

---

# Detailed Conference Programme

---

## 29 May 2014

9:00-9:40 **Keynote Speech**  
*Hassan Sawaf: Language Technology for Commerce, the eBay Way*

Chairperson: Khalid Choukri  
*Eldborg*

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

### **O17 • Infrastructures for LRs**

Chairperson: Nancy Ide

*Nordurljos*

9:45-10:05 ELRA's Consolidated Services for the HLT Community  
*Victoria Arranz, Khalid Choukri, Valérie Mapelli and Hélène Mazo*

10:05-10:25 The Strategic Impact of META-NET on the Regional, National and International Level  
*Georg Rehm, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audronė Bielevičienė, Lars Borin, António Branco, Gerhard Budin, Nicoletta Calzolari, Walter Daelemans, Radovan Garabik, Marko Grobelnik, Carmen Garcia-Mateo, Josef van Genabith, Jan Hajic, Inma Hernaez, John Judge, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Linden, Bernardo Magnini, Joseph Mariani, John McNaught, Maite Melero, Monica Monachini, Asuncion Moreno, Jan Odijk, Maciej Ogrodniczuk, Piotr Pezik, Stelios Piperidis, Adam Przepiórkowski, Eiríkur Rögnvaldsson, Michael Rosner, Bolette Pedersen, Inguna Skadina, Koenraad De Smedt, Marko Tadić, Paul Thompson, Dan Tufiş, Tamás Váradi, Andrejs Vasiljevs, Kadri Vider and Jolanta Zabarskaite*

10:25-10:45 The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars  
*Erhard Hinrichs and Steven Krauwer*

10:45-11:05 META-SHARE: One Year After  
*Stelios Piperidis, Harris Papageorgiou, Christian Spurk, Georg Rehm, Khalid Choukri, Olivier Hamon, Nicoletta Calzolari, Riccardo Del Gratta, Bernardo Magnini and Christian Girardi*

11:05-11:25 New Directions for Language Resource Development and Distribution  
*Christopher Cieri, Denise DiPersio, Mark Liberman, Andrea Mazzucchi, Stephanie Strassel and Jonathan Wright*

### **O18 • Speech Resources Annotation**

Chairperson: Satoshi Nakamura

*Silfurberg A*

9:45-10:05 Designing the Latvian Speech Recognition Corpus  
*Mārcis Pinnis, Ilze Auziņa and Kārlis Goba*

10:05-10:25 A Corpus of Spontaneous Speech in Lectures: the KIT Lecture Corpus for Spoken Language Processing and Translation  
*Eunah Cho, Sarah Fünfer, Sebastian Stüker and Alex Waibel*

- 10:25-10:45 The Pragmatic Annotation of a Corpus of Academic Lectures  
*Sian Alsop and Hilary Nesi*
- 10:45-11:05 HESITA(te) in Portuguese  
*Sara Candeias, Dirce Celorico, Jorge Proença, Arlindo Veiga, Carla Lopes and Fernando Perdigão*
- 11:05-11:25 VOCE Corpus: Ecologically Collected Speech Annotated with Physiological and Psychological Stress Assessments  
*Ana Aguiar, Mariana Kaiseler, Hugo Meinedo, Pedro Almeida, Mariana Cunha and Jorge Silva*

### **O19 • Summarisation**

Chairperson: Horacio Saggion

*Silfurberg B*

- 9:45-10:05 The Impact of Cohesion Errors in Extraction Based Summaries  
*Evelina Rennes and Arne Jonsson*
- 10:05-10:25 Out in the Open: Finding and Categorising Errors in the Lexical Simplification Pipeline  
*Matthew Shardlow*
- 10:25-10:45 LQVSumm: a Corpus of Linguistic Quality Violations in Multi-Document Summarization  
*Annemarie Friedrich, Marina Valeeva and Alexis Palmer*
- 10:45-11:05 Summarizing News Clusters on the Basis of Thematic Chains  
*Natalia Loukachevitch and Aleksey Alekseev*
- 11:05-11:25 A Repository of State of the Art and Competitive Baseline Summaries for Generic News Summarization  
*Kai Hong, John Conroy, Benoît Favre, Alex Kulesza, Hui Lin and Ani Nenkova*

### **O20 • Grammar, Lexicon and Morphology**

Chairperson: Lori Levin

*Kaldalon*

- 9:45-10:05 The Interplay Between Lexical and Syntactic Resources in Incremental Parsebanking  
*Victoria Rosén, Petter Haugereid, Martha Thunes, Gyri S. Losnegaard and Helge Dyvik*
- 10:05-10:25 An Efficient Language Independent Toolkit for Complete Morphological Disambiguation  
*László Laki and György Orosz*
- 10:25-10:45 Language Resource Addition: Dictionary or Corpus?  
*Shinsuke Mori and Graham Neubig*
- 10:45-11:05 Utilizing Constituent Structure for Compound Analysis  
*Kristín Bjarnadóttir and Jón Daðason*
- 11:05-11:25 Word Semantic Similarity for Morphologically Rich Languages  
*Kalliopi Zervanou, Elias Iosif and Alexandros Potamianos*

9:45 - 11:25 **Poster Sessions - Poster Area (1)**

### **P23 • Collaborative Resource Construction**

Chairperson: Christian Chiarcos

Digital Library 2.0: Source of Knowledge and Research Collaboration Platform  
*Włodzimierz Gruszczyński and Maciej Ogrodniczuk*

Exploiting Networks in Law

*Livio Robaldo, Guido Boella, Luigi Di Caro and Andrea Viola*

Guampa: a Toolkit for Collaborative Translation

*Alex Rudnick, Taylor Skidmore, Alberto Samaniego and Michael Gasser*

The Halliday Centre Tagger: an Online Platform for Semi-automatic Text Annotation and Analysis

*Billy T.M. Wong, Ian C. Chow, Jonathan J. Webster and Hengbin Yan*

Modeling, Managing, Exposing, and Linking Ontologies with a Wiki-based Tool

*Mauro Dragoni, Alessio Bosca, Matteo Casu and Andi Rexha*

Propa-L: a Semantic Filtering Service from a Lexical Network Created using Games With A Purpose

*Mathieu Lafourcade and Karën Fort*

Open Philology at the University of Leipzig

*Frederik Baumgardt, Giuseppe Celano, Gregory R. Crane, Stella Dee, Maryam Foradi,*

*Emily Franzini, Greta Franzini, Monica Lent, Maria Moritz and Simona Stoyanova*

LexTerm Manager: Design for an Integrated Lexicography and Terminology System

*Joshua Elliot, Logan Kearsley, Jason Housley and Alan Melby*

RESTful Annotation and Efficient Collaboration

*Jonathan Wright*

## **P24 • Corpora and Annotation**

Chairperson: Maria Gavrilidou

Collecting Natural SMS and Chat Conversations in Multiple Languages: the BOLT Phase 2 Corpus

*Zhiyi Song, Stephanie Strassel, Haejoong Lee, Kevin Walker, Jonathan Wright, Jennifer Garland, Dana Fore, Brian Gainor, Preston Cabe, Thomas Thomas, Brendan Callahan and Ann Sawyer*

The Slovak Categorized News Corpus

*Daniel Hladek, Jan Stas and Jozef Juhar*

TUKE-Bnews-SK: Slovak Broadcast News Corpus Construction and Evaluation

*Matus Pleva and Jozef Juhar*

Sublanguage Corpus Analysis Toolkit: a Tool for Assessing the Representativeness and Sublanguage Characteristics of Corpora

