



University of Wolverhampton
**Research Institute in Information and
 Language Processing**
 Research Group in Computational
 Linguistics



Bulgarian Academy of Sciences
Institute for Parallel Processing
 Linguistic Modelling Department

Association for Computational Linguistics - Bulgaria

Programme of
 the 7th International Conference on

**Recent Advances
 in Natural Language Processing**

RANLP-2009

September 12 – 18, 2009

Borovets, Bulgaria
 Hotel SAMOKOV

BULGARIA, SERDICA and PRESLAV halls



RANLP-2009 TUTORIALS, 12-13 September 2009

RODINA and BULGARIA halls

12 September, Saturday, RODINA hall	13 September, Sunday, RODINA hall
<p>10⁰⁰-13⁴⁰</p> <p>Kiril Simov and Petya Osenova</p> <p><i>In the NLP world of Knowledge Nets</i></p> <p>11⁰⁰-11²⁰ Coffee break 12²⁰-12⁴⁰ Break</p>	<p>9⁰⁰-12⁴⁰</p> <p>Roberto Navigli</p> <p><i>Graph-based Word Sense Disambiguation and Discrimination</i></p> <p>10⁰⁰-10²⁰ Coffee break 11²⁰-11⁴⁰ Break</p>
13 ⁴⁰ -15 ⁰⁰ Lunch break	12 ⁴⁰ -14 ⁰⁰ Lunch break
<p>15⁰⁰-18⁴⁰</p> <p>Constantin Orasan</p> <p><i>Automatic Summarisation in the Information Age</i></p> <p>16⁰⁰-16²⁰ Coffee break 17²⁰-17⁴⁰ Break</p>	<p>14⁰⁰-17⁴⁰</p> <p>Kevin Bretonnel Cohen</p> <p><i>Biomedical Natural Language Processing: BioNLP</i></p> <p>15⁰⁰-15²⁰ Coffee break 16²⁰-16⁴⁰ Break</p>

18⁴⁵ – Welcome cocktail in the lobby

Monday, September 14

8⁴⁵-9⁰⁰ Opening (BULGARIA hall)

9⁰⁰-10⁰⁰ **Ricardo Baeza-Yates: Towards Semantic Search (invited talk)**,
BULGARIA hall

Bulgaria hall	Serdica hall
<p><i>Semantics</i></p> <p>10⁰⁰-10³⁰ Unsupervised Extraction of False Friends from Parallel Bi-Texts Using the Web as a Corpus <i>Svetlin Nakov, Preslav Nakov and Elena Paskaleva</i></p> <p>10³⁰-11⁰⁰ Identifying Semantic Relations in Context: Near-misses and Overlaps <i>Alla Rozovskaya and Roxana Girju</i></p>	<p><i>Parsing</i></p> <p>10⁰⁰-10³⁰ Lexicalized Semi-incremental Dependency Parsing <i>Hany Hassan, Khalil Sima'an and Andy Way</i></p> <p>10³⁰-11⁰⁰ Co-Parsing with Competitive Models <i>Lidia Khmylko, Kilian Foth and Wolfgang Menzel</i></p>

11⁰⁰-11³⁰ Coffee Break, BULGARIA Hall lobby

Bulgaria hall	Serdica hall
<p><i>Sentiment analysis</i></p> <p>11³⁰-12⁰⁰ Classification of Opinions with Non-affective Adverbs and Adjectives <i>Marina Sokolova and Guy Lapalme</i></p> <p>12⁰⁰-12²⁵ Multi-entity Sentiment Scoring <i>Karo Moilanen and Stephen Pulman</i></p> <p>12²⁵-12⁵⁰ Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach <i>Vassiliki Rentoumi, George Giannakopoulos, Vangelis Karkaletsis and George Vouros</i></p>	<p><i>Lexicon, dictionary</i></p> <p>11³⁵-12⁰⁰ A morphological and syntactic wide-coverage lexicon for Spanish: The Leffe <i>Miguel Angel Molinero, Benoît Sagot and Lionel Nicolas</i></p> <p>12⁰⁰-12²⁵ Exploring the Vector Space Model for Finding Verb Synonyms in Portuguese <i>Luís Sarmento, Paula Carvalho and Eugénio Oliveira</i></p> <p>12²⁵-12⁵⁰ Grouping synonyms by definitions <i>Ingrid Falk, Claire Gardent, Evelyne Jacquey and Fabienne Venant</i></p>

Monday, September 14

12⁵⁰-14³⁰ Lunch Break

14³⁰-15³⁰ **Kevin Bretonnel Cohen: Paradigms for Evaluation in Natural Language Processing (invited talk)**, BULGARIA hall

