Introduction

This is the tenth meeting of the Pacific Association for Computational Linguistics. It is a pleasure to return to Melbourne, the site of the first such meeting. We hope that this meeting will help to bring together language technology workers from the entire pacific region.

This year, we received 96 submissions, 55 from Asia (40 from Japan), 23 from Australia and New Zealand, 12 from Europe and 6 from North America. After a thorough reviewing process, we selected 39 papers to be presented at the conference. There were so many high quality papers that we organized the program into two parallel sessions.

The papers selected range over a variety of topics, including Machine Translation, Parsing, Named Entity Extraction and Terminology. In keeping with the theme of “Deep processing: Abstracting away from the surface”, the two most common themes are Lexical Semantics (including WSD) and Question Answering.

We would like to thank all the authors for submitting their research and the members of the program committee for their careful reviews and useful suggestions to the authors. We would also like to thank the three invited speakers: Ann Copestake, Robert Dale, and Yuji Matsumoto.

In order to further the advancement of science by removing unnecessary legal and technical barriers to scientific collaboration and innovation all the papers have been released under the Creative Commons Attribution 3.0 unported license (http://creativecommons.org/licenses/by/3.0/). This allows others to copy, use, distribute, transmit and display the papers in these proceedings as well as to make and distribute derivative works, subject to proper attribution of authorship. It is our hope that these proceedings will thus become the raw material for further research.

We are grateful to the National Institute of Information and Communications Technology, Japan (NICT), National ICT Australia (NICTA), and the University of Melbourne for sponsoring PACLING-2007. In addition, the conference would not have been possible without the tireless efforts of the local chairs, Timothy Baldwin and David Martinez. Finally we would like to thank the PACLING President, Shun Ishizaki for his help in the overall organization.

Francis Bond
National Institute of Information and Communications Technology, Japan
PACLING 2007 Program Chair
Organizers

Program Chair:
Francis Bond (NICT, Japan)

Local Chairs:
Timothy Baldwin (University of Melbourne, Australia)
David Martinez (University of Melbourne, Australia)

Program Committee:
Emily Bender (University of Washington, USA)
Steven Bird (University of Melbourne, Australia)
Sandra Carberry (University of Delaware, USA)
Lawrence Cavedon (NICTA, Australia)
Nick Cercone (York University, Canada)
James Curran (University of Sydney, Australia)
Robert Dale (Macquarie University, Australia)
Rebecca Dridan (University of the Saarlands, Germany)
Kevin Duh (University of Washington, USA)
Tsutomu Endo (Kyushu Institute of Technology, Japan)
Laurel Fais (University of British Columbia, Canada)
Dan Flickinger (Stanford University, USA)
John Fry (San Jose State University, USA)
Sanae Fujita (NTT, Japan)
Chikara Hashimoto (Yamagata University, Japan)
Yoshihiko Hayashi (Osaka University, Japan)
Lars Hellan (NTNU, Norway)
Graeme Hirst (University of Toronto, Canada)
Young-Sook Hwang (ETRI, Korea)
Satoru Ikebara (Tottori University, Japan)
Kentaro Inui (Nara Institute of Science and Technology, Japan)
Pierre Isabelle (University of Quebec, Canada)
Shun Ishizaki (Keio University, Japan)
Daisuke Kawahara (NICT, Japan)
Jong-Bok Kim (Kyung Hee University, Korea)
Ali Knott (University of Otago, New Zealand)
Kiyoshi Kogure (ATR, Japan)
Rob Malouf (San Diego State University, USA)
Yuval Marom (GAPbuster Worldwide, Australia)
Hirosi Masuuchi (Fuji Xerox, Japan)
Robert Mercer (University of Western Ontario, Canada)
Adam Meyers (New York University, USA)
Diego Molla (Macquarie University, Australia)
Masaki Murata (NICT, Japan)
Akiyo Nadamoto (NICT, Japan)
Ryo Nagata (Hyogo University of Teacher Education, Japan)
Hiromi Nakaiwa (NTT Corp., Japan)
Tsuneko Nakazawa (Tokyo University, Japan)
Shigeko Nariyama (University of Melbourne, Australia)
Hwee-Tou Ng (National University of Singapore, Singapore)
Eric Nichols (NAIST, Japan)
Yoshihiko Nitta (Nihon University, Japan)
Stephan Oepen (University of Oslo, Norway)
Kentaro Ogura (NTT Software, Japan)
Cécile Paris (CSIRO, Australia)
David Powers (Flinders University, Australia)
Hiroshi Sakaki (Meisei University, Japan)
Ulrich Schaefer (DFKI, Germany)
Virach Sornlertlamvanich (NICT, Thailand)
Nicola Stokes (NICTA, Australia)
Thepchai Supnithi (NECTEC, Thailand)
Hisami Suzuki (Microsoft Research, USA)
Kumiko Tanaka-Ishii (University of Tokyo, Japan)
Takaaki Tanaka (NTT, Japan)
Enya-Kong Tan (Penang University, Malaysia)
Ben Waldron (NTNU, Norway)
Xiaohong Wu (NICT, Japan)
Kazuhide Yamamoto (Nagaoka University of Technology, Japan)
Jaehyung Yang (Kangnam University, Korea)
Yì Zhang (DFKI, Germany)
Yújie Zhang (NICT, Japan)

