

===== THIS IS A DRAFT! Errors have been corrected in the final version =====

## Table of Contents

### Trends and Opportunities

Has Computational Linguistics Become More Applied? (Invited paper) . . . . .	1
<i>Kenneth Church</i>	
Opportunities for Natural Language Processing Research in Education (Invited Paper) . . . . .	6
<i>Jill Burstein</i>	

### Linguistic Knowledge Representation Formalisms

Information Structure in a Formal Framework: Uni.cation-Based Combinatory Categorical Grammar . . . . .	28
<i>Maarika Traat</i>	
A Karaka Based Annotation Scheme for English . . . . .	41
<i>Ashwini Vaidya, Samar Husain, Prashanth Mannem, and Dipti Misra Sharma</i>	

### Corpus Analysis and Lexical Resources

Substring Statistics . . . . .	53
<i>Kyoji Umemura and Kenneth Church</i>	
Evaluation of the Syntactic Annotation in EPEC, the Reference Corpus for the Processing of Basque . . . . .	72
<i>Larraitz Uria, Ainara Estarrona, Izaskun Aldezabal, Maria Jes ´us Aranzabe, Arantza D ´yaz de Ilarraza, and Mikel Iruskieta</i>	
Reducing Noise in Labels and Features for a Real World Dataset: Application of NLP Corpus Annotation Methods. . . . .	86
<i>Rebecca J. Passonneau, Cynthia Rudin, Axinia Radeva, and Zhi An Liu</i>	
Unsupervised Classi.cation of Verb Noun Multi-Word Expression Tokens . . . . .	98
<i>Mona T. Diab and Madhav Krishna</i>	

### Extraction of Lexical Knowledge

Semantic Mapping for Related Term Identi.cation . . . . .	111
<i>Rafael E. Banchs</i>	
An Improved Automatic Term Recognition Method for Spanish . . . . .	125
<i>Alberto Barr ´on-Cede˜no, Gerardo Sierra, Patrick Drouin, and Sophia Ananiadou</i>	
Bootstrapping a Verb Lexicon for Biomedical Information Extraction . . .	137
<i>Giulia Venturi, Simonetta Montemagni, Simone Marchi, Yutaka Sasaki, Paul Thompson, John McNaught, and Sophia Ananiadou</i>	
VIII Table of Contents	
TermeX: A Tool for Collocation Extraction . . . . .	149
<i>Davor Delaˆj, Zoran Krleˆza, Jan ˆSnajder, Bojana Dalbelo Baˆsi´c, and Frane ˆSari´c</i>	

### Morphology and Parsing

Guessers for Finite-State Transducer Lexicons . . . . .	158
<i>Kristel Lind'en</i>	
Combining Language Modeling and Discriminative Classification for Word Segmentation (Invited Talk) . . . . .	170
<i>Dekang Lin</i>	
Formal Grammar for Hispanic Named Entities Analysis . . . . .	183
<i>Grettel Barcel'ó, Eduardo Cendejas, Grigori Sidorov, and Igor A. Bolshakov</i>	
Automatic Extraction of Clause Relationships from a Treebank . . . . .	195
<i>Oldrich Krç'uzá and Vladislav Kubojn</i>	
A General Method for Transforming Standard Parsers into Error-Repair Parsers . . . . .	207
<i>Carlos G'omez-Rodr'íguez, Miguel A. Alonso, and Manuel Vilares</i>	

### **Semantics**

Topic-Focus Articulation from the Semantic Point of View . . . . .	220
<i>Marie Duiz'ý</i>	
The Value of Weights in Automatically Generated Text Structures . . . . .	233
<i>Dana Dann'ells</i>	
AORTE for Recognizing Textual Entailment . . . . .	245
<i>Reda Sibliñi and Leila Kosseim</i>	

### **Word Sense Disambiguation**

Semi-supervised Word Sense Disambiguation Using the Web as Corpus . . . . .	256
<i>Rafael Guzm'an-Cabrera, Paolo Rosso, Manuel Montes-y-G'omez, Luis Villaseñor-Pineda, and David Pinto-Avenidaño</i>	
Semi-supervised Clustering for Word Instances and Its Effect on Word Sense Disambiguation . . . . .	266
<i>Kazunari Sugiyama and Manabu Okumura</i>	
Alleviating the Problem of Wrong Coreferences in Web Person Search . . . . .	280
<i>Octavian Popescu and Bernardo Magnini</i>	
Improved Unsupervised Name Discrimination with Very Wide Bigrams and Automatic Cluster Stopping . . . . .	294
<i>Ted Pedersen</i>	
Table of Contents IX	

