| Time         | Session | Chair                  | Room               | Topic                                                                                              |
|-------------|---------|------------------------|--------------------|----------------------------------------------------------------------------------------------------|
| 9.00-9.40   | Keynote | Chair: Jan Odijk       |                    | Sören Auer: The Web of Data: Decentralized, collaborative, interlinked and interoperable            |
| 9.40-9.45   |         | 5 minutes break        |                    |                                                     |
| Session: P15 - Semantic Annotation | Chair: Aline Villavicencio | Poster Area (1) |
|-----------------------------------|---------------------------|----------------|
| Arnulphy Béatrice, Tannier Xavier and Vilnat Anne | Event Nominals: Annotation Guidelines and a Manually Annotated Corpus in French |
| Maria Aloni, Andreas van Cranenburgh, Raquel Fernandez and Marta Szajdjer | Building a Corpus of Indefinite Uses Annotated with Fine-grained Semantic Functions |
| Antônio Branco, Catarina Carvalheiro, Silvia Pereira, Sara Silveira, João Silva, Sérgio Castro and João Graça | A PropBank for Portuguese: the CINTIL-PropBank |
| Ashwini Vaidya, Jinho D. Choi, Martha Palmer and Bhuvana Narasimhan | Empty Argument Insertion in the Hindi PropBank |
| Pierre Bouillon, Elisabetta Jezek, Chiara Melloni and Aurélie Picton | Annotating Qualia Relations in Italian and French Complex Nominals |
| Juliette Thuillier and Laurence Danos | Semantic annotation of French corpora: animacy and verb semantic classes |
| Josef Ruppenhofer and Ines Rehbein | Yes we can!? Annotating English modal verbs |
| Mehdi Manshadi, James Allen and Mary Swift | An Annotation Scheme for Quantifier Scope Disambiguation |
| Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa | Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure |
| Kyoko Ohara | Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English |

| Session: P16 - Document Classification, Text Categorisation | Chair: Serge Sharoff | Poster Area (1) |
|----------------------------------------------------------|---------------------|----------------|
| Mike Kestemont, Claudia Peersman, Benny De Decker, Guy De Pauw, Kim Luyckx, Roser Morante, Frederik Vaassen, Janneke van de Loo and Walter Daelemans | The Netlog Corpus. A Resource for the Study of Flemish Dutch Internet Language |
| Kseniya Zablotskaya, Fernando Fernández Martínez and Wolfgang Minker | Investigating Verbal Intelligence Using the TF-IDF Approach |
| Sanja Štajner and Ruslan Mitkov | Diachronic Changes in Text Complexity in 20th Century English Language: An NLP Approach |
| Tommaso Fornaciari and Massimo Poesio | DeCour: a corpus of DEceptive statements in Italian COURts |
| Amalia Todirascu, Sebastian Pado, Jennifer Krisch, Max Kisselew and Ulrich Heid | French and German Corpora for Audience-based Text Type Classification |
| Borut Sluban, Senja Polliak, Roel Coesmans and Nada Lavrac | Irregularity Detection in Categorized Document Corpora |
| Carmen Dayrell, Arnaldo Candido Jr., Gabriel Lima, Danilo Machado Jr., Ann Copestake, Valèria Feltrim, Stella Tagnin and Sandra Aluisio | Rhetorical Move Detection in English Abstracts: Multi-label Sentence Classifiers and their Annotated Corpora |
| Andrea Varga, Daniel Preotiuc-Pietro and Fabio Ciravegna | Unsupervised document zone identification using probabilistic graphical models |

| Session: P17 - Grammar and Syntax | Chair: Eleni Efthimiou | Poster Area (1) |
|-----------------------------------|-----------------------|----------------|
| Erhard Hinrichs and Thomas Zastrow | Automatic Annotation and Manual Evaluation of the Diachronic German Corpus TüBa-D/DC |
| Authors | Title |
|---------|-------|
| Hongsuck Seo, Kyusong Lee, Gary Geunbae Lee, Soo-Ok Kweon and Hae-Ri Kim | Grammatical Error Annotation for Korean Learners of Spoken English |
| Heiki-Jaan Kaalep and Kadri Muischnek | Robust clause boundary identification for corpus annotation |
| Patrick Ziering, Sina Zarrieß and Jonas Kuhn | A Corpus-based Study of the German Recipient Passive |
| Zygmunt V etulani | Wordnet Based Lexicon Grammar for Polish |
| Montserrat Arza, José M. García-Miguel, Francisco Campillo and Miguel Cuevas-Alono | A Galician Syntactic Corpus with Application to Intonation Modeling |
| Hiroaki Sato | A Search Tool for FrameNet Constructicon |
| Markus Dickinson and Scott Ledbetter | Annotating Errors in a Hungarian Learner Corpus |
| Stefan Bott, Horacio Saggion and Simon Mille | Text Simplification Tools for Spanish |
| Antske Fokkens, Tania Avgustinova and Yi Zhang | CLIMB grammars: three projects using metagrammar engineering |
| Peteris Paikens and Normunds Gruzitis | An implementation of a Latvian resource grammar in Grammatical Framework |
| Shaqfat Mumtaz Virk and Elnaz Abolahar | An Open Source Persian Computational Grammar |
| Paula Butterly and Andrew Caines | Reclassifying subcategorization frames for experimental analysis and stimulus generation |
| Andrew Caines and Paula Butterly | Annotating progressive aspect constructions in the spoken section of the British National Corpus |