Invited Speakers:

Ann Copestake (University of Cambridge, UK)
Robert Dale (Macquarie University, Australia)
Yuji Matsumoto (NAIST, Japan)
**Table of Contents**

*Applying Robust Semantics*  
Ann Copestake ................................................................. 1

*Pinyomi: Dictionary lookup via orthographic associations*  
Lars Yencken, Zhihui Jin and Kumiko Tanaka-Ishii .......................... 13

*Graphical query for linguistic treebanks*  
Steven Bird and Haejoong Lee .................................................. 22

*Corpus Statistics Approaches to Discriminating Among Near-Synonyms*  
Mary Gardiner and Mark Dras .................................................... 31

*Detecting Compositionality of English Verb-Particle Constructions using Semantic Similarity*  
Su Nam Kim and Timothy Baldwin .............................................. 40

*Sentence and Token Splitting Based On Conditional Random Fields*  
Katrin Tomanek, Joachim Wermter and Udo Hahn ................................. 49

*Using Conditional Random Fields for Clause Splitting*  
Vinh Van Nguyen, Minh Le Nguyen and Akira Shimazu .......................... 58

*Incorporating Global Information into Supervised Learning for Chinese Word Segmentation*  
Hai Zhao and Chunyu Kit .......................................................... 66

*Combining Similarity-based Approaches for Semantic Tagging of Unknown Words*  
Fumiyo Fukumoto and Yoshimi Suzuki .......................................... 75

*Using Contexts of One Trillion Words for WSD*  
Tobias Hawker .............................................................................. 85

*Word Sense Disambiguation in Contextual Dynamic Network Using Associative Concept Dictionary*  
Jun Okamoto and Shun Ishizaki ................................................... 94

*Large-Scale Supervised Models for Noun Phrase Bracketing*  
David Vadas and James Curran .................................................... 104

*Empirical study on Parsing Chinese Based on Collins Model*  
Hailong Cao, Yujie Zhang and Hitoshi Isahara .................................. 113

*Learning-Based Argument Structure Analysis of Event-Nouns in Japanese*  
Mamoru Komachi, Ryu Iida, Kentaro Inui and Yuji Matsumoto ................... 120

*Interpreting Noun Compounds using Bootstrapping and Sense Collocation*  
Su Nam Kim and Timothy Baldwin ................................................ 129

*Semantic Analysis of Entity Contexts towards Open Named Entity Classification on the Web*  
Xuan-Hieu Phan, Le-Minh Nguyen, Cam-Tu Nguyen and Susumu Horiguchi .................. 137

