

CONTENTS

Editors' Foreword	ix
-------------------	----

I. COMPUTATION FOR LINGUISTICS

<i>John Nerbonne</i> Linguistic challenges for computationalists	1
---	---

II. INFORMATION EXTRACTION & INDEXING

<i>Ralph Grishman</i> NLP: An information extraction perspective	17
<i>Florian Seydoux & Jean-Cédric Chappelier</i> Semantic indexing using minimum redundancy cut in ontologies	25
<i>Niraj Aswani, Valentin Tablan, Kalina Bontcheva & Hamish Cunningham</i> Indexing and querying linguistic metadata and document content	35
<i>Irina Matveeva, Gina-Anne Levow, Ayman Farahat & Christiaan Royer</i> Term representation with Generalized Latent Semantic Analysis	45

III. PARSING

<i>Ming-Wei Chang, Quang Do & Dan Roth</i> Multilingual dependency parsing: A pipeline approach	55
<i>Sandra Kübler</i> How does treebank annotation influence parsing? Or how not to compare apples and oranges	79
<i>Laura Alonso, Joan Antoni Capilla, Irene Castellón, Ana Fernández-Montraveta & Gloria Vázquez</i> The SenSem project: Syntactico-semantic annotation of sentences in Spanish	89

IV. ANAPHORA & REFERRING EXPRESSIONS

Robert Dale

- Generating referring expressions: Past, present and future 99

Erhard W. Hinrichs, Katja Filippova & Holger Wunsch

- A data-driven approach to pronominal anaphora resolution for German 115

V. CLASSIFICATION

Nicolas Nicolov & Franco Salvetti

- Efficient spam analysis for weblogs through URL segmentation 125

Filip Ginter, Sampo Pyysalo & Tapio Salakoski

- Document classification using semantic networks
with an adaptive similarity measure 137

Rada Mihalcea & Samer Hassan

- Text summarization for improved text classification 147

Caroline Sporleder & Alex Lascarides

- Exploiting linguistic cues to classify rhetorical relations 157

VI. TEXTUAL ENTAILMENT & QUESTION ANSWERING

Milen Kouylekov & Bernardo Magnini

- Tree edit distance for textual entailment 167

Jörg Tiedemann

- A genetic algorithm for optimising information retrieval
with linguistic features in question answering 177

Vasile Rus & Arthur Graesser

- Lexico-syntactic subsumption for textual entailment 187

Courtney Corley, Andras Csomai & Rada Mihalcea

- A knowledge-based approach to text-to-text similarity 197

VII. ONTOLOGIES*Keiji Shinzato & Kentaro Torisawa*

A simple WWW-based method for semantic word class acquisition 207

Eduard Barbu & Verginica Barbu Mititelu

Automatic building of Wordnets 217

VIII. MACHINE TRANSLATION*Stelios Piperidis, Panagiotis Dimitrakis & Irene Balta*

Lexical transfer selection using annotated parallel corpora 227

Victoria Arranz, Elisabet Comelles & David Farwell

Multi-perspective evaluation of the FAME speech-to-speech translation system for Catalan, English and Spanish 237

Dániel Varga, Péter Halácsy, András Kornai, Viktor Nagy, László Németh & Viktor Trón

Parallel corpora for medium density languages 247

IX. CORPORA*Anne De Roeck*

The role of data in NLP: The case for dataset profiling 259

Anne De Roeck, Avik Sarkar & Paul H. Garthwaite

Even very frequent function words do not distribute homogeneously 267

Lucia Specia, Maria das Graças V. Nunes & Mark Stevenson

Exploiting parallel texts to produce a multilingual sense tagged corpus for word sense disambiguation 277

Francis Chantree, Alistair Willis, Adam Kilgarriff & Anne de Roeck

Detecting dangerous coordination ambiguities using word distribution 287

List and Addresses of Contributors 297**Index of Subjects and Terms 303**