INTERNATIONAL CONFERENCE

RECENT ADVANCES IN

NATURAL LANGUAGE PROCESSING

P R O C E E D I N G S

Edited by
Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, Nicolas Nicolov, Nikolai Nikolov

Borovets, Bulgaria

14-16 September 2009

INTERNATIONAL CONFERENCE
RECENT ADVANCES IN
NATURAL LANGUAGE PROCESSING'2009

PROCEEDINGS

Borovets, Bulgaria
14-16 September 2009

ISSN 1313-8502

Designed and Printed by INCOMA Ltd.
Shoumen, BULGARIA

ORGANISERS AND SPONSORS

The International Conference RANLP–2009 is organised by:

Linguistic Modelling Department, Institute for Parallel Processing (IPP),
Bulgarian Academy of Sciences (BAS)

Association for Computational Linguistics, Bulgaria

and

Research Group in Computational Linguistics, University of Wolverhampton, U.K.

RANLP–2009 is partially supported by:

Ontotext Semantic Technology Lab, Bulgaria

Research Institute of Information and Language Processing,
University of Wolverhampton, UK

IPP-BAS (BIS-21 Centre of Excellence)

Association for Computational Linguistics, Bulgaria

The team behind RANLP–2009:

Galia Angelova	Bulgarian Academy of Sciences, Bulgaria, OC Chair
Kalina Bontcheva	University of Sheffield, U.K.
Ruslan Mitkov	University of Wolverhampton, U.K., PC Chair
Nicolas Nicolov	Umbria, Inc., Boulder, U.S.A.
Nikolai Nikolov	INCOMA Ltd., Shoumen, Bulgaria
Kiril Simov	Bulgarian Academy of Sciences, Bulgaria, Workshop Coordinator

PROGRAMME COMMITTEE CHAIR

Ruslan Mitkov (*University of Wolverhampton, UK*)

PROGRAMME COMMITTEE

Roberto Basili (*Univ. of Roma "Tor Vergata", Italy*)
Marco De Boni (*Unilever, UK*)
Kalina Bontcheva (*University of Sheffield, UK*)
António Branco (*University of Lisbon, Portugal*)
Michael Carl (*Copenhagen Business School, Denmark*)
Eugene Charniak (*Brown University, USA*)
Kevin Cohen (*University of Colorado School of Medicine, USA*)
Gloria Corpas (*University of Malaga, Spain*)
Dan Cristea (*"Al. I. Cuza" University of Iasi, Romania*)
Gael Dias (*Univ. of Beira Interior, Portugal*)
Arantza Diaz de Ilarraza (*Un. of Basque Country, Spain*)
Atefeh Farzindar (*NLP Technologies, Canada*)
Robert Gaizauskas (*University of Sheffield, UK*)
Alexander Gelbukh (*Nat. Polytechnic Inst., Mexico*)
Ralph Grishman (*New York University, USA*)
Le An Ha (*University of Wolverhampton, UK*)
Johann Haller (*IAI Saarbrücken, Germany*)
Patrick Hanks (*Charles University, Czech Republic*)
Erhard Hinrichs (*University of Tübingen, Germany)*
Veronique Hoste (*University College Ghent, Belgium*)
Diana Inkpen (*University of Ottawa, Canada*)
Richard J. Evans (*University of Wolverhampton, UK*)
Dimitar Kazakov (*University of York, UK*)
Alma Kharrat (*Microsoft, USA*)
Udo Kruschwitz (*University of Essex, UK*)
Hristo Krushkov (*Plovdiv University, Bulgaria*)
Sandra Kuebler (*Indiana University, USA*)
Shalom Lappin (*King's College London, UK*)
Montse Maritxalar (*Univ. of the Basque Country, Spain*)
M. Antonia Marti (*University of Barcelona, Spain*)
Patricio Martinez-Barco (*Univ. of Alicante, Spain*)
Yuji Matsumoto (*NAIST, Japan*)
Diana Maynard (*University of Sheffield, UK*)
Wolfgang Menzel (*University of Hamburg, Germany*)
Rada Mihalcea (*University of North Texas, USA*)
Andrei Mikheev (*Infogistics Ltd & Daxtra Tech. Ltd, UK*)
Paola Monachesi (*Utrecht University, The Netherlands*)
Andres Montonyo (*University of Alicante, Spain*)
Rafael Munoz Guillena (*University of Alicante, Spain*)
Preslav Nakov (*National Univ. of Singapore, Singapore*)
Roberto Navigli (*Univ. di Roma La Sapienza, Italy*)
John Nerbonne (*Univ. of Groningen, The Netherlands*)
Michael Oakes (*University of Sunderland, UK*)
Kemal Oflazer *(Carnegie Mellon University, Qatar)*
Constantin Orasan (*University of Wolverhampton, UK*)
Petya Osenova (*Bulgarian Academy of Sciences, Bulgaria*)
Manuel Palomar (*University of Alicante, Spain*)
Viktor Pekar (*Oxford University Press, UK*)
Stelios Piperidis (*ILSP, Greece*)
Massimo Poesio (*Univ. of Trento, Italy & Univ. of Essex, UK*)
John Prager (*IBM, USA*)
Gabor Proszeky (*MorphoLogic, Hungary*)
Paul Rayson (*Lancaster University, UK*)
Horacio Rodriguez (*Technical Univ. of Catalonia, Spain*)
Satoshi Sekine (*New York University, USA*)
Khalil Sima'an (*University of Amsterdam, The Netherlands*)
Kiril Simov (*Bulgarian Academy of Sciences, Bulgaria*)
Thamar Solorio (*Univ. of Alabama at Birmingham, USA*)
Lucia Specia (*University of Wolverhampton, UK*)
Ralf Steinberger (*European Commission - Joint Research Centre, Italy*)
Joel Tetreault (*University of Pittsburgh, UK*)
L. Alfonso Urena Lopez (*University of Jaen, Spain*)
Karin Verspoor (*University of Colorado School of Medicine, USA*)
Yorick Wilks (*University of Sheffield, UK*)
Michael Zock (*LIF, CNRS, France*)

