

Exploring the archives for textual entry points to speech

– Experiences of interdisciplinary collaboration in making cultural heritage accessible for research

Rickard Domeij¹, Jens Edlund², Gunnar Eriksson¹, Per Fallgren², David House², Eva Lindström¹, Susanne Nylund Skog¹ and Jenny Öqvist¹

¹ Institute of Language and Folklore, Sweden

² KTH Royal Institute of Technology, Sweden
rickard.domeij@isof.se

Abstract. *Tilltal* (*Tillgängligt kulturarv för forskning i tal*, ‘Accessible cultural heritage for speech research’) is a multidisciplinary and methodological project undertaken by the Institute of Language and Folklore, KTH Royal Institute of Technology, and The Swedish National Archives in cooperation with the National Language Bank and SWE-CLARIN [1]. It aims to provide researchers better access to archival audio recordings using methods from language technology.

The project comprises three case studies and one activity and usage study. In the case studies, actual research agendas from three different fields (ethnology, sociolinguistics and interaction analysis) serve as a basis for identifying procedures that may be simplified with the aid of digital tools. In the activity and usage study, we are applying an activity-theoretical approach with the aim of involving researchers and investigating how they use – and would like to be able to use – the archival resources at ISOF. Involving researchers in participatory design ensures that digital solutions are suggested and evaluated in relation to the requirements expressed by researchers engaged in specific research tasks [2].

In this paper we focus on one of the case studies, which investigates the process by which personal experience narratives are transformed into cultural heritage [3], and account for our results in exploring how different types of text material from the archives can be used to find relevant sections of the audio recordings. Finally, we discuss what lessons can be learned, and what conclusions can be drawn, from our experiences of interdisciplinary collaboration in the project.

Keywords: found data, archive speech, interdisciplinary collaboration, participatory design.

Copyright 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1 Introduction

The project *Tilltal* (*Tillgängligt kulturarv för forskning i tal*, ‘Accessible cultural heritage for speech research’) applies speech technology methods to archival material, with the overall goal of making archived speech data more accessible to research in the social sciences and humanities (SSH) [1],[4]. The project is undertaken by the Institute of Language and Folklore (ISOF), KTH Royal Institute of Technology, and The Swedish National Archives in cooperation with the National Language Bank and SWE-CLARIN [1].

Swedish memory institutions hold enormous collections of recorded speech, but they are rarely used, due to the lack of efficient methods of handling archival audio material. The archives of ISOF in Gothenburg, Lund, Umeå, and Uppsala are home to one of the largest folklore and dialects collections in the country. As part of this collection, the archives hold something like 25,000 hours of recorded speech, collected over more than 100 years to document dialectal variation in Swedish (mostly), on the one hand, and beliefs and practices of traditional (and later contemporary) society across the country, on the other. The recordings are heterogeneous, representing a multitude of speakers, topics, genres, and localities, as well as physically reflecting changing recording techniques over the years. ISOF is dedicated to preserving and continuously adding to the dialect and folklore collections, and to supporting research into Swedish folk culture and varieties of spoken Swedish. The audio materials represent an enormously rich resource for research of many kinds; however, they are sadly underutilised, because of the lack of convenient ways of locating and accessing audio content. Currently, audio recordings have to be played in real time to be analysed. We see the use of digital tools as a promising way of facilitating exploration of the wealth of recordings.

The objectives of the *Tilltal* project is to identify areas where language and speech technology can be applied to simplify researchers’ work process, providing ways of browsing the contents and conveniently accessing audio materials to work directly from the recordings, rather than through transcripts. This work proceeds from three case studies, where actual research agendas are pursued by researchers from different disciplines (ethnology, sociolinguistics and interaction analysis) and is grounded in close collaboration between them and language and speech technologists. In this way, digital solutions are tailored to specific needs, and evaluated against those needs in actual usage. The project also includes an activity and usage study to investigate and document aspects of the collaboration as such, the work process of the researchers with and without digital aids, as well as the impact of the specific archival setting as a framework for the research conducted there.