*Irina Temnikova, William A. Baumgartner Jr., Negacy D. Hailu, Ivelina Nikolova, Tony McEnery, Adam Kilgarriff, Galia Angelova and K. Bretonnel Cohen*

The Hungarian Gigaword Corpus

*Csaba Oravecz, Tamás Váradi and Bálint Sass*

The SETimes.HR Linguistically Annotated Corpus of Croatian

*Željko Agić and Nikola Ljubešić*

caWaC - a Web Corpus of Catalan and its Application to Language Modeling and Machine Translation

*Nikola Ljubešić and Antonio Toral*

ACTIV-ES: a Comparable, Cross-Dialect Corpus of "everyday" Spanish from Argentina, Mexico, and Spain

*Jerid Francom, Mans Hulden and Adam Ussishkin*

Language Editing Dataset of Academic Texts

*Vidas Daudaravicius*

Annotating the Focus of Negation in Japanese Text  
*Suguru Matsuyoshi, Ryo Otsuki and Fumiyo Fukumoto*

A Corpus of Participant Roles in Contentious Discussions  
*Siddharth Jain, Archana Bhatia, Angelique Rein and Eduard Hovy*

## **P25 • Machine Translation**

Chairperson: Holger Schwenk

CFT13: a Resource for Research into the Post-editing Process  
*Michael Carl, Mercedes Martínez García and Bartolomé Mesa-Lao*

Not an Interlingua, but Close: Comparison of English AMRs to Chinese and Czech  
*Nianwen Xue, Ondrej Bojar, Jan Hajic, Martha Palmer, Zdenka Uresova and Xiuhong Zhang*

On Complex Word Alignment Configurations  
*Miriam Kaeshammer and Anika Westburg*

Shata-Anuvadak: Tackling Multiway Translation of Indian Languages  
*Anoop Kunchukuttan, Abhijit Mishra, Rajen Chatterjee, Ritesh Shah and Pushpak Bhattacharyya*

Automatic Annotation of Machine Translation Datasets with Binary Quality Judgements  
*Marco Turchi and Matteo Negri*

A Large-Scale Evaluation of Pre-editing Strategies for Improving User-Generated Content Translation  
*Violeta Seretan, Pierrette Bouillon and Johanna Gerlach*

Rule-based Reordering Space in Statistical Machine Translation  
*Nicolas Pécheux, Alexander Allauzen and François Yvon*

Hindi to English Machine Translation: using Effective Selection in Multi-Model SMT  
*Kunal Sachdeva, Rishabh Srivastava, Sambhav Jain and Dipti Sharma*

## **P26 • Parallel Corpora**

Chairperson: Dan Tufiş

Benchmarking of English-Hindi Parallel Corpora  
*Jayendra Rakesh Yeka, Prasanth Kolachina and Dipti Misra Sharma*

Transliteration and Alignment of Parallel Texts from Cyrillic to Latin  
*Petic Mircea and Daniela Gifu*

Exploiting Catenae in a Parallel Treebank Alignment  
*Manuela Sanguinetti, Cristina Bosco and Loredana Cupi*

SwissAdmin: a Multilingual Tagged Parallel Corpus of Press Releases  
*Yves Scherrer, Luka Nerima, Lorenza Russo, Maria Ivanova and Eric Wehrli*

UM-Corpus: a Large English-Chinese Parallel Corpus for Statistical Machine Translation  
*Liang Tian, Derek F. Wong, Lidia S. Chao, Paulo Quaresma, Francisco Oliveira and Lu Yi*

Quality Estimation for Synthetic Parallel Data Generation  
*Raphael Rubino, Antonio Toral, Nikola Ljubešić and Gema Ramírez-Sánchez*

Billions of Parallel Words for Free: Building and using the EU Bookshop Corpus  
*Raivis Skadiņš, Jörg Tiedemann, Roberts Rozis and Daiga Deksnė*

The AMARA Corpus: Building Parallel Language Resources for the Educational Domain  
*Ahmed Abdelali, Francisco Guzman, Hassan Sajjad and Stephan Vogel*

Incorporating Alternate Translations into English Translation Treebank  
*Ann Bies, Justin Mott, Seth Kulick, Jennifer Garland and Colin Warner*

Dual Subtitles as Parallel Corpora  
*Shikun Zhang, Wang Ling and Chris Dyer*

Aligning Parallel Texts with InterText  
*Pavel Vondřička*

## **P27 • Sign Language**

Chairperson: Thomas Hanke

Expanding N-gram Analytics in ELAN and a Case Study for Sign Synthesis  
*Rosalee Wolfe, John McDonald, Larwan Berke and Marie Stumbo*

SLMotion - an Extensible Sign Language Oriented Video Analysis Tool  
*Matti Karppa, Ville Viitaniemi, Marcos Luzardo, Jorma Laaksonen and Tommi Jantunen*

S-pot - a Benchmark in Spotting Signs Within Continuous Signing  
*Ville Viitaniemi, Tommi Jantunen, Leena Savolainen, Matti Karppa and Jorma Laaksonen*

A Colloquial Corpus of Japanese Sign Language: Linguistic Resources for Observing Sign Language Conversations  
*Mayumi Bono, Kouhei Kikuchi, Paul Cibulka and Yutaka Osugi*

Exploring Factors that Contribute to Successful Fingerspelling Comprehension  
*Leah Geer and Jonathan Keane*

Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather  
*Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt and Hermann Ney*

The Use of a FileMaker Pro Database in Evaluating Sign Language Notation Systems  
*Julie Hochgesang*

A New Framework for Sign Language Recognition Based on 3D Handshape Identification and Linguistic Modeling  
*Mark Dilsizian, Polina Yanovich, Shu Wang, Carol Neidle and Dimitris Metaxas*

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

## **O21 • Collaborative Resources (2)**

Chairperson: Thierry Declerck

### ***Nordurljos***

11:45-12:05 Annotating Arguments: the NOMAD Collaborative Annotation Tool  
*Georgios Petasis*

12:05-12:25 Pivot-based Multilingual Dictionary Building using Wiktionary  
*Judit ács*

12:25-12:45 Terminology Resources and Terminology Work Benefit from Cloud Services  
*Tatiana Gornostay and Andrejs Vasiljevs*

12:45-13:05 Crowdsourcing for the Identification of Event Nominals: an Experiment  
*Rachele Sprugnoli and Alessandro Lenci*

## **O22 • Conversational (2)**

Chairperson: Dafydd Gibbon

### ***Silfurberg A***

- 11:45-12:05 Combining Elicited Imitation and Fluency Features for Oral Proficiency Measurement  
*Deryle Lonsdale and Carl Christensen*
- 12:05-12:25 On the Use of a Fuzzy Classifier to Speed Up the Sp ToBI Labeling of the Glissando Spanish Corpus  
*David Escudero, Aguilar-Cuevas Lourdes, González-Ferreras César, Gutiérrez-González Yurena and Valentín Cardeñoso-Payo*
- 12:25-12:45 The RATS Collection: Supporting HLT Research with Degraded Audio Data  
*David Graff, Kevin Walker, Stephanie Strassel, Xiaoyi Ma, Karen Jones and Ann Sawyer*
- 12:45-13:05 Eliciting and Annotating Uncertainty in Spoken Language  
*Heather Pon-Barry, Stuart Shieber and Nicholas Longenbaugh*