REVIEWERS

In addition to the members of the Programme Committee, the following colleagues were involved in the reviewing process:

Naveed Afzal (*University of Wolverhampton, UK*)
Itziar Aldabe (*Univ. of the Basque Country, Spain*)
Afra Alishahi (*University of Saarland, Gemany*)
Alexandra Balahur-Dobrescu (*Univ. Alicante, Spain and European Commission – Joint Research Centre, Italy*)
Verginica Barbu Mititelu (*Rom. Academy, Romania*)
Leonor Becerra-Bonache (*Univ. Rovira i Virgili, Spain*)
Lamia Hadrich Belguith (*University of Sfax, Tunisia*)
Svetla Boytcheva (*State Univ. LS and IT, Bulgaria*)
Guadalupe Aguado de Cea (*Polytechn. Univ. Madrid*)
Rosita Chan (*Ins. Comercial and Univ. of Panama*)
Atanas Chanev (*Bulgaria*)
Andras Csomai (*University of North Texas, USA*)
Iustin Dornescu (*University of Wolverhampton, UK*)
Robert M. Foster (*University of Wolverhampton, UK*)
Kuzman Ganchev (*University of Pennsylvania, USA*)
Lisette Garcia Moya (*University Jaume I, Spain*)
Kallirroi Georgila (*Univ. of Southern California, USA*)
Jose M. Gomez (*University of Alicante, Spain*)
Laura Hasler (*University of Wolverhampton, UK*)
Jesus M. Hermida (*Univ. of Alicante, Spain*)
Adrian Iftene (*"Al. I. Cuza" University of Iasi, Romania*)
Iustina Ilisei (*University of Wolverhampton, UK*)
Radu Ion (*Romanian Academy, Romania*)
Heng Ji (*City University of New York, USA*)
Jason Kessler (*Indiana University, USA*)
Kiril Kolev (*University of Wolverhampton, UK*)
Natalia Konstantinova (*Univ. of Wolverhampton, UK*)
Zornitsa Kozareva (*USC ISI, California, USA*)
Elina Lagoudaki (*Imperial College London, UK)*
Els Lefever (*University College Ghent, Belgium*)
Fang Li (*University of Wolverhampton, UK*)
Elena Lloret (*University of Alicante, Spain*)
Annie Louis (*University of Pennsylvania, USA*)
Oier Lopez de Lacalle (*Univ. of the Basque Country, Spain*)
Lieve Macken (*University College Ghent, Belgium*)
Ruslana Margova (*Journal 'Geomedia', Bulgaria*)
Dalila Mekhaldi (*University of Wolverhampton, UK*)
Neil Millar (*Lancaster University, UK*)
Arturo Montejo Raez (*University of Jaen, Spain*)
Rumen Moraliyski (*Univ. of Beira Interior, Portugal*)
Andrea Mulloni (*University of Wolverhampton, UK*)
Shiyan Ou (*Nanjing University, China*)
Slav Petrov (*Google Research New York, USA*)
Ionut Pistol (*"Al. I. Cuza" University of Iasi, Romania*)
Emily Pitler (*University of Pennsylvania, USA*)
Natalia Ponomareva (*Univ. of Wolverhampton, UK*)
Jelena Prokic (*Univ. of Groningen, The Netherlands*)
Prokopis Prokopidis (*ILSP / R.C. "Athena", Greece*)
Georgiana Puscasu (*Univ. of Wolverhampton, UK*)
Marta Recasens (*University of Barcelona, Spain*)
Luz Rello (*University of Wolverhampton, UK*)
Julie Renahy (*Université de Franche-Comté, France*)
Estela Saquete Boro (*University of Alicante, Spain*)
Armando Suarez Cueto (*University of Alicante, Spain*)
Horacio Saggion (*University of Sheffield, UK*)
Doaa Samy (*Cairo University, Egypt*)
Smriti Singh (*Indian Inst. of Technology, Bombay, India*)
Yvonne Skalban (*University of Wolverhampton, UK*)
Veselin Stoyanov (*Cornell University, USA*)
Ang Sun (*New York University, USA*)
Irina Temnikova (*University of Wolverhampton, UK*)
Rafael M. Terol (*Univ. of Alicante, Spain*)
Diana Trandabat (*"Al. I. Cuza" Univ. of Iasi, Romania*)
Andrea Varga (*University of Wolverhampton, UK*)
Sonia Vazquez Perez (*University of Alicante, Spain*)
Cristina Vertan (*University of Hamburg, Germany*)
Roman Yangarber (*University of Helsinki, Finland*)
Nick Webb (*University at Albany, SUNY, USA*)
Jakub Zavrel (*Textkernel, The Netherlands*)
Kalliopi Zervanou (*ILK, Tilburg Univ., The Netherlands*)

