TillTal – making cultural heritage accessible for speech research

Johanna Berg
Digisam
The Swedish National Archives, Sweden
Johanna.Berg@riksarkivet.se

Rickard Domeij
The Language Council of Sweden
The Institute for Language and Folklore, Sweden
Rickard.Domeij@sprakochfolkminnen.se

Jens Edlund
Speech, Music and Hearing
KTH Royal Institute of Technology, Sweden

edlund@speech.kth.se

Gunnar Eriksson
The Language Council of Sweden
The Institute for Language and Folklore, Sweden
gunnar.eriksson@sprakochfolkminnen.se

David House
Speech, Music and Hearing
KTH Royal Institute of Technology, Sweden
davidh@speech.kth.se

Zofia Malisz
Speech, Music and Hearing
KTH Royal Institute of Technology, Sweden
malisz@kth.se

Susanne Nylund Skog
The Folklore Archives
The Institute for Language and Folklore, Sweden
Susanne.nylund.skog@sprakochfolkminnen.se

Jenny Öqvist
The Folklore Archives
The Institute for Language and Folklore, Sweden
jenny.öqvist@sprakochfolkminnen.se

Abstract

This paper announces the new Swedish research project TillTal, a cross-disciplinary project aiming to improve collaborations between SSH research and speech technology and to make Swedish speech archives more accessible to researchers. The project proposal was a direct result of Swe-Clarin efforts to boost speech in SSH, and will start in the beginning of 2017. Here, we provide the background and motivation for the project as well as the project's outline and goals.

1 Introduction

This paper presents a new Swedish research project, Accessible Cultural Heritage for Speech Research (TillTal). The project won funding in 2016 and will commence in the beginning of 2017. TillTal takes speech and spoken language as its starting point, and the fact that speech is a severely underutilized material in the cultural heritage archives. Recorded speech constitutes a hidden treasure of vast amounts of information of great interest to SSH research.

The project proposal was a result of an ongoing collaboration between several Swe-Clarin in a working group for speech, with a focus on the utilization of speech data rather than written language, and followed a series of meetings, a comprehensive survey of Swedish public speech resources, and workshops on the utilization of speech data in SSH research.

The project will develop existing speech technology methods and to the extent it is doable, adapt them to archive data. It will test these methods in three full SSH research projects, while paying strict attention to their fruitfulness and monitoring the manner in which they are adopted by SSH researchers. The project is expected to lead to results on a number of levels and in a number of fields.

This paper presents the background and motivation for the project, its goals and expected outcome as well as obstacle we expect to encounter. We hope that the project results will not only be valuable as they are, but will lead to similar projects in other countries and other archives.

2 Background

2.1 The Swe-Clarin Working Group for Speech

Within the Swedish CLARIN organization (Swe-Clarin), the Working Group for Speech consists of participants from three member sites: the Institute for Language and Folklore (ISOF), KTH Royal Institute of Technology (KTH), and the Swedish National Archives (DIGISAM). The group aims to counterbalance to the rather heavy focus on written materials in CLARIN, and promote speech based SSH research in CLARIN activities. The Swedish National Archives and its Digisam section holds profound knowledge of digital archives – in particular audio recordings and film – in Swedish national archives. DIGISAM also act as consultants and advisors on issues of digitization and IP rights. KTH's department for Speech, Music and Hearing, a CLARIN K-centre for speech research (CLARIN-SPEECH), is one of the oldest speech labs in existence. In addition to speech engineers, it hosts SSH researchers such as human-human conversation and interaction analysts. Finally, the Institute for Language and Folklore hosts the Language Council of Sweden, which works closely with language change and with language usage questions. It is also a very large archive: over 13000 hours of digitized speech recordings and an equal amount that has yet to be digitized. Finally, the Institute hosts a number of in-house SSH researchers.

2.2 Speech data in Swedish national archives

In 2015, KTH performed a survey of speech resources and speech data in public and government agencies and national archives on behalf of the Swedish Post and Telecom Authority (PTS). The results (PTS, 2016) pointed to vast quantities of speech data, but also highlighted a range of problems in making these data sets useful for speech research and speech technology development. Amongst these, issues of intellectual property ownership, privacy and integrity, and legal issues relating to the (often unknown) content of the recordings rank highly, unsurprisingly. A more surprising result, perhaps, was the finding that virtually none of these data sets were used in any research or development of any kind. The result is that thousands – millions even – of hours of speech data sits tucked away on shelves. One of the main reasons for this is simply that the sheer quantity and complexity of the data is daunting.

2.3 SSH research collaborations in CLARIN and Swe-Clarin

It is well-known that achieving in practice the collaborations with SSH researchers that are at the core of CLARIN's efforts has proven quite difficult. This is true for text data, and perhaps even more so for speech recordings. The working group has made a number of efforts to alleviate this problem.

During a workshop in 2015 (Swe-Clarin and the spoken language I – Research collaborations between *resource holders, speech technologists, and researchers within the human and social sciences*, Stockholm, 2015-11-16), the concept of research triplets was introduced and tested. The idea was to put together groups of three partners: an SSH researcher, a speech technologist, and a data holder. Putting together the engineer and the SSH researcher is obvious, but often, the resource holders are left out, which in our experience leads to numerous problems as researchers and engineers make assumptions about the data instead of involving people who are familiar with it. We then wanted to take the SSH researchers' current work procedure as the starting point. In the workshop, researchers described their work in small groups (triplets), allowing data holders and speech technologists to suggest ways in which their process could be facilitated by large speech data sets and speech technology. The hands-on task was to come up with one suggestion for a research project per group. The method was a success, and the three SSH research projects that reside within the framework of TillTal were conceived at that meeting.

2.4 *Research and the Archives*

In 2015 and 2016, Stiftelsen Riksbankens Jubileumsfond and Vitterhetsakademin, two Swedish funding agencies, cofounded the call *Forskningen och Samlingarna* (loosely "Research and the Archives"). The

call was a directed effort targeting specifically collaborations between researchers and archives that would lead to an increase in the use of data in national archives.

2.5 A joint proposal

In light of this background, it seemed inevitable that the participating members of the Working Group for Speech should send in a joint proposal targeting the increased accessibility of the extremely underused and vast resource that archived speech constitutes. The proposal, headed ISOF, was submitted and awarded funding in 2016, and the project starts in 2017.

3 Project outline

Speech technology has come a long way. Today, many techniques such as automatic speech recognition or speaker verification, voice quality analysis, and analysis of interaction patters such as chronograms, are mature and work reliably and with quantifiable errors. Tis holds true, however, as long as the speech recordings they are applied to meet a number of requirements. For example, the recording circumstances should be known, the quality even and of a certain standard, the vocabulary known or predictable, and so on.

TillTal is designed as a framework in which existing speech technology methods – mainly mature, proven methods – are applied to a new kind of speech materials: those that are found in archives. We are confident that these methods can be used fruitfully to some extent, but exactly what will work and what will not is impossible to tell without experimentation – there are simply too many unknowns.

In order to keep the experiments and their results valid and meaningful, we steer clear of artificial experimental tasks and focus instead on real, current research questions. The use of real research questions also allows us to add another key element, a research question on a different level: to study and document how the new methods are received and used by the SSH researchers involved. Within the project, three full research studies aimed at different types of speech related SSH research make up the use cases.

3.1 From stories to cultural heritage

This study investigates a collection created by Karl Gösta Gilstring (1915-1986), consisting of 8000 original letters, and 250 hours of recorded speech (mainly interviews made by Gillstring). The collection is the largest Nordic collection by a single researcher in modern time. In addition to the original SSH research, the study will act as a use case for experimenting with speech technology methods that provide new entry points to the speech data and relate these to the texts.

3.2 Linguistic variation in time and space

This study investigates linguistic variation and change in speech materials. Currently, variation and change is often captured by the researcher, who painstakingly annotates the phenomena of interest, for example all occurrences of a specific vowel. Here, we will test and evaluate different methods of finding these occurrences automatically, which would allow for larger data sets by orders of magnitude, whilst simultaneously freeing up the researchers' labelling time for other, more challenging pursuits.

3.3 Interaction patterns over time and type of conversation

In studies of conversation, researchers often focus on the flow of the conversation and the relation between different contributions to this flow: Who speaks when? How are the speaker changes managed? How is common ground achieved? How are attitudes towards the spoken signalled? Interaction models based on chronograms – simple diagrams showing segments of speech and silence over time – have proven a successful approach to study these kind of questions. Generally, these models are built automatically, but how well the methods used work on archive data is unknown.

4   Project goals and impact

TillTal is a cross-disciplinary project that encompasses goals on a number of levels and in several fields. Within the projects lifetime, we have speech technology goals that deal with the modification of existing technology to better work with archive data and more importantly, the description of what works and what does not together with quantified error levels. From a HMI standpoint, the tracking and documentation of the uptake of new methods will be a real contribution. And from an SSH standpoint, three different research areas will be furthered as a direct result of TillTal.

The direct long-term goal is to make the Swedish speech archives more accessible in general, and to SSH researchers in particular. We hope to achieve this not only by describing methods by which speech technology can be used to reach SSH research goals, but also by providing examples of fruitful interdisciplinary collaborations. Much of the work in the project will not be strongly language specific, and we hope that it will be useful to research based on other languages as well.

5   Conclusion

We have announced the new Swedish research project TillTal. The project proposal was a direct result of efforts to boost SSH collaborations with speech technologists in Swe-Clarin, and will start in the beginning of 2017. We look forward to reporting first results shortly after.