Tapio Salakoski
Filip Ginter
Sampo Pyysalo
Tapio Pahikkala (Eds.)

LNAI 4139

Advances in Natural Language Processing

5th International Conference on NLP, FinTAL 2006
Turku, Finland, August 2006
Proceedings

Springer

Tapio Salakoski Filip Ginter
Sampo Pyysalo Tapio Pahikkala (Eds.)

Advances in Natural Language Processing

5th International Conference on NLP, FinTAL 2006
Turku, Finland, August 23-25, 2006
Proceedings

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Tapio Salakoski
Filip Ginter
Sampo Pyysalo
Tapio Pahikkala

University of Turku
Department of Information Technology
20014 Turku, Finland
E-mail: {Tapio.Salakoski, Filip.Ginter, Sampo.Pyysalo, Tapio.Pahikkala}@it.utu.fi

Library of Congress Control Number: 2006930503

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 3-540-37334-9 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-37334-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11816508 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 4139

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 4139: T. Salakoski, F. Ginter, S. Pyysalo, T. Pahikkala (Eds.), Advances in Natural Language Processing. XIV, 771 pages. 2006.

Vol. 4114: D.-S. Huang, K. Li, G.W. Irwin (Eds.), Computational Intelligence, Part II. XXVII, 1337 pages. 2006.

Vol. 4108: J.M. Borwein, W.M. Farmer (Eds.), Mathematical Knowledge Management. VIII, 295 pages. 2006.

Vol. 4106: T.R. Roth-Berghofer, M.H. Göker, H. A. Güvenir (Eds.), Advances in Case-Based Reasoning. XIV, 566 pages. 2006.

Vol. 4099: Q. Yang, G. Webb (Eds.), PRICAI 2006: Trends in Artificial Intelligence. XXVIII, 1263 pages. 2006.

Vol. 4093: X. Li, O.R. Zaiane, Z. Li (Eds.), Advanced Data Mining and Applications. XXI, 1110 pages. 2006.

Vol. 4092: J. Lang, F. Lin, J. Wang (Eds.), Knowledge Science, Engineering and Management. XV, 664 pages. 2006.

Vol. 4088: Z.-Z. Shi, R. Sadananda (Eds.), Agent Computing and Multi-Agent Systems. XVII, 827 pages. 2006.

Vol. 4068: H. Schärfe, P. Hitzler, P. Øhrstrøm (Eds.), Conceptual Structures: Inspiration and Application. XI, 455 pages. 2006.

Vol. 4065: P. Perner (Ed.), Advances in Data Mining. XI, 592 pages. 2006.

Vol. 4062: G. Wang, J.F. Peters, A. Skowron, Y. Yao (Eds.), Rough Sets and Knowledge Technology. XX, 810 pages. 2006.

Vol. 4049: S. Parsons, N. Maudet, P. Moraitis, I. Rahwan (Eds.), Argumentation in Multi-Agent Systems. XIV, 313 pages. 2006.

Vol. 4048: L. Goble, J.-J.C.. Meyer (Eds.), Deontic Logic and Artificial Normative Systems. X, 273 pages. 2006.

Vol. 4045: D. Barker-Plummer, R. Cox, N. Swoboda (Eds.), Diagrammatic Representation and Inference. XII, 301 pages. 2006.

Vol. 4031: M. Ali, R. Dapoigny (Eds.), Advances in Applied Artificial Intelligence. XXIII, 1353 pages. 2006.

Vol. 4029: L. Rutkowski, R. Tadeusiewicz, L.A. Zadeh, J.M. Zurada (Eds.), Artificial Intelligence and Soft Computing – ICAISC 2006. XXI, 1235 pages. 2006.

Vol. 4027: H.L. Larsen, G. Pasi, D. Ortiz-Arroyo, T. Andreasen, H. Christiansen (Eds.), Flexible Query Answering Systems. XVIII, 714 pages. 2006.

Vol. 4021: E. André, L. Dybkjær, W. Minker, H. Neumann, M. Weber (Eds.), Perception and Interactive Technologies. XI, 217 pages. 2006.

Vol. 4020: A. Bredenfeld, A. Jacoff, I. Noda, Y. Takahashi (Eds.), RoboCup 2005: Robot Soccer World Cup IX. XVII, 727 pages. 2006.

Vol. 4013: L. Lamontagne, M. Marchand (Eds.), Advances in Artificial Intelligence. XIII, 564 pages. 2006.