PROGRAMME COMMITTEE COORDINATOR

Ivelina Nikolova (*Bulgarian Academy of Sciences, Bulgaria*)

PROGRAMME COMMITTEE SUPPORT

Natalia Konstantinova (*University of Wolverhampton, UK*)
Irina Temnikova (*University of Wolverhampton, UK*)

Table of Contents

Unsupervised Relation Extraction for Automatic Generation of Multiple-Choice Questions
Naveed Afzal and Viktor Pekar 1

Summary Generation for Toponym-referenced Images using Object Type Language Models
Ahmet Aker and Robert Gaizauskas 6

Prepositional Phrase Attachment in Shallow Parsing
Vincent Van Asch and Walter Daelemans 12

A Comparative Study of Open Domain and Opinion Question Answering Systems for Factual and Opinionated Queries
Alexandra Balahur, Ester Boldrini, Andrés Montoyo and Patricio Martínez-Barco 18

Acquisition of Common Sense Knowledge for Basic Level Concepts
Eduard Barbu 23

Unsupervised Knowledge Extraction for Taxonomies of Concepts from Wikipedia
Eduard Barbu and Massimo Poesio 28

Exploring Treebank Transformations in Dependency Parsing
Kepa Bengoetxea and Koldo Gojenola 33

Contextual Salience in Query-based Summarization
Wauter Bosma 39

Integrating Document Structure into a Multi-Document Summarizer
Aurélien Bossard and Thierry Poibeau 45

Cross-Linguistic Sentiment Analysis: From English to Spanish
Julian Brooke, Milan Tofiloski and Maite Taboada 50

The Influence of Text Pre-processing on Plagiarism Detection
Zdenek Ceska and Chris Fox 55

Combining Finite State and Corpus-based Techniques for Unknown Word Prediction
Kostadin Cholakov and Gertjan van Noord 60

Prototype-based Active Learning for Lemmatization
Walter Daelemans, Hendrik J. Groenewald and Gerhard B. van Huyssteen 65

From Partial toward Full Parsing
Heshaam Faili 71

Grouping Synonyms by Definitions
Ingrid Falk, Claire Gardent, Evelyne Jacquey and Fabienne Venant 76

Singular Value Decomposition for Feature Selection in Taxonomy Learning
Francesca Fallucchi and Fabio Massimo Zanzottto 82

Improving Text Segmentation by Combining Endogenous and Exogenous Methods
Olivier Ferret 88

Edlin: an Easy to Read Linear Learning Framework
Kuzman Ganchev and Georgi Georgiev 94

Exploiting the Use of Prior Probabilities for Passage Retrieval in Question Answering
Surya Ganesh and Vasudeva Varma 99

Exploiting Structure and Content of Wikipedia for Query Expansion in the Context
Surya Ganesh and Vasudeva Varma 103

Text Content and Task Performance in the Evaluation of a Natural Language Generation System
Albert Gatt and François Portet 107

Feature-Rich Named Entity Recognition for Bulgarian Using Conditional Random Fields
Georgi Georgiev, Preslav Nakov, Kuzman Ganchev, Petya Osenova and Kiril Simov 113

Uncertainty Detection for Information Extraction
Bénédicte Goujon 118

Learning to Identify Educational Materials
Samer Hassan and Rada Mihalcea 123

Lexicalized Semi-incremental Dependency Parsing
Hany Hassan, Khalil Sima'an and Andy Way 128

Identification of Parallel Text Pairs Using Fingerprints
Martin Hassel and Hercules Dalianis 135

Stochastic Definite Clause Grammars
Christian Theil Have 139

Topic-based Multi-Document Summarization with Probabilistic Latent Semantic Analysis
Leonhard Hennig 144

Detection of Opinions and Facts. A Cognitive Approach
Yann Vigile Hoareau, Adil El-Ghali and Charles Tijus 150

Evaluating the Impact of Morphosyntactic Ambiguity in Grammatical Error Detection
Arantza Díaz de Ilarraza, Koldo Gojenola and Maite Oronoz 155

Fast Boosting-based Part-of-Speech Tagging and Text Chunking with Efficient Rule Representation for Sequential Labeling
Tomoya Iwakura 161

Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges
Heng Ji, Ralph Grishman, Zheng Chen and Prashant Gupta 166

Co-Parsing with Competitive Models
Lidia Khmylko, Kilian A. Foth and Wolfgang Menzel 173

Robust Compositional Polarity Classification
Manfred Klenner, Stefanos Petrakis and Angela Fahrni 180

Feature Subset Selection in Conditional Random Fields for Named Entity Recognition
Roman Klinger and Christoph M. Friedrich 185

User's Choice of Precision and Recall in Named Entity Recognition
Roman Klinger and Christoph M. Friedrich 192

Semi-Supervised Learning for Word Sense Disambiguation: Quality vs. Quantity
Sandra Kübler and Desislava Zhekova 197

Treelex Meets Adjectival Tables
Anna Kupść 203

Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm
Egoitz Laparra and German Rigau 208

Sampling-based Multilingual Alignment
Adrien Lardilleux and Yves Lepage 214

Using Semantic Networks to Identify Temporal Expressions from Semantic Roles
Hector Llorens, Borja Navarro and Estela Saquete 219

The Design of an Experiment in Anaphora Resolution for Referring Expressions Generation
Diego Jesus de Lucena and Ivandré Paraboni 225

A Model for the Cross-Modal Influence of Visual Context upon Language Procesing
Patrick McCrae 230

Bimodal Corpora Terminology Extraction: Another Brick in the Wall
Claudiu Mihăilă and Dalila Mekhaldi 236

Exploiting Latent Semantic Relations in Highly Linked Hypertext for Information Retrieval in Wikis
Tristan Miller, Bertin Klein and Elisabeth Wolf 241

Large Vocabulary Continuous Speech Recognition for Bulgarian
Petar Mitankin, Stoyan Mihov and Tinko Tinchev 246

Diacritization for Real-World Arabic Texts
Emad Mohamed and Sandra Kübler 251

Multi-entity Sentiment Scoring
Karo Moilanen and Stephen Pulman 258

A Morphological and Syntactic Wide-coverage Lexicon for Spanish: The Leffe
Miguel A. Molinero, Benoît Sagot and Lionel Nicolas 264

How Limited is the Limit?
Prakash Mondal 270

Dependency Parsing and Semantic Role Labeling as a Single Task
Roser Morante, Vincent Van Asch and Antal van den Bosch 275

Structured Output Learning with Polynomial Kernel
Hajime Morita, Hiroya Takamura and Manabu Okumura 281

Unsupervised Word Sense Induction from Multiple Semantic Spaces with Locality Sensitive Hashing
Claire Mouton, Guillaume Pitel, Gaël de Chalendar and Anne Vilnat 287

Unsupervised Extraction of False Friends from Parallel Bi-Texts Using the Web as a Corpus
Svetlin Nakov, Preslav Nakov and Elena Paskaleva 292

Evaluating Term Extraction
Adeline Nazarenko and Haïfa Zargayouna 299

Question Answering over Structured Data: an Entailment-Based Approach to Question Analysis
Matteo Negri and Milen Kouylekov 305

A Semi-supervised Approach for Generating a Table-of-Contents
Viet Cuong Nguyen, Le Minh Nguyen and Akira Shimazu 312

Towards Efficient Production of Linguistic Resources: the Victoria Project
Lionel Nicolas, Miguel A. Molinero, Benoît Sagot, Elena Trigo, Éric De la Clergerie, Miguel Pardo, Jacques Farré and Joan Miquel Vergés 318

A Classification-driven Approach to Document Planning
Rafael Oliveira, Eder Novais, Roberto Araujo and Ivandré Paraboni 324

Interactive Machine Translation Based on Partial Statistical Phrase-based Alignments
Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta 330

Topic Modeling of Research Fields: An Interdisciplinary Perspective
Michael Paul and Roxana Girju 337

An Interaction Grammar of Interrogative and Relative Clauses in French
Guy Perrier 343

Comparing Statistical Similarity Measures for Stylistic Multivariate Analysis
Marius Popescu and Liviu P. Dinu 349

From Bag of Languages to Family Trees From Noisy Corpus
Taraka Rama and Anil Kumar Singh 355

Language-Independent Sentiment Analysis Using Subjectivity and Positional Information
Veselin Raychev and Preslav Nakov 360

All Words Unsupervised Semantic Category Labeling for Hindi
Siva Reddy, Abhilash Inumella, Rajeev Sangal and Soma Paul 365

Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach
Vassiliki Rentoumi, George Giannakopoulos, Vangelis Karkaletsis and George A. Vouros 370

Global Evaluation of Random Indexing through Swedish Word Clustering Compared to the Peoples Dictionary of Synonyms
Magnus Rosell, Martin Hassel and Viggo Kann 376

Identifying Semantic Relations in Context: Near-misses and Overlaps
Alla Rozovskaya and Roxana Girju 381

Statistical Confidence Measures for Probabilistic Parsing
Ricardo Sánchez-Sáez, Joan-Andreu Sánchez and José Miguel Benedí Ruíz 388

