

Språkbanken 2018: Research Resources for Text, Speech, & Society

Lars Borin¹, Markus Forsberg¹, Jens Edlund², Rickard Domeij³

¹University of Gothenburg, Gothenburg, Sweden

²KTH Royal Institute of Technology, Stockholm, Sweden

³The Institute for Language and Folklore, Stockholm, Sweden

{lars.borin, markus.forsberg}@svenska.gu.se, edlund@speech.kth.se, Rickard.Domeij@sprakochfolkminnen.se

Abstract. We introduce an expanded version of the Swedish research resource Språkbanken (the Swedish Language Bank). In 2018, Språkbanken, which has supported national and international research for over four decades, adds two branches, one focusing on speech and one on societal aspects of language, to its existing organization, which targets text.

Keywords: Infrastructure, speech, text, society

1 Background

By now, Språkbanken (the Swedish Language Bank) is a well-known institution, both amongst Swedish and international researchers and amongst interested laymen. Established as early as 1975, following Sture Allén’s groundbreaking work in corpus linguistics, the Språkbanken research unit at the University of Gothenburg’s department of Swedish is nationally and internationally acknowledged.

Traditionally, its work focuses on methodologies for handling written Swedish, and on the development of linguistic resources and tools for Swedish text. The resulting language resources are made available both to researchers in language technology and other disciplines, and to the general public.

2 The expanded Språkbanken

In 2018, the University of Gothenburg joins forces with KTH Royal Institute of Technology (KTH) and the Institute for Language and Folklore (ISOF), both in Stockholm, to expand Språkbanken. The Gothenburg branch remains as *Språkbanken Text*. KTH’s Speech group, founded in 1951 by seminal phonetician Gunnar Fant, sets up *Språkbanken Tal* (Speech), aiming to provide infrastructure, data and processes for speech technology and speech related research. Finally, ISOF’s Swedish Language Council creates Språkbanken Sam (Society), which will support research around the languages and dialects in Sweden, as well as other parts of the intangible cultural heritage in Sweden.

3 Relevance to the Digital Humanities

The extended Språkbanken is highly relevant to Digital Humanities. Språkbanken operates within the same community as the Swe-Clarin¹, the Swedish membership in the European Clarin ERIC². Through Swe-Clarin, all partners are already strong proponents of the use of language and speech technology in the Digital Humanities. The strengthened resources of a formal infrastructure will boost this work.

4 Acknowledgements

The expanded Språkbanken is supported by the Swedish Research Council's infrastructure programme, which will provide funding from 2018 to 2024 in a first stage. The new infrastructure collaboration builds on years of collaboration between the three parties on many fronts. All three are active members of Swe-Clarin and have contributed to Meta-Net, and also collaborate both on research and infrastructure projects.

¹ <https://sweclarin.se/>

² <https://www.clarin.eu/>