**Session: P18 - Digital Libraries**

Chair: Monica Monachini

| Authors | Title |
|---------|-------|
| Jordi Adell, Antonio Bonafonte, Antonio Cardenal, Marta R. Costa-jussà, José A. R. Fonollosa, Asunción Moreno, Eva Navas and Eduardo R. Banga | BUCEADOR, a multi-language search engine for digital libraries |
| Ranka Stanković, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić | A tool for enhanced search of multilingual digital libraries of e-journals |
| Benjamin Weitz and Ulrich Schäfer | A Graphical Citation Browser for the ACL Anthology |
| Eleftheria Ahtaridis, Christopher Cieri and Denise DiPersio | LDC Language Resource Database: Building a Bibliographic Database |
| Eneko Agirre, Ander Barrena, Oier Lopez de Lacalle, Aitor Soroa, Samuel Fernando and Mark Stevenson | Matching Cultural Heritage items to Wikipedia |

**11.25-11.45 Coffee break**

**Session: O21 - Speech Corpora and Tools**

Chair: Robrecht Comeyne

| Time | Authors | Title |
|------|---------|-------|
| 11.45-12.05 | Maria Eskevich, Gareth J.F. Jones, Martha Larson and Roeland Ordelman | Creating a Data Collection for Evaluating Rich Speech Retrieval |
| 12.05-12.25 | Petya Osenova and Kirl Simov | The Political Speech Corpus of Bulgarian |
| 12.25-12.45 | Brigite Bigi | SPPAS: a tool for the phonetic segmentation of speech |
| 12.45-13.05 | Brigite Bigi, Pauline Péri and Roxane Bertrand | Orthographic Transcription: which enrichment is required for phonetization? |

