

CALCS 2025

Computational Approaches to Linguistic Code-Switching

Proceedings of the Workshop

May 3, 2025

©2025 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-053-0

Introduction

Welcome to the proceedings of the seventh edition of the workshop on computational approaches for linguistic code-switching (CALCS-2025)! Code-switching is a common phenomenon in the multilingual communities where multilingual speakers communicate by moving back and forth between the languages they speak when communicating with other multilingual speakers. This year the workshop is being held in Albuquerque, New Mexico, USA on May 3rd, 2025 at NAACL.

This workshop series brings together experts and practitioners that are currently working on different aspects of code-switching with a special focus on motivating tighter collaborations between speech and text researchers. We received 13 regular workshop submissions, of which we accepted 5, 2 non-archival, and 1 shared task paper. Our workshop also aims to motivate new research and energize the community to take on the challenges posed by code-switching data.

The workshop program includes short talks from regular workshop submissions and keynote speakers. We also have a stellar invited speaker program with a keynote talk by Sunayana Sitaram, Monojit Choudhury, and Alham Fikri Aji. We would like to thank the NAACL workshop organizers for their help during the organization of the workshop. It would have been great to see everyone face to face in Albuquerque and we hope that you join us on May 3rd and that you enjoy the program we put together.

Let's talk code-switching in the desert!

The Workshop Organizers

Organizing Committee

Organizer

Genta Indra Winata, Capital One AI Foundations

Sudipta Kar, Oracle Health AI

Marina Zhukova, University of California, Santa Barbara

Thamar Solorio, University of Houston and Mohamed bin Zayed University of Artificial Intelligence

Xi Ai, National University of Singapore

Injy Hamed, Mohamed bin Zayed University of Artificial Intelligence

Mahardika Krisna Ihsani, Mohamed bin Zayed University of Artificial Intelligence

Derry Tanti Wijaya, Monash University Indonesia and Boston University

Garry Kuwanto, Boston University

Program Committee

Program Chairs

Genta Indra Winata, Capital One AI Foundations
Sudipta Kar, Amazon
Marina Zhukova, University of California, Santa Barbara
Thamar Solorio, University of Houston and Mohamed bin Zayed University of Artificial Intelligence
Xi Ai, National University of Singapore
Injy Hamed, Mohamed bin Zayed University of Artificial Intelligence
Mahardika Krisna Ihsani, Mohamed bin Zayed University of Artificial Intelligence
Derry Tanti Wijaya, Monash University Indonesia and Boston University
Garry Kuwanto, Boston University

Panelist

Ruo Chen Zhang
Sunayana Sitaram, Microsoft
Monojit Choudhury, Mohamed bin Zayed University of Artificial Intelligence
Alham Aji, Mohamed bin Zayed University of Artificial Intelligence

Invited Speakers

Sunayana Sitaram, Microsoft
Monojit Choudhury, Mohamed bin Zayed University of Artificial Intelligence
Alham Aji, Mohamed bin Zayed University of Artificial Intelligence

Keynote Talk

Sunayana Sitaram
Microsoft



Bio: Sunayana Sitaram is a Principal Researcher at Microsoft Research India. Her research goal is to make AI more inclusive to everyone on the planet. Her current area of research is on measuring and improving the performance of Large Language Models on non-English languages. Sunayana also serves as the director of the MSR India Research Fellow program, that currently houses around 65 Research Fellows and exposes bright young researchers to a world-class research environment to prepare them for careers in research, engineering and entrepreneurship. Prior to joining MSRI as a Post Doc Researcher, Sunayana completed her MS and PhD at the Language Technologies Institute, Carnegie Mellon University in 2015. Sunayana’s research has been published in top NLP and Speech conferences including ACL, EMNLP, Interspeech, ICASSP and she regularly serves in the organizing committee of these conferences.

Keynote Talk

Monojit Choudhury

Mohamed bin Zayed University of Artificial Intelligence



Bio: Monojit Choudhury is a Professor of Natural Language Processing at MBZUAI. His research focuses on the intersection of language technology and society, examining how foundation models learn and (mis)represent linguistic and cultural diversity. He investigates the impact of representational disparities on technology use and their broader implications for linguistic and cultural dynamics. A key objective of his work is to develop fair and equitable language technologies that contribute to a more inclusive future. Additionally, he explores the use of generative AI for large-scale quantitative investigations and modeling of cultural phenomena.

Keynote Talk

Alham Fikri Aji

Mohamed bin Zayed University of Artificial Intelligence



Bio: Alham Fikri Aji is an Assistant Professor at MBZUAI and an Adjunct Faculty member at Monash Indonesia. He earned his Ph.D. from the University of Edinburgh’s Institute for Language, Cognition, and Computation, where he focused on improving the training and inference speed of machine translation under the supervision of Dr. Kenneth Heafield and Dr. Rico Sennrich. His research explores multilingual, low-resource, and low-compute NLP, with a particular emphasis on developing multilingual large language models and building NLP resources for underrepresented languages, especially Indonesian.

Table of Contents

<i>EuskañolDS: A Naturally Sourced Corpus for Basque-Spanish Code-Switching</i> Maite Heredia, Jeremy Barnes and Aitor Soroa	1
<i>The Impact of Code-switched Synthetic Data Quality is Task Dependent: Insights from MT and ASR</i> Injy Hamed, Thang Vu and Nizar Habash	6
<i>Beyond Monolingual Limits: Fine-Tuning Monolingual ASR for Yoruba-English Code-Switching</i> Oreoluwa Boluwatife Babatunde, Victor Tolulope Olufemi, Emmanuel Bolarinwa, Kausar Yetunde Moshood and Chris Chinenye Emezue	18
<i>Where and How Do Languages Mix? A Study of Spanish-Guaraní Code-Switching in Paraguay</i> Olga Kellert and Nemika Tyagi	26
<i>Tongue-Tied: Breaking LLMs Safety Through New Language Learning</i> Bibek Upadhyay and Vahid Behzadan	32
<i>LexiLogic@CALCS 2025: Predicting Preferences in Generated Code-Switched Text</i> Pranav Gupta, Souvik Bhattacharyya, Niranjana Kumar M and Billodal Roy	48