Proceedings of LR4SSHOC: Workshop about Language Resources for the SSH Cloud, pages 42–44
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Crossing the SSH Bridge with Interview Data

Henk van den Heuvel
CLS/CLST, Radboud University
Erasmusplein 1, Nijmegen, the Netherlands
h.vandenheuvel@let.ru.nl

Abstract

Spoken audio data, such as interview data, is a scientific instrument used by researchers in various disciplines crossing the boundaries of social sciences and humanities. In this paper, we will have a closer look at a portal designed to perform speech-to-text conversion on audio recordings through Automatic Speech Recognition (ASR) in the CLARIN infrastructure. Within the cluster cross-domain EU project SSHOC the potential value of such a linguistic tool kit for processing spoken language recording has found uptake in a webinar about the topic, and in a task addressing audio analysis of panel survey data. The objective of this contribution is to show that the processing of interviews as a research instrument has opened up a fascinating and fruitful area of collaboration between Social Sciences and Humanities (SSH).

Keywords: SSH, interview data, automatic speech recognition, NLP, spoken language processing

1. Introduction: Cross Disciplinary Use of Interview Data

Spoken audio data, such as interview data, is a scientific instrument used by researchers in various disciplines. These disciplines span the social sciences and the humanities. An oral historian will typically approach a recorded interview as an intersubjective account of a past experience, whereas another historian might consider the same source of interest only because of the factual information it conveys. A social scientist is likely to try to discover common themes and similarities and differences across a whole set of interviews, whereas a computational linguist will rely on counting frequencies and detecting collocations and co-occurrences, for similar purposes. On the other hand sociologists who interview, often seek to understand their interviewees in the same way as (oral) historians (Scagliola et al., 2020).

Then the question arises how the various disciplines can benefit from the large amount of freely available transcription, annotation, linguistic and emotion recognition tools. We should take into account that most scholars are not familiar with each other's approaches, and hesitate to take up technology. When software is used, it is often proprietary and binds scholars to a particular set of practices.

To clear the situation a multidisciplinary international community of experts organised a series of hands-on workshops with scholars who work with interview data, and tested the reception of a number of digital tools that are used at various stages of the research process. We engaged with tools for transcription, for annotation, for analysis and for emotion recognition. The workshops were held at Oxford, Utrecht, Arezzo, Munich, Utrecht and Sofia between 2016 and 2019, and were mostly sponsored by CLARIN. Participants were recruited among communities of historians, social science scholars, linguists, speech technologists, phonologists, archivists and information scientists. The website https://oralhistory.eu/ was set up to communicate across disciplinary borders. For a full account of experiences, we refer to Scagliola et al., (2020).

Through these workshops it became ever clearer that, despite different scientific methods of analysis used by these researchers, core processing methods of this kind of data are cross-disciplinary. Creating transcriptions with appropriate level of detail is one of the initial and most important steps in the spoken audio data analysis, but this step can also be very time-consuming. This is why researchers can greatly benefit from at least partial automation of the transcription process. However, choosing high-quality tools and learning how to use them is not always a straightforward process, and researchers can quickly lose their enthusiasm for automation for the fear of that the automation process might be too complex or non-transparent.

In this paper, we will have a closer look at a portal designed to perform speech-to-text conversion on audio recordings of interviews through Automatic Speech Recognition (ASR) with an option to manually correct the text output (section 2). Then we will point to a number of options to apply NLP analysis tools to the resulting text (section 3), and finally we will address activities organized in the SSHOC project[1] to set up a bridge spanning the SSH communities using ASR for recorded audio materials (section 4).

2. Interview Data and ASR

Automatic speech recognition (ASR) has reached a performance level where, under favorable acoustic conditions, a quality of transcriptions can be achieved that is a sufficient starting point for many researchers to start subsequent (domain specific) text analysis (labelling and encoding on). An additional advantage of using ASR for transcription purposes is that the output comes with time stamps of the words locating them in the original audio

