



Transcribing Oral History Recordings Using the Transcription Portal

Christoph Draxler¹, Julian Pömp¹, Henk van den Heuvel², Fabio Ardolino³, Arjan van Hessen⁴

¹Institute of Phonetics and Speech Processing, LMU Munich, Germany

²Centre for Language Studies, Radboud University, Netherlands

³Dipartimento di Filologia e Critica delle Letterature Antiche e Moderne, Siena University, Italy

⁴Utrecht University, Netherlands

draxler|j.poemp@phonetik.uni-muenchen.de, H.vandenHeuvel@let.ru.nl,
fabio.ardolino@unisi.it, A.J.vanHessen@uu.nl

Abstract

The Transcription Portal is a web-based service for multilingual orthographic transcription of speech, notably Oral History interviews. It is targeted at non-technical users: it provides a simple and intuitive GUI, supports several languages, and the workflow is pre-configured. Currently, the workflow consists of three steps: 1) automatic speech recognition, 2) manual correction of the transcript, 3) data export. Summarization and translation are planned. We demonstrate the portal on a set of historical Italian interviews on the Ravensbrück concentration camps.

Index Terms: Oral History, multilingual speech recognition, user interface

1. Transcription Portal

1.1. Introduction

A large body of material in Oral History (OH) consists of interviews – often in highly dialectal speech without standard orthography, in a more or less formal style, and recorded in undocumented locations or noisy environments and on analog media. Where a standard orthography is available, word error rates > 50% are typical even for the most advanced ASR systems; for dialects without orthography, performance is even worse.

For reliable transcriptions, the results of automatic speech recognition (ASR) have to be checked manually using transcription editors. However, most such editors were designed for linguistic or phonetic speech data, i. e. for multi-tier annotation of relatively short recordings, and as standalone applications with a steep learning curve. In other domains, e. g. journalism or market research, transcription editors with an easy-to-use UI were developed, but they often lack the necessary features for scientific speech processing, i. e. word-based time-alignment or access to external signal processing tools or services.

The Transcription Portal makes available current speech recognition technology in a seamless combination with manual transcription to non-technical users. Thus, it provides a simple GUI based on a table metaphor with drag&drop functionality. Further key features are giving users the choice of service providers, and access to innovative SLP and NLP services.

1.2. Walkthrough of the Transcription Portal UI

To use the Transcription Portal open <https://clarin.phonetik.uni-muenchen.de/apps/TranscriptionPortal/>. Users must authenticate as members of an academic research institution. The Transcription Portal displays a 1-2-3 Quickstart page (fig. 1).

1. Drag the audio files from the local file system to the Quickstart page and check the tasks from the workflow (fig. 2).

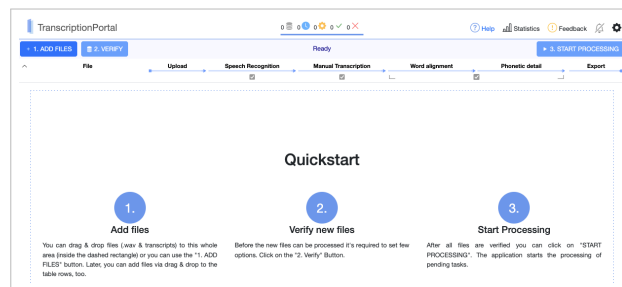


Figure 1: 1-2-3 Quickstart page showing the sequence of tasks in the workflow

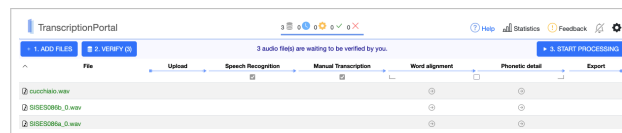


Figure 2: View the uploaded files and select the workflow tasks by setting the check marks in the column heads

2. Select the input language (currently, Dutch, English, German, and Italian are supported) and the ASR provider; both commercial and academic providers are available (fig. 3). Click on the provider logo to display a synopsis of the main terms of use and privacy policy.
3. Click on "3. Start Processing" to upload the files and start the selected tasks. A turning cog wheel icon in the table cell shows that this task is currently running, a check mark indicates that it has finished successfully (fig. 4).
4. Click on a check mark icon to display a preview of the outcome of the current task, e. g. the raw transcript as generated by the ASR (fig. 5). The preview formats available include plain text, CSV, TextGrid, SRT, VTT, and a JSON format.
5. If manual transcription has been selected as part of the workflow, the transcription editor Octra opens inside the window, showing the current audio waveform and transcript (fig. 6).
6. Export the transcript to the local file system: select the desired output format and click on the download button (fig. 7).

Currently, the Portal uses locally installed versions of whisper and Fraunhofer ASR, Radboud University's ASR web service, plus limited versions of Google, IBM and EML ASR. Automatic word alignment via the MAUS web service is an option in the manual correction of the transcript, or as a separate step in the workflow [1].