**Session: O22 - Machine Translation and Evaluation (2)**

Chair: François Yvon

| Time | Authors | Title |
|------|---------|-------|
| 11.45-12.05 | Sandra Weiss and Lars Ahrenberg | Error profiling for evaluation of machine-translated text: a Polish-English case study |
| Time          | Title                                                                 | Authors                                                                 | Details                                                      |
|--------------|----------------------------------------------------------------------|------------------------------------------------------------------------|--------------------------------------------------------------|
| 12.05-12.25  | Chunqi Shi, Donghui Lin, Masahiko Shimada and Toru Ishida            | Two Phase Evaluation for Selecting Machine Translation Services       |
| 12.25-12.45  | Lorenza Russo, Sharid Loaiciga and Asheesh Gulati                    | Italian and Spanish Null Subjects. A Case Study Evaluation in an MT Perspective |
| 12.45-13.05  | Sara Stymne and Lars Ahrenberg                                      | On the practice of error analysis for machine translation evaluation  |
| **Session: O23 - Semantic Resources** | Chair: Zygmunt Vetulani | Room: Topkapı B                                                      |
| 11.45-12.05  | Janine Pimentel                                                      | Identifying equivalents of specialized verbs in a bilingual comparable corpus of judgments: A frame-based methodology |
| 12.05-12.25  | Alessandra Zarcone and Stefan Rued                                  | Logical metonymies and qualia structures: an annotated database of logical metonymies for German |
| 12.25-12.45  | Iris Hendrickx, Amalia Mendes and Silvia Mencarelli                  | Modality in Text: a Proposal for Corpus Annotation                     |
| 12.45-13.05  | Pablo Mendes, Max Jakob and Christian Bizer                         | DBpedia: A Multilingual Cross-domain Knowledge Base                    |
| **Session: O24 - Trends in Corpora** | Chair: tba | Room: Dolmabahçe A                                               |
| 11.45-12.05  | Annie Louis and Ani Nenkova                                         | A corpus of general and specific sentences from news                   |
| 12.05-12.25  | Gozde Ozbal, Carlo Strapparava and Marco Guerini                    | Brand Pitt: A Corpus to Explore the Art of Naming                      |
| 12.25-12.45  | Jonathon Read, Dan Flickinger, Rebecca Dridan, Stephan Oepen and Lilja Øvrelid | The WeSearch Corpus, Treebank, and Treecache -- A Comprehensive Sample of User-Generated Content |
| 12.45-13.05  | Masashi Inoue and Toshiki Akagi                                     | Collecting humorous expressions from a community-based question-answering-service corpus |
| **Poster Sessions: 11.45-13.05** | | Poster Area (2)                                               |
| **Session: P19 - Treebanks** | Chair: Menno van Zaanen | |
| 12.05-12.25  | Seth Kulick, Ann Bies and Justin Mott                               | Further Developments in Treebank Error Detection Using Derivation Trees |
| 12.25-12.45  | Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies and Nianwen Xue | Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures |
| 12.45-13.05  | Mohamed Maamouri, Ann Bies and Seth Kulick                           | Expanding Arabic Treebank to Speech: Results from Broadcast News       |
| 12.25-12.45  | Magali Sanches Duran and Sandra Maria Alustio                        | Propbank-BR: a Brazilian Treebank annotated with semantic role labels   |
| 12.45-13.05  | Yi Zhang, Rui Wang and Yu Chen                                      | Joint Grammar and Treebank Development for Mandarin Chinese with HPSG  |
| 12.05-12.25  | Annette Rios and Anne Göhring                                       | A tree is a Baum is an árbol is a sach’a: Creating a trilingual treebank |
| 12.05-12.25  | Miriam Kaeshammer and Vera Demberg                                   | German and English Treebanks and Lexica for Tree-Adjoining Grammars    |
| 12.05-12.25  | Loganathan Ramasamy and Zdeněk Žabokrtský                             | Prague Dependency Style Treebank for Tamil                             |
| 12.05-12.25  | Patricia Gonçalves, Rita Santos and António Branco                   | Treebanking by Sentence and Tree Transformation: Building a Treebank to support Question Answering in Portuguese |
| 12.05-12.25  | Dasa Berovic, Zeljko Agic and Marko Tadic                            | Croatian Dependency Treebank: Recent Development and Initial Experiments |
| 12.25-12.45  | Rahul Agarwal, Bharat Ram Ambai and Anil Kumar Singh                 | A GUI to Detect and Correct Errors in Hindi Dependency Treebank        |
| 12.25-12.45  | Masood Ghayoomi                                                      | From Grammar Rule Extraction to Treebanking: A Bootstrapping Approach  |
| 12.25-12.45  | Montserrat Marimon, Beatriz Fisas, Núria Bel, Marta Villegas, Jorge Vivaldi, Sergei Torner, Mercé Lorente, Silvia Vázquez and Marta Villegas | The IULA Treebank                                                       |
| 12.25-12.45  | Atro Voutilainen, Kristiina Muhonen, Tanja Purtonen and Krister Linden | Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3            |
| Session: P20 - Parsing | Chair: Antonio Moreno-Sandoval |
|------------------------|-------------------------------|
| Cristina Bosco, Manuela Sanguinetti and Leonardo Lesmo | The Parallel-TUT: a multilingual and multiformat treebank |
| Teresa Lynn, Ozlem Cetinoglu, Jennifer Foster, Elaine Uí Dhonnchadhá, Mark Dras and Josef van Genabith | Irish Treebanking and Parsing: A Preliminary Evaluation |
| Mohammed Attia, Khaled Shaalan, Lamia Tounsi and Josef van Genabith | Automatic Extraction and Evaluation of Arabic LFG Resources |
| Kristiina Muuronen and Tanja Purtonen | Rule-Based Detection of Clausal Coordinate Ellipsis |
| Gulsen Eryigit | The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish |
| Stasinios Konstantopoulos, Valia Kordoni, Nicola Cancedda, Vangelis Karkaletsis, Dietrich Klakow and Jean-Michel Renders | Task-Driven Linguistic Analysis based on an Underspecified Features Representation |
| Malin Ahlberg and Ramona Enache | "Combining Language Resources Into A Grammar-Driven Swedish Parser" |
| Eiríkur Rögnvaldsson, Anton Karl Ingason, Einar Freyr Sigurðsson and Joel Wallenberg | The Icelandic Parsed Historical Corpus (IcePaHC) |
| Anita Alicante, Cristina Bosco, Anna Corazza and Alberto Lavelli | A treebank-based study on the influence of Italian word order on parsing performance |
| Dong Wang and Fei Xia | Effort of Genre Variation and Prediction of System Performance |

| Session: P21 - Information Extraction (2) | Chair: Paul Buitelaar |
|----------------------------------------|---------------------|
| Michael Tepper, Daniel Capurro, Fei Xia, Lucy Vanderwende and Meliha Yetisgen-Yıldız | Statistical Section Segmentation in Free-Text Clinical Records |
| Ramona Bongelli, Carla Canestrari, Baria Riccioni, Andrzej Zuczkowski, Cinzia Buldorini, Ricardo Pietrobon, Alberto Lavelli and Bernardo Magnini | A Corpus of Scientific Biomedical Texts Spanning over 168 Years Annotated for Uncertainty |
| Cristina Mota, Alberto Simões, Cláudia Freitas, Luís Costa and Diana Santos | Págico: Evaluating Wikipedia-based information retrieval in Portuguese |
| Danica Damljanovic, Udo Kruschwitz, M-Dyaa Albakour, Johann Petrak and Mihai Lupu | Applying Random Indexing to Structured Data to Find Contextually Similar Words |
| Horacio Saggion and Sandra Szasz | The CONCISUS Corpus of Event Summaries |
| Gracinda Carvalho, David Martins de Matos and Vitor Rocio | Building and Exploring Semantic Equivalences Resources |
| Marc Verhagen and James Pustejovsky | The TARSQI Toolkit |
| Gabriella Pardelli, Manuela Sassi, Sara Goggi and Stefania Biagioni | From medical language processing to BioNLP domain |
| Romaric Besançon, Olivier Ferret and Ludovic Jean-Louis | Evaluation of a Complex Information Extraction Application in Specific Domain |
| Hannah Kermes | A methodology for the extraction of information about the usage of formulaic expressions in scientific texts |
| André Santos, José João Almeida and Nuno Carvalho | Structural alignment of plain text books |
| Gerald Schneider, Fabio Rinaldi and Simon Clematide | Dependency parsing for interaction detection in pharmacogenomics |
William Black, Rob Procter, Steven Gray and Sophia Ananiadou  
A data and analysis resource for an experiment in text mining a collection of micro-blogs on a political topic

