COLING 2012:
Some facts and figures

Pushpak Bhattacharyya
CSE Dept.,
IIT Bombay
Organized by

- **CFILT**, CSE Department, IIT Bombay, and
- **TDIL**, DIT, Ministry of Communication and Information Technology
- Under the umbrella of ICCL (International Committee on Computational Linguistics)
Committees

- **Program Chairs:** Prof. Martin Kay, Stanford University and Prof. Christian Boitet, University Joseph Fourier, France
- **Organizing Chairs:** Prof. Pushpak Bhattacharyya, IIT Bombay and Prof. Rajeev Sangal, IIIT Hyderabad
- **Workshop Chair:** Prof. Laurent Besacier, University Joseph Fourier
- **Tutorial Chair:** Prof. Sadao Kurohashi, Kyoto University, Japan
- **Publication Chair:** Prof. Roger Evans, University of Brighton, UK
Date, Venue etc.

- Main Conference: 10-14 Dec, 2012
- Pre-conference workshops and tutorials: 8, 9 Dec, 2012
- Post-conference workshops: 15 Dec, 2012
- Venue: VMCC and convocation hall, IIT Bombay
- 1000+ submissions
- Conf URL: http://www.coling2012-iitb.org/
Our sponsors

- **Diamond**
  - TCS
  - LDC-IL, CIIL, Mysore

- **Gold**
  - Microsoft Research
  - Baidu

- **Silver**
  - Yahoo!
  - IBM
  - ezDi
  - enago
COLING2012 PARTICIPATION

Total Registration: 763
Total attendance: 612

<table>
<thead>
<tr>
<th></th>
<th>Main Conference</th>
<th>Tutorial</th>
<th>Workshop</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conference</td>
<td>390</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tutorial</td>
<td>19</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Workshop</td>
<td>133</td>
<td>64</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>606</td>
<td>99</td>
<td>336</td>
</tr>
</tbody>
</table>
Number of accepted papers in various categories (as in the proceedings)

- Oral (173)
- Poster (138)
- Demo (66)
- Reserve (22)
Best paper award

“Accurate Unbounded Dependency Recovery using Generalized Categorial Grammars”, Luan Nguyen¹, Marten Van Schijndel² and William Schuler²

1: University of Minnesota, Dept. of Computer Science and Engineering, Minneapolis, MN
2: The Ohio State University, Dept. of Linguistics, Columbus, OH

lnguyen@cs.umn.edu, vanschm@ling.osu.edu, schuler@ling.osu.edu
Workshops (1/3)

- **WS1**: “Advances in discourse analysis and its computational aspects”: Eva Hajicova
- **WS2**: “Second Workshop on Advances in Text Input Methods (WTIM 2)”: Kalika Bali, Monojit Choudhury, Yoh Okuno
- **WS3**: “Eye-tracking and Natural Language Processing”: Michael Carl, Pushpak Bhattacharya, Kamal Kumar Choudhary
- **WS4**: “3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP)”: Virach Sornlertlamvanich, Abbas Malik
- **WS5**: “Cognitive Aspects of the Lexicon (CogALex-III)”: Michael Zock, Reinhard Rapp
- **WS6**: “10th Workshop on Asian Language Resources”: Ruvan Weerasinghe, Rachel Edita O. Roxas, Virach Sornlertlamvanich, Sarmad Hussain
Workshops (2/3)

- WS7: “2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012)”: Sivaji Bandyopadhyay, Manabu Okumura
- WS8: “Sixth Workshop on Analytics for Noisy Unstructured Text Data”: Lipika Dey, Daniel Lopresti, Christoph Ringlstetter, Shourya Roy, L. Venkata Subramaniam
- WS9: “Information Extraction & Entity Analytics on Social Media Data”: Sriram Raghavan, Ganesh Ramakrishnan
- WS11: “Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT (ML4HMT-12 WS and Shared Task)”: Josef van Genabith, Toni Badia, Christian Federmann
Workshops (3/3)