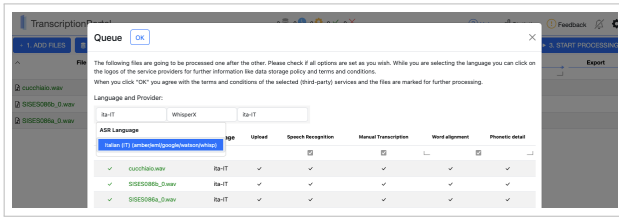


Figure 3: Select both the language of the audio recording and the ASR provider

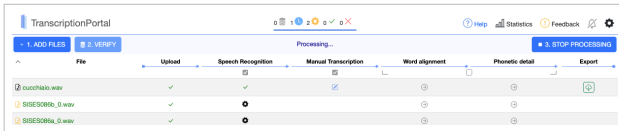


Figure 4: View current state of a task

2. Italian interviews

2.1. Content of the recordings

The recordings used in this study are part of the oral archive compiled by Anna Maria Bruzzone, a secondary school teacher and independent historian, whose work focused extensively on memory studies, OH, and female political deportation during the Nazi regime. The specific interviews considered here were conducted in 1976, using a portable cassette recorder, during visits to the private homes of the interviewees. The corpus includes four interviews, featuring five Italian women who survived the Ravensbrück concentration camp. The total duration of the recorded material is approximately 18 hours of spoken discourse, which comprises detailed autobiographical accounts, personal reflections, and valuable linguistic and communicative insights into the experience of deportation and internment [2].

2.2. Digitization

The digitization process was carried out by the Centro di Sonologia Computazionale (CSC) at the University of Padua. Each analog tape was digitized to preservation master files in uncompressed .wav format (96000 Hz, 24-bit), alongside digital images of the original media and detailed metadata documenting the technical specifications and the transfer process.

2.3. Availability

CLARIN ERIC has featured a specific resource family page for OH recordings: <https://www.clarin.eu/resource-families/oral-history-corpora>. The corpora of interviews on this page have been selected

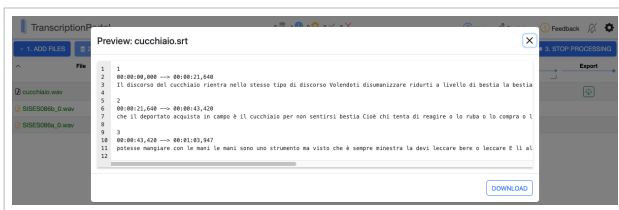


Figure 5: Preview raw ASR-generated transcript (here: SRT-format)

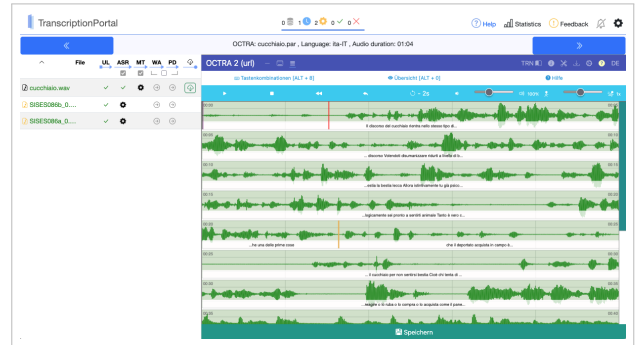


Figure 6: Manual correction of the transcript using the embedded transcription editor

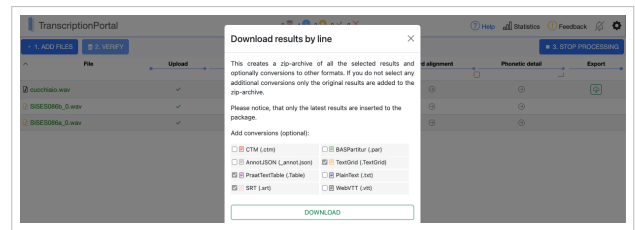


Figure 7: Final export of transcripts to different formats

with the aim of offering the possibility to compare data that covers similar topics [3]. The digitized Ravensbrück OH corpus is accessible through The Language Archive (TLA), hosted by the Max Planck Institute for Psycholinguistics in Nijmegen: https://archive.mpi.nl/tla/islandora/object/tla%3A1839_76375cde_a68e_4c87_8539_513c3a63e308.

3. Future Work

The Transcription Portal is further developed in the ATRIUM project (<https://atrium-research.eu/>). In this project we will add transformer-based ASR services to replace commercially based ASR engines now provided in the portal, and we will support additional languages (e.g. Spanish and Czech). We will include improved speaker diarization facilities by comparing standard pyannote whisper solutions with e.g. NVIDIA Nemo. Further additions will include translation of transcripts (typically into English) and summarization, all based on open source (large) language models.

4. References

- [1] T. Kisler, F. Schiel, and H. Sloetjes, "Signal Processing Via Web Services: The Use Case Webmaus," in *Proceedings Digital Humanities*, Hamburg, 2012, pp. 30–34.
- [2] S. Calamai, S. Scagliola, F. Ardolino, C. Draxler, A. van Hossen, and H. van den Heuvel, "Ravensbrück Interviews: How to Curate Legacy Data to Make it CLARIN Compliant," in *Selected Papers from the CLARIN Annual Conference 2021*, ser. Linköping Electronic Conference Proceedings 189, 2022.
- [3] F. de Jong, D. Van Uytvanck, F. Frontini, A. van den Bosch, D. Fišer, and A. Witt, "The European Research Infrastructure CLARIN, Today and Tomorrow," in *CLARIN: The Infrastructure for Language Resources*, ser. Digital Linguistics. De Gruyter, 2022, vol. 1, pp. 31–57. [Online]. Available: <https://doi.org/10.1515/9783110767377>