In this paper, we focus on one of the case studies, which investigates the process by which personal letters and recordings were transformed into archived records of scientific knowledge and how, in that process, they came to be regarded as cultural heritage [3]. We account for our experiences and results while exploring how different types of text material from the archives can be used to find relevant sections of the audio recordings, as a complement to speech technological methods.

2 The humanities research problem

Three case studies from different fields guide the Tilltal efforts to design technological solutions in support of research on speech archive recordings: 1. ‘From personal experience narratives to cultural heritage’ (folkloristics/ethnology); 2. ‘Linguistic variation in time and space’ (sociolinguistics); and 3. ‘Interaction patterns over time and type of conversation’ (interaction analysis).

In this paper we will focus on the first case study, From personal experience narratives to cultural heritage, which is undertaken by Susanne Nylund Skog. The objective is to analyse how narratives of personal experience, as expressed in letters and recorded interviews, are transformed into ethnological records, and through the very process of sorting, cataloguing and incorporation into the structures of an archival collection come to form the warp and weft of the fabric that is understood as our cultural heritage [3].

The Uppsala archive of dialects and folklore at ISOF, where the research case reported in this paper is taking place, is over 100 years old. As with many (most?) archives, its history is an integral part of it, and layers upon layers of preferences, priorities, ideologies and presumably even inattentions of past archivists, as well as the particulars of collections of external origin that have been incorporated into the archive over time, are inextricably embedded into the setting of any research done there. Locating records pertaining to any particular research topic often requires detailed and extensive knowledge of specific archival principles of organisation.

The study is performed on a subset of the large collection on Swedish folklore – The Gilstring collection. Karl Gösta Gilstring (1915–1986), a clergyman, high school teacher and enthusiastic private collector of folklore, spent over 50 years assembling what is regarded as the largest folklore collection by a single Scandinavian researcher in modern times. It consists of more than 8,000 original letters and 250 hours of recordings (mainly interviews conducted by Gilstring himself), from which Gilstring made some 70,000 folklore records, divided into approximately one hundred collections by parish, each organised by subject matter, dealing with folk tales and traditional beliefs as well as rural everyday life and practices.

The focus of interest are the motivations and scientific premises used by Gilstring when itemising and classifying his data, typically collected in narrative form, into his various subject matter categories. His principles are revealed through studying the differences between unedited audio recordings (and letters) and the edited written versions of the same information as it later appears in the collection.

For a sub-study of manageable size, Nylund Skog investigated the contributions by one of Gilstring’s 700 long-term correspondents, a Swedish American called Carl Nelson, who emigrated in 1896, aged 18. Nelson and Gilstring exchanged over 600 letters, primarily in the years 1962–1966, on a wide range of topics all pertaining to his memories of folktales and traditional lore from his native parish of Oskar in Småland, Sweden.

Large parts of Gilstring’s collection are meticulously organised, but when Nylund Skog tried to get an overview of the correspondence, she found that whole batches of letters were not in their proper place in the chronological sequence. Moreover, several

unmarked binders full of correspondence were found near the rest of the collection, and just putting the letters into order took quite some time.

Carl Nelson was recorded by Barbro Klein (then Barbro Sklute) in 1967, retelling some of the content he had given Gilstring in writing. In the recording, Nelson skips from one story to another, and often goes back to comment on a story he had told earlier. Given the untidy structure and the length of the recording, it took Nylund Skog two weeks to go through it, listening, stopping to make notes in a Word document (indicating the time in the recording where the item of interest occurred), listening some more, stopping to write, listening again, going back, listening, stopping again, and so forth. Her notes come to 90 A4 pages, and that is without even attempting a full transcription of the content.