Exploring the Vector Space Model for Finding Verb Synonyms in Portuguese
Luís Sarmento, Paula Carvalho and Eugénio Oliveira 393

A Unified Method for Extracting Simple and Multiword Verbs with Valence Information and Application for Hungarian
Bálint Sass 399

Combining Lexical Resources for Contextual Synonym Expansion
Ravi Sinha and Rada Mihalcea 404

String Distance-Based Stemming of the Highly Inflected Croatian Language
Jan Šnajder and Bojana Dalbelo Bašić 411

Classification of Emotion Words in Russian and Romanian Languages
Marina Sokolova and Victoria Bobicev 416

Classification of Opinions with Non-affective Adverbs and Adjectives
Marina Sokolova and Guy Lapalme 421

Amharic Part-of-Speech Tagger for Factored Language Modeling
Martha Yifiru Tachbelie and Wolfgang Menzel 428

Improving Unsegmented Statistical Dialogue Act Labelling
Vicent Tamarit, Carlos-D. Martínez-Hinarejos and José Miguel Benedí Ruíz 434

Three Issues in Cross-Language Frame Information Transfer
Sara Tonelli and Emanuele Pianta 441

A Study on Linking Wikipedia Categories to Wordnet Synsets using Text Similarity
Antonio Toral, Óscar Ferrández, Eneko Agirre and Rafael Muñoz 449

Ontology Engineering and Knowledge Extraction for Cross-Lingual Retrieval
Jantine Trapman and Paola Monachesi 455

A Method to Restrict the Blow-up of Hypotheses of a Non-disambiguated
Jernej Vičič, Petr Homola and Vladislav Kuboň 460

Sources of Performance in CRF Transfer Training: a Business Name-tagging Case Study
Marc Vilain, Jonathan Huggins and Ben Wellner 465

Extracting Synonyms from Dictionary Definitions
Tong Wang and Graeme Hirst 471

Instance Sampling Methods for Pronoun Resolution
Holger Wunsch, Sandra Kübler and Rachael Cantrell 478

Approximate Matching for Evaluating Keyphrase Extraction
Torsten Zesch and Iryna Gurevych 484

Too Many Mammals: Improving the Diversity of Automatically Recognized Terms
Ziqi Zhang, Lei Xia, Mark A. Greenwood and José Iria 490

RANLP-09 Programme, 14 September 2009

8^{45}- 9^{00} Opening

9^{00}-10^{00} **Ricardo Baeza-Yates:** *Towards Semantic Search* (*invited talk*)

Morning session 1a	Morning session 1b
Semantics	*Parsing*
10^{00}-10^{30} **Unsupervised Extraction of False Friends from Parallel Bi-Texts Using the Web as a Corpus** *Svetlin Nakov, Preslav Nakov and Elena Paskaleva*	10^{00}-10^{30} **Lexicalized Semi-Incremental Dependency Parsing** *Hany Hassan, Khalil Sima'an and Andy Way*
10^{30}-11^{00} **Identifying Semantic Relations in Context: Near-misses and Overlaps** *Alla Rozovskaya and Roxana Girju*	10^{30}-11^{00} **Co-Parsing with Competitive Models** *Lidia Khmylko, Kilian Foth and Wolfgang Menzel*

11^{00}-11^{30} Coffee break

Morning session 2a	Morning session 2b
Sentiment analysis	*Lexicon, dictionary*
11^{30}-12^{00} **Classification of Opinions with Non-affective Adverbs and Adjectives** *Marina Sokolova and Guy Lapalme*	11^{35}-12^{00} **A morphological and syntactic wide-coverage lexicon for Spanish: The Leffe** *Miguel Angel Molinero, Benoît Sagot and Lionel Nicolas*
12^{00}-12^{25} **Multi-entity Sentiment Scoring** *Karo Moilanen and Stephen Pulman*	12^{00}-12^{25} **Exploring the Vector Space Model for Finding Verb Synonyms in Portuguese** *Luís Sarmento, Paula Carvalho and Eugénio Oliveira*
12^{25}-12^{50} **Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach** *Vassiliki Rentoumi, George Giannakopoulos, Vangelis Karkaletsis and George Vouros*	12^{25}-12^{50} **Grouping synonyms by definitions** *Ingrid Falk, Claire Gardent, Evelyne Jacquey and Fabienne Venant*

12^{50}-14^{30} Lunch

14^{30}-15^{30} **Kevin Bretonnel Cohen:** *Paradigms for Evaluation in Natural Language Processing* (*invited talk*)

