

# Table of Contents

## Language Resources

A Distributed Database System for Developing Ontological and Lexical Resources in Harmony . . . . .	1
<i>Aleš Horák, Piek Vossen, and Adam Rambousek</i>	
Verb Class Discovery from Rich Syntactic Data . . . . .	16
<i>Lin Sun, Anna Korhonen and Yuval Krymolowski</i>	
Growing TreeLex . . . . .	28
<i>Anna Kupść and Anne Abeillé</i>	
Acquisition of Elementary Synonym Relations from Biological Structured Terminology . . . . .	40
<i>Thierry Hamon and Natalia Grabar</i>	
A Comparison of Co-occurrence and Similarity Measures as Simulations of Context . . . . .	52
<i>Stefan Bordag</i>	
Various Criteria of Collocation Cohesion in Internet: Comparison of Resolving Power . . . . .	64
<i>Igor A. Bolshakov, Elena I. Bolshakova, Alexey P. Kotlyarov, and Alexander Gelbukh</i>	
Why Don't Romanians Have a Five O'clock Tea, Nor Halloween, But Have a Kind of Valentines Day? . . . . .	73
<i>Corina Forăscu</i>	

## Best Student Paper Award

SIGNUM: A Graph Algorithm for Terminology Extraction . . . . .	85
<i>Axel-Cyrille Ngonga Ngomo</i>	

## Morphology and Syntax

Arabic Morphology Parsing Revisited . . . . .	96
<i>Suhel Jaber and Rodolfo Delmonte</i>	
A Probabilistic Model for Guessing Base Forms of New Words by Analogy . . . . .	106
<i>Krister Lindén</i>	

Unsupervised and Knowledge-Free Learning of Compound Splits and Periphrases . . . . .	117
<i>Florian Holz and Chris Biemann</i>	
German Decomposing in a Difficult Corpus . . . . .	128
<i>Enrique Alfonseca, Slaven Bilac, and Stefan Pharies</i>	
Clause Boundary Identification Using Conditional Random Fields . . . . .	140
<i>R. Vijay Sundar Ram and Sobha Lalitha Devi</i>	

## Semantics and Discourse

### Invited Paper

Natural Language as the Basis for Meaning Representation and Inference . . . . .	151
<i>Ido Dagan, Roy Bar-Haim, Idan Szpektor, Iddo Greental, and Eyal Shnarch</i>	
Layer Structures and Conceptual Hierarchies in Semantic Representations for NLP . . . . .	171
<i>Hermann Helbig, Ingo Glöckner, and Rainer Osswald</i>	
Deep Lexical Semantics . . . . .	183
<i>Jerry R. Hobbs</i>	
On Ontology Based Abduction for Text Interpretation . . . . .	194
<i>Irma Sofía Espinosa Peraldi, Atila Kaya, Sylvia Melzer, and Ralf Möller</i>	
Analysis of Joint Inference Strategies for the Semantic Role Labeling of Spanish and Catalan . . . . .	206
<i>Mihai Surdeanu, Roser Morante, and Lluís Màrquez</i>	
A Preliminary Study on the Robustness and Generalization of Role Sets for Semantic Role Labeling . . . . .	219
<i>Beñat Zepirain, Eneko Agirre, and Lluís Màrquez</i>	
XTM: A Robust Temporal Text Processor . . . . .	231
<i>Caroline Hagège and Xavier Tannier</i>	

### Invited Paper

What We Are Talking about and What We Are Saying about It . . . . .	241
<i>Eva Hajičová</i>	
Trusting Politicians' Words (for Persuasive NLP) . . . . .	263
<i>Marco Guerini, Carlo Strapparava, and Oliviero Stock</i>	

Sense Annotation in the Penn Discourse Treebank .....	275
<i>Eleni Miltsakaki, Livio Robaldo, Alan Lee, and Aravind Joshi</i>	

## Word Sense Disambiguation and Named Entity Recognition

A Semantics-Enhanced Language Model for Unsupervised Word Sense Disambiguation .....	287
<i>Show-de Lin and Karin Verspoor</i>	

## Best Paper Award, 1st Place

Discovering Word Senses from Text Using Random Indexing .....	299
<i>Niladri Chatterjee and Shiwali Mohan</i>	

Domain Information for Fine-Grained Person Name Categorization .....	311
<i>Zornitsa Kozareva, Sonia Vazquez, and Andres Montoyo</i>	

Language Independent First and Last Name Identification in Person Names .....	322
<i>Octavian Popescu and Bernardo Magnini</i>	

Mixing Statistical and Symbolic Approaches for Chemical Names Recognition .....	334
<i>Florian Boudin, Juan Manuel Torres-Moreno, and Marc El-Bèze</i>	

