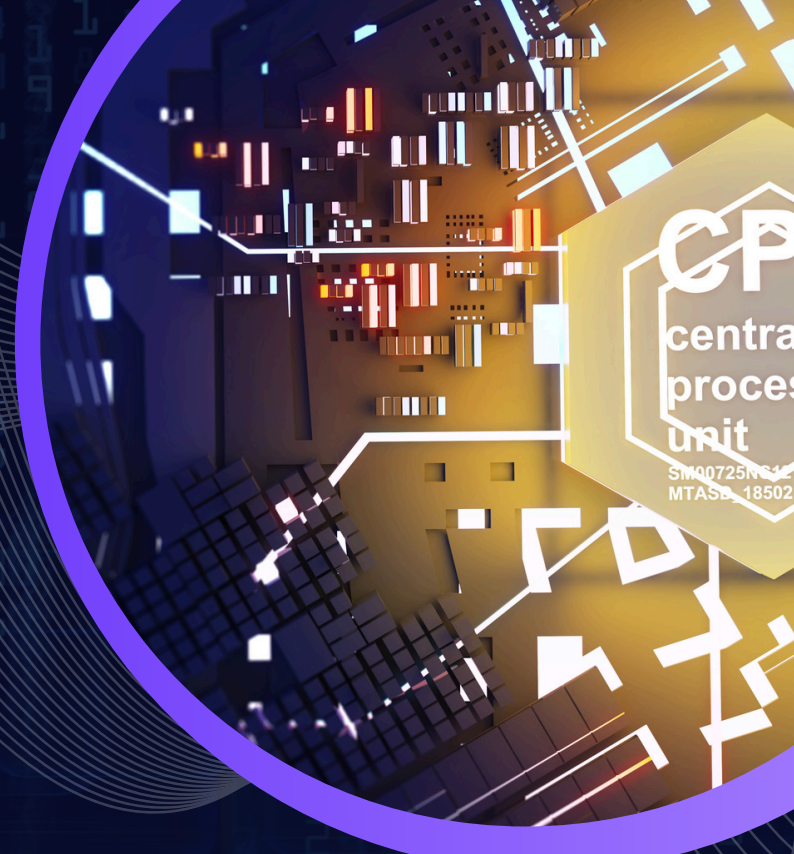


SNLP 2020



Proceedings

Symposium on

Natural Language Processing 2020

20th October, 2020

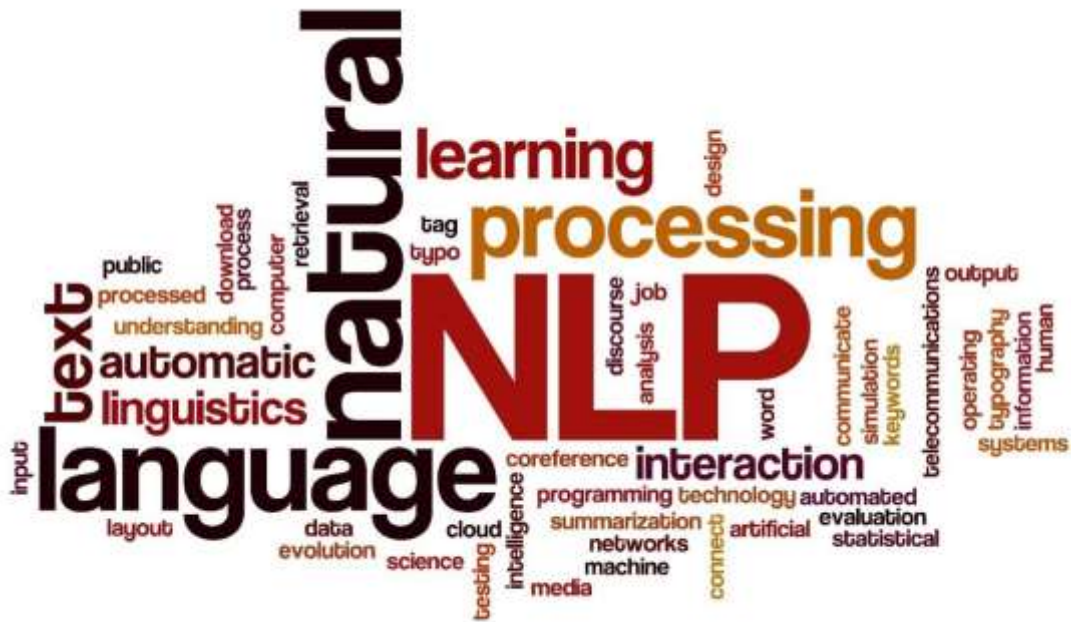


Proceedings

Symposium on Natural Language

Processing 2020

(SNLP 2020)



National Language Processing Centre University of
Moratuwa Sri Lanka

Cataloguing-in-Publication Data Approved by the
National Library and Documentation Services Board

Symposium on Natural Language Processing 2020 (SNLP)

(20th October 2020 : Moratuwa-Sri Lanka)

Proceedings / ed. Uthayasanker Thayasivam. - Moratuwa : University
of Moratuwa, 2024

21 p. : ill. ; 30 cm.

ISBN 978-955-9027-86-7

i . 006.35 DDC 23

ii. Thayasivam, Uthayasanker(ed.)

iii. Title

1. Natural language processing(Computer science)
2. Artificial intelligence
3. Machine learning

Preface

We are pleased to present this collection of abstracts for the *Symposium on Natural Language Processing 2020* hosted by the National Language Processing Centre at the University of Moratuwa, Sri Lanka. This proceeding reflects the cutting-edge research and innovative contributions of the authors in the fields of natural language processing (NLP), machine learning, and semantic technologies, with a special focus on low-resource languages such as Sinhala and Tamil.

