour and Georges Linares

How to Use Less Features and Reach Better Performance in Author Gender Identification
Juan Soler and Leo Wanner

Genres in the Prague Discourse Treebank
Lucie Poláková, Pavlína Jínová and Jiří Mírovský

Data Mining with Shallow vs. Linguistic Features to Study Diversification of Scientific Registers
Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes, Ekaterina Lapshinova-Koltunski, Noam Ordan and Elke Teich

Detecting Document Structure in a Very Large Corpus of UK Financial Reports
Mahmoud El-Haj, Paul Rayson, Steve Young and Martin Walker

Designing and Evaluating a Reliable Corpus of Web Genres via Crowd-Sourcing
Noushin Rezapour Asheghi, Serge Sharoff and Katja Markert

Locating Requests among Open Source Software Communication Messages
Ioannis Korkontzelos and Sophia Ananiadou

Sockpuppet Detection in Wikipedia: a Corpus of Real-World Deceptive Writing for Linking Identities
Tamar Solorio, Ragib Hasan and Mainul Mizan

P20 • FrameNet

Chairperson: Alessandro Lenci

Reusing Swedish FrameNet for Training Semantic Roles

Ildikó Pilán and Elena Volodina

Discovering Frames in Specialized Domains

Marie-Claude L'Homme, Benoît Robichaud and Carlos Subirats Rüggeberg

Developing a French FrameNet: Methodology and First Results

Marie Candito, Pascal Amsili, Lucie Barque, Farah Benamara, Gaël de Chalendar, Marianne Djemaa, Pauline Haas, Richard Huyghe, Yvette Yannick Mathieu, Philippe Muller, Benoît Sagot and Laure Vieu

P21 • Semantics

Chairperson: Peter Anick

Corpus-based Computation of Reverse Associations

Reinhard Rapp

First Approach toward Semantic Role Labeling for Basque

Haritz Salaberri, Olatz Arregi and Beñat Zapirain

Constructing a Corpus of Japanese Predicate Phrases for Synonym/Antonym Relations

Tomoko Izumi, Tomohide Shibata, Hisako Asano, Yoshihiro Matsuo and Sadao Kurohashi

Distributed Distributional Similarities of Google Books over the Centuries

Martin Riedl, Richard Steuer and Chris Biemann

Lexical Substitution Dataset for German

Kostadin Cholakov, Chris Biemann, Judith Eckle-Kohler and Iryna Gurevych

Buy One Get One Free: Distant Annotation of Chinese Tense, Event Type and Modality

Nianwen Xue and Yuchen Zhang

Latent Semantic Analysis Models on Wikipedia and TASA

Dan Stefanescu, Rajendra Banjade and Vasile Rus

Annotation of Computer Science Papers for Semantic Relation Extraction

Yuka Tateisi, Yo Shidahara, Yusuke Miyao and Akiko Aizawa

Automatic Extraction of Synonyms for German Particle Verbs from Parallel Data with Distributional Similarity as a Re-Ranking Feature

Moritz Wittmann, Marion Weller and Sabine Schulte im Walde

DBpedia Domains: Augmenting DBpedia with Domain Information

Gregor Tütze, Volha Bryl, Căcilia Zirn and Simone Paolo Ponzetto

Classifying Inconsistencies in DBpedia Language Specific Chapters

Elena Cabrio, Serena Villata and Fabien Gandon

P22 • Speech Resources

Chairperson: Giuseppe Riccardi

The Database for Spoken German – DGD2

Thomas Schmidt

The EASR Corpora of European Portuguese, French, Hungarian and Polish Elderly Speech
Annika Hämäläinen, Jairo Avelar, Sílvia Rodrigues, Miguel Sales Dias, Artur Kolesiński, Tibor Fegyó, Géza Németh, Petra Csobánka, Karine Lan and David Hewson

GRASS: the Graz Corpus of Read and Spontaneous Speech

Barbara Schuppler, Martin Hagmueller, Juan A. Morales-Cordovilla and Hannes Pessentheiner

Design and Development of an RDB Version of the Corpus of Spontaneous Japanese

Hanae Koiso, Yasuharu Den, Ken'ya Nishikawa and Kikuo Maekawa

Designing a Bilingual Speech Corpus for French and German Language Learners: a Two-Step Process

Camille Fauth, Anne Bonneau, Frank Zimmerer, Juergen Trouvain, Bistra Andreeva, Vincent Colotte, Dominique Fohr, Denis Jouvet, Jeanin Jügler, Yves Laprie, Odile Mella and Bernd Möbius

Semi-Automatic Annotation of the UCU Accents Speech Corpus

Rosemary Orr, Marijn Huijbregts, Roeland van Beek, Lisa Teunissen, Kate Backhouse and David van Leeuwen

A Corpus of European Portuguese Child and Child-directed Speech

Ana Lúcia Santos, Michel Génèreux, Aida Cardoso, Celina Agostinho and Silvana Abalada

The SSPNet-Mobile Corpus: Social Signal Processing over Mobile Phones.

Anna Polychroniou, Hugues Salamin and Alessandro Vinciarelli

Annotation Pro + TGA: Automation of Speech Timing Analysis

Katarzyna Klessa and Dafydd Gibbon

The Munich Biovoice Corpus: Effects of Physical Exercising, Heart Rate, and Skin Conductance on Human Speech Production

Björn Schuller, Felix Friedmann and Florian Eyben

20:00

Welcome Reception at Harpa