### **Machine Translation and Multilingualism**

Enriching Statistical Translation Models Using a Domain-Independent Multilingual Lexical Knowledge Base . . . . .	306
<i>Miguel Garc'ía, Jesús Gim'enez, and Llu'ís M'arquez</i>	
Exploiting Parallel Treebanks to Improve Phrase-Based Statistical Machine Translation . . . . .	318
<i>John Tinsley, Mary Hearne, and Andy Way</i>	
Cross-Language Frame Semantics Transfer in Bilingual Corpora . . . . .	332
<i>Roberto Basili, Diego De Cao, Danilo Croce, Bonaventura Coppola, and Alessandro Moschitti</i>	
A Parallel Corpus Labeled Using Open and Restricted Domain Ontologies . . . . .	346
<i>Ester Boldrini, Sergio Ferr'andez, Ruben Izquierdo,</i>	

<i>David Tom'as, and Jose Luis Vicedo</i>	
Language Identification on the Web: Extending the Dictionary	
Method . . . . .	357
<i>Radim J. Rehe'ujrek and Milan Kolkus</i>	

### **Information Extraction and Text Mining**

Business Specific Online Information Extraction from German	
Websites . . . . .	369
<i>Yeong Su Lee and Michaela Geierhos</i>	
Low-Cost Supervision for Multiple-Source Attribute Extraction . . . . .	382
<i>Joseph Reisinger and Marius Pasca</i>	
An Integrated Architecture for Processing Business Documents in	
Turkish . . . . .	394
<i>Serif Adali, A. Coskun Sonmez, and Mehmet Gokturk</i>	
Detecting Protein-Protein Interactions in Biomedical Texts Using a	
Parser and Linguistic Resources. . . . .	406
<i>Gerold Schneider, Kaarel Kaljurand, and Fabio Rinaldi</i>	
Learning to Learn Biological Relations from a Small Training Set . . . . .	418
<i>Laura Alonso i Alemany and Santiago Bruno</i>	
Using a Bigram Event Model to Predict Causal Potential . . . . .	430
<i>Brandon Beamer and Roxana Girju</i>	
Semantic-Based Temporal Text-Rule Mining . . . . .	442
<i>Kjetil N'orv'e'ag and Ole Kristian Fivelstad</i>	
Generating Executable Scenarios from Natural Language . . . . .	456
<i>Michal Gordon and David Harel</i>	
Determining the Polarity and Source of Opinions Expressed in Political	
Debates. . . . .	468
<i>Alexandra Balahur, Zornitsa Kozareva, and Andr'es Montoyo</i>	
X Table of Contents	

### **Information Retrieval and Text Comparison**

Query Translation and Expansion for Searching Normal and	
OCR-Degraded Arabic Text . . . . .	481
<i>Tarek Elghazaly and Aly Fahmy</i>	
NLP for Shallow Question Answering of Legal Documents Using	
Graphs . . . . .	498
<i>Alfredo Monroy, Hiram Calvo, and Alexander Gelbukh</i>	
Semantic Clustering for a Functional Text Classification Task. . . . .	509
<i>Thomas Lippincott and Rebecca Passonneau</i>	
Reducing the Plagiarism Detection Search Space on the Basis of the	
Kullback-Leibler Distance . . . . .	523
<i>Alberto Barr'on-Cede'no, Paolo Rosso, and Jos'e-Miguel Bened'ıy</i>	
Empirical Paraphrasing of Modern Greek Text in Two Phases: An	
Application to Steganography . . . . .	535
<i>Katia Lida Kermanidis and Emmanouil Magkos</i>	
BorderFlow: A Local Graph Clustering Algorithm for Natural Language	
Processing . . . . .	547
<i>Axel-Cyrille Ngonga Ngomo and Frank Schumacher</i>	
Generalized Mongue-Elkan Method for Approximate Text String	
Comparison . . . . .	559
<i>Sergio Jimenez, Claudia Becerra, Alexander Gelbukh, and</i>	
<i>Fabio Gonzalez</i>	

### **Text Summarization**

Estimating Risk of Picking a Sentence for Document Summarization . . .	571
<i>Chandan Kumar, Prasad Pingali, and Vasudeva Varma</i>	
The Decomposition of Human-Written Book Summaries . . . . .	582
<i>Hakan Ceylan and Rada Mihalcea</i>	

**Applications to the Humanities**

Linguistic Ethnography: Identifying Dominant Word Classes in Text . . . . .	594
<i>Rada Mihalcea and Stephen Pulman</i>	

<b>Author Index</b> . . . . .	603
-------------------------------	-----