The workshop is sponsored by:

ACL/SIGPARSE
Institute of Computational Linguistics of Peking University
Intel Corporation (China)
Organization

General Chair
Harry Bunt, ITK, Tilburg University, The Netherlands

Program Chair
Giorgio Satta, DEI, University of Padua, Italy

Local Organizing Committee
Fuliang Weng (Co-Chair), Robert Bosch Corp., USA
Shiwen Yu (Co-Chair), Peking University, Beijing, China
Youqi Cao, Chinese Academy of Sciences, Beijing, China
Qunxiu Chen, Tsinghua University, Beijing, China
Houfeng Wang (Secretariat), Peking University, Beijing, China

Program Committee
Shuo Bai, Ministry of Information Industry, China
Eric Brill, Microsoft Research, USA
Harry Bunt, Tilburg University, The Netherlands
Bob Carpenter, SpeechWorks International Inc., USA
John Carroll, University of Sussex, UK
Ken Church, AT&T Labs - Research, USA
Mark Johnson, Brown University, USA
Aravind K. Joshi, University of Pennsylvania, USA
Ronald Kaplan, Xerox PARC, USA
Martin Kay, Xerox PARC, USA
Sadao Kurohashi, Kyoto University, Japan
Alon Lavie, Carnegie-Mellon University, USA
Yuji Matsumoto, Nara Institute of Science and Technology, Japan
Paola Merlo, University of Geneva, Switzerland
Mark-Jan Nederhof, University of Groningen, The Netherlands, and DFKI GmbH, Germany
Anton Nijholt, University of Twente, The Netherlands
Christer Samuelsson, Inzigo, Canada
Satoshi Sekine, New York University, USA
Virach Sornlertlamvanich, NECTEC, Thailand
Mark Steedman, University of Edinburgh, UK
Oliviero Stock, ITC-IRST, Italy
Hozumi Tanaka, Tokyo Institute of Technology, Japan
Masaru Tomita, Keio University, Japan
Hans Uszkoreit, DFKI GmbH and Universitat des Saarlandes, Germany
Eric Villemonte de la Clergerie, INRIA, France
Vijay-Shanker, University of Delaware, USA
David Weir, University of Sussex, UK
Mats Wiren, Telia Research, Sweden
Dekai Wu, Hong Kong University of Science and Technology, China
Tiejun Zhao, Harbin Institute of Technology, China

**Additional Refereeing**
Many thanks to Peter Asveld (University of Twente, The Netherlands), Ciprian Chelba (Microsoft Research, USA), David Chiang (University of Pennsylvania, USA), Mark Dras (IRCS, USA), Alberto Lavelli (ITC-IRST, Italy), Rudi Lutz (University of Sussex, UK), Yao Meng (Harbin Institute of Technology, China), Rieks op den Akker (University of Twente, The Netherlands), Anoop Sarkar (University of Pennsylvania, USA) and Gertjan van Noord (University of Groningen, The Netherlands).
Preface

IWPT’01 is the seventh in a series of workshops which have become the major occasion where specialists in parsing meet to present and discuss the advances in the field. The series started in 1989 when Masaru Tomita organized the International Workshop on Parsing Technology at Carnegie-Mellon University, with a second part in a conference resort. This workshop was very successful, and resulted in the book Current Issues in Parsing Technologies (Kluwer, Boston 1991).

Inspired by this success Masaru Tomita organized the Second International Workshop on Parsing Technology, IWPT’91, in Cancun, Mexico. By then, it seemed clear that these workshops could become an attractive biennial event for the parsing community, and Masaru Tomita and I together organized the third IWPT in 1993 at Tilburg University, the Netherlands, with again a second part in a conference resort (in the Belgian Ardennes mountains). IWPT’93 lead to the follow-up volume Recent Advances in Parsing Technology, Harry Bunt and Masaru Tomita, eds., published by Kluwer in 1996.

