

COLING 2018

**The 27th International Conference  
on Computational Linguistics**

**Proceedings of the First International Workshop on Language  
Cognition and Computational Models (LCCM-2018)**

August 20, 2018  
Santa Fe, New Mexico, USA

Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-1-948087-57-5

## Introduction

Welcome to the COLING-2018 Workshop on Language, Cognition and Computational Models!

Language as a communication tool is one of the key attributes of human society. It is also what distinguishes human communication from most of the other species. Language is, arguably, also what shapes our view of the world. However, language is a complex and intricate tool developed and is continuously evolving over thousands of years, influenced by usage, demographics, and socio-cultural factors. The study of language communication, comprehension and its complex interaction with thought is a rapidly expanding multi-disciplinary and challenging field of research. This growth comes from both its domain and its interdisciplinary nature that confluences cognitive science, computer science, neuroscience, linguistics, psycholinguistics, psychology and many other fields. The development of increasingly sophisticated tools are making it possible to studying different brain activities. A plethora of works have been done studying the representation, organization and processing of language in the human mind. Despite such huge efforts, a coherent picture is yet to emerge. We are yet to go a long-way to develop holistic computational models and make up for the scarcity of corpora in variety of languages.

In addition, each language possess a beauty and uniqueness of its own, and demands a customized approach to understand its intricate relationship with speakers. We especially, encourage works in low resourced and less studied languages and our workshop aims to provide a suitable platform to those less articulated voices.

The goal of this workshop is to bring together researchers working in the field of linguistics, cognitive science, computer science and the intersection of these areas, together and provide a venue for the multidisciplinary discussion of theoretical and practical research for computational models of language and cognition. This knowledge does not only answer one of the primary aspects of cognitive science, but also is useful for designing better NLP systems based on the understood principles. The focus centers around recent advances on cognitively motivated computational models for language representation, organization, processing, acquisition, comprehension and evolution. Given the lack of large standardized corpora for this area of research, we are also interested in developing public data sets for the area and various languages..



## Organizers

**Manjira Sinha, Accenture AI Labs, India:** is associate principal artificial intelligence at Accenture AI Labs. Prior to that she was a research scientist in the Text and Graph Analytics research group at Conduent Labs India (formerly known as Xerox Research Centre India) for 3 years. She is currently working in NLP for Healthcare. She has also worked on cross-domain Text categorization, Social Media analysis for Urban Informatics, Knowledge Extraction, and Quality Analysis for Call center Interactions. Manjira has a Ph.D. in Computer Science from the Indian Institute of Technology Kharagpur. She is also visiting faculty in Indian Institute of Information Technology Kalyani. Her areas of interest include Language Comprehension and Psycholinguistics, Natural Language Processing, Assistive Technology and Human Computer Interaction.

<https://www.linkedin.com/in/manjira-sinha-8554b157/>

**Tirthankar Dasgupta, Innovation Labs, Tata Consultancy Services Limited, India:** is a research scientist at Innovation Labs, Tata Consultancy Services Ltd., India in the Text Analytics and Web Intelligence group. He holds a Ph.D. in computer science from Indian Institute of Technology, Kharagpur. His research interests span natural language processing, computational psycholinguistics, machine learning and human computer interaction. He has organized a number of workshops in the area of assistive technology and natural language processing. He is also an active organizing member and regional coordinator of Panini Linguistic Olympiad in India. He was an organizing member of International Linguistic Olympiad 2016 in India.