*Recognizing Names by the Meaning of Defining Syntactic Structures*  
Li Yang and Steven Abney .......................................................... 145
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>A FrameNet Approach to Connectives: The Polysemy of While</td>
<td>Satoru Uchida and Seiko Fujii</td>
</tr>
<tr>
<td>Resolution of Coordination Ellipses in Biological Named Entities with Conditional Random Fields</td>
<td>Ekaterina Buyko, Katrin Tomanek and Udo Hahn</td>
</tr>
<tr>
<td>Minimising semantic drift with Mutual Exclusion Bootstrapping</td>
<td>James Curran, Tara Murphy and Bernhard Scholz</td>
</tr>
<tr>
<td>Analogical Mapping Method and Semantic Categorization of Japanese Compound and Complex Sentence Patterns</td>
<td>Satoru Ikehara, Masato Tokuhisa and Jin’ichi Murakami</td>
</tr>
<tr>
<td>Bridging the gap: Thai - Thai Sign Machine Translation</td>
<td>Srisavakon Dangsaart, Kanlaya Naruedomkul, Nick Cercone and Booncharoen Sirinaovakul</td>
</tr>
<tr>
<td>Preparing, Restructuring, and Augmenting a French Treebank: Lexicalised Parsers or Coherent Treebanks?</td>
<td>Natalie Schluter and Josef van Genabith</td>
</tr>
<tr>
<td>Improving the complement/adjunct distinction in CCGbank</td>
<td>Matthew Honnibal and James Curran</td>
</tr>
<tr>
<td>Evaluating Feature Types for Encoding Clinical Notes</td>
<td>Jon Patrick, Yitao Zhang and Yefeng Wang</td>
</tr>
<tr>
<td>Global Revision in Summarisation: Generating Novel Sentences with Prim’s Algorithm</td>
<td>Stephen Wan, Robert Dale, Mark Dras and Cécile Paris</td>
</tr>
<tr>
<td>Modelling and Exploring the Network Structure of Terminology by Potts Spin Glass Model</td>
<td>Kyo Kageura and Takeshi Abekawa</td>
</tr>
<tr>
<td>Determining the Unithood of Word Sequences: Mutual Information and Independence Measure</td>
<td>Wilson Wong, Wei Liu and Mohammed Bennamoun</td>
</tr>
<tr>
<td>A case study of comparison of several methods for corpus-based speech intention identification</td>
<td>Kazutaka Shimada, Kaoru Iwashita and Tsutomu Endo</td>
</tr>
<tr>
<td>Author Profiling for English Emails</td>
<td>Dominique Estival, Tanja Gaustad, Ben Hutchinson, Son Bao Pham and Will Radford</td>
</tr>
<tr>
<td>Towards a Semantic Analysis of Bahasa Indonesia for Question Answering</td>
<td>Septina Dian Larasati and Ruli Manurung</td>
</tr>
<tr>
<td>Query Recommendation from Search Engine Query Logs Using Query Enrichment</td>
<td>Shisanu Tongchim, Virach Sornlertlamvanich and Hitoshi Isahara</td>
</tr>
<tr>
<td>Automagically Inferring the Source Words of Lexical Blends</td>
<td>Paul Cook and Suzanne Stevenson</td>
</tr>
<tr>
<td>Discovering Word Senses for Polysemous Words Using Feature Domain Similarity</td>
<td>Noriko Tomuro, Kyoko Kanzaki and Hitoshi Isahara</td>
</tr>
</tbody>
</table>
Building a clustered semantic network for an Entire Large Dictionary of Japanese
Hiroyuki Akama, Jaeyoung Jung, Maki Miyake, Kenji Yosihasi and Kikuko Nishina ........... 308

A Named Entity Recogniser for Question Answering
Menno van Zaanen and Diego Mollá ................................................................. 317

Effect of combining different Web search engines on Web question-answering
Tatsunori Mori, Akira Kanai, Madoka Ishioroshi and Mitsuru Sato ......................... 325

What to Classify and How: Experiments in question classification for Japanese
Rebecca Dridan and Timothy Baldwin .............................................................. 333

Sentence retrieval for extracting biomedical knowledge
Tara McIntosh and James Curran ................................................................. 342

Centering in Chinese Anaphor Resolution: A Parametric Study
Manjuan Duan ............................................................................................... 350
Conference Program

Wednesday, September 19, 2007

9:45–10:00 Opening Remarks (Th 3)

Session 1a: Querying (Th 3)  Session 1b: Lexical Semantics (Th 2.06)

10:00–10:30 Pinyin: Dictionary lookup via orthographic associations
Lars Yencken, Zhihui Jin and Kumiko Tanaka-Ishii  Detecting Compositionality of English Verb-Particle Constructions using Semantic Similarity
Su Nam Kim and Timothy Baldwin

10:30–11:00 Corpus Statistics Approaches to Discriminating Among Near-Synonyms
Mary Gardiner and Mark Dras  Graphical query for linguistic treebanks
Steven Bird and Haejoong Lee

11:00–11:30 Coffee break

Invited Talk (Th 3)

11:30–12:45 The Generation of Referring Expressions: Where We’ve Been, How We Got Here, and Where We’re Going
Invited Talk by Robert Dale

12:45–14:00 Lunch

Session 2a: Segmentation (Th 3)  Session 2b: WSD (Th 2.06)

14:00–14:30 Sentence and Token Splitting Based On Conditional Random Fields
Katrin Tomanek, Joachim Wermter and Udo Hahn  Combining Similarity-based Approaches for Semantic Tagging of Unknown Words
Fumiyo Fukumoto and Yoshimi Suzuki

14:30–15:00 Using Conditional Random Fields for Clause Splitting
Vinh Van Nguyen, Minh Le Nguyen and Akira Shimazu  Using Contexts of One Trillion Words for WSD
Tobias Hawker

15:00–15:30 Incorporating Global Information into Supervised Learning for Chinese Word Segmentation
Hai Zhao and Chunyu Kit  Word Sense Disambiguation in Contextual Dynamic Network Using Associative Concept Dictionary
Jun Okamoto and Shun Ishizaki

