

Papers Index

SESSION O1 - Corpora for Machine Translation

Iñaki San Vicente and Iker Manterola , PaCo2: A Fully Automated tool for gathering Parallel Corpora from the Web.....	1
Mark Fishel, Ondřej Bojar and Maja Popović , Terra: a Collection of Translation Error-Annotated Corpora.....	7
Ahmet Aker, Evangelos Kanoulas and Robert Gaizauskas , A light way to collect comparable corpora from the Web.....	15
Volha Petukhova, Rodrigo Agerri, Mark Fishel, Sergio Penkale, Arantza del Pozo, Mirjam Sepesy Maucec, Andy Way, Panayota Georgakopoulou and Martin Volk , SUMAT: Data Collection and Parallel Corpus Compilation for Machine Translation of Subtitles.....	21
Daniele Pighin, Lluís Màrquez and Lluís Formiga , The FAUST Corpus of Adequacy Assessments for Real-World Machine Translation Output.....	29

SESSION O2 - Infrastructures and Strategies for LRs (1)

Stelios Piperidis , The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions.....	36
Riccardo Del Gratta, Francesca Frontini, Francesco Rubino, Irene Russo and Nicoletta Calzolari , The Language Library: supporting community effort for collective resource production	43
Khalid Choukri, Victoria Arranz, Olivier Hamon and Jungyeul Park , Using the International Standard Language Resource Number: Practical and Technical Aspects.....	50
Valérie Mapelli, Victoria Arranz, Matthieu Carré, Hélène Mazo, Djamel Mostefa and Khalid Choukri , ELRA in the heart of a cooperative HLT world.....	55
Christopher Cieri, Marian Reed, Denise DiPersio and Mark Liberman , Twenty Years of Language Resource Development and Distribution: A Progress Report on LDC Activities.....	60

SESSION O3 - Semantics

Dan Moldovan and Eduardo Blanco , Polaris: Lymba's Semantic Parser.....	66
Sylvia Springorum, Sabine Schulte im Walde and Antje Roßdeutscher , Automatic classification of German “””an”””particle verbs.....	73
Livio Robaldo and Jakub Szymanik , Pragmatic identification of the witness sets.....	81
Orphée De Clercq, Veronique Hoste and Paola Monachesi , Evaluating automatic cross-domain Dutch semantic role annotation.....	88
Benoît Robichaud , Logic Based Methods for Terminological Assessment.....	94

SESSION O4 - Speech corpora

Luis Javier Rodriguez-Fuentes, Mikel Penagarikano, Amparo Varona, Mireia Diez and	
--	--

German Bordel , KALAKA-2: a TV Broadcast Speech Database for the Recognition of Iberian Languages in Clean and Noisy Environments.....	99
Tommaso Raso, Heliana Mello and Maryualê Malvessi Mittmann , The C-ORAL-BRASIL I: Reference Corpus for Spoken Brazilian Portuguese.....	106
Guillaume Gravier, Gilles Adda, Niklas Paulsson, Matthieu Carré, Aude Giraudel and Olivier Galibert , The ETAPE corpus for the evaluation of speech-based TV content processing in the French language.....	114
Daniel Stein and Bela Usabaev , Automatic Speech Recognition on a Firefighter TETRA Broadcast Channel.....	119
Anthony Rousseau, Paul Deléglise and Yannick Estève , TED-LIUM: an Automatic Speech Recognition dedicated corpus.....	125

SESSION P1 - Anaphora and Coreference

Abdul-Baqee Sharaf and Eric Atwell , QurAna: Corpus of the Quran annotated with Pronominal Anaphora.....	130
Stefanie Dipper, Melanie Seiss and Heike Zinsmeister , The Use of Parallel and Comparable Data for Analysis of Abstract Anaphora in German and English.....	138
Lucie Poláková, Pavlína Jínová and Jiří Mírovský , Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component.....	146
Luz Rello and Iria Gayo , A Portuguese-Spanish Corpus Annotated for Subject Realization and Referentiality.....	154
Marilisa Amoia, Kerstin Kunz and Ekaterina Lapshinova-Koltunski , Coreference in Spoken vs. Written Texts: a Corpus-based Analysis.....	158
Marta Recasens, M. Antònia Martí and Constantin Orasan , Annotating Near-Identity from Coreference Disagreements.....	165
Thomas Kaspersson, Christian Smith, Henrik Danielsson and Arne Jönsson , This also affects the context - Errors in extraction based summaries.....	173
Natsuko Nakagawa and Yasuharu Den , Annotation of anaphoric relations and topic continuity in Japanese conversation.....	179
Olga Uryupina and Massimo Poesio , Domain-specific vs. Uniform Modeling for Coreference Resolution.....	187
Mateusz Kopeć and Maciej Ogrodniczuk , Creating a Coreference Resolution System for Polish	192

SESSION P2 - Tools, Systems and Evaluation

Felix Burkhardt , Fast Labeling and Transcription with the Speechalyzer Toolkit.....	196
Bart Jongejan , Automatic annotation of head velocity and acceleration in Anvil.....	201
Przemyslaw Lenkiewicz, Binyam Gebrekidan Gebre, Oliver Schreer, Stefano Masneri, Daniel Schneider and Sebastian Tschöpel , AVATech – automated annotation through audio and video analysis.....	209
Henk van den Heuvel, Eric Sanders, Robin Rutten, Stef Scagliola and Paula Witkamp ,	

An Oral History Annotation Tool for INTER-VIEWS	215
Han Sloetjes and Aarthy Somasundaram , ELAN development, keeping pace with communities' needs	219
Michał Marcińczuk, Jan Kocoń and Bartosz Broda , Inforex – a web-based tool for text corpus management and semantic annotation	224
Binyam Gebrekidan Gebre, Peter Wittenburg and Przemyslaw Lenkiewicz , Towards Automatic Gesture Stroke Detection	231
Thomas Schmidt , EXMARaLDA and the FOLK tools – two toolsets for transcribing and annotating spoken language	236
Leonardo Campillos Llanos , Designing a search interface for a Spanish learner spoken corpus: the end-user's evaluation	241

SESSION P3 - Lexical Resources

Satoshi Sato , Dictionary Look-up with Katakana Variant Recognition	249
Karin Friberg Heppin and Maria Toporowska Gronostaj , The Rocky Road towards a Swedish FrameNet - Creating SweFN	256
Marie-Claude L'Homme and Janine Pimentel , Capturing syntactico-semantic regularities among terms: An application of the FrameNet methodology to terminology	262
David Graff and Mohamed Maamouri , Developing LMF-XML Bilingual Dictionaries for Colloquial Arabic Dialects	269
Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek and Christian M. Meyer , UBY-LMF – A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF	275
František Cvrček, Karel Pala and Pavel Rychlý , Legal electronic dictionary for Czech ..	283
Amir Hazem and Emmanuel Morin , Adaptive Dictionary for Bilingual Lexicon Extraction from Comparable Corpora	288
Jennifer Williams and Graham Katz , A New Twitter Verb Lexicon for Natural Language Processing	293

SESSION P4 - Annotation and Corpora

Ritesh Kumar , Challenges in the development of annotated corpora of computer-mediated communication in Indian Languages: A Case of Hindi	299
Christian Chiarcos , Ontologies of Linguistic Annotation: Survey and perspectives	303
Johanka Spoustová and Miroslav Spousta , A High-Quality Web Corpus of Czech	311
Xavier Tannier , WebAnnotator, an Annotation Tool for Web Pages	316
Chi-Hsin Yu, Yi-jie Tang and Hsin-Hsi Chen , Development of a Web-Scale Chinese Word N-gram Corpus with Parts of Speech Information	320
Dominique Fohr and Odile Mella , CoALT: A Software for Comparing Automatic Labelling Tools	325
Valentina Bartalesi Lenzi, Giovanni Moretti and Rachele Sprugnoli , CAT: the CELCT Annotation Tool	333

Radu Ion, Elena Irimia, Dan Ștefănescu and Dan Tufiș , ROMBAC: The Romanian Balanced Annotated Corpus	339
Ismail El Maarouf and Jeanne Villaneau , A French Fairy Tale Corpus syntactically and semantically annotated	345
Carlos Morell, Jorge Vivaldi and Núria Bel , Iula2Standoff: a tool for creating standoff documents for the IULACT.....	351
Frederic Landragin, Thierry Poibeau and Bernard Victorri , ANALEC: a New Tool for the Dynamic Annotation of Textual Data	357
Georgios Petasis , The SYNC3 Collaborative Annotation Tool.....	363
Heba Elfardy and Mona Diab , Simplified guidelines for the creation of Large Scale Dialectal Arabic Annotations	371

SESSION O5 - Crowdsourcing (Special Session)

Arno Scharl, Marta Sabou, Stefan Gindl, Walter Rafelsberger and Albert Weichselbraun , Leveraging the Wisdom of the Crowds for the Acquisition of Multilingual Language Resources	379
Anoop Kunchukuttan, Shourya Roy, Pratik Patel, Kushal Ladha, Somya Gupta, Mitesh M. Khapra and Pushpak Bhattacharyya , Experiences in Resource Generation for Machine Translation through Crowdsourcing	384
Elena Filatova , Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing ..	392
Luís Marujo, Anatole Gershman, Jaime Carbonell, Robert Frederking and João P. Neto , Supervised Topical Key Phrase Extraction of News Stories using Crowdsourcing, Light Filtering and Co-reference Normalization	399

SESSION O6 - Dialogue and Multimodality

Kristiina Jokinen and Graham Wilcock , Constructive Interaction for Talking about Interesting Topics	404
Florian Nothdurft and Wolfgang Minker , Using multimodal resources for explanation approaches in intelligent systems	411
Shota Yamasaki, Hirohisa Furukawa, Masafumi Nishida, Kristiina Jokinen and Seiichi Yamamoto , Multimodal Corpus of Multi-party Conversations in Second Language	416
Takenobu Tokunaga, Ryu Iida, Asuka Terai and Naoko Kuriyama , The REX corpora: A collection of multimodal corpora of referring expressions in collaborative problem solving dialogues	422
Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis and David Traum , ISO 24617-2: A semantically-based standard for dialogue annotation	430

SESSION O7 - Machine Translation and Language Resources (1)

Inguna Skadina, Ahmet Aker, Nikos Mastropavlos, Fangzhong Su, Dan Tufiș, Mateja Verlic, Andrejs Vasiljevs, Bogdan Babych, Paul Clough, Robert Gaizauskas, Nikos	
--	--

Glaros, Monica Lestari Paramita and Mārcis Pinnis , Collecting and Using Comparable Corpora for Statistical Machine Translation.....	438
Casey Redd Kennington, Martin Kay and Annemarie Friedrich , Suffix Trees as Language Models.....	446
Ralf Steinberger, Andreas Eisele, Szymon Kłoczek, Spyridon Pilos and Patrick Schlüter , DGT-TM: A freely available Translation Memory in 22 languages.....	454
Reinhard Rapp, Serge Sharoff and Bogdan Babych , Identifying Word Translations from Comparable Documents Without a Seed Lexicon.....	460
Gideon Kotzé, Vincent Vandeghinste, Scott Martens and Jörg Tiedemann , Large aligned treebanks for syntax-based machine translation.....	467

SESSION O8 - Corpus Processing and Infrastructure

Lars Borin, Markus Forsberg and Johan Roxendal , Korp – the corpus infrastructure of Språkbanken.....	474
Jonathan Wright, Kira Griffitt, Joe Ellis, Stephanie Strassel and Brendan Callahan , Annotation Trees: LDC’s customizable, extensible, scalable, annotation infrastructure.....	479
Roland Schäfer and Felix Bildhauer , Building Large Corpora from the Web Using a New Efficient Tool Chain.....	486
Young-Min Kim, Patrice Bellot, Elodie Faath and Marin Dacos , Annotated Bibliographical Reference Corpora in Digital Humanities.....	494
Jan Pomikálek, Miloš Jakubíček and Pavel Rychlý , Building a 70 billion word corpus of English from ClueWeb.....	502

SESSION P5 - Information Extraction (1)