[1] https://sshopencloud.eu/

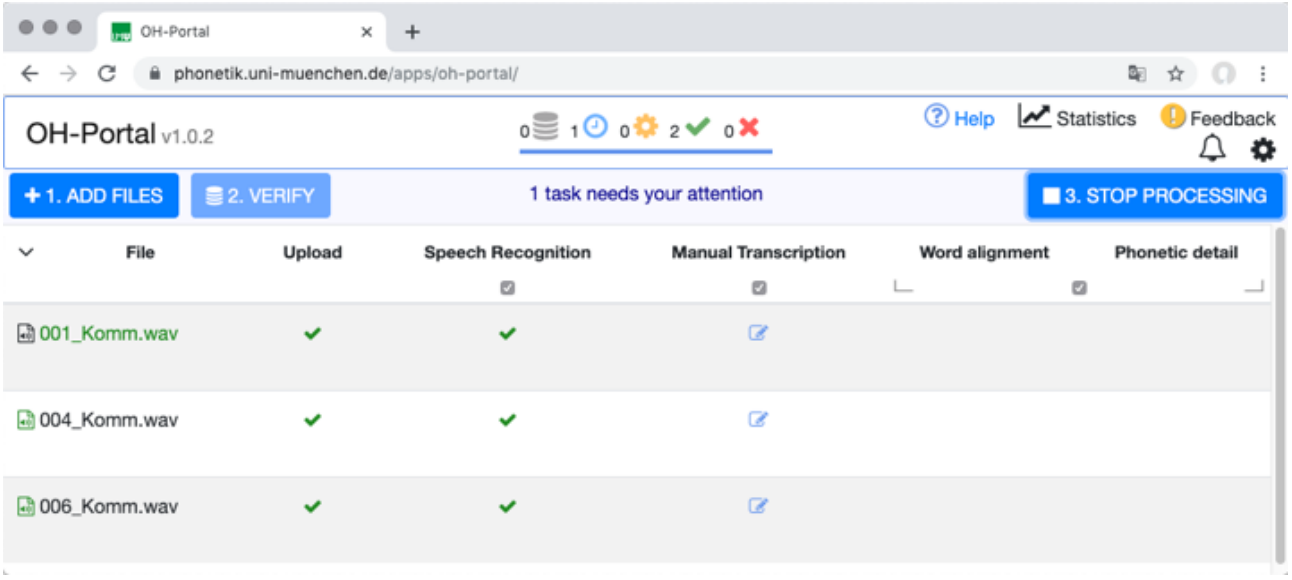

Figure 1. Screenshot of OH Portal with three audio files. The files were uploaded and processed by ASR, and are now awaiting manual correction of the transcript.

stream and permitting seamless subtitling of audio and video recordings.

Draxler et al. (2020) describe a webportal developed for the CLARIN ERIC[2] where researchers can upload audio recordings, use ASR engines for a variety of languages to obtain text transcriptions of the recordings, and to manually correct the transcriptions and realign the corrected transcripts with the audio files. The portal is accessible via a login at https://clarin.phonetik.uni-muenchen.de/apps/oh-portal/. Upon entering the portal the user sees the screen depicted in Figure 1, showing the three phases in the transcription process mentioned above. Draxler et al. (2020) gives a detailed account of the various processing steps, the user agreements for the available speech recognisers (also with respect to privacy issues), the technical limitations of the portal, the performance one may expect, and guidelines for making audio recordings that are suitable for ASR processing.

3. From ASR output to NLP

As pointed out in Draxler et al. (2020) we are well aware of the relevance of tools for follow up analyses after the speech to text conversion in the portal. The current workflow implemented by the OH portal is derived from the requirements of speech technology development. However, the requirements of oral historians but also of humanities scholars and social scientists are different. Studying the interaction between two people who construct meaning via a dialogue, requires retrieving high-level information from the recordings, it is not only about 'what is said' but also about 'how it is said'. Scholars want to know: what is the major topic of the recording, what emotions can be observed, what are the named entities, what can be said about the regional background of the speaker, what relationships exist between historical data and audio recordings, etc. Trained human transcribers may extract this information, but this is a time-consuming manual process. Topic modelling, sentiment analysis, named entity recognition, dialect modelling and information extraction or summarization are all active research areas in computational linguistics and speech processing.