- WS12: “Speech and Language Processing Tools in Education”: Radhika Mamidi, Kishore Prahallad
- WS14: “Question Answering for Complex Domains”: Nanda Kambhatla Sachindra Joshi, Ganesh Ramakrishnan, Kiran Kate, Priyanka Agrawal
Tutorials

- T1: “Temporal Information Extraction and Shallow Temporal Reasoning”: Dan Roth, Heng Ji, Taylor Cassidy, Quang Do
- T2: “Exploiting Web Data Sources for Advanced NLP”: Gerard de Melo
- T3: “Multimodal Corpora”: Patrizia Paggio, Dirk Heylen, Costanza Navarretta
- T5: “Open-domain Conversations with Humanoid Robots”: Kristiina Jokinen, Graham Wilcock
- T6: “Revisiting Dimensionality Reduction Techniques for NLP”: Jagadeesh Jagarlamudi, Raghavendra Udupa
Tutorial Participation

- Temporal Information Extraction and Shallow Temporal Reasoning: 25%
- Exploiting Web Data Sources for Advanced NLP: 29%
- Multimodal Corpora: 10%
- Revisiting Dimensionality Reduction Techniques for NLP: 15%
- Open-domain Conversations with Humanoid Robots: 12%
- The Hindi/Urdu Treebank: New Frontiers in Hindi and Urdu NLP: 9%
Workshop Participation

- Advances in discourse analysis and its computational aspects: 8%
- Second Workshop on Advances in Text Input Methods: 6%
- Eye-tracking and Natural Language Processing: 5%
- 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP): 8%
- Cognitive Aspects of the Lexicon (CogALex-III): 7%
- 10th Workshop on Asian Language Resources: 5%
- Second Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012): 8%
- Sixth Workshop on Analytics for Noisy Unstructured Text Data: 9%
- Information Extraction & Entity Analytics on Social Media Data: 10%
- Machine Translation and Parsing in Indian Languages (MTPIL-2012): 8%
- Speech and Language Processing Tools in Education: 6%
- Question Answering for Complex Domains: 6%
- Reordering for Statistical Machine Translation: 4%
- Second Workshop on Applying Machine Learning Techniques to Optimise the Division of Labour in Hybrid MT: 4%
- First International Workshop on Optimization Techniques for Human Language Technology: 6%
COLING 2012 Participation by Nationality

- Europe, 217
- USA+North America+ South America, 82
- Asia + Africa +AUS+Nz, 454

- India 251
- United States 70
- Japan 58
- China 53
- Germany 49
- France 36
- United Kingdom 24
- Ireland 17
- Australia 14
- Czech Republic 15
- Sweden 12
- Taiwan 12
- Hong Kong 11
- Canada 10
- Russia 9
- Spain 9
- Italy 8
- Korea, South 8
- Poland 6
- Singapore 6
- Denmark 5
- Hungary 5
- Croatia 4
- Portugal 4
- Vietnam 4
- Egypt 3
- Finland 3
- Iran 3
- Pakistan 3
- Others 41
Participation by Nationality