3 Suggested digital solutions

Automatic speech recognition (ASR) has developed rapidly in recent years, but it is still a long way from managing heavily non-standard speech from a great number of different speakers, often recorded in less than optimal conditions. In other words, the vast majority of the audio materials cannot be searched directly. One big obstacle to using the audio collections, therefore, is how to locate portions of interest. The obvious answer is text – any notes, summaries, transcriptions, subject topic indices, recording logs, etc. which references the content (with indication of the point of time in the recording where that content is to be found) constitute potential keys to opening the treasure chest of recordings.

The archive does contain many such referencing texts but they are not always easily found and connected with the relevant recording, and there is great heterogeneity in the types and richness of documents referring to a recording. Only a small portion of the recordings has full transcriptions. Within the set of recordings emanating from one project, there may be summaries with rough time stamps of the majority of the recordings, but for others merely a mention in a recording journal from a field trip. Furthermore, the accession numbers, which constitute the nearest thing to object IDs in the collections, are inconsistently rendered, which is of course an obstacle to handy searches even among materials that have been scanned and OCR processed.

For parts of the Gilstring Collection, some promising sets of documents pointing to recordings were identified, but when trying to retrieve the recordings it transpired that that they had not been preserved – presumably, Gilstring re-used the media after extracting the information he wanted.

Needless to say, to make full use of text-based pointers into audio materials, both text and audio need to be digitised. This is no mean task, but the process is underway at least as regards audio. Many text materials are not readily machine readable even if scanned, for instance having copious scribbled comments on the page or being handwritten altogether, or consisting of poor or ill-aligned photocopies, or carbon copies of typed text, and so forth. A variety of methods are being explored within ISOF generally, and within Tilltal.

While the Tilltal project aims to suggest and implement tools to aid researchers working with archival audio materials, a large part of the purpose of the project is to analyse the setting, research processes and potential areas for technological assistance, to identify more wide-ranging measures than can be fully implemented in the current project. Some of these preliminary results, therefore, are as yet more conceptual in nature, while others have been put into practice, at least on a testing scale.

3.1 Bunches

One outcome of analysing the archival structures from the point of view of improving access to audio recordings is the concept of “bunches”. The idea is for each recording to be linked to all its related documents and any other relevant items (cf. [3]), of whatever format, in the collections (and perhaps beyond) into virtual bunches. Accessing any of the component parts would bring up the others, including things like correspondence about a recorded item, or a photo of the speaker. Figure 1 illustrates a bunch with some of its possible components.

Bunches have to be coherent units, but not too tightly bundled. They must be open enough to allow additions such as Nylund Skog’s (see 2 above) annotations of instances of laughter and repetitions, and of the beginnings of new narrative sequences, or to link in other related archival items uncovered at a later date.

Crucially, they must also permit “sideways” exploring, using most any component as a starting point for finding other materials, belonging in other bunches, for instance using the location where a recording was made to find more recordings from that place, or subject topic tags to find other mentions of the same, and so on.

In the present paper we focus on ways of accessing audio and treat bunches as being centered on recordings, but of course any member of a bunch could serve as the center of another bunch. With each such change of viewpoint, the other associated items in the bunch come to function as metadata for the current focus item.

We are still in the process of identifying the best way of digitally representing these underlying relations between archival items, for instance looking into linked data approaches.

It must be admitted that the obstacles to large-scale encoding of bunches are not just technological – as set out above, the structure of the archives often makes it rather hard to find and reliably connect physical records that belong together; as a consequence the relations of digitised materials are not always evident either.

Nevertheless, bunches are a valid and important concept, and making the interrelations between items explicit adds significant value to each, in expressing the archival context, as well as permitting structured browsing of the holdings.



Fig. 1. Illustration of a “bunch”, in this case centered on a recording with the accession number Bd 1711.

3.2 Digital research environment

Taken together, the bunches just described will form a multidimensional network of archival records (/items/text excerpts/index entries/commentary, etc.) encoded for their various properties and their relations to other records (/items/...).

One of the project aims is to design a prototype system for users to interact with the data, a digital environment made available to the research community that permits exploration of the enriched dataset through searches and navigation.