## **O23 • Text Mining**

Chairperson: Lucia Specia

### ***Silfurberg B***

- 11:45-12:05 The Meta-knowledge of Causality in Biomedical Scientific Discourse  
*Claudiu Mihăilă and Sophia Ananiadou*
- 12:05-12:25 Co-clustering of Bilingual Datasets as a Mean for Assisting the Construction of Thematic Bilingual Comparable Corpora  
*Guiyao Ke and Pierre-Francois Marteau*
- 12:25-12:45 NewsReader: Recording History from Daily News Streams  
*Piek Vossen, German Rigau, Luciano Serafini, Pim Stouten, Francis Irving and Willem Van Hage*
- 12:45-13:05 Identification of Technology Terms in Patents  
*Peter Anick, Marc Verhagen and James Pustejovsky*

## **O24 • Document Classification**

Chairperson: Robert Frederking

### ***Kaldalon***

- 11:45-12:05 Cross-Language Authorship Attribution  
*Dasha Bogdanova and Angeliki Lazaridou*
- 12:05-12:25 Learning from Domain Complexity  
*Robert Remus and Dominique Ziegelmayr*
- 12:25-12:45 Using Word Familiarities and Word Associations to Measure Corpus Representativeness  
*Reinhard Rapp*
- 12:45-13:05 A Modular System for Rule-based Text Categorisation  
*Marco Del Tredici and Malvina Nissim*

**P28 • Information Extraction**

Chairperson: Diana Maynard

Extracting News Web Page Creation Time with DCTFinder

*Xavier Tannier*

Information Extraction from German Patient Records via Hybrid Parsing and Relation Extraction Strategies

*Hans-Ulrich Krieger, Christian Spurr, Hans Uszkoreit, Feiyu Xu, Yi Zhang, Frank Müller and Thomas Tolxdorff*

Media Monitoring and Information Extraction for the Highly Inflected Agglutinative Language Hungarian

*Júlia Pajzs, Ralf Steinberger, Maud Ehrmann, Mohamed Ebrahim, Leonida Della Rocca, Stefano Bucci, Eszter Simon and Tamás Váradi*

Creating a Gold Standard Corpus for the Extraction of Chemistry-Disease Relations from Patent Texts

*Antje Schlaf, Claudia Bobach and Matthias Irmer*

T2K<sup>2</sup>: a System for Automatically Extracting and Organizing Knowledge from Texts

*Felice Dell'Orletta, Giulia Venturi, Andrea Cimino and Simonetta Montemagni*

Freepal: A Large Collection of Deep Lexico-Syntactic Patterns for Relation Extraction

*Johannes Kirschnick, Alan Akbik and Holmer Hensen*

Ranking Job Offers for Candidates: Learning Hidden Knowledge from Big Data

*Marc Poch, Núria Bel, Sergio Espeja and Felipe Navio*

Hot Topics and Schisms in NLP: Community and Trend Analysis with Saffron on ACL and LREC Proceedings

*Paul Buitelaar, Georgeta Bordea and Barry Coughlan*

Textual Emigration Analysis (TEA)

*Andre Blessing and Jonas Kuhn*

**P29 • Lexicons**

Chairperson: Nianwen Xue

WordNet-Wikipedia-Wiktionary: Construction of a Three-way Alignment

*Tristan Miller and Iryna Gurevych*

xLiD-Lexica: Cross-Lingual Linked Data Lexica

*Lei Zhang, Michael Färber and Achim Rettinger*

Turkish Resources for Visual Word Recognition

*Begum Erten, Cem Bozsahin and Deniz Zeyrek*

Computer-aided Quality Assurance of an Icelandic Pronunciation Dictionary

*Martin Jansche*

Bring vs. MTRoget: Evaluating Automatic Thesaurus Translation

*Lars Borin, Jens Allwood and Gerard de Melo*

Bilingual Dictionary Induction as an Optimization Problem

*Wushouer Mairidan, Toru Ishida, Donghui Lin and Katsutoshi Hirayama*

Enriching the "Senso Comune" Platform with Automatically Acquired Data  
*Tommaso Caselli, Laure Vieu, Carlo Strapparava and Guido Vetere*

MUHIT: a Multilingual Harmonized Dictionary  
*Sameh Alansary*

Language Resources for French in the Biomedical Domain  
*Aurelie Neveol, Julien Grosjean, Stéfan Darmoni and Pierre Zweigenbaum*

Gold-Standard for Topic-specific Sentiment Analysis of Economic Texts  
*Pyry Takala, Pekka Malo, Ankur Sinha and Oskar Ahlgren*

### **P30 • Large Projects and Infrastructural Issues**

Chairperson: Yohei Murakami

A Decade of HLT Agency Activities in the Low Countries: from Resource Maintenance (BLARK) to Service Offerings (BLAISE)  
*Peter Spyns and Remco van Veenendaal*

CLARA: a New Generation of Researchers in Common Language Resources and their Applications  
*Koenraad De Smedt, Erhard Hinrichs, Detmar Meurers, Inguna Skadina, Bolette Pedersen, Costanza Navarretta, Núria Bel, Krister Linden, Marketa Lopatkova, Jan Hajic, Gisle Andersen and Przemyslaw Lenkiewicz*

Encompassing a Spectrum of LT Users in the CLARIN-DK Infrastructure  
*Lina Henriksen, Dorte Haltrup Hansen, Bente Maegaard, Bolette Sandford Pedersen and Claus Povlsen*

Legal Aspects of Text Mining  
*Maarten Truyens and Patrick Van Eecke*

CLARIN-NL: Major Results  
*Jan Odijk*

An Innovative World Language Centre: Challenges for the Use of Language Technology  
*Auður Hauksdóttir*

Facing the Identification Problem in Language-Related Scientific Data Analysis.  
*Joseph Mariani, Christopher Cieri, Gil Francopoulo, Patrick Paroubek and Marine Delaborde*

Taalportaal: an Online Grammar of Dutch and Frisian  
*Frank Landsbergen, Carole Tiberius and Roderik Dernison*

### **P31 • Opinion Mining and Reviews Analysis**

Chairperson: Manfred Stede

The USAGE Review Corpus for Fine Grained Multi Lingual Opinion Analysis  
*Roman Klinger and Philipp Cimiano*

PACE Corpus: a Multilingual Corpus of Polarity-Annotated Textual Data from the Domains Automotive and CELLphone  
*Christian Haenig, Andreas Niekler and Carsten Wuensch*

Adapting Freely Available Resources to Build an Opinion Mining Pipeline in Portuguese  
*Patrik Lambert and Carlos Rodriguez-Penagos*

The NewSoMe Corpus: a Unifying Opinion Annotation Framework Across Genres and in Multiple Languages  
*Roser Saurí, Judith Domingo and Toni Badia*



The Dangerous Myth of the Star System  
*André Bittar, Luca Dini, Sigrid Maurel and Mathieu Ruhlmann*

A Corpus of Comparisons in Product Reviews  
*Wiltrud Kessler and Jonas Kuhn*

### **P32 • Social Media Processing**

Chairperson: Fei Xia

Finding Romanized Arabic Dialect in Code-Mixed Tweets  
*Clare Voss, Stephen Tratz, Jamal Laoudi and Douglas Briesch*

Hashtag Occurrences, Layout and Translation: a Corpus-driven Analysis of Tweets  
Published by the Canadian Government  
*Fabrizio Gotti, Phillippe Langlais and Atefeh Farzindar*

Clustering Tweets using Wikipedia Concepts  
*Guoyu Tang, Yunqing Xia, Weizhi Wang, Raymond Lau and Fang Zheng*

An Arabic Twitter Corpus for Subjectivity and Sentiment Analysis  
*Eshrag Refaee and Verena Rieser*

