

Sixth International Joint Conference on  
Natural Language Processing



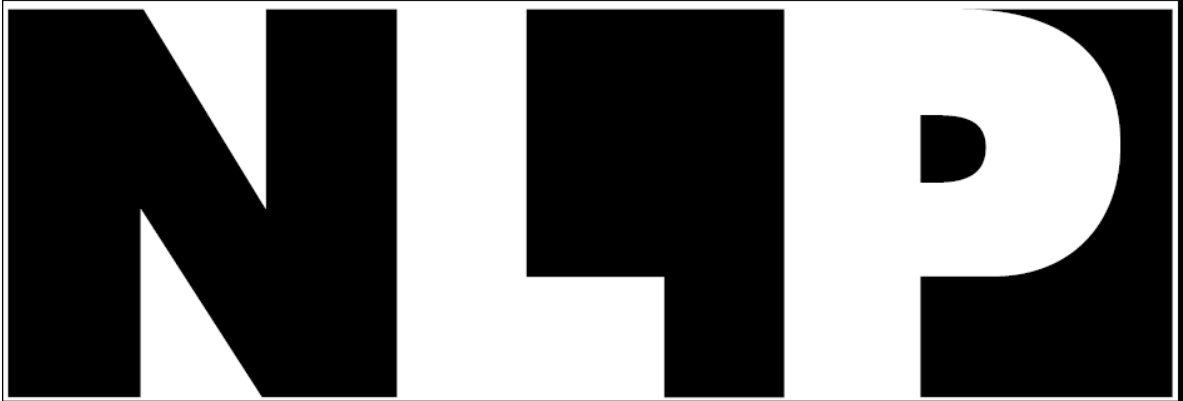
**Proceedings of the 11th Workshop on Asian Language  
Resources**



We wish to thank our sponsors and supporters!

Platinum Sponsors

---



[www.anlp.jp](http://www.anlp.jp)

Silver Sponsors

---



[www.google.com](http://www.google.com)

Bronze Sponsors

---



[www.rakuten.com](http://www.rakuten.com)

Supporters

---



**NAGOYA CONVENTION  
& VISITORS BUREAU**

[Nagoya Convention & Visitors Bureau](http://Nagoya Convention & Visitors Bureau)

We wish to thank our organizers!

Organizers

---



[Asian Federation of Natural Language Processing \(AFNLP\)](#)



[Toyohashi University of Technology](#)

©2013 Asian Federation of Natural Language Processing

ISBN978-4-9907348-4-8

## Preface

It is a pleasure for us to carry on with the mantle of Asian Language Resources Workshop which is in its 11th incarnation this year. The workshop is a satellite event of IJCNL 2013 being held at Nagoya, Japan, 14-18 October, 2013. These days, lexical resources form a critical component of NLP systems. Even though statistical, ML driven approaches are the ruling paradigm in many sub areas of NLP, the "accuracy plateau" or the saturation is often overcome only with the deployment of lexical resources.

In this year's ALR workshop, there were 15 submissions of which 10 were accepted after rigorous double blind review. The topics of the papers form a rich panorama with sentiment analysis, annotation, parsing, bilingual dictionary, semantics and so on. Languages too are diverse covering Punjabi, Bangla, Hindi, Malayalam, Vietnamese and Chinese amongst others. We hope the proceedings of the workshop will be a valuable addition to knowledge and technique of processing Asian Languages.

Pushpak Bhattacharaya (organizing chair)

Key-Sun Choi (workshop chair)

**Organizers:**

Pushpak Bhattacharyya (Chair), IIT Bombay, India  
Key-Sun Choi (Chair), KAIST, South Korea  
Laxmi Kashyap , IIT Bombay, India  
Prof. Malhar Kulkarni , IIT Bombay, India  
Mitesh Khapra, IBM Research Lab, India  
Salil Joshi, IBM Research Lab, India  
Brijesh Bhatt, IIT Bombay, India  
Sudha Bhingardive (Co-organizer), IIT Bombay, India  
Samiulla Shaikh, IIT Bombay, India