Afternoon session 1a	Afternoon session 1b
Summarisation	*Grammars, POS tagging*
15^{30}-15^{55} **A Semi-supervised Approach for Generating a Table-of-Contents** *Viet Cuong Nguyen, Le Minh Nguyen and Akira Shimazu*	15^{30}-15^{55} **An Interaction Grammar of Interrogative & Relative Clauses in French** *Guy Perrier*
15^{55}-16^{20} **Summary Generation for Toponym-referenced Images using Object Type Language Models** *Ahmet Aker and Robert Gaizauskas*	15^{55}-16^{20} **Evaluating the Impact of Morphosyntactic Ambiguity in Grammatical Error Detection** *Arantza Díaz de Ilarraza, Koldo Gojenola and Maite Oronoz*
16^{20}-16^{45} **Contextual Salience in Query-based Summarization** *Wauter Bosma*	16^{20}-16^{45} **Amharic Part-of-Speech Tagger for Factored Language Modeling** *Martha Yifiru Tachbelie and Wolfgang Menzel*

16^{45}-17^{15} Coffee break

17^{15}-18^{30} ***Poster presentations: session 1***

RANLP-09 CONFERENCE POSTERS

Global Evaluation of Random Indexing through Swedish Word Clustering Compared to the People's Dictionary of Synonyms, *Magnus Rosell, Martin Hassel and Viggo Kann*

Exploiting Latent Semantic Relations in Highly Linked Hypertext for Information Retrieval in Wikis, *Tristan Miller, Bertin Klein and Elisabeth Wolf*

From Partial toward Full Parsing, *Heshaam Faili*

Unsupervised Word Sense Induction from Multiple Semantic Spaces with Locality Sensitive Hashing, *Claire Mouton, Guillaume Pitel, Gaël de Chalendar and Anne Vilnat*

A Model for the Cross-Modal Influence of Visual Context upon Language Processing, *Patrick McCrae*

Detection of Opinions and Facts. A Cognitive Approach
Yann Vigile Hoareau, Adil El-Ghali and Charles Tijus

STUDENT RESEARCH WORKSHOP POSTERS

Normalized Accessor Variety in Chinese Word Segmentation Based on Conditional Random Fields, *Saike He, Tao Zheng Zxhang, Xue Bai, Xiaojie Wang and Yuan Dong*

LOGICON: A System for Extracting Semantic Structure using Partial Parsing, *Kais Dukes*

An Evaluation of Output Quality of Machine Translation Program, *Mitra Shahabi*

Pronunciation Modeling for Dialectal Arabic Speech Recognition, *Hassan Al-Haj and Roger Hsiao*

Hierarchical Discourse Parsing based on Similarity Metrics, *Ravikiran Vadlapudi, Poornima Malepati and Suman Yelati*

Does Language Shape Thought? Time Estimation in Speakers of English and Persian, *Omid Tabatabaei, Ali Reza Ahmadi and Bahar Assarzadegan*

A Study of Machine Learning Algorithms for Recognizing Textual Entailment, *Julio Javier Castillo*

Exploring Context Variation and Lexicon Coverage in Projection-based Approach for Term Translation, *Raphaël Rubino*

ZAC.PB: An Annotated Corpus for Zero Anaphora Resolution in Portuguese, *Simone Pereira*

A Rule Based Approach to the Identification of Spanish Zero Pronouns, *Luz Rello and Iustina Ilisei*

Context Driven XML Retrieval, *Aneliya Tincheva*

Framework for Using a Natural Language Approach to Object Identification, *Mosa Elbendak*

Improving the Output from Software that Generates Multiple Choice Question (MCQ) Test Items Automatically using Controlled Rhetorical Structure Theory, *Robert Foster*

18^{30}-19^{15} Demonstration of Ontotext software products

Integrating Language Technologies in Modern Semantic Processing
Atanas Kiryakov and Georgi Georgiev, Ontotext AD, SIRMA group

RANLP-09 Programme, 15 September 2009

9^{00}-10^{00} **Mirella Lapata:** *Vector-based Models of Semantic Composition* (*invited talk)*

Morning session 1a	Morning session 1b
Applications: MT and QA	*Semantics*
10^{00}-10^{30} **Interactive Machine Translation Based on Partial Statistical Phrase-based Alignments** *Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta*	10^{00}-10^{30} **Extracting Synonyms from Dictionary Definitions** *Tong Wang and Graeme Hirst*
10^{30}-11^{00} **Question Answering over Structured Data: an Entailment-Based Approach to Question Analysis** *Matteo Negri and Milen Kouylekov*	10^{30}-11^{00} **Combining Lexical Resources for Contextual Synonym Expansion** *Ravi Sinha and Rada Mihalcea*

11^{00}-11^{30} Coffee break

Morning session 2a	Morning session 2b
Parsing	*Term extraction*
11^{30}-11^{55} **Prepositional Phrase Attachment in Shallow Parsing** ***Vincent Van Asch and Walter Daelemans***	11^{30}-11^{55} **Evaluating Term Extraction** *Adeline Nazarenko and Haïfa Zargayouna*
11^{55}-12^{20} **Exploring Treebank Transformations in Dependency Parsing** *Kepa Bengoetxea and Koldo Gojenola*	11^{55}-12^{20} **Too Many Mammals: Improving the Diversity of Automatically Recognized Terms** *Ziqi Zhang, Lei Xia, Mark A. Greenwood and José Iria*
12^{20}-12^{45} **Dependency Parsing and Semantic Role Labeling as a Single Task** *Roser Morante, Vincent Van Asch and Antal van den Bosch*	

