

# **11th International Conference on Natural Language Processing (ICON-2014)**

Goa, India  
18 – 21 December 2014

ISBN: 978-1-5108-3786-7

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2014) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2017)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Keynote Lecture 1: Complexity of Dependency Representations for Natural Languages</i> Aravind K Joshi .....	1
<i>SMT from Agglutinative Languages: Use of Suffix Separation and Word Splitting</i> Prakash B. Pimpale, Raj Nath Patel and Sasikumar M.....	2
<i>Tackling Close Cousins: Experiences In Developing Statistical Machine Translation Systems For Marathi And Hindi</i> Raj Dabre, Jyotesh Choudhari and Pushpak Bhattacharyya.....	11
<i>Correlating decoding events with errors in Statistical Machine Translation</i> Eleftherios Avramidis and Maja Popović .....	20
<i>Supertag Based Pre-ordering in Machine Translation</i> Rajen Chatterjee, Anoop Kunchukuttan and Pushpak Bhattacharyya .....	30
<i>Duration Modeling by Multi-Models based on Vowel Production characteristics</i> V Ramu Reddy, Parakrant Sarkar and K Sreenivasa Rao .....	39
<i>Voice Activity Detection using Temporal Characteristics of Autocorrelation Lag and Maximum Spectral Amplitude in Sub-bands</i> Sivanand Achanta, Nivedita Chennupati, Vishala Pannala, Mansi Rankawat and Kishore Prahallad	48
<i>Use of GPU and Feature Reduction for Fast Query-by-Example Spoken Term Detection</i> Gautam Mantena and Kishore Prahallad .....	56
<i>Influence of Mother Tongue on English Accent</i> G. Radha Krishna and R. Krishnan .....	63
<i>Keynote Lecture 2: Text Analysis for identifying Entities and their mentions in Indian languages</i> Sobha L .....	68
<i>HinMA: Distributed Morphology based Hindi Morphological Analyzer</i> Ankit Bahuguna, Lavita Talukdar, Pushpak Bhattacharyya and Smriti Singh .....	69
<i>Roles of Nominals in Construing Meaning at the Level of Discourse</i> Soumya Sankar Ghosh and Samir Karmakar .....	76
<i>Anou Tradir: Experiences In Building Statistical Machine Translation Systems For Mauritian Languages – Creole, English, French</i> Raj Dabre, Aneerav Sukhoo and Pushpak Bhattacharyya .....	82
<i>How Sentiment Analysis Can Help Machine Translation</i> Santanu Pal, Braja Gopal Patra, Dipankar Das, Sudip Kumar Naskar, Sivaji Bandyopadhyay and Josef van Genabith .....	89

<i>Introduction to Synskarta: An Online Interface for Synset Creation with Special Reference to Sanskrit</i> Hanumant Redkar, Jai Paranjape, Nilesh Joshi, Irawati Kulkarni, Malhar Kulkarni and Pushpak Bhattacharyya .....	95
<i>LMSim : Computing Domain-specific Semantic Word Similarities Using a Language Modeling Ap- proach</i> Sachin Pawar, Swapnil Hingmire and Girish K. Palshikar .....	101
<i>Multiobjective Optimization and Unsupervised Lexical Acquisition for Named Entity Recognition and Classification</i> Govind, Asif Ekbal and Chris Biemann .....	107
<i>Improving the accuracy of pronunciation lexicon using Naive Bayes classifier with character n-gram as feature: for language classified pronunciation lexicon generation</i> Aswathy P V, Arun Gopi, Sajini T and Bhadrans V K .....	113
<i>Learning phrase-level vocabulary in second language using pictures/gestures and voice</i> Lavanya Prahallad, Prathyusha Danda and Radhika Mamidi .....	119
<i>Creating a PurposeNet Ontology: An insight into the issues encountered during ontology creation</i> Rishabh Srivastava and Soma Paul .....	126
<i>Bundeli Folk-Song Genre Classification with kNN and SVM</i> Ayushi Pandey and Indranil Dutta .....	133
<i>Word net based Method for Determining Semantic Sentence Similarity through various Word Senses</i> Madhuri A. Tayal, M. M. Raghuvanshi and Latesh Malik .....	139
<i>Identification of Karaka relations in an English sentence</i> Sai Kiran Gorthi, Ashish Palakurthi, Radhika Mamidi and Dipti Misra Sharma .....	146
<i>A Sentiment Analyzer for Hindi Using Hindi Senti Lexicon</i> Raksha Sharma and Pushpak Bhattacharyya .....	150
<i>A Sandhi Splitter for Malayalam</i> Devadath V V, Litton J Kurisinkel, Dipti Misra Sharma and Vasudeva Varma .....	156
<i>PaCMan : Parallel Corpus Management Workbench</i> Diptesh Kanojia, Manish Shrivastava, Raj Dabre and Pushpak Bhattacharyya .....	162
<i>How to Know the Best Machine Translation System in Advance before Translating a Sentence?</i> Bibekananda Kundu and Sanjay Kumar Choudhury .....	167
<i>A Domain-Restricted, Rule Based, English-Hindi Machine Translation System Based on Dependency Parsing</i> Pratik Desai, Amit Sangodkar and Om P. Damani .....	177
<i>Translation of TO infinitives in Anusaaraka Platform: an English Hindi MT system</i> Akshar Bharati, Sukhada and Soma Paul .....	186