## Anaphora and Co-reference

Portuguese Pronoun Resolution: Resources and Evaluation .....	344
<i>Ramon Ré Moya Cuevas, Willian Yukio Honda, Diego Jesus de Lucena, Ivandré Paraboni, and Patrícia Rufino Oliveira</i>	

Semantic and Syntactic Features for Dutch Coreference Resolution .....	351
<i>Iris Hendrickx, Veronique Hoste, and Walter Daelemans</i>	

## Machine Translation and Parallel Corpora

### Invited Paper

Stat-XFER: A General Search-Based Syntax-Driven Framework for Machine Translation .....	362
<i>Alon Lavie</i>	

**Invited Paper**

Statistical Machine Translation into a Morphologically Complex Language .....	376
<i>Kemal Oflazer</i>	
Translation Paraphrases in Phrase-Based Machine Translation .....	388
<i>Francisco Guzmán and Leonardo Garrido</i>	
n-Best Reranking for the Efficient Integration of Word Sense Disambiguation and Statistical Machine Translation .....	399
<i>Lucia Specia, Baskaran Sankaran, and Maria das Graças Volpe Nunes</i>	
Learning Finite State Transducers Using Bilingual Phrases .....	411
<i>Jorge González, Germán Sanchis, and Francisco Casacuberta</i>	
Learning Spanish-Galician Translation Equivalents Using a Comparable Corpus and a Bilingual Dictionary .....	423
<i>Pablo Gamallo Otero and José Ramon Pichel Campos</i>	
Context-Based Sentence Alignment in Parallel Corpora .....	434
<i>Ergun Biçici</i>	
Bilingual Segmentation for Alignment and Translation .....	445
<i>Chung-Chi Huang, Wei-Teh Chen, and Jason S. Chang</i>	
Dynamic Translation Memory: Using Statistical Machine Translation to Improve Translation Memory Fuzzy Matches .....	454
<i>Ergun Biçici and Marc Dymetman</i>	
Identification of Transliterated Foreign Words in Hebrew Script .....	466
<i>Yoav Goldberg and Michael Elhadad</i>	

**Natural Language Generation**

Innovative Approach for Engineering NLG Systems: The Content Determination Case Study .....	478
<i>Marco Fonseca, Leonardo Junior, Alexandre Melo, and Hendrik Macedo</i>	

**Speech Recognition**

Comparison of Different Modeling Units for Language Model Adaptation for Inflected Languages .....	488
<i>Tanel Alumäe</i>	

## Information Retrieval and Question Answering

Word Distribution Analysis for Relevance Ranking and Query Expansion .....	500
<i>Patricio Galeas and Bernd Freisleben</i>	
Hybrid Method for Personalized Search in Scientific Digital Libraries ...	512
<i>Thanh-Trung Van and Michel Beigbeder</i>	
Alignment-Based Expansion of Textual Database Fields.....	522
<i>Piroska Lendvai</i>	
Detecting Expected Answer Relations through Textual Entailment .....	532
<i>Matteo Negri, Milen Kouylekov, and Bernardo Magnini</i>	
Improving Question Answering by Combining Multiple Systems Via Answer Validation .....	544
<i>Alberto Téllez-Valero, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda, and Anselmo Peñas</i>	

## Text Classification

Evaluation of Internal Validity Measures in Short-Text Corpora.....	555
<i>Diego Ingaramo, David Pinto, Paolo Rosso, and Marcelo Errecalde</i>	

## Text Summarization

Arabic/English Multi-document Summarization with CLASSY—The Past and the Future .....	568
<i>Judith D. Schlesinger, Dianne P. O’Leary, and John M. Conroy</i>	

## Best Paper Award, 3rd Place

Lexical Cohesion Based Topic Modeling for Summarization .....	582
<i>Gonenc Ercan and Ilyas Cicekli</i>	
Terms Derived from Frequent Sequences for Extractive Text Summarization .....	593
<i>Yulia Ledeneva, Alexander Gelbukh, and René Arnulfo García-Hernández</i>	

## Spell Checking and Authoring Aid

Real-Word Spelling Correction with Trigrams: A Reconsideration of the Mays, Damerau, and Mercer Model .....	605
<i>Amber Wilcox-O’Hearn, Graeme Hirst, and Alexander Budanitsky</i>	

**Best Paper Award, 2nd Place**

Non-interactive OCR Post-correction for Giga-Scale Digitization  
Projects ..... 617  
*Martin Reynaert*

Linguistic Support for Revising and Editing ..... 631  
*Cerstin Mahlow and Michael Piotrowski*

The Role of PP Attachment in Preposition Generation ..... 643  
*John Lee and Ola Knutsson*

EFL Learner Reading Time Model for Evaluating Reading  
Proficiency ..... 655  
*Katsunori Kotani, Takehiko Yoshimi, Takeshi Kutsumi,  
Ichiko Sata, and Hitoshi Isahara*

**Author Index** ..... 665