(in comparison) IJ CNLP2011

<table>
<thead>
<tr>
<th>Country</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>61</td>
</tr>
<tr>
<td>China</td>
<td>51</td>
</tr>
<tr>
<td>United States</td>
<td>35</td>
</tr>
<tr>
<td>India</td>
<td>28</td>
</tr>
<tr>
<td>France</td>
<td>17</td>
</tr>
<tr>
<td>Germany</td>
<td>14</td>
</tr>
<tr>
<td>Thailand</td>
<td>12</td>
</tr>
<tr>
<td>Taiwan</td>
<td>9</td>
</tr>
<tr>
<td>Singapore</td>
<td>8</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>8</td>
</tr>
<tr>
<td>Ireland</td>
<td>6</td>
</tr>
<tr>
<td>Australia</td>
<td>5</td>
</tr>
<tr>
<td>Canada</td>
<td>5</td>
</tr>
<tr>
<td>Italy</td>
<td>4</td>
</tr>
<tr>
<td>Korea</td>
<td>4</td>
</tr>
<tr>
<td>Phillipines</td>
<td>4</td>
</tr>
<tr>
<td>Netherlands</td>
<td>3</td>
</tr>
<tr>
<td>Spain</td>
<td>3</td>
</tr>
<tr>
<td>Iran</td>
<td>2</td>
</tr>
<tr>
<td>Pakistan</td>
<td>2</td>
</tr>
<tr>
<td>Sweden</td>
<td>2</td>
</tr>
<tr>
<td>Austria</td>
<td>1</td>
</tr>
<tr>
<td>Belgium</td>
<td>1</td>
</tr>
<tr>
<td>Brazil</td>
<td>1</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
</tr>
<tr>
<td>Ghana</td>
<td>1</td>
</tr>
<tr>
<td>Greece</td>
<td>1</td>
</tr>
<tr>
<td>Hungary</td>
<td>1</td>
</tr>
<tr>
<td>Latvia</td>
<td>1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>1</td>
</tr>
<tr>
<td>Morocco</td>
<td>1</td>
</tr>
<tr>
<td>New Zealand</td>
<td>1</td>
</tr>
<tr>
<td>Norway</td>
<td>1</td>
</tr>
<tr>
<td>Portugal</td>
<td>1</td>
</tr>
<tr>
<td>Qatar</td>
<td>1</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>1</td>
</tr>
<tr>
<td>Swaziland</td>
<td>1</td>
</tr>
<tr>
<td>Vietnam</td>
<td>1</td>
</tr>
</tbody>
</table>
Invited Talks

- “The adaptive brain: acquiring a complex cognitive skill in complex contexts ”: Prof. Barbara Moser-Mercer, University of Geneva (10th Dec, Day 1)
- “Digital Book, Digital Library, and Natural Language Processing”: Prof. Makato Nagao, Japan (11th Dec, Day 2)
- “Minimum Description Length as the basis of Panini's grammar”: Prof. Paul Kiparsky, Stanford University (13th Dec, Day 4)
- “NLP from Paninian Perspective”: Prof. Dipti Mishra Sharma, IIIT Hyderabad (14th Dec, Day 5)

(Day 3 was excursion day)
Social Events

- Reception: 10\textsuperscript{th} evening (performances by delegates and organizers)
- Banquet: 11\textsuperscript{th} evening: Renaissance
- Excursion: 12\textsuperscript{th} whole day: Karla and Bhaja caves
- Indian classical music: 13\textsuperscript{th} evening: Prasad Padhye (tabla) and Nayan Ghosh (Sitar)
Area wise representation of accepted papers

- Indian language technology: 5%
- Underresourced languages: 5%
- Morphology & POS tagging: 6%
- Grammar and formalisms: 2%
- Parsing: 4%
- Semantics: 5%
- Discourse and pragmatics: 4%
- Coreference resolution: 3%
- Ontologies and terminology: 3%
- Textual entailment: 1%
- Resources and annotation: 7%
- Psychological and neurological modelling: 2%
- Empirical machine translation: 6%
- Summarization: 5%
- Word sense disambiguation: 4%
- Speech recognition and synthesis: 2%
- Question answering: 3%
- Natural language generation: 3%
- Named entity recognition: 3%
- Information retrieval: 4%
- Information & content extraction: 5%
- Hybrid man+machine architectures: 1%
- Expert or hybrid machine translation: 3%
- Deployment of NLP-based applications, software integration & quality: 5%
- Textual entailment: 1%
- Expert or hybrid machine translation: 3%
Area and Countrywise statistics

Area wise statistics
Canada 20%
China 10%
Egypt 10%
Germany 10%
Hong Kong Special administrative region of China 10%
Hungary 10%
Poland 10%
Singapore 10%
United States 10%

Grammar and Formalisms
Discourse and pragmatics

- United States: 28%
- China: 17%
- France: 11%
- United Kingdom: 6%
- Hong Kong Special Administrative Region of China: 6%
- Japan: 6%
- Germany: 6%
- Australia: 5%
- Czech Republic: 5%
- Estonia: 5%
- Finland: 5%
- United States: 6%
- Hong Kong Special Administrative Region of China: 5%
- Japan: 6%
- United Kingdom: 6%
- Germany: 6%
- Australia: 5%
- Czech Republic: 5%
- Estonia: 5%
- Finland: 5%
Coreference resolution

