8th International Conference on Language Resources and Evaluation 2012

(LREC-2012)

Istanbul, Turkey
21-27 May 2012

Volume 1 of 5

ISBN: 978-1-62276-504-1

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2012) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2012)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

Volume 1

PaCo^2: A Fully Automated Tool for Gathering Parallel Corpora from the Web 1
Iñaki San Vicente, Iker Manterola

Terra: A Collection of Translation Error-Annotated Corpora 7
Mark Fishel, Ondrej Bojar, Maja Popovic

A Light Way to Collect Comparable Corpora from the Web 15
Ahmet Aker, Evangelos Kanoulas, Robert Gaizauskas

SUMAT: Data Collection and Parallel Corpus Compilation for Machine Translation of Subtitles 21
Volha Petukhova, Rodrigo Agerri, Mark Fishel, Sergio Penkale, Arantza Del Pozo, Mirjam Sepesy Maucec, Andy Way, Yota Georgakopoulou, Martin Volk

A Corpus of Adequacy Assessments for Real-World Machine Translation Output 29
Daniele Pighin, Lluís Màrquez, Lluís Formiga

The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions 36
Stelios Piperidis

The Language Library: Supporting Community Effort for Collective Resource Production 43
Nicoletta Calzolari, Riccardo Del Gratta, Francesca Frontini, Francesco Rubino, Irene Russo

Practical and Technical Aspects of Using the International Standard Language Resource Number 50
Jungyeul Park, Victoria Arranz, Olivier Hamon, Khalid Choukri

ELRA in the Heart of a Cooperative HLT World 55
Valérie Mapelli, Victoria Arranz, Matthieu Carré, Hélène Mazo, Djamel Mostefa, Khalid Choukri

Twenty Years of Language Resource Development and Distribution: A Progress Report on LDC Activities 60
Christopher Cieri, Marian Reed, Denise Dipersio, Mark Liberman

Polaris: Lymba's Semantic Parsing 66
Dan Moldovan, Eduardo Blanco

Automatic Classification of German "an" Particle Verbs 73
Sylvia Springorum, Sabine Schulte Im Walde, Antje Roßdeutscher

Pragmatic Identification of the Witness Sets 81
Livio Robaldo, Jakub Szymanik

Evaluating Automatic Cross-domain Dutch Semantic Role Annotation 88
Orphée De Clercq, Veronique Hoste, Paola Monachesi

Logic and Graph Based Methods for Terminological Assessment 94
Benoît Robichaud

KALAKA-2: A TV Broadcast Speech Database for the Recognition of Iberian Languages in Clean and Noisy Environments 99
Luis Javier Rodriguez-Fuentes, Mikel Penagarikano, Amparo Varona, Mireia Diez, German Bordel

The C-ORAL-BRASIL I: Reference Corpus for Spoken Brazilian Portuguese 106
Tommaso Raso, Heliana Mello, Maryualê M. Mittmann

The ETAPE Corpus for the Evaluation of Speech-based TV Content Processing in the French Language 114
Guillaume Gravier, Gilles Adda, Niklas Paulsson, Matthieu Carré, Aude Giraudel, Olivier Galibert

Automatic Speech Recognition on a Firefighter TETRA Broadcast Channel 119
Daniel Stein, Bela Usabaev

TED-LIUM: An Automatic Speech Recognition Dedicated Corpus 125
Anthony Rousseau, Paul Deléglise, Yannick Estève

QurAna: Corpus of the Quran Annotated with Pronominal Anaphora 130
Abdul-Baquee Sharaf, Eric Atwell

Using Parallel and Comparable Data for Abstract Anaphora Resolution in German and English 138
Heike Zinsmeister, Melanie Seiss, Stefanie Dipper

Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component 146
Lucie Poláková, Pavlína Jínová, Jirí Mírovský

A Comparable Portuguese-Spanish Corpus with Ellipsis Annotations 154
Luz Rello, Iria Gayo

Coreference in Spoken vs. Written Texts: A Corpus-based Analysis 158
Marilisa Amoia, Kerstin Kunz, Ekaterina Lapshinova-Koltunski

Annotating Near-Identity from Coreference Disagreements 165
Marta Recasens, M. Antònia Martí, Constantin Orasan

This Also Affects the Context - Errors in Extraction Based Summaries 173
Thomas Kaspersson, Christian Smith, Henrik Danielsson, Arne Jönsson

Annotation of Anaphoric Relations and Topic Continuity in Japanese Conversation 179
Natsuko Nakagawa, Yasuharu Den

Domain-specific vs. Uniform Modeling for Coreference Resolution 187
Olga Uryupina, Massimo Poesio

Creating a Coreference Resolution System for Polish 192
Mateusz Kopec, Maciej Ogrodniczuk

Fast Labeling and Transcription with the Speechalyzer Toolkit 196
Felix Burkhardt
Automatic Annotation of Head Velocity and Acceleration in Anvil 201
Bart Jongejan
AVATecH – Automated Annotation Through Audio and Video Analysis 209
Przemyslaw Lenkiewicz, Binyam Gebrekidan Gebre, Oliver Schreer, Stefano Masneri, Daniel Schneider, Sebastian Tschöpel
An Oral History Annotation Tool for INTER-VIEWs 215
Henk Van Den Heuvel, Eric Sanders, Robin Rutten, Stef Scagliola, Paula Witkamp
ELAN Development, Keeping Pace with Communities' Needs 219
Han Sloetjes, Aarthy Somasundaram
Inforex — A Web-based Tool for Text Corpus Management and Semantic Annotation 224
Michal Marcinczuk, Jan Kocon, Bartosz Broda
Towards Automatic Gesture Stroke Detection 231
Binyam Gebrekidan Gebre, Peter Wittenburg, Przemyslaw Lenkiewicz
EXMARaLDA and the FOLK Tools – Two Toolsets for Transcribing and Annotating Spoken Language 236
Thomas Schmidt
Designing a Search Interface for a Spanish Learner Oral Corpus: The End-user's Evaluation 241
Leonardo Campillos Llanos
Dictionary Look-up with Katakana Variant Recognition 249
Satoshi Sato
The Rocky Road towards a Swedish FrameNet - Creating SweFN 256
Karin Friberg Heppin, Maria Toporowska Gronostaj
Capturing Syntactico-semantic Regularities Among Terms: An Application of the FrameNet Methodology to Terminology 262
Marie-Claude L'Homme, Janine Pimentel
Developing LMF-XML Bilingual Dictionaries for Colloquial Arabic Dialects 269
David Graff, Mohamed Maamouri
UBY-LMF — A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF 275
Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer
Legal Electronic Dictionary for Czech 283
František Cvrcek, Karel Pala, Pavel Rychlý
Adaptive Dictionary for Bilingual Lexicon Extraction from Comparable Corpora 288
Amir Mohamed Hazem, Emmanuel Morin
A New Twitter Verbal Lexicon for Natural Language Processing 293
Jennifer Williams, Graham Katz
Challenges in the Development of Annotated Corpora of Computer-mediated Communication in Indian Languages: A Case of Hindi 299
Ritesh Kumar
Ontologies of Linguistic Annotation: Survey and Perspectives 303
Christian Chiarcos
A High-Quality Web Corpus of Czech 311
Johanka Spoustová, Miroslav Spousta
WebAnnotator, an Annotation Tool for Web Pages 316
Xavier Tannier
Development of a Web-Scale Chinese Word N-gram Corpus with Parts of Speech Information 320
Chi-Hsin Yu, Yi-Jie Tang, Hsin-Hsi Chen
CoALT: A Software for Comparing Automatic Labelling Tools 325
Dominique Fohr, Odile Mella
CAT: The CELCT Annotation Tool 333
Valentina Bartalesi Lenzi, Giovanni Moretti, Rachele Sprugnoli
ROMBAC - The Romanian Balanced Annotated Corpus 339
Radu Ion, Elena Irimia, Dan Stefanescu, Dan Tufis
A French Fairy Tale Corpus Syntactically and Semantically Annotated 345
Ismaïl El Maarouf, Jeanne Villaneau
Iula2Standoff: A Tool for Creating Standoff Documents for the IULACT 351
Carlos Morell, Jorge Vivaldi, Núria Bel
ANALEC: A New Tool for the Dynamic Annotation of Textual Data 357
Frederic Landragin, Thierry Poibeau, Bernard Victorri
The SYNC3 Collaborative Annotation Tool 363
Georgios Petasis
Simplified Guidelines for the Creation of Large Scale Dialectal Arabic Annotations 371
Mona Diab, Heba Elfardy
Leveraging the Wisdom of the Crowds for the Acquisition of Multilingual Language Resources 379
Arno Scharl, Marta Sabou, Stefan Gindl, Walter Rafelsberger, Albert Weichselbraun
Experiences in Resource Generation for Machine Translation through Crowdsourcing 384
Anoop Kunchukuttan, Shourya Roy, Pratik Patel, Kushal Ladha, Somya Gupta, Mitesh M. Khapra, Pushpak Bhattacharyya
Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing 392
Elena Filatova

Supervised Topical Key Phrase Extraction of News Stories using Crowdsourcing, Light Filtering and Co-reference Normalization 399
Luís Marujo, Anatole Gershman, Jaime Carbonell, Robert Frederking, João P. Neto

Constructive Interaction for Talking about Interesting Topics 404
Kristiina Jokinen, Graham Wilcock

Using Multimodal Resources for Explanation Approaches in Technical Systems 411
Florian Nothdurft, Wolfgang Minker

Multimodal Corpus of Multi-party Conversations in Foreign Language 416
Shota Yamazaki, Hirohisa Furukawa, Masafumi Nishida, Kristiina Jokinen, Seiichi Yamamoto

The REX Corpora: A Collection of Multimodal Corpora of Referring Expressions in Collaborative Problem Solving Dialogue 422
Takenobu Tokunaga, Ryu IIda, Asuka Terai, Naoko Kuriyama

ISO 24617-2: A Semantically-based Standard for Dialogue Annotation 430
Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, David Traum

Collecting and Using Comparable Corpora for Statistical Machine Translation 438
Inguna Skadina, Ahmet Aker, Nikos Glaros, Fangzhong Su, Dan Tufis, Mateja Verlic, Andrejs Vasiljevs, Bogdan Babych, Paul Clough, Robert Gaizauskas, Nikos Glaros, Monica Lestari Paramita, Marcis Pinnis

Suffix Trees as Language Models 446
Casey Redd Kennington, Martin Kay, Annemarie Friedrich

DGT-TM: A Freely Available Translation Memory in 22 Languages 454
Ralf Steinberger, Andreas Eisele, Szymon Klocek, Spyridon Pilos, Patrick Schlüter

Identifying Word Translations from Comparable Documents without a Seed Lexicon 460
Reinhard Rapp, Serge Sharoff, Bogdan Babych

Large Aligned Treebanks for Syntax-based Machine Translation 467
Gideon Kotzé, Vincent Vandeghinste, Scott Martens, Jörg Tiedemann

Korp – The Corpus Infrastructure of Språkbanken 474
Lars Borin, Markus Forsberg, Johan Roxendal

Annotation Trees: LDC's Customizable, Extensible, Scalable, Annotation Infrastructure 479
Jonathan Wright, Kira Griffitt, Joe Ellis, Stephanie Strassel, Brendan Callahan

Building Large Corpora from the Web Using a New Efficient Tool Chain 486
Roland Schäfer, Felix Bildhauer

Annotated Bibliographical Reference Corpora in Digital Humanities 494
Young-Min Kim, Patrice Bellot, Elodie Faath, Marin Dacos

Building a 70 Billion Word Corpus of English from Clueweb 502
Jan Pomikálek, Miloš Jakubícek, Pavel Rychlý

A Gold Standard for Relation Extraction in the Food Domain 507
Michael Wiegand, Benjamin Roth, Eva Lasarcyk, Stephanie Köser, Dietrich Klakow

Textual Characteristics for Language Engineering 515
Mathias Bank, Robert Remus, Martin Schierle

Automatically Extracting Procedural Knowledge from Instructional Texts using Natural Language Processing 520
Ziqi Zhang, Philip Webster, Victoria Uren, Andrea Varga, Fabio Ciravegna

Evolution of Event Designation in Media: Preliminary Study 528
Xavier Tannier, Véronique Moriceau, Béatrice Arnulphy, Ruixin He

CLTC: A Chinese-English Cross-lingual Topic Corpus 532
Yunqing Xia, Guoyu Tang, Peng Jin, Xia Yang