In Scagliola et al. (2020) we presented an overview of NLP analysis packages used in the workshops. These include lemmatizers, syntactic parsers, named entity recognizers, auto-summarizers, tools for detecting concordances/n-grams and semantic correlations. Participants were then given a live demo of the software tools and then some step by step guided exercises with data. Linguistic tools introduced were

- Voyant (https://voyant-tools.org/), a lightweight text analysis tool that yields output on the fly
- the Stanford CoreNLP (https://stanfordnlp.github.io/CoreNLP/), a linguistic tool that can automatically tag words in a number of different ways, such as recognizing part of speech, type of proper noun, numeric quantities, and more
- Autosummarizer (http://autosummarizer.com/), a website which uses AI to automatically produce summaries of texts.
- TXM, a more complex tool for 'textometry', a methodology allowing quantitative and qualitative analysis of textual corpora, by combining developments in lexometric and statistical research with corpus technologies (http://textometrie.ens-lyon.fr/?lang=en). It allows for a more granular analysis of language features, requiring the integration of a specific language model, the splitting of speakers, the conversion of data into computer readable XML language, and the lemmatization of the data.

2 http://clarin.eu/

4. Interviews, ASR and SSHOC

Within the EU SSHOC[3] project (which is focused on cooperation of the SSH communities in sharing data and tools) the potential value of CLARIN's linguistic tool kit for processing spoken language recording has found uptake in general in organizing webinars about the topic, and more specifically in Task 4.4 which addresses *Voice recorded interviews and audio analysis*. In this task we aim to introduce specific questions LISS Panel surveys to which participants can respond with audio recordings. These questions typically relate to more general opinions on for instance finances, ethics, and politics. A use case will be started for Dutch in which the audio recordings will be transcribed using the Dutch ASR and processed using further NLP tools such as for summarization, topic detection and, possibly, automatic translation. Special care will be given to GDPR compliant data collection and processing (Emery et al., 2019).

In order to raise awareness for the potential benefits of ASR for the transcription of audio recordings a webinar was organized by the dissemination team of SSHOC in which the background of the portal was addressed, followed by a tutorial on how to use it. A blogpost[4] about the webinar was published together with a Youtube podcast[5].
There were 172 viewers of the webinar. The majority of participants came from the EU countries, but the webinar was also followed by some participants from countries outside Europe (i.e. the USA, several African countries, China, etc.). The great majority (approx. 70 %), belonged to categories "Researchers, Research Networks and Communities" and "Universities and research performing institutions". These two categories were followed by "Research libraries and archives", "Research and e-infrastructures" and "Private sector and industry players". Their representation accounted to approximately 20 % of the entire audience. The remaining categories, "Policy making organizations", "Research funding organizations" and "Civil society and citizen scientists" were represented only by a few participants (approx. 10 %). These numbers show the enormous interest for and potential of spoken language processing in a wide variety of scientific disciplines.

As a follow up of the webinar we started organising four weekly QA sessions during which users of the OH portal can contact us in an interactive session based on pre-submitted issues that they come up e.g. in using the portal for their research.

3 https://sshopencloud.eu/

4 https://www.sshopencloud.eu/news/sshoc-webinar-clarin-hands-tutorial-transcribing-interview-data

5 https://www.youtube.com/watch?v=X6bFGJpMjVQ&t=6s

5. Conclusion

Our experiences with the various workshops and the webinar have convinced the Oral History working group (see section 1) that the processing of interviews as a research instrument has opened up a fascinating area of collaboration between humanities scholars and social scientists. Research tools such as the OH portal appear to appeal to great variety of researchers across academic disciplines. Building the appropriate tools requires a lot of "overbridging" talk by ICT developers in the Digital Humanities, but the fruits we see growing from that tree are certainly worth the efforts.

6. Bibliographical References

Draxler, C., Van den Heuvel. H., Van Hessen, A., Calamai, S., Corti, L., Scagliola, S. (2020). A CLARIN Transcription Portal for Interview Data. *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC2020)*.

Emery, T., Luijckx, R., Vanden Heuvel, H., (2019). Guidelines for the integration of Audio Capture data in Survey Interviews. *D4.12 of the SSHOC project*. https://zenodo.org/record/3631169#.Xo2N3_0za70

Scagliola, S., Corti, L., Calamai, S., Karrouche, N., Beeken, J., Van Hessen, A., Draxler, C., Van den Heuvel, H., and Broekhuizen M., (2020) Cross disciplinary overtures with interview data: Integrating digital practices and tools in the scholarly workflow. *Proceedings CLARIN Annual Conference, Leipzig, October 2019.*

Van den Heuvel, H., Draxler, C., Van Hessen, A., Corti, L., Scagliola, S., Calamai, S., Karouche, N. (2019). A Transcription Portal for Oral History Research and Beyond. *Digital Humanities 2019, Utrecht, 9-12 July 2019*. https://dev.clariah.nl/files/dh2019/boa/0854.html