BULGARIA hall	SERDICA hall
<p><i>Summarisation</i></p> <p>15³⁰-15⁵⁵ A Semi-supervised Approach for Generating a Table-of-Contents <i>Viet Cuong Nguyen, Le Minh Nguyen and Akira Shimazu</i></p> <p>15⁵⁵-16²⁰ Summary Generation for Toponym-referenced Images using Object Type Language Models <i>Ahmet Aker and Robert Gaizauskas</i></p> <p>16²⁰-16⁴⁵ Contextual Saliency in Query-based Summarization <i>Wauter Bosma</i></p>	<p><i>Grammars, POS tagging</i></p> <p>15³⁰-15⁵⁵ An Interaction Grammar of Interrogative & Relative Clauses in French <i>Guy Perrier</i></p> <p>15⁵⁵-16²⁰ Evaluating the Impact of Morphosyntactic Ambiguity in Grammatical Error Detection <i>Arantza Díaz de Ilaraza, Koldo Gojenola and Maite Oronoz</i></p> <p>16²⁰-16⁴⁵ Amharic Part-of-Speech Tagger for Factored Language Modeling <i>Martha Yifiru Tachbelie and Wolfgang Menzel</i></p>

16⁴⁵ - Coffee Break, BULGARIA Hall lobby

<p>17⁰⁰ - 18³⁰</p> <p>Poster Session 1</p> <p>Bulgaria hall lobby</p>

18³⁰-19¹⁵ Demonstration of Ontotext software products, BULGARIA Hall

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

9⁰⁰-10⁰⁰ **Mirella Lapata**: *Vector-based Models of Semantic Composition (invited talk)*, BULGARIA hall

12⁴⁵-14³⁰ Lunch Break

14³⁰-15³⁰ **Shalom Lappin**: *Restricting Probability Distributions to Expand the Class of Learnable Languages (invited talk)*, BULGARIA hall

Bulgaria hall	Serdica hall
<p><i>Applications: MT and QA</i></p> <p>10⁰⁰-10³⁰ Interactive Machine Translation Based on Partial Statistical Phrase-based Alignments <i>Daniel Ortiz-Martínez, Ismael García-Varea and Francisco Casacuberta</i></p> <p>10³⁰-11⁰⁰ Question Answering over Structured Data: an Entailment-Based Approach to Question Analysis <i>Matteo Negri and Milen Kouylekov</i></p>	<p><i>Semantics</i></p> <p>10⁰⁰-10³⁰ Extracting Synonyms from Dictionary Definitions <i>Tong Wang and Graeme Hirst</i></p> <p>10³⁰-11⁰⁰ Combining Lexical Resources for Contextual Synonym Expansion <i>Ravi Sinha and Rada Mihalcea</i></p>

BULGARIA hall	SERDICA hall
<p><i>Word-sense disambiguation, temporal processing</i></p> <p>15³⁰-15⁵⁵ Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm <i>Egoitz Laparra and German Rigau</i></p> <p>15⁵⁵-16²⁰ Semi-Supervised Learning for Word Sense Disambiguation: Quality vs. Quantity <i>Sandra Kübler and Desislava Zhekova</i></p> <p>16²⁰-16⁴⁵ Using Semantic Networks to Identify Temporal Expressions from Semantic Roles <i>Hector Llorens, Borja Navarro and Estela Saquete</i></p>	<p><i>Language learning</i></p> <p>15³⁰-15⁵⁵ Singular Value Decomposition for Feature Selection in Taxonomy Learning <i>Francesca Fallucchi and Fabio Massimo Zanzotto</i></p> <p>15⁵⁵-16²⁰ Prototype-based Active Learning for Lemmatization <i>Walter Daelemans, Hendrik J. Groenewald and Gerhard B. van Huyssteen</i></p> <p>16²⁰-16⁴⁵ Sources of Performance in CRF Transfer Training: a Business Name-tagging Case Study <i>Marc Vilain, Jonathan Huggins and Ben Wellner</i></p>

11⁰⁰-11³⁰ Coffee Break, BULGARIA Hall lobby

16⁴⁵ - Coffee Break, BULGARIA Hall lobby

Bulgaria hall	Serdica hall
<p><i>Parsing</i></p> <p>11³⁰-11⁵⁵ Prepositional Phrase Attachment in Shallow Parsing <i>Vincent Van Asch and Walter Daelemans</i></p> <p>11⁵⁵-12²⁰ Exploring Treebank Transformations in Dependency Parsing <i>Kepa Bengoetxea and Koldo Gojenola</i></p> <p>12²⁰-12⁴⁵ Dependency Parsing and Semantic Role Labeling as a Single Task <i>Roser Morante, Vincent Van Asch and Antal van den Bosch</i></p>	<p><i>Term extraction</i></p> <p>11³⁰-11⁵⁵ Evaluating Term Extraction <i>Adeline Nazarenko and Haïfa Zargayouna</i></p> <p>11⁵⁵-12²⁰ Too Many Mammals: Improving the Diversity of Automatically Recognized Terms <i>Ziqi Zhang, Lei Xia, Mark A. Greenwood and José Iria</i></p>