The research included in this collection spans a diverse range of topics, including but not limited to named entity recognition, machine translation, sentiment analysis, and automated text summarization. These studies not only advance the field of NLP for underrepresented languages but also demonstrate the potential of applying deep learning, hybrid models, and other innovative techniques to overcome unique linguistic challenges.

We hope that the insights and advancements presented in these abstracts will foster discussions, inspire collaboration, and contribute to the collective growth of NLP technologies globally. We are confident that the work presented here will be of interest to researchers, academics, and professionals striving to push the boundaries of semantic and natural language processing.

On behalf of the organizing committee, we extend our sincere thanks to all the contributors and reviewers for their efforts and commitment to advancing this vibrant field of research.

**Dr. Uthayasanker Thayasivam,
National Language Processing Center,
University of Moratuwa**

Editorial Board

Editor in Chief

Dr. Uthayasanker Thayasivam

Co-Editor

Ms. Charitha Rathnayaka

Academic Staff at NLPC

Prof. Gihan Dias, Senior lecturer, Director of NLPC

Prof. Sanath Jayasena, Senior lecturer

Dr. Surangika Ranthunga, Senior lecturer

Dr. Uthayasanker Thayasivam, Senior lecturer

Keynote Speaker

Prof. Asif Ekbal, Senior lecturer, Department of Computer Science and Engineering, IIT Patna