15:30–16:00 Coffee break

Session 3a: Parsing (Th 3)  Session 3b: Lexical Semantics (Th 2.06)

16:00–16:30 Large-Scale Supervised Models for Noun Phrase Bracketing
David Vadas and James R. Curran  Learning-Based Argument Structure Analysis of Event-Nouns in Japanese
Mamoru Komachi, Ryu Iida, Kentaro Inui and Yuji Matsumoto

16:30–17:00 Empirical study on Parsing Chinese Based on Collins’ Model
Hailong Cao, Yujie Zhang and Hitoshi Ishahara  Interpreting Noun Compounds using Bootstrapping and Sense Collocation
Su Nam Kim and Timothy Baldwin

RECEPTION (17:30–19:00)
Thursday, September 20, 2007

10:00–10:30  **Session 4a: Named Entities (Th 3)**  
*Semantic Analysis of Entity Contexts towards Open Named Entity Classification on the Web*
Xuan-Hieu Phan, Le-Minh Nguyen, Cam-Tu Nguyen and Susumu Horiguchi

10:30–11:00  **Session 4b: Coordination/Conjunction (Th 2.06)**  
*A FrameNet Approach to Connectives: The Polysemy of While*
Satoru Uchida and Seiko Fujii

11:00–11:30  **Coffee break**

11:30–12:45  **Invited Talk (Th 3)**  
*Applying robust semantics*
Invited Talk by Ann Copestake

12:45–14:00  **Lunch**

14:00–14:30  **Session 5: Lexical Semantics (Th 3)**  
*Minimising semantic drift with Mutual Exclusion Bootstrapping*
James Curran, Tara Murphy and Bernhard Scholz

14:30–15:00  **Session 6a: Machine Translation (Th 3)**  
*Analogical Mapping Method and Semantic Categorization of Japanese Compound and Complex Sentence Patterns*
Satoru Ikehara, Masato Tokuhisa and Jin’ichi Murakami

15:00–15:30  **Session 6b: Parsing (Th 2.06)**  
*Preparing, Restructuring, and Augmenting a French Treebank: Lexicalised Parsers or Coherent Treebanks?*
Natalie Schluter and Josef van Genabith

15:30–16:00  **Coffee break**

16:00–16:30  **Session 7a: Terminology (Th 3)**  
*Modelling and Exploring the Network Structure of Terminology by Potts Spin Glass Model*
Kyo Kageura and Takeshi Abekawa

16:30–17:00  **Session 7b: Meta Information (Th 2.06)**  
*Determining the Unithood of Word Sequences: Mutual Information and Independence Measure*
Wilson Wong, Wei Liu and Mohammed Bennamoun

**Banquet (19:00–21:00)**
Session 8a: Meta Information (Th 3)
10:00–10:30  A case study of comparison of several methods for corpus-based speech intention identification
Kazutaka Shimada, Kaoru Iwashita and Tsutomu Endo

Session 8b: Question Answering (Th 2.06)
10:30–11:00  Towards a Semantic Analysis of Bahasa Indonesia for Question Answering
Septina Dian Larasati and Ruli Manurung

10:30–11:00  Author Profiling for English Emails
Dominique Estival, Tanja Gaustad, Ben Hutchinson, Son Bao Pham and Will Radford

11:00–11:30  Coffee break

Invited Talk (Th 3)
11:30–12:45  Word Construction and Multi-layered Sentence Analysis
Invited Talk by Yuji Matsumoto

Session 9a: Lexical Semantics (Th 3)
14:00–14:30  Automagically Inferring the Source Words of Lexical Blends
Paul Cook and Suzanne Stevenson

Session 9b: Question Answering (Th 2.06)
14:30–15:00  A Named Entity Recogniser for Question Answering
Menno van Zaanen and Diego Mollá

14:30–15:00  Discovering Word Senses for Polysemous Words Using Feature Domain Similarity
Noriko Tomuro, Kyoko Kanzaki and Hitoshi Isahara

15:00–15:30  Building a clustered semantic network for an Entire Large Dictionary of Japanese
Hiroyuki Akama, Jaeyoung Jung, Maki Miyake, Kenji Yoshihashi and Kikuko Nishina

15:30–16:00  Coffee break

Session 10a: Information Extraction (Th 3)
16:00–16:30  Sentence retrieval for extracting biomedical knowledge
Tara McIntosh and James Curran

Session 10b: Anaphor Resolution (Th 2.06)
16:30–17:00  Centering in Chinese Anaphor Resolution: A Parametric Study
Manjuan Duan

16:30–17:00  An Introduction to Research at NICT

17:00–17:30  An Introduction to Research at NICTA

17:30–17:45  Closing Remarks (Th 3)