

Lecture Notes in Artificial Intelligence 5221

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Bengt Nordström Aarne Ranta (Eds.)

Advances in Natural Language Processing

6th International Conference, GoTAL 2008
Gothenburg, Sweden, August 25-27, 2008
Proceedings

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada

Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Bengt Nordström

Aarne Ranta

Chalmers University of Technology

Department of Computer Science and Engineering

41296 Göteborg, Sweden

E-mail: {bengt, aarne}@chalmers.se

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.7, F.4.2-3, I.2, H.3, I.7

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-540-85286-7 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-85286-5 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2008

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper SPIN: 12463534 06/3180 5 4 3 2 1 0

Preface

This volume contains the papers presented at GoTAL 2008, the 6th International Conference on Natural Language Processing, held on August 25–27, 2008, at Chalmers University of Technology in Gothenburg, Sweden. GoTAL was the sixth conference in the TAL series, preceded by FracTAL 1997 (Université de Franche-Comté, Besançon, France), VexTAL 1999 (Università Ca’ Foscari di Venezia, Venice, Italy), PorTAL 2002 (Universidade do Algarve, Faro, Portugal), EsTAL 2004 (Universitat d’Alacant, Alicante, Spain), and FinTAL 2006 (University of Turku, Turku, Finland).

The conference received 107 submissions. Each submission was reviewed by three programme committee members or external reviewers. The committee finally accepted 44 papers to be presented at the conference and included in the proceedings. The conference programme also included three invited talks, which are the first three papers in this volume.

We are grateful to the programme committee members and the external reviewers for their careful and punctual work. The staff in the local organization team at Chalmers helped in a very professional way. The sponsors contributed, in particular, to the social programme planned for the conference. The invited speakers – Johan Bos, Lori Lamel, and Joakim Nivre – gave the scientific programme the broad, yet focused, profile that we wanted to achieve. And finally, it is essentially the authors of the submissions that created the substance of the conference and this volume, with all their good papers and their cooperative attitude.

The EasyChair software was used throughout the reviewing and editing process. It saved us a lot of work by doing exactly the things that could be automated, in exactly the ways we expected.

June 2008

Bengt Nordström
Aarne Ranta

Organization

Local Organization

Björn Bringert
Håkan Burden
Rebecca Cyrén
Markus Forsberg
Harald Hammarström
Tiina Rankanen

Programme Committee

Olli Aaltonen	University of Helsinki, Finland
Walid El Abed	Nestle Corp., Switzerland
Jan Alexandersson	DFKI, Germany
Jorge Baptista	University of Algarve, Portugal
Tilman Becker	DFKI, Germany
Chris Biemann	Powerset, USA
Patrick Blackburn	INRIA Lorraine
Lars Borin	University of Gothenburg, Sweden
Johan Bos	University of Rome “La Sapienza,” Italy
Johan Boye	SpeechAct, Sweden
Caroline Brun	Xerox Corp., France
Sylviane Cardey	University of Franche-Comté, France
Lauri Carlson	University of Helsinki, Finland
Rolf Carlson	KTH, Sweden
Alexander Clark	Royal Holloway University of London, UK
Robin Cooper	University of Gothenburg, Sweden
Walter Daelemans	University of Antwerp, Belgium
Rodolfo Delmonte	University of Venice, Italy
Elisabet Engdahl	University of Gothenburg, Sweden
Jan van Eijck	CWI Amsterdam, The Netherlands
Filip Ginter	University of Turku, Finland
Peter Greenfield	University of Franche-Comté, France
Philippe de Groote	INRIA Lorraine, France
Viggo Kann	KTH, Sweden
Kimmo Koskenniemi	University of Helsinki, Finland
Hans Leiß	LMU Munich, Germany
Oliver Lemon	University of Edinburgh, UK
Patricio Martinez Barco	University of Alicante, Spain
Adeline Nazarenko	University Paris-Nord, France

Joakim Nivre	Växjö University and Uppsala University, Sweden
Bengt Nordström	Chalmers University of Technology, Sweden
Pierre Nugues	University of Lund, Sweden
Guy Perrier	INRIA Lorraine, France
Elisabete Ranchhod	University of Lisbon, Portugal
Aarne Ranta	University of Gothenburg, Sweden (Chair)
Manny Rayner	University of Geneva, Switzerland
Tapio Salakoski	University of Turku, Finland
Karl-Michael Schneider	Textkernel, The Netherlands
Rolf Schwitter	Macquarie University, Australia
José Luis Vicedo	University of Alicante, Spain
Simo Vihjanen	Lingsoft Ltd., Finland
Annie Zaenen	Palo Alto Research Center, USA