Further design specifications of the backend of this platform are being investigated, such as the type of permanent resource IDs to be used, options for effective file storage formats and retrieval methods, and most efficient query procedures.

The envisaged system will further provide a set of custom tools to aid access to audio content. For instance, it will be possible to go straight from a specific annotation in text or a subject topic card to listening to the relevant portion of a recording, and there will be ways of exploring content with the help of visualisation. Project members at KTH are developing an ASR system, which will be able to assist by making many of the recordings searchable in text format.

A system of this type would correspond well with the wishes expressed by folklore researchers for a system that can handle different types of related data resources – recorded interviews, letters, notes and summaries from data collectors and researchers, questionnaires, etc. – as parts of one connected collection, rather than as isolated resources.

Some of the digital tools, pre-existing or designed within the project, are described below.

3.3 Audio navigation tools

In response to Nylund Skog's time-consuming work with analysing the recording of Carl Nelson (see cf. [3]), a prototype tool was designed to link notations of time in her document to the correct parts of the recording, allowing her to simply click on the time stamp to re-listen to the audio that an annotation concerns. This is helpful also in that she can reorder her notes as she wishes, without worrying that abandoning chronological order will make it harder to retrieve the sound.

In the project, larger-scale trials with automatic alignment have been made on archive materials, linking brief text descriptions of the contents or subject topic into sections of audio recordings, primarily for two lots of data, so that the right portion of audio can be accessed directly via the text records, which in turn can be found by searching. The first lot of data is the subject topic catalogue for recordings (which had been digitised by retyping). The other lot emanate from the project Hem ('Home') where participants were interviewed about their attitudes to their accommodation situation, and the interviews were summarised in chunks. Both sets have time stamps, allowing automatic linking.

These materials, along with other materials belonging with them in bunches, are expected to form the basis of the prototype environment for accessing archive data as described above.

Another tool explored is the transcription and annotation software ELAN [12], developed with a view to linguistic analysis of video and audio recordings of speech (or sign language). Leaving video aside, the audio track is displayed in raw waveform, which forms a time line. The wave form can be segmented and annotations entered for each segment on an annotation tier below (see bottom left image in Fig.1]. Any

number of tiers may be added and then hidden/displayed and reordered, so that there is no problem having mark-up for quite varied sets of properties. For instance, someone may want to tag all speech disfluencies, then the same person or someone else wants to annotate all uses of demonstratives, or references to mythical beings, or turn-taking patterns between interviewer and respondent, and so forth.

A small portion of archive recordings have full, word-by-word transcriptions of content, and in those cases we have successfully been able to use one of the web services for automatic annotation accessible through ELAN to align text to audio, namely the WebMAUS forced aligner [13].

Annotation tiers can easily be exported and added to bunches, forming another source of time-stamped textual entry points to audio.

The audio browser Edyson [14], developed within the project, also produces time-stamped annotations that may serve in the same way.

4 The collaboration experience – observations, conclusions and recommendations

The activity and usage study conducted within the Tilltal project is a crucial part of the collaboration process, and is presented briefly below.

The activity and usage study should be seen as an overarching design process that involves all three case studies in the project. It has two parts, using activity theory and use case modelling, respectively, as methods of investigation.

The more general part of the study aims to understand and describe the bigger picture of (using tools for) collecting, processing and making spoken narratives available for research at the ISOF, with the aid of Activity theory (for an overview, see [7]) as the theoretical framework. This involves considering contextual factors surrounding the core activities: the community in which the activities take place (at the different archives), the division of labor between actors (e.g., IT staff, researchers and archivists) and the rules and conventions that govern the activities (such as intellectual property rights, privacy laws and research ethics). Field diaries, observations, interviews, group discussions and workshops are used for data collection.