A Resource-light Approach to Phrase Extraction for English and German Documents from the Patent Domain and User Generated Content 538
Julia Maria Schulz, Daniela Becks, Christa Womser-Hacker, Thomas Mandl

An Evaluation of the Effect of Automatic Preprocessing on Syntactic Parsing for Biomedical Relation Extraction 544
Md. Faisal Mahbub Chowdhury, Alberto Lavelli

Evaluation of Unsupervised Information Extraction 552
Wei Wang, Romaric Besançon, Olivier Ferret, Brigitte Grau

Extraction of Unmarked Quotations in Newspapers: A Study Based on Direct Speech Extraction Systems 559
Stéphanie Weiser, Patrick Watrin

NgramQuery - Smart Information Extraction from Google N-gram using External Resources 563
Martin Aleksandrov, Carlo Strapparava

A Voting Scheme to Detect Semantic Underspecification 569
Héctor Martínez Alonso, Núria Bel, Bolette Sandford Pedersen

A Comparative Evaluation of Word Sense Disambiguation Algorithms for German 576
Verena Henrich, Erhard Hinrichs

DutchSemCor: Targeting the Ideal Sense-tagged Corpus 584
Piek Vossen, Attila Görög, Rubén Izquierdo, Antal Van Den Bosch

Mapping WordNet Synsets to Wikipedia Articles 590
Samuel Fernando, Mark Stevenson

A New Semantically Annotated Corpus with Syntax-based and Cross-lingual Senses 597
Myriam Rakho, Éric Laporte, Matthieu Constant

Detection of Peculiar Word Sense by Distance Metric Learning with Labeled Examples 601
Minoru Sasaki, Hiroyuki Shinnou

Using Semi-experts to Derive Judgments on Word Sense Alignment: A Pilot Study 605
Soojeong Eom, Markus Dickinson, Graham Katz
ATLIS: Identifying Locational Information in Text Automatically 612
John Vogel, Marc Verhagen, James Pustejovsky
Semi-Supervised Technical Term Tagging With Minimal User Feedback 617
Behrang Qasemizadeh, Paul Buitelaar, Tianqi Chen, Georgeta Bordea
Linguistic Knowledge for Specialized Text Production 622
Miriam Buendía-Castro, Beatriz Sánchez-Cárdenas
In the Same Boat and Other Idiomatic Seafaring Expressions 627
Rita Marinelli, Laura Cignoni
Association Norms of German Noun Compounds 632
Sabine Schulte Im Walde, Susanne Borgwaldt, Ronny Jauch
Medical Term Extraction in an Arabic Medical Corpus 640
Doaa Samy, Antonio Moreno-Sandoval, Conchi Bueno-Díaz, Marta Garrote-Salazar, José M. Guirao
Evaluating the Impact of External Lexical Resources into a CRF-based Multiword Segmenter and Part-of-Speech Tagger 646
Matthieu Constant, Isabelle Tellier
Adapting and Evaluating a Generic Term Extraction Tool 651
Anita Gojun, Ulrich Heid, Bernd Weißbach, Carola Loth, Insa Mingers
Evaluation of Classification Algorithms and Features for Collocation Extraction in Croatian 657
Mladen Karan, Jan Šnajder, Bojana Dalbelo Bašic
The Quaero Evaluation Campaign on Term Extraction 663
Thibault Mondary, Adeline Nazarenko, Haïfa Zargayouna, Sabine Barreaux
Statistical Measures for the Acceptability and Semi-productivity of Persian Light Verb Constructions 670
Shiva Taslimipoor, Afsaneh Fazly, Ali Hamzeh
Identifying Multi-Word Expressions in Statistical Machine Translation 674
Dhouha Bouamor, Nasredine Semmar, Pierre Zweigenbaum
Detecting Japanese Compound Functional Expressions using Canonical/Derivational Relation 680
Takafumi Suzuki, Yusuke Abe, Itsuki Toyota, Takehito Utsuro, Suguru Matsuyoshi, Masatoshi Tsuchiya
Building a Database of French Frozen Adverbial Phrases 685
Aude Grezka, Céline Poudat
German Verb Patterns and Their Implementation in an Electronic Dictionary 693
Marc Luder
Risk Analysis and Prevention: LELIE, a Tool Dedicated to Procedure and Requirement Authoring 698
Flore Barcellini, Marie Garnier, Corinne Grosse, Patrick Saint-Dizier
A Framework for Spelling Correction in Persian Language Using Noisy Channel Model 706
Mohammad Hoseyn Sheykholeslam, Behrouz Minaei-Bidgoli, Hossein Juzi
Conventional Orthography for Dialectal Arabic 711
Nizar Habash, Mona Diab, Owen Rambow
Arabic Word Generation and Modelling for Spell Checking 719
Khaled Shaalan, Mohammed Attia, Pavel Pecina, Younes Samih, Josef Van Genabith
Similarity Ranking as Attribute for Machine Learning Approach to Authorship Identification 726
Jan Rygl, Aleš Horák
Spell Checking for Chinese 730
Shaohua Yang, Hai Zhao, Xiaolin Wang, Bao-Liang Lu
Spell Checking in Spanish: The Case of Diacritic Accents 737
Jordi Atserias, Maria Fuentes, Rogelio Nazar, Irene Renau
Incorporating an Error Corpus into a Spellchecker for Maltese 743
Michael Rosner, Albert Gatt, Andrew Attard, Jan Joachimsen
A Rule-based Morphological Analyzer for Murrinh-Patha 751
Melanie Seiss
Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages 759
Dirk Goldhahn, Thomas Eckart, Uwe Quasthoff
„Rendering Endangered Lexicons Interoperable through Standards Harmonization": The RELISH project 766
Helen Aristar-Dry, Sebastian Drude, Jost Gippert, Irina Nevskaya, Menzo Windhouwer
Measuring Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms 771
Ryan Georgi, Fei Xia, William Lewis
Measuring Interlanguage: Native Language Identification with L1-influence Metrics 779
Julian Brooke, Graeme Hirst
Distractorless Authorship Verification 785
John Noecker Jr., Michael Ryan
Correlation between Similarity Measures for Wikipedia 790
Monica Lestari Paramita, Paul Clough, Ahmet Aker, Robert Gaizauskas
JRC Eurovoc Indexer JEX - A Freely Available Multi-label Categorisation Tool 798
Ralf Steinberger, Mohamed Ebrahim, Marco Turchi
Annotations for Power Relations on Email Threads 806
Vinodkumar Prabhakaran, Huzaifa Neralwala, Owen Rambow, Mona Diab
A Corpus for Research on Deliberation and Debate 812
Marilyn Walker, Jean Fox Tree, Pranav Anand, Rob Abbott, Joseph King

Agreement and Disagreement in Threaded Discussion 818
Jacob Andreas, Sara Rosenthal, Kathleen McKeown
Evaluation of Discourse Relation Annotation in the Hindi Discourse Relation Bank 823
Sudheer Kolachina, Rashmi Prasad, Dipti Misra Sharma, Aravind Joshi

Volume 2

Using Verb Subcategorization for Word Sense Disambiguation 829
Will Roberts, Valia Kordoni
Applying Cross-lingual WSD to Wordnet Development 833
Marianna Apidianaki, Benoît Sagot
A Prototype Tool to Discover Missing Wikipedia Inter-language Links by means of Cross-lingual Word Sense Disambiguation 841
Els Lefever, Veronique Hoste, Martine De Cock
Unsupervised Word Sense Disambiguation with Multilingual Representations 847
Erwin Fernandez-Ordonez, Rada Mihalcea, Samer Hassan
First Steps towards the Semi-automatic Development of a Wordformation-based Lexicon of Latin 852
Marco Passarotti, Francesco Mambrini
PoliMorf: A (Not So) New Open Morphological Dictionary for Polish 860
Marcin Wolinski, Marcin Milkowski, Maciej Ogrodniczuk, Adam Przepiórkowski, Lukasz Szalkiewicz
Unsupervised Acquisition of Concatenative Morphology 865
Lionel Nicolas, Jacques Farré, Cécile Darme
Annotating and Learning Morphological Segmentation of Egyptian Colloquial Arabic 873
Emad Mohamed, Behrang Mohit, Kemal Oflazer
First Results in a Study Evaluating Pre-labeling and Correction Propagation for Machine-Assisted Syriac Morphological Analysis 878
Paul Felt, Eric Ringger, Kevin Seppi, Kristian Heal, Robbie Haertel, Deryle Lonsdale
Evaluating Hebbian Self-Organizing Memories for Lexical Representation and Access 886
Marcello Ferro, Claudia Marzi, Claudia Caudai, Vito Pirrelli
A Morphological Analyzer For Wolof Using Finite-State Techniques 894
Cheikh M. Bamba Dione
IDENTIC Corpus: Morphologically Enriched Indonesian-English Parallel Corpus 902
Septina Dian Larasati
The Romanian Neuter Examined Through A Two-Gender N-Gram Classification System 907
Liviu P. Dinu, Vlad Niculae, Octavia-Maria Sulea
UniDic for Early Middle Japanese: An Electrical Dictionary for Morphological Analysis of Classical Japanese 911
Toshinobu Ogiso, Mamoru Komachi, Yasuharu Den, Yuji Matsumoto
Recognition of Polish Derivational Relations Based on Supervised Learning Scheme 916
Maciej Piasecki, Radoslaw Ramocki, Marek Maziarz
Reconstructing the Diachronic Morphology of Romanian from Dictionary Citations 923
Dan Cristea, Radu Simionescu, Gabriela Haja
Generation of Verbal Stems in Derivationally Rich Language 928
Nives Mikelic Preradovic, Krešimir Šojat, Marko Tadic
A Morphological Transducer for Kyrgyz 934
Jonathan Washington, Mirlan Ipasov, Francis Tyers
AnIta: A Powerful Morphological Analyser for Italian 941
Fabio Tamburini, Matias Melandri
A Structural View of Topic and Focus Marking in Italian 948
Gloria Gagliardi, Edoardo Lombardi Vallauri, Fabio Tamburini
Confrontation of Two Models of Language for the Automatic Phonetic Labeling of an Unknown Ethnic Language of the South-asia: The Case of Mo Piu 956
Geneviève Caelen-Haumont, Sethserey Sam
MISTRAL: A Melody Intonation Speaker Tonal Range Semi-automatic Analysis using Variable Levels 963
Binh Hai Pham, Benoît Weber, Geneviève Caelen-Haumont, Do-Dat Tran
Comparing Performance of Different Set-covering Strategies for Linguistic Content Optimization in Speech Corpora 969
Nelly Barbot, Olivier Boeffard, Arnaud Delhay
Towards Fully Automatic Annotation of Audio Books for TTS 975
Olivier Boeffard, Laure Charonnat, Sébastien Le Maguer, Damien Lolive, Gaëlle Vidal
Statistical Evaluation of Pronunciation Encoding 981
Iris Merkus, Florian Schiel
Annotating a Corpus of Human Interaction with Prosodic Profiles – Focusing on Mandarin Repair/Disfluency 986
Helen Kaiyun Chen
Prediction of Non-Linguistic Information of Spontaneous Speech from the Prosodic Annotation: Evaluation of the X-JToBI System 991
Kikuo Maekawa
Prosomarker: A Prosodic Analysis Tool Based on Optimal Pitch Stylization and Automatic Syllabification 997
Antonio Origlia, Iolanda Alfano