TweetNorm\_es: an Annotated Corpus for Spanish Microtext Normalization  
*Iñaki Alegria, Nora Aranberri, Pere Comas, Victor Fresno, Pablo Gamallo, Luís Padró, Iñaki San Vicente, Jordi Turmo and Arkaitz Zubiaga*

TweetCaT: a Tool for Building Twitter Corpora of Smaller Languages  
*Nikola Ljubešić, Darja Fišer and Tomaž Erjavec*

A German Twitter Snapshot  
*Tatjana Scheffler*

### **P33 • Treebanks**

Chairperson: Montserrat Marimón

The Procedure of Lexico-Semantic Annotation of Skladnica Treebank  
*Elżbieta Hajnicz*

Deep Syntax Annotation of the Sequoia French Treebank  
*Marie Candito, Guy Perrier, Bruno Guillaume, Corentin Ribeyre, Karèn Fort, Djamel Seddah and Eric de la Clergerie*

Projection-based Annotation of a Polish Dependency Treebank  
*Alina Wróblewska and Adam Przepiórkowski*

Croatian Dependency Treebank 2.0: New Annotation Guidelines for Improved Parsing  
*Željko Agić, Daša Berović, Danijela Merkler and Marko Tadić*

Correcting and Validating Syntactic Dependency in the Spoken French Treebank Rhapsodie  
*Rachel Bawden, Marie-Amélie Botalla, Kim Gerdes and Sylvain Kahane*

Because Size Does Matter: the Hamburg Dependency Treebank  
*Kilian A. Foth, Arne Köhn, Niels Beuck and Wolfgang Menzel*

HamleDT 2.0: Thirty Dependency Treebanks Stanfordized  
*Rudolf Rosa, Jan Mašek, David Mareček, Martin Popel, Daniel Zeman and Zdeněk Žabokrtský*

Bidirectional Converter Between Syntactic Annotations: from French Treebank Dependencies to PASSAGE Annotations, and back

*Munshi Asadullah, Patrick Paroubek and Anne Vilnat*

Developing an Egyptian Arabic Treebank: Impact of Dialectal Morphology on Annotation and Tool Development

*Mohamed Maamouri, Ann Bies, Seth Kulick, Michael Ciul, Nizar Habash and Ramy Eskander*

13:05-13:10 **5-minute break**

13:10-13:30 **Icelandic Invited Talk:**  
***Þórhallur Eyþórsson: Icelandic Quirks: Testing Linguistic Theories and Language Technology***

Chairperson: Eiríkur Rögnvaldsson

*Nordurljos*

13:30-14:55 **Lunch break**

### **O25 • Machine Translation and Evaluation (2)**

Chairperson: Toru Ishida

*Nordurljos*

- 14:55-15:15 The taraXÜ Corpus of Human-Annotated Machine Translations  
*Eleftherios Avramidis, Aljoscha Burchardt, Sabine Hunsicker, Maja Popović, Cindy Tscherwinka, David Vilar and Hans Uszkoreit*
- 15:15-15:35 Measuring the Impact of Spelling Errors on the Quality of Machine Translation  
*Irina Galinskaya, Valentin Gusev, Elena Mescheryakova and Mariya Shmatova*
- 15:35-15:55 A Quality-based Active Sample Selection Strategy for Statistical Machine Translation  
*Varvara Logacheva and Lucia Specia*
- 15:55-16:15 A Comparison of MT Errors and ESL Errors  
*Homa B. Hashemi and Rebecca Hwa*
- 16:15-16:35 VERTa: Facing a Multilingual Experience of a Linguistically-based MT Evaluation  
*Elisabet Comelles, Jordi Atserias, Victoria Arranz, Irene Castellon and Jordi Sesé*

### **O26 • Computer Aided Language Learning**

Chairperson: Justus Roux

*Silfurberg A*

- 14:55-15:15 ASR-based CALL Systems and Learner Speech Data: New Resources and Opportunities for Research and Development in Second Language Learning  
*Catia Cucchiari, Steve Bodnar, Bart Penning de Vries, Roeland Van Hout and Helmer Strik*
- 15:15-15:35 The Dutch LESLLA Corpus  
*Eric Sanders, Ineke van de Craats and Vanja de Lint*
- 15:35-15:55 Student Achievement and French Sentence Repetition Test Scores  
*Deryle Lonsdale and Benjamin Millard*

- 15 :55-16 :15 Using a Serious Game to Collect a Child Learner Speech Corpus  
*Claudia Baur, Manny Rayner and Nikos Tsourakis*
- 16 :15-16 :35 Sprinter: Language Technologies for Interactive and Multimedia Language Learning  
*Renlong Ai, Marcela Charfuelan, Walter Kasper, Tina Klüwer, Hans Uszkoreit, Feiyu Xu, Sandra Gasber and Philip Gienandt*

### **O27 • Information Extraction (1)**

Chairperson: Sophia Ananiadou

#### ***Silfurberg B***

- 14 :55-15 :15 Automatic Semantic Relation Extraction from Portuguese Texts  
*Leonardo Sameshima Taba and Helena Caseli*
- 15 :15-15 :35 Estimation of Speaking Style in Speech Corpora Focusing on Speech Transcriptions  
*Raymond Shen and Hideaki Kikuchi*
- 15 :35-15 :55 Annotating Clinical Events in Text Snippets for Phenotype Detection  
*Prescott Klassen, Fei Xia, Lucy Vanderwende and Meliha Yetisgen*
- 15 :55-16 :15 Annotating Inter-Sentence Temporal Relations in Clinical Notes  
*Jennifer D'Souza and Vincent Ng*
- 16 :15-16 :35 Characterizing and Predicting Bursty Events: the Buzz Case Study on Twitter  
*Mohamed Morchid, Georges Linares and Richard Dufour*

### **O28 • Lexicon**

Chairperson: Núria Bel

#### ***Kaldalon***

- 14 :55-15 :15 Building Domain Specific Bilingual Dictionaries  
*Lucas Hilgert, Lucelene Lopes, Artur Freitas, Renata Vieira, Denise Hogetop and Aline Vanin*
- 15:15-15:35 DeLex, a Freely-available, Large-scale and Linguistically Grounded Morphological Lexicon for German  
*Benoît Sagot*
- 15:35-15:55 Walenty: Towards a Comprehensive Valence Dictionary of Polish  
*Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski and Marek Świdziński*
- 15:55-16:15 A Language-independent Approach to Extracting Derivational Relations from an Inflectional Lexicon  
*Marion Baranes and Benoît Sagot*
- 16:15-16:35 Automatic Expansion of the MRC Psycholinguistic Database Imageability Ratings  
*Ting Liu, Kit Cho, G. Aaron Broadwell, Samira Shaikh, Tomek Strzalkowski, John Lien, Sarah Taylor, Laurie Feldman, Boris Yamrom, Nick Webb, Umit Boz, Ignacio Cases and Ching-Sheng Lin*

14:55 - 16:35

### **Poster Sessions - Poster Area (1)**

#### **P34 • Corpora and Annotation**

Chairperson: Zygmunt Vetulani

A Conventional Orthography for Tunisian Arabic  
*Inès Zribi, Rahma Boujelbane, Abir Masmoudi, Mariem Ellouze, Lamia Belguith and Nizar Habash*

Large Scale Arabic Error Annotation: Guidelines and Framework  
*Wajdi Zaghouani, Behrang Mohit, Nizar Habash, Ossama Obeid, Nadi Tomeh, Alla Rozovskaya, Noura Farra, Sarah Alkuhlani and Kemal Oflazer*

Flow Graph Corpus from Recipe Texts  
*Shinsuke Mori, Hirokuni Maeta, Yoko Yamakata and Tetsuro Sasada*