13.05-13.10 5 minutes break

13.10-13.30 Invited Talk
Chair: Mehmed Özkan
Kemal Oflazer
The Turkish Language and its Challenges for Language Processing

13.30-14.55 Lunch break

| Session: O25 - Multimodal Corpora (2) | Chair: Nick Campbell | Room: Anadolu Auditorium |
|---------------------------------------|----------------------|--------------------------|
| 14.55-15.15                           | Ibrahim Saygin Topkaya and Hakan Erdogan | SUTAV: A Turkish Audio-Visual Database |
| 15.15-15.35                           | Costanza Navarretta and Patrizia Paggio | Multimodal Behaviour and Feedback in Different Types of Interaction |
| 15.35-15.55                           | Carlo Strapparava, Rada Mihalcea and Alberto Battocchi | A Parallel Corpus of Music and Lyrics Annotated with Emotions |
| 15.55-16.15                           | Merlin Teodosia Suarez, Jocelynn Cu and Madeleine Sta. Maria | Building a Multimodal Laughter Database for Emotion Recognition |
| 16.15-16.35                           | Dimitra Anastasiou | A Speech and Gesture Spatial Corpus in Assisted Living |

| Session: O26 - Child Language Corpus | Chair: Massimo Poesio | Room: Topkapı A |
|-------------------------------------|----------------------|----------------|
| 14.55-15.15                         | Priti Aggarwal, Ron Artstein, Jillian Gerten, Athanasios Katsamanis, Shrikanth Narayan, Angela Nazarian and David Traum | The Twins Corpus of Museum Visitor Questions |
| 15.15-15.35                         | Hyejin Hong, Sunhee Kim and Minhwa Chung | Korean Children's Spoken English Corpus and an Analysis of its Pronunciation Variability |
| 15.35-15.55                         | Marie Tahan, Agnes Delaborde and Laurence Devillers | Corpus of Children Voices for Mid-level Markers and Affect Bursts Analysis |
| 15.55-16.15                         | Aline Villavicencio, Beracah Yankama, Marco Iliart and Robert Berwick | A large scale annotated child language construction database |
| 16.15-16.35                         | Brian MacWhinney | Morphosyntactic Analysis of the CHILDES and TalkBank Corpora |

| Session: O27 - MultiWord Expressions | Chair: Emanuele Pianta | Room: Topkapı B |
|-------------------------------------|----------------------|----------------|
| 14.55-15.15                         | Veronika Vincze | Light Verb Constructions in the Szeged ParallelFX English–Hungarian Parallel Corpus |
| 15.15-15.35                         | Antoni Gurrutxaga and Iñaki Alegria | Measuring the compositionalty of NV expressions in Basque by means of distributional similarity techniques |
| 15.35-15.55                         | Marion Weller and Ulrich Heid | Analyzing and Aligning German compound nouns |
| 15.55-16.15                         | Natalia Loukachevitch | Automatic Term Recognition Needs Multiple Evidence |
| 16.15-16.35                         | Roman Kurc, Maciej Piasecki and Bartosz Broda | Constraint Based Description of Polish Multiword Expressions |

| Session: O28 - Sign Language | Chair: Javier Camanero | Room: Dolmabahçe A |
|-----------------------------|----------------------|----------------|
| 14.55-15.15                 | Dimitris Metaxas, Bo Liu, Fei Yang, Peng Yang, Nicholas Michael and Carol Neidle | Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking |
| 15.15-15.35                 | Matti Karppa, Tommi Jantunen, Ville Vitianiemi, Jorma Laaksonen, Birgitta Burger and Danny De Weerdt | Comparing computer vision analysis of signed language video with motion capture recordings |
| 15.35-15.55                 | Annelies Braffort and Leila Boutora | DEGELS1: A comparable corpus of French Sign Language and co-speech gestures |
| Time       | Presenter | Title                                                                 |
|------------|-----------|----------------------------------------------------------------------|
| 15.55-16.15 | Matilde GONZALEZ, Michael FILHOL and Christophe COLLET | Semi-Automatic Sign Language Corpora Annotation using Lexical Representations of Signs |
| 16.15-16.35 | Umar Shoaib, Nadeem Ahmad, Paolo Prinetto and Gabriele Tiotto | A platform-independent user-friendly dictionary from Italian to LIS |

**Poster Sessions: 14.55-16.35**

**Poster Area (1)**

**Session: P22 - Part-of-Speech Tagging**

**Chair: Reinhard Rapp**

- Slav Petrov, Dipanjan Das and Ryan McDonald: A Universal Part-of-Speech Tagset
- Atro Vuotilaainen: Improving corpus annotation productivity: a method and experiment with interactive tagging
- Andrea Gesmundo and Tanja Samardzic: Lemmatising Serbian as Category Tagging with Bidirectional Sequence Classification
- Souhir Gahbiche-Braham, Hélène Bonneau-Maynard, Thomas Lavergne and François Yvon: Joint Segmentation and POS Tagging for Arabic Using a CRF-based Classifier
- Mans Hulden and Jerid Francom: Boosting statistical tagger accuracy with simple rule-based grammars
- Maarten Janssen: NeoTag: a POS Tagger for Grammatical Neologism Detection
- Francesco Rubino, Francesca Frontini and Valeria Quochi: Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser

**Session: P23 - Machine Translation (1)**

**Chair: Philippe Langlais**

- Bruno Cartoni and Thomas Meyer: Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies
- Marianna J. Martindale: Can Statistical Post-Editing with a Small Parallel Corpus Save a Weak MT Engine?
- Sanja Seljan, Marija Brkić and Tomislav Vikić: BLEU Evaluation of Machine-Translated English-Croatian Legislation
- Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi: Chinese Characters Mapping Table of Japanese, Traditional Chinese and Simplified Chinese
- Juan Pablo Martínez Cortés, Jim O’Regan and Francis Tyers: Free/Open Source Shallow-Transfer Based Machine Translation for Spanish and Aragonese
- Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović and Daniel Zeman: Automatic MT Error Analysis: Hjerson Helping Addicter
- Amit Sangodkar and Om Damani: Re-ordering Source Sentences for SMT
- Ángela Costa, Tiago Luís, Joana Ribeiro, Ana Cristina Mendes and Luisa Coheur: An English-Portuguese parallel corpus of questions: translation guidelines and application in SMT
- Mehmet Talha Çakmak, Süleyman Acar and Gülşen Eryiğit: Word Alignment for English-Turkish Language Pair
- Radu Ion: PEXACC: A Parallel Sentence Mining Algorithm from Comparable Corpora
- Eleftherios Avramidis, Marta R. Costa-Jussá, Christian Federmann, Josef van Genabith, Maite Melero and Pavel Pecina: A Richly Annotated, Multilingual Parallel Corpus for Hybrid Machine Translation
- Stephen Grimes, Katherine Peterson and Xuansong Li: Automatic word alignment tools to scale production of manually aligned parallel texts
- Carla Parra Escartín: Design and compilation of a specialized Spanish-German parallel corpus
- Jörg Tiedemann, Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen and Matthias Zump: A Distributed Resource Repository for Cloud-Based Machine Translation

**Session: P24 - Corpus Creation, Processing, Usage (1)**

**Chair: Takenobu Tokunaga**

- Jörg Tiedemann: Parallel Data. Tools and Interfaces in OPUS
| Title                                                                 | Authors                                                                                                                                                                                                 | Abstract                                                                                                                                                                                                 |
|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Maciej Ogrodniczuk: The Polish Sejm Corpus                           | Lieve Macken, Veronique Hoste, Marielle Leijten and Luuk Van Waes                                                                                                                                       | From keystrokes to annotated process data: Enriching the output of Inputlog with linguistic information                                                                                                         |
| Maristella Agosti, Birgit Alber, Giorgio Maria Di Nunzio, Marco Dussin, Stefan Rabanus and Alessandra Tomaselli: A Curated Database for Linguistic Research: The Test Case of Cimbrian Varieties |                                                                                               |                                                                                                                                                                                                    |
| Michiel Génèreux, Iris Hendricks and Amália Mendes: Introducing the Reference Corpus of Contemporary Portuguese Online | Mojgan Seraji, Beáta Megyesi and Joakim Nivre                                                                                                                                                    | A Basic Language Resource Kit for Persian                                                                                                                                                              |
| Eric Sanders: Collecting and Analysing Chats and Tweets in SoNaR      |                                                                                               |                                                                                                                                                                                                    |
| Tomaz Erjavec: The goo300k corpus of historical Slovene               | Mathieu-Henri Falco, Véronique Moriceau and Anne Vilnat                                                                                                                                          | Kitten: a tool for normalizing HTML and extracting its textual content                                                                                                                                |
| Maristella Agosti, Birgit Alber, Giorgio Maria Di Nunzio, Marco Dussin, Stefan Rabanus and Alessandra Tomaselli: A Curated Database for Linguistic Research: The Test Case of Cimbrian Varieties |                                                                                               |                                                                                                                                                                                                    |
| Alessandro Panunzi, Marco Fabbri, Massimo Monegía, Lorenzo Gregori and Samuele Paladini: A Basic Language Resource Kit for Persian |                                                                                               |                                                                                                                                                                                                    |
| Brett Drury and Jose Joao Almeida: The Minho Quotation Resource       |                                                                                               |                                                                                                                                                                                                    |
| Elena Frick, Carsten Schnober and Piotr Bański: Evaluating Query Languages for a Corpus Processing System |                                                                                               |                                                                                                                                                                                                    |
| Session: P25 - Evaluation Methodologies                              | Mathieu Lafourcade                                                                            |                                                                                                                                                                                                    |
| Abdul-Baquee Sharaf and Eric Atwell: QurSim: A corpus for evaluation of relatedness in short texts | Christina Feilmayr, Birgit Pröll and Elisabeth Linsmayer: EVALJEX – A Proposal for an Extended Evaluation Methodology for Information Extraction Systems   |                                                                                                                                                                                                    |
| Patrick Paroubeck and Xavier Tanner: A Rough Set Formalization of Quantitative Evaluation with Ambiguity | Thomas Eckart, Uwe Quasthoff and Dirk Goldhahn: The Influence of Corpus Quality on Statistical Measurements on Language Resources |                                                                                                                                                                                                    |
| Olga Babko-Malaya, Greg Milette, Michael Schneider and Sarah Scogin: Identifying Nuggets of Information in GALE Distillation Evaluation | Chieh-Jen Wang, Shuk-Man Cheng, Lung-Hao Lee, Hsin-Hsi Chen, Wen-shen Liu, Pei-Wen Huang and Shih-Peng Lin: NTUSocialRec: An Evaluation Dataset Constructed from Microblogs for Recommendation Applications in Social Networks |                                                                                                                                                                                                    |
| Chieh-Jen Wang, Shuk-Man Cheng, Lung-Hao Lee, Hsin-Hsi Chen, Wen-shen Liu, Pei-Wen Huang and Shih-Peng Lin: NTUSocialRec: An Evaluation Dataset Constructed from Microblogs for Recommendation Applications in Social Networks |                                                                                               |                                                                                                                                                                                                    |