Michael Wiegand, Benjamin Roth, Eva Lasarczyk, Stephanie Köser and Dietrich Klakow , A Gold Standard for Relation Extraction in the Food Domain.....	507
Mathias Bank, Robert Remus and Martin Schierle , Textual Characteristics for Language Engineering.....	515
Ziqi Zhang, Philip Webster, Victoria Uren, Andrea Varga and Fabio Ciravegna , Automatically Extracting Procedural Knowledge from Instructional Texts using Natural Language Processing	520
Xavier Tannier, Véronique Moriceau, Béatrice Arnulphy and Ruixin He , Evolution of Event Designation in Media: Preliminary Study.....	528
Yunqing Xia, Guoyu Tang, Peng Jin and Xia Yang , CLTC: A Chinese-English Cross-lingual Topic Corpus.....	532
Julia Maria Schulz, Daniela Becks, Christa Womser-Hacker and Thomas Mandl , A Resource-light Approach to Phrase Extraction for English and German Documents from the Patent Domain and User Generated Content.....	538
Md. Faisal Mahbub Chowdhury and Alberto Lavelli , An Evaluation of the Effect of Automatic Preprocessing on Syntactic Parsing for Biomedical Relation Extraction.....	544
Wei Wang, Romaric Besançon, Olivier Ferret and Brigitte Grau , Evaluation of Unsuper-	

vised Information Extraction.....	552
Stéphanie Weiser and Patrick Watrin , Extraction of unmarked quotations in Newspapers	559
Martin Aleksandrov and Carlo Strapparava , NgramQuery - Smart Information Extraction from Google N-gram using External Resources.....	563

SESSION P6 - Word Sense Disambiguation and Evaluation

Héctor Martínez Alonso, Núria Bel and Bolette Sandford Pedersen , A voting scheme to detect semantic underspecification.....	569
Verena Henrich and Erhard Hinrichs , A Comparative Evaluation of Word Sense Disambiguation Algorithms for German.....	576
Piek Vossen, Attila Görög, Rubén Izquierdo and Antal Van den Bosch , DutchSemCor: Targeting the ideal sense-tagged corpus.....	584
Samuel Fernando and Mark Stevenson , Mapping WordNet synsets to Wikipedia articles.	590
Myriam Rakho, Éric Laporte and Matthieu Constant , A new semantically annotated corpus with syntactic-semantic and cross-lingual senses.....	597
Minoru Sasaki and Hiroyuki Shinnou , Detection of Peculiar Word Sense by Distance Metric Learning with Labeled Examples.....	601
Soojeong Eom, Markus Dickinson and Graham Katz , Using semi-experts to derive judgments on word sense alignment: a pilot study.....	605
John Vogel, Marc Verhagen and James Pustejovsky , ATLAS: Identifying Locational Information in Text Automatically.....	612

SESSION P7 - Multiword Expressions and Term Extraction

Behrang QasemiZadeh, Paul Buitelaar, Tianqi Chen and Georgeta Bordea , Semi-Supervised Technical Term Tagging With Minimal User Feedback.....	617
Miriam Buendía-Castro and Beatriz Sánchez-Cárdenas , Linguistic knowledge for specialized text production.....	622
Rita Marinelli and Laura Cignoni , In the same boat and other idiomatic seafaring expressions	627
Sabine Schulte im Walde, Susanne Borgwaldt and Ronny Jauch , Association Norms of German Noun Compounds.....	632
Doaa Samy, Antonio Moreno-Sandoval, Conchi Bueno-Díaz, Marta Garrote-Salazar and José M. Guirao , Medical Term Extraction in an Arabic Medical Corpus.....	640
Matthieu Constant and Isabelle Tellier , Evaluating the Impact of External Lexical Resources into a CRF-based Multiword Segmenter and Part-of-Speech Tagger.....	646
Anita Gojun, Ulrich Heid, Bernd Weißbach, Carola Loth and Insa Mingers , Adapting and evaluating a generic term extraction tool.....	651
Mladen Karan, Jan Šnajder and Bojana Dalbelo Bašić , Evaluation of Classification Algorithms and Features for Collocation Extraction in Croatian.....	657
Thibault Mondary, Adeline Nazarenko, Haïfa Zargayouna and Sabine Barreaux , The Quaero Evaluation Initiative on Term Extraction.....	663

Shiva Taslimipoor, Afsaneh Fazly and Ali Hamzeh , Using Noun Similarity to Adapt an Acceptability Measure for Persian Light Verb Constructions.....	670
Dhouha Bouamor, Nasredine Semmar and Pierre Zweigenbaum , Identifying bilingual Multi-Word Expressions for Statistical Machine Translation	674
Takafumi Suzuki, Yusuke Abe, Itsuki Toyota, Takehito Utsuro, Suguru Matsuyoshi and Masatoshi Tsuchiya , Detecting Japanese Compound Functional Expressions using Canonical/Derivational Relation	680
Aude Grezka and Céline Poudat , Building a database of French frozen adverbial phrases .	685
Marc Luder , German Verb Patterns and Their Implementation in an Electronic Dictionary ..	693

SESSION P8 - Authoring Tools, Proofing

Flore Barcellini, Camille Albert, Corinne Grosse and Patrick Saint-Dizier , Risk Analysis and Prevention: LELIE, a Tool dedicated to Procedure and Requirement Authoring.....	698
Mohammad Hoseyn Sheykholeslam, Behrouz Minaei-Bidgoli and Hossein Juzi , A Framework for Spelling Correction in Persian Language Using Noisy Channel Model.....	706
Nizar Habash, Mona Diab and Owen Rambow , Conventional Orthography for Dialectal Arabic	711
Khaled Shaalan, Mohammed Attia, Pavel Pecina, Younes Samih and Josef van Genabith , Arabic Word Generation and Modelling for Spell Checking	719
Jan Rygl and Aleš Horák , Similarity Ranking as Attribute for Machine Learning Approach to Authorship Identification	726
Shaohua Yang, Hai Zhao, Xiaolin Wang and Bao-liang Lu , Spell Checking for Chinese	730
Jordi Atserias, Maria Fuentes, Rogelio Nazar and Irene Renau , Spell Checking in Spanish: The Case of Diacritic Accents.....	737
Michael Rosner, Albert Gatt, Andrew Attard and Jan Joachimsen , Incorporating an Error Corpus into a Spellchecker for Maltese	743

SESSION O9 - Endangered Languages

Melanie Seiss , A Rule-based Morphological Analyzer for Murrinh-Patha.....	751
Dirk Goldhahn, Thomas Eckart and Uwe Quasthoff , Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages.....	759
Helen Aristar-Dry, Sebastian Drude, Menzo Windhouwer, Jost Gippert and Irina Nevskaya , "Rendering Endangered Lexicons Interoperable through Standards Harmonization": the RELISH project	766
Ryan Georgi, Fei Xia and William Lewis , Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms	771

SESSION O10 - Document Classification, Text Categorisation

Julian Brooke and Graeme Hirst , Measuring Interlanguage: Native Language Identification with L1-influence Metrics	779
John Noecker Jr and Michael Ryan , Distractorless Authorship Verification.....	785

Monica Lestari Paramita, Paul Clough, Ahmet Aker and Robert Gaizauskas, Correlation between Similarity Measures for Inter-Language Linked Wikipedia Articles.....	790
Ralf Steinberger, Mohamed Ebrahim and Marco Turchi, JRC Eurovoc Indexer JEX - A freely available multi-label categorisation tool.....	798

SESSION O11 - Discourse (1)

Vinodkumar Prabhakaran, Huzaifa Neralwala, Owen Rambow and Mona Diab, Annotations for Power Relations on Email Threads.....	806
Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott and Joseph King, A Corpus for Research on Deliberation and Debate.....	812
Jacob Andreas, Sara Rosenthal and Kathleen McKeown, Annotating Agreement and Disagreement in Threaded Discussion.....	818
Sudheer Kolachina, Rashmi Prasad, Dipti Misra Sharma and Aravind Joshi, Evaluation of Discourse Relation Annotation in the Hindi Discourse Relation Bank.....	823

SESSION O12 - Word Sense Disambiguation

Will Roberts and Valia Kordoni, Using Verb Subcategorization for Word Sense Disambiguation	829
Marianna Apidianaki and Benoît Sagot, Applying cross-lingual WSD to wordnet development	833
Els Lefever, Veronique Hoste and Martine De Cock, Discovering Missing Wikipedia Inter-language Links by means of Cross-lingual Word Sense Disambiguation.....	841
Erwin Fernandez-Ordonez, Rada Mihalcea and Samer Hassan, Unsupervised Word Sense Disambiguation with Multilingual Representations.....	847

SESSION P9 - Morphology

Marco Passarotti and Francesco Mambrini, First Steps towards the Semi-automatic Development of a Wordformation-based Lexicon of Latin.....	852
Marcin Woliński, Marcin Miłkowski, Maciej Ogrodniczuk and Adam Przepiórkowski, PoliMorf: a (not so) new open morphological dictionary for Polish.....	860
Lionel Nicolas, Jacques Farré and Cécile Darme, Unsupervised acquisition of concatenative morphology.....	865
Emad Mohamed, Behrang Mohit and Kemal Oflazer, Annotating and Learning Morphological Segmentation of Egyptian Colloquial Arabic.....	873
Paul Felt, Eric Ringger, Kevin Seppi, Kristian Heal, Robbie Haertel and Deryle Lonsdale, First Results in a Study Evaluating Pre-annotation and Correction Propagation for Machine-Assisted Syriac Morphological Analysis.....	878
Claudia Marzi, Marcello Ferro, Claudia Caudai and Vito Pirrelli, Evaluating Hebbian Self-Organizing Memories for Lexical Representation and Access.....	886
Cheikh M. Bamba Dione, A Morphological Analyzer For Wolof Using Finite-State Techniques	894

Septina Dian Larasati , IDENTIC Corpus: Morphologically Enriched Indonesian-English Parallel Corpus	902
Liviu P. Dinu, Vlad Niculae and Octavia-Maria Şulea , The Romanian Neuter Examined Through A Two-Gender N-Gram Classification System.....	907
Toshinobu Ogiso, Mamoru Komachi, Yasuharu Den and Yuji Matsumoto , UniDic for Early Middle Japanese: a Dictionary for Morphological Analysis of Classical Japanese.....	911
Maciej Piasecki, Radosław Ramocki and Marek Maziarz , Recognition of Polish Derivational Relations Based on Supervised Learning Scheme	916
Dan Cristea, Radu Simionescu and Gabriela Haja , Reconstructing the Diachronic Morphology of Romanian from Dictionary Citations	923
Krešimir Šojat, Nives Mikelić Preradović and Marko Tadić , Generation of Verbal Stems in Derivationally Rich Language.....	928
Jonathan Washington, Mirlan Ipasov and Francis Tyers , A finite-state morphological transducer for Kyrgyz.....	934
Fabio Tamburini and Matias Melandri , AnIta: a powerful morphological analyser for Italian	941

SESSION P10 - Prosody and Phonetics

Gloria Gagliardi, Edoardo Lombardi Vallauri and Fabio Tamburini , A topologic view of Topic and Focus marking in Italian	948
Geneviève Caelen-Haumont and Sethserey Sam , Comparison between two models of language for the automatic phonetic labeling of an undocumented language of the South-Asia: the case of Mo Piu.....	956
Benoît Weber, Geneviève Caelen-Haumont, Binh Hai Pham and Do-Dat Tran , MIS-TRAL+: A Melody Intonation Speaker Tonal Range semi-automatic Analysis using variable Levels	963
Nelly Barbot, Olivier Boeffard and Arnaud Delhay , Comparing performance of different set-covering strategies for linguistic content optimization in speech corpora.....	969
Olivier Boeffard, Laure Charonnat, Sébastien Le Maguer and Damien Lolive , Towards Fully Automatic Annotation of Audio Books for TTS	975
Iris Merkus and Florian Schiel , Statistical Evaluation of Pronunciation Encoding	981
Helen Kaiyun Chen , Annotating a corpus of human interaction with prosodic profiles – focusing on Mandarin repair/disfluency	986
Kikuo Maekawa , Prediction of Non-Linguistic Information of Spontaneous Speech from the Prosodic Annotation: Evaluation of the X-JToBI system	991
Antonio Origlia and Iolanda Alfano , Prosomarker: a prosodic analysis tool based on optimal pitch stylization and automatic syllabification.....	997
David Doukhan, Sophie Rosset, Albert Rilliard, Christophe d’Alessandro and Martine Adda-Decker , Designing French Tale Corpora for Entertaining Text To Speech Synthesis ...	1003
Claire Brierley, Majdi Sawalha and Eric Atwell , Open-Source Boundary-Annotated Corpus for Arabic Speech and Language Processing.....	1011
Luc Boruta and Justyna Jastrzebska , A Phonemic Corpus of Polish Child-Directed Speech	1017

SESSION P11 - Language Resource Infrastructures (1)

