

2 The Workshop

This workshop aims to take stock of our efforts to break down disciplinary boundaries by offering scholars the opportunity to apply, experiment and exchange digital tools that have been developed in the realm of Digital Humanities. We would also like to share our insights with a wider audience of scholars coming from different disciplines. Point of departure is a webportal that we have developed for automatic speech to text conversion (transcription) by Automatic Speech Recognition (ASR).

This half-day workshop is open for everyone interested in the processing of spoken narratives for research purposes (scholars) or for archiving and opening up interview data (archivists). It will provide cross-disciplinary knowledge exchange between all participants by presenting the following elements:

- Outline of the various approaches of different disciplines that work with spoken data and how they use annotation, text analysis and emotion extraction tools to process their data [6]
- Outline and hands-on components of the open source tool the T-Chain (Fig. 2). Participants can process their own audio-clips in Dutch, German, Italian or English. and explore how to further process the ASR output [7, 8]
- Connect the ASR-results to existing written text sources (books, newspapers, blogs)
- Short lectures by invited speakers
- Figure out how to strengthen and expand the community around Interview Data and Technology

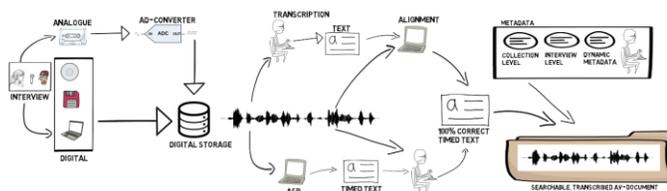


Figure 2: schematic view of the T-Chain: processing of spoken documents from recording/digitization into transcribed digital files including metadata

3 Workshop website and user participation

The workshop has a maximum of 25 participants. The publicity, interaction (questions, remarks, registration with special wishes/questions concerning the workshop) with the potential participants, will go through the website:

<https://oralhistory.eu/workshops/icmi>. Participants who are interested in contributing to the workshop, can submit their proposals via the same website.

The organizers will ask at least two workshop participants to give a presentation about their own working methods,

experiences, and wishes regarding the elaboration of the interview process. The selection will be made by the organizers.

4 Programme

9:00 – 9:30	Introduction and short presentation on ‘Digital Humanities approaches to interview data - can historians, linguists and social scientists share tools?’
9:30 – 11:00	Preparing your audio-data, uploading the audio to the portal and automatic recognizing the speech. Correcting the ASR-results Downloading the (corrected) results and improving the readability
11:00 – 11:15	Coffee break
11:15 – 11:45	Invited speaker 1 incl. QA
11:45 – 12:15	Invited speaker 2 incl. QA
12:15 – 12:45	General discussion and close of meeting

ACKNOWLEDGMENTS

The conference organisers gratefully acknowledge the support by the European CLARIN initiative.

REFERENCES

- [1] Oral History and Technology Collaboration. 2018. Oral History and Technology group. Website: <https://oralhistory.eu>.
- [2] L. Corti, N. Fielding, 2016. Opportunities From the Digital Revolution: Implications for Researching, Publishing, and Consuming Qualitative Research. SAGE Open. <https://doi.org/10.1177/2158244016678912>.
- [3] B. A. Lanman and L. M. Wendling, 2006. Preparing the Next Generation of Oral Historians: An Anthology of Oral History Education. Altamira Press.
- [4] F. de Jong and A. van Hessen and T. Petrovic and S. Scagliola, 2014. Croatian Memories: speech, meaning and emotions in a collection of interviews on experiences of war and trauma, Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).
- [5] S. Scagliola and L. Corti, 2018. Oral History under scrutiny in München - Cross disciplinary overtures between linguists, historians and social scientists, CLARIN website Blog.
- [6] K. P. Truong and G. J. Westerhof, and S. Lamers and F. de Jong and A. Sools, 2013. Emotional expression in oral history narratives: comparing results of automated verbal and nonverbal analyses, Proceedings of the Workshop on Computational Models of Narrative, CMN 2013, 2013 Workshop on Computational Models of Narrative. CMN 2013 Hamburg, Germany.
- [7] H. Van den Heuvel and A. van Hessen and S. Scagliola and C. Draxler, 2017. Transcribing Oral History Audio Recordings – the Transcription Chain Workflow. Poster at the CLARIN EU Conference, Budapest, September 18/19- 2017.
- [8] C. Draxler and H. Van den Heuvel and A. van Hessen and S. Calamai and L. Corti and S. Scagliola, 2020. A CLARIN Transcription Portal for Interview Data. Proceedings of LREC 2020, Marseille