Data from that part of the study feeds into the next, which is an iterative, dialogical design process, where requirements from the SSH researchers meet proposals from the language technologists, following the principles of participatory design [10]. The aim is to further specify and model researchers' needs of information and usable tools for seeking information in speech material: What kind of information is needed to answer a certain research question? To what extent can the language technology of today provide that information? Is the quality level of the delivered information sufficient for the researcher? What would be the optimal way to report the results to the researcher? What degree of interactivity with the result reporting system is convenient for the researcher?

To give an example from the usage and activity study, one of the researchers (Nylund Skog) has been keeping a fairly detailed journal of her activities in the research undertaken in the first case study, such as practical efforts to locate materials in

the archive (specifically the Gilstring Collection as sketched above), her interactions with the materials found, as well as the process of shaping and defining her thinking on the research topic in response to that interaction. This has been quite instrumental for the language technologists to understand her specific research process and practices so as to be able to suggest relevant technological aids and in some cases create trial versions (cf. 3.3 above).

We believe that systematic documentation and analysis of the collaboration process can improve the quality of digital humanities research projects in several ways. Firstly, and most obviously, such meta-analysis makes the collaboration process itself more tangible and observable. This can be of great help in creating a more constructive and rewarding collaboration process. Although this of course is not a goal in itself; rather, the improved collaboration then becomes a platform for achieving the specific aims of the digital humanities research project, and providing valid and reliable digital solutions to do so.

Furthermore, systematically studying the collaboration process in its own right not only benefits the project itself – it also opens up possibilities to disseminate the gathered knowledge to the scientific community.

From the earliest planning stages of the Tilltal project, it was clear that the interdisciplinary aspects of the project needed to be taken seriously. We expected that collaboration between participants from such different academic traditions would entail some difficulties in communication, or at least require mutual readjustments. Several steps have been taken to ensure that all participants are on the same page. There are frequent meetings to make sure everyone is up to speed with what is going on, often with a subset of participants but several times a year all project members get together. We have also held several offsite workshops of a couple of days, in camp school spirit, which have moved the project forward significantly. The workshops have allowed project members to meet for longer continuous time blocks to present work in progress and to jointly explore archive data, with ample time for discussions and reflections.

Project members recognise that the well-worn distinction between quantitative and qualitative approaches holds to some extent for the scholars in the project too, despite the fact that the language technologists have a strong linguistics background and SSH academic culture is hardly alien to them. One thing that has become clear to them is how centrally important the concept of “context” is to the SSH researchers in the project; both in terms of data points being of no interest unless accompanied by rich context information, and in the sense that the archival setting itself constitutes a context with strong bearing on the research that can be conducted there. Although the language technologists expected the archive collections to be heterogeneous and the organisation complex, they had not fathomed the level and scope of heterogeneity and complexity – there is no valid comparison with, for example, a text corpus as a body of data from which a well-formulated query will extract all relevant items (with or without context). Meanwhile, the SSH researchers have gained a better understanding of what language and speech technology can – and cannot – do to facilitate their work. Continuously evaluating and discussing proposals for digital solutions (as part

of the participatory design process) has also allowed for new research questions to be formulated jointly by various constellations of project members.

5 Concluding remarks

A fully digitised archive, with a user-friendly interface and handy tools enable whole new ways of interacting with and understanding archive data. In the physical archive, items are largely bound to whatever position they happen to occupy through events in the history of the archive. In the digital version, any item can be taken from its customary place, regarded from different angles, tentatively combined with other items in new ways and reshuffled over and over. There is no need to choose between chronology, topic, informant, location, etc. – you can have them all. Any property can become the path to finding more content.

The concept of bunches, implemented in the rich digital research environment with its tools, promise exciting advances in the research methods available in an archival setting. They could not have come about without the intense focus on cross-disciplinary exchange of methods and ideas implemented in the Tilltal project.

6 Acknowledgement

The Tilltal project has funding for the period 2017–2020 as part of the Collections and Research Programme, financed by the Royal Swedish Academy of Letters, History and Antiquities and the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond) (SAF16-0917:1).