Recent Developments in DeReKo  
*Marc Kupietz and Harald Lünzen*

Why Chinese Web-as-Corpus is Wacky? Or: How Big Data is Killing Chinese Corpus Linguistics  
*Shu-Kai Hsieh*

Extending HeidelbergTime for Temporal Expressions Referring to Historic Dates  
*Jannik Strötgen, Thomas Bögel, Julian Zell, Ayser Armiti, Tran Van Canh and Michael Gertz*

A 500 Million Word POS-Tagged Icelandic Corpus  
*Thomas Eckart, Erla Hallsteinsdóttir, Sigrún Helgadóttir, Uwe Quasthoff and Dirk Goldhahn*

Building the Sense-Tagged Multilingual Parallel Corpus  
*Shan Wang and Francis Bond*

A Hindi-English Code-Switching Corpus  
*Anik Dey and Pascale Fung*

KoKo: an L1 Learner Corpus for German  
*Andrea Abel, Aivars Glaznieks, Lionel Nicolas and Egon Stemle*

On Paraphrase Identification Corpora  
*Vasile Rus, Rajendra Banjade and Mihai Lintean*

Construction and Annotation of a French Folkstale Corpus  
*Anne Garcia-Fernandez, Anne-Laure Ligozat and Anne Vilnat*

Statistical Analysis of Multilingual Text Corpus and Development of Language Models  
*Shyam Sundar Agrawal, Abhimanyu, Shweta Bansal and Minakshi Mahajan*

Building a Dataset for Summarization and Keyword Extraction from Emails  
*Vanessa Loza, Shibamouli Lahiri, Rada Mihalcea and Po-Hsiang Lai*

## **P35 • Grammar and Syntax**

Chairperson: Tamás Váradi

Language CoLLAGE: Grammatical Description with the LinGO Grammar Matrix  
*Emily M. Bender*

To Pay or to Get Paid: Enriching a Valency Lexicon with Diatheses  
*Anna Verneřová, Václava Kettnerová and Marketa Lopatkova*

The Ellogon Pattern Engine: Context-free Grammars over Annotations  
*Georgios Petasis*

Extracting a Bilingual Semantic Grammar from FrameNet-annotated Corpora  
*Dana Dannells and Normunds Gruzītis*

Relating Frames and Constructions in Japanese FrameNet  
*Kyoko Ohara*

MultiVal - Towards a Multilingual Valence Lexicon  
*Lars Hellan, Dorothee Beermann, Tore Bruland, Mary Esther Kropp Dakubu and Montserrat Marimon*

A Vector Space Model for Syntactic Distances Between Dialects  
*Emanuele Di Buccio, Giorgio Maria Di Nunzio and Gianmaria Silvello*

Resources in Conflict: a Bilingual Valency Lexicon vs. a Bilingual Treebank vs. a Linguistic Theory

*Jana Sindlerova, Zdenka Uresova and Eva Fucikova*

### **P36 • Metaphors**

Chairperson: Walter Daelemans

A Multi-Cultural Repository of Automatically Discovered Linguistic and Conceptual Metaphors  
*Samira Shaikh, Tomek Strzalkowski, Ting Liu, George Aaron Broadwell, Boris Yamrom, Sarah Taylor, Laurie Feldman, Kit Cho, Umit Boz, Ignacio Cases, Yuliya Peshkova and Ching-Sheng Lin*

Two Approaches to Metaphor Detection

*Brian MacWhinney and Davida Fromm*

Mining Online Discussion Forums for Metaphors

*Andrew Gargett and John Barnden*

### **P37 • Named Entity Recognition**

Chairperson: German Rigau

Simple Effective Microblog Named Entity Recognition: Arabic as an Example

*Kareem Darwish and Wei Gao*

Biomedical Entity Extraction using Machine-Learning Based Approaches

*Cyril Grouin*

NoSta-D Named Entity Annotation for German: Guidelines and Dataset

*Darina Benikova, Chris Biemann and Marc Reznicek*

Comparison of the Impact of Word Segmentation on Name Tagging for Chinese and Japanese

*Haibo Li, Masato Hagiwara, Qi Li and Heng Ji*

HFST-SweNER - a New NER Resource for Swedish

*Dimitrios Kokkinakis, Jyrki Niemi, Sam Hardwick, Krister Lindén and Lars Borin*

Crowdsourcing and Annotating NER for Twitter #drift

*Hege Fromreide, Dirk Hovy and Anders Sjøgaard*

Clustering of Multi-Word Named Entity Variants: Multilingual Evaluation

*Guillaume Jacquet, Maud Ehrmann and Ralf Steinberger*

Comparative Analysis of Portuguese Named Entities Recognition Tools

*Daniela Amaral, Evandro Fonseca, Lucelene Lopes and Renata Vieira*

Generating a Resource for Products and Brandnames Recognition. Application to the Cosmetic Domain.

*Cédric Lopez, Frédérique Segond, Olivier Hondermarck, Paolo Curtoni and Luca Dini*

Named Entity Corpus Construction using Wikipedia and DBpedia Ontology

*Younggyun Hahn, Jungyeul Park, Kyungtae Lim, Youngsik Kim, Dosam Hwang and Key-Sun Choi*

Exploring the Utility of Coreference Chains for Improved Identification of Personal Names

*Andrea Glaser and Jonas Kuhn*

Named Entity Tagging a Very Large Unbalanced Corpus: Training and Evaluating NE Classifiers

*Joachim Bingel and Thomas Haider*

### **P38 • Question Answering**

Chairperson: António Branco

REFRACTIVE: an Open Source Tool to Extract Knowledge from Syntactic and Semantic Relations

*Peter Exner and Pierre Nugues*

Overview of Todai Robot Project and Evaluation Framework of its NLP-based Problem Solving

*Akira Fujita, Akihiro Kameda, Ai Kawazoe and Yusuke Miyao*

Annotating Question Decomposition on Complex Medical Questions

*Kirk Roberts, Kate Masterton, Marcelo Fiszman, Halil Kilicoglu and Dina Demner-Fushman*

JUST.ASK, a QA System that Learns to Answer New Questions from Previous Interactions

*Sérgio Curto, Ana C. Mendes, Pedro Curto, Luísa Coheur and Angela Costa*

Extraction of Daily Changing Words for Question Answering

*Kugatsu Sadamitsu, Ryuichiro Higashinaka and Yoshihiro Matsuo*

LinkedHealthAnswers: Towards Linked Data-Driven Question Answering for the Health Care Domain

*Artem Ostankov, Florian Röhrbein and Ulli Waltinger*

A Tool Suite for Creating Question Answering Benchmarks

*Axel-Cyrille Ngonga Ngomo, Norman Heino, René Speck and Prodrimos Malakasiotis*

### **P39 • Speech Resources**

Chairperson: Henk van den Heuvel

The DIRHA Simulated Corpus

*Luca Cristoforetti, Mirco Ravanelli, Maurizio Omologo, Alessandro Sosi, Alberto Abad, Martin Hagnmueller and Petros Maragos*

Euronews: a Multilingual Speech Corpus for ASR

*Roberto Gretter*

Towards Multilingual Conversations in the Medical Domain: Development of Multilingual Medical Data and a Network-based ASR System

*Sakriani Sakti, Keigo Kubo, Sho Matsumiya, Graham Neubig, Tomoki Toda, Satoshi Nakamura, Fumihiko Adachi and Ryosuke Isotani*

The Slovene BNSI Broadcast News Database and Reference Speech Corpus GOS: Towards the Uniform Guidelines for Future Work