Text and Speech Corpora for Text-To-Speech Synthesis of Tales 1003
David Doukhan, Sophie Rosset, Albert Rilliard, Christophe D'Alessandro, Martine Adda-Decker
Open-Source Boundary-Annotated Corpus for Arabic Speech and Language Processing 1011
Claire Brierley, Majdi Sawalha, Eric Atwell
A Phonemic Corpus of Polish Child-Directed Speech 1017
Luc Boruta, Justyna Jastrzebska
Smooth Sailing for STEVIN 1021
Peter Spyns, Elisabeth D'Halleweyn
Semantic Metadata Mapping in Practice: The Virtual Language Observatory 1029
Dieter Van Uytvanck, Herman Stehouwer, Lari Lampen
Aspects of a Legal Framework for Language Resource Management 1035
Aditi Sharma Grover, Annamart Nieman, Gerhard Van Huyssteen, Justus Roux
Introducing the Swedish Kelly-list, a New Free e-resource for Swedish 1040
Elena Volodina, Sofie Johansson Kokkinakis
Texto4Science: A Quebec French Database of Annotated Short Text Messages 1047
Philippe Langlais, Patrick Drouin, Amélie Paulus, Eugénie Rompré Brodeur, Florent Cottin
Recent Developments in CLARIN-NL 1055
Jan Odijk
A Metadata Editor to Support the Description of Linguistic Resources 1061
Emanuel Dima, Erhard Hinrichs, Christina Hoppermann, Thorsten Trippel, Claus Zinn
Fivehundredmillionandone Tokens. Loading the AAC Container with Text Resources for Text Studies. 1067
Hanno Biber, Evelyn Breiteneder
The Common Orthographic Vocabulary of the Portuguese Language: A Set of Common Open Lexical Resources for a Pluricentric Language 1071
José Pedro Ferreira, Maarten Janssen, Gladis Barcellos De Almeida, Margarita Correia, Gilvan Müller De Oliveira
Creation of Shared Language Resource Repository in the Nordic and Baltic Countries 1076
Andrejs Vasiljevs, Markus Forsberg, Tatiana Gornostay, Dorte Hansen, Kristín Jóhannsdóttir, Gunn Lyse, Krister Lindén, Lene Offersgaard, Sussi Olsen, Bolette Pedersen, Eiríkur Rögnvaldsson, Inguna Skadina, Koenraad De Smedt, Roberts Rozis
The LRE Map. Harmonising Community Descriptions of Resources 1084
Nicoletta Calzolari, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo, Claudia Soria
The META-SHARE Metadata Schema for the Description of Language Resources 1090
Maria Gavrilidou, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz, Valerie Mapelli
Towards Automation in Using Multi-modal Language Resources: Compatibility and Interoperability for Multi-modal Features in Kachako 1098
Yoshinobu Kano
The REPERE Corpus: A Multimodal Corpus for Person Recognition 1102
Aude Giraudel, Matthieu Carré, Valérie Mapelli, Juliette Kahn, Olivier Galibert, Ludovic Quintard
Polish Multimodal Corpus – A Collection of Referential Gestures 1108
Magdalena Lis
An Audiovisual Political Speech Analysis Incorporating Eye-tracking and Perception Data 1114
Stefan Scherer, Georg Layher, John Kane, Heiko Neumann, Nick Campbell
Eye Tracking as a Tool for Machine Translation Error Analysis 1121
Sara Stymne, Henrik Danielsson, Sofia Bremin, Hongzhan Hu, Johanna Karlsson, Anna Prytz Lillkull, Martin Wester
Involving Language Professionals in the Evaluation of Machine Translation 1127
Eleftherios Avramidis, Aljoscha Burchardt, Christian Federmann, Maja Popovic, Cindy Tscherwinka, David Vilar
An Analysis (and an Annotated Corpus) of User Responses to Machine Translation Output 1131
Daniele Pighin, Lluís Màrquez, Jonathan May
Challenges in the TAC-KBP Slot Filling Task 1137
Bonan Min, Ralph Grishman
Evaluating Machine Reading Systems through Comprehension Tests 1143
Anselmo Peñas, Eduard Hovy, Pamela Forner, Álvaro Rodrigo, Richard Sutcliffe, Corina Forascu, Caroline Sporleder
Chinese-English CLIR in Biomedicine Using the Extended CMeSH Terms to Expand Queries 1148
Xinkai Wang, Paul Thompson, Sophia Ananiadou
Towards a User-Friendly Platform for Building Language Resources based on Web Services 1156
Marc Poch, Antonio Toral, Olivier Hamon, Valeria Quochi, Núria Bel
Web Service Integration Platform for Polish Linguistic Resources 1164
Maciej Ogrodniczuk, Michal Lenart
Classifying Standard Linguistic Processing Functionalities based on the Fundamental Data Operation Types 1169
Yoshihiko Hayashi, Chiharu Narawa
A Multilingual Natural Stress Emotion Database 1174
Xin Zuo, Tian Li, Pascale Fung
Method for Collection of Acted Speech Using Various Situation Scripts 1179
Takahiro Miyajima, Hideaki Kikuchi, Katsuhiko Shirai, Shigeki Okawa
Annotating Opinions in German Political News 1183
Kristina Adson, Hong Li, Tal Kirshboim, Xiwen Cheng, Feiyu Xu
Hindi Subjective Lexicon (HSL): A Lexical Resource for Hindi Adjective Polarity Classification 1189
Akshat Bakliwal, Piyush Arora, Vasudeva Varma

The I3MEDIA Speech Database: A Trilingual Annotated Corpus for the Analysis and Synthesis of Emotional Speech 1197
Juan María Garrido, Yesika Laplaza, Montse Marquina, Andrea Pearman, José Gregorio Escalada, Miguel Ángel Rodríguez, Ana Armenta

A Hierarchical Approach with Feature Selection for Emotion Recognition from Speech 1203
Panagiotis Giannoulis, Gerasimos Potamianos

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

Fine-grained German Sentiment Analysis on Social Media 1215
Saeedeh Momtazi

"You Seem Aggressive!" Monitoring Anger in a Practical Application 1221
Felix Burkhardt

Mining Sentiment Words from Microblogs for Predicting Writer-Reader Emotion Transition 1226
Yi-Jie Tang, Hsin-Hsi Chen

Bootstrapping Sentiment Labels For Unannotated Documents With Polarity PageRank 1230
Christian Scheible, Hinrich Schütze

Learning Categories and their Instances by Contextual Features 1235
Antje Schlaf, Robert Remus

Rembrandt - A Named-entity Recognition Framework 1240
Nuno Cardoso

An Adaptive Framework for Named Entity Combination 1244
Bogdan Sacaleanu, Günter Neumann

Rule-based Entity Recognition and Coverage of SNOMED CT in Swedish Clinical Text 1250
Maria Skeppstedt, Maria Kvist, Hercules Dalianis

Latvian and Lithuanian Named Entity Recognition with TildeNER 1258
Marcis Pinnis

Tree-Structured Named Entity Recognition on OCR Data: Analysis, Processing and Results 1266
Marco Dinarelli, Sophie Rosset

Aleda, a Free Large-scale Entity Database for French 1273
Benoît Sagot, Rosa Stern

Evaluating the Impact of Phrase Recognition on Concept Tagging 1277
Pablo Mendes, Joachim Daiber, Rohana Rajapakse, Felix Sasaki, Christian Bizer

Adaptive Speech Recognition for Intuitive Model-based Spoken Dialogues 1281
Tobias Heinroth, Maximilian Grotz, Florian Nothdurft, Wolfgang Minker

Relating Dominance of Dialogue Participants with Their Verbal Intelligence Scores 1289
Kseniya Zablotskaya, Umair Rahim, Fernando Fernandez Martinez, Wolfgang Minker

The Coding and Annotation of Multimodal Dialogue Acts 1293
Volha Petukhova, Harry Bunt

Using DiAML and ANVIL for Multimodal Dialogue Annotations 1301
Harry Bunt, Michael Kipp, Volha Petukhova

A Scalable Architecture For Web Deployment of Spoken Dialogue Systems 1309
Matthew Fuchs, Nikos Tsourakis, Manny Rayner

A Corpus for Gesture-Controlled Mobile Spoken Dialogue Systems 1315
Nikos Tsourakis, Manny Rayner

A Corpus of Spontaneous Multi-party Conversation in Bosnian Serbo-Croatian and British English 1323
Emina Kurtic, Bill Wells, Guy J. Brown, Timothy Kempton, Ahmet Aker

Speech & Multimodal Resources: The Herme Database of Spontaneous Multimodal Human-Robot Dialogues 1328
Jing Guang Han, Emer Gilmartin, Celine Delooze, Brian Vaughan, Nick Campbell

Annotation of Response Tokens and Their Triggering Expressions in Japanese Multi-party Conversations 1332
Yasuharu Den, Hanae Koiso, Katsuya Takanashi, Nao Yoshida

Syntactic Annotation of Spontaneous Speech: Application to Call-center Conversation Data 1338
Thierry Bazillon, Melanie Delplano, Frederic Bechet, Alexis Nasr, Benoit Favre

DECODA: A Call-centre Human-human Spoken Conversation Corpus 1343
Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, Eric Arbillot

Resource Evaluation for Usable Speech Interfaces: Utilizing Human–Human Dialogue 1348
Pepi Stavropoulou, Dimitris Spiliotopoulos, Georgios Kouroupetroglou

3rd Party Observer Gaze for a Continuous Measure of Dialogue Flow 1354
Jens Edlund, Simon Alexandersson, Jonas Beskow, Lisa Gustavsson, Mattias Heldner, Anna Hjalmarsson, Petter Kallionen, Ellen Marklund

Pursing Power in Arabic On-line Discussion Forums 1359
Marc Tomlinson, David B. Bracewell, Mary Draper, Zewar Almissour, Ying Shi, Jeremy Bensley

Causal Analysis of Task Completion Errors in Spoken Music Retrieval Interactions 1365
Sunao Hara, Norihide Kitaoka, Kazuya Takeda

An Annotated Corpus of Film Dialogue for Learning and Characterizing Character Style 1373
Marilyn Walker, Grace Lin, Jennifer E. Sawyer

The FLaReNet Strategic Language Resource Agenda 1379
Claudia Soria, Nuria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan Odijk, Stelios Piperidis, Valeria Quochi, Nicoletta Calzolari

Standardizing a Component Metadata Infrastructure 1387
Daan Broeder, Dieter Van Uytvanck, Maria Gavrilidou, Thorsten Trippel

Citing On-line Language Resources 1391
Daan Broeder, Dieter Van Uytvanck, Gunter Senft

An Analytical Model of Language Resource Sustainability 1395
Khalid Choukri, Victoria Arranz

On Using Linked Data for Language Resource Sharing in the Long Tail of the Localisation Market 1403
David Lewis, Alexander O'Connor, Andrzej Zydron, Gerd Sjögren, Rahzeb Choudhury

Evaluation of Online Dialogue Policy Learning Techniques 1410
Alexandros Papangelis, Vangelis Karkaletsis, Fillia Makedon

The Acquisition and Dialog Act Labeling of the EDECAN-SPORTS Corpus 1416
Lluis-F. Hurtado, Fernando Garcia, Emilio Sanchis, Encarna Segarra

Developing and Evaluating an Emergency Scenario Dialogue Corpus 1421
Jolanta Bachan

Building and Exploiting a Corpus of Dialog Interactions between French Speaking Virtual and Human Agents 1428
Lina M. Rojas-Barahona, Alejandra Lorenzo, Claire Gardent

Robustness and Adaptation of Spoken Language Understanding Systems Among Languages and Domains: The PORTMEDIA Project 1436
Fabrice Lefèvre, Djamel Mostefa, Laurent Besacier, Yannick Estève, Matthieu Quignard, Nathalie Camelin, Lina Rojas-Barahona, Bassam Jabaian, Benoit Favre

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

Automatic Lexical Semantic Classification of Nouns 1448
Núria Bel, Lauren Romeo, Muntsa Padró

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

Rapid Creation of Large-scale Corpora and Frequency Dictionaries 1462
Attila Zséder, Gábor Recski, Dániel Varga, András Kornai

Boosting the Coverage of a Semantic Lexicon by Automatically Extracted Event Nominalizations 1466
Kata Gábor, Marianna Apidianaki, Benoît Sagot, Eric Villemonte De La Clergerie

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

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

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

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

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

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

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

A PropBank for Portuguese: The CINTIL-PropBank 1516
António Branco, Catarina Carvalheiro, Mariana Avelãs, Clara Pinto, Sílvia Pereira, Francisco Costa, Sara Silveira, João Silva, Sérgio Castro, João Graça

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

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

Semantic Annotation of French Corpora: Animacy and Verb Semantic Classes 1533
Juliette Thuilier, Laurence Danlos

Yes We Can!? Annotating English Modal Verbs 1538
Josef Ruppenhofer, Ines Rehbein

An Annotation Scheme for Quantifier Scope Disambiguation 1546
Mehdi Manshadi, Eric Meinhardt, James Allen, Mary Swift

Building Japanese Predicate-argument Structure Corpus using Lexical Conceptual Structure 1554
Yuichiroh Matsubayashi, Yusuke Miyao, Akiko Aizawa

Semantic Annotations in Japanese FrameNet: Comparing Frames in Japanese and English 1559
Kyoko Ohara

ConanDoyle-neg: Annotation of Negation Cues and Their Scope in Conan Doyle Stories 1563
Roser Morante, Walter Daelemans

The Netlog Corpus. A Resource for the Study of Flemish Dutch Internet Language 1569
Mike Kestemont, Claudia Peersman, Benny De Decker, Guy De Pauw, Kim Luyckx, Roser Morante, Frederik Vaassen, Janneke Van De Loo, Walter Daelemans

