Voices from Ravensbrück. Towards the creation of an oral and multilingual resource family

Silvia Calamai
Università di Siena
Siena, Italy
silvia.calamai@unisi.it

Jeannine Beeken
University of Essex
Colchester, United Kingdom
jeannine.beeken@essex.ac.uk

Max Broekhuizen
Erasmus School of History
Rotterdam, The Netherlands
maksbroekhuizen@gmail.com

Christoph Draxler
Ludwig Maximilian University
Munich, Germany
draxler@phonetik.uni-muenchen.de

Arjan van Hessen
University of Twente
Enschede, The Netherlands
a.j.vanhessen@utwente.nl

Henk van den Heuvel
Radboud University
Nijmegen, The Netherlands
h.vandenheuvel@let.ru.nl

Stefania Scagliola
University of Luxembourg
Luxembourg
scagliolas@gmail.com

Abstract

The paper describes the main features of a planned pilot project aimed at creating a new corpus type entitled 'narratives' in the CLARIN resource family tree. For this, a new multilingual corpus will bring together and curate existing interviews of survivors of the Ravensbrück concentration camp. This corpus will be unique in that its interviews cover the same topic, are narrated in different languages, involve the same interaction style, are enriched with transcripts and other types of annotation, and that it will be curated according to CLARIN compliant standards.

1 CLARIN Resource families and oral history

The Clarin Resource Family is a user-friendly overview per data type of the available language resources in the CLARIN infrastructure aimed at the needs of researchers from digital humanities, social sciences and human language technologies.[1] Within the corpus group of resources, there is only one entry for spoken corpora, which contains 90 data sets mainly targeted at phonetic, linguistic and speech technology research. We argue that a new type of entry is needed: 'narratives', covering oral history interviews and other types of narrative discourse. Interviews, aside from oral history, are a central object of research in a broad variety of fields such as anthropology, psychology, literary studies, sociology, health studies, education, linguistics and cognitive science, but there seem to be scarce opportunities for comparison and cross-fertilisation.

It is our contention that oral history interviews represent an under-utilized but potentially promising data outside the realm of history, and that the CLARIN infrastructure is the ideal 'house' for this type of 'family' to illustrate its multidisciplinary potential. Our curation project aims at probing this position, by bringing together similar interview data – retrospective spoken narratives of women – from different countries and languages – about the same topic: experiences of war and trauma in concentration camp Ravensbrück. We will first discuss the characteristics of oral history and opportunities for digital humanities approaches, then elaborate on the data that is currently available and how it will be

---

[1] https://www.clarin.eu/resource-families

complemented with other languages. Finally, we describe the workflow that we envision for collecting, enriching and publishing this type of material.

2 Oral History

The key characteristics of oral history is that its content is *multilayered*, that it can be appreciated in *multimodal* ways (i.e., seeing the interview, hearing the recording, reading the transcript), and that it is *co-created*.

An oral history interview offers information about a personal past, but from the perspective of interactions within a social context (e.g., the family, the village, the military unit, the company). The interviewer elicits memories of the past, and the composition of what is uttered reveals the social, economic, political and cultural dimensions of someone's life. Aside from what is said, one can reflect on what is not said, what is repeated again and again, or how the narrative is organized.

A different type of interaction than the one that occurred in the past, is the one that takes place during the interview: the mediating process between interviewer and interviewee who together are co-creating an historical source. Typical questions are: What is the relation between narrator and interviewee during the interview process? Does their background play a role? Do they accommodate their speech with each other? How is the relation between standard and vernacular speech in the course of the interview?

With regard to modalities, we have two representations of the encounter: (1) the audio-signal, which holds not only the verbatim content, but also speaker-specific physiological information which can only be accessed in an objective way using digital tools (segmental and suprasegmental acoustics, silence, emotions, speech rate and timing), and (2) the textual representation of what is uttered, which is characterized by the transcription convention. Both kinds of data have value as singular source, as well as in the form of a collection (similarities and differences).

With regard to ethical and legal issues and re-use of data, oral history data has a different background than spoken corpora that are specifically created to study language. This is evident when considering recordings that were made before the digital era, on the basis of consent for traditional use of the source, which meant consulting it at the premises of the archive under guidance of a responsible archivist. The web has profoundly changed the practice of consulting, sharing, copying, editing, and giving credits to this type of data. But there are still gaps between the archival realm, where massive amounts of spoken accounts can be found, and the realm of digital humanities, where such data can be accessed and processed in new ways.