External Reviewers

Krasimir Angelov	Richard Johansson
Björn Bringert	Lauri Karttunen
Håkan Burden	Janna Khagai
Maud Ehrmann	Marco Kuhlmann
Samuel Eleutério	Joseph Le Roux
Oscar Ferrandez	Peter Ljunglöf
Sergio Ferrández	Beáta Megyesi
Andrei Filip	Borja Navarro Colorado
Bruno Guillaume	Magnus Rosell
Markus Forsberg	Fernando Ruiz-Rico
Caroline Hagege	Markus Saers
Harald Hammarström	Estela Saquete
Rubén Izquierdo Beviá	Kamel Smaïli
Guillaume Jacquet	David Tomás
Kristofer Johannisson	Marcus Uneson

Sponsors

Centre for Language Technology, Gothenburg
City of Göteborg
Lingsoft Ltd., Helsinki

Table of Contents

Formal Semantics in the Real World	1
<i>Johan Bos</i>	
Speech Processing for Audio Indexing	4
<i>Lori Lamel and Jean-Luc Gauvain</i>	
Sorting Out Dependency Parsing	16
<i>Joakim Nivre</i>	
“I Know What You Feel”: Analyzing the Role of Conjunctions in Automatic Sentiment Analysis	28
<i>Ritesh Agarwal, T.V. Prabhakar, and Sugato Chakrabarty</i>	
Automatic Annotation of Direct Reported Speech in Arabic and French, According to a Semantic Map of Enunciative Modalities	40
<i>Motasem Alrahabi and Jean-Pierre Desclés</i>	
Type-Theoretical Bulgarian Grammar	52
<i>Krasimir Angelov</i>	
A Compact Arabic Lexical Semantics Language Resource Based on the Theory of Semantic Fields	65
<i>Mohamed Attia, Mohsen Rashwan, Ahmed Ragheb, Mohamed Al-Badrashiny, Husein Al-Basoumy, and Sherif Abdou</i>	
Automatically Extracting Personal Name Aliases from the Web	77
<i>Danushka Bollegala, Taiki Honma, Yutaka Matsuo, and Mitsuru Ishizuka</i>	
An Efficient Statistical Approach for Automatic Organic Chemistry Summarization	89
<i>Florian Boudin, Juan-Manuel Torres-Moreno, and Patricia Velázquez-Morales</i>	
Augmenting Word Space Models for Word Sense Discrimination Using an Automatic Thesaurus	100
<i>Hiram Calvo</i>	
Plagiarism Detection Based on Singular Value Decomposition	108
<i>Zdenek Ceska</i>	
Networking Multiword Units	120
<i>Mathieu Constant and Patrick Watrin</i>	

Searching for Part of Speech Tags That Improve Parsing Models	126
<i>Martín Ariel Domínguez and Gabriel Infante-Lopez</i>	
A POS-Based Word Prediction System for the Persian Language	138
<i>Masood Ghayoomi and Ehsan Daroodi</i>	
A Graph Based Method for Building Multilingual Weakly Supervised Dependency Parsers	148
<i>Jagadeesh Gorla, Anil Kumar Singh, Rajeev Sangal, Karthik Gali, Samar Husain, and Sriram Venkatapathy</i>	
A Web-Based Self-training Approach for Authorship Attribution	160
<i>Rafael Guzmán-Cabrera, Manuel Montes-y-Gómez, Paolo Rosso, and Luis Villaseñor-Pineda</i>	
Parsing Discontinuous Phrase Structure with Grammatical Functions . . .	169
<i>Johan Hall and Joakim Nivre</i>	
How Can the Term Compositionality be Useful for Acquiring Elementary Semantic Relations?	181
<i>Thierry Hamon and Natalia Grabar</i>	
Training Statistical Language Models from Grammar-Generated Data: A Comparative Case-Study	193
<i>Beth Ann Hockey, Manny Rayner, and Gwen Christian</i>	
A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI)	205
<i>Anton Karl Ingason, Sigrún Helgadóttir, Hrafn Loftsson, and Eiríkur Rögnvaldsson</i>	
Semantic Roles in Valency Lexicon of Czech Verbs: Verbs of Communication and Exchange	217
<i>Václava Kettnerová, Markéta Lopatková, and Klára Hrstková</i>	
Automatic Generation of Frequent Case Forms of Query Keywords in Text Retrieval	222
<i>Kimmo Kettunen</i>	
Definition Extraction with Balanced Random Forests	237
<i>Lukasz Kobyliński and Adam Przepiórkowski</i>	
Reviewing and Evaluating Automatic Term Recognition Techniques . . .	248
<i>Ioannis Korkontzelos, Ioannis P. Klapaftis, and Suresh Manandhar</i>	
Finding Text Boundaries and Finding Topic Boundaries: Two Different Tasks?	260
<i>Alexandre Labadié and Violaine Prince</i>	