Investigating Verbal Intelligence Using the TF-IDF Approach 1573
Kseniya Zablotskaya, Fernando Fernandez Martinez, Wolfgang Minker

Diachronic Changes of Text Complexity in 20th Century English Language: NLP Approach 1577
Sanja Štajner, Ruslan Mitkov
DeCour: A Corpus of DEceptive Statements in Italian COURts 1585
Tommaso Fornaciari, Massimo Poesio
French and German Corpora for Audience-based Text Type Classification 1591
Amalia Todirascu, Sebastian Pado, Jennifer Krisch, Max Kisselew, Ulrich Heid
Irregularity Detection in Categorized Document Corpora 1598
Borut Sluban, Senja Pollak, Roel Coesemans, Nada Lavrac
Rhetorical Move Detection in English Abstracts: Multi-label Sentence Classifiers and their Annotated Corpora 1604
Carmen Dayrell, Arnaldo Candido Jr., Gabriel Lima, Danilo Machado Jr., Ann Copestake, Valéria Feltrim, Stella Tagnin, Sandra Aluisio
Unsupervised Document Zone Identification Using Probabilistic Graphical Models 1610
Andrea Varga, Daniel Preotiuc-Pietro, Fabio Ciravegna
Improving K-Nearest Neighbor Efficacy for Farsi Text Classification 1618
Behrooz Minaei-Bidgoli, Mohammad Hossein Elahimanesh, Hossein Malekinezhad
Automatic Annotation and Manual Evaluation of the Diachronic German Corpus TüBa-D/DC 1622
Erhard Hinrichs, Thomas Zastrow
Grammatical Error Annotation for Korean Learners of Spoken English 1628
Hongsuck Seo, Kyusong Lee, Gary Geunbae Lee, Soo-Ok Kweon, Hae-Ri Kim
Robust Clause Boundary Identification for Corpus Annotation 1632
Heiki-Jaan Kaalep, Kadri Muischnek
A Corpus-based Study of the German Recipient Passive 1637
Patrick Ziering, Sina Zarrieß, Jonas Kuhn
Wordnet Based Lexicon Grammar for Polish 1645
Zygmunt Vetulani
A Galician Syntactic Corpus with Application to Intonation Modeling 1650
Montserrat Arza, José M. García-Miguel, Francisco Campillo, Miguel Cuevas-Alonso
A Search Tool for FrameNet Constructicon 1655
Hiroaki Sato

Volume 3

Annotating Errors in a Hungarian Learner Corpus 1659
Markus Dickinson, Scott Ledbetter
Text Simplification Tools for Spanish 1665
Stefan Bott, Horacio Saggion, Simon Mille
CLIMB Grammars: Three Projects Using Metagrammar Engineering 1672
Antske Fokkens, Tania Avgustinova, Yi Zhang
An Implementation of a Latvian Resource Grammar in Grammatical Framework 1680
Peteris Paikens, Normunds Gruzitis
An Open Source Persian Computational Grammar 1686
Shafqat Mumtaz Virk, Elnaz Abolahrar
Reclassifying Subcategorization Frames for Experimental Analysis and Stimulus Generation 1694
Paula Buttery, Andrew Caines
Annotating Progressive Aspect Constructions in the Spoken Section of the British National Corpus 1699
Andrew Caines, Paula Buttery
BUCEADOR, a Multi-language Search Engine for Digital Libraries 1705
Jordi Adell, Antonio Bonafonte, Antonio Cardenal, Marta Ruiz, José A. R. Fonollosa, Asunción Moreno, Eva Navas, Eduardo R. Banga
A Tool for Enhanced Search of Multilingual Digital Libraries of E-journals 1710
Ranka Stankovic, Cvetana Krstev, Ivan Obradovic, Aleksandra Trtovac, Miloš Utvic
A Graphical Citation Browser for the ACL Anthology 1718
Benjamin Weitz, Ulrich Schäfer
LDC Language Resource Papers Catalog: Building a Bibliographic Database 1723
Eleftheria Ahtaridis, Christopher Cieri, Denise Dipersio
Matching Cultural Heritage Items to Wikipedia 1729
Eneko Agirre, Ander Barrena, Oier Lopez De Lacalle, Aitor Soroa, Samuel Fernando, Mark Stevenson
Creating a Data Collection for Evaluating Rich Speech Retrieval 1736
Maria Eskevich, Gareth J. F. Jones, Martha Larson, Roeland Ordelman
The Political Speech Corpus of Bulgarian 1744
Petya Osenova, Kiril Simov
SPPAS: A Tool for the Phonetic Segmentation of Speech 1748
Brigitte Bigi
Orthographic Transcription: Which Enrichment is Required for Phonetization? 1756
Brigitte Bigi, Pauline Peri, Roxane Bertrand
Error Profiling for Task-based Evaluation of Machine-translated Text: A Polish-English Case Study 1764
Sandra Weiss, Lars Ahrenberg
Two Phase Evaluation for Selecting Machine Translation Services 1771
Chunqi Shi, Donghui Lin, Masahiko Shimada, Toru Ishida

Italian and Spanish Null Subjects. A Case Study Evaluation in an MT Perspective. ... 1779
Lorenza Russo, Sharid Loáiciga, Asheesh Gulati

On the Practice of Error Analysis for Machine Translation Evaluation ... 1785
Sara Stymne, Lars Ahrenberg

Identifying Equivalents of Specialized Verbs in a Bilingual Comparable Corpus of Judgments: A Frame-based Methodology ... 1791
Janine Pimentel

Logical Metonymies and Qualia Structures: An Annotated Database of Logical Metonymies for German ... 1799
Alessandra Zarcone, Stefan Rued

Modality in Text: A Proposal for Corpus Annotation ... 1805
Iris Hendrickx, Amália Mendes, Silvia Mencarelli

DBpedia: A Multilingual Cross-domain Knowledge Base ... 1813
Pablo Mendes, Max Jakob, Christian Bizer

A Corpus of General and Specific Sentences from News ... 1818
Annie Louis, Ani Nenkova

Brand Pitt: A Corpus to Explore the Art of Naming ... 1822
Gozde Ozbal, Carlo Strapparava, Marco Guerini

TheWeSearch Corpus, Treebank, and Treecache: A Comprehensive Sample of User-Generated Content ... 1829
Jonathon Read, Rebecca Dridan, Stephan Oepen, Lilja Øvrelid

Collecting Humorous Expressions from a Community-based Question-Answering-Service Corpus ... 1836
Masashi Inoue, Toshiki Akagi

Further Developments in Treebank Error Detection Using Derivation Trees ... 1840
Seth Kulick, Ann Bies, Justin Mott

Parallel Aligned Treebanks at LDC: New Challenges Interfacing Existing Infrastructures ... 1848
Xuansong Li, Stephanie Strassel, Stephen Grimes, Safa Ismael, Mohamed Maamouri, Ann Bies, Nianwen Xue

Expanding Arabic Treebank to Speech: Results from Broadcast News ... 1856
Mohamed Maamouri, Ann Bies, Seth Kulick

Propbank-Br: A Brazilian Treebank Annotated with Semantic Role Labels ... 1862
Magali Sanches Duran, Sandra Maria Aluísio

Joint Grammar and Treebank Development for Mandarin Chinese with HPSG ... 1868
Yi Zhang, Rui Wang, Yu Chen

A Tree is a Baum is an Árbol is a Sach'a: Creating a Trilingual Treebank ... 1874
Annette Rios, Anne Göhring

German and English Treebanks and Lexica for Tree-Adjoining Grammars ... 1880
Miriam Kaeshammer, Vera Demberg

Prague Dependency Style Treebank for Tamil ... 1888
Loganathan Ramasamy, Zdenek Žabokrtský

Treebanking by Sentence and Tree Transformation: Building a Treebank to Support Question Answering in Portuguese ... 1895
Patricia Gonçalves, Rita Santos, António Branco

Croatian Dependency Treebank: Recent Development and Initial Experiments ... 1902
Dasa Berovic, Zeljko Agic, Marko Tadic

A GUI to Detect and Correct Errors in Hindi Dependency Treebank ... 1907
Rahul Agarwal, Bharat Ram Ambati, Anil Kumar Singh

From Grammar Extraction to Treebanking: A Bootstrapping Approach ... 1912
Masood Ghayoomi

The IULA Treebank ... 1920
Montserrat Marimon, Beatríz Fisas, Núria Bel, Marta Villegas, Jorge Vivaldi, Sergi Torner, Mercè Lorente

Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3 ... 1927
Atro Voutilainen, Kristiina Muhonen, Tanja Purtonen, Krister Linden

The Parallel-TUT: A Multilingual and Multiformat Treebank ... 1932
Cristina Bosco, Manuela Sanguinetti, Leonardo Lesmo

Irish Treebanking and Parsing: A Preliminary Evaluation ... 1939
Teresa Lynn, Ozlem Cetinoglu, Jennifer Foster, Elaine Uí Dhonnchadha, Mark Dras, Josef Van Genabith

Automatic Extraction and Evaluation of Arabic LFG Resources ... 1947
Mohammed Attia, Khaled Shaalan, Lamia Tounsi, Josef Van Genabith

Rule-Based Detection of Clausal Coordinate Ellipsis ... 1955
Kristiina Muhonen, Tanja Purtonen

The Impact of Automatic Morphological Analysis & Disambiguation on Dependency Parsing of Turkish ... 1960
Gulsen Eryigit

Task-Driven Linguistic Analysis based on an Underspecified Features Representation ... 1966
Stasinos Konstantopoulos, Valia Kordoni, Nicola Cancedda, Vangelis Karkaletsis, Dietrich Klakow, Jean-Michel Renders

Combining Language Resources Into a Grammar-driven Swedish Parser ... 1971
Malin Ahlberg, Ramona Enache

The Icelandic Parsed Historical Corpus (IcePaHC) ... 1977
Eiríkur Rögnvaldsson, Anton Karl Ingason, Einar Freyr Sigurðsson, Joel Wallenberg

A Treebank-based Study on the Influence of Italian Word Order on Parsing Performance ... 1985
Anita Alicante, Cristina Bosco, Anna Corazza, Alberto Lavelli

Effort of Genre Variation and Prediction of System Performance ... 1993
Dong Wang, Fei Xia

Statistical Section Segmentation in Free-Text Clinical Records 2001
Michael Tepper, Daniel Capurro, Fei Xia, Lucy Vanderwende, Meliha Yetisgen-Yildiz

A Corpus of Scientific Biomedical Texts Spanning over 168 Years Annotated for Uncertainty 2009
Alberto Lavelli, Bernardo Magnini, Ramona Bongelli, Carla Canestrari, Ilaria Riccioni, Cinzia Buldorini, Ricardo Pietrobon, Andrzej Zuczkowski

Págico: Evaluating Wikipedia-based Information Retrieval in Portuguese 2015
Cristina Mota, Alberto Simões, Cláudia Freitas, Luís Costa, Diana Santos

Applying Random Indexing to Structured Data to Find Contextually Similar Words 2023
Danica Damljanovic, Udo Kruschwitz, M-Dyaa Albakour, Johann Petrak, Mihai Lupu

The CONCISUS Corpus of Event Summaries 2031
Horacio Saggion, Sandra Szasz

Building and Exploring Semantic Equivalences Resources 2038
Gracinda Carvalho, David Martins De Matos, Vitor Rocio

The TARSQI Toolkit 2043
Marc Verhagen, James Pustejovsky

From Medical Language Processing to BioNLP Domain 2049
Sara Goggi, Manuela Sassi, Gabriella Pardelli, Stefania Biagioni

Evaluation of a Complex Information Extraction Application in Specific Domain 2056
Romaric Besançon, Olivier Ferret, Ludovic Jean-Louis

A Methodology for the Extraction of Information About the Usage of Formulaic Expressions in Scientific Texts 2064
Hannah Kermes

Structural Alignment of Plain Text Books 2069
André Santos, José João Almeida, Nuno Carvalho

Dependency Parsing for Interaction Detection in Pharmacogenomics 2075
Gerold Schneider, Fabio Rinaldi, Simon Clematide

A Data and Analysis Resource for an Experiment in Text Mining a Collection of Micro-blogs On a Political Topic. 2083
William Black, Rob Procter, Steven Gray, Sophia Ananiadou

A Universal Part-of-Speech Tagset 2089
Slav Petrov, Dipanjan Das, Ryan McDonald

Improving Corpus Annotation Productivity: A Method and Experiment with Interactive Tagging 2097
Atro Voutilainen

