

SIGTURK 2024

**The First Workshop on Natural Language Processing for
Turkic Languages (SIGTURK 2024)**

Proceedings of the Workshop

August 15, 2024

©2024 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 979-8-89176-140-7

Preface

Welcome to the First Workshop on Natural Language Processing for Turkic Languages (SIGTURK 2024), held on August 15, 2024, in Bangkok, Thailand, and online.

This inaugural workshop received 25 submissions, out of which 9 papers were accepted as archival publications and 7 as non-archival. All 16 accepted papers will be presented as posters in addition to oral presentations.

We are excited to bring together researchers working on NLP for Turkic languages and hope this workshop will foster collaboration and advance the field. The program includes invited talks, oral presentations, and a poster session showcasing the latest work in this area.

We thank all authors for their submissions, the program committee for their thorough reviews, and our invited speakers for sharing their expertise. We look forward to engaging discussions and new connections made at SIGTURK 2024.

SIGTURK 2024 Organizers

Organizing Committee

Organizers

Duygu Ataman, New York University

Mehmet Oguz Derin

Sardana Ivanova, University of Helsinki

Abdullatif Köksal, Ludwig-Maximilians-Universität München

Jonne Sälevä, Brandeis University

Deniz Zeyrek, Middle East Technical University

Program Committee

Organizers

Duygu Ataman, New York University
Mehmet Oguz Derin
Sardana Ivanova, University of Helsinki
Abdullatif Köksal, Ludwig-Maximilians-Universität München
Deniz Zeyrek, Middle East Technical University
Jonne Sälevä, Brandeis University

Reviewers

Duygu Ataman, New York University
Necva Bölücü, CSIRO
Cagri Coltekin, University of Tuebingen
Mehmet Oguz Derin, Mehmet Oguz Derin
Orhan Firat, Google
Omer Goldman, Bar Ilan University
Mammad Hajili, Microsoft
Sardana Ivanova, University of Helsinki
Murathan Kurfali
Abdullatif Köksal, Ludwig-Maximilians-Universität München
Constantine Lignos, Brandeis University
Aziza Mirsaidova
Saliha Muradoglu
Jonne Sälevä, Brandeis University
Francis M. Tyers, Indiana University, Bloomington
Jonathan Washington, Swarthmore College
Deniz Zeyrek, Middle East Technical University
Arzucan Özgür, Boğaziçi University
Adnan Öztürel, Google
Gözde Gül Şahin, Koç University

Table of Contents

<i>Unsupervised Learning of Turkish Morphology with Multiple Codebook VQ-VAE</i> Müge Kural and Deniz Yuret	1
<i>Open foundation models for Azerbaijani language</i> Jafar Isbarov, Kavsar Huseynova, Elvin Mammadov and Mammad Hajili	18
<i>ImplicaTR: A Granular Dataset for Natural Language Inference and Pragmatic Reasoning in Turkish</i> Mustafa Kürşat Halat and Ümit Atlamaz	29
<i>A coreference corpus of Turkish situated dialogs</i> Faruk Büyüktekin and Umut Özge	42
<i>Do LLMs Recognize me, When I is not me: Assessment of LLMs Understanding of Turkish Indexical Pronouns in Indexical Shift Contexts</i> Metehan Oğuz, Yusuf Umut Ciftci and Yavuz Faruk Bakman	53
<i>Towards a Clean Text Corpus for Ottoman Turkish</i> Fatih Burak Karagöz, Berat Doğan and Şaziye Betül Özateş	62
<i>Turkish Delights: a Dataset on Turkish Euphemisms</i> Hasan Can Biyik, Patrick Lee and Anna Feldman	71
<i>Do LLMs Speak Kazakh? A Pilot Evaluation of Seven Models</i> Akylbek Maxutov, Ayan Myrzakhmet and Pavel Braslavski	81
<i>Intelligent Tutor to Support Teaching and Learning of Tatar</i> Alsu Zakirova, Jue Hou, Anisia Katinskaia, Anh-Duc Vu and Roman Yangarber	92

Program

Thursday, August 15, 2024

- 09:00 - 09:10 *Opening remarks*
- 09:10 - 10:30 *Morning session 1: Invited speech*
- 10:30 - 11:00 *Coffee break*
- 11:00 - 12:30 *Morning session 2: Oral presentations*
- 12:30 - 13:30 *Lunch*
- 13:30 - 14:50 *Afternoon session 1: Oral presentations*
- 14:50 - 15:30 *Afternoon session 2: Invited speech*
- 15:30 - 16:00 *Coffee break*
- 16:00 - 17:00 *Poster session*
- 17:00 - 17:40 *Afternoon session 3: Invited speech*
- 17:40 - 18:00 *Closing remarks*