Peter Spyns and Elisabeth D’Halleweyn , Smooth Sailing for STEVIN	1021
Dieter van Uytvanck, Herman Stehouwer and Lari Lampen , Semantic metadata mapping in practice: the Virtual Language Observatory	1029
Aditi Sharma Grover, Annamart Nieman, Gerhard Van Huyssteen and Justus Roux , Aspects of a Legal Framework for Language Resource Management	1035
Elena Volodina and Sofie Johansson Kokkinakis , Introducing the Swedish Kelly-list, a new lexical e-resource for Swedish	1040
Philippe Langlais, Patrick Drouin, Amélie Paulus, Eugénie Rompré Brodeur and Florent Cottin , Texto4Science: a Quebec French Database of Annotated Short Text Messages .	1047
Jan Odijk , Recent Developments in CLARIN-NL	1055
Emanuel Dima, Christina Hoppermann, Erhard Hinrichs, Thorsten Trippel and Claus Zinn , A Metadata Editor to Support the Description of Linguistic Resources	1061
Hanno Biber and Evelyn Breiteneder , Fivehundredmillionandone Tokens. Loading the AAC Container with Text Resources for Text Studies	1067
José Pedro Ferreira, Maarten Janssen, Gladis Barcellos de Oliveira, Margarita Correia and Gilvan Müller de Oliveira , The Common Orthographic Vocabulary of the Portuguese Language: a set of open lexical resources for a pluricentric language	1071
Andrejs Vasiljevs, Markus Forsberg, Tatiana Gornostay, Dorte Haltrup Hansen, Kristín Jóhannsdóttir, Gunn Lyse, Krister Lindén, Lene Offersgaard, Sussi Olsen, Bolette Pedersen, Eiríkur Rögnvaldsson, Inguna Skadina, Koenraad De Smedt, Ville Oksanen and Roberts Rozis , Creation of an Open Shared Language Resource Repository in the Nordic and Baltic Countries	1076
Nicoletta Calzolari, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo and Claudia Soria , The LRE Map. Harmonising Community Descriptions of Resources	1084
Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papa-georgiou, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz and Valérie Mapelli , The META-SHARE Metadata Schema for the Description of Language Resources	1090
Yoshinobu Kano , Towards automation in using multi-modal language resources: compatibility and interoperability for multi-modal features in Kachako	1098

SESSION O13 - Multimodal Corpora (1)

Aude Giraudel, Matthieu Carré, Valérie Mapelli, Juliette Kahn, Olivier Galibert and Ludovic Quintard , The REPERE Corpus : a multimodal corpus for person recognition	1102
Magdalena Lis , Polish Multimodal Corpus – a collection of referential gestures	1108
Stefan Scherer, Georg Layher, John Kane, Heiko Neumann and Nick Campbell , An audiovisual political speech analysis incorporating eye-tracking and perception data	1114

SESSION O14 - Machine Translation and Evaluation (1)

Sara Stymne, Henrik Danielsson, Sofia Bremin, Hongzhan Hu, Johanna Karlsson, Anna Prytz Lillkull and Martin Wester, Eye Tracking as a Tool for Machine Translation Error Analysis 1121

Eleftherios Avramidis, Aljoscha Burchardt, Christian Federmann, Maja Popović, Cindy Tscherwinka and David Vilar, Involving Language Professionals in the Evaluation of Machine Translation..... 1127

Daniele Pighin, Lluís Màrquez and Jonathan May, An Analysis (and an Annotated Corpus) of User Responses to Machine Translation Output 1131

SESSION O15 - Information Extraction and Question Answering

Bonan Min and Ralph Grishman, Challenges in the Knowledge Base Population Slot Filling Task..... 1137

Anselmo Peñas, Eduard Hovy, Pamela Forner, Álvaro Rodrigo, Richard Sutcliffe, Corina Forascu and Caroline Sporleder, Evaluating Machine Reading Systems through Comprehension Tests 1143

Xinkai Wang, Paul Thompson, Jun’ichi Tsujii and Sophia Ananiadou, Biomedical Chinese-English CLIR Using an Extended CMeSH Resource to Expand Queries 1148

SESSION O16 - Web Services

Marc Poch, Antonio Toral, Olivier Hamon, Valeria Quochi and Núria Bel, Towards a User-Friendly Platform for Building Language Resources based on Web Services 1156

Maciej Ogrodniczuk and Michał Lenart, Web Service integration platform for Polish linguistic resources 1164

Yoshihiko Hayashi and Chiharu Narawa, Classifying Standard Linguistic Processing Functionalities based on Fundamental Data Operation Types 1169

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

Xin Zuo, Tian Li and Pascale Fung, A Multilingual Natural Stress Emotion Database... 1174

Takahiro Miyajima, Hideaki Kikuchi, Katsuhiko Shirai and Shigeki Okawa, Method for Collection of Acted Speech Using Various Situation Scripts 1179

Hong Li, Xiwen Cheng, Kristina Adson, Tal Kirshboim and Feiyu Xu, Annotating Opinions in German Political News..... 1183

Akshat Bakliwal, Piyush Arora and Vasudeva Varma, Hindi Subjective Lexicon: A Lexical Resource for Hindi Adjective Polarity Classification 1189

Juan María Garrido, Yesika Laplaza, Montse Marquina, Andrea Pearman, José Gregorio Escalada, Miguel Ángel Rodríguez and Ana Armenta, The I3MEDIA speech database: a trilingual annotated corpus for the analysis and synthesis of emotional speech..... 1197

Panagiotis Giannoulis and Gerasimos Potamianos, A hierarchical approach with feature selection for emotion recognition from speech 1203

Alexandra Balahur and Jesús M. Hermida, Extending the EmotiNet Knowledge Base to Improve the Automatic Detection of Implicitly Expressed Emotions from Text 1207

Saeedeh Momtazi , Fine-grained German Sentiment Analysis on Social Media	1215
Felix Burkhardt , “You Seem Aggressive!” Monitoring Anger in a Practical Application	1221
Yi-jie Tang and Hsin-Hsi Chen , Mining Sentiment Words from Microblogs for Predicting Writer-Reader Emotion Transition	1226
Christian Scheible and Hinrich Schütze , Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank	1230

SESSION P13 - Named Entity Recognition

Antje Schlaf and Robert Remus , Learning Categories and their Instances by Contextual Features	1235
Nuno Cardoso , Rembrandt - a named-entity recognition framework	1240
Bogdan Sacaleanu and Günter Neumann , An Adaptive Framework for Named Entity Combination	1244
Maria Skeppstedt, Maria Kvist and Hercules Dalianis , Rule-based Entity Recognition and Coverage of SNOMED CT in Swedish Clinical Text	1250
Mārcis Pinnis , Latvian and Lithuanian Named Entity Recognition with TildeNER	1258
Marco Dinarelli and Sophie Rosset , Tree-Structured Named Entity Recognition on OCR Data: Analysis, Processing and Results	1266
Benoît Sagot and Rosa Stern , Aleda, a free large-scale entity database for French	1273
Pablo Mendes, Joachim Daiber, Rohana Rajapakse, Felix Sasaki and Christian Bizer , Evaluating the Impact of Phrase Recognition on Concept Tagging	1277

SESSION P14 - Dialogue

Tobias Heinroth, Maximilian Grotz, Florian Nothdurft and Wolfgang Minker , Adaptive Speech Understanding for Intuitive Model-based Spoken Dialogues	1281
Kseniya Zablotskaya, Umair Rahim, Fernando Fernández Martínez and Wolfgang Minker , Relating Dominance of Dialogue Participants with their Verbal Intelligence Scores	1289
Volha Petukhova and Harry Bunt , The coding and annotation of multimodal dialogue acts	1293
Harry Bunt, Michael Kipp and Volha Petukhova , Using DiAML and ANVIL for multimodal dialogue annotations	1301
Matthew Fuchs, Nikos Tsourakis and Manny Rayner , A Scalable Architecture For Web Deployment of Spoken Dialogue Systems	1309
Nikos Tsourakis and Manny Rayner , A Corpus for a Gesture-Controlled Mobile Spoken Dialogue System	1315
Emina Kurtic, Bill Wells, Guy J. Brown, Timothy Kempton and Ahmet Aker , A Corpus of Spontaneous Multi-party Conversation in Bosnian Serbo-Croatian and British English	1323
Jing Guang Han, Emer Gilmartin, Celine DeLooze, Brian Vaughan and Nick Campbell , The Herme Database of Spontaneous Multimodal Human-Robot Dialogues	1328
Yasuharu Den, Hanae Koiso, Katsuya Takanashi and Nao Yoshida , Annotation of response tokens and their triggering expressions in Japanese multi-party conversations	1332

Thierry Bazillon, Melanie Deplano, Frederic Bechet, Alexis Nasr and Benoit Favre, Syntactic annotation of spontaneous speech: application to call-center conversation data	1338
Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori and Eric Arbillot, DECODA: a call-centre human-human spoken conversation corpus	1343
Pepi Stavropoulou, Dimitris Spiliotopoulos and Georgios Kouroupetroglou, Resource Evaluation for Usable Speech Interfaces: Utilizing Human-Human Dialogue	1348
Jens Edlund, Simon Alexandersson, Jonas Beskow, Lisa Gustavsson, Mattias Heldner, Anna Hjalmarsson, Petter Kallionen and Ellen Marklund, 3rd party observer gaze as a continuous measure of dialogue flow	1354
Marc Tomlinson, David Bracewell, Mary Draper, Zewar Almissour, Ying Shi and Jeremy Bensley, Pursing power in Arabic on-line discussion forums	1359
Sunao Hara, Norihide Kitaoka and Kazuya Takeda, Causal analysis of task completion errors in spoken music retrieval interactions	1365
Marilyn Walker, Grace Lin and Jennifer Sawyer, An Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style	1373

SESSION O17 - Infrastructures and Strategies for LRs (2)

Claudia Soria, Núria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan Odijk, Stelios Piperidis, Valeria Quochi and Nicoletta Calzolari, The FLaReNet Strategic Language Resource Agenda	1379
Daan Broeder, Dieter van Uytvanck, Maria Gavrilidou, Thorsten Trippel and Menzo Windhouwer, Standardizing a Component Metadata Infrastructure	1387
Daan Broeder, Dieter van Uytvanck and Gunter Senft, Citing on-line Language Resources	1391
Khalid Choukri and Victoria Arranz, An Analytical Model of Language Resource Sustainability	1395
David Lewis, Alexander O'Connor, Andrzej Zydroń, Gerd Sjögren and Rahzeb Choudhury, On Using Linked Data for Language Resource Sharing in the Long Tail of the Localisation Market	1403

SESSION O18 - Dialogue

Alexandros Papangelis, Vangelis Karkaletsis and Fillia Makedon, Evaluation of Online Dialogue Policy Learning Techniques	1410
Lluis-F. Hurtado, Fernando Garcia, Emilio Sanchis and Encarna Segarra, The acquisition and dialog act labeling of the EDECAN-SPORTS corpus	1416
Jolanta Bachan, Developing and evaluating an emergency scenario dialogue corpus	1421
Lina M. Rojas-Barahona, Alejandra Lorenzo and Claire Gardent, Building and Exploiting a Corpus of Dialog Interactions between French Speaking Virtual and Human Agents	1428
Fabrice Lefèvre, Djamel Mostefa, Laurent Besacier, Yannick Estève, Matthieu Quignard, Nathalie Camelin, Benoit Favre, Bassam Jabaian and Lina M. Rojas-Barahona,	

Leveraging study of robustness and portability of spoken language understanding systems across languages and domains: the PORTMEDIA corpora 1436

SESSION O19 - Resource Creation and Acquisition

Xabier Saralegi, Iker Manterola and Iñaki San Vicente, Building a Basque-Chinese Dictionary by Using English as Pivot 1443

Núria Bel, Lauren Romeo and Muntsa Padró, Automatic lexical semantic classification of nouns 1448

Ahmet Aker, Mahmoud El-Haj, M-Dyaa Albakour and Udo Kruschwitz, Assessing Crowdsourcing Quality through Objective Tasks..... 1456

Attila Zséder, Gábor Recski, Dániel Varga and András Kornai, Rapid creation of large-scale corpora and frequency dictionaries..... 1462

Kata Gábor, Marianna Apidianaki, Benoît Sagot and Éric Villemonte de la Clergerie, Boosting the Coverage of a Semantic Lexicon by Automatically Extracted Event Nominalizations
1466

SESSION O20 - Corpus and Annotation

Karën Fort, Claire François, Olivier Galibert and Maha Ghribi, Analyzing the Impact of Prevalence on the Evaluation of a Manual Annotation Campaign..... 1474

Donia Scott, Rossano Barone and Rob Koeling, Corpus Annotation as a Scientific Task 1481

Stephen Wattam, Paul Rayson and Damon Berridge, Document Attrition in Web Corpora: an Exploration 1486

Anil Kumar Singh, A Concise Query Language with Search and Transform Operations for Corpora with Multiple Levels of Annotation 1490

Paola Velardi, Roberto Navigli, Stefano Faralli and Juana Maria Ruiz-Martinez, A New Method for Evaluating Automatically Learned Terminological Taxonomies 1498

SESSION P15 - Semantic Annotation

Béatrice Arnulphy, Xavier Tannier and Anne Vilnat, Event Nominals: Annotation Guidelines and a Manually Annotated Corpus in French 1505

Maria Aloni, Andreas van Cranenburgh, Raquel Fernandez and Marta Sznajder, Building a Corpus of Indefinite Uses Annotated with Fine-grained Semantic Functions 1511