Vol. 4012: T. Washio, A. Sakurai, K. Nakajima, H. Takeda, S. Tojo, M. Yokoo (Eds.), New Frontiers in Artificial Intelligence. XIII, 484 pages. 2006.

Vol. 4008: J.C. Augusto, C.D. Nugent (Eds.), Designing Smart Homes. XI, 183 pages. 2006.

Vol. 4005: G. Lugosi, H.U. Simon (Eds.), Learning Theory. XI, 656 pages. 2006.

Vol. 3978: B. Hnich, M. Carlsson, F. Fages, F. Rossi (Eds.), Recent Advances in Constraints. VIII, 179 pages. 2006.

Vol. 3963: O. Dikenelli, M.-P. Gleizes, A. Ricci (Eds.), Engineering Societies in the Agents World VI. XII, 303 pages. 2006.

Vol. 3960: R. Vieira, P. Quaresma, M.d.G.V. Nunes, N.J. Mamede, C. Oliveira, M.C. Dias (Eds.), Computational Processing of the Portuguese Language. XII, 274 pages. 2006.

Vol. 3955: G. Antoniou, G. Potamias, C. Spyropoulos, D. Plexousakis (Eds.), Advances in Artificial Intelligence. XVII, 611 pages. 2006.

Vol. 3949: F. A. Savacı (Ed.), Artificial Intelligence and Neural Networks. IX, 227 pages. 2006.

Vol. 3946: T.R. Roth-Berghofer, S. Schulz, D.B. Leake (Eds.), Modeling and Retrieval of Context. XI, 149 pages. 2006.

Vol. 3944: J. Quiñonero-Candela, I. Dagan, B. Magnini, F. d'Alché-Buc (Eds.), Machine Learning Challenges. XIII, 462 pages. 2006.

Vol. 3930: D.S. Yeung, Z.-Q. Liu, X.-Z. Wang, H. Yan (Eds.), Advances in Machine Learning and Cybernetics. XXI, 1110 pages. 2006.

Vol. 3918: W.K. Ng, M. Kitsuregawa, J. Li, K. Chang (Eds.), Advances in Knowledge Discovery and Data Mining. XXIV, 879 pages. 2006.

Vol. 3913: O. Boissier, J. Padget, V. Dignum, G. Lindemann, E. Matson, S. Ossowski, J.S. Sichman, J. Vázquez-Salceda (Eds.), Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems. XII, 259 pages. 2006.

Vol. 3910: S.A. Brueckner, G.D.M. Serugendo, D. Hales, F. Zambonelli (Eds.), Engineering Self-Organising Systems. XII, 245 pages. 2006.

Vol. 3904: M. Baldoni, U. Endriss, A. Omicini, P. Torroni (Eds.), Declarative Agent Languages and Technologies III. XII, 245 pages. 2006.

Vol. 3900: F. Toni, P. Torroni (Eds.), Computational Logic in Multi-Agent Systems. XVII, 427 pages. 2006.

Vol. 3899: S. Frintrop, VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search. XIV, 216 pages. 2006.

Vol. 3898: K. Tuyls, P.J. 't Hoen, K. Verbeeck, S. Sen (Eds.), Learning and Adaption in Multi-Agent Systems. X, 217 pages. 2006.

Vol. 3891: J.S. Sichman, L. Antunes (Eds.), Multi-Agent-Based Simulation VI. X, 191 pages. 2006.

Vol. 3890: S.G. Thompson, R. Ghanea-Hercock (Eds.), Defence Applications of Multi-Agent Systems. XII, 141 pages. 2006.

Vol. 3885: V. Torra, Y. Narukawa, A. Valls, J. Domingo-Ferrer (Eds.), Modeling Decisions for Artificial Intelligence. XII, 374 pages. 2006.

Vol. 3881: S. Gibet, N. Courty, J.-F. Kamp (Eds.), Gesture in Human-Computer Interaction and Simulation. XIII, 344 pages. 2006.

Vol. 3874: R. Missaoui, J. Schmidt (Eds.), Formal Concept Analysis. X, 309 pages. 2006.

Vol. 3873: L. Maicher, J. Park (Eds.), Charting the Topic Maps Research and Applications Landscape. VIII, 281 pages. 2006.

Vol. 3863: M. Kohlhase (Ed.), Mathematical Knowledge Management. XI, 405 pages. 2006.

Vol. 3862: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 267 pages. 2006.