12^{45}-14^{30} Lunch

14^{30}-15^{30} **Shalom Lappin:** *Restricting Probability Distributions to Expand the Class of Learnable Languages (invited talk)*

Afternoon session 1a	Afternoon session 1b
Word-sense disambiguation, temporal processing	*Language learning*
15^{30}-15^{55} **Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm** *Egoitz Laparra and German Rigau*	15^{30}-15^{55} **Singular Value Decomposition for Feature Selection in Taxonomy Learning** *Francesca Fallucchi and Fabio Massimo Zanzotto*
15^{55}-16^{20} **Semi-Supervised Learning for Word Sense Disambiguation: Quality vs. Quantity** *Sandra Kübler and Desislava Zhekova*	15^{55}-16^{20} **Prototype-based Active Learning for Lemmatization** *Walter Daelemans, Hendrik J. Groenewald and Gerhard B. van Huyssteen*
16^{20}-16^{45} **Using Semantic Networks to Identify Temporal Expressions from Semantic Roles** *Hector Llorens, Borja Navarro and Estela Saquete*	16^{20}-16^{45} **Sources of Performance in CRF Transfer Training: a Business Name-tagging Case Study** *Marc Vilain, Jonathan Huggins and Ben Wellner*

16^{45}-17^{15} Coffee break

17^{00} - 18^{30}	*Student Research Workshop Oral Presentations*
Poster presentations: session 2 **RANLP-09 CONFERENCE POSTERS**	17^{15}-17^{45} **A Two-stage Bootstrapping Algorithm for Relation Extraction** *Ang Sun* 17^{45}-18^{15} **Mink: An Incremental Data-Driven Dependency Parser with Integrated Conversion to Semantics** *Rachael Cantrell* 18^{15}-18^{45} **Event Ordering. Temporal Annotation on Top of the BulTreeBank** *Laska Laskova*

Poster session 2: **RANLP-09 CONFERENCE POSTERS**

String Distance-based Stemming of the Highly Inflected Croatian Language, *Jan Šnajder and Bojana Dalbelo Bašić*

All Words Unsupervised Semantic Category Labeling for Hindi, *Siva Reddy, Abhilash Inumella, Rajeev Sangal and Soma Paul*

The Influence of Text Pre-processing on Plagiarism Detection, *Zdenek Ceska and Chris Fox*

Combining Finite State and Corpus-based Techniques for Unknown Word Prediction, *Kostadin Cholakov and Gertjan van Noord*

Identification of Parallel Text Pairs Using Fingerprints, *Martin Hassel and Hercules Dalianis*

Unsupervised Relation Extraction for Automatic Generation of Multiple-Choice Questions, *Naveed Afzal and Viktor Pekar*

Bimodal Corpora Terminology Extraction: Another Brick in the Wall, *Claudiu Mihăilă and Dalila Mekhaldi*

Integrating Document Structure into a Multi-Document Summarizer, *Aurélien Bossard and Thierry Poibeau*

Classification of Emotion Words in Russian and Romanian Languages, *Marina Sokolova and Victoria Bobicev*

Cross-Linguistic Sentiment Analysis: From English to Spanish
Julian Brooke, Milan Tofiloski and Maite Taboada

A Comparative Study of Open Domain and Opinion Question Answering Systems for Factual and Opinionated Queries, *Alexandra Balahur, Ester Boldrini, Andrés Montoyo and Patricio Martínez-Barco*

Sampling-based Multilingual Alignment, *Adrien Lardilleux and Yves Lepage*

Language-Independent Sentiment Analysis Using Subjectivity and Positional Information, *Veselin Raychev and Preslav Nakov*

Ontology Engineering and Knowledge Extraction for Crosslingual Retrieval, *Jantine Trapman and Paola Monachesi*

User's Choice of Precision and Recall in Named Entity Recognition, *Roman Klinger and Christoph M. Friedrich*

Learning to Identify Educational Materials, *Samer Hassan and Rada Mihalcea*

RANLP-09 Programme, 16 September 2009

9^{00}-10^{00} **Walter Daelemans:** *Robust Features for Computational Stylometry* (*invited talk)*

Morning session 1a	Morning session 1b
Dialogue, lexical resources	*Named Entity Recognition, Diacritization*
10^{00}-10^{30} **Improving Unsegmented Statistical Dialogue Act Labelling** *Vicent Tamarit, Carlos-D. Martínez-Hinarejos and José-Miguel Benedí Ruíz*	10^{00}-10^{30} **Feature Subset Selection in Conditional Random Fields for Named Entity Recognition** *Roman Klinger and Christoph M. Friedrich*
10^{30}-11^{00} **Three Issues in Cross-language Frame Information Transfer** *Sara Tonelli and Emanuele Pianta*	10^{30}-11^{00} **Diacritization for Real-World Arabic Texts** *Emad Mohamed and Sandra Kübler*