António Branco, Catarina Carvalheiro, Sílvia Pereira, Sara Silveira, João Silva, Sérgio Castro and João Graça, A PropBank for Portuguese: the CINTIL-PropBank 1516

Ashwini Vaidya, Jinho D. Choi, Martha Palmer and Bhuvana Narasimhan, Empty Argument Insertion in the Hindi PropBank 1522

Pierrette Bouillon, Elisabetta Jezek, Chiara Melloni and Aurélie Picton, Annotating Qualia Relations in Italian and French Complex Nominals 1527

Juliette Thuilier and Laurence Danlos, Semantic annotation of French corpora: animacy and verb semantic classes 1533

Josef Ruppenhofer and Ines Rehbein, Yes we can!? Annotating English modal verbs.... 1538

Mehdi Manshadi, James Allen and Mary Swift, An Annotation Scheme for Quantifier Scope Disambiguation	1546
Yuichiroh Matsubayashi, Yusuke Miyao and Akiko Aizawa, Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure	1554
Kyoko Ohara, Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English	1559
Roser Morante and Walter Daelemans, ConanDoyle-neg: Annotation of negation cues and their scope in Conan Doyle stories	1563

SESSION P16 - Document Classification, Text Categorisation

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	1569
Kseniya Zablotskaya, Fernando Fernández Martínez and Wolfgang Minker, Investigating Verbal Intelligence Using the TF-IDF Approach	1573
Sanja Štajner and Ruslan Mitkov, Diachronic Changes in Text Complexity in 20th Century English Language: An NLP Approach	1577
Tommaso Fornaciari and Massimo Poesio, DeCour: a corpus of DEceptive statements in Italian COURts	1585
Amalia Todirascu, Sebastian Pado, Jennifer Krisch, Max Kisselew and Ulrich Heid, French and German Corpora for Audience-based Text Type Classification	1591
Borut Sluban, Senja Pollak, Roel Coesemans and Nada Lavrac, Irregularity Detection in Categorized Document Corpora	1598
Carmen Dayrell, Arnaldo Candido Jr., Gabriel Lima, Danilo Machado Jr., Ann Copes-take, Valéria Feltrim, Stella Tagnin and Sandra Aluisio, Rhetorical Move Detection in English Abstracts: Multi-label Sentence Classifiers and their Annotated Corpora	1604
Andrea Varga, Daniel Preotiuc-Pietro and Fabio Ciravegna, Unsupervised document zone identification using probabilistic graphical models	1610
Mohammad Hossein Elahimanesh, Behrouz Minaei and Hossein Malekinezhad, Improving K-Nearest Neighbor Efficacy for Farsi Text Classification	1618

SESSION P17 - Grammar and Syntax

Erhard Hinrichs and Thomas Zastrow, Automatic Annotation and Manual Evaluation of the Diachronic German Corpus TüBa-D/DC	1622
Hongsuck Seo, Kyusong Lee, Gary Geunbae Lee, Soo-Ok Kweon and Hae-Ri Kim, Grammatical Error Annotation for Korean Learners of Spoken English	1628
Heiki-Jaan Kaalep and Kadri Muischnek, Robust clause boundary identification for corpus annotation	1632
Patrick Ziering, Sina Zarriß and Jonas Kuhn, A Corpus-based Study of the German Recipient Passive	1637
Zygmunt Vetulani, Wordnet Based Lexicon Grammar for Polish	1645

Montserrat Arza, José M. García-Miguel, Francisco Campillo and Miguel Cuevas - Alonso , A Galician Syntactic Corpus with Application to Intonation Modeling	1650
Hiroaki Sato , A Search Tool for FrameNet Constructicon.....	1655
Markus Dickinson and Scott Ledbetter , Annotating Errors in a Hungarian Learner Corpus	1659
Stefan Bott, Horacio Saggion and Simon Mille , Text Simplification Tools for Spanish..	1665
Antske Fokkens, Tania Avgustinova and Yi Zhang , CLIMB grammars: three projects using metagrammar engineering	1672
Peteris Paikens and Normunds Gruzitis , An implementation of a Latvian resource grammar in Grammatical Framework.....	1680
Shafqat Mumtaz Virk and Elnaz Abolahrar , An Open Source Persian Computational Grammar	1686
Paula Buttery and Andrew Caines , Reclassifying subcategorization frames for experimental analysis and stimulus generation	1694
Andrew Caines and Paula Buttery , Annotating progressive aspect constructions in the spoken section of the British National Corpus	1699

SESSION P18 - Digital Libraries

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	1705
Ranka Stanković, Cvetana Krstev, Ivan Obradović, Aleksandra Trtovac and Miloš Utvić , A tool for enhanced search of multilingual digital libraries of e-journals.....	1710
Benjamin Weitz and Ulrich Schäfer , A Graphical Citation Browser for the ACL Anthology	1718
Eleftheria Ahtaridis, Christopher Cieri and Denise DiPersio , LDC Language Resource Database: Building a Bibliographic Database	1723
Eneko Agirre, Ander Barrena, Oier Lopez de Lacalle, Aitor Soroa, Samuel Fernando and Mark Stevenson , Matching Cultural Heritage items to Wikipedia.....	1729

SESSION O21 - Speech Corpora and Tools

Maria Eskevich, Gareth J.F. Jones, Martha Larson and Roeland Ordelman , Creating a Data Collection for Evaluating Rich Speech Retrieval	1736
Petya Osenova and Kiril Simov , The Political Speech Corpus of Bulgarian.....	1744
Brigitte Bigi , SPPAS: a tool for the phonetic segmentation of speech.....	1748
Brigitte Bigi, Pauline Péri and Roxane Bertrand , Orthographic Transcription: which enrichment is required for phonetization?	1756

SESSION O22 - Machine Translation and Evaluation (2)

Sandra Weiss and Lars Ahrenberg , Error profiling for evaluation of machine-translated text: a Polish-English case study	1764
---	------

Chunqi Shi, Donghui Lin, Masahiko Shimada and Toru Ishida, Two Phase Evaluation for Selecting Machine Translation Services.....	1771
Lorenza Russo, Sharid Loáiciga and Asheesh Gulati, Italian and Spanish Null Subjects. A Case Study Evaluation in an MT Perspective.....	1779
Sara Stymne and Lars Ahrenberg, On the practice of error analysis for machine translation evaluation.....	1785

SESSION O23 - Semantic Resources

Janine Pimentel, Identifying equivalents of specialized verbs in a bilingual comparable corpus of judgments: A frame-based methodology.....	1791
Alessandra Zarcone and Stefan Rued, Logical metonymies and qualia structures: an annotated database of logical metonymies for German.....	1799
Iris Hendrickx, Amália Mendes and Silvia Mencarelli, Modality in Text: a Proposal for Corpus Annotation.....	1805
Pablo Mendes, Max Jakob and Christian Bizer, DBpedia: A Multilingual Cross-domain Knowledge Base.....	1813

SESSION O24 - Trends in Corpora

Annie Louis and Ani Nenkova, A corpus of general and specific sentences from news.....	1818
Gozde Ozbal, Carlo Strapparava and Marco Guerini, Brand Pitt: A Corpus to Explore the Art of Naming.....	1822
Jonathon Read, Dan Flickinger, Rebecca Dridan, Stephan Oepen and Lilja Øvrelid, The WeSearch Corpus, Treebank, and Treecache – A Comprehensive Sample of User-Generated Content	1829
Masashi Inoue and Toshiki Akagi, Collecting humorous expressions from a community-based question-answering-service corpus.....	1836

SESSION P19 - Treebanks

Seth Kulick, Ann Bies and Justin Mott, Further Developments in Treebank Error Detection Using Derivation Trees.....	1840
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.....	1848
Mohamed Maamouri, Ann Bies and Seth Kulick, Expanding Arabic Treebank to Speech: Results from Broadcast News.....	1856
Magali Sanches Duran and Sandra Maria Aluísio, Propbank-Br: a Brazilian Treebank annotated with semantic role labels.....	1862
Yi Zhang, Rui Wang and Yu Chen, Joint Grammar and Treebank Development for Mandarin Chinese with HPSG.....	1868
Annette Rios and Anne Göhring, A tree is a Baum is an árbol is a sach'a: Creating a trilingual treebank.....	1874

Miriam Kaeshammer and Vera Demberg , German and English Treebanks and Lexica for Tree-Adjoining Grammars	1880
Loganathan Ramasamy and Zdeněk Žabokrtský , Prague Dependency Style Treebank for Tamil 1888	
Patricia Gonçalves, Rita Santos and António Branco , Treebanking by Sentence and Tree Transformation: Building a Treebank to support Question Answering in Portuguese.....	1895
Dasa Berovic, Zeljko Agic and Marko Tadić , Croatian Dependency Treebank: Recent Development and Initial Experiments	1902
Rahul Agarwal, Bharat Ram Ambati and Anil Kumar Singh , A GUI to Detect and Correct Errors in Hindi Dependency Treebank	1907
Masood Ghayoomi , From Grammar Rule Extraction to Treebanking: A Bootstrapping Approach	1912
Montserrat Marimon, Beatriz Fisas, Núria Bel, Marta Villegas, Jorge Vivaldi, Sergi Torner, Mercè Lorente, Silvia Vázquez and Marta Villegas , The IULA Treebank	1920
Atro Voutilainen, Kristiina Muhonen, Tanja Purtonen and Krister Lindén , Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3	1927
Cristina Bosco, Manuela Sanguinetti and Leonardo Lesmo , The Parallel-TUT: a multilingual and multiformat treebank.....	1932
Teresa Lynn, Ozlem Cetinoglu, Jennifer Foster, Elaine Uí Dhonnchadha, Mark Dras and Josef van Genabith , Irish Treebanking and Parsing: A Preliminary Evaluation	1939

SESSION P20 - Parsing

Mohammed Attia, Khaled Shaalan, Lamia Tounsi and Josef van Genabith , Automatic Extraction and Evaluation of Arabic LFG Resources	1947
Kristiina Muhonen and Tanja Purtonen , Rule-Based Detection of Clausal Coordinate Ellipsis	1955
Gülşen Eryiğit , The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish	1960
Stasinou Konstantopoulos, Valia Kordoni, Nicola Cancedda, Vangelis Karkaletsis, Dietrich Klakow and Jean-Michel Renders , Task-Driven Linguistic Analysis based on an Under-specified Features Representation	1966
Malin Ahlberg and Ramona Enache , Combining Language Resources Into A Grammar-Driven Swedish Parser	1971
Eiríkur Rögnvaldsson, Anton Karl Ingason, Einar Freyr Sigurðsson and Joel Wallenberg , The Icelandic Parsed Historical Corpus (IcePaHC).....	1977
Anita Alicante, Cristina Bosco, Anna Corazza and Alberto Lavelli , A treebank-based study on the influence of Italian word order on parsing performance	1985
Dong Wang and Fei Xia , Effort of Genre Variation and Prediction of System Performance	1993

SESSION P21 - Information Extraction (2)

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

Yildiz , Statistical Section Segmentation in Free-Text Clinical Records	2001
Ramona Bongelli, Carla Canestrari, Ilaria 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.....	2009
Cristina Mota, Alberto Simões, Cláudia Freitas, Luís Costa and Diana Santos , Páxico: Evaluating Wikipedia-based information retrieval in Portuguese.....	2015
Danica Damljanovic, Udo Kruschwitz, M-Dyaa Albakour, Johann Petrak and Mihai Lupu , Applying Random Indexing to Structured Data to Find Contextually Similar Words..	2023
Horacio Saggion and Sandra Szasz , The CONCUSUS Corpus of Event Summaries	2031
Gracinda Carvalho, David Martins de Matos and Vitor Rocio , Building and Exploring Semantic Equivalences Resources	2038
Marc Verhagen and James Pustejovsky , The TARSQI Toolkit	2043
Gabriella Pardelli, Manuela Sassi, Sara Goggi and Stefania Biagioni , From medical language processing to BioNLP domain	2049
Romarc Besançon, Olivier Ferret and Ludovic Jean-Louis , Evaluation of a Complex Information Extraction Application in Specific Domain	2056
Hannah Kermes , A methodology for the extraction of information about the usage of formulaic expressions in scientific texts.....	2064
André Santos, José João Almeida and Nuno Carvalho , Structural alignment of plain text books.....	2069
Gerold Schneider, Fabio Rinaldi and Simon Clematide , Dependency parsing for interaction detection in pharmacogenomics	2075
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.....	2083