16.35-16.55 Coffee break

| Session: O29 - Language Generation and Paraphrasing                 | Chair: Robert Dale                                                                         | Room: Anadolu Auditorium                                                                                                                                                                                                 |
|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 16.55-17.15                                                         | Houda Bouamour, Aurélien Max, Gabriel Ilouz and Anne Vilnat: A contrastive review of paraphrase acquisition techniques |                                                                                                                                                                                                                     |
| 17.15-17.35                                                         | Matteo Negri, Yashar Mehdad, Alessandro Marchetti, Danilo Giampiccolo and Luisa Bentivogli: Chinese Whispers: Cooperative Paraphrase Acquisition |                                                                                                                                                                                                                     |
| 17.35-17.55                                                         | Hideki Shima and Teruko Mitamura: Diversifiable Bootstrapping for Acquiring High-Coverage Paraphrase Resource |                                                                                                                                                                                                                     |
| 17.55-18.15                                                         | Sebastian Varges, Heike Bieler, Manfred Stede, Lukas C. Faulstich, Kristin Irsig and Malik Atalla: SemScribe: Natural Language Generation for Medical Reports |                                                                                                                                                                                                                     |
| Session: O30 - Computer Aided Language Learning                     | Chair: Justus Roux                                                                          | Room: Topkapı A                                                                                                                                                                                                     |

| Time       | Presenters/Participants                                                                 | Title                                                                 |
|------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| 16.55-17.15| Hitokazu Matsushita and Deryle Lonsdale                                                  | Item Development and Scoring for Japanese Oral Proficiency Testing   |
| 17.15-17.35| Manny Rayner, Pierrette Bouillon and Johanna Gerlach                                    | Evaluating Appropriateness Of System Responses In A Spoken CALL Game |
| 17.35-17.55| Antonio Moreno-Sandoval, Leonardo Campillos, Yang Dong, Emi Takamori, José M. Guirao, Paula Gozalo, Chieko Kimura, Kengo Matsui and Marta Garrote | Spontaneous Speech Corpora for language learners of Spanish, Chinese and Japanese |
| 17.55-18.15| Helmer Strik, Jozef Colpaert, Joost Van Doremalen and Catia Cucchiari                   | The DISCO ASR-based CALL system: practicing L2 oral skills and beyond |
|            | **Session: O31 - Discourse (2)**                                                        | **Room:** Topkapı B                                                  |
|            | **Chair:** Aravind Joshi                                                                | **A Tool for Extracting Conversational Implications**                |
| 16.55-17.15| Marta Tatu and Dan Moldovan                                                             | **Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns** |
| 17.15-17.35| Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni and Sandrine Zufferley | **Annotating Story Timelines as Temporal Dependency Structures**     |
| 17.35-17.55| Steven Bethard, Olekxandr Kolomiëts and Marie-Francine Moërs                            | **An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus** |
| 17.55-18.15| Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cecile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paul Perry-Woodley, Laurent Prevet, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret and Laure Vieu | **HamleDT: To Parse or Not to Parse?**                                |
| 17.15-17.35| Elsa Tolone, Benoît Sagot and Éric Villemonte de la Clergerie                          | **Evaluating and improving syntactic lexica by plugging them within a parser** |
| 17.35-17.55| Thomas Proisl and Peter Uhrig                                                           | **Efficient Dependency Graph Matching with the IMS Open Corpus Workbench** |
| 17.55-18.15| Miguel Ballesteros and Joakim Nivre                                                     | **MaltOptimizer: A System for MaltParser Optimization**               |
|            | **Session: O32 - Syntax and Parsing**                                                   | **Room:** Dolmabahçe A                                               |
| 16.55-17.15| Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Zabokrtský and Jan Hajič | **HamleDT: To Parse or Not to Parse?**                                |
| 17.15-17.35| Elsa Tolone, Benoît Sagot and Éric Villemonte de la Clergerie                          | **Evaluating and improving syntactic lexica by plugging them within a parser** |
| 17.35-17.55| Thomas Proisl and Peter Uhrig                                                           | **Efficient Dependency Graph Matching with the IMS Open Corpus Workbench** |
| 17.55-18.15| Miguel Ballesteros and Joakim Nivre                                                     | **MaltOptimizer: A System for MaltParser Optimization**               |
|            | **Poster Sessions: 16.55-18.15**                                                        | **Poster Area (2)**                                                  |
| 16.55-18.15| **Session: P26 - Multilinguality**                                                      | **Poster Area (2)**                                                  |
|            | **Chair:** Gil Francopoulo                                                               | **Representing the Translation Relation in a Bilingual Wordnet**      |
|            | Jyrki Niemi and Krister Lindén                                                          | **Building a multilingual parallel corpus for human users**          |
|            | Alexandr Rosen and Martin Vavrin                                                       | **Hungar: A Hungarian–Russian Parallel Corpus**                       |
|            | Martina Katalin Szabó, Veronika Vincze and István Nagy T.                               | **Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics** |
|            | Gilles Sérasset                                                                        | **Dbnary: Wiktionary as a LMF based Multilingual RDF network**       |
|            | Lluís Padró and Evgeny Stanilovský                                                     | **FreeLing 3.0: Towards Wider Multilinguality**                      |
|            | Svetla Koeva, Ivelina Stoyanova, Rositsa Dekova, Borislav Rizov and Angel Genov         | **Bulgarian X-language Parallel Corpus**                             |
|            | Enakő Héja and Dávid Takács                                                           | **Automatically Generated Online Dictionaries**                       |
| Session: P27 - Question Answering and Summarisation | Chair: Horacio Saggion |
|---------------------------------------------------|------------------------|
| Christian Smith, Henrik Danielsson and Arne Jönsson | A good space: Lexical predictors in word space evaluation |
| Ulrich Andersen, Anna Braasch, Lina Henrikse, Csaba Huszka, Anders Johamnmen, Lars Kayser, Bente Maegaard, Ole Norgaard, Stefan Schulz and Jürgen Wedekind | Creation and use of Language Resources in a Question-Answering eHealth System |

