

Trends in Parsing Technology

Text, Speech and Language Technology

VOLUME 43

Series Editors

Nancy Ide, *Vassar College, New York*

Jean Véronis, *Université de Provence and CNRS, France*

Editorial Board

Harald Baayen, *Max Planck Institute for Psycholinguistics, The Netherlands*

Kenneth W. Church, *Microsoft Research Labs, Redmond WA, USA*

Judith Klavans, *Columbia University, New York, USA*

David T. Barnard, *University of Regina, Canada*

Dan Tufis, *Romanian Academy of Sciences, Romania*

Joaquim Llisterri, *Universitat Autònoma de Barcelona, Spain*

Stig Johansson, *University of Oslo, Norway*

Joseph Mariani, *LIMSI-CNRS, France*

For further volumes:

<http://www.springer.com/series/6636>

Trends in Parsing Technology

Dependency Parsing, Domain Adaptation, and Deep Parsing

Edited by

Harry Bunt

Tilburg University, The Netherlands

Paola Merlo

University of Geneva, Switzerland

and

Joakim Nivre

Växjö University and Uppsala University, Sweden

Editors

Harry Bunt
Tilburg University
Tilburg Center for Cognition and
Communication (TiCC) and
Dept. of Communication
and Information Sciences
Warandelaan 2
5000 LE Tilburg
Netherlands
Harry.Bunt@uvt.nl

Paola Merlo
Université de Genève
Dépt. Linguistique
rue de Candolle 2
1211 Genève
Switzerland
paola.merlo@unige.ch

Joakim Nivre
Växjö University
Uppsala University
Pimpstensvägen 16
752 67 Uppsala
Sweden
joakim.nivre@lingfil.uu.se

ISSN 1386-291X

ISBN 978-90-481-9351-6

e-ISBN 978-90-481-9352-3

DOI 10.1007/978-90-481-9352-3

Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010936679

© Springer Science+Business Media B.V. 2010

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

1	Current Trends in Parsing Technology	1
	Paola Merlo, Harry Bunt, and Joakim Nivre	
2	Single Malt or Blended? A Study in Multilingual Parser Optimization	19
	Johan Hall, Jens Nilsson, and Joakim Nivre	
3	A Latent Variable Model for Generative Dependency Parsing	35
	Ivan Titov and James Henderson	
4	Dependency Parsing and Domain Adaptation with Data-Driven LR Models and Parser Ensembles	57
	Kenji Sagae and Jun-ichi Tsujii	
5	Dependency Parsing Using Global Features	69
	Tetsuji Nakagawa	
6	Dependency Parsing with Second-Order Feature Maps and Annotated Semantic Information	87
	Massimiliano Ciaramita and Giuseppe Attardi	
7	Strictly Lexicalised Dependency Parsing	105
	Qin Iris Wang, Dale Schuurmans, and Dekang Lin	
8	Favor Short Dependencies: Parsing with Soft and Hard Constraints on Dependency Length	121
	Jason Eisner and Noah A. Smith	
9	Corrective Dependency Parsing	151
	Keith Hall and Václav Novák	

10	Inducing Lexicalised PCFGs with Latent Heads	169
	Detlef Prescher	
11	Self-Trained Bilexical Preferences to Improve Disambiguation Accuracy	183
	Gertjan van Noord	
12	Are Very Large Context-Free Grammars Tractable?	201
	Pierre Boullier and Benoît Sagot	
13	Efficiency in Unification-Based <i>N</i>-Best Parsing	223
	Yi Zhang, Stephan Oepen, and John Carroll	
14	HPSG Parsing with a Supertagger	243
	Takashi Ninomiya, Takuya Matsuzaki, Yusuke Miyao, Yoshimasa Tsuruoka, and Jun-ichi Tsujii	
15	Evaluating the Impact of Re-training a Lexical Disambiguation Model on Domain Adaptation of an HPSG Parser	257
	Tadayoshi Hara, Yusuke Miyao, and Jun-ichi Tsujii	
16	Semi-supervised Training of a Statistical Parser from Unlabeled Partially-Bracketed Data	277
	Rebecca Watson, Ted Briscoe, and John Carroll	
	Index	293

Contributors

Giuseppe Attardi Università di Pisa, I-56127, Pisa, Italy, attardi@di.unipi.it

Pierre Boullier INRIA-Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex, France, pierre.boullier@inria.fr