SESSION P22 - Part-of-Speech Tagging

Slav Petrov, Dipanjan Das and Ryan McDonald , A Universal Part-of-Speech Tagset ..	2089
Atro Voutilainen , Improving corpus annotation productivity: a method and experiment with interactive tagging.....	2097
Andrea Gesmundo and Tanja Samardzic , Lemmatising Serbian as Category Tagging with Bidirectional Sequence Classification.....	2103
Souhir Gahbiche-Braham, H�el�ene Bonneau-Maynard, Thomas Lavergne and Fran�ois Yvon , Joint Segmentation and POS Tagging for Arabic Using a CRF-based Classifier	2107
Mans Hulden and Jerid Francom , Boosting statistical tagger accuracy with simple rule-based grammars	2114
Maarten Janssen , NeoTag: a POS Tagger for Grammatical Neologism Detection.....	2118
Francesco Rubino, Francesca Frontini and Valeria Quochi , Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser	2125

SESSION P23 - Machine Translation (1)

Bruno Cartoni and Thomas Meyer , Extracting Directional and Comparable Corpora from a	
--	--

Multilingual Corpus for Translation Studies.....	2132
Marianna J. Martindale , Can Statistical Post-Editing with a Small Parallel Corpus Save a Weak MT Engine?.....	2138
Sanja Seljan, Marija Brkić and Tomislav Vičić , BLEU Evaluation of Machine-Translated English-Croatian Legislation.....	2143
Chenhui Chu, Toshiaki Nakazawa and Sadao Kurohashi , Chinese Characters Mapping Table of Japanese, Traditional Chinese and Simplified Chinese.....	2149
Juan Pablo Martínez Cortés, Jim O’Regan and Francis Tyers , Free/Open Source Shallow-Transfer Based Machine Translation for Spanish and Aragonese.....	2153
Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović and Daniel Zeman , Automatic MT Error Analysis: Hjerson Helping Addicter.....	2158
Amit Sangodkar and Om Damani , Re-ordering Source Sentences for SMT.....	2164
Ângela Costa, Tiago Luís, Joana Ribeiro, Ana Cristina Mendes and Luísa Coheur , An English-Portuguese parallel corpus of questions: translation guidelines and application in SMT	2172
Mehmet Talha Çakmak, Süleyman Acar and Gülşen Eryiğit , Word Alignment for English-Turkish Language Pair.....	2177
Radu Ion , PEXACC: A Parallel Sentence Mining Algorithm from Comparable Corpora.....	2181
Eleftherios Avramidis, Marta R. Costa-Jussà, Christian Federmann, Josef van Genabith, Maite Melero and Pavel Pecina , A Richly Annotated, Multilingual Parallel Corpus for Hybrid Machine Translation.....	2189
Stephen Grimes, Katherine Peterson and Xuansong Li , Automatic word alignment tools to scale production of manually aligned parallel texts.....	2194
Carla Parra Escartín , Design and compilation of a specialized Spanish-German parallel corpus	2199
Jörg Tiedemann, Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen and Matthias Zumpe , A Distributed Resource Repository for Cloud-Based Machine Translation.....	2207

SESSION P24 - Corpus Creation, Processing, Usage (1)

Jörg Tiedemann , Parallel Data, Tools and Interfaces in OPUS.....	2214
Maciej Ogrodniczuk , The Polish Sejm Corpus.....	2219
Lieve Macken, Veronique Hoste, Marielle Leijten and Luuk Van Waes , From keystrokes to annotated process data: Enriching the output of Inputlog with linguistic information.....	2224
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.....	2230
Michel Génèreux, Iris Hendrickx and Amália Mendes , Introducing the Reference Corpus of Contemporary Portuguese Online.....	2237
Mojgan Seraji, Beáta Megyesi and Joakim Nivre , A Basic Language Resource Kit for Persian	2245
Eric Sanders , Collecting and Analysing Chats and Tweets in SoNaR.....	2253
Tomaz Erjavec , The goo300k corpus of historical Slovene.....	2257

Mathieu-Henri Falco, Véronique Moriceau and Anne Vilnat, Kitten: a tool for normalizing HTML and extracting its textual content	2261
Maaske Treurniet, Orphée De Clercq, Henk van den Heuvel and Nelleke Oostdijk, Collection of a corpus of Dutch SMS	2268
Alessandro Panunzi, Marco Fabbri, Massimo Moneglia, Lorenzo Gregori and Samuele Paladini, RIDIRE-CPI: an Open Source Crawling and Processing Infrastructure for Supervised Web-Corpora Building	2274
Brett Drury and José João Almeida, The Minho Quotation Resource	2280
Elena Frick, Carsten Schnober and Piotr Bański, Evaluating Query Languages for a Corpus Processing System	2286

SESSION P25 - Evaluation Methodologies

Abdul-Baquee Sharaf and Eric Atwell, QurSim: A corpus for evaluation of relatedness in short texts	2295
Christina Feilmayr, Birgit Pröll and Elisabeth Linsmayr, EVALIEX – A Proposal for an Extended Evaluation Methodology for Information Extraction Systems	2303
Patrick Paroubek and Xavier Tannier, A Rough Set Formalization of Quantitative Evaluation with Ambiguity	2311
Thomas Eckart, Uwe Quasthoff and Dirk Goldhahn, The Influence of Corpus Quality on Statistical Measurements on Language Resources	2318
Olga Babko-Malaya, Greg Milette, Michael Schneider and Sarah Scogin, Identifying Nuggets of Information in GALE Distillation Evaluation	2322
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	2328

SESSION O25 - Multimodal Corpora (2)

Ibrahim Saygin Topkaya and Hakan Erdogan, SUTAV: A Turkish Audio-Visual Database	2334
Costanza Navarretta and Patrizia Paggio, Multimodal Behaviour and Feedback in Different Types of Interaction	2338
Carlo Strapparava, Rada Mihalcea and Alberto Battocchi, A Parallel Corpus of Music and Lyrics Annotated with Emotions	2343
Merlin Teodosia Suarez, Jocelynn Cu and Madelene Sta. Maria, Building a Multimodal Laughter Database for Emotion Recognition	2347
Dimitra Anastasiou, A Speech and Gesture Spatial Corpus in Assisted Living	2351

SESSION O26 - Child Language Corpus

Priti Aggarwal, Ron Artstein, Jillian Gerten, Athanasios Katsamanis, Shrikanth Narayanan, Angela Nazarian and David Traum, The Twins Corpus of Museum Visitor Questions ...	2355
Hyejin Hong, Sunhee Kim and Minhwa Chung, Korean Children’s Spoken English Corpus and an Analysis of its Pronunciation Variability	2362

Marie Tahon, Agnes Delaborde and Laurence Devillers, Corpus of Children Voices for Mid-level Markers and Affect Bursts Analysis	2366
Aline Villavicencio, Beracah Yankama, Marco Idiart and Robert Berwick, A large scale annotated child language construction database	2370
Brian MacWhinney, Morphosyntactic Analysis of the CHILDES and TalkBank Corpora ...	2375

SESSION O27 - MultiWord Expressions

Veronika Vincze, Light Verb Constructions in the SzegedParalellFX English–Hungarian Parallel Corpus	2381
Antton Gurrutxaga and Iñaki Alegria, Measuring the compositionality of NV expressions in Basque by means of distributional similarity techniques	2389
Marion Weller and Ulrich Heid, Analyzing and Aligning German compound nouns	2395
Natalia Loukachevitch, Automatic Term Recognition Needs Multiple Evidence	2401
Roman Kurc, Maciej Piasecki and Bartosz Broda, Constraint Based Description of Polish Multiword Expressions	2408

SESSION O28 - Sign Language

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	2414
Matti Karppa, Tommi Jantunen, Ville Viitaniemi, Jorma Laaksonen, Birgitta Burger and Danny De Weerd, Comparing computer vision analysis of signed language video with motion capture recordings	2421
Annelies Braffort and Leïla Boutora, DEGELS1: A comparable corpus of French Sign Language and co-speech gestures	2426
Matilde Gonzalez, Michael Filhol and Christophe Collet, Semi-Automatic Sign Language Corpora Annotation using Lexical Representations of Signs	2430
Umar Shoaib, Nadeem Ahmad, Paolo Prinetto and Gabriele Tiotto, A platform-independent user-friendly dictionary from Italian to LIS	2435

SESSION P26 - Multilinguality

Jyrki Niemi and Krister Lindén, Representing the Translation Relation in a Bilingual Wordnet	2439
Alexandr Rosen and Martin Vavřín, Building a multilingual parallel corpus for human users	2447
Martina Katalin Szabó, Veronika Vincze and István Nagy T., HunOr: A Hungarian–Russian Parallel Corpus	2453
Kanika Gupta, Monojit Choudhury and Kalika Bali, Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics	2459
Gilles Sérasset, Dbnary: Wiktionary as a LMF based Multilingual RDF network	2466
Lluís Padró and Evgeny Stanilovsky, FreeLing 3.0: Towards Wider Multilinguality	2473

Svetla Koeva, Ivelina Stoyanova, Rositsa Dekova, Borislav Rizov and Angel Genov, Bulgarian X-language Parallel Corpus	2480
Enikő Héja and Dávid Takács, Automatically Generated Online Dictionaries	2487
Costanza Navarretta, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen and Patrizia Paggio, Feedback in Nordic First-Encounters: a Comparative Study	2494
Yu Chen and Andreas Eisele, MultiUN v2: UN Documents with Multilingual Alignments	2500
Zahurul Islam and Alexander Mehler, Customization of the Europarl Corpus for Translation Studies	2505
Thierry Declerck, Karlheinz Mörth and Piroska Lendvai, Accessing and standardizing Wiktionary lexical entries for the translation of labels in Cultural Heritage taxonomies	2511
Ying Li, Yue Yu and Pascale Fung, A Mandarin-English Code-Switching Corpus	2515
Paulo Fernandes, Lucelene Lopes, Carlos A. Prolo, Afonso Sales and Renata Vieira, A Fast, Memory Efficient, Scalable and Multilingual Dictionary Retriever	2520
Aitor Gonzalez-Agirre, Egoitz Laparra and German Rigau, Multilingual Central Repository version 3.0	2525

SESSION P27 - Question Answering and Summarisation

Christian Smith, Henrik Danielsson and Arne Jönsson, A good space: Lexical predictors in word space evaluation	2530
Ulrich Andersen, Anna Braasch, Lina Henriksen, Csaba Huszka, Anders Johannsen, Lars Kayser, Bente Maegaard, Ole Norgaard, Stefan Schulz and Jürgen Wedekind, Creation and use of Language Resources in a Question-Answering eHealth System	2536
Atsushi Fujii, Yuya Fujii and Takenobu Tokunaga, Effects of Document Clustering in Modeling Wikipedia-style Term Descriptions	2543
Silvia Quarteroni, Vincenzo Guerrisi and Pietro La Torre, Evaluating Multi-focus Natural Language Queries over Data Services	2547
Maria Fuentes, Horacio Rodríguez and Jordi Turmo, Summarizing a multimodal set of documents in a Smart Room	2553

SESSION P28 - Multimodal Corpus for Interaction

Dietmar Rösner, Jörg Frommer, Rafael Friesen, Matthias Haase, Julia Lange and Mirko Otto, LAST MINUTE: a Multimodal Corpus of Speech-based User-Companion Interactions .	2559
Karén Fort and Vincent Claveau, Annotating Football Matches: Influence of the Source Medium on Manual Annotation	2567
Stephanie Strassel, Amanda Morris, Jonathan Fiscus, Christopher Caruso, Haejoong Lee, Paul Over, James Fiumara, Barbara Shaw, Brian Antonishek and Martial Michel, Creating HAVIC: Heterogeneous Audio Visual Internet Collection	2573
Charlotte Alazard, Corine Astésano and Michel Billières, MULTIPHONIA: a MULTImodal database of PHONetics teaching methods in classroom InterActions	2578

SESSION 29 - Ontologies

Egoitz Laparra, German Rigau and Piek Vossen, Mapping WordNet to the Kyoto ontology	2584
Kugatsu Sadamitsu, Kuniko Saito, Kenji Imamura and Yoshihiro Matsuo, Constructing a Class-Based Lexical Dictionary using Interactive Topic Models	2590
Verginica Barbu Mititelu, Adding Morpho-semantic Relations to the Romanian Wordnet	2596
Julien Seinturier, Elisabeth Murisasco, Emmanuel Bruno and Philippe Blache, An ontological approach to model and query multimodal concurrent linguistic annotations	2602
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	2606
Inga Gheorghita and Jean-Marie Pierrel, Towards a methodology for automatic identification of hypernyms in the definitions of large-scale dictionary	2614
John McCrae, Elena Montiel-Ponsoda and Philipp Cimiano, Collaborative semantic editing of linked data lexica	2619
Christophe Roche, Ontoterminology: How to unify terminology and ontology into a single paradigm	2626
Alexandre Denis, Ingrid Falk, Claire Gardent and Laura Perez-Beltrachini, Representation of linguistic and domain knowledge for second language learning in virtual worlds	2631
Petya Osenova, Kiril Simov, Laska Laskova and Stanislava Kancheva, A Treebank-driven Creation of an OntoValence Verb lexicon for Bulgarian	2636
Elisa Bianchi, Mirko Tavosanis and Emiliano Giovannetti, Creation of a bottom-up corpus-based ontology for Italian Linguistics	2641
Matteo Abrate and Clara Bacciu, Visualizing word senses in WordNet Atlas	2648