Table of Contents

Preface	2
Editorial Board	3
Keynote Address	6
Recent Achievements	7
Event Schedule and Session Overview	8
Abstracts	10
Issues in Employing Metaphorically Equivalent Phrases according to Cognitive Translation Hypothesis in Translating Sinhala Text into English to Train an AI Translation Software	10
Authors: Lakmini Herath (lakminiherath@live.com), Buddhika Ranaweera (buddhiran@gmail.com)	10
Automatic Sinhala News Text Summarizer	11
Author: Sajini Tennakoon.....	11
Named Entity Boundary Detection for Religious Unhealthy Statements in Social Media	12
Author: Prasan Yapa	12
Old Sinhala Newspaper Reader for People with Visual Impairment	13
Authors: Vindya Perera (vindipererauom@gmail.com) ¹ , P.P.A Gayashan (p.p.a.gayashan@gmail.com) ² , Shashiwadana Nirmani (shashiwadanannirmani14@gmail.com) ³ ...	13
An Improved kNN Algorithm using K-means and fastText to Predict Sentiments Expressed in Tamil Texts.....	14
Authors: Sajeetha Thavareesan ¹ , Sinnathamby Mahesan ²	14
Smart Mail Sorting System for Sinhala Handwritten Addresses.....	15
Authors: W.D.R. Fernando (dinal.16@itfac.mrt.ac.lk), S.D.P. Bhashani (piumikadasanayake@gmail.com), L.P.V.H. Ranasingha (ranasinghahanshika@gmail.com).....	15
Assessing the Influence of Structural Ambiguity in Technical Translation with Special Reference to Administrative Documents in Sri Lanka.....	16
Authors: Nipuni Sewwandi ¹ , Harsha Amarasinghe ²	16
Effectiveness of Grammatically Correct Translation from English into Sinhalese: With Special Reference to the Annual Report 2019/20 of Sri Lankan Airlines	17
Author: D. D. I. M. B. Gunathilaka (ddimbg@gmail.com).....	17
Lexical Uncertainties Confronted by Government Translators When translating Legal Documents from Sinhala to English.....	18
Author: J.A.M Hansani (madukahansani89@gmail.com).....	18
Polysemy in Technical Translations	19
Author: S. A. T. A. Senevirathne	19

A Study of the Language Practices of the Government 20
Languages Proficiency in Sri Lanka 20
Author: P. S. Manthirathna 20

Keynote Address

Thank you for the invitation to speak at SNLP 2020. It's an honor to be here among so many innovative minds pushing the boundaries of Natural Language Processing and Conversational AI. Today, I'll be discussing how we can advance the field by teaching conversational machines with a focus on affective understanding and human-like interaction.

To start, we'll look at the foundational elements of *Conversational AI* and the concept of *Grounding Conversation*—the process of ensuring that conversations are relevant, coherent, and contextually anchored. As conversational agents become more present in our daily lives, there's an increasing need to induce human qualities such as courteousness, sentiment, and emotional responsiveness. I'll share techniques we're using to embed these traits into conversational models, allowing for more natural and empathetic interactions.

Next, we'll explore the role of *personalization in conversational agents*, enabling them to adapt to individual user preferences and styles. I'll also touch on *multi-modal response generation*, a crucial development that allows conversational machines to incorporate multiple forms of input—like text, images, and voice—to generate responses that are richer and more context-aware.

To wrap up, I'll summarize the advancements we've discussed and share an overview of our recent research, showcasing how these principles are coming to life in our latest projects. By embedding affects into conversational machines, we're not just building better AI—we're building systems that can understand, engage, and genuinely connect with people.

My best wishes go with organizers and active participants.

Prof. Asif Ekbal,

Senior lecturer, Department of Computer Science and Engineering, IIT Patna

Recent Achievements

- NLPC has won the Digital Social Impact award at the e-Suwabhimani 2018 under the category of Government and Citizen Engagement.
- Won AHEAD, World Bank fund in 2019
- NLPC organized Summer School in 2019 collaboratively with University of Konstanz, Germany, University of Moratuwa and University of Sri Jayewardenepura with 14 participants from various South Asian countries and NLPC/CL academics and pioneers resource people from Qatar and Germany
- NLPC Invited Prof Pushpak Bhattacharyya, is a computer scientist and a Professor at Computer Science and Engineering Department, IIT Bombay, Former Director, IIT Patna and conducted workshop collaboratively with SLIIT in 2019
- Under DAAD - NSF Sri Lanka – based Personnel Exchange Programme won DAAD- NSF PPP grant. Invited Ph.D student to University of Moratuwa from Konstanz and sent Ph.D student to University of Konstanz. Organized a workshop at University of Moratuwa in 2019
- Constantly organizing NLP Symposium. This will be the 7 th SNLP organized by NLPC
- NLPC launched the translation support systems between Sinhala → English and English → Sinhala on 1st July 2020 while SiTa Computer-Assisted Translation System which assists in the translation of documents from Sinhala to Tamil and vice versa since 2018.
- NLPC is planning to include English to Tamil and vice versa in the future.

