LREC-COLING 2024

The First Workshop on Holocaust Testimonies as Language Resources HTRes@LREC-COLING-2024

Workshop Proceedings

Editors

Isuri Anuradha, Martin Wynne, Francesca Frontini, Alistair Plum

21 May, 2024
Torino, Italia

Proceedings of the First Workshop on Holocaust Testimonies as Language Resources: HTRes@LREC-COLING-2024

Introduction

Holocaust testimonies serve as a bridge between people today and history's darkest chapters, providing a connection to profound experiences of survivors. Testimonies stand as the primary source of information that describe the Holocaust, offering first-hand accounts and personal narratives of those who experienced it. The majority of testimonies are captured in an oral format, as survivors vividly explain and share their personal experiences and observations from that time period.

The creation of accessible, comprehensive, and well-annotated Holocaust testimony collections is of paramount importance to our society. However, transforming Holocaust testimonies into a machine-processable digital format and publishing according to the standards and best practices of spoken and oral corpora can be a difficult task. At the same time these collections empower researchers and historians to validate the accuracy of socially and historically significant information, enabling them to share critical insights and trends derived from these data.

This workshop aimed to investigate how language technologies and, in particular, techniques and tools from natural language processing and corpus linguistics can greatly contribute to the exploration, analysis, dissemination and preservation of Holocaust testimonies.

The workshop was open to diverse communities and disciplines. The archivists who curate and digitize archives, producing digital editions that may also be enriched with various levels of annotation and structured knowledge; the NLP experts who process and in some case also make use of testimonies to develop specialised language models for speech and text processing; and finally historians who analyse the digital archives, via interfaces or using advanced text mining techniques, to discover new insights, or, for example, to retrace the biographical trajectories of witnesses.

The accepted papers touch upon all the aforementioned topics. Del Grosso et al and Beniere present two cases of publication of digital editions making use of the TEI standard, while Dermentzi et al show how the same manually curated editions can be used to train NLP models (in this case for NER). Flinz and Leonardi turn their attention in particular to the treatment of dates and places, while Liu and Mattingly (in a paper presented at the conference but not published in these proceedings) show how a typology of places can be extracted using NLP techniques. Draxler et al illustrate how the use of speech technologies can facilitate the curation of oral history materials, while Ifergan et al show how narratives can be analysed with the use of topic modelling. Vitali and Brazzo show how structured knowledge can be used to identify deportation trajectories, while, finally, Wagner et al show how these trajectories can also be extracted automatically in a structured format, and then visualised, interrogating a Large Language Model with a zero shot approach.

All these contributions illustrate well how techniques and best practices from the NLP and Language Resources and Technology domains are now mature enough to be relatively easily applied to Holocaust testimonies and readily applied to enlighten new aspects and dimensions of research. At the same time these testimonies can be language resources of great value, providing language technologists with a testbeds for domain adaptation of existing models and applications. The workshop was sparked by a collaboration between CLARIN, the Language Resources and Technology Iinfrastructure, and EHRI, the European Holocaust Research infrastructure. This is the third in an ongoing series of collaborative workshop, following on from earlier ones held at King's College London in 2023 (Making Holocaust Oral Testimonies More Usable as Research Data) and Charles University Prague in 2024 (Natural Language Processing Meets Holocaust Archives), and the first to take place attached to a conference

devoted to NLP or language resources and tools.

Two keynote speakers were invited by the two infrastructures, Michal Frankl, coordinator of EHRI in the Czech Republic, highlighted in his talk the requirements of historians in terms of digital resources, while Silvia Calamai, member of the CLARIN-IT national consortium and one of the coordinators of the CLARIN funded Voices from Ravensbruck project showed how holocaust testimonies can be considered as important language resources also for language studies, in particular for their value as multilingual corpora documenting very interesting sociolinguistic and variety aspects.

