

Storspigg-TBI: A parallel read speech Swedish corpus

Christina Tännander^{1,2}, Jim O'Regan¹, Jens Edlund¹

¹ Speech, Music & Hearing, KTH Royal Institute of Technology, Sweden

² MTM, Swedish Agency for Accessible Media

christina.tannander@mtm.se, joregan@kth.se, edlund@speech.kth.se

Abstract

We present Storspigg-TBI, a multi-speaker Swedish read speech corpus comprising 1 000 recordings produced by 99 professional narrators, released by Språkbanken Tal and MTM (Swedish Agency for Accessible Media). Following Språkbanken Tal tradition, resources are named after fish, this time Storspigg, the three-spined stickleback (*Gasterosteus aculeatus*). This is a shoaling fish that here symbolises a shoal of voices. The TBI suffix represents the first sub-corpus of this type and stands for “Talboksinformationskorpusen”.



Introduction

The Storspigg corpus collection targets corpora containing the same or similar text read by many people or in many ways. These corpora are helpful in particular to speech-oriented science in that they facilitate comparisons and investigations into variation in speech. The first installation in Storspigg, Storspigg-TBI (“Talboksinformationskorpusen”), includes in all 1 000 readings of an introductory message for talking books by 99 different professional narrators.

The texts are uncopyrighted artifacts from the talking book production of the Swedish Agency for Accessible Media (MTM). MTM is a Swedish government agency with two main missions: to act as a knowledge centre for accessible text and to adapt text into accessible formats, such as Braille and talking books, for people with print impairments (Sveriges Riksdag, 2025). MTM's digital library, Legimus, includes more than 150 000 digital talking books, around 114 000 in Swedish and 21 000 in English. Most talking books include a spoken introductory message containing information about the copyright law and the borrower's responsibilities, as well as the number of pages and heading structure. These messages do not contain any copyrighted material, and a legal assessment in 2019 came to the conclusion that the recordings could be made publicly available. A presentation of the corpus focusing on the preparations, such as the legal procedure, data selection procedure, and corpus adoption procedure was published in (Tännander et al., 2026). The corpus is available at <https://github.com/sprakbankental/storspigg>.

The text

The read text is largely similar, but not identical, across recordings, since it contains book-specific information: the number of pages, the number of heading levels, and the narrator's name and recording company. A typical Swedish text is seen in Example 1 and an English (not yet released) in Example 2.

Example 1: A typical Swedish text

Information om upphovsrättslagen och om talboken. Denna bok är framställd för användare av anpassade medier enligt sjuttonde paragrafen Upphovsrättslagen. Olaga spridning eller överföring beivras. Talboken har X sidor och rubriker på Y nivåer. Boken är inläst för Myndigheten för tillgängliga medier MTM år YYYY. Inläsare är NARRATOR vid COMPANY. Denna talbok innehåller även elektronisk text som är tillgänglig via läsprogram i dator. Slut på informationen.

Example 2: English version (not yet released):

Information about Swedish copyright law and this talking book. This talking book has been produced for users of adapted media in accordance with section seventeen of Swedish copyright law. Illegal distribution or transmission will be prosecuted. The talking book is X pages long and has headings at Y level(s). It is produced by the Swedish Agency for Accessible Media, MTM, in YYYY. Narrated by NARRATOR at COMPANY. The e-text of the book is provided and can be accessed using text display software. End of information.

Data selection

1 000 Swedish information messages were selected according to a handful of criteria: the message should be in Swedish, produced by MTM between 2011 and 2018, and contain the target text without any additional, potentially copyright-protected, material. The last criterion was enforced by retaining only files between 15 and 75 seconds in duration, and by using the speech recognition system Whisper (Radford et al., 2022) to verify that each recording approximately matched the target message described above.

Data adaptation

The corpus has been adopted as a Språkbanken Tal corpus using its standard corpus ingestion procedure which includes legal clearance and registration, non-lossy conversion to standard formats, detailed format registration, and initial transcription. Pedigree and derivations between adopted, converted, and curated versions are recorded within a separate provenance system rather than in filenames. No dynamic-range or amplitude normalisation was applied, preserving the original studio characteristics, but all files comply with the infrastructure’s baseline requirements for archival stability.

Initial data description and analysis

Table 1 shows some statistics of the dataset. The audio format is 16-bit linear PCM WAV mono, 22 050 Hz.

Table 1. