

Proceedings of The  
13th International Conference on Parsing Technologies  
**IWPT-2013**

November 27–29, 2013

Nara, Japan



# Contents

<b>Front matter</b> . . . . .	<b>iii</b>
<b>Discontinuous Parsing with an Efficient and Accurate DOP Model</b> . . . . . <i>Andreas van Cranenburgh and Rens Bod</i>	<b>7</b>
<b>An Efficient Typed Feature Structure Index: Theory and Implementation</b> . . . . . <i>Bernd Kiefer and Hans-Ulrich Krieger</i>	<b>17</b>
<b>Unsupervised Learning of Bilingual Categories in Inversion Transduction Grammar Induction</b> . . . . . <i>Markus Saers and Dekai Wu</i>	<b>26</b>
<b>Comparative Evaluation of Argument Extraction Algorithms in Discourse Relation Parsing</b> . . . . . <i>Evgeny Stepanov and Giuseppe Riccardi</i>	<b>36</b>
<b>Improved Chinese Parsing Using Named Entity Cue</b> . . . . . <i>Dongchen Li, Xiantao Zhang and Xihong Wu</i>	<b>45</b>
<b>Improving a symbolic parser through partially supervised learning</b> . . . . . <i>Eric De La Clergerie</i>	<b>54</b>
<b>On Different Approaches to Syntactic Analysis Into Bi-Lexical Dependencies. An Empirical Comparison of Direct, PCFG-Based, and HPSG-Based Parsers</b> . . . . . <i>Angelina Ivanova, Stephan Oepen, Rebecca Dridan, Dan Flickinger and Lilja vrelid</i>	<b>63</b>
<b>Generalization of Words for Chinese Dependency Parsing</b> . . . . . <i>Xianchao Wu, Jie Zhou, Yu Sun, Zhanyi Liu, Dianhai Yu, Hua Wu and Haifeng WANG</i>	<b>73</b>
<b>Dynamic-oracle Transition-based Parsing with Calibrated Probabilistic Output</b> . . . . . <i>Yoav Goldberg</i>	<b>82</b>
<b>Dependency Structure for Incremental Parsing of Japanese and Its Application</b> . . . . . <i>Tomohiro Ohno and Shigeki Matsubara</i>	<b>91</b>
<b>Active Learning for Dependency Parsing by A Committee of Parsers</b> . . . . . <i>Saeed Majidi and Gregory Crane</i>	<b>98</b>
<b>Development of Amharic Grammar Checker Using Morphological Features of Words and N-Gram Based Probabilistic Methods</b> . . . . .	<b>106</b>

*Aynadis Temesgen and Yaregal Assabie*

**LCFRS binarization and debinarization for directional parsing . . . . . 113**

*Wolfgang Maier*

**Towards Fully Lexicalized Dependency Parsing for Korean . . . . . 120**

*Jungyeul Park, Daisuke Kawahara, Sadao Kurohashi and Key-Sun Choi*

**Document Parsing: Towards Realistic Syntactic Analysis . . . . . 127**

*Rebecca Dridan and Stephan Oepen*

**Neo-Davidsonian Semantics in Lexicalized Grammars . . . . . 134**

*Petr Homola*

**Effective Parsing for Human Aided NLP Systems . . . . . 141**

*Naman Jain and Sambhav Jain*

## Preface

Welcome to the 13th International Conference on Parsing Technologies in the beautiful ancient city of Nara, Japan. This conference continues the tradition of biennial conferences organized by SIGPARSE, ACL's Special Interest Group on Parsing. The first International Workshop on Parsing Technologies (IWPT) took place in 1989 in Philadelphia, and was followed by successful biennial workshops in Cancun ('91); Tilburg ('93); Prague ('95); Boston ('97); Trento (2000); Beijing (2001); Nancy (2003); and Vancouver (2005), after which the name was changed from 'workshop' to 'conference', while retaining the abbreviation 'IWPT'. Subsequent IWPT conferences were held in Prague (2007), Paris (2009) and Dublin (2011). Over time, these conferences have developed more and more as the primary specialized forum for research on natural language parsing.

Based on contributions to IWPT workshops and conferences, five books on parsing have been published of which the latest one, based on IWPT 2007 and '09, was published by Springer in 2010 as *Trends in Parsing Technology*, edited by Harry Bunt, Paola Merlo and Joakim Nivre. Selected revised papers from IWPT 2011 have been accepted for publication in a special issue of the *Journal for Logic and Computation* which is scheduled to appear in April 2014; see <http://logcom.oxfordjournals.org/content/early/recent> for online pre-publication of these papers.

This year we received a total of 28 valid submissions, Of these, 9 were accepted as long papers and 8 as short papers. All accepted papers are published in these proceedings and presented at the conference either as a long talk or as a short talk and a poster

In addition to the contributed papers, IWPT 2013 as usual features invited talks on topics relevant to natural language parsing. This year we are delighted to welcome two distinguished Japanese invited speakers: Jun'ichi Tsujii (Microsoft Research) and Taro Watanabe (National Institute for Communication Technologies NiCT).

Organizing IWPT 2013 would not have been possible without the dedicated work of a number of people. We would like to thank the local organizing committee, chaired by Yuji Matsumoto and including Kevin Duh and Sorami Hisamoto, for an outstanding job in taking care of the local and practical organization of this conference. We also thank Kai Zhao for his excellent work in setting up and maintaining the IWPT 2013 website. We are extremely grateful to the members of the program committee, who spent precious time on reviewing submitted papers in the middle of the summer holiday season. Many thanks are also due to the sponsors whose support helped to make IWPT 2013 possible: NAIST - the Nara Institute of Science and Technology, Springer Publishers, and the Nara Visitors Bureau. Finally, we like to thank the directors of the NAIST, NiCT, and NTT Communication Sciences laboratories for the post-conference visit to these labs that the IWPT 2013 participants are all invited to.