| Effect of Document Clustering in Modeling Wikipedia-style Term Descriptions |
| Silvia Quarteroni, Vincenzo Guerrisi and Pietro La Torre |

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

| Session: P28 - Multimodal Corpus for Interaction | Chair: Vangelis Karkaletsis |
|---------------------------------------------------|------------------------|
| Dietmar Rösner, Jörg Frommer, Rafael Friesen, Matthias Haase, Julia Lange and Mirko Otto | LAST MINUTE: a Multimodal Corpus of Speech-based User-Companion Interactions |
| Karén Fort and Vincent Claveau | Annotating Football Matches: Influence of the Source Medium on Manual Annotation |
| Stephanie Strassel, Amanda Morris, Jonathan Fiscus, Christopher Caruso, Heojoong Lee, Paul Over, James Fiumara, Barbara Shaw, Brian Antonishek and Martial Michel | Creating HAVIC: Heterogeneous Audio Visual Internet Collection |

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

| Session: 29 - Ontologies | Chair: Paola Velardi |
|---------------------------------------------------|------------------------|
| Egoitz Laparra, German Rigau and Piek Vossen | Mapping WordNet to the Kyoto ontology |
| Kugatsu Sadamitsu, Kuniko Saito, Kenji Imamura and Yoshihiko Matsuo | Constructing a Class-Based Lexical Dictionary using Interactive Topic Models |
| Verginica Barbu Mititelu | Adding Morpho-semantic Relations to the Romanian Wordnet |
| Julien Scintuier, Elisabeth Murissasco, Emmanuel Bruno and Philippe Blache | An ontological approach to model and query multimodal concurrent linguistic annotations |
| Massimo Moneglia, Monica Monachini, Omar Calabrese, Alessandro Panunzi, Francesca Frontini, Gloria Gagliardi and Irene Russo | The IMAGACT Cross-linguistic Ontology of Action. A new infrastructure for natural language disambiguation |
| Inga Gheorghita and Jean-Marie Pierrel | Towards a methodology for automatic identification of hypernyms in the definitions of large-scale dictionary |
| John McCrae, Elena Montiel-Ponsoda and Philipp Cimiano | Collaborative semantic editing of linked data lexica |
|--------------------------------------------------------|---------------------------------------------------|
| Christophe Roche                                      | Ontoterminology: How to unify terminology and ontology into a single paradigm |
| Alexandre Denis, Ingrid Falk, Claire Gardent and Laura Perez-Beltrachini | Representation of linguistic and domain knowledge for second language learning in virtual worlds |
| Petya Osenova, Kiril Simov, Laska Laskova and Stanislava Kancheva | A Treebank-driven Creation of an OntoValence Verb lexicon for Bulgarian |
| Elisa Bianchi, Mirko Tavosanis and Emiliano Giovannetti | Creation of a bottom-up corpus-based ontology for Italian Linguistics |
| Matteo Abrate and Clara Bacchi | Visualizing word senses in WordNet Atlas |