SESSION O29 - Language Generation and Paraphrasing

Houda Bouamor, Aurélien Max, Gabriel Illouz and Anne Vilnat, A contrastive review of paraphrase acquisition techniques	2653
Matteo Negri, Yashar Mehdad, Alessandro Marchetti, Danilo Giampiccolo and Luisa Bentivogli, Chinese Whispers: Cooperative Paraphrase Acquisition	2659
Hideki Shima and Teruko Mitamura, Diversifiable Bootstrapping for Acquiring High-Coverage Paraphrase Resource	2666
Sebastian Vargas, Heike Bieler, Manfred Stede, Lukas C. Faulstich, Kristin Irsig and Malik Atalla, SemScribe: Natural Language Generation for Medical Reports	2674

SESSION O30 - Computer Aided Language Learning

Hitokazu Matsushita and Deryle Lonsdale, Item Development and Scoring for Japanese Oral Proficiency Testing	2682
Manny Rayner, Pierrette Bouillon and Johanna Gerlach, Evaluating Appropriateness Of System Responses In A Spoken CALL Game	2690
Antonio Moreno-Sandoval, Leonardo Campillos Llanos, Yang Dong, Emi Takamori, José M. Guirao, Paula Gozalo, Chieko Kimura, Kengo Matsui and Marta Garrote-Salazar,	

Spontaneous Speech Corpora for language learners of Spanish, Chinese and Japanese.....	2695
Helmer Strik, Jozef Colpaert, Joost Van Doremalen and Catia Cucchiarini , The DISCO ASR-based CALL system: practicing L2 oral skills and beyond.....	2702

SESSION O31 - Discourse (2)

Marta Tatu and Dan Moldovan , A Tool for Extracting Conversational Implicatures.....	2708
Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni and Sandrine Zufferey , Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns	2716
Steven Bethard, Oleksandr Kolomiyets and Marie-Francine Moens , Annotating Story Timelines as Temporal Dependency Structures.....	2721
Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cecile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paul Pery-Woodley, Laurent Prevot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret and Laure Vieu , An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus.....	2727

SESSION O32 - Syntax and Parsing

Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský and Jan Hajič , HamleDT: To Parse or Not to Parse?.....	2735
Elsa Tolone, Benoît Sagot and Éric Villemonte de la Clergerie , Evaluating and improving syntactic lexica by plugging them within a parser.....	2742
Thomas Proisl and Peter Uhrig , Efficient Dependency Graph Matching with the IMS Open Corpus Workbench.....	2750
Miguel Ballesteros and Joakim Nivre , MaltOptimizer: A System for MaltParser Optimization	2757

SESSION P30 - Discourse

Kristiina Jokinen and Silvi Tenjes , Investigating Engagement - intercultural and technological aspects of the collection, analysis, and use of the Estonian Multiparty Conversational video data	2764
Patrick Saint-Dizier , DISLOG: A logic-based language for processing discourse structures .	2770
Sarah Bourse and Patrick Saint-Dizier , A Repository of Rules and Lexical Resources for Discourse Structure Analysis: the Case of Explanation Structures	2778
Stefania Degaetano-Ortlieb, Ekaterina Lapshinova-Koltunski and Elke Teich , Feature Discovery for Diachronic Register Analysis: a Semi-Automatic Approach.....	2786
Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli , Improving the Recall of a Discourse Parser by Constraint-based Postprocessing	2791
Elizabeth Baran, Yaqin Yang and Nianwen Xue , Annotating dropped pronouns in Chinese newswire text.....	2795
Magdalena Rysova , Alternative Lexicalizations of Discourse Connectives in Czech.....	2800

Utku Şirin, Ruket Çakıcı and Deniz Zeyrek, METU Turkish Discourse Bank Browser..	2808
David Elson, DramaBank: Annotating Agency in Narrative Discourse.....	2813
Gisela Redeker, Ildikó Berzlánovich, Nynke van der Vliet, Gosse Bouma and Markus Egg, Multi-Layer Discourse Annotation of a Dutch Text Corpus.....	2820
Iskandar Keskes, Farah Benamara and Lamia Hadrach Belguith, Clause-based Discourse Segmentation of Arabic Texts.....	2826
Mariana Gomes, Ana Guilherme, Leonor Tavares and Rita Marquilhas, Project FLY: a multidisciplinary project within Linguistics	2833
Ching-Sheng Lin, Zumurut Akcam, Samira Shaikh, Sharon Small, Ken Stahl, Tomek Strzalkowski and Nick Webb, Revealing Contentious Concepts Across Social Groups.....	2838

SESSION P31 - Lexical Acquisition

Tommaso Caselli, Francesco Rubino, Francesca Frontini, Irene Russo and Valeria Quochi, Customizable SCF Acquisition in Italian.....	2842
Gregor Thurmair, Vera Aleksic and Christoph Schwarz, Large Scale Lexical Analysis	2849
Elsa Tolone, Stavroula Voyatzi, Claude Martineau and Matthieu Constant, Extending the adverbial coverage of a French morphological lexicon.....	2856
Somayeh Bagherbeygi and Mehrnoush Shamsfard, Corpus based Semi-Automatic Extraction of Persian Compound Verbs and their Relations	2863

SESSION P32 - Corpus Creation, Processing, Usage (2)

Ting Liu, Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stromer-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.....	2868
Ivana Tanasijević, Biljana Sikimić and Gordana Pavlović-Lažetić, Multimedia database of the cultural heritage of the Balkans.....	2874
Rania Al-Sabbagh and Roxana Girju, YADAC: Yet another Dialectal Arabic Corpus....	2882
Yves Scherrer and Bruno Cartoni, The Trilingual ALLEGRA Corpus: Presentation and Possible Use for Lexicon Induction	2890
Martin Reynaert, Ineke Schuurman, Veronique Hoste, Nelleke Oostdijk and Maarten van Gompel, Beyond SoNaR: towards the facilitation of large corpus building efforts	2897
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.....	2905
Kurt Eberle, Kerstin Eckart, Ulrich Heid and Boris Haselbach, A Tool/Database Interface for Multi-Level Analyses.....	2912
Djamel Mostefa, Khalid Choukri, Sylvie Brunessaux, Karim Boudahmane, New language resources for the Pashto language.....	2917
Şenay Kafkas, Ian Lewin, David Milward, Erik van Mulligen, Jan Kors, Udo Hahn and Dietrich Rebolz-Schuhmann, CALBC: Releasing the Final Corpora.....	2923

Martin Majliš and Zdeněk Žabokrtský, Language Richness of the Web 2927

SESSION P33 - Web Services

Markus Forsberg and Torbjörn Lager, Cloud Logic Programming for Integrating Language Technology Resources..... 2935

Marc Kemps-Snijders, Matthijs Brouwer, Jan Pieter Kunst and Tom Visser, Dynamic web service deployment in a cloud environment 2941

Bharat Ram Ambati, Siva Reddy and Adam Kilgarriff, Word Sketches for Turkish ... 2945

Chunqi Shi, Donghui Lin and Toru Ishida, Service Composition Scenarios for Task-Oriented Translation 2951

Aleksandar Savkov, Laska Laskova, Stanislava Kancheva, Petya Osenova and Kiril Simov, Linguistic Analysis Processing Line for Bulgarian 2959

Victoria Arranz and Olivier Hamon, On the Way to a Legal Sharing of Web Applications in NLP..... 2965

Rafal Rak, Andrew Rowley and Sophia Ananiadou, Collaborative Development and Evaluation of Text-processing Workflows in a UIMA-supported Web-based Workbench..... 2971

Javier Caminero, Mari Carmen Rodríguez, Jean Vanderdonckt, Fabio Paternò, Joerg Rett, Dave Raggett, Jean-Loup Comeliau and Ignacio Marín, The SERENOA Project: Multidimensional Context-Aware Adaptation of Service Front-Ends 2977

SESSION O33 - Semantics from Corpora

Alex Judea, Vivi Nastase and Michael Strube, Concept-based Selectional Preferences and Distributional Representations from Wikipedia Articles 2985

Elias Iosif, Maria Giannoudaki, Eric Fosler-Lussier and Alexandros Potamianos, Associative and Semantic Features Extracted From Web-Harvested Corpora 2991

Octavian Popescu, Building a Resource of Patterns Using Semantic Types..... 2999

SESSION O34 - Authoring and Related Tools

Irina Temnikova, Constantin Orasan and Ruslan Mitkov, CLCM - A Linguistic Resource for Effective Simplification of Instructions in the Crisis Management Domain and its Evaluations 3007

Robert Dale and George Narroway, A Framework for Evaluating Text Correction..... 3015

Paul Rodrigues and C. Anton Rytting, Typing Race Games as a Method to Create Spelling Error Corpora 3019

SESSION O35 - Word Sense Annotation and Disambiguation

Rebecca J. Passonneau, Collin F. Baker, Christiane Fellbaum and Nancy Ide, The MASC Word Sense Corpus..... 3025

Darja Fišer, Nikola Ljubešić and Ozren Kubelka, Addressing polysemy in bilingual lexicon extraction from comparable corpora 3031

Gerard de Melo, Collin F. Baker, Nancy Ide, Rebecca J. Passonneau and Christiane Fellbaum, Empirical Comparisons of MASC Word Sense Annotations..... 3036

SESSION O36 - Time and Space

Hector Llorens, Leon Derczynski, Robert Gaizauskas and Estela Saquete, TIMEN: An Open Temporal Expression Normalisation Resource..... 3044

Kirk Roberts, Travis Goodwin and Sanda M. Harabagiu, Annotating Spatial Containment Relations Between Events..... 3052

James Pustejovsky and Jessica Moszkowicz, The Role of Model Testing in Standards Development: The Case of ISO-Space..... 3060

SESSION O37 - Subjectivity and Emotions

Jörg Frommer, Bernd Michaelis, Dietmar Rösner, Andreas Wendemuth, Rafael Friesen, Matthias Haase, Manuela Kunze, Rico Andrich, Julia Lange, Axel Panning and Ingo Siegert, Towards Emotion and Affect Detection in the Multimodal LAST MINUTE Corpus . 3064

Isa Maks and Piek Vossen, Building a fine-grained subjectivity lexicon from a web corpus 3070

Veronica Perez-Rosas, Carmen Banea and Rada Mihalcea, Learning Sentiment Lexicons in Spanish..... 3077

Tommaso Caselli, Irene Russo and Francesco Rubino, Assigning Connotation Values to Events..... 3082

Balamuraliar, Aditya Joshi and Pushpak Bhattacharyya, Cost and Benefit of Using WordNet Senses for Sentiment Analysis..... 3090

SESSION O38 - Named Entities

Xuansong Li, Stephanie Strassel, Heng Ji, Kira Griffitt and Joe Ellis, Linguistic Resources for Entity Linking Evaluation: from Monolingual to Cross-lingual..... 3098

Dawn Lawrie, James Mayfield, Paul McNamee and Douglas Oard, Creating and Curating a Cross-Language Person-Entity Linking Collection..... 3106

Keith J. Miller, Elizabeth Schroeder Richerson, Sarah McLeod, James Finley and Aaron Schein, International Multicultural Name Matching Competition: Design, Execution, Results, and Lessons Learned..... 3111

K Saravanan, Monojit Choudhury, Raghavendra Udupa and A Kumaran, An Empirical Study of the Occurrence and Co-Occurrence of Named Entities in Natural Language Corpora 3118

Olivier Galibert, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum and Ludovic Quintard, Extended Named Entities Annotation on OCRed Documents: From Corpus Constitution to Evaluation Campaign..... 3126

SESSION O39 - Treebanks and Syntax

Wolfgang Seeker and Jonas Kuhn, Making Ellipses Explicit in Dependency Conversion for a German Treebank..... 3132

Hongzhi Xu, Helen Kaiyun Chen, Chu-Ren Huang, Qin Lu, Dingxu Shi and Tin-Shing Chiu , A Grammar-informed Corpus-based Sentence Database for Linguistic and Computational Studies	3140
Tafseer Ahmed, Miriam Butt, Annette Hautli and Sebastian Sulger , A Reference Dependency Bank for Analyzing Complex Predicates	3145
Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová and Zdeněk Žabokrtský , Announcing Prague Czech-English Dependency Treebank 2.0	3153
Liesbeth Augustinus, Vincent Vandeghinste and Frank Van Eynde , Example-Based Treebank Querying	3161