In 1994, the Special Interest Group on Parsing (SIGPARSE) was set up within the Association for Computational Linguistics (ACL), with the primary aim to provide continuity to the IWPT series. Under the auspices of SIGPARSE, IWPT’95 took place in the Czech Republic, IWPT’97 at MIT in Boston/Cambridge, and the ’99 workshop took place in February 2000 in Trento, Italy. IWPT’97 was also the basis of a third volume in the sequence of books on parsing: Advances in Probabilistic and Other Parsing Technologies, edited by Harry Bunt and Anton Nijholt, was published by Kluwer in 2000.

Since parsing, like language technology as a whole, more and more pursued at a global scale, it seemed right to move IWPT to a location in Asia in 2001. Thanks to the efforts of our colleagues in China, it has turned out to be possible to organize IWPT’01 in the beautiful setting of Beijing. I would like to thank Fuliang Weng and Shiwen Yu, co-chairs of the IWPT’01 Local Organization Committee, and Youqi Cao, Qunxiu Chen and Houfeng Wang, the members of this committee, for their efforts in organizing IWPT’01 in Beijing. Thanks are also due to the members of the Program Committee, for their careful reviewing of submitted papers. I would like to say a particular word of thanks to two people: Program Chair Giorgio Satta and former Program Chair John Carroll. Giorgio Satta did much more than chairing the Program Committee and managing the review process. He very actively participated in thinking about all aspects of making IWPT’01 a good conference, and he did so during the summer where he was also a program chair of the ACL’01 conference. John Caroll gave us advice on many issues concerning the reviewing, invited speakers, the workshop program, and any problem that came up during the preparation of IWPT’01. Giorgio and John, thank you very much; IWPT’01 owes a great deal to you.