<i>Determining Trustworthiness in E-Commerce Customer Reviews</i> Dhruv Gupta and Asif Ekbal .....	196
<i>Naturalistic Audio-Visual Emotion Database</i> Sudarsana Reddy Kadiri, P. Gangamohan, V.K. Mittal and B. Yegnanarayana .....	206
<i>Discriminating Neutral and Emotional Speech using Neural Networks</i> Sudarsana Reddy Kadiri, P. Gangamohan and B. Yegnanarayana .....	214
<i>Keynote Lecture 3: Modeling NonPropositional Semantics</i> Lori Levin .....	222
<i>Text Readability in Hindi: A Comparative Study of Feature Performances Using Support Vectors</i> Manjira Sinha, Tirthankar Dasgupta and Anupam Basu .....	223
<i>Sangam: A Perso-Arabic to Indic Script Machine Transliteration Model</i> Gurpreet Singh Lehal and Tejinder Singh Saini .....	232
<i>AutoParSe: An Automatic Paradigm Selector For Nouns in Konkani</i> Shilpa Desai, Neenad Desai, Jyoti Pawar and Pushpak Bhattacharyya .....	240
<i>Continuum models of semantics for language discovery</i> Deepali Semwal, Sunakshi Gupta and Amitabha Mukerjee .....	249
<i>Syllables as Linguistic Units?</i> Amitabha Mukerjee and Prashant Jalan .....	258
<i>Accurate Identification of the Karta (Subject) Relation in Bangla</i> Arnab Dhar and Sudeshna Sarkar .....	267
<i>Manipuri Chunking: An Incremental Model with POS and RMWE</i> Kishorjit Nongmeikapam, Thiyam Ibungomacha Singh, Ngariyanbam Mayekleima Chanu and Sivaji Bandyopadhyay .....	277
<i>Segmentation of Navya-Nyya Expressions</i> Arjuna S. R. and Amba Kulkarni .....	287
<i>Handling Plurality in Bengali Noun Phrases</i> Biswanath Barik and Sudeshna Sarkar .....	295
<i>Making Verb Frames for Bangla Vector Verbs</i> Sanjukta Ghosh .....	305
<i>English to Punjabi Transliteration using Orthographic and Phonetic Information</i> Kamaljeet Kaur and Parminder Singh .....	315
<i>Evaluating Two Annotated Corpora of Hindi Using a Verb Class Identifier</i> Neha Dixit and Narayan Choudhary .....	321

*Hindi Word Sketches*

Anil Krishna Eragani, Varun Kuchib Hotla, Dipti Misra Sharma, Siva Reddy and Adam Kilgarriff  
328

*Extracting and Selecting Relevant Corpora for Domain Adaptation in MT*

Lars Bungum ..... 336

*Merging Verb Senses of Hindi WordNet using Word Embeddings*

Sudha Bhingardive, Ratish Puduppully, Dhirendra Singh and Pushpak Bhattacharyya ..... 344

*Hierarchical Recursive Tagset for Annotating Cooking Recipes*

Sharath Reddy Gunamgari, Sandipan Dandapat and Monojit Choudhury ..... 353

*Named Entity Based Answer Extraction form Hindi Text Corpus Using n-grams*

Lokesh Kumar Sharma and Namita Mittal ..... 362

*“ye word kis lang ka hai bhai?” Testing the Limits of Word level Language Identification*

Spandana Gella, Kalika Bali and Monojit Choudhury ..... 368

*Identifying Languages at the Word Level in Code-Mixed Indian Social Media Text*

Amitava Das and Björn Gambäck ..... 378

*Unsupervised Detection and Promotion of Authoritative Domains for Medical Queries in Web Search*

Manoj K. Chinnakotla, Rupesh K. Mehta and Vipul Agrawal ..... 388

*Significance of Paralinguistic Cues in the Synthesis of Mathematical Equations*

Venkatesh Potluri, SaiKrishna Rallabandi, Priyanka Srivastava and Kishore Prahallad ..... 395