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 automatized, 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
VIII Organization

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 Khegai
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
Formal Semantics in the Real World .......................................................... 1
   Johan Bos

Speech Processing for Audio Indexing ...................................................... 4
   Lori Lamel and Jean-Luc Gauvain

Sorting Out Dependency Parsing ............................................................... 16
   Joakim Nivre

“I Know What You Feel”: Analyzing the Role of Conjunctions in
Automatic Sentiment Analysis ............................................................... 28
   Ritesh Agarwal, T.V. Prabhakar, and Sugato Chakrabarty

Automatic Annotation of Direct Reported Speech in Arabic and
French, According to a Semantic Map of Enunciative Modalities ............... 40
   Motasem Alrahabi and Jean-Pierre Desclés

Type-Theoretical Bulgarian Grammar ....................................................... 52
   Krasimir Angelov

A Compact Arabic Lexical Semantics Language Resource Based on the
Theory of Semantic Fields ........................................................................ 65
   Mohamed Attia, Mohsen Rashwan, Ahmed Ragheb, Mohamed Al-Badrashiny,
   Husein Al-Basoumy, and Sherif Abdou

Automatically Extracting Personal Name Aliases from the Web ............... 77
   Danushka Bollegala, Taiki Honma, Yutaka Matsuo, and Mitsuru Ishizuka

An Efficient Statistical Approach for Automatic Organic Chemistry
Summarization ............................................................................................ 89
   Florian Boudin, Juan-Manuel Torres-Moreno, and Patricia Velázquez-Morales

Augmenting Word Space Models for Word Sense Discrimination Using
an Automatic Thesaurus ............................................................................ 100
   Hiram Calvo

Plagiarism Detection Based on Singular Value Decomposition ................. 108
   Zdenek Ceska

Networking Multiword Units ..................................................................... 120
   Matthieu Constant and Patrick Watrin
Searching for Part of Speech Tags That Improve Parsing Models ........ 126
   Martín Ariel Domínguez and Gabriel Infante-Lopez

A POS-Based Word Prediction System for the Persian Language ........ 138
   Masood Ghayoomi and Ehsan Daroodi

A Graph Based Method for Building Multilingual Weakly Supervised
Dependency Parsers ............................................................... 148
   Jagadeesh Gorla, Anil Kumar Singh, Rajeev Sangal, Karthik Gali,
   Samar Husain, and Sriram Venkatapathy

A Web-Based Self-training Approach for Authorship Attribution ........ 160
   Rafael Guzmán-Cabrera, Manuel Montes-y-Gómez,
   Paolo Rosso, and Luis Villaseñor-Pineda

Parsing Discontinuous Phrase Structure with Grammatical Functions ... 169
   Johan Hall and Joakim Nivre

How Can the Term Compositionality be Useful for Acquiring
Elementary Semantic Relations? ............................................. 181
   Thierry Hamon and Natalia Grabar

Training Statistical Language Models from Grammar-Generated Data:
A Comparative Case-Study ..................................................... 193
   Beth Ann Hockey, Manny Rayner, and Gwen Christian

A Mixed Method Lemmatization Algorithm Using a Hierarchy of
Linguistic Identities (HOLI) .................................................... 205
   Anton Karl Ingason, Sigrún Helgadóttir, Hrafn Loftsson, and
   Eiríkur Rögnvaldsson

Semantic Roles in Valency Lexicon of Czech Verbs: Verbs of
Communication and Exchange ............................................... 217
   Václava Kettnerová, Markéta Lopatková, and Klára Hrstková

Automatic Generation of Frequent Case Forms of Query Keywords in
Text Retrieval ................................................................. 222
   Kimmo Kettunen

Definition Extraction with Balanced Random Forests .................... 237
   Łukasz Kobyláński and Adam Przepiórkowski

Reviewing and Evaluating Automatic Term Recognition Techniques ... 248
   Ioannis Korkontzelos, Ioannis P. Klapaftis, and Suresh Manandhar

Finding Text Boundaries and Finding Topic Boundaries: Two Different
Tasks? .................................................................................... 260
   Alexandre Labadié and Violaine Prince
Tamil Question Classification Using Morpheme Features .................. 272
   S. Lakshmana Pandian and T.V. Geetha

Analogical Translation of Medical Words in Different Languages ....... 284
   Philippe Langlais, François Yvon, and Pierre Zweigenbaum

Improving Chinese Pronominal Anaphora Resolution by Extensive
Feature Representation and Confidence Estimation ..................... 296
   Tyne Liang and Dian-Song Wu

A Grammar Formalism for Specifying ISU-Based Dialogue Systems .... 303
   Peter Ljunglöf and Staffan Larsson

Word Sense Disambiguation of Farsi Homographs Using Thesaurus and
Corpus .................................................................................. 315
   Raheleh Makki and Mohammad Mehdi Homayounpour

Local Rephrasing Suggestions for Supporting the Work of Writers ...... 324
   Aurélien Max

Interactive Multilingual Web Applications with Grammatical
Framework ............................................................................. 336
   Moisés Salvador Meza Moreno and Björn Bringert

Dependency Parsing by Transformation and Combination ............... 348
   Jens Nilsson and Joakim Nivre

Using Constraints over Finite Sets of Integers for Range Concatenation
Grammar Parsing ..................................................................... 360
   Yannick Parmentier and Wolfgang Maier

Analyzing Argumentative Structures in Procedural Texts ............... 366
   Lionel Fontan and Patrick Saint-Dizier

Natural Language Processing Across Time: An Empirical Investigation
on Italian .................................................................................. 371
   Marco Pennacchiotti and Fabio Massimo Zanzotto

Statistical Surface Realisation of Portuguese Referring Expressions ..... 383
   Daniel Bastos Pereira and Ivandré Paraboni

Classification-Based Filtering of Semantic Relatedness in Hyponymy
Extraction .............................................................................. 393
   Maciej Piasecki, Stanisław Szpakowicz, Michał Marciniiczuk, and
   Bartosz Broda

Similarity of Names Across Scripts: Edit Distance Using Learned Costs
of N-Grams ............................................................................. 405
   Bruno Pouliquen
Table of Contents

Turkish Language Resources: Morphological Parser, Morphological Disambiguator and Web Corpus ................................................. 417
\textit{Hasim Sak, Tunga Gung\"or, and Murat Sara\c{c}lar}

A Graph Partitioning Approach to Entity Disambiguation Using Uncertain Information ................................................................. 428
\textit{Emili Sapena, Llu\'is Padr\'o, and Jordi Turmo}

Arabic Named Entity Recognition from Diverse Text Types .......... 440
\textit{Khaled Shaalan and Hafsa Raza}

A Noun-Predicate Bigram-Based Similarity Measure for Lexical Relations ..................................................................................... 452
\textit{Hyopil Shin and Insik Cho}

German Compounds in Factored Statistical Machine Translation ...... 464
\textit{Sara Stymne}

A Reordering Model for Phrase-Based Machine Translation .......... 476
\textit{Vinh Van Nguyen, Thai Phuong Nguyen, Akira Shimazu, and Minh Le Nguyen}

Interruption, Resumption and Domain Switching in In-Vehicle Dialogue ...................................................................................... 488
\textit{Jessica Villing, Cecilia Holtelius, Staffan Larsson, Anders Lindstr\"om, Alexander Seward, and Nina \AA berg}

Finite Matters: Verbal Features in Data-Driven Parsing of Swedish .... 500
\textit{Lilja \O vrelid}

Author Index .................................................................................. 511