Harry Bunt
SIGPARSE Officer and IWPT’01 General Chair
Program

October 17

08:30-09:30  Registration
09:30-09:40  Opening
09:40-10:40  Invited Talk: The XTAG Project at Penn
             Aravind K. Joshi
10:40-11:10  Coffee/Tea
11:10-11:50  On Compilation of the Quick-Check Filter for Feature Structure Unification
             Liviu Ciortuz
11:50-12:30  A Novel Probabilistic Model for Link Unification Grammar
             Fuliang Weng, Naiyong Jin, Jie Meng, Yujia Zhu
12:30-14:00  Lunch
14:00-14:40  Grammar Induction by MDL-Based Distributional Classification
             Yikun Guo, Fuliang Weng, Lide Wu
14:40-15:20  Robust Data Oriented Parsing of Speech Utterances
             Khalil Sima’an
15:20-16:00  Automatic Detection of Prosody Phrase Boundaries for Text-to-Speech System
             Xin Lv, Tie-jun Zhao, Zhan-yi Liu, Mu-yun Yang
16:00-16:30  Coffee/Tea
16:30-17:00  Poster Session
17:00-18:00  Poster Presentations
             A Multi-Input Dependency Parser
             Salah Aït-Mokhtar, Jean-Pierre Chanod, Claude Roux
             Property Grammars: A Flexible Constraint-Based Approach to Parsing
             Philippe Blache, Jean-Marie Balfourier
             Automatic Grammar Partitioning for Syntactic Parsing
             Po Chui Luk, Fuliang Weng, Helen Meng
             On Specialised Compilation of Rules in Unification Grammars
             Liviu Ciortuz
             An Approach to Parsing Vietnamese Noun Compounds
             Dinh Dien, Hoang Kiem
             The Implementation Process of a Statistical Parser for Brazilian Portuguese
             Andréia Gentil Bonfante, Maria das Graças Volpe Nunes
             Efficient Sentence Parsing with Language Specific Features: A Case Study of Czech
             Ales Horšk, Pavel Smrž
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
<th>Speaker(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:30-10:10</td>
<td>Parsing and Hypergraphs</td>
<td>Dan Klein, Christopher D. Manning</td>
</tr>
<tr>
<td>10:10-10:40</td>
<td>Coffee/Tea</td>
<td></td>
</tr>
<tr>
<td>10:40-11:20</td>
<td>High Precision Extraction of Grammatical Relations</td>
<td>John Carroll, Ted Briscoe</td>
</tr>
<tr>
<td>11:20-12:00</td>
<td>A Minimalist Implementation of Verb Subcategorization</td>
<td>Sourabh Niyogi</td>
</tr>
<tr>
<td>12:00-12:40</td>
<td>Parsing the CHILDES Database: Methodology and Lessons Learned</td>
<td>Kenji Sagae, Alon Lavie, Brian MacWhinney</td>
</tr>
<tr>
<td>12:40-14:00</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>14:00-18:30</td>
<td>Excursion</td>
<td></td>
</tr>
<tr>
<td>19:00-21:00</td>
<td>Banquet</td>
<td></td>
</tr>
<tr>
<td><strong>October 19</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09:00-10:00</td>
<td>Invited Talk: <em>Issues in Extracting Information from the Web</em></td>
<td>William W. Cohen</td>
</tr>
<tr>
<td>10:00-10:40</td>
<td>Unsupervised POS-Tagging Improves Parsing Accuracy and Parsing Efficiency</td>
<td>Robbert Prins, Gertjan van Noord</td>
</tr>
<tr>
<td>10:40-11:10</td>
<td>Coffee/Tea</td>
<td></td>
</tr>
<tr>
<td>11:10-11:50</td>
<td>Generalized ε-Skip Discriminating-Reverse Parsing on Graph-Structured Stack</td>
<td>José Fortes Gálvez, Jacques Farré</td>
</tr>
<tr>
<td>11:50-12:30</td>
<td>On the Complexity of Some Extensions of RCG Parsing</td>
<td>Eberhard Bertsch, Mark-Jan Nederhof</td>
</tr>
<tr>
<td>12:30-14:00</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>14:00-14:40</td>
<td>Bidirectional Automata for Tree Adjoining Grammars</td>
<td>Miguel A. Alonso, Víctor J. Díaz, Manuel Vilares</td>
</tr>
<tr>
<td>14:40-15:20</td>
<td>Bidirectional Ascendant Parsing for Natural Language Processing</td>
<td>Ştefan Andrei</td>
</tr>
<tr>
<td>15:20-16:00</td>
<td>Probabilistic Modelling of Island-Driven Parsing</td>
<td>Alicia Ageno, Horacio Rodríguez</td>
</tr>
<tr>
<td>16:00-16:30</td>
<td>Coffee/Tea</td>
<td></td>
</tr>
</tbody>
</table>
16:30-17:00 Poster Session

17:00-18:00 Poster Presentations

Efficient Incremental Dependency Parsing
Yoshihide Kato, Shigeki Matsubara, Katsuhiko Toyama, Yasuyoshi Inagaki

Direct Bottom-Up Chart Parsing Using Abbreviated Context-Free Grammars
Guido Minnen, William Thompson, Harry Bliss, Dale Russell

Word-Order Relaxations & Restrictions within a Dependency Grammar
Martin Plátek, Tomáš Holan, Vladislav Kuboň, Karel Oliva

Inside-Outside Estimation Meets Dynamic EM
Detlef Prescher

Knowledge Acquisition from a Text by a Linguistic and Statistical Method
Jean-David Sta, Yun-Chuang Chiao

Buffered Shift-Reduce Parsing
Bing Swen

How Much Will a RE-Based Preprocessor Help a Statistical Parser?
Daniel Zeman

18:00-18:10 Closing
Contents

Organization iii
Preface v
Program vii
Contents xi

Invited Talks

Issues in Extracting Information from the Web
William W. Cohen ................................................................. 3

Parameter Estimation for Statistical Parsing Models: Theory and Practice of
Distribution-Free Methods
Michael Collins .............................................................. 4

The XTAG Project at Penn
Aravind K. Joshi ............................................................... 16