Event Schedule and Session Overview

Session 1

Introduction	Dr. Uthayasankar Thayasivam
Welcome & Address speech	Prof. Gihan Dias (Director of NLPC)
Welcome & Address speech	Dr. Indika Perera (Head of Department of CSE)
Introduction of the Keynote speaker	Prof. Sanath Jayasena
Keynote speech	Prof. Asif Ekbal
Coffee break	

Session 2

Machine Learning Session	Dr. Surangika Ranathunga
Issues in Employing Metaphorically Equivalent Phrases according to Cognitive Translation Hypothesis in Translating Sinhala Text into English to Train an AI Translation Software	Lakmini Herath, Buddhika Ranaweera
Automatic Sinhala News Text Summarizer	Sajini Tennakoon
Named Entity Boundary Detection for Religious Unhealthy Statements in Social Media	Prasan Yapa
Old Sinhala Newspaper Reader for People With Visual Impairment	Vindya Perera
An Improved kNN Algorithm using K-means and fastText to Predict Sentiments Expressed in Tamil Texts	Sajeetha Thavareesan
Smart Mail Sorting System for Sinhala HandWritten Addresses	W.D.R Fernando, S.D.P. Bhashini, L.P.V.H. Ranasingha
Handling Out-of-Vocabulary Problem in Sinhala-English Domain-Specific NMT	Aloka Fernando
Using back translation to improve domain specific English-Sinhala	Koshiya Epaliyana

Neural Machine Translation	
Inducing bilingual dictionaries for Sinhala-English languages by using modern word embedding models such as fasttext, word2vec or glovec.	Anushika Liyanage
Multi-task and Transfer Learning for Sinhala-Tamil-English Neural Machine Translation	Sarubi Thillainathan
Sub wording techniques for Neural Machine Translation between English and Tamil.	Janarthanasarma Baskarakurukkal
Named Entity Recognition for Sinhala, Tamil and English Text.	Rameela Azeez
Automatic Post Editing (APE) for the improvement of Machine Translation systems.	Chathuri Jayaweera. Anosha Ignatius
Computational Linguistics	Prof. Gihan Dias
Assessing the Influence of Structural Ambiguity in Technical Translation with Special Reference to Administrative Documents in Sri Lanka.	Nipuni Sewwandi Harsha Amarasinghe
Effectiveness of Grammatically Correct Translation from English into Sinhalese: With Special Reference to the Annual Report 2019/20 of SriLankan Airlines.	D. D. I. M. B. Gunathilaka
Lexical Uncertainties Confronted by Government Translators When translating Legal Documents from Sinhala to English	J.A.M Hansani
Polysemy in Technical Translations	S. A. T. A. Senevirathne
A Study of the Language Practices of the Government Service: With Reference to the Examination for Official Languages Proficiency in Sri Lanka	P. S. Manthirathna Chamila Liyanage
Conversion of Concise Sinhala Dictionary to structured digital form	Prof. Gihan Dias
A Study of Sinhala Word Morphology.	Indu Herath
A Corpus Based Study on the Usage of ‘Wisin’ () Particle in Sinhala Official Letters	D.M. Saman Peramuna
Development of Grammar Error Checker for Sinhala Language.	Kalindu Kumarasinghe
How I developed a Universal Dependency Parser for Tamil?	Sarveswaran K