Ted Briscoe Computer Laboratory, University of Cambridge, Cambridge, UK, Ted.Briscoe@cl.cam.ac.uk

Harry Bunt Tilburg University, Tilburg Center for Cognition and Communication (TiCC) and Department of Communication and Information Sciences, Tilburg, The Netherlands, harry.bunt@uvt.nl

John Carroll Department of Informatics, University of Sussex, UK, J.A.Carroll@sussex.ac.uk

Massimiliano Ciaramita Yahoo! Research, S-08018, Barcelona, Catalonia, Spain, massi@alumni.brown.edu

Jason Eisner Johns Hopkins University, Baltimore, MD, USA, jason@cs.jhu.edu

Johan Hall Växjö University, Växjö, Sweden, johan.hall@vxu.se

Keith Hall Google Research, Zurich, Switzerland, kbhall@google.com

Tadayoshi Hara Department of Computer Science, Faculty of Information Science and Technology, University of Tokyo, Tokyo 113-0033, Japan, harasan@is.s.u-tokyo.ac.jp

James Henderson Department of Computer Science, University of Geneva, Geneva, Switzerland, james.henderson@unige.ch

Dekang Lin Google Inc., Mountain View, CA 94043, USA, lindek@google.com

Takuya Matsuzaki Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, matuzaki@is.s.u-tokyo.ac.jp

Paola Merlo University of Geneva, Geneva, Switzerland, Paola.Merlo@unige.ch

Yusuke Miyao Department of Computer Science, Faculty of Information Science and Technology, University of Tokyo, Tokyo, 113-0033, Japan, yusuke@nii.ac.jp

Tetsuji Nakagawa Knowledge Creating Communication Research Center, National Institute of Information and Communications Technology, Kyoto 619-0289, Japan, tnaka@nict.go.jp

Jens Nilsson Växjö University, Växjö, Sweden, jens.nilsson@vxu.se

Takashi Ninomiya Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime, Japan, ninomiya@cs.ehime-u.ac.jp

Joakim Nivre Uppsala University, Uppsala, Sweden, joakim.nivre@lingfil.uu.se

Václav Novák Charles University, Prague, Czech Republic, novak@ufal.mff.cuni.cz

Stephan Oepen Department of Informatics, University of Oslo, Oslo, Norway, oe@ifi.uio.no

Detlef Prescher 76307 Karlsbad, Czech Republic, prescher@snlp.de

Kenji Sagae Institute for Creative Technologies, University of Southern California, Marina del Rey, CA 90292, USA, sagae@ict.usc.edu

Benoît Sagot INRIA-Rocquencourt, Domaine de Voluceau, 78153 Le Chesnay Cedex, France, benoit.sagot@inria.fr

Dale Schuurmans Department of Computing Science, University of Alberta, Edmonton, AB, Canada T6G 2E8, dale@cs.ualberta.ca

Noah A. Smith Carnegie Mellon University, Pittsburgh, PA, USA, nasmith@cs.cmu.edu

Ivan Titov Cluster of Excellence, MMC, Saarland University, Saarbrücken, Germany, titov@mmci.uni-saarland.de

Jun-ichi Tsujii Department of Computer Science, Faculty of Information Science and Technology, University of Tokyo, Tokyo, 113-0033, Japan; School of Computer Science, University of Manchester, Manchester, UK; National Center for Text Mining (NaCTeM), tsujii@is.s.u-tokyo.ac.jp

Yoshimasa Tsuruoka School of Information Science, Japan Advanced Institute of Science and Technology (JAIST), 1-1 Asahidai, Nomi, Ishikawa, Japan, tsuruoka@jaist.ac.jp

Gertjan van Noord Faculty of Arts, University of Groningen, 9700 AS Groningen, The Netherlands, G.J.M.van.Noord@rug.nl

Qin Iris Wang Yahoo! Inc., Santa Clara, CA 95054, USA, wangqin@gmail.com

Rebecca Watson Computer Laboratory, University of Cambridge, Cambridge, UK, Rebecca.Watson@cl.cam.ac.uk

Yi Zhang Language Technology Laboratory, Department of Computational Linguistics, Saarland University, DFKI GmbH, Saarbrücken, Germany, yzhang@coli.uni-sb.de