*Andrej Zgank, Ana Zwitter Vitez and Darinka Verdonik*

Aix Map Task Corpus: the French Multimodal Corpus of Task-oriented Dialogue

*Jan Gorisch, Corine Astésano, Ellen, Gurman Bard, Brigitte Bigi and Laurent Prévot*

CORILGA: a Galician Multilevel Annotated Speech Corpus for Linguistic Analysis

*Carmen Garcia-Mateo, Antonio Cardenal, Xose Luis Regueira, Elisa Fernández Rei, Marta Martínez, Roberto Seara, Rocío Varela and Noemí Basanta*

Basque Speecon-like and Basque SpeechDAT MDB-600: Speech Databases for the Development of ASR Technology for Basque

*Igor Odriozola, Inma Hernaez, María Inés Torres, Luis Javier Rodríguez-Fuentes, Mikel Penagarikano and Eva Navas*

New Bilingual Speech Databases for Audio Diarization

*David Tavaréz, Eva Navas, Daniel Erro, Ibon Saratzaga and Inma Hernaez*

Erlangen-CLP: a Large Annotated Corpus of Speech from Children with Cleft Lip and Palate  
*Tobias Bocklet, Andreas Maier, Korbinian Riedhammer, Ulrich Eysholdt and Elmar Nöth*  
The Development of the Multilingual LUNA Corpus for Spoken Language System Porting  
*Evgeny Stepanov, Giuseppe Riccardi and Ali Orkan Bayer*

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

### **O29 • Sentiment Analysis (2)**

Chairperson: Frédérique Segond

#### *Nordurljos*

- 16:55-17:15 CLIPS Stylometry Investigation (CSI) Corpus: a Dutch Corpus for the Detection of Age, Gender, Personality, Sentiment and Deception in Text  
*Ben Verhoeven and Walter Daelemans*
- 17:15-17:35 Building and Modelling Multilingual Subjective Corpora  
*Motaz Saad, David Langlois and Kamel Smaili*
- 17:35-17:55 Author-specific Sentiment Aggregation for Polarity Prediction of Reviews  
*Subhabrata Mukherjee and Sachindra Joshi*
- 17:55-18:15 Meta-Classifiers Easily Improve Commercial Sentiment Detection Tools  
*Mark Cieliebak, Oliver Dürr and Fatih Uzdilili*

### **O30 • Multimodality**

Chairperson: Michael Kipp

#### *Silfurberg A*

- 16:55-17:15 AusTalk: an Audio-Visual Corpus of Australian English  
*Dominique Estival, Steve Cassidy, Felicity Cox and Denis Burnham*
- 17:15-17:35 From Synsets to Videos: Enriching ItalWordNet Multimodally  
*Roberto Bartolini, Valeria Quochi, Irene De Felice, Irene Russo and Monica Monachini*
- 17:35-17:55 A Multimodal Dataset for Deception Detection  
*Veronica Perez-Rosas, Rada Mihalcea, Alexis Narvaez and Mihai Burzo*
- 17:55-18:15 The Distress Analysis Interview Corpus of Human and Computer Interviews  
*Jonathan Gratch, Ron Artstein, Gale Lucas, Giota stratou, Stefan Scherer, Angela Nazarian, Rachel Wood, Jill Boberg, David DeVault, Stacy Marsella, David Traum, Albert "Skip" Rizzo and Louis-Philippe Morency*

### **O31 • Under-resourced Languages**

Chairperson: Andrejs Vasiljevs

#### *Silfurberg B*

- 16:55-17:15 Modern Chinese Helps Archaic Chinese Processing: Finding and Exploiting the Shared Properties  
*Yan Song and Fei Xia*
- 17:15-17:35 Linguistic Landscaping of South Asia using Digital Language Resources: Genetic vs. Areal Linguistics  
*Lars Borin, Anju Saxena, Taraka Rama and Bernard Comrie*
- 17:35-17:55 PanLex: Building a Resource for Panlingual Lexical Translation  
*David Kamholz, Jonathan Pool and Susan Colowick*
- 17:55-18:15 Enriching ODIN  
*Fei Xia, William Lewis, Michael Wayne Goodman, Joshua Crowgey and Emily M. Bender*

## **O32 • Parallel Corpora**

Chairperson: Patrizia Paggio

### ***Kaldalon***

- 16:55-17:15 Creating a Massively Parallel Bible Corpus  
*Thomas Mayer and Michael Cysouw*
- 17:15-17:35 DCEP - Digital Corpus of the European Parliament  
*Najeh Hajlaoui, David Kolovratnik, Jaakko Väyrynen, Ralf Steinberger and Daniel Varga*
- 17:35-17:55 Innovations in Parallel Corpus Search Tools  
*Martin Volk, Johannes Graën and Elena Callegaro*
- 17:55-18:15 An Open-Source Heavily Multilingual Translation Graph Extracted from Wiktionaries and Parallel Corpora  
*Valérie Hanoka and Benoît Sagot*

16:55 - 18:15

## **Poster Sessions - Poster Area (2)**

### **P40 • Lexicons**

Chairperson: Yoshihiko Hayashi

Mapping the Lexique des Verbes du français (Lexicon of French Verbs) to a NLP Lexicon using Examples

*Bruno Guillaume, Karèn Fort, Guy Perrier and Paul Bédaride*

Text Readability and Word Distribution in Japanese

*Satoshi Sato*

High Quality Word Lists as a Resource for Multiple Purposes

*Uwe Quasthoff, Dirk Goldhahn, Thomas Eckart, Erla Hallsteinsdóttir and Sabine Fiedler*

ISLEX - a Multilingual Web Dictionary

*Bórdís Úlfarsdóttir*

Automatic Mapping Lexical Resources: a Lexical Unit as the Keystone

*Eduard Bejček, Kettnerová Václava and Marketa Lopatkova*

Towards Electronic SMS Dictionary Construction: an Alignment-based Approach

*Cédric Lopez, Reda Bestandji, Mathieu Roche and Rachel Panckhurst*

Bilingual Dictionaries for all EU Languages

*Ahmet Aker, Monica Paramita, Marcis Pinnis and Robert Gaizauskas*

Automatic Acquisition of Urdu Nouns (along with Gender and Irregular Plurals)

*Tafseer Ahmed Khan*

NomLex-PT: a Lexicon of Portuguese Nominalizations

*Valeria de Paiva, Livy Real, Alexandre Rademaker and Gerard de Melo*

### **P41 • Parsing**

Chairperson: Simonetta Montemagni

Sentence Rephrasing for Parsing Sentences with OOV Words

*Hen-Hsen Huang, Huan-Yuan Chen, Chang-Sheng Yu, Hsin-Hsi Chen, Po-Ching Lee and Chun-Hsun Chen*



Pruning the Search Space of the Wolof LFG Grammar using a Probabilistic and a Constraint Grammar Parser

*Cheikh M. Bamba Dione*

How Could Veins Speed Up the Process of Discourse Parsing

*Elena Mitocariu, Daniel Anechitei and Dan Cristea*

Parsing Heterogeneous Corpora with a Rich Dependency Grammar

*Achim Stein*

Treelet Probabilities for HPSG Parsing and Error Correction

*Angelina Ivanova and Gertjan van Noord*

Self-training a Constituency Parser using n-gram Trees

*Arda Celebi and Arzucan Özgür*

A Gold Standard Dependency Corpus for English

*Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer and Chris Manning*

DiscoSuite - A Parser Test Suite for German Discontinuous Structures

*Wolfgang Maier, Miriam Kaeshammer, Peter Baumann and Sandra Kübler*

## **P42 • Part-of-Speech Tagging**

Chairperson: Krister Lindén

SWIFT Aligner, a Multifunctional Tool for Parallel Corpora: Visualization, Word Alignment, and (Morpho)-Syntactic Cross-Language Transfer

