SIGTURK 2026

The Second Workshop on Natural Language Processing for Turkic Languages

Proceedings of the Workshop

March 29, 2026

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 979-8-89176-370-8

Preface by the General Chair

Welcome to the Second Workshop on Natural Language Processing for Turkic Languages (SIGTURK 2026), held on March 29, 2026, in Rabat, Morocco.

This workshop received 28 submissions, out of which 20 papers were accepted as archival publications. Out of 20 accepted papers, we invited 8 papers as oral presentation during the workshop and 11 invited as poster presentations.

We are excited to bring together researchers working on NLP for Turkic languages and hope this workshop will foster further collaborations and advance the field. This year's participants made contributions by introducing new datasets and tools, presenting novel approaches to train LLMs and their evaluation, and diverse applications of models on Turkic languages.

We thank all authors for their submissions and the program committee for their thorough reviews. We look forward to engaging discussions and new connections made at SIGTURK 2026.

Kemal Oflazer, General Chair
Abdullatif Köksal and Onur Varol, Program Co-Chairs

Organizing Committee

General Chair

Kemal Oflazer, Carnegie Mellon University, USA

Program Committee Co-Chairs

Abdullatif Köksal, Google Deepmind, UK
Onur Varol, Sabanci University, Türkiye

Publicity Chair

Jonne Sälevä, Brandeis University, USA

Shared-Task Chair

Gözde Gül Şahin, FAU Erlangen-Nürnberg, Germany

Program Committee

Program Chairs

Abdullatif Köksal, Google Deepmind
Onur Varol, Sabanci University

Reviewers

Emre Can Acikgoz, Ilseyar Alimova, Mehmet Fatih Amasyali, İnanç Arın

Nimet Beyza Bozdag, Cem H. Bozsahin, Necva Bölücü

Cagri Coltekin

A. Seza Doğruöz

Gülşen Eryiğit

Tunga Gungor

Dilek Hakkani-Tür

Jafar Isbarov

Dilara Keküllüoğlu, Aykut Koc, Murathan Kurfali, Abdullatif Köksal

Constantine Lignos

Arzucan Özgür, Adnan Öztürel

Anar Rzayev

Lütfi Kerem Senel

A. Cüneyd Tantug, Cagri Toraman, Gokhan Tur

Jonathan Washington

Reyyan Yeniterzi, Suveyda Yeniterzi, Deniz Yuret

Kerem Zaman, Deniz Zeyrek

Table of Contents

Program

Sunday, March 29, 2026

09:00 - 09:15 *Opening Remarks*

09:15 - 10:00 *Session 1*

SindBERT, the Sailor: Charting the Seas of Turkish NLP
Raphael Schmitt and Stefan Schweter

Building a Turkish Large Language Model via Continual Pre-Training and Parameter-Efficient Adaptation
Alperen Enes Bayar, Mert Ege, Gökhan Yurtalan, Alper Karamanlioglu, Berkan Demirel and Ramazan Gokberk Cinbis

When Semantic Overlap Is Not Enough: Cross-Lingual Euphemism Transfer Between Turkish and English
Hasan Can Biyik, Libby Barak, Jing Peng and Anna Feldman

10:00 - 10:45 *Invited Talk - Mirac Suzgun, Stanford University*

10:45 - 11:00 *Coffee Break*

11:00 - 12:30 *Session 2*

TUNE: A Task For Turkish Machine Unlearning For Data Privacy
Doruk Benli, Ada Canoğlu, Nehir İlkim Gönençer and Dilara Keküllüoğlu

TurkBench: A Benchmark for Evaluating Turkish Large Language Models
Cagri Toraman, Ahmet Kaan Sever, Ayşe Aysu Cengiz, Elif Ecem Arslan, Görkem Sevinç, Sarp Kantar, Mete Mert Birdal, Yusuf Faruk Güldemir, Ali Buğra Kanburoğlu, Sezen Felekoğlu, Birsen Şahin Kütük, Büşra Tufan, Elif Genç, Serkan Coşkun, Gupse Ekin Demir, Muhammed Emin Arayıcı, Olgun Dursun, Onur Gungor, Susan Üsküdarlı, Abdullah Topraksoy and Esra Darıcı

Modelling the Morphology of Verbal Paradigms: A Case Study in the Tokenization of Turkish and Hebrew
Giuseppe Samo and Paola Merlo

Beyond the Token: Correcting the Tokenization Bias in XAI via Morphologically-Aligned Projection
Muhammet Anil Yagiz and Fahrettin Horasan

BIRDTurk: Adaptation of the BIRD Text-to-SQL Dataset to Turkish
Burak Aktaş, Mehmet Can Baytekin, Süha Kağan Köse, Ömer İlbilgi, Elif Özge Yılmaz, Cagri Toraman and Bilge Kaan Görür

Sunday, March 29, 2026 (continued)

Overview of the SIGTURK 2026 Shared Task: Terminology-Aware Machine Translation for English–Turkish Scientific Texts
Ali Gebeşçe, Abdulfattah Safa, Ege Uğur Amasya and Gözde Gül Şahin

14:00 - 15:30 *Poster*

TR-EduVSum: A Turkish-Focused Dataset and Consensus Framework for Educational Video Summarization
Figen Eğin and Aytuğ Onan

SarcasTürk: Turkish Context-Aware Sarcasm Detection Dataset
Niyazi Ahmet Metin, Sevde Yılmaz, Osman Enes Erdoğdu, Elif Sude Meydan, Oğul Sümer and Dilara Keküllüoğlu

RAGTurk: Best Practices for Retrieval Augmented Generation in Turkish
Süha Kağan Köse, Mehmet Can Baytekin, Burak Aktaş, Bilge Kaan Görür, Evren Ayberk Munis, Deniz Yılmaz, Muhammed Yusuf Kartal and Cagri Toraman

A Morphology-Aware Evaluation of Turkish Syntax in Large Language Models
Ezgi Başar and Arianna Bisazza

OCRTurk: A Comprehensive OCR Benchmark for Turkish
Deniz Yılmaz, Evren Ayberk Munis, Cagri Toraman, Süha Kağan Köse, Burak Aktaş, Mehmet Can Baytekin and Bilge Kaan Görür

A Unified Turkic Idiom Understanding Benchmark: Idiom Detection and Semantic Retrieval Across Five Turkic Languages
Gözde Aslantaş and Tunga Gungor

Language Matters: Target-Language Supervision for Political Bias Detection in Turkish News
Umut Ozbagriacik and Haim Dubossarsky

Benchmarking Hate Speech Detection in Azerbaijani with Turkish Cross-Lingual Transfer and Transformer Models
Tural Alizada and Haim Dubossarsky

From Lemmas to Dependencies: What Signals Drive Light Verbs Classification?
Sercan Karakas and Yusuf Şimşek