SESSION O40 - Semantic Lexicons and Semantic Annotation

Valentin I. Spitkovsky and Angel X. Chang , A Cross-Lingual Dictionary for English Wikipedia Concepts	3168
Silvie Cinková, Martin Holub, Adam Rambousek and Lenka Smejkalová , A database of semantic clusters of verb usages	3176
Ernesto William De Luca , Is it Useful to Support Users with Lexical Resources? A User Study. 3184	
Natalia Konstantinova, Sheila C.M. de Sousa, Noa P. Cruz, Manuel J. Maña, Maite Taboada and Ruslan Mitkov , A review corpus annotated for negation, speculation and their scope	3190
Valerio Basile, Johan Bos, Kilian Evang and Noortje Venhuizen , Developing a large semantically annotated corpus	3196

SESSION P34 Corpus Creation, Processing, Usage (3)

Ronaldo Martins , Le Petit Prince in UNL	3201
Christian Chiarcos , A generic formalism to represent linguistic corpora in RDF and OWL/DL 3205	
Silvia Pareti , A Database of Attribution Relations	3213
Bartosz Broda, Michał Marcińczuk, Marek Maziarz, Adam Radziszewski and Adam Wardyński , KPWr: Towards a Free Corpus of Polish	3218
Yeşim Aksan, Mustafa Aksan, Ahmet Koltuksuz, Taner Sezer, Ümit Mersinli, Umut Ufuk Demirhan, Hakan Yilmazer, Gülsüm Atasoy, Seda Öz, İpek Yıldız and Özlem Kurtoğlu , Construction of the Turkish National Corpus (TNC)	3223
Jirka Hana, Alexandr Rosen, Barbora Štindlová and Petr Jäger , Building a learner corpus 3228	
Isabella Poggi, Francesca D’Errico and Giovanna Leone , Pedagogical stances and their multimodal signals	3233
Tsuyoshi Okita , Annotated Corpora for Word Alignment between Japanese and English and its Evaluation with MAP-based Word Aligner	3241

Djamé Seddah, Marie Candito, Benoit Crabbé and Enrique Henestroza Anguiano, Ubiquitous Usage of a Broad Coverage French Corpus: Processing the Est Republicain corpus 3249

SESSION P35 - Language Resource Infrastructures (2)

Herman Stehouwer, Matej Durco, Eric Auer and Daan Broeder, Federated Search: Towards a Common Search Infrastructure. 3255

Willem Elbers, Daan Broeder and Dieter van Uytvanck, Proper Language Resource Centers 3260

Sebastian Drude, Daan Broeder, Paul Trilsbeek and Peter Wittenburg, The Language Archive – a new hub for language resources 3264

Kais Dukes and Eric Atwell, LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis 3268

James Clarke, Vivek Srikumar, Mark Sammons and Dan Roth, An NLP Curator (or: How I Learned to Stop Worrying and Love NLP Pipelines) 3276

Marta Villegas, N ria Bel, Carlos Gonzalo, Amparo Moreno and Nuria Simelio, Using Language Resources in Humanities research. 3284

Sebastian Nordhoff and Harald Hammarstr m, Glottolog/Langdoc: Increasing the visibility of grey literature for low-density languages 3289

Steve Cassidy, Michael Haugh, Pam Peters and Mark Fallu, The Australian National Corpus: National Infrastructure for Language Resources 3295

Christian Federmann, Ioanna Giannopoulou, Christian Girardi, Olivier Hamon, Dimitris Mavroeidis, Salvatore Minutoli and Marc Schr der, META-SHARE v2: An Open Network of Repositories for Language Resources including Data and Tools 3300

Alessio Bosca, Luca Dini, Milen Kouylekov and Marco Trevisan, Linguagrid: a network of Linguistic and Semantic Services for the Italian Language. 3304

SESSION P36 - Speech Synthesis

 naki Sainz, Daniel Erro, Eva Navas, Inma Hern ez, Jon S nchez, Ibon Saratxaga and Igor Odriozola, Versatile Speech Databases for High Quality Synthesis for Basque. 3308

Dietmar Schabus, Michael Pucher and Gregor Hofer, Building a synchronous corpus of acoustic and 3D facial marker data for adaptive audio-visual speech synthesis 3313

Alistair Conkie, Thomas Okken, Yeon-Jun Kim and Giuseppe Di Fabrizio, Building Text-To-Speech Voices in the Cloud. 3317

Emilia Garcia Casademont, Antonio Bonafonte and Asunci n Moreno, Building Synthetic Voices in the META-NET Framework. 3322

Nur-Hana Samsudin and Mark Lee, Building Text-to-Speech Systems for Resource Poor Languages. 3327

Eva Szekely, Joao Paulo Cabral, Mohamed Abou-Zleikha, Peter Cahill and Julie Carson-Berndsen, Evaluating expressive speech synthesis from audiobook corpora for conversational phrases 3335

SESSION P37 - Speech Resources

Panikos Heracleous, Carlos Ishi, Takahiro Miyashita and Norihiro Hagita, Body-conductive acoustic sensors in human-robot communication	3340
Lucie Válková, Martina Waclawičová and Michal Křen, Balanced data repository of spontaneous spoken Czech.....	3345
R.P. Clapham, L. van der Molen, R.J.JH. van Son, M. van den Brekel and F.J.M. Hilgers, NKI-CCRT Corpus - Speech Intelligibility Before and After Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy	3350
Thomas Ulrich Christiansen and Peter Juel Henrichsen, Sense Meets Nonsense - Sense Meets Nonsense - a dual-layer Danish speech corpus for perception studies.....	3356
Peter Juel Henrichsen and Marcus Uneson, SMALLWorlds – Multilingual Content-Controlled Monologues.....	3362
Alexander Schmitt, Stefan Ultes and Wolfgang Minker, A Parameterized and Annotated Spoken Dialog Corpus of the CMU Let’s Go Bus Information System	3369
Sergey Zablotkiy, Alexander Shvets, Maxim Sidorov, Eugene Semenkin and Wolfgang Minker, Speech and Language Resources for LVCSR of Russian.....	3374
Dae-Lim Choi, Bong-Wan Kim, Yeon-Whoa Kim, Yong-Ju Lee, Yongnam Um and Minhwa Chung, Dysarthric Speech Database for Development of QoLT Software Technology	3378
Eckhard Bick, Heliana Mello, Alessandro Panunzi and Tommaso Raso, The annotation of the C-ORAL-BRASIL oral through the implementation of the Palavras Parser	3382
Janne Bondi Johannessen, Joel Priestley, Kristin Hagen, Anders Nøklestad and André Lynum, The Nordic Dialect Corpus.....	3387
Dafydd Gibbon, ULex: new data models and a mobile environment for corpus enrichment..	3392
Kengo Ohta, Masatoshi Tsuchiya and Seiichi Nakagawa, Developing Partially-Transcribed Speech Corpus from Edited Transcriptions.....	3399
Xiaoyi Ma, LDC Forced Aligner.....	3405
Sebastian Stüker, Florian Kraft, Christian Mohr, Teresa Herrmann, Eunah Cho and Alex Waibel, The KIT Lecture Corpus for Speech Translation.....	3409
Shyam Agrawal, Shweta Sinha, Pooja Singh and Jesper Olson, Development of Text and Speech database for Hindi and Indian English specific to Mobile Communication environment	3415

SESSION O41 - Machine Translation and Language Resources (2)

Víctor M. Sánchez-Cartagena, Miquel Esplà-Gomis and Juan Antonio Pérez-Ortiz, Source-Language Dictionaries Help Non-Expert Users to Enlarge Target-Language Dictionaries for Machine Translation.....	3422
Christian Federmann, Eleftherios Avramidis, Marta R. Costa-Jussà, Josef van Genabith, Maite Melero and Pavel Pecina, The ML4HMT Workshop on Optimising the Division of Labour in Hybrid Machine Translation.....	3430
Maria Holmqvist, Sara Stymne, Lars Ahrenberg and Magnus Merkel, Alignment-based reordering for SMT.....	3436
Monica Gavrilă, Walther v. Hahn and Cristina Vertan, Same domain different discourse style - A case study on Language Resources for data-driven Machine Translation.....	3441

Hidetsugu Nanba, Toshiyuki Takezawa, Kiyoko Uchiyama and Akiko Aizawa, Automatic Translation of Scholarly Terms into Patent Terms Using Synonym Extraction Techniques 3447

SESSION O42 - WordNets

Sanni Nimb and Bolette Sandford Pedersen, Towards a richer wordnet representation of properties 3452

Aitor Gonzalez-Agirre, Mauro Castillo and German Rigau, A proposal for improving WordNet Domains 3457

Fernando Castilho, Roger Granada, Breno Meneghetti, Leonardo Carvalho and Renata Vieira, Corpus+WordNet thesaurus generation for ontology enriching 3463

Benoît Sagot and Darja Fišer, Cleaning noisy wordnets 3468

Valérie Hanoka and Benoît Sagot, Wordnet extension made simple: A multilingual lexicon-based approach using wiki resources 3473

SESSION O43 - Text Mining

Mathias Bank and Martin Schierle, A Survey of Text Mining Architectures and the UIMA Standard 3479

Diana Maynard and Mark A. Greenwood, Large Scale Semantic Annotation, Indexing and Search at The National Archives 3487

Georgeta Bordea, Sabrina Kirrane, Paul Buitelaar and Bianca Pereira, Expertise Mining for Enterprise Content Management 3495

Elias Iosif and Alexandros Potamianos, SemSim: Resources for Normalized Semantic Similarity Computation Using Lexical Networks 3499

Raheel Nawaz, Paul Thompson and Sophia Ananiadou, Identification of Manner in Bio-Events 3505

SESSION O44 - Evaluation of Systems and Application

Ioana Vasilescu, Martine Adda-Decker and Lori Lamel, Cross-lingual studies of ASR errors: paradigms for perceptual evaluations 3511

Kallirroi Georgila, Alan Black, Kenji Sagae and David Traum, Practical Evaluation of Human and Synthesized Speech for Virtual Human Dialogue Systems 3519

Tomoyosi Akiba, Hiromitsu Nishizaki, Kiyooki Aikawa, Tatsuya Kawahara and Tomoko Matsui, Designing an Evaluation Framework for Spoken Term Detection and Spoken Document Retrieval at the NTCIR-9 SpokenDoc Task 3527

Tina Kluewer, Feiyu Xu, Peter Adolphs and Hans Uszkoreit, Evaluation of the KomParse Conversational Non-Player Characters in a Commercial Virtual World 3535

Marcello Federico, Sebastian Stüker, Luisa Bentivogli, Michael Paul, Mauro Cettolo, Teresa Herrmann, Jan Niehues and Giovanni Moretti, The IWSLT 2011 Evaluation Campaign on Automatic Talk Translation 3543

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

Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger and Michael Wiegand, MLSA – A Multi-layered Reference Corpus for German Sentiment Analysis	3551
Silvia Vázquez and Núria Bel, A Classification of Adjectives for Polarity Lexicons Enhancement	3557
Jorge Carrillo de Albornoz, Laura Plaza and Pablo Gervás, SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis	3562
Tom De Smedt and Walter Daelemans, “Vreselijk mooi!” (terribly beautiful): A Subjectivity Lexicon for Dutch Adjectives	3568
Rasmus Sundberg, Anders Eriksson, Johan Bini and Pierre Nugues, Visualizing Sentiment Analysis on a User Forum	3573
Erik Cambria, Yunqing Xia and Amir Hussain, Affective Common Sense Knowledge Acquisition for Sentiment Analysis	3580

SESSION P39 - Language Resource Infrastructures (3)

Emanuel Dima, Verena Henrich, Erhard Hinrichs, Marie Hinrichs, Christina Hoppermann, Thorsten Trippel, Thomas Zastrow and Claus Zinn, A Repository for the Sustainable Management of Research Data	3586
Maciej Ogrodniczuk, Piotr Pezik and Adam Przepiórkowski, Towards a comprehensive open repository of Polish language resources	3593
Lars Borin, Markus Forsberg, Leif-Jöran Olsson and Jonatan Uppström, The open lexical infrastructure of Språkbanken	3598
Christian Chiarcos, Sebastian Hellmann, Sebastian Nordhoff, Steven Moran, Richard Littauer, Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek and Christian M. Meyer, The Open Linguistics Working Group	3603
Silke Scheible, Richard J. Whitt, Martin Durrell and Paul Bennett, GATEtoGerManC: A GATE-based Annotation Pipeline for Historical German	3611
Nicolas Hernandez, Tackling interoperability issues within UIMA workflows	3618