*Timur Gilmanov, Olga Scrivner and Sandra Kübler*

The CLE Urdu POS Tagset

*Saba Urooj, Sarmad Hussain, Asad Mustafa, Rahila Parveen, Farah Adeeba, Tafseer Ahmed Khan, Miriam Butt and Annette Hautli*

Using Stem-Templates to Improve Arabic POS and Gender/Number Tagging

*Kareem Darwish, Ahmed Abdelali and Hamdy Mubarak*

The LIMA Multilingual Analyzer Made Free: FLOSS Resources Adaptation and Correction

*Gaël de Chalendar*

A Tagged Corpus and a Tagger for Urdu

*Bushra Jawaid, Amir Kamran and Ondrej Bojar*

Correcting Errors in a New Gold Standard for Tagging Icelandic Text

*Sigrún Helgadóttir, Hrafn Loftsson and Eiríkur Rögnvaldsson*

PoliTā: a Multitagger for Polish

*Lukasz Kobylński*

## **P43 • Semantics**

Chairperson: Marc Verhagen

Polysemy Index for Nouns: an Experiment on Italian using the PAROLE SIMPLE CLIPS Lexical Database

*Francesca Frontini, Valeria Quochi, Sebastian Padó, Monica Monachini and Jason Utt*

Comparing Similarity Measures for Distributional Thesauri

*Muntsa Padró, Marco Idiart, Aline Villavicencio and Carlos Ramisch*

Reconstructing the Semantic Landscape of Natural Language Processing

*Elisa Omodei, Jean-Philippe Cointet and Thierry Poibeau*

Compounds and Distributional Thesauri

*Olivier Ferret*

UnixMan Corpus: a Resource for Language Learning in the Unix Domain

*Kyle Richardson and Jonas Kuhn*

Multilingual EXtended WordNet Knowledge Base: Semantic Parsing and Translation of Glosses

*Tatiana Erekhinskaya, Meghana Satpute and Dan Moldovan*

Relation Inference in Lexical Networks ... with Refinements

*Manel Zarrouk and Mathieu Lafourcade*

Extracting Semantic Relations from Portuguese Corpora using Lexical-Syntactic Patterns

*Raquel Amaro*

An Analysis of Ambiguity in Word Sense Annotations

*David Jurgens*

PropBank: Semantics of New Predicate Types

*Claire Bonial, Julia Bonn, Kathryn Conger, Jena D. Hwang and Martha Palmer*

Semi-Supervised Methods for Expanding Psycholinguistics Norms by Integrating Distributional Similarity with the Structure of WordNet

*Michael Mohler, Marc Tomlinson, David Bracewell and Bryan Rink*

A Graph-based Approach for Computing Free Word Associations

*Gemma Bel Enguix, Reinhard Rapp and Michael Zock*

A Hierarchical Taxonomy for Classifying Hardness of Inference Tasks

*Martin Gleize and Brigitte Grau*

## **P44 • Speech Recognition and Synthesis**

Chairperson: Denise DiPersio

Speech Recognition Web Services for Dutch

*Joris Pelemans, Kris Demuyne, Hugo Van hamme and Patrick Wambacq*

Morpho-Syntactic Study of Errors from Speech Recognition System

*Maria Goryainova, Cyril Grouin, Sophie Rosset and Ioana Vasilescu*

Human Annotation of ASR Error Regions: is "gravity" a Sharable Concept for Human Annotators?

*Daniel Luzzati, Cyril Grouin, Ioana Vasilescu, Martine Adda-Decker, Eric Bilinski, Nathalie Camelin, Juliette Kahn, Carole Lailler, Lori Lamel and Sophie Rosset*

Development of a TV Broadcasts Speech Recognition System for Qatari Arabic

*Mohamed Elmahdy, Mark Hasegawa-Johnson and Eiman Mustafawi*

Automatic Long Audio Alignment and Confidence Scoring for Conversational Arabic Speech

*Mohamed Elmahdy, Mark Hasegawa-Johnson and Eiman Mustafawi*

The WaveSurfer Automatic Speech Recognition Plugin

*Giampiero Salvi and Niklas Vanhainen*

A Toolkit for Efficient Learning of Lexical Units for Speech Recognition

*Matti Varjokallio and Mikko Kurimo*

Using Audio Books for Training a Text-to-Speech System

*Aimilios Chalamandaris, Pirros Tsiakoulis, Sotiris Karabetos and Spyros Raptis*

### **5-minute break**

### **O33 • Linked Data and Semantic Web**

Chairperson: Guadalupe Aguado-de-Cea

#### *Nordurljos*

- 18:20-18:40 TMO - the Federated Ontology of the Trendminer Project  
*Hans-Ulrich Krieger and Thierry Declerck*
- 18:40-19:00 A Collection of Scholarly Book Reviews from the Platforms of Electronic Sources in Humanities and Social Sciences OpenEdition.org  
*Chahinez Benkoussas, Hussam Hamdan, Patrice Bellot, Frédéric Béchet and Elodie Faath*
- 19:00-19:20 A Meta-data Driven Platform for Semi-automatic Configuration of Ontology Mediators  
*Manuel Fiorelli, Maria Teresa Pazienza and Armando Stellato*

### **O34 • Dialogue (1)**

Chairperson: Francesco Cutugno

#### *Silfurberg A*

- 18:20-18:40 Twente Debate Corpus - A Multimodal Corpus for Head Movement Analysis  
*Bayu Rahayudi, Ronald Poppe and Dirk Heylen*
- 18:40-19:00 A Multimodal Corpus of Rapid Dialogue Games  
*Maike Paetzel, David Nicolas Racca and David DeVault*
- 19:00-19:20 The Tutorbot Corpus - A Corpus for Studying Tutoring Behaviour in Multiparty Face-to-Face Spoken Dialogue  
*Maria Koutsombogera, Samer Al Moubayed, Bajibabu Bollepalli, Ahmed Hussen Abdelaziz, Martin Johansson, José David Aguas Lopes, Jekaterina Novikova, Catharine Oertel, Kalin Stefanov and Gül Varol*

### **O35 • Word Sense Annotation and Disambiguation**

Chairperson: Patrik Lambert

#### *Silfurberg B*

- 18:20-18:40 Exploiting Portuguese Lexical Knowledge Bases for Answering Open Domain Cloze Questions Automatically  
*Hugo Gonçalo Oliveira, Inês Coelho and Paulo Gomes*
- 18:40-19:00 Single Classifier Approach for Verb Sense Disambiguation Based on Generalized Features  
*Daisuke Kawahara and Martha Palmer*
- 19:00-19:20 Annotating the MASC Corpus with BabelNet  
*Andrea Moro, Roberto Navigli, Francesco Maria Tucci and Rebecca J. Passonneau*

### **O36 • Legal and Ethical Issues**

Chairperson: Christopher Cieri

#### *Kaldalon*

- 18:20-18:40 The Liability of Service Providers in e-Research Infrastructures: Killing the Messenger?  
*Pawel Kamocki*
- 18:40-19:00 Evaluating Corpora Documentation with regards to the Ethics and Big Data Charter  
*Alain Couillault, Karën Fort, Gilles Adda and Hugues de Mazancourt*
- 19:00-19:20 Disclose Models, Hide the Data - How to Make Use of Confidential Corpora without Seeing Sensitive Raw Data  
*Erik Faessler, Johannes Hellrich and Udo Hahn*