Tamil Question Classification Using Morpheme Features	272
<i>S. Lakshmana Pandian and T.V. Geetha</i>	
Analogical Translation of Medical Words in Different Languages	284
<i>Philippe Langlais, François Yvon, and Pierre Zweigenbaum</i>	
Improving Chinese Pronominal Anaphora Resolution by Extensive Feature Representation and Confidence Estimation	296
<i>Tyne Liang and Dian-Song Wu</i>	
A Grammar Formalism for Specifying ISU-Based Dialogue Systems	303
<i>Peter Ljunglöf and Staffan Larsson</i>	
Word Sense Disambiguation of Farsi Homographs Using Thesaurus and Corpus	315
<i>Raheleh Makki and Mohammad Mehdi Homayounpour</i>	
Local Rephrasing Suggestions for Supporting the Work of Writers	324
<i>Aurélien Max</i>	
Interactive Multilingual Web Applications with Grammatical Framework	336
<i>Moisés Salvador Meza Moreno and Björn Bringert</i>	
Dependency Parsing by Transformation and Combination	348
<i>Jens Nilsson and Joakim Nivre</i>	
Using Constraints over Finite Sets of Integers for Range Concatenation Grammar Parsing	360
<i>Yannick Parmentier and Wolfgang Maier</i>	
Analyzing Argumentative Structures in Procedural Texts	366
<i>Lionel Fontan and Patrick Saint-Dizier</i>	
Natural Language Processing Across Time: An Empirical Investigation on Italian	371
<i>Marco Pennacchiotti and Fabio Massimo Zanzotto</i>	
Statistical Surface Realisation of Portuguese Referring Expressions	383
<i>Daniel Bastos Pereira and Ivandré Paraboni</i>	
Classification-Based Filtering of Semantic Relatedness in Hypernymy Extraction	393
<i>Maciej Piasecki, Stanisław Szpakowicz, Michał Marcińczuk, and Bartosz Broda</i>	
Similarity of Names Across Scripts: Edit Distance Using Learned Costs of N-Grams	405
<i>Bruno Pouliquen</i>	

Turkish Language Resources: Morphological Parser, Morphological Disambiguator and Web Corpus	417
<i>Hasım Sak, Tunga Güngör, and Murat Saraçlar</i>	
A Graph Partitioning Approach to Entity Disambiguation Using Uncertain Information	428
<i>Emili Sapena, Lluís Padró, and Jordi Turmo</i>	
Arabic Named Entity Recognition from Diverse Text Types	440
<i>Khaled Shaalan and Hafsa Raza</i>	
A Noun-Predicate Bigram-Based Similarity Measure for Lexical Relations	452
<i>Hyopil Shin and Insik Cho</i>	
German Compounds in Factored Statistical Machine Translation	464
<i>Sara Stymne</i>	
A Reordering Model for Phrase-Based Machine Translation	476
<i>Vinh Van Nguyen, Thai Phuong Nguyen, Akira Shimazu, and Minh Le Nguyen</i>	
Interruption, Resumption and Domain Switching in In-Vehicle Dialogue	488
<i>Jessica Villing, Cecilia Holtelius, Staffan Larsson, Anders Lindström, Alexander Seward, and Nina Åberg</i>	
Finite Matters: Verbal Features in Data-Driven Parsing of Swedish	500
<i>Lilja Øvrelid</i>	
Author Index	511