- Belgium: 9%
- China: 8%
- Germany: 17%
- India: 17%
- Italy: 8%
- Japan: 8%
- Poland: 8%
- United States: 25%
China 25%
Japan 25%
United States 50%

Textual entailment
Psychological and Neurological Modelling

- United States: 50%
- Japan: 25%
- China: 13%
- Australia: 12%
Empirical Machine Translation

- China: 31%
- United States: 13%
- United Kingdom: 9%
- France: 9%
- Germany: 13%
- Iran (Islamic Republic of): 5%
- Spain: 4%
- Republic of Korea: 4%
- Qatar: 4%
- Macau: 4%
- Ireland: 4%
Deployment of NLP applications

- United States: 5%
- United Kingdom: 5%
- Taiwan: 15%
- French Federation: 5%
- Romania: 5%
- Republic of Korea: 10%
- Qatar: 5%
- Norway: 5%
- Japan: 15%
- Germany: 5%
- India: 5%
- China: 5%
- Japan: 15%
- Canada: 5%
- Norway: 5%
Expert or hybrid machine translation

- China: 22%
- France: 22%
- United States: 7%
- Belgium: 7%
- Denmark: 7%
- Greece: 7%
- Germany: 7%
- Japan: 14%
- Taiwan: 7%
Hybrid man+machine architectures & human factors

- Czech Republic: 34%
- Iran (Islamic Republic of): 33%
- France: 33%
Australia 10%
China 10%
Germany 15%
India 5%
Iran (Islamic Republic of) 10%
Ireland 5%
Italy 5%
Japan 5%
Romania 5%
Singapore 5%
Taiwan 5%
Turkey 5%
Information retrieval

- China: 28%
- India: 17%
- United States: 6%
- Spain: 6%
- Ireland: 11%
- Pakistan: 6%
- Japan: 5%
- Germany: 11%
- France: 5%
- Hong Kong: 5%
Pie chart showing named entity recognition with the following countries and percentages:

- United States: 9%
- Canada: 9%
- China: 9%
- France: 9%
- United Arab Emirates: 9%
- United Kingdom: 9%
- Turkey: 9%
- India: 9%
- Germany: 28%
United Kingdom 27%
France 28%
Germany 18%
Argentina 9%
Australia 9%
Canada 9%
Question answering

- China: 34%
- United States: 17%
- Japan: 17%
- United Kingdom: 8%
- Italy: 8%
- India: 8%
- Australia: 8%
Sentiment and text classification

- United States: 29%
- Germany: 18%
- India: 11%
- United Kingdom: 7%
- Singapore: 4%
- Russian Federation: 7%
- Portugal: 4%
- Japan: 4%
- Italy: 4%
- Ireland: 3%
- Canada: 3%
- China: 3%
- Denmark: 3%
Speech recognition and synthesis

- India: 14%
- Japan: 43%
- Republic of Korea: 14%
- United States: 29%
Canada 9%
China 48%
Greece 5%
India 9%
Japan 9%
Singapore 5%
Sweden 5%
United States 10%
Summarization
Message for students

- Clear goal
- Clear Technique
- Clear evaluation
- Clear error analysis and path to improve
Message (2/2)

- Guide us in straight path: holy Qur’an
- Lead us not unto temptation: holy Bible
- Asato ma sat gamaya: Upanishad
Thank you
Material given in the registration kit

- Bag
- Conference handbook
- USB containing
  - Main conference proceedings
  - Workshop proceedings
  - Tutorial notes
  - Help desk (internet, map, eating joints etc)
  - Excursion pamphlet
- Delegate badges with coupons
- Pen
- Pad
Observations

- High Quality papers, tutorials and workshops
- India’s presence: commendable
  - 219 out of 251 registered
  - 42 papers from India (main conf) out of 193 (Oral, Poster and Demo)