5 minutes break

| Session: O33 - Semantics from Corpora | Chair: Maria Teresa Pazienza | Room: Anadolu Auditorium |
|-------------------------------------|-------------------------------|--------------------------|
| 18.20-18.40 Alex Judea, Vivi Nastase and Michael Strube | Concept-based Selectional Preferences and Distributional Representations from Wikipedia Articles |
| 18.40-19.00 Elias Iosif, Maria Giannoudaki, Eric Fosler-Lussier and Alexandros Potamianos | Associative and Semantic Features Extracted From Web-Harvested Corpora |
| 19.00-19.20 Octavian Popescu | Buildind a Resource of Patterns Using Semantic Types |

| Session: O34 - Authoring and Related Tools | Chair: Bernardo Magnini | Room: Topkapi A |
|--------------------------------------------|-------------------------|------------------|
| 18.20-18.40 Irina Temnikova, Constantin Orasan and Ruslan Mitkov | CLCM - A Linguistic Resource for Effective Simplification of Instructions in the Crisis Management Domain and its Evaluations |
| 18.40-19.00 Robert Dale and George Narroway | A Framework for Evaluating Text Correction |
| 19.00-19.20 Paul Rodrigues and C. Anton Rytting | Typing Race Games as a Method to Create Spelling Error Corpora |

| Session: O35 - Word Sense Annotation and Disambiguation | Chair: Yoshiliko Hayashi | Room: Topkapi B |
|----------------------------------------------------------|-------------------------|------------------|
| 18.20-18.40 Rebecca Passonneau, Collin Baker, Christiane Fellbaum and Nancy Ide | The MASC Word Sense Corpus |
| 18.40-19.00 Darja Fišer, Nikola Ľubešić and Ozren Kubelka | Addressing polysemy in bilingual lexicon extraction from comparable corpora |
| 19.00-19.20 Gerard de Melo, Collin F. Baker, Nancy Ide, Rebecca J. Passonneau and Christiane Fellbaum | Empirical Comparisons of MASC Word Sense Annotations |

| Session: O36 - Time and Space | Chair: Piek Vossen | Room: Dolmabahçe A |
|-----------------------------|-------------------|------------------|
| 18.20-18.40 Hector Llorens, Leon Derczynski, Robert Gaizauskas and Estela Saquete | TIMEN: An Open Temporal Expression Normalisation Resource |
| 18.40-19.00 Kirk Roberts, Travis Goodwin and Sanda Harabagiu | Annotating Spatial Containment Relations Between Events |
| 19.00-19.20 James Pustejovsky and Jessica Moszkowicz | The Role of Model Testing in Standards Development: The Case of ISO-Space |

| Poster Sessions: 18.20-18.40 | Poster Area (1) |
|-----------------------------|----------------|

| Session: P30 - Discourse | Chair: David Traum |
|--------------------------|-------------------|
| Kristiina Jokinen and Silvi Tenjes | Investigating Engagement - intercultural and technological aspects of the collection, analysis, and use of the Estonian Multiparty Conversational video data |
| Patrick Saint-Dizier | DISLOG: A logic-based language for processing discourse structures |
| Sarah Bourse and Patrick Saint-Dizier | A Repository of Rules and Lexical Resources for Discourse Structure Analysis: the Case of Explanation Structures |
| Session: P31 - Lexical Acquisition | Chair: Pierre Zweigenbaum |
|-----------------------------------|--------------------------|
| Tommaso Caselli, Francesco Rubino, Francesca Forntini, Irene Russo and Valeria Quochi | Customizable SCF Acquisition in Italian |
| Gregor Thurmair, Vera Aleksic and Christoph Schwarz | Large Scale Lexical Analysis |
| Elsa Tolone, Stavroula Voyatzis, Claude Martineau and Matthieu Constant | Extending the adverbial coverage of a French morphological lexicon |
| Somayeh Bagherbeygi and Mehrnoosh Shamsfard | Corpus based Semi-Automatic Extraction of Persian Compound Verbs and their Relations |

| Session: P32 - Corpus Creation, Processing, Usage (2) | Chair: Shyam Agrawal |
|------------------------------------------------------|---------------------|
| Ting Liu, Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stomer-Galley, Sarah Taylor, Umit Boz, Xiaoi Ren and Jingsi Wu | Extending the MPC corpus to Chinese and Urdu - A Multiparty Multi-Lingual Chat Corpus for Modeling Social Phenomena in Language |
| Ivana Tanasijević, Biljana Sikimić and Gordana Pavlović-Lazetić | Multimedia database of the cultural heritage of the Balkans |
| Rania Al-Sabbagh and Roxana Girju | YADAC: Yet another Dialectal Arabic Corpus |
| Yves Scherrer and Bruno Cartoni | The Trilingual ALLEGRA Corpus: Presentation and Possible Use for Lexicon Induction |
| Martin Reynaert, Ineke Schuurman, Veronique Hoste, Nelleke Oostdijk and Maarten van Gompel | Beyond SoNaR: towards the facilitation of large corpus building efforts |
| Piotr Bański, Peter M. Fischer, Elena Frick, Erik Ketzan, Marc Kupietz, Carsten Schnober, Oliver Schonefeld and Andreas Witt | The New IDS Corpus Analysis Platform: Challenges and Prospects |
| Kurt Eberle, Kerstin Eckart, Ulrich Heid and Boris Haselbach | A Tool/Database Interface for Multi-Level Analyses |
| Djamel Mostefa | New language resources for the Pashto language |
| Şenay Kafkas, Ian Lewin, David Milward, Erik van Mulligen, Jan Kors, Udo Hahn and Dietrich Rebholz-Schuhmann | CALBC: Releasing the Final Corpora |
| Martín Majliš and Zdeněk Žabokrtský | Language Richness of the Web |

**Session: P33 - Web Services**  
Chair: Nuria Bel

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

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

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

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

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

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

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

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