BULGARIA hall lobby	PRESLAV hall
<p>17⁰⁰ - 18³⁰</p> <p>Poster Session 2</p>	<p><i>Student Research Workshop: oral presentations</i></p> <p>17¹⁵-17⁴⁵ A Two-stage Bootstrapping Algorithm for Relation Extraction <i>Ang Sun</i></p> <p>17⁴⁵-18¹⁵ Mink: An Incremental Data-Driven Dependency Parser with Integrated Conversion to Semantics <i>Rachael Cantrell</i></p> <p>18¹⁵-18⁴⁵ Event Ordering. Temporal Annotation on Top of the BulTreeBank <i>Laska Laskova</i></p>

19:45 – 2 am Conference reception and disco, Samokov hotel

9⁰⁰-10⁰⁰ **Walter Daelemans: Robust Features for Computational Stylometry (invited talk)**, BULGARIA hall

12⁴⁵-14³⁰ Lunch Break

Bulgaria hall	Serdica hall
<i>Dialogue, lexical resources</i>	<i>Named Entity Recognition, Diacritization</i>
10 ⁰⁰ -10 ³⁰ Improving Unsegmented Statistical Dialogue Act Labelling <i>Vicent Tamarit, Carlos-D. Martínez-Hinarejos and José-Miguel Benedí Ruíz</i>	10 ⁰⁰ -10 ³⁰ Feature Subset Selection in Conditional Random Fields for Named Entity Recognition <i>Roman Klinger and Christoph M. Friedrich</i>
10 ³⁰ -11 ⁰⁰ Three Issues in Cross-language Frame Information Transfer <i>Sara Tonelli and Emanuele Pianta</i>	10 ³⁰ -11 ⁰⁰ Diacritization for Real-World Arabic Texts <i>Emad Mohamed and Sandra Kübler</i>

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

11⁰⁰-11³⁰ Coffee Break, BULGARIA Hall lobby

15⁴⁵ - Coffee Break, BULGARIA Hall lobby

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

<p>16⁰⁰ - 17³⁰</p> <p>Poster Session 3</p> <p>Bulgaria hall lobby</p>

17³⁰-18³⁰ **Massimo Poesio: Conceptual Knowledge: Evidence from Corpora and the Brain (invited talk)**, BULGARIA hall

18³⁰ Closure

Monday, September 14

17⁰⁰-18³⁰ Poster Session 1, BULGARIA hall lobby

RANLP 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

Saïke 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

Ambiguous Arabic Words Disambiguation: The Results

Laroussi Merhben, Anis Zouaghi and Mounir Zrigui

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 Assar zadegan

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

Effect of Minimal Semantics on Dependency Parsing

Bharat Ram Ambati, Pujitha Gade, Chaitanya GSK and Samar Husain

Tuesday, September 15

17⁰⁰-18³⁰ **Poster Session 2, BULGARIA** hall lobby

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

Wednesday, September 16

16⁰⁰-17³⁰ **Poster Session 3, BULGARIA** hall lobby

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

Bénédictte 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

International Workshop – 14-15 September

RANLP-2009 Student Research Workshop

*Organised by Irina Temnikova, Ivelina Nikolova and
Natalia Konstantinova*

<http://lml.bas.bg/ranlp2009/stud-ranlp09>

14 September, Poster presentations 17⁰⁰-18³⁰ Bulgaria Hall lobby

15 September, Oral presentations 17¹⁵-18⁴⁵ Preslav Hall

International Workshops – 17 September

W1: Multilingual Resources, Technologies and Evaluation for Central and Eastern European Languages

(full-day workshop)

*Organised by Cristina Vertan, Stelios Piperidis, Elena Paskaleva and
Milena Slavcheva*

<http://www.c-phil.uni-hamburg.de/view/Main/RanlpBalkan09>

W2: Adaptation of Language Resources and Technology to New Domains

(full-day workshop)

*Organised by Nuria Bel, Erhard Hinrichs, Kiril Simov and
Petya Osenova*

<http://www.bultreebank.org/AdaptLRTtoND/>

W3: Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning

(full-day workshop)

Organised by Viktor Pekar, Iustina Narcisa Ilisei, and Silvia Bernardini

<http://pers-www.wlv.ac.uk/~in0963/ranlp>

International Workshops – 17 September

W4: Events in Emerging Text Types (eETTs)

(full-day workshop)

Organised by Constantin Orasan, Laura Hasler and Corina Forascu

<http://clg.wlv.ac.uk/events/eETTs/index.php>

19⁰⁰ – Welcome cocktail for the workshops in the lobby

International Workshops – 18 September

W5: 1st Workshop on Definition Extraction

(full-day workshop)

Organised by Gerardo Eugenio Sierra Martínez

<http://www.iling.unam.mx/wde>

W6: Biomedical Information Extraction

(full-day workshop)

*Organised by Guergana Savova, Vangelis Karkaletsis
and Galia Angelova*

<http://lml.bas.bg/ranlp2009/biomedicalIE/index.htm>