SESSION P40 - Knowledge and Ontologies

Anne-Kathrin Schumann, Knowledge-Rich Context Extraction and Ranking with KnowPipe	3626
Maria Teresa Paziienza, Noemi Scarpato and Armando Stellato, Application of a Semantic Search Algorithm to Semi-Automatic GUI Generation	3631
Roldano Cattoni, Francesco Corcoglioniti, Christian Girardi, Bernardo Magnini, Luciano Serafini and Roberto Zanolli, The KnowledgeStore: an Entity-Based Storage System	3639
Bartosz Broda, Marek Maziarz and Maciej Piasecki, Tools for plWordNet Development. Presentation and Perspectives	3647
Silvia Moraes and Vera Lima, Combining Formal Concept Analysis and semantic information for building ontological structures from texts : an exploratory study	3653
Menzo Windhouwer, RELcat: a Relation Registry for ISOcat data categories	3661
Eric Charton and Michel Gagnon, A disambiguation resource extracted from Wikipedia for	

semantic annotation	3665
Guido Boella, Luigi di Caro, Llio Humphreys, Livio Robaldo and Leon van der Torre, NLP Challenges for Eunomos a Tool to Build and Manage Legal Knowledge	3672
Robert Speer and Catherine Havasi, Representing General Relational Knowledge in ConceptNet 5	3679

SESSION P41 - Semantics

Alain Joubert and Mathieu Lafourcade, A new dynamic approach for lexical networks evaluation	3687
Roberta Catizone, Louise Guthrie, Arthur Thomas and Yorick Wilks, LIE: Leadership, Influence and Expertise	3692
Richard Johansson, Karin Friberg Heppin and Dimitrios Kokkinakis, Semantic Role Labeling with the Swedish FrameNet	3697
Pedro Fialho, Sérgio Curto, Ana Cristina Mendes and Luísa Coheur, Extending a wordnet framework for simplicity and scalability	3701
Boris Haselbach, Wolfgang Seeker and Kerstin Eckart, German “”nach””-Particle Verbs in Semantic Theory and Corpus Data	3706
Alessandro Lenci, Gabriella Lapesa and Giulia Bonansinga, LexIt: A Computational Resource on Italian Argument Structure	3712
Alessandro Lenci, Simonetta Montemagni, Giulia Venturi and Maria Grazia Cutrullà, Enriching the ISST-TANL Corpus with Semantic Frames	3719

SESSION P42 - Temporal Information

Francisco Costa and António Branco, TimeBankPT: A TimeML Annotated Corpus of Portuguese	3727
Angel X. Chang and Christopher Manning, SUTime: A library for recognizing and normalizing time expressions	3735
André Bittar, Caroline Hagège, Véronique Moriceau, Xavier Tannier and Charles Teissèdre, Temporal Annotation: A Proposal for Guidelines and an Experiment with Inter-annotator Agreement	3741
Jannik Strötgen and Michael Gertz, Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards	3746
Leon Derczynski, Hector Llorens and Estela Saquete, Massively Increasing TIMEX3 Resources: A Transduction Approach	3754
Corina Forascu and Dan Tufiş, Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information	3762

SESSION P43 - Sign Language

Zoya Gavrilov, Stan Sclaroff, Carol Neidle and Sven Dickinson, Detecting Reduplication in Videos of American Sign Language	3767
---	------

Nedelina Ivanova and Olle Eriksen , BiBiKit - A Bilingual Bimodal Reading and Writing Tool for Sign Language Users	3774
Fabrizio Borgia, Claudia S. Bianchini, Patrice Dalle and Maria De Marsico , Resource production of written forms of Sign Languages by a user-centered editor, SWift (SignWriting improved fast transcriber)	3779
Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater and Hermann Ney , RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus	3785

SESSION O45 - New Media (Special Session)

Eleanor Clark and Kenji Araki , Two Database Resources for Processing Social Media English Text	3790
Maite Melero, Marta R. Costa-Jussà, Judith Domingo, Montse Marquina and Martí Quixal , Holaaa!! writin like u talk is kewl but kinda hard 4 NLP	3794
William J. Corvey, Sudha Verma, Sarah Vieweg, Martha Palmer and James H. Martin , Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events	3801
Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie and Sanda M. Harabagiu , EmpaTweet: Annotating and Detecting Emotions on Twitter	3806

SESSION O46 - Semantics, Knowledge and Ontologies

Nava Maroto, Marie-Claude L’Homme and Amparo Alcina , Semantic Relations Established by Specialized Processes Expressed by Nouns and Verbs: Identification in a Corpus by means of Syntactico-semantic Annotation	3814
Jorge Vivaldi, Luis Adrián Cabrera-Diego, Gerardo Sierra and María Pozzi , Using Wikipedia to Validate the Terminology found in a Corpus of Basic Textbooks	3820
Maria Teresa Pazienza, Armando Stellato and Andrea Turbati , PEARL: ProjECTION of Annotations Rule Language, a Language for Projecting (UIMA) Annotations over RDF Knowledge Bases	3828
Peter Exner and Pierre Nugues , Constructing Large Proposition Databases	3836
Montse Cuadros, Lluís Padró and German Rigau , Highlighting relevant concepts from Topic Signatures	3841

SESSION O47 - Segmentation, Tagging, Parsing

Agnieszka Patejuk and Adam Przepiórkowski , Towards an LFG parser for Polish: An exercise in parasitic grammar development	3849
Yan Song and Fei Xia , Using a Goodness Measurement for Domain Adaptation: A Case Study on Chinese Word Segmentation	3853
Daniel Bauer, Hagen Fürstenau and Owen Rambow , The Dependency-Parsed FrameNet Corpus	3861

Majdi Sawalha, Claire Brierley and Eric Atwell, Predicting Phrase Breaks in Classical and Modern Standard Arabic Text	3868
Sudheer Kolachina and Prasanth Kolachina, Parsing Any Domain English text to CoNLL dependencies	3873

SESSION O48 - Named Entities and Subjectivity

Udo Hahn, Elena Beisswanger, Ekaterina Buyko, Erik Faessler, Jenny Traumüller, Susann Schröder and Kerstin Hornbostel, Iterative Refinement and Quality Checking of Annotation Guidelines — How to Deal Effectively with Semantically Sloppy Named Entity Types, such as Pathological Phenomena	3881
Danuta Ploch, Leonhard Hennig, Angelina Duka, Ernesto William De Luca and Sahin Albayrak, GerNED: A German Corpus for Named Entity Disambiguation	3886
Ian Lewin, Şenay Kafkas and Dietrich Rebholz-Schuhmann, Centroids: Gold standards with distributional variation	3894
Yulan He, Hassan Saif, Zhongyu Wei and Kam-Fai Wong, Quantising Opinions for Political Tweets Analysis	3901
Muhammad Abdul-Mageed and Mona Diab, AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis	3907

SESSION P44 - Machine Translation (2)

Saab Mansour and Hermann Ney, Arabic-Segmentation Combination Strategies for Statistical Machine Translation	3915
Ondřej Bojar, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel and Aleš Tamchyna, The Joy of Parallelism with CzEng 1.0	3921
Takanori Kusumoto and Tomoyosi Akiba, Statistical Machine Translation without Source-side Parallel Corpus Using Word Lattice and Phrase Extension	3929
Lambert Patrik, Holger Schwenk and Frédéric Blain, Automatic Translation of Scientific Documents in the HAL Archive	3933
Georgi Iliev and Angel Genov, Expanding Parallel Resources for Medium-Density Languages for Free	3937
Elisabet Comelles, Jordi Atserias, Victoria Arranz and Irene Castellón, VERTa: Linguistic features in MT evaluation	3944
Zhiyi Song, Safa Ismael, Stephen Grimes, David Doermann and Stephanie Strassel, Linguistic Resources for Handwriting Recognition and Translation Evaluation	3951
Fangzhong Su and Bogdan Babych, Development and Application of a Cross-language Document Comparability Metric	3956
Claire Jaja, Douglas Briesch, Jamal Laoudi and Clare Voss, Assessing Divergence Measures for Automated Document Routing in an Adaptive MT System	3963
Ananthakrishnan Ramanathan and Karthik Visweswariah, A Study of Word-Classing for MT Reordering	3971

João Silva, Luísa Coheur, Ângela Costa and Isabel Trancoso , Dealing with unknown words in statistical machine translation.....	3977
Wilker Aziz, Sheila Castilho and Lucia Specia , PET: a Tool for Post-editing and Assessing Machine Translation.....	3982
Chris Irwin Davis , Tajik-Farsi Persian Transliteration Using Statistical Machine Translation	3988
Emma Barker and Robert Gaizauskas , Assessing the Comparability of News Texts	3996

SESSION P45 - Natural Language Generation

Hilder Pereira, Eder Novais, Andre Mariotti and Ivandre Paraboni , Corpus-based Referring Expressions Generation.....	4004
Eder Novais, Ivandre Paraboni and Douglas Silva , Portuguese Text Generation from Large Corpora	4010
Sigrid Klerke and Anders Søgaard , DSim, a Danish Parallel Corpus for Text Simplification	4015
Violeta Seretan , Acquisition of Syntactic Simplification Rules for French.....	4019
Anja Belz and Albert Gatt , A Repository of Data and Evaluation Resources for Natural Language Generation.....	4027
Eric Kow and Anja Belz , LG-Eval: A Toolkit for Creating Online Language Evaluation Experiments	4033

SESSION P46 - Crowdsourcing

Chris Biemann , Turk Bootstrap Word Sense Inventory 2.0: A Large-Scale Resource for Lexical Substitution.....	4038
Marion Potet, Emmanuelle Esperança-Rodier, Laurent Besacier and Hervé Blanchon , Collection of a Large Database of French-English SMT Output Corrections	4043
Jirka Hana and Barbora Hladka , Getting more data – Schoolkids as annotators	4049
Anna Rumshisky, Nick Botchan, Sophie Kushkuley and James Pustejovsky , Word Sense Inventories by Non-Experts.	4055

SESSION P47 - Text Mining and Text Entailment

Anton Leuski, Carsten Eickhoff, James Ganis and Victor Lavrenko , The BladeMistress Corpus: From Talk to Action in Virtual Worlds.....	4060
Alvin Grissom II and Yusuke Miyao , Annotating Factive Verbs.....	4068
Mahdi Khademian, Kaveh Taghipour, Saab Mansour and Shahram Khadivi , A Holistic Approach to Bilingual Sentence Fragment Extraction from Comparable Corpora	4073
Mohammad Fazleh Elahi and Paola Monachesi , An Examination of Cross-Cultural Similarities and Differences from Social Media Data with respect to Language Use.....	4080
Seniz Demir, Ilknur Durgar El-Kahlout, Erdem Unal and Hamza Kaya , Turkish Paraphrase Corpus	4087
Rui Wang and Shuguang Li , Constructing a Question Corpus for Textual Semantic Relations	4092

Alexandra Roshchina, John Cardiff and Paolo Rosso, Evaluating the Similarity Estimator component of the TWIN Personality-based Recommender System..... 4098

SESSION P48 - Speech/Multimodal Tools, Systems, Applications

Michael Kipp, Annotation Facilities for the Reliable Analysis of Human Motion..... 4103

Michael Carl, Translog-II: a Program for Recording User Activity Data for Empirical Reading and Writing Research..... 4108

Giovanni Costantini, Andrea Paoloni and Massimiliano Todisco, Intelligibility assessment in forensic applications..... 4113

David Tavaréz, Eva Navas, Daniel Erro and Ibon Saratxaga, Strategies to Improve a Speaker Diarisation Tool..... 4117

Igor Odriozola, Eva Navas, Inma Hernández, Iñaki Sainz, Ibon Saratxaga, Jon Sánchez and Daniel Erro, Using an ASR database to design a pronunciation evaluation system in Basque 4122

Francesco Cutugno, Vincenza Anna Leano and Antonio Origlia, W-PhAMT: A web tool for phonetic multilevel timeline visualization..... 4127

Amalia Zahra and Julie Carson-Berndsen, English to Indonesian Transliteration to Support English Pronunciation Practice..... 4132

Joao Paulo Cabral, Mark Kane, Zeeshan Ahmed, Mohamed Abou-Zleikha, Eva Szekely, Amalia Zahra, Kalu Ogbureke, Peter Cahill, Julie Carson-Berndsen and Stephan Schlogl, Rapidly Testing the Interaction Model of a Pronunciation Training System via Wizard-of-Oz. 4136

Bernhard Brüning, Christian Schnier, Karola Pitsch and Sven Wasmuth, PAMOCAT: Automatic retrieval of specified postures..... 4143

HunOr A Hungarian-Russian Parallel corpus comprises approximately 800 thousand words. CoNLL 2017 Shared Task Hungarian data Automatic segmentation, tokenization and morphological and syntactic annotations of raw texts from the Common Crawl. CSS10 A Collection of Single Speaker Speech Datasets for 10 Languages including Hungarian. In this paper, we present HunOr, the first multi-domain Hungarian-Russian parallel corpus. Some of the corpus texts have been manually aligned and split into sentences, besides, named entities also have been annotated while the other parts are automatically aligned at the sentence level and they are POS-tagged as well. The corpus contains texts from the domains literature, official language use and science, however, we would like to add texts from the news domain to the corpus. In the future

CONTINUE READING.