Vol. 3849: I. Bloch, A. Petrosino, A.G.B. Tettamanzi (Eds.), Fuzzy Logic and Applications. XIV, 438 pages. 2006.

Vol. 3848: J.-F. Boulicaut, L. De Raedt, H. Mannila (Eds.), Constraint-Based Mining and Inductive Databases. X, 401 pages. 2006.

Vol. 3847: K.P. Jantke, A. Lunzer, N. Spyratos, Y. Tanaka (Eds.), Federation over the Web. X, 215 pages. 2006.

Vol. 3835: G. Sutcliffe, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XIV, 744 pages. 2005.

Vol. 3830: D. Weyns, H. V.D. Parunak, F. Michel (Eds.), Environments for Multi-Agent Systems II. VIII, 291 pages. 2006.

Vol. 3817: M. Faundez-Zanuy, L. Janer, A. Esposito, A. Satue-Villar, J. Roure, V. Espinosa-Duro (Eds.), Nonlinear Analyses and Algorithms for Speech Processing. XII, 380 pages. 2006.

Vol. 3814: M. Maybury, O. Stock, W. Wahlster (Eds.), Intelligent Technologies for Interactive Entertainment. XV, 342 pages. 2005.

Vol. 3809: S. Zhang, R. Jarvis (Eds.), AI 2005: Advances in Artificial Intelligence. XXVII, 1344 pages. 2005.

Vol. 3808: C. Bento, A. Cardoso, G. Dias (Eds.), Progress in Artificial Intelligence. XVIII, 704 pages. 2005.

Vol. 3802: Y. Hao, J. Liu, Y.-P. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma, Y.-C. Jiao (Eds.), Computational Intelligence and Security, Part II. XLII, 1166 pages. 2005.

Vol. 3801: Y. Hao, J. Liu, Y.-P. Wang, Y.-m. Cheung, H. Yin, L. Jiao, J. Ma, Y.-C. Jiao (Eds.), Computational Intelligence and Security, Part I. XLI, 1122 pages. 2005.

Vol. 3789: A. Gelbukh, Á. de Albornoz, H. Terashima-Marín (Eds.), MICAI 2005: Advances in Artificial Intelligence. XXVI, 1198 pages. 2005.

Vol. 3782: K.-D. Althoff, A. Dengel, R. Bergmann, M. Nick, T.R. Roth-Berghofer (Eds.), Professional Knowledge Management. XXIII, 739 pages. 2005.

Vol. 3763: H. Hong, D. Wang (Eds.), Automated Deduction in Geometry. X, 213 pages. 2006.

Vol. 3755: G.J. Williams, S.J. Simoff (Eds.), Data Mining. XI, 331 pages. 2006.

Vol. 3735: A. Hoffmann, H. Motoda, T. Scheffer (Eds.), Discovery Science. XVI, 400 pages. 2005.

Vol. 3734: S. Jain, H.U. Simon, E. Tomita (Eds.), Algorithmic Learning Theory. XII, 490 pages. 2005.

Vol. 3721: A.M. Jorge, L. Torgo, P.B. Brazdil, R. Camacho, J. Gama (Eds.), Knowledge Discovery in Databases: PKDD 2005. XXIII, 719 pages. 2005.

Vol. 3720: J. Gama, R. Camacho, P.B. Brazdil, A.M. Jorge, L. Torgo (Eds.), Machine Learning: ECML 2005. XXIII, 769 pages. 2005.

Vol. 3717: B. Gramlich (Ed.), Frontiers of Combining Systems. X, 321 pages. 2005.

Vol. 3702: B. Beckert (Ed.), Automated Reasoning with Analytic Tableaux and Related Methods. XIII, 343 pages. 2005.

Vol. 3698: U. Furbach (Ed.), KI 2005: Advances in Artificial Intelligence. XIII, 409 pages. 2005.

Vol. 3690: M. Pěchouček, P. Petta, L.Z. Varga (Eds.), Multi-Agent Systems and Applications IV. XVII, 667 pages. 2005.

Vol. 3684: R. Khosla, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part IV. LXXIX, 933 pages. 2005.

Vol. 3683: R. Khosla, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. LXXX, 1397 pages. 2005.

Vol. 3682: R. Khosla, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. LXXIX, 1371 pages. 2005.

Vol. 3681: R. Khosla, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LXXX, 1319 pages. 2005.

Vol. 3673: S. Bandini, S. Manzoni (Eds.), AI*IA 2005: Advances in Artificial Intelligence. XIV, 614 pages. 2005.