**Program Committee:**

Virach Sornlertlamvanich, NECTEC, Thailand  
Kemal Oflazer, Carnegie Mellon University-Qatar, Qatar  
Suresh Manandhar, University of York, Heslington, York  
Philipp Cimiano, University of Bielefeld  
Sadao Kurohashi, Kyoto University, Japan  
Niladri Sekhar Dash, Indian Statistical Institute, Kolkata, India  
Niladri Chatterjee, IIT Delhi, India  
Sudeshna Sarkar, IIT Kharagpur, India  
Ganesh Ramakrishnan, IIT Bombay, India  
Arulmozi S., Thanjavur University, India  
Jyoti Pawar, Goa University, India  
Panchanan Mohanty, University of Hyderabad, India  
Kalika Bali, Microsoft Research, India  
Monojit Choudhury, Microsoft Research, India  
Malhar Kulkarni , IIT Bombay, India  
Girish Nath Jha, JNU, India  
Amitava Das, Samsung Research, India  
Ananthakrishnan Ramanathan, IBM Research Lab, India  
Prasenjit Majumder, DAIICT, Gandhinagar, Kolkata India  
Asif Ekbal, Jadavpur University, India  
Dipti Misra Sharma, IIIT Hyderabad, India  
Sivaji Bandyopadhyaya, Jadavpur University, India  
Kashyap Papat, IIT Bombay, India  
Manish Shrivastava, IIT Bombay, India  
Raj Dabre, IIT Bombay, India  
Balamurali A, IIT Bombay, India  
Vasudevan N, IIT Bombay, India  
Abhijit Mishra, IIT Bombay, India  
Aditya Joshi, IIT Bombay, India  
Ritesh Shah, IIT Bombay, India  
Anoop Kunchookuttan, IIT Bombay, India  
Subhabrata Mukherjee, IIT Bombay, India  
Sobha Nair, AUKBC, India





## Table of Contents

<i>EVBCorpus - A Multi-Layer English-Vietnamese Bilingual Corpus for Studying Tasks in Comparative Linguistics</i>	
Quoc Hung Ngo, Werner Winiwarter and Bartholomäus Wloka . . . . .	1
<i>Building the Chinese Open Wordnet (COW): Starting from Core Synsets</i>	
Shan Wang and Francis Bond . . . . .	10
<i>Detecting Missing Annotation Disagreement using Eye Gaze Information</i>	
Koh Mitsuda, Ryu Iida and Takenobu Tokunaga . . . . .	19
<i>Valence alternations and marking structures in a HPSG grammar for Mandarin Chinese</i>	
Janna Lipenkova . . . . .	27
<i>Event and Event Actor Alignment in Phrase Based Statistical Machine Translation</i>	
Anup Kolya, Santanu Pal, Asif Ekbal and Sivaji Bandyopadhyay . . . . .	36
<i>Sentiment Analysis of Hindi Reviews based on Negation and Discourse Relation</i>	
Namita Mittal, Basant Agarwal, Garvit Chouhan, Nitin Bania and Prateek Pareek . . . . .	45
<i>Annotating Legitimate Disagreement in Corpus Construction</i>	
Billy T.M. Wong and Sophia Y.M. Lee . . . . .	51
<i>A Hybrid Statistical Approach for Named Entity Recognition for Malayalam Language</i>	
Jisha P Jayan, Rajeev R R and Elizabeth Sherly . . . . .	58
<i>Designing a Generic Scheme for Etymological Annotation: a New Type of Language Corpora Annotation</i>	
Niladri Sekhar Dash and Mazhar Mehdi Hussain . . . . .	64
<i>UNL-ization of Punjabi with IAN</i>	
Vaibhav Agarwal and Parteek Kumar . . . . .	72



# Conference Program

## Monday, 14 October 2013

- 9:30-10.00 Inauguration
- 10.00-10.30 Keynote speech - Knowledge-Intensive Structural NLP in the Era of Big Data by Prof. Sadao Kurohashi
- 10.30-11.00 Tea break
- 11.00–11.30 *EVBCorpus - A Multi-Layer English-Vietnamese Bilingual Corpus for Studying Tasks in Comparative Linguistics*  
Quoc Hung Ngo, Werner Winiwarter and Bartholomäus Wloka
- 11.30–12.00 *Building the Chinese Open Wordnet (COW): Starting from Core Synsets*  
Shan Wang and Francis Bond
- 12.00–12.30 *Detecting Missing Annotation Disagreement using Eye Gaze Information*  
Koh Mitsuda, Ryu Iida and Takenobu Tokunaga
- 12.30-13.30 Lunch break
- 13.30–14.00 *Valence alternations and marking structures in a HPSG grammar for Mandarin Chinese*  
Janna Lipenkova
- 14.00–14.30 *Event and Event Actor Alignment in Phrase Based Statistical Machine Translation*  
Anup Kolya, Santanu Pal, Asif Ekbal and Sivaji Bandyopadhyay
- 14.30–15.00 *Sentiment Analysis of Hindi Reviews based on Negation and Discourse Relation*  
Namita Mittal, Basant Agarwal, Garvit Chouhan, Nitin Bania and Prateek Pareek
- 15.00–15.30 *Annotating Legitimate Disagreement in Corpus Construction*  
Billy T.M. Wong and Sophia Y.M. Lee
- 15.30–16.00 *A Hybrid Statistical Approach for Named Entity Recognition for Malayalam Language*  
Jisha P Jayan, Rajeev R R and Elizabeth Sherly
- 16.00-16.30 Tea break

**Monday, 14 October 2013 (continued)**

16.30–17.00 *Designing a Generic Scheme for Etymological Annotation: a New Type of Language Corpora Annotation*

Niladri Sekhar Dash and Mazhar Mehdi Hussain

17.00–17:30 *UNL-ization of Punjabi with IAN*

Vaibhav Agarwal and Parteek Kumar