Enjoy the conference and its setting!

Harry Bunt      Khalil Sima'an and Liang Huang  
General Chair   Program Chairs

**Organizers:**

General Chair Harry Bunt (Tilburg University, The Netherlands)

**Program Chairs**

Chair Khalil Simaan (University of Amsterdam, Netherlands)

Co-chair Liang Huang (City University of New York, USA)

**Local Organization Committee**

Chair Yuji Matsumoto (Nara Institute of Science and Technology, Japan)

Co-chair Kevin Duh (Nara Institute of Science and Technology, Japan)

Co-chair Sorami Hisamoto (Nara Institute of Science and Technology, Japan)

**Program Committee:**

Srinivas Bangalore (AT&T Labs-Research, USA)  
 Philippe Blache (CNRS/Provence University, France)  
 Harry Bunt (Tilburg University, Netherlands)  
 Aoife Cahill (Educational Testing Services, USA)  
 John Carroll (University of Sussex, UK)  
 Eugene Charniak (Brown University, USA)  
 David Chiang (USC Information Sciences Institute, USA)  
 Eric Villemonte de la Clergerie (INRIA, Rocquencourt, France)  
 Jason Eisner (Johns Hopkins University, USA)  
 Jennifer Foster (Dublin City University, Ireland)  
 Josef van Genabith (Dublin City University, Ireland)  
 Yoav Goldberg (Bar Ilan University, Israel)  
 Carlos Gomez-Rodriguez (University of La Corua, Spain)  
 Nizar Habash (Columbia University)  
 James Henderson (University of Geneva, Switzerland)  
 Julia Hockenmaier (University of Illinois at Urbana-Champaign, USA)  
 Zhongqiang Huang (BBN Technologies, USA)  
 Mark Johnson (Macquarie University, Australia)  
 Ronald Kaplan (Nuance Communications, USA)  
 Daisuke Kawahara (Kyoto University, Japan)  
 Martin Kay (Stanford University, USA)  
 Terry Koo (Google Inc, USA)  
 Sandra Kuebler (Indiana University, USA)  
 Marco Kuhlmann (Uppsala University, Sweden)  
 Jonas Kuhn (University of Stuttgart, Germany)  
 Sadao Kurohashi (Kyoto University, Japan)  
 Alon Lavie (Carnegie-Mellon University, USA)  
 Yuji Matsumoto (Nara Institute of Science and Technology, Japan)  
 David McClosky (IBM T. J. Watson Research Center, USA)  
 Ryan McDonald (Google Inc, USA)  
 Yusuke Miyao (University of Tokyo, Japan)  
 Mark-Jan Nederhof (University of St. Andrews, UK)  
 Joakim Nivre (University of Uppsala, Sweden)  
 Stephan Oepen (University of Oslo, Norway)

Gerald Penn (University of Toronto, Canada)  
Slav Petrov (Google Inc, USA)  
Brian Roark (Google Inc, USA)  
Kenji Sagae (Institute for Creative Technologies, USA)  
Djam Seddah (University Paris-Sorbonne)  
Vijay Shanker (University of Delaware, USA)  
Noah Smith (Carnegie Mellon University, USA)  
Mark Steedman (University of Edinburgh, UK)  
Ivan Titov (Saarland University, Germany)  
Reut Tsarfaty (Uppsala University, Sweden)  
Dekai Wu (Hong Kong University of Science and Technology, China)  
Hao Zhang (Google Inc, USA)  
Yi Zhang (Saarland University, Germany)  
Yue Zhang (Singapore University of Technology and Design, Singapore)

**Invited Speaker:**

Junichi Tsujii, Microsoft Research (MSR)  
Taro Watanabe, National Institute for Communication Technologies (NICT), Japan

## Invited Talks

### Semantic Processing – The Depth and Width

**Jun'ichi Tsujii**

Microsoft Research

#### **Abstract**

Accumulation of large amounts of structured or semi-structured facts and knowledge, such as FreeBase, Yago, Wikipedia, etc., will make semantics-based Natural Language Processing plausible in practical application settings. In this talk, I would like to address crucial problems involved in wide and deep semantic processing. While the ability of making inferences would be crucial, logically rigorous frameworks, which are required to solve problems intelligently, would not be what we need in Natural Language Processing for semantics-based information access. What we need to do first is to make linkages between textual expressions and structured facts/knowledge. We can extend word-centered concepts such as synonymy to more general concepts such as paraphrases of larger units of linguistic expressions such as phrases, clauses, etc. We discuss several on-going projects in this direction of semantic processing.

### Grammar Induction for Machine Translation

**Taro Watanabe**

National Institute for Communication Technologies

Kyoto, Japan

#### **Abstract**

Unsupervised methods for inducing bilingual correspondence are important components in machine translation. In this talk I would like to focus on automatically inducing grammatical information from bilingual data. Based on non-parametric Bayesian methods, first, an inversion transduction grammar is learnt through binary branching in two languages. The recursive splitting strategy is combined with the hierarchical Pitman-Yor process to memorize all the granularities of phrasal rules. Next, I would discuss part-of-speech tag induction given dependency trees in one language to improve the performance of machine translation. In particular the monolingual infinite tree model is extended to a bilingual scenario by emitting a source word with its aligned target words, either jointly or independently, from each hidden state of a source-side dependency tree. Finally, I would like to present some work in progress for machine translation.