Vol. 3662: C. Baral, G. Greco, N. Leone, G. Terracina (Eds.), Logic Programming and Nonmonotonic Reasoning. XIII, 454 pages. 2005.

Vol. 3661: T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, T. Rist (Eds.), Intelligent Virtual Agents. XIII, 506 pages. 2005.

Vol. 3658: V. Matoušek, P. Mautner, T. Pavelka (Eds.), Text, Speech and Dialogue. XV, 460 pages. 2005.

¥718.00元

Preface

The research papers in this volume comprise the proceedings of FinTAL 2006, a Natural Language Processing conference continuing the TAL series of events: FracTAL 1997 at Université de Franche-Comté in Besançon, France; VexTAL 1999 at Università Ca' Foscari di Venezia in Venice, Italy; PorTAL 2002 at Universidade do Algarve in Faro, Portugal; and EsTAL 2004 at Universitat d'Alacant in Alicante, Spain. The main goals of the TAL conferences have been to bring together the international NLP community, to strengthen local NLP research, and to provide a forum for discussion of new NLP research and applications.

FinTAL 2006, organized by Turku Centre for Computer Science (TUCS) in Turku, Finland, also contributed to the goals mentioned above, further increasing the high international standing of the TAL conference series. We called for submissions both from academia and industry on any topic that is of interest to the NLP community, particularly encouraging research emphasizing multidisciplinary aspects of NLP and the interplay between linguistics, computer science and application domains such as biomedicine, communication systems, public services, and educational technology.

As a response, we received as many as 150 submissions from 38 countries in Europe, Asia, Africa, and the Americas. The manuscripts were reviewed by three members of FinTAL Program Committee, composed of researchers in the field, or external reviewers designated by the PC members. The PC members as well as the external reviewers are gratefully acknowledged in the following pages for their valuable contribution.

We would like to express here our gratitude to Turku Centre for Computer Science (TUCS), University of Turku, and Åbo Akademi University, as well as to our sponsors the city of Turku, Nokia, Lingsoft, PARC, and Sanako. We also thank our keynote speakers, the highly esteemed scholars Fred Karlsson, Lauri Karttunen, and Igor Mel'čuk. Last but obviously not least, we would like to thank all the individuals involved in organizing the event; without you it would not have been possible at all.

The careful evaluation of the submissions finally led to the selection of 72 papers to be presented at the conference and published in this volume. It is our firm belief that the accepted papers provide a significant contribution to the advance of science and technology. We hope that you will enjoy reading the articles and find them inspiring for your work, whether in basic NLP research or in the development of human language technology applications.

Turku, Finland, June 2006

Tapio Salakoski

Organization

FinTAL 2006 was organized by Turku Centre for Computer Science (TUCS) in conjunction with the University of Turku and Åbo Akademi University.

Program Committee

Tapio Salakoski	University of Turku, Finland (Chair)
Olli Aaltonen	University of Turku, Finland
Walid El Abed	Nestlé Corp., Switzerland
Jorge Baptista	University of Faro, Portugal
Florence Beaujard	Airbus Corp., France
Krzysztof Bogacki	University of Warsaw, Poland
Caroline Brun	Xerox Corp., France
Sylviane Cardey	University of Franche-Comte, France
Nigel Collier	National Institute of Informatics, Japan
Walter Daelemans	University of Antwerp, Belgium
Rodolfo Delmonte	University of Venice, Italy
Pasi Fränti	University of Joensuu, Finland
Peter Greenfield	University of Franche-Comté, France
Jari Kangas	Nokia Research Center, Finland
Kimmo Koskenniemi	University of Helsinki, Finland
Kyoko Kuroda	Shimane College, Japan
Hsiang-I Lin	National Taiwan University (NTU), Taiwan
Nuno Mamede	University of Lisbon, Portugal
Patricio Martínez-Barco	University of Alicante, Spain
Einar Meister	Tallinn University of Technology, Estonia
Rada Mihalcea	University of North Texas, USA
Leonel Ruiz Miyares	University of Santiago de Cuba, Cuba
Adeline Nazarenko	University Paris-Nord, France
Elisabete Ranchhod	University of Lisbon, Portugal
Karl-Michael Schneider	Textkernel BV, Amsterdam, The Netherlands
Rolf Schwitter	Macquarie University, Australia
John Tait	University of Sunderland, UK
José Luis Vicedo	University of Alicante, Spain
Simo Vihjanen	Lingsoft Ltd., Finland
Roman Yangarber	University of Helsinki, Finland