References

1. Johanna Berg, Rickard Domeij, Jens Edlund, Gunnar Eriksson, Per Fallgren, David House, Eva Lindström, Erik Magnusson Petzell, Zofia Malisz, Susanne Nylund Skog and Jenny Öqvist. 2019. Making Archival Speech Recordings Accessible for Research – A Report from the Tilltal Project. In Skott & Strandberg (eds): Svenska landsmål och svenskt folkliv 2018, pp. 171–178. The Royal Gustavus Adolphus Academy for Swedish Folk Culture, Uppsala.
2. Berg, Johanna, Domeij, Rickard, Edlund, Jens, Eriksson, Gunnar, House, David, Malisz, Zofia, Nylund Skog, Susanne & Öqvist, Jenny. 2017. Involving users and collaborating between disciplines in making cultural heritage accessible for research. Paper presented at CLARIN Annual Conference 18–20 september 2017, Budapest, Hungary.
3. Nylund Skog, Susanne 2018. From personal letters to scientific knowledge: The creation of archived records in a tradition archive. In: Lauri Harvilahti, Audun Kjus, Cliona O’Carroll, Susanne Österlund-Pötzsch, Fredrik Skott and Rita Treija (eds.) Visions and Traditions: Knowledge Production and Tradition Archives Helsinki: Academia Scientiarum Fennica, FFC 315.

4. Berg, Johanna, Domeij, Rickard, Edlund, Jens, Eriksson, Gunnar, House, David, Malisz, Zofia, Nylund Skog, Susanne & Öqvist, Jenny. 2016. Tilltal – making cultural heritage accessible for speech research. Paper presented at CLARIN Annual Conference 26–28 October 2016, Aix-en-Provence, France.
5. Malisz, Zofia, Öqvist, Jenny, Fallgren, Per, Edlund, Jens & House, David. 2017. Visualising vocalic variability in space and time – automatic exploration of “found data”. Paper presented at the 47th Poznań Linguistic Meeting, 18–20 September 2017, Adam Mickiewicz University, Poznań, Poland.
6. Prescott, Andrew & Hughes, Lorna. 2018. Why do we digitize? The case for slow digitization. In: *Archive journal* (Special issue: Digital Medieval Manuscript Cultures, edited by Michael Hanrahan & Bridget Whearty, September 2018.) Retrieved from: <http://www.archivejournal.net/essays/why-do-we-digitize-the-case-for-slow-digitization/>
7. Kaptelinin, V. (2013). Activity Theory. In M. Soegaard & R. F. Dam (Eds.), *Encyclopedia of Human-Computer Interaction*. Interaction Design Foundation.
8. Dagsson, Trausti & Skott, Fredrik. 2018. Digitalt Kulturarv – ett digitalt folkarkiv [Blog post]. Retrieved from: <https://sweclarin.se/swe/digitalt-kulturarv—ett-digitalt-folkminnesarkiv>
9. Borin, Lars, Forsberg, Markus, Edlund, Jens & Domeij, Rickard. 2018. Språkbanken 2018: Research Resources for Text, Speech, & Society. Poster DHN In: Mäkelä, Eetu, Tolonen, Mikko & Tuominen, Jouni (eds.) *Digital Humanities in the Nordic Countries 3rd Conference*. 504–506. Retrieved from: <http://ceur-ws.org/Vol-2084/poster7.pdf>
10. Kensing, F., & Blomberg, J.: Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*, 7(3-4), 167–185.
11. Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations. In: *Proceedings from the IEEE Symposium on Visual Languages*, Boulder, CO, USA, 1996, pp. 336–343. doi: 10.1109/VL.1996.545307
12. ELAN (Version 5.2) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from <https://tla.mpi.nl/tools/tla-tools/elan/>
13. <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMAUSBasic>
14. Fallgren, P. How to annotate 100 hours in 45 minutes. Retrieved from: https://www.iscaspeech.org/archive/Interspeech_2019/pdfs/1648.pdf