Lemmatising Serbian as Category Tagging with Bidirectional Sequence Classification 2103
Andrea Gesmundo, Tanja Samardzic

Joint Segmentation and POS Tagging for Arabic Using a CRF-based Classifier 2107
Souhir Gahbiche-Braham, Hélène Bonneau-Maynard, Thomas Lavergne, François Yvon

Boosting Statistical Tagger Accuracy with Simple Rule-based Grammars 2114
Mans Hulden, Jerid Francom

POS Tagging for Grammaticalization and Grammatical Neologism Detection 2118
Maarten Janssen

Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser 2125
Francesco Rubino, Francesca Frontini, Valeria Quochi

Extracting Directional and Comparable Corpora from a Multilingual Corpus for Translation Studies 2132
Bruno Cartoni, Thomas Meyer

Can Statistical Post-Editing with a Small Parallel Corpus Save a Weak MT Engine? 2138
Marianna J. Martindale

BLEU Evaluation of Machine-Translated English-Croatian Legislation 2143
Sanja Seljan, Marija Brkic, Tomislav Vicic

Chinese Characters Mapping Table of Japanese, Traditional Chinese and Simplified Chinese 2149
Chenhui Chu, Toshiaki Nakazawa, Sadao Kurohashi

Free/Open Source Shallow-Transfer Based Machine Translation for Spanish and Aragonese 2153
Juan Pablo Martínez Cortés, Jim O'Regan, Francis Tyers

Automatic MT Error Analysis: Hjerson Helping Addicter 2158
Jan Berka, Ondrej Bojar, Mark Fishel, Maja Popovic, Daniel Zeman

Re-ordering Source Sentences for SMT 2164
Amit Sangodkar, Om Damani

An English-Portuguese Parallel Corpus of Questions: Translation Guidelines and Application in SMT 2172
Ângela Costa, Tiago Luís, Joana Ribeiro, Ana Cristina Mendes, Luísa Coheur

Word Alignment for English-Turkish Language Pair 2177
Mehmet Talha Çakmak, Süleyman Acar, Gülsen Eryigit

PEXACC: A Parallel Data Mining Algorithm from Comparable Corpora 2181
Radu Ion

A Richly Annotated, Multilingual Parallel Corpus for Hybrid Machine Translation 2189
Eleftherios Avramidis, Marta R. Costa-Jussa, Christian Federmann, Josef Van Genabith, Maite Melero, Pavel Pecina

Automatic Word Alignment Tools to Scale Production of Manually Aligned Parallel Text 2194
Stephen Grimes, Katherine Peterson, Xuansong Li

Design and Compilation of a Specialized Parallel Corpus Spanish-German 2199
Carla Parra Escartín

A Distributed Resource Repository for Cloud-Based Machine Translation 2207
Jörg Tiedemann, Dorte Haltrup Hansen, Lene Offersgaard, Sussi Olsen, Matthias Zumpe

Parallel Data, Tools and Interfaces in OPUS 2214
Jörg Tiedemann
The Polish Sejm Corpus 2219
Maciej Ogrodniczuk
From Keystrokes to Annotated Process Data: Enriching the Output of Inputlog with Linguistic Information 2224
Lieve Macken, Veronique Hoste, Marielle Leijten, Luuk Van Waes
A Curated Database for Linguistic Research: The Test Case of Cimbrian Varieties 2230
Maristella Agosti, Birgit Alber, Giorgio Maria Di Nunzio, Marco Dussin, Stefan Rabanus, Alessandra Tomaselli
Introducing the Reference Corpus of Contemporary Portuguese 2237
Michel Généreux, Iris Hendrickx, Amália Mendes
A Basic Language Resource Kit for Persian 2245
Mojgan Seraji, Beáta Megyesi, Joakim Nivre
Collecting and Analysing Chats and Tweets in SoNaR 2253
Eric Sanders
Reference Corpus of Historical Slovene 2257
Tomaž Erjavec
Kitten: A Tool for Normalizing HTML and Extracting Its Textual Content 2261
Mathieu-Henri Falco, Véronique Moriceau, Anne Vilnat
Collection of a Corpus of Dutch SMS 2268
Maaske Treurniet, Orphée De Clercq, Henk Van Den Heuvel, Nelleke Oostdijk
RIDIRE-CPI: An Open Source Crawling and Processing Infrastructure for Supervised Web-Corpora Building 2274
Alessandro Panunzi, Marco Fabbri, Massimo Moneglia, Lorenzo Gregori, Samuele Paladini
The Minho Quotation Resource 2280
Brett Drury, J. J. Almeida
Evaluating Query Languages for a Corpus Processing System 2286
Elena Frick, Carsten Schnober, Piotr Banski
QurSim: A Corpus for Evaluation of Relatedness in Short Texts 2295
Abdul-Baquee Sharaf, Eric Atwell
EVALIEX – A Proposal for an Extended Evaluation Methodology for Information Extraction Systems 2303
Christina Feilmayr, Birgit Pröll, Elisabeth Linsmayr
A Rough Set Formalization of Quantitative Evaluation with Ambiguity 2311
Patrick Paroubek, Xavier Tannier
The Influence of Corpus Quality on Statistical Measurements on Language Resources 2318
Thomas Eckart, Uwe Quasthoff, Dirk Goldhahn
Identifying Nuggets of Information in GALE Distillation Evaluation 2322
Olga Babko-Malaya, Greg Milette, Michael Schneider, Sarah Scogin
NTUSocialRec: An Evaluation Dataset Constructed from Microblogs for Recommendation Applications in Social Networks 2328
Chieh-Jen Wang, Shuk-Man Cheng, Lung-Hao Lee, Hsin-Hsi Chen, Wen-Shen Liu, Pei-Wen Huang, Shih-Peng Lin
SUTAV: A Turkish Audio-Visual Database 2334
Ibrahim Saygin Topkaya, Hakan Erdogan
Multimodal Behaviour and Feedback in Different Types of Interaction 2338
Costanza Navarretta, Patrizia Paggio
A Parallel Corpus of Music and Lyrics Annotated with Emotions 2343
Carlo Strapparava, Rada Mihalcea, Alberto Battocchi
Building a Multimodal Laughter Database for Emotion Recognition 2347
Merlin Teodosia Suarez, Jocelynn Cu, Madelene Sta. Maria
A Speech and Gesture Spatial Corpus in Assisted Living 2351
Dimitra Anastasiou
The Twins Corpus of Museum Visitor Questions 2355
Priti Aggarwal, Ron Artstein, Jillian Gerten, Athanasios Katsamanis, Shrikanth Narayanan, Angela Nazarian, David Traum
Korean Children's Spoken English Corpus and an Analysis of its Pronunciation Variability 2362
Hyejin Hong, Sunhee Kim, Minhwa Chung
Corpus of Children Voices for Mid-level Social Markers and Affect Bursts Analysis 2366
Marie Tahon, Agnes Delaborde, Laurence Devillers
A Large Scale Annotated Child Language Construction Database 2370
Aline Villavicencio, Beracah Yankama, Robert Berwick, Marco A. P. Idiart
Morphosyntactic Analysis of the CHILDES and TalkBank Corpora 2375
Brian Macwhinney
Light Verb Constructions in the SzegedParalellFX English—Hungarian Parallel Corpus 2381
Veronika Vincze
Measuring the Compositionality of NV Expressions in Basque by Means of Distributional Similarity Techniques 2389
Antton Gurrutxaga, Iñaki Alegria
Analyzing and Aligning German Compound Nouns 2395
Marion Weller, Ulrich Heid
Automatic Term Recognition Needs Multiple Evidence 2401
Natalia Loukachevitch
Constraint Based Description of Polish Multiword Expressions 2408
Roman Kurc, Maciej Piasecki, Bartosz Broda

Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking 2414
Nicholas Michael, Bo Liu, Fei Yang, Dimitris Metaxas, Carol Neidle, Peng Yang

Comparing Computer Vision Analysis of Signed Language Video with Motion Capture Recordings 2421
Matti Karppa, Tommi Jantunen, Ville Viitaniemi, Jorma Laaksonen, Birgitta Burger, Danny De Weerdt

DEGELS1: A Comparable Corpus of French Sign Language and Co-speech Gestures 2426
Annelies Braffort, Leïla Boutora

Semi-Automatic Sign Language Corpora Annotation using Lexical Representations of Signs 2430
Matilde Gonzalez, Michael Filhol, Christophe Collet

A Platform-independent User-friendly Dictionary from Italian to LIS 2435
Umar Shoaib, Gabriele Tiotto, Nadeem Ahmad, Paolo Prinetto

Representing the Translation Relation in a Bilingual Wordnet 2439
Jyrki Niemi, Krister Lindén

Building a Multilingual Parallel Corpus for Human Users 2447
Alexandr Rosen, Martin Vavrín

HunOr: A Hungarian–Russian Parallel Corpus 2453
Martina Katalin Szabó, Veronika Vincze, István Nagy

Mining Hindi-English Transliteration Pairs from Online Hindi Lyrics 2459
Kanika Gupta, Monojit Choudhury, Kalika Bali

Dbnary: Wiktionary as a LMF based Multilingual RDF Network 2466
Gilles Sérasset

FreeLing 3.0: Towards Wider Multilinguality 2473
Lluís Padró, Evgeny Stanilovsky

Bulgarian X-language Parallel Corpus 2480
Svetla Koeva, Ivelina Stoyanova, Rositsa Dekova, Borislav Rizov, Angel Genov

Automatically Generated Online Dictionaries 2487
Eniko Héja, Dávid Takács

Volume 4

Feedback in Nordic First-Encounters: A Comparative Study 2494
Costanza Navarretta, Elisabeth Ahlsén, Jens Allwood, Kristiina Jokinen, Patrizia Paggio

MultiUN v2: UN Documents with Multilingual Alignments 2500
Yu Chen, Andreas Eisele

Customization of the Europarl Corpus for Translation Studies 2505
Zahurul Islam, Alexander Mehler

Accessing and Standardizing Wiktionary Lexical Entries for the Translation of Labels in Cultural Heritage Taxonomies 2511
Thierry Declerck, Karlheinz Mörth, Piroska Lendvai

A Mandarin-English Code-switching Corpus 2515
Ying Li, Yue Yu, Pascale Fung

A Fast, Memory Efficient, Scalable and Multilingual Dictionary Retriever 2520
Paulo Fernandes, Lucelene Lopes, Carlos A. Prolo, Afonso Sales, Renata Vieira

Multilingual Central Repository Version 3.0 2525
Aitor González, Egoitz Laparra, German Rigau

A Good Space: Lexical Predictors in Word Space Evaluation 2530
Christian Smith, Henrik Danielsson, Arne Jönsson

Creation and use of Language Resources in a Question-Answering eHealth System 2536
Ulrich Andersen, Anna Braasch, Lina Henriksen, Csaba Huszka, Anders Johannsen, Lars Kayser, Bente Maegaard, Ole Norgaard, Stefan Schulz, Jürgen Wedekind

Effects of Document Clustering in Modeling Wikipedia-style Term Descriptions 2543
Atsushi Fujii, Yuya Fujii, Takenobu Tokunaga

Evaluating Multi-focus Natural Language Queries over Data Services 2547
Silvia Quarteroni, Vincenzo Guerrisi, Pietro La Torre

Summarizing a Multi-source Set of Documents in a Smart Room 2553
Maria Fuentes, Horacio Rodríguez, Jordi Turmo

LAST MINUTE: A Multimodal Corpus of Speech-based User-Companion Interactions 2559
Dietmar Rösner, Jörg Frommer, Rafael Friesen, Matthias Haase, Julia Lange, Mirko Otto

Annotating Football Matches: Influence of the Source Medium on Manual Annotation 2567
Karën Fort, Vincent Claveau

Creating HAVIC: The Heterogeneous Audio Visual Internet Collection 2573
Stephanie Strassel, Amanda Morris, Jonathan Fiscus, Christopher Caruso, Haejoong Lee, Paul Over, James Fiumara, Barbara Shaw, Brian Antonishek, Martial Michel

MULTIPHONIA: A MULTImodal Database of PHONetics Teaching Methods in Classroom InterActions. 2578
Charlotte Alazard, Corine Astésano, Michel Billières

Mapping WordNet to the Kyoto Ontology 2584
Egoitz Laparra, German Rigau, Piek Vossen

Constructing a Class-Based Lexical Dictionary using Interactive Topic Models 2590
Kugatsu Sadamitsu, Kuniko Saito, Kenji Imamura, Yoshihiro Matsuo