<https://www.linkedin.com/in/tirthankar-dasgupta-89b0551/>

## Programme Committee

Narayanan Srinivasan, Centre of Behavioural and Cognitive Sciences , University of Allahabad  
Monojit Choudhury, Microsoft Research India  
Pabitra Mitra, Indian Institute of Technology Kharagpur (IIT), India  
Dipti Mishra Sharma, International Institute of Information Technology Hyderabad (IIIT-H), India  
Ayesha Kidwai, Jawaharlal Nehru University, India  
Jiaul Paik, Indian Institute of Technology Kharagpur (IIT), India  
Rajlakshmi Guha, Indian Institute of Technology Kharagpur (IIT), India.  
Priyanka Sinha, TCS Innovation Labs, India  
Lipika Dey, TCS Innovation Labs  
Kalika Bali, Microsoft Research, India  
Amitava Das, IIIT Sri City  
Sunandan Chakraborty, New York University  
Sandya Mannarswamy, Conduent Labs India  
Vaishna Narang, Jawaharlal Nehru University, India  
Rupsa Saha, TCS Innovation Labs, India  
Moumita Saha, TCS Innovation Labs, India  
Bornini Lahiri, Jadavpur University, India Ritesh Kumar, Bhim Rao Ambedkar University, India  
Dripta Piplai, IIT Kharagpur

## Table of Contents

<i>A Compositional Bayesian Semantics for Natural Language</i>	
Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis and Shalom Lappin . . . . .	1
<i>Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP</i>	
JT Wolohan, Misato Hiraga, Atreyee Mukherjee, Zeeshan Ali Sayyed and Matthew Millard . . . .	11
<i>An OpenNMT Model to Arabic Broken Plurals</i>	
Elsayed Issa . . . . .	22
<i>Enhancing Cohesion and Coherence of Fake Text to Improve Believability for Deceiving Cyber Attackers</i>	
Prakruthi Karuna, Hemant Purohit, Ozlem Uzuner, Sushil Jajodia and Rajesh Ganesan . . . . .	31
<i>Addressing the Winograd Schema Challenge as a Sequence Ranking Task</i>	
Juri Opitz and Anette Frank . . . . .	41
<i>Finite State Reasoning for Presupposition Satisfaction</i>	
Jacob Collard . . . . .	53
<i>Language-Based Automatic Assessment of Cognitive and Communicative Functions Related to Parkinson's Disease</i>	
Lesley Jessiman, Gabriel Murray and McKenzie Braley . . . . .	63
<i>Can spontaneous spoken language disfluencies help describe syntactic dependencies? An empirical study</i>	
M. KURDI . . . . .	75
<i>Word-word Relations in Dementia and Typical Aging</i>	
Natalia Arias-Trejo, Aline Minto-García, Diana I. Luna-Umanzor, Alma E. Ríos-Ponce, Balderas-Pliego Mariana and Gemma Bel-Enguix . . . . .	85
<i>Part-of-Speech Annotation of English-Assamese code-mixed texts: Two Approaches</i>	
Ritesh Kumar and Manas Jyoti Bora . . . . .	94





## Conference Program

*A Compositional Bayesian Semantics for Natural Language*

Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis and Shalom Lappin

*Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP*

JT Wolohan, Misato Hiraga, Atreyee Mukherjee, Zeeshan Ali Sayyed and Matthew Millard

*An OpenNMT Model to Arabic Broken Plurals*

Elsayed Issa

*Enhancing Cohesion and Coherence of Fake Text to Improve Believability for Deceiving Cyber Attackers*

Prakruthi Karuna, Hemant Purohit, Ozlem Uzuner, Sushil Jajodia and Rajesh Ganesan

*Addressing the Winograd Schema Challenge as a Sequence Ranking Task*

Juri Opitz and Anette Frank

*Finite State Reasoning for Presupposition Satisfaction*

Jacob Collard

*Language-Based Automatic Assessment of Cognitive and Communicative Functions Related to Parkinson's Disease*

Lesley Jessiman, Gabriel Murray and McKenzie Braley

*Can spontaneous spoken language disfluencies help describe syntactic dependencies? An empirical study*

M. KURDI

*Word-word Relations in Dementia and Typical Aging*

Natalia Arias-Trejo, Aline Minto-García, Diana I. Luna-Umanzor, Alma E. Ríos-Ponce, Balderas-Pliego Mariana and Gemma Bel-Enguix

*Part-of-Speech Annotation of English-Assamese code-mixed texts: Two Approaches*

Ritesh Kumar and Manas Jyoti Bora