Local Organizers

Filip Ginter
Sampo Pyysalo
Hanna Suominen
Tapio Pahikkala
Tomi 'bgt' Mäntylä
Irmeli Laine
Christel Donner

Reviewers

Adeline Nazarenko
Agnes Sandor
Amanda Bouffier
Annu Paganus
Anssi Yli-Jyrä
Antoine Doucet
Antoine Rozenknop
Armando Suárez-Cueto
Borja Navarro
Caroline Brun
Caroline Hagege
Christopher Stokoe
Cristina Mota
David Martins de Matos
David Tomás
Davy Weissenbacher
Diamantino Caseiro
Duygu Can
Einar Meister
Elisabete Ranchhod
Evgeni Tsivtsivadze
Evgenia Chernenko
Felipe Sánchez-Martínez
Filip Ginter
Florence Beaujard
Graham Wilcock
Guillaume Bouchard
Guillaume Jacquet
Hanna Suominen
Hervé Déjean
Hsiang-I Lin
Hugo Meinedo
Ismo Kärkkäinen
Jari Kangas
Jean-Sébastien Tisserand
Joana L. Paulo
Joao Cabral
John Tait
Jorge Baptista
Jorma Boberg
José Luis Vicedo
Jouni Järvinen
Juhani Saastamoinen
Jussi Hakokari
Jussi Piitulainen
Jyri Paakkulainen
Karl-Michael Schneider
Kimmo Koskenniemi
Krister Lindén
Krzysztof Bogacki
Kyoko Kuroda
Leonel Ruiz Miyares
Luisa Coheur
Manuel Célio Conceição
Marc Dymetman
Marketta Hiissa
Maud Ehrmann
Michael Oakes
Navid Atar Sharghi
Nigel Collier
Nuno Mamede
Olli Aaltonen
Paloma Moreda
Pasi Fränti
Patricio Martínez-Barco
Paula Carvalho
Peter Greenfield
Rada Mihalcea
Rafael Carrasco
Rafael Muñoz
Ricardo Ribeiro
Rodolfo Delmonte
Rolf Schwitter
Roman Yangarber
Sampo Pyysalo
Shao Fen Liang
Simo Vihjanen
Siobhan Devlin
Sylviane Cardey
Séverine Vienney
Tanja Kavander
Tapio Pahikkala
Tapio Salakoski
Thierry Poibeau
Timo Honkela
Timo Knuutila
Tomi 'bgt' Mäntylä
Tony Mullen
Tuomo Saarni
Ville Hautamäki
Walid El Abed
Walter Daelemans
Xiaohong Wu

Conference Sponsors

Table of Contents

Keynote Addresses

Research Papers

Recursion in Natural Languages

Fred Karlsson

Department of General Linguistics
P.O. Box 9, FI-00014 University of Helsinki, Finland
fgk@ling.helsinki.fi

Abstract. The received view is that there are no grammatical constraints on clausal embedding complexity in sentences in languages of the 'Standard Average European' (SAE) type like English, Finnish, and Russian. The foremost proponent of this thesis is Noam Chomsky. This hypothesis of unbounded clausal embedding complexity is closely related to the hypothesis of unbounded syntactic recursion.

Psycholinguistic experimentation in the 1960's established that there are clear performance-related preferences especially regarding center-embedding. The acceptability of repeated center-embeddings (nesting) below depth 1 steeply decreases with each successive level of embedding.

Not much corpus-based work has been done to find out what the empirical 'facts' of clausal embedding complexity are. I have conducted extensive corpus studies of English, Finnish, German, Latin, and Swedish, with the aim of determining the most complex clausal embedding patterns actually used. The basic constraint on nested center-embedding in written language turns out to be two (with a marginal cline to three), in spoken language one. There are further specific restrictions on which types of clauses may be nested. The practical limit of final embedding (right-branching) is five. Repeated initial embedding (left-branching) of clauses below depth two is not possible.

These written language constraints were reached already in Sumerian, Akkadian, and Latin along with the advent of written language and have remained the same ever since.

The constraints on center-embedding imply that SAE syntax is finite-state, type 3 in the Chomsky hierarchy. Clause-level recursion is thus not unbounded. The special case of right-branching relative clauses is rather an instance of depth-preserving iteration.

T. Salakoski et al. (Eds.): FinTAL 2006, LNAI 4139, p. 1, 2006.