Adding Morpho-semantic Relations to the Romanian Wordnet 2596
Verginica Barbu Mititelu

An Ontological Approach to Model and Query Multimodal Concurrent Linguistic Annotations 2602
Julien Seinturier, Elisabeth Murisasco, Emmanuel Bruno, Philippe Blache

The IMAGACT Cross-linguistic Ontology of Action. A New Infrastructure for Natural Language Disambiguation 2606
Massimo Moneglia, Monica Monachini, Omar Calabrese, Alessandro Panunzi, Francesca Frontini, Gloria Gagliardi, Irene Russo

Towards a Methodology for Automatic Identification of Hypernyms in the Definitions of Large-scale Dictionary 2614
Inga Gheorghita, Jean-Marie Pierrel

Collaborative Semantic Editing of Linked Data Lexica 2619
John McCrae, Elena Montiel-Ponsoda, Philipp Cimiano

Ontoterminology: How to Unify Terminology and Ontology Into a Single Paradigm 2626
Christophe Roche

Representation of Linguistic and Domain Knowledge for Second Language Learning in Virtual Worlds 2631
Alexandre Denis, Ingrid Falk, Claire Gardent, Laura Perez-Beltrachini

A Treebank-driven Creation of an OntoValence Verb lexicon for Bulgarian 2636
Petya Osenova, Kiril Simov, Laska Laskova, Stanislava Kancheva

Creation of a Bottom-up Corpus-based Ontology for Italian Linguistics 2641
Elisa Bianchi, Mirko Tavosanis, Emiliano Giovannetti

Visualizing Word Senses in WordNet Atlas 2648
Matteo Abrate, Clara Bacciu

A Contrastive Review of Paraphrase Acquisition Techniques 2653
Houda Bouamor, Aurélien Max, Gabriel Illouz, Anne Vilnat

Chinese Whispers: Cooperative Paraphrase Acquisition 2659
Matteo Negri, Yashar Mehdad, Alessandro Marchetti, Danilo Giampiccolo, Luisa Bentivogli

Diversified Bootstrapping for Acquiring High-Coverage Paraphrase Resource 2666
Hideki Shima, Teruko Mitamura

SemScribe: Natural Language Generation for Medical Reports 2674
Sebastian Varges, Heike Bieler, Manfred Stede, Lukas C. Faulstich, Kristin Irsig, Malik Atalla

Item Development and Scoring for Japanese Oral Proficiency Testing 2682
Hitokazu Matsushita, Deryle Lonsdale

Evaluating Appropriateness Of System Responses In A Spoken CALL Game 2690
Manny Rayner, Pierrette Bouillon, Johanna Gerlach

Spontaneous Speech Corpora for Language Learners of Spanish, Chinese and Japanese 2695
Antonio Moreno-Sandoval, Leonardo Campillos, Yang Dong, Emi Takamori, José M. Guirao, Paula Gozalo, Chieko Kimura, Kengo Matsui, Marta Garrote

The DISCO ASR-based CALL System: Practicing L2 Oral Skills and Beyond 2702
Helmer Strik, Jozef Colpaert, Joost Van Doremalen, Catia Cucchiarini

A Tool for Extracting Conversational Implicatures 2708
Marta Tatu, Dan Moldovan

Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns 2716
Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni, Sandrine Zufferey

Annotating Story Timelines as Temporal Dependency Structures 2721
Bethard Steven, Oleksandr Kolomiyets, Marie-Francine Moens

An Empirical Resource for Discovering Cognitive Principles of Discourse Organization: The ANNODIS Corpus 2727
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, Laure Vieu

HamleDT: To Parse or Not to Parse? 2735
David Marecek, Martin Popel, Loganathan Ramasamy, Jan Štepánek, Daniel Zeman, Zdenek Žabokrtský, Jan Hajic

Evaluating and Improving Syntactic Lexica by Plugging Them Within a Parser 2742
Elsa Tolone, Benoît Sagot, Éric Villemonte De La Clergerie

Efficient Dependency Graph Matching with the IMS Open Corpus Workbench 2750
Thomas Proisl, Peter Uhrig

MaltOptimizer: A System for MaltParser Optimization 2757
Miguel Ballesteros, Joakim Nivre

Investigating Engagement - Intercultural and Technological Aspects of the Collection, Analysis, and Use of the Estonian Multiparty Conversational Video Data 2764
Kristiina Jokinen, Mare Koit

DISLOG: A Logic-based Language for Processing Discourse Structures 2770
Patrick Saint-Dizier

A Repository of Rules and Lexical Resources for Discourse Structure Analysis: The Case of Explanation Structures 2778
Sarah Bourse, Patrick Saint-Dizier

Feature Discovery for Diachronic Register Analysis: A Semi-Automatic Approach 2786
Stefania Degaetano-Ortlieb, Ekaterina Lapshinova-Koltunski, Elke Teich

Improving the Recall of a Discourse Parser by Constraint-based Postprocessing 2791
Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi, Sara Tonelli

Annotating Dropped Pronouns in Chinese Newswire Text 2795
Elizabeth Baran, Yaqin Yang, Nianwen Xue

Alternative Lexicalizations of Discourse Connectives in Czech 2800
Magdalena Rysova

METU Turkish Discourse Bank Browser 2808
Utku Sirin, Ruket Çakici, Deniz Zeyrek
DramaBank: Annotating Agency in Narrative Discourse 2813
David Elson
Multi-Layer Discourse Annotation of a Dutch Text Corpus 2820
Gisela Redeker, Ildikó Berzlánovich, Nynke Van Der Vliet, Gosse Bouma, Markus Egg
Clause-based Discourse Segmentation of Arabic Texts 2826
Iskandar Keskes, Farah Benamara, Lamia Hadrich Belguith
Project CARDS and FLY: A Multidisciplinary Project within Linguistics 2833
Mariana Gomes, Ana Guilherme, Leonor Tavares, Rita Marquilhas
Revealing Contentious Concepts Across Social Groups 2838
Zumrut Akcam, Ching-Sheng Lin, Samira Shaikh, Sharon Small, Ken Stahl, Tomek Strzalkowski, Nick Webb
Flexible Acquisition of Verb Subcategorization Frames in Italian 2842
Tommaso Caselli, Francesca Frontini, Valeria Quochi, Francesco Rubino, Irene Russo
Large Scale Lexical Analysis 2849
Gregor Thurmair, Vera Aleksic, Christoph Schwarz
Extending the Adverbial Coverage of a French Morphological Lexicon 2856
Elsa Tolone, Stavroula Voyatzi, Claude Martineau, Matthieu Constant
Corpus based Semi-Automatic Extraction of Persian Compound Verbs and their Relations 2863
Somayeh Bagherbeygi, Mehrnoush Shamsfard
Extending the MPC Corpus to Chinese and Urdu - A Multiparty Multi-Lingual Chat Corpus for Modeling Social Phenomena in Language 2868
Ting Liu, Samira Shaikh, Tomek Strzalkowski, Aaron Broadwell, Jennifer Stromer-Galley, Sarah Taylor, Umit Boz, Xiaoai Ren, Jingsi Wu
Multimedia Database of the Cultural Heritage of the Balkans 2874
Ivana Tanasijevic, Biljana Sikimic, Gordana Pavlovic-Lažetic
YADAC: Yet another Dialectal Arabic Corpus 2882
Rania Al-Sabbagh, Roxana Girju
The ALLEGRA Corpus: A Trilingual Resource for Romansh, an Under-represented Language of Switzerland 2890
Yves Scherrer, Bruno Cartoni
Beyond SoNaR: Towards the Facilitation of Large Corpus Building Efforts 2897
Martin Reynaert, Ineke Schuurman, Veronique Hoste, Nelleke Oostdijk, Maarten Van Gompel
The New IDS Corpus Analysis Platform: Challenges and Prospects 2905
Piotr Banski, Peter M. Fischer, Elena Frick, Erik Ketzan, Marc Kupietz, Carsten Schnober, Oliver Schonefeld, Andreas Witt
A Tool/Database Interface for Multi-level Analyses 2912
Kurt Eberle, Kerstin Eckart, Ulrich Heid, Boris Haselbach
New Language Resources for the Pashto Language 2917
Djamel Mostefa, Khalid Choukri, Sylvie Brunessaux, Karim Boudahmane
CALBC: Releasing the Final Corpora 2923
Senay Kafkas, Ian Lewin, David Milward, Erik Van Mulligen, Jan Kors, Udo Hahn, Dietrich Rebholz-Schuhmann
Language Richness of the Web 2927
Martin Majliš, Zdenek Žabokrtský
Cloud Logic Programming for Integrating Language Technology Resources 2935
Markus Forsberg, Torbjörn Lager
Dynamic Web Service Deployment in a Cloud Environment 2941
Marc Kemps-Snijders, Matthijs Brouwer, Janpieter Kunst, Tom Visser
Word Sketches for Turkish 2945
Bharat Ram Ambati, Siva Reddy, Adam Kilgarriff
Service Composition Scenarios for Task-Oriented Translation 2951
Chunqi Shi, Donghui Lin, Toru Ishida
Linguistic Analysis Processing Line for Bulgarian 2959
Aleksandar Savkov, Laska Laskova, Stanislava Kancheva, Petya Osenova, Kiril Simov
On the Way to a Legal Sharing of Web Applications in NLP 2965
Victoria Arranz, Olivier Hamon
Collaborative Development and Evaluation of Text-processing Workflows in a UIMA-supported Web-based Workbench 2971
Rafal Rak, Andrew Rowley, Sophia Ananiadou
The SERENOA Project: Multidimensional Context-Aware Adaptation of Service Front-Ends 2977
Javier Caminero, Mari Carmen Rodríguez, Jean Vanderdonckt, Fabio Paternò, Joerg Rett, Dave Raggett, Jean-Loup Comeliau, Ignacio Marín
Concept-based Selectional Preferences and Distributional Representations from Wikipedia Articles 2985
Alex Judea, Vivi Nastase, Michael Strube
Associative and Semantic Features Extracted From Web-Harvested Corpora 2991
Elias Iosif, Maria Giannoudaki, Eric Fosler-Lussier, Alexandros Potamianos
Building a Resource of Patterns using Semantic Types 2999
Octavian Popescu
CLCM - A Linguistic Resource for Effective Simplification of Instructions in the Crisis Management Domain and its Evaluations 3007
Irina Temnikova, Constantin Orasan, Ruslan Mitkov

