EMNLP 2023

**Proceedings of the CRAC 2023 Shared Task on
Multilingual Coreference Resolution**

at

**The Sixth Workshop on Computational Models of
Reference, Anaphora and Coreference (CRAC 2023)**

CRAC 2023, an EMNLP 2023 Workshop
December 7, 2023
Singapore

Preface

This volume contains papers describing the CRAC 2023 Shared Task on Multilingual Coreference Resolution and the participating systems. The public edition of the multilingual collection CorefUD 1.1 was used as the source of training and evaluation data, spanning 17 datasets for 12 languages, namely Catalan, Czech, English, French, German, Hungarian, Lithuanian, Norwegian, Polish, Russian, Spanish, and Turkish. Shared task participants were supposed to identify mentions in texts and to predict coreference relations between the identified mentions; only identity coreference is considered in this shared task.

7 systems participated in the shared mask. In this volume, system description papers delivered by 4 teams are presented, preceded with an overview paper describing in more detail the task itself, the input data, the baseline system, the main evaluation metric, and global performance comparisons.

This year's shared task follows up on the first edition of the shared task held with CRAC 2022. The number of languages as well as the number of participating teams has grown, and we can only hope that this will become a trend.

Finally, we would like to thank all the participants for their efforts, and program committee members for reviewing the submitted manuscripts. In addition, we would like to thank all authors of the involved coreference datasets for making the results of their work publicly accessible.

<div align="right">

November 2023
Maciej Ogrodniczuk, Zdeněk Žabokrtský
on behalf of the shared task organizers

</div>

Shared task specification

https://ufal.mff.cuni.cz/corefud/crac23

Shared task organizers

- Charles University (Prague, Czechia):
    - Anna Nedoluzhko
    - Michal Novák
    - Martin Popel
    - Zdeněk Žabokrtský
    - Daniel Zeman

- Institute of Computer Science, Polish Academy of Sciences (Warsaw, Poland):
    - Maciej Ogrodniczuk

- University of West Bohemia (Pilsen, Czechia):
    - Miloslav Konopík
    - Ondřej Pražák
    - Jakub Sido

Program Committee

- Veronique Hoste
- Miloslav Konopík
- Anna Nedoluzhko
- Vincent Ng
- Michal Novák
- Massimo Poesio
- Martin Popel
- Ondrej Prazak
- Jakub Sido
- Daniel Zeman

Invited Talk

The CRAC 2023 Shared Task on Multilingual Coreference Resolution

Milan Straka, Charles University, Czech Republic

Abstract

In a manner consistent with development in various domains of natural language processing, the performance of coreference resolution systems has been exhibiting a consistent improvement over recent years. With coreference resolution being a complex structured prediction problem, quite a few approaches have been put forth, encompassing auto-/non-autoregressive decoding, diverse mention representation, and pretrained language models of varying size and kind. In this talk, I seek to offer a review of prominent approaches and assess and compare them with a high degree of independence. Furthermore, owing to the CorefUD initiative providing datasets in many languages, I aim to empirically quantify the impact of multilingual and crosslingual transfer on the performance of the best system of the CRAC 2023 Shared Task on Multilingual Coreference Resolution.

Speaker Bio

Milan Straka is an assistant professor at the Institute of Formal and Applied Linguistics at the Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic. He is the (co-)author of several shared-task-winning NLP tools like UDPipe, a morphosyntactic analyzer for currently 72 languages; PERIN, a semantic parser; and CorPipe, the winner of CRAC 2022 and 2023 shared tasks on multilingual coreference resolution. His further research interests include named entity recognition, named entity linking, grammar error correction, and multilingual models in general.

Table of Contents

Shared Task Session Program

Thursday, December 7, 2023

Welcome

9:00–9:05   *Opening and Welcome*

CRAC Shared Task Invited Talk

9:05–10:00   *Recent Computational Approaches to Coreference Resolution*
Milan Straka

CRAC Shared Task Overview

10:00–10:30   *Findings of the Second Shared Task on Multilingual Coreference Resolution*
Zdeněk Žabokrtský, Miloslav Konopík, Anna Nedoluzhko, Michal Novák, Maciej Ogrodniczuk, Martin Popel, Ondřej Pražák, Jakub Sido, Daniel Zeman and Yilun Zhu

Short Break

10:30–11:00   *Coffee Break*

CRAC Shared Task System Description Session

11:00–11:20   *Multilingual coreference resolution: Adapt and Generate*
Natalia Skachkova, Tatiana Anikina and Anna Mokhova

11:20–11:40   *Neural End-to-End Coreference Resolution using Morphological Information*
Tuğba Pamay Arslan, Kutay Acar and Gülşen Eryiğit

11:40–12:00   *ÚFAL CorPipe at CRAC 2023: Larger Context Improves Multilingual Coreference Resolution*
Milan Straka

12:00–12:20   *McGill at CRAC 2023: Multilingual Generalization of Entity-Ranking Coreference Resolution Models*
Ian Porada and Jackie Chi Kit Cheung

Closing of the Workshop

12:20–12:30   *Closing Remarks*