Organising Committee

Martin Wynne, Oxford University, UK
Francesca Frontini,CNR-Istituto di Linguistica Computazionale "A. Zampolli", Italy & CLARIN ERIC
Isuri Anuradha, Lancaster University, UK
Alistair Plum, University of Luxembourg, Luxembourg
Paul Rayson, Lancaster University, UK
Ingo Frommholz, University of Wolverhampton, UK
Ruslan Mitkov, Lancaster University, UK

Programme Committee

Angelo Mario Del Grosso, CNR-Istituto di Linguistica Computazionale "A. Zampolli", Italy
Arjan van Hessen, Radboud University
Estelle Bunout, University of Luxembourg, Luxembourg
Eveline Wandl-Vogt, Austrian Academy of Sciences, Vienna
Federico Boschetti, CNR-Istituto di Linguistica Computazionale "A. Zampolli", Italy
Gabor Toth, University of Luxembourg, Luxembourg
Henk van den Heuvel, Radboud University & CLARIN ERIC
Ian Gregory, Lancaster University, UK
Ignatius Ezeani, Lancaster University, UK
Jan Svec, University of West Bohemia
Johannes-Dieter Steinert, University of Wolverhampton, UK
Le An Ha, University of Wolverhampton, UK
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences, Poland
Maciej Piasecki, Wroclaw University of Science and Technology, Poland
Maria Dermentzi, King's College London, UK
Martin Bulin, University of West Bohemia, Czech Republic
Patricia Murrieta-Flores, Lancaster University, UK
Rachel Pistol, King's College London, UK
Stefanie Rauch, Wiener Holocaust Library, UK
Renana Keydar, The Hebrew University of Jerusalem, Israel
Tim Cole, University of Bristol, UK

Invited Speakers

Michal Frankl, Masaryk Institute and Archives of the Czech Academy of Sciences
Silvia Calamai, University of Siena, Italy

Table of Contents

Workshop Program

21 May 2024

9.30–9.45   **Welcome and Introduction**

9:45–10:30   **First Keynote Speech: Talking about Holocaust research (what type of digital resources they want? And why?) by Michal Frankl**

10:30–11:00   **Coffee Break**

11.00–12:00   **First session: Editions and their exploitation**

The Impact of Digital Editing on the Study of Holocaust Survivors' Testimonies in the context of Voci dall'Inferno Project
Angelo Mario Del Grosso, Marina Riccucci and Elvira Mercatanti

TEI Specifications for a Sustainable Management of Digitized Holocaust Testimonies
Sarah Bénière, Floriane Chiffoleau and Laurent Romary

Repurposing Holocaust-Related Digital Scholarly Editions to Develop Multilingual Domain-Specific Named Entity Recognition Tools
Maria Dermentzi and Hugo Scheithauer

12.00–13.00   **Second paper session: Dates and Places**

Dates and places as points of attachment for memorial contents in the ISW corpus: 1938 as a turning point
Carolina Flinz and simona leonardi

Creating a Typology of Places to Annotate Holocaust Testimonies Through Machine Learning
Christine Liu and William J.B. Mattingly

21 May 2024 (continued)

13.00–14.00	**Lunch Break**

14.00– 15.15   **Second Keynote Speech: Silvia Calamai, Università degli Studi di Siena, member of the CLARIN-IT consortium The Voices from Ravensbruck project: Bringing together what is dispersed**

15.15– 16.00   **Third session: Testimonies and Narratives**

Speech Technology Services for Oral History Research
Christoph Draxler, Henk van den Heuvel, Arjan van Hessen, Pavel Ircing and Jan Lehečka

Identifying Narrative Patterns and Outliers in Holocaust Testimonies Using Topic Modeling
Maxim Ifergan, Omri Abend, Renana Keydar and Amit Pinchevski

16.00– 16.30   **Coffee Break**

16.30– 17.10   **Fourth session: Traces and networks**

Tracing the deportation to define Holocaust geometries. The exploratory case of Milan
Giovanni Pietro Vitali and Laura Brazzo

Zero-shot Trajectory Mapping in Holocaust Testimonies
Eitan Wagner, Renana Keydar and Omri Abend

21 May 2024 (continued)

17.10–
17.50
     Pannel Discussion

17.50–
18.00
     Final Remark