A Framework for Evaluating Text Correction 3015
Robert Dale, George Narroway
Typing Race Games as a Method to Create Spelling Error Corpora 3019
Paul Rodrigues, C. Anton Rytting
The MASC Word Sense Corpus 3025
Rebecca Passonneau, Collin Baker, Christiane Fellbaum, Nancy Ide
Addressing Polysemy in Bilingual Lexicon Extraction from Comparable Corpora 3031
Darja Fišer, Nikola Ljubešic
Empirical Comparisons of MASC Word Sense Annotations 3036
Gerard De Melo, Collin F. Baker, Nancy Ide, Rebecca J. Passonneau, Christiane Fellbaum
TIMEN: An Open Temporal Expression Normalization Resource 3044
Hector Llorens, Leon Derczynski, Robert Gaizauskas, Estela Saquete
Annotating Spatial Containment Relations Between Events 3052
Kirk Roberts, Travis Goodwin, Sanda Harabagiu
The Role of Model Testing in Standards Development: The Case of ISO-Space 3060
James Pustejovsky, Jessica Moszkowicz
Towards Emotion and Affect Detection in the Multimodal LAST MINUTE Corpus 3064
Jörg Frommer, Bernd Michaelis, Dietmar Rösner, Andreas Wendemuth, Rafael Friesen, Matthias Haase, Manuela Kunze, Rico Andrich, Julia Lange, Axel Panning, Ingo Siegert
Building a Fine-grained Subjectivity Lexicon from a Web Corpus 3070
Isa Maks, Piek Vossen
Learning Sentiment Lexicons in Spanish 3077
Veronica Perez-Rosas, Carmen Banea, Rada Mihalcea
Assigning Connotation Values to Events 3082
Tommaso Caselli, Irene Russo, Francesco Rubino
Cost and Benefit of Using WordNet Senses for Sentiment Analysis 3090
A. R. Balamurali, Aditya Joshi, Pushpak Bhattacharyya
Linguistic Resources for Entity Linking Evaluation: From Monolingual to Cross-lingual 3098
Xuansong Li, Stephanie Strassel, Heng Ji, Kira Griffitt, Joe Ellis
Creating and Curating a Cross-Language Entity Linking Collection 3106
Dawn Lawrie, James Mayfield, Paul McNamee, Douglas Oard
International Multicultural Name Matching Competition: Design, Execution, Results, and Lessons Learned 3111
Keith J. Miller, Elizabeth Schroeder Richerson, Sarah McLeod, James Finley
An Empirical Study of the Occurrence and Co-Occurrence of Named Entities in Natural Language Corpora 3118
K. Saravanan, Monojit Choudhury, Raghavendra Udupa, A. Kumaran
Extended Named Entities Annotation on OCRized Documents: From Corpus Constitution to Evaluation Campaign 3126
Olivier Galibert, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum, Ludovic Quintard
Making Ellipses Explicit in Dependency Conversion for a German Treebank 3132
Wolfgang Seeker, Jonas Kuhn
A Grammar-informed Corpus-based Sentence Database for Linguistic and Computational Studies 3140
Hongzhi Xu, Helen Kaiyun Chen, Chu-Ren Huang, Qin Lu, Tin-Shing Chiu, Dingxu Shi
A Reference Dependency Bank for Analyzing Complex Predicates 3145
Tafseer Ahmed, Miriam Butt, Annette Hautli, Sebastian Sulger
Announcing Prague Czech-English Dependency Treebank 2.0 3153
Ondrej Bojar, Jan Hajic, Eva Hajicová, Jarmila Panevová, Petr Sgall, Silvie Cinková, Eva Fucíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jirí Semecký, Jana Šindlerová, Jan Štepánek, Josef Toman, Zdenka Urešová, Zdenek Žabokrtský
Example-Based Treebank Querying 3161
Liesbeth Augustinus, Vincent Vandeghinste, Frank Van Eynde
A Cross-Lingual Dictionary for English Wikipedia Concepts 3168
Valentin I. Spitkovsky, Angel X. Chang
A Database of Semantic Clusters of Verb Usages 3176
Silvie Cinková, Martin Holub, Lenka Smejkalová, Adam Rambousek
Is it Useful to Support Users with Lexical Resources? A User Study. 3184
Ernesto William De Luca
A Review Corpus Annotated for Negation, Speculation and Their Scope 3190
Natalia Konstantinova, Sheila C. M. De Sousa, Noa P. Cruz, Manuel J. Maña, Maite Taboada, Ruslan Mitkov
Developing a Large Semantically Annotated Corpus 3196
Valerio Basile, Johan Bos, Kilian Evang, Noortje Venhuizen
Le Petit Prince in UNL 3201
Ronaldo Martins
A Generic Formalism to Represent Linguistic Corpora in RDF and OWL/DL 3205
Christian Chiarcos
A Database of Attribution Relations 3213
Silvia Pareti
InfiKorp: Towards a Free Corpus of Polish 3218
Bartosz Broda, Michal Marcinczuk, Marek Maziarz, Adam Radziszewski, Adam Wardynski
Construction of the Turkish National Corpus (TNC) 3223
Yesim Aksan, Mustafa Aksan, Ahmet Koltuksuz, Taner Sezer, Ümit Mersinli, Umut Ufuk Demirhan, Hakan Yilmazer, Özlem Kurtoglu, Gülsüm Atasoy, Seda Öz, Ipek Yildiz

Building a Learner Corpus 3228
Jirka Hana, Alexandr Rosen, Barbora Štindlová, Petr Jäger
Pedagogical Stances and Their Multimodal Signals. 3233
Giovanna Leone, Francesca D'Errico, Isabella Poggi
Annotated Corpora for Word Alignment between Japanese and English and its Evaluation with MAP-based Word Aligner 3241
Tsuyoshi Okita
Ubiquitous Usage of a Broad Coverage French Corpus: Processing the Est Republicain Corpus 3249
Djamé Seddah, Marie Candito, Benoit Crabbé, Enrique Henestroza Anguiano
Federated Search: Towards a Common Search Infrastructure 3255
Herman Stehouwer, Matej Durco, Eric Auer, Daan Broeder
Proper Language Resource Centers 3260
Willem Elbers, Daan Broeder, Dieter Van Uytvanck
The Language Archive – A New Hub for Language Resources 3264
Sebastian Drude, Daan Broeder, Paul Trilsbeek, Peter Wittenburg
LAMP: A Multimodal Web Platform for Collaborative Linguistic Analysis 3268
Kais Dukes, Eric Atwell
An NLP Curator (or: How I Learned to Stop Worrying and Love NLP Pipelines) 3276
James Clarke, Vivek Srikumar, Mark Sammons, Dan Roth
Using Language Resources in Humanities Research 3284
Marta Villegas, Nuria Bel, Carlos Gonzalo, Amparo Moreno, Nuria Simelio
Glottolog/Langdoc: Increasing the Visibility of Grey Literature for Low-density Languages 3289
Sebastian Nordhoff, Harald Hammarström
The Australian National Corpus: National Infrastructure for Language Resources 3295
Steve Cassidy, Michael Haugh, Pam Peters, Mark Fallu
META-SHARE v2: An Open Network of Repositories for Language Resources Including Data and Tools 3300
Christian Federmann, Ioanna Georgantopoulos, Christian Girardi, Olivier Hamon, Dimitris Mavroeidis, Salvatore Minutoli, Marc Schröder
Linguagrid: A Network of Linguistic and Semantic Services for the Italian Language. 3304
Alessio Bosca, Luca Dini, Milen Kouylekov, Marco Trevisan
Versatile Speech Databases for High Quality Synthesis for Basque 3308
Iñaki Sainz, Daniel Erro, Eva Navas, Inma Hernáez, Jon Sánchez, Ibon Saratxaga, Igor Odriozola
Building a Synchronous Corpus of Acoustic and 3D Facial Marker Data for Adaptive Audio-visual Speech Synthesis 3313
Dietmar Schabus, Michael Pucher, Gregor Hofer
Building Text-To-Speech Voices in the Cloud 3317
Alistair Conkie, Thomas Okken, Yeon-Jun Kim, Giuseppe Di Fabbrizio

Volume 5

Building Synthetic Voices in the META-NET Framework 3322
Emília Garcia Casademont, Antonio Bonafonte, Asunción Moreno
Building Text-to-Speech Systems for Resource Poor Languages 3327
Nur-Hana Samsudin, Mark Lee
Evaluating Expressive Speech Synthesis from Audiobook Corpora for Conversational Phrases 3335
Eva Szekely, Joao Cabral, Mohamed Abou-Zleikha, Peter Cahill, Julie Carson-Berndsen
Body-conductive Acoustic Sensors in Human-robot Communication 3340
Panikos Heracleous, Carlos Ishi, Takahiro Miyashita, Norihiro Hagita
Balanced Data Repository of Spontaneous Spoken Czech 3345
Lucie Válková, Martina Waclawicová, Michal Kren
NKI-CCRT Corpus - Speech Intelligibility Before and After Advanced Head and Neck Cancer Treated with Concomitant Chemoradiotherapy 3350
R. P. Clapham, L. Van Der Molen, R. J. J. H. Van Son, M. Van Den Brekel, F. J. M. Hilgers
Sense Meets Nonsense - A Dual-layer Danish Speech Corpus for Perception Studies 3356
Thomas Ulrich Christiansen, Peter Juel Henrichsen
SMALLWorlds — A Multi-lingual Speech Corpus for Cognitive Research 3362
Peter Juel Henrichsen, Marcus Uneson
A Parameterized and Annotated Corpus of the CMU Let's Go Bus Information System 3369
Alexander Schmitt, Stefan Ultes, Wolfgang Minker
Speech and Language Resources for LVCSR of Russian 3374
Sergey Zablotskiy, Alexander Shvets, Maxim Sidorov, Eugene Semenkin, Wolfgang Minker
Dysarthric Speech Database for Development of QoLT Software Technology 3378
Dae-Lim Choi, Bong-Wan Kim, Yeon-Whoa Kim, Yong-Ju Lee, Yongnam Um, Minhwa Chung
The Annotation of the C-ORAL-BRASIL Oral Through the Implementation of the Palavras Parser 3382
Eckhard Bick, Heliana Mello, Alessandro Panunzi, Tommaso Raso
The Nordic Dialect Corpus 3387
Janne Bondi Johannessen, Joel Priestley, Kristin Hagen, Anders Nøklestad, André Lynum
ULex: New Data Models and a Mobile Environment for Corpus Enrichment. 3392
Dafydd Gibbon

Developing Partially-Transcribed Speech Corpus from Edited Transcriptions 3399
Kengo Ohta, Masatoshi Tsuchiya, Seiichi Nakagawa

LDC Forced Aligner 3405
Xiaoyi Ma

The KIT Lecture Corpus for Speech Translation 3409
Sebastian Stüker, Teresa Herrmann, Florian Kraft, Christian Mohr, Alex Waibel, Eunah Cho

Development of Text and Speech Database for Hindi and Indian English Specific to Mobile Communication Environment 3415
Shyam Agrawal, Shweta Sinha, Pooja Singh, Jesper Olson

Source-Language Dictionaries Help Non-Expert Users to Enlarge Target-Language Dictionaries for Machine Translation 3422
Víctor M. Sánchez-Cartagena, Miquel Esplà-Gomis, Juan Antonio Pérez-Ortiz

The ML4HMT Workshop on Optimising the Division of Labour in Hybrid Machine Translation 3430
Christian Federmann, Eleftherios Avramidis, Marta R. Costa-Jussa, Josef Van Genabith, Maite Melero, Pavel Pecina

Alignment-based Reordering for SMT 3436
Maria Holmqvist, Sara Stymne, Lars Ahrenberg, Magnus Merkel

Same Domain Different Discourse Style - A Case Study on Language Resources for Data-driven Machine Translation 3441
Monica Gavrila, Walther V. Hahn, Cristina Vertan

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

Towards a Richer Wordnet Representation of Properties 3452
Sanni Nimb, Bolette Sandford Pedersen

A Proposal for Improving WordNet Domains 3457
Aitor González, German Rigau, Mauro Castillo

Corpus+WordNet Thesaurus Generation for Ontology Enriching 3463
Fernando Castilho, Roger Granada, Breno Meneghetti, Leonardo Carvalho, Renata Vieira

Cleaning Noisy Wordnets 3468
Benoît Sagot, Darja Fišer

Wordnet Extension Made Simple: A Multilingual Lexicon-based Approach Using Wiki Resources 3473
Valérie Hanoka, Benoît Sagot

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

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

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

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

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

Cross-lingual Studies of ASR Errors: Paradigms for Perceptual Evaluations 3511
Ioana Vasilescu, Martine Adda-Decker, Lori Lamel

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

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

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

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

MLSA - A Multi-layered Reference Corpus for German Sentiment Analysis 3551
Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger, Michael Wiegand

A Classification of Adjectives for Polarity Lexicons Enhancement 3557
Silvia Vázquez, Núria Bel

SentiSense: An Easily Scalable Concept-based Affective Lexicon for Sentiment Analysis 3562
Jorge Carrillo De Albornoz, Laura Plaza, Pablo Gervás

"Vreselijk mooi!" (Terribly Beautiful): A Subjectivity Lexicon for Dutch Adjectives. 3568
Tom De Smedt, Walter Daelemans

Visualizing Sentiment Analysis on a User Forum 3573
Rasmus Sundberg, Anders Eriksson, Johan Bini, Pierre Nugues

Collecting, Interpreting and Exploiting Affective Common Sense Knowledge 3580
Erik Cambria, Amir Hussain, Yunqing Xia

A Repository for the Sustainable Management of Research Data 3586
Emanuel Dima, Verena Henrich, Erhard Hinrichs, Marie Hinrichs, Christina Hoppermann, Thorsten Trippel, Thomas Zastrow, Claus Zinn