Papers

Probabilistic Modelling of Island-Driven Parsing
Alicia Ageno and Horacio Rodríguez ........................................ 31

Bidirectional Automata for Tree Adjoining Grammars
Miguel A. Alonso, Victor J. Díaz and Manuel Vilares .................. 42

Bidirectional Ascendant Parsing for Natural Language Processing
Ștefan Andrei .............................................................. 54

On the Complexity of Some Extensions of RCG Parsing
Eberhard Bertsch and Mark-Jan Nederhof ................................. 66

High Precision Extraction of Grammatical Relations
John Carroll and Ted Briscoe ............................................... 78

On Compilation of the Quick-Check Filter for Feature Structure Unification
Liviu Ciortuz ................................................................. 90

Generalized ε-Skip Discriminating-Reverse Parsing on Graph-Structured Stack
José Fortes Gálvez and Jacques Farré .................................. 101

Grammar Induction by MDL-Based Distributional Classification
Yikun Guo, Fuliang Weng and Lide Wu ................................. 112

Parsing and Hypergraphs
Dan Klein and Christopher D. Manning .................................. 123

Automatic Detection of Prosody Phrase Boundaries for Text-to-Speech System
Xin Lv, Tie-jun Zhao, Zhan-yi Liu and Mu-yun Yang .................. 135

A Minimalist Implementation of Verb Subcategorization
Sourabh Niyogi .............................................................. 142

Unsupervised POS-Tagging Improves Parsing Accuracy and Parsing Efficiency
Robbert Prins and Gertjan van Noord ................................. 154
Parsing the CHILDES Database: Methodology and Lessons Learned
Kenji Sagae, Alon Lavie and Brian MacWhinney .................................................. 166
Robust Data Oriented Parsing of Speech Utterances
Khalil Sima’an ................................................................. 177
A Novel Probabilistic Model for Link Unification Grammar
Fuliang Weng, Naiyong Jin, Jie Meng and Yujia Zhu........................................... 189

Posters
A Multi-Input Dependency Parser
Salah Aït-Mokhtar, Jean-Pierre Chanod and Claude Roux .................................. 201
Property Grammars: A Flexible Constraint-Based Approach to Parsing
Philippe Blache and Jean-Marie Balfourier ................................................... 205
On Specialised Compilation of Rules in Unification Grammars
Liviu Ciortuz ........................................................................... 209
An Approach to Parsing Vietnamese Noun Compounds
Dinh Dien and Hoang Kiem ............................................................................. 213
The Implementation Process of a Statistical Parser for Brazilian Portuguese
Andréia Gentil Bonfante and Maria das Graças Volpe Nunes ............................... 217
Efficient Sentence Parsing with Language Specific Features: A Case Study of Czech
Aleš Horák and Pavel Smrž ............................................................................. 221
Efficient Incremental Dependency Parsing
Yoshihide Kato, Shigeki Matsubara, Katsuhiko Toyama and Yasuyoshi Inagaki .... 225
Automatic Grammar Partitioning for Syntactic Parsing
Po Chui Luk, Fuliang Weng and Helen Meng ................................................... 229
Direct Bottom-Up Chart Parsing Using Abbreviated Context-Free Grammars
Guido Minnen, William Thompson, Harry Bliss and Dale Russell .................... 233
Word-Order Relaxations & Restrictions within a Dependency Grammar
Martin Plátek, Tomáš Holan, Vladislav Kuboň and Karel Oliva ............................ 237
Inside-Outside Estimation Meets Dynamic EM
Detlef Prescher .................................................................................. 241
Knowledge Acquisition from a Text by a Linguistic and Statistical Method
Jean-David Sta and Yun-Chuang Chiao .................................................... 245
Buffered Shift-Reduce Parsing
Bing Swen .................................................................................. 249
How Much Will a RE-Based Preprocessor Help a Statistical Parser?
Daniel Zeman .................................................................................. 253

Author Index .................................................................................. 257