**P45 • Anaphora and Coreference**

Chairperson: Costanza Navarretta

Corpus for Coreference Resolution on Scientific Papers

*Panot Chaimongkol, Akiko Aizawa and Yuka Tateisi*

ParCor 1.0: a Parallel Pronoun-Coreference Corpus to Support Statistical MT

*Liane Guillou, Christian Hardmeier, Aaron Smith, Jörg Tiedemann and Bonnie Webber*

The DARE Corpus: a Resource for Anaphora Resolution in Dialogue Based Intelligent Tutoring Systems

*Nobal Niraula, Vasile Rus, Rajendra Banjade, Dan Stefanescu, William Baggett and Brent Morgan*

CROMER: a Tool for Cross-Document Event and Entity Coreference

*Christian Girardi, Manuela Speranza, Rachele Sprugnoli and Sara Tonelli*

Coreference Resolution for Latvian

*Arturs Znotins and Peteris Paikens*

An Exercise in Reuse of Resources: Adapting General Discourse Coreference Resolution for Detecting Lexical Chains in Patent Documentation

*Nadjet Bouayad-Agha, Alicia Burga, Gerard Casamayor, Joan Codina, Rogelio Nazar and Leo Wanner*

The Extended DIRNDL Corpus as a Resource for Coreference and Bridging Resolution

*Anders Björkelund, Kerstin Eckart, Arndt Riestler, Nadja Schaffler and Katrin Schweitzer*

Multilingual Corpora with Coreferential Annotation of Person Entities

*Marcos Garcia and Pablo Gamallo*

Polish Coreference Corpus in Numbers

*Maciej Ogrodniczuk, Mateusz Kopeć and Agata Savary*

**P46 • Information Extraction and Information Retrieval**

Chairperson: Dimitrios Kokkinakis

French Resources for Extraction and Normalization of Temporal Expressions with HeidelTime

*Véronique Moriceau and Xavier Tammier*

Multilingual Test Sets for Machine Translation of Search Queries for Cross-Lingual Information Retrieval in the Medical Domain

*Zdenka Uresova, Jan Hajic, Pavel Pecina and Ondrej Dusek*

Semantic Approaches to Software Component Retrieval with English Queries

*Huijing Deng and Grzegorz Chrupala*

Annotating Relation Mentions in Tabloid Press

*Hong Li, Sebastian Krause, Feiyu Xu, Hans Uszkoreit, Robert Hummel and Veselina Mironova*

Construction of Diachronic Ontologies from People's Daily of Fifty Years

*Shaoda He, Xiaojun Zou, Liumingjing Xiao and Junfeng Hu*

Use of Unsupervised Word Classes for Entity Recognition: Application to the Detection of Disorders in Clinical Reports

*Maria Evangelia Chatzimina, Cyril Grouin and Pierre Zweigenbaum*

The Weltmodell: a Data-Driven Commonsense Knowledge Base  
*Alan Akbik and Thilo Michael*

Discovering and Visualising Stories in News  
*Marieke van Erp, Gleb Satyukov, Piek Vossen and Marit Nijssen*

A Large Scale Database of Strongly-related Events in Japanese  
*Tomohide Shibata, Shotaro Kohama and Sadao Kurohashi*

ClearTK 2.0: Design Patterns for Machine Learning in UIMA  
*Steven Bethard, Philip Ogren and Lee Becker*

## **P47 • Language Identification**

Chairperson: Michael Rosner

Vocabulary-based Language Similarity using Web Corpora  
*Dirk Goldhahn and Uwe Quasthoff*

Automatic Language Identity Tagging on Word and Sentence-Level in Multilingual Text Sources: a Case-Study on Luxembourgish  
*Thomas Lavergne, Gilles Adda, Martine Adda-Decker and Lori Lamel*

VarClass: an Open-Source Language Identification Tool for Language Varieties  
*Marcos Zampieri and Binyam Gebre*

Native Language Identification using Large, Longitudinal Data  
*Xiao Jiang, Yufan Guo, Jeroen Geertzen, Dora Alexopoulou, Lin Sun and Anna Korhonen*

On the Romance Languages Mutual Intelligibility  
*Liviu Dinu and Alina Maria Ciobanu*

## **P48 • Morphology**

Chairperson: Karel Pala

Heuristic Hyper-minimization of Finite State Lexicons  
*Senka Drobac, Krister Lindén, Tommi Pirinen and Miikka Silfverberg*

Crowd-sourcing Evaluation of Automatically Acquired, Morphologically Related Word Groupings  
*Claudia Borg and Albert Gatt*

Morphological Parsing of Swahili using Crowdsourced Lexical Resources  
*Patrick Littell, Kaitlyn Price and Lori Levin*

Chasing the Perfect Splitter: a Comparison of Different Compound Splitting Tools  
*Carla Parra Escartín*

Generating and using Probabilistic Morphological Resources for the Biomedical Domain  
*Vincent Claveau and Ewa Kijak*

Using Resource-Rich Languages to Improve Morphological Analysis of Under-Resourced Languages  
*Peter Baumann and Janet Pierrehumbert*

Turkish Treebank as a Gold Standard for Morphological Disambiguation and its Influence on Parsing  
*Ozlem Cetinoglu*

CroDeriV: a New Resource for Processing Croatian Morphology  
*Krešimir Šojat, Matea Srebačić, Marko Tadić and Tin Pavelić*

DerivBase.hr: a High-Coverage Derivational Morphology Resource for Croatian  
*Jan Šnajder*

Finite-State Morphological Transducers for Three Kypchak Languages  
*Jonathan Washington, Ilnar Salimzyanov and Francis Tyers*

### **P49 • Multimodality**

Chairperson: Volker Steinbiss

Representing Multimodal Linguistic Annotated Data  
*Brigitte Bigi, Tatsuya Watanabe and Laurent Prévot*

Single-Person and Multi-Party 3D Visualizations for Nonverbal Communication Analysis  
*Michael Kipp, Levin Freiherr von Hollen, Michael Christopher Hrstka and Franziska Zamponi*

The AV-LASYN Database: a Synchronous Corpus of Audio and 3D Facial Marker  
Data for Audio-Visual Laughter Synthesis  
*Huseyin Cakmak, Jerome Urbain, Thierry Dutoit and Joelle Tilmanne*

Linking Pictographs to Synsets: Sclera2Cornetto  
*Vincent Vandeghinste and Ineke Schuurman*

The MMASCS Multi-Modal Annotated Synchronous Corpus of Audio, Video, Facial  
Motion and Tongue Motion Data of Normal, Fast and Slow Speech  
*Dietmar Schabus, Michael Pucher and Phil Hoole*

Mining a Multimodal Corpus for Non-Verbal Behavior Sequences Conveying Attitudes  
*Mathieu Chollet, Magalie Ochs and Catherine Pelachaud*

The IMAGACT Visual Ontology. an Extendable Multilingual Infrastructure for the  
Representation of Lexical Encoding of Action  
*Massimo Moneglia, Susan Brown, Francesca Frontini, Gloria Gagliardi, Fahad Khan,  
Monica Monachini and Alessandro Panunzi*

Multimodal Dialogue Segmentation with Gesture Post-Processing  
*Kodai Takahashi and Masashi Inoue*

The D-ANS Corpus: the Dublin-Autonomous Nervous System Corpus of Biosignal and  
Multimodal Recordings of Conversational Speech  
*Shannon Hennig, Ryad Chellali and Nick Campbell*