Towards a Comprehensive Open Repository of Polish Language Resources 3593
Maciej Ogrodniczuk, Piotr Pezik, Adam Przepiórkowski
The Open Lexical Infrastructure of Språkbanken 3598
Lars Borin, Markus Forsberg, Leif-Jöran Olsson, Jonatan Uppström
The Open Linguistics Working Group 3603
Christian Chiarcos, Sebastian Hellmann, Sebastian Nordhoff, Steven Moran, Richard Littauer, Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek, Christian M. Meyer
GATEtoGerManC: A GATE-based Annotation Pipeline for Historical German 3611
Silke Scheible, Richard J. Whitt, Martin Durrell, Paul Bennett
Tackling Interoperability Issues Within UIMA Work-flows 3618
Nicolas Hernandez
Knowledge-Rich Context Candidate Extraction and Ranking with KnowPipe 3626
Anne-Kathrin Schumann
Application of a Semantic Search Algorithm to Semi-Automatic GUI Generation 3631
Maria Teresa Pazienza, Noemi Scarpato, Armando Stellato
The KnowledgeStore: An Entity-Based Storage System 3639
Roldano Cattoni, Francesco Corcoglioniti, Christian Girardi, Bernardo Magnini, Luciano Serafini, Roberto Zanoli
Tools for plWordNet Development. Presentation and Perspectives 3647
Bartosz Broda, Marek Maziarz, Maciej Piasecki
Combining Formal Concept Analysis and Semantic Information for Building Ontological Structures from Texts : An Exploratory Study 3653
Silvia Moraes, Vera Lima
RELcat: A Relation Registry for ISOcat Data Categories 3661
Menzo Windhouwer
A Disambiguation Resource for Semantic Annotation 3665
Eric Charton, Michel Gagnon
NLP Challenges for Eunomos a Tool to Build and Manage Legal Knowledge 3672
Guido Boella, Luigi Di Caro, Llio Humphreys, Livio Robaldo, Leon Van Der Torre
Representing General Relational Knowledge in ConceptNet 5 3679
Robert Speer, Catherine Havasi
A New Dynamic Approach for Lexical Networks Evaluation 3687
Alain Joubert, Mathieu Lafourcade
LIE: Leadership, Influence and Expertise 3692
Roberta Catizone, Louise Guthrie, Arthur Thomas, Yorick Wilks
Semantic Role Labeling with the Swedish FrameNet 3697
Richard Johansson, Karin Friberg Heppin, Dimitrios Kokkinakis
Extending a Wordnet Framework for Simplicity and Scalability 3701
Pedro Fialho, Sérgio Curto, Ana Cristina Mendes, Luísa Coheur
German "nach"-Particle Verbs in Semantic Theory and Corpus Data 3706
Boris Haselbach, Wolfgang Seeker, Kerstin Eckart
LexIt: A Computational Resource on Italian Argument Structure 3712
Alessandro Lenci, Gabriella Lapesa, Giulia Bonansinga
Enriching the ISST-TANL Corpus with Semantic Frames 3719
Alessandro Lenci, Simonetta Montemagni, Giulia Venturi, Maria Rosaria Cutrullà
TimeBankPT: A TimeML Annotated Corpus of Portuguese 3727
Francisco Costa, António Branco
SUTime: A Library for Recognizing and Normalizing Time Expressions 3735
Angel Chang, Christopher Manning
Temporal Annotation: A Proposal for Guidelines and an Experiment with Inter-annotator Agreement 3741
André Bittar, Caroline Hagège, Véronique Moriceau, Xavier Tannier, Charles Tesseidre
Temporal Tagging on Different Domains: Challenges, Strategies, and Gold Standards 3746
Jannik Strötgen, Michael Gertz
Massively Increasing TIMEX3 Resources: A Transduction Approach 3754
Leon Derczynski, Estela Saquete, Hector Llorens
Romanian TimeBank: An Annotated Parallel Corpus for Temporal Information 3762
Corina Forascu, Dan Tufis
Detecting Reduplication in Videos of American Sign Language 3767
Zoya Gavrilov, Stan Sclaroff, Carol Neidle, Sven Dickinson
A Bilingual Bimodal Reading and Writing Tool for Sign Language Users 3774
Nedelina Ivanova, Olle Eriksen
Resource Production of Written Forms of Sign Languages by a User-centered Editor, SWift (SignWriting Improved Fast Transcriber) 3779
Fabrizio Borgia, Claudia S. Bianchini, Patrice Dalle, Maria De Marsico
RWTH-PHOENIX-Weather: A Large Vocabulary Sign Language Recognition and Translation Corpus 3785
Jens Forster, Christoph Schmidt, Thomas Hoyoux, Oscar Koller, Uwe Zelle, Justus Piater, Hermann Ney
Two Database Resources for Processing Social Media English Text 3790
Eleanor Clark, Kenji Araki
Holaaa!! Writin Like U Talk is Kewl But Kinda Hard 4 NLP 3794
Maite Melero, Judith Domingo, Montse Marquina, Martí Quixal

Foundations of a Multilayer Annotation Framework for Twitter Communications During Crisis Events 3801
William J. Corvey, Sarah Vieweg, Sudha Verma, Martha Palmer, James H. Martin

EmpaTweet: Annotating and Detecting Emotions on Twitter 3806
Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, Sanda M. Harabagiu

Semantic Relations Established by Processes Expressed by Nouns and Verbs: Identification in a Corpus by Means of Syntaxico-semantic Annotation 3814
Nava Maroto García, Marie-Claude L'Homme, Amparo Alcina

Using Wikipedia to Validate the Terminology found in a Corpus of Basic Textbooks 3820
Jorge Vivaldi, Luis Adrián Cabrera-Diego, Gerardo Sierra, María Pozzi

PEARL: ProjEction of Annotations Rule Language, a Language for Projecting (UIMA) Annotations over RDF Knowledge Bases 3828
Maria Teresa Pazienza, Armando Stellato, Andrea Turbati

Constructing Large Proposition Databases 3836
Peter Exner, Pierre Nugues

Highlighting Relevant Concepts from Topic Signatures 3841
Montse Cuadros, Lluís Padró, German Rigau

Towards an LHG Parser for Polish: An Exercise in Parasitic Grammar Development 3849
Agnieszka Patejuk, Adam Przepiórkowski

The Effectiveness of Unsupervised Learning on Domain Adaptation: A Case Study on Chinese Word Segmentation 3853
Yan Song, Fei Xia

The Dependency-parsed FrameNet Corpus 3861
Daniel Bauer, Hagen Fürstenau, Owen Rambow

Predicting Phrase Breaks in Classical and Modern Standard Arabic Text 3868
Majdi Sawalha, Claire Brierley, Eric Atwell

Parsing Any Domain English text to CoNLL Dependencies 3873
Sudheer Kolachina, Prasanth Kolachina

Iterative Refinement of Annotation Guidelines for Semantically Fuzzy Named Entity Types, such as Pathological Phenomena 3881
Elena Beisswanger, Ekaterina Buyko, Erik Faessler, Jennifer Traumüller, Susann Schröder, Udo Hahn

GerNED: A Corpus in German for Named Entity Disambiguation 3886
Danuta Ploch, Leonhard Hennig, Angelina Duka, Ernesto William De Luca, Sahin Albayrak

Centroids: Gold Standards with Distributional Variation 3894
Ian Lewin, Senay Kafkas, Dietrich Rebholz-Schuhmann

Quantising Opinions for Political Tweets Analysis 3901
Yulan He, Hassan Saif, Zhongyu Wei, Kam-Fai Wong

AWATIF: A Multi-Genre Corpus for Modern Standard Arabic Subjectivity and Sentiment Analysis 3907
Muhammad Abdul-Mageed, Mona Diab

Arabic-Segmentation Combination Strategies for Statistical Machine Translation 3915
Saab Mansour, Hermann Ney

The Joy of Parallelism with CzEng 1.0 3921
Ondrej Bojar, Zdenek Žabokrtský, Ondrej Dušek, Petra Galuščáková, Martin Majliš, David Marecek, Jirí Maršík, Michal Novák, Martin Popel, Aleš Tamchyna

Statistical Machine Translation without Source-side Parallel Corpus using Word Lattice and Phrase Extension 3929
Takanori Kusumoto, Tomoyoshi Akiba

Automatic Translation of Scientific Documents in the HAL Archive 3933
Lambert Patrik, Holger Schwenk, Frédéric Blain

Expanding Parallel Resources for Medium-Density Languages for Free 3937
Georgi Iliev, Angel Genov

VERTa: Linguistic Features in MT Evaluation 3944
Elisabet Comelles, Jordi Atserias, Victoria Arranz, Irene Castellón

Linguistic Resources for Handwriting Recognition and Translation Evaluation 3951
Zhiyi Song, Safa Ismael, Steven Grimes, David Doermann, Stephanie Strassel

Development and Application of a Cross-language Document Comparability Metric 3956
Fangzhong Su, Bogdan Babych

Assessing Divergence Measures for Automated Document Routing in an Adaptive MT System 3963
Claire Jaja, Douglas Briesch, Jamal Laoudi, Clare Voss

A Study of Word-Classing for MT Reordering 3971
Ananthakrishnan Ramanathan, Karthik Visweswariah

Dealing with Unknown Words in Statistical Machine Translation 3977
João Silva, Luísa Coheur, Ângela Costa, Isabel Trancoso

PET: A Tool for Post-editing and Assessing Machine Translation 3982
Wilker Aziz, Sheila C. M. De Sousa, Lucia Specia

Tajik-Farsi Persian Transliteration Using Statistical Machine Translation 3988
Chris Irwin Davis

Assessing the Comparability of News Texts 3996
Emma Barker, Rob Gaizauskas

Corpus-based Referring Expressions Generation 4004
Hilder Pereira, Eder Novais, Andre Mariotti, Ivandre Paraboni

Portuguese Text Generation from Large Corpora 4010
Eder Novais, Ivandre Paraboni, Douglas Silva
Danish Parallel Corpus for Text Simplification 4015
Sigrid Klerke, Anders Søgaard
Acquisition of Syntactic Text Simplification Rules for French 4019
Violeta Seretan
A Repository of Data and Evaluation Resources for Natural Language Generation 4027
Anja Belz, Albert Gatt
LG-Eval: A Toolkit for Creating Online Language Evaluation Experiments 4033
Eric Kow, Anja Belz
Turk Bootstrap Word Sense Inventory 2.0: A Large-Scale Resource for Lexical Substitution 4038
Chris Biemann
Collection of a Large Database of French-English SMT Output Corrections 4043
Marion Potet, Emmanuelle Esperança-Rodier, Laurent Besacier, Hervé Blanchon
Getting More Data — Schoolkids As Annotators 4049
Jirka Hana, Barbora Hladka
Word Sense Inventories by Non-Experts. 4055
Anna Rumshisky, Nick Botchan, Sophie Kushkuley, James Pustejovsky
The BladeMistress Corpus: From Talk to Action in Virtual Worlds 4060
Anton Leuski, Carsten Eickhoff, James Ganis, Victor Lavrenko
Annotating Factive Verbs 4068
Alvin Grisson II, Yusuke Miyao
A Holistic Approach to Bilingual Sentence Fragment Extraction from Comparable Corpora 4073
Mahdi Khademian, Kaveh Taghipour, Shahram Khadivi, Saab Mansour
An Examination of Cross-Cultural Similarities and Differences from Social Media Data with Respect to Language Use 4080
Mohammad Fazleh Elahi, Paola Monachesi
Turkish Paraphrase Corpus 4087
Seniz Demir, Ilknur Durgar El-Kahlout, Erdem Unal, Hamza Kaya
Constructing a Question Corpus for Textual Semantic Relations 4092
Rui Wang, Shuguang Li
Evaluating the Similarity Estimator Component of the TWIN Personality-based Recommender System 4098
Alexandra Roshchina, John Cardiff, Paolo Rosso
Annotation Facilities for the Reliable Analysis of Human Motion 4103
Michael Kipp
Translog-II: A Program for Recording User Activity Data for Empirical Translation Process Research 4108
Michael Carl
Intelligibility Assessment in Forensic Applications 4113
Giovanni Costantini, Andrea Paoloni, Massimiliano Todisco
Strategies to Improve a Speaker Diarisation Tool 4117
David Tavarez, Eva Navas, Daniel Erro, Ibon Saratxaga
Using an ASR Database to Design a Pronunciation Evaluation System in Basque 4122
Igor Odriozola, Eva Navas, Inma Hernáez, Iñaki Sainz, Ibon Saratxaga, Jon Sánchez, Daniel Erro
W-PhAMT: A Web Tool for Phonetic Multilevel Timeline Visualization 4127
Francesco Cutugno, Vincenza Anna Leano, Antonio Origlia
English to Indonesian Transliteration to Support English Pronunciation Learning 4132
Amalia Zahra, Julie Carson-Berndsen
Rapidly Testing the Interaction Model of a Pronunciation Training System via Wizard-of-Oz 4136
Joao Paulo Cabral, Mark Kane, Zeeshan Ahmed, Mohamed Abou-Zleikha, Eva Szekely, Amalia Zahra, Kalu Ogbureke, Peter Cahill, Julie Carson-Berndsen, Stephan Schlogl
PAMOCAT: Automatic Retrieval of Specified Postures 4143
Bernhard Brüning, Christian Schnier, Karola Pitsch, Sven Wasmuth
Author Index