11^{00}-11^{30} Coffee break

Morning session 2a	Morning session 2b
Topic segmentation and modelling	*Natural Language Generation, keyphrase extraction*
11^{30}-11^{55} **Improving Text Segmentation by Combining Endogenous and Exogenous Methods** *Olivier Ferret*	11^{30}-11^{55} **Text Content and Task Performance in the Evaluation of a Natural Language Generation System** *Albert Gatt and François Portet*
11^{55}-12^{20} **Topic Modeling of Research Fields: An Interdisciplinary Perspective** *Michael Paul and Roxana Girju*	11^{55}-12^{20} **A Classification-driven Approach to Document Planning** *Rafael Oliveira, Eder Novais, Roberto Araujo and Ivandré Paraboni*
12^{20}-12^{45} **Topic-based Multi-Document Summarization with Probabilistic Latent Semantic Analysis** *Leonhard Hennig*	12^{20}-12^{45} **Approximate Matching for Evaluating Keyphrase Extraction** *Torsten Zesch and Iryna Gurevych*

12^{45}-14^{30} Lunch

Afternoon session 1a	Afternoon session 1b
Similarity, Information Extraction	*Language resources, language models, anaphora resolution*
14^{30}-14^{55} **A Study on Linking Wikipedia Categories to Wordnet Synsets using Text Similarity** *Antonio Toral, Óscar Ferrández, Eneko Agirre and Rafael Muñoz*	14^{30}-14^{55} **Towards Efficient Production of Linguistic Resources: the Victoria Project** *Lionel Nicolas, Miguel Molinero, Benoît Sagot, Elena S. Trigo, Éric de la Clergerie, Miguel Al. Pardo, Jacques Farré and Joan Miquel Vergés*
14^{55}-15^{20} **Comparing Statistical Similarity Measures for Stylistic Multivariate Analysis** *Marius Popescu and Liviu Dinu*	14^{55}-15^{20} **Structured Output Learning with Polynomial Kernel** *Hajime Morita, Hiroya Takamura and Manabu Okumura*
15^{20}-15^{45} **Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges** *Heng Ji, Ralph Grishman, Zheng Chen and Prashant Gupta*	15^{20}-15^{45} **Instance Sampling Methods for Pronoun Resolution** *Holger Wunsch, Sandra Kübler and Rachael Cantrell*

15^{45} - Coffee Break

16^{00} - 17^{30} *Poster presentations: session 3*

RANLP-09 CONFERENCE POSTERS

17^{30}-18^{30} **Massimo Poesio:** *Conceptual Knowledge: Evidence from Corpora and the Brain (invited talk)*

18^{30} **Closure**

Poster session 3: **RANLP-09 CONFERENCE POSTERS**

Statistical Confidence Measures for Probabilistic Parsing, *Ricardo Sánchez-Sáez, Joan Andreu Sánchez and José-Miguel Benedí*

Robust Compositional Polarity Classification, *Manfred Klenner, Stefanos Petrakis and Angela Fahrni*

Fast Boosting-based Part-of-Speech Tagging and Text Chunking with Efficient Rule Representation for Sequential Labeling, *Tomoya Iwakura*

Unsupervised Knowledge Extraction for Taxonomies of Concepts from Wikipedia, *Eduard Barbu and Massimo Poesio*

A Unified Method for Extracting Simple and Multiword Verbs with Valence Information and Application for Hungarian, *Bálint Sass*

The Design of an Experiment in Anaphora Resolution for Referring Expressions Generation, *Diego Jesus de Lucena and Ivandré Paraboni*

How Limited is the Limit? *Prakash Mondal*

Acquisition of Common Sense Knowledge for Basic Level Concepts, *Eduard Barbu*

Uncertainty Detection for Information Extraction, *Bénédicte Goujon*

Stochastic Definite Clause Grammars, *Christian Theil Have*

A Method to Restrict the Blow-up of Hypotheses of a Non-disambiguated Shallow Machine Translation System, *Jernej Vičič, Petr Homola and Vladislav Kuboň*

From Bag of Languages to Family Trees from Noisy Corpus, *Taraka Rama and Anil Kumar Singh*

Treelex Meets Adjectival Tables, *Anna Kupść*

Feature-Rich Named Entity Recognition for Bulgarian Using Conditional Random Fields, *Georgi Georgiev, Preslav Nakov, Kuzman Ganchev, Petya Osenova and Kiril Simov*

Exploiting the Use of Prior Probabilities for Passage Retrieval in Question Answering, *Surya Ganesh and Vasudeva Varma*

Exploiting Structure and Content of Wikipedia for Query Expansion in the Context of Question Answering, *Surya Ganesh and Vasudeva Varma*

Edlin: an Easy to Read Linear Learning Framework, *Kuzman Ganchev and Georgi Georgiev*

Large Vocabulary Continuous Speech Recognition for Bulgarian, *Petar Mitankin, Stoyan Mihov and Tinko Tinchev*