The issue is that the traditional way of treating the interview as a source is limited to a one-to-one encounter: you can only listen to one signal or read one transcript at a time. This can be facilitated at a distance via the web in a secure way, but in order to detect signals or patterns we need access to sets of files processed with software. In a way we are taking Alessandro Portelli's adagium to 'bring back orality to oral history' one step further. He pledged for shifting the focus of research from the transcript to the auditive features of the source: pay more attention to tone and rhythm of a voice (Portelli, 1981).

We strive to go back to the signal itself and discover patterns in speech and non-verbal features with the help of digital tools. Technically this does not pose problems; a digital copy can be made at once, but with regard to privacy, copyright and access control, special measures have to be implemented to guarantee that the interviews will be treated in a respectful way (Calamai et al. 2019). Once this becomes standard practice for scholars, it will open up a 'bonanza' of data that is waiting to be understood in 'new ways'.

3 The historical context of Ravensbrück and its suitability for our objective

Ravensbrück was a German concentration camp built in 1939 located in northern Germany, 90 km north of Berlin. It stands out as the only camp with only female prisoners, all with very different backgrounds. It was intended for social outcasts, Romani people, political dissenters, foreign resistance fighters, the disabled and other so-called 'inferior beings'. Some 130,000 women from 20 different nationalities passed through it: about 48,500 from Poland, 28,000 from the Soviet Union, 24,000 from Germany and Austria, 8,000 from France, and thousands from other countries. More than 20,000 among them were Jewish, and 80 percent were political prisoners. Many forced labour prisoners were employed by Siemens & Halske. From 1942 to 1945, medical experiments were undertaken as well.

Narratives about being imprisoned in Ravensbrück describe a unique experience, which is limited to one and the same location, and one and the same time period. Moreover, it is gender specific. What makes this type of narrative the ideal data for comparison is that the camp was populated by women from all over Europe. They returned to their countries and moulded their recollections of the past within the specific social-cultural context of that country.

Although a number of historical studies have been published about Ravensbrück, the donation of the entire oral archive of Anna Maria Bruzzone to the University of Siena in 2016 opened up an opportunity for cross-disciplinary multilingual research, with a more specific set of variables that can be compared with data from the many oral history projects that have been set up about experiences in Ravensbrück in other countries.

4   Data that is currently available

Web research and consultation of a number of authors has yielded enough data and commitment to be able to search for profiles that are similar to those of the five narrators in Bruzzone's archive. We intend to first complement the Italian interviews with interviews in English, Dutch and German. In the long run we intend to expand the project to Eastern Europe, especially to Poland and Russia, where many survivors came from.

Bruzzone's Ravensbrück interviews consist of 14 audio cassettes, with a total duration of about 18 hours and 20 minutes. The analogue audio cassettes were digitized according to IASA standards (.wav format, 96000 Hz, 24 bit). The archive contains 4 long interviews with 5 survivors. For her publication, Anna Maria Bruzzone transcribed the recordings step by step, writing everything she was hearing (unfortunately, the handwritten transcriptions were lost [2]). Later, the transcripts were adapted for printing. In 2016, the book was translated to German.

For Dutch, we have access to interviews held with Dutch Ravensbrück internees between 2007 and 2010, for a PhD study on the memory of Ravensbrück by Susan Hogervorst (2010). In case this material may pose legal problems, due to the absence of consent from narrators who have deceased, we can fall back on the many project-generated interview collections that are publicly available in archives. These interviews have been created in the wake of the 50th anniversary of the second World War from the 1990s onwards, and their proliferation has been strongly influenced by the availability of digital technology and a push towards presence on the web. In the Netherlands this has resulted in the online resource *Getuigenverhalen*[3] which contains 3 interviews on Ravensbrück. The Visual History Archive contains 5 interviews in Dutch about Ravensbrück. The United States Holocaust Memorial Museum contains 1 interview in Dutch about Ravensbrück.

With regard to interviews in German, we have access to the *Videoarchiv "Die Frauen von Ravensbrück"* (50 interviews), *Österreichische Lagergemeinschaft Ravensbrück und Freundinnen* (34 interviews). In English, we have the *Visual History Archive* (20 interviews) and the *Imperial War Museum GB* (8 interviews).

We are confident that in some of the countries involved in the CLARIN network, additional similar material might be found, in further languages.

5   Towards a CLARIN Resource Family for Oral History

The planned project consists of two phases, *curation* and *expansion*. In the curation phase, five Italian interviews from the Bruzzone collection, will be revised, we will prepare the hosting of this corpus in a CLARIN Centre and generate the metadata according to an appropriate CMDI profile for oral history. From this perspective, data that was created within a historical framework will be described in a format and structure that adheres to CLARIN methodology. Furthermore, we will search for other existing collections of oral history interviews on Ravensbrück in other languages, notably Dutch, English and German, with the goal of collecting data that is comparable to the Italian data – with regard to length (about 15-20 hours), amount of narrators (4-6), and interview approach (chronological structured in-depth interview probing for details).

---

[2] Hopefully, with digital repositories this will never happen again.
[3] http://getuigenverhalen.nl/home

In the expansion phase, we will curate and enrich the collected data by adding transcriptions, time stamps at word level, phonetic and suprasegmental information. We will also add annotations on e.g. the use of specific language, emotions expressed in non-verbal modes of communication (i.e., laugh, pauses and silences, breathing). Aside from exploring the content, we will also document the legal and technological challenges of reuse of interview data to create a new collection. Working on legacy data means indeed testing the feasibility of the re-use of AV data, within the framework of FAIR open data, taking into account the particularity of the interview as source of knowledge, and identifying technological and legal obstacles for the dissemination of such material.

The completion of these tasks will yield a unique resource family of oral history interviews, curated along CLARIN compliant standards, covering the same topic, narrated in different languages, and enriched with transcriptions and other types of annotation (e.g. pauses and word durations for emotion analysis).

Summary

The 'Voices from Ravensbrück' project will compile and make accessible a unique multilingual corpus for oral history research. It will demonstrate that interview data can – and should be – reused and interpreted in different ways with the help of digital tools. Furthermore, it will make available an important source of oral history material to other research fields such as linguistics, sociolinguistics, phonetics, and speech technology, fostering the exchange between these fields – there is a lot these fields can learn from each other. Finally, by including the corpus in the CLARIN Resources Family, both the visibility of these resources as a whole, and the individual corpora within, will increase.

References

Beccaria Rolfi, L., Bruzzone, A.M. 2020. *Le donne di Ravensbrück. Testimonianze di deportate politiche italiane*, Torino, Einaudi.

Beccaria Rolfi, L., Bruzzone, A.M. 2016. *Als Italienerin in Ravensbrück. Politische Gefangene berichten über ihre Deportation und ihre Haft im Frauen-Konzentrationslager*, Herausgegeben von Johanna Kootz, Metropol Verlag Berlin.

Calamai, S., Kolletzek C., Kelli, A. 2019. *Towards a protocol for the curation and dissemination of vulnerable people archives*. Selected papers from the CLARIN Annual Conference 2018, Pisa, 8-10 October 2018, edited by I. Skadina, M. Eskevich, 2018, Series: Linköping Electronic Conference Proceedings Issue, Linköping University Electronic Press, Linköpings universitet: 28-38 https://ep.liu.se/ecp/159/003/ecp18159003.pdf.

Draxler, C., Van den Heuvel, H., Van Hessen, A., Calamai, S., Corti, L., and Scagliola, S. 2020. A CLARIN *Transcription Portal for Interview Data. Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC2020)*. pp. 3346-3352 http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.411.pdf

Hogervorst, S. *Onwrikbare Herinnering. Herinneringsculturen van Ravensbrück in Europa*. 2010. Uitgeverij Verloren, Hilversum.

Portelli, C. 1981. *On the peculiarities of oral history.* History Workshop Journal, 12: 96-107. https://doi.org/10.1093/hwj/12.1.96

Scagliola, S., Corti, L., Calamai, S., Karrouche, N., Beeken, J., van Hessen, A., Draxler, Chr., van den Heuvel, H., Broekhuizen, M., and Truong, K.. 2020. *Cross disciplinary overtures with interview data: Integrating digital practices and tools in the scholarly workflow*. In: Simov, K., & Eskevich, M.: Selected Papers from the CLARIN Annual Conference 2019 Leipzig, 30 September - 2 October 2019. Linköping Electronic Conference Proceedings 172:15, 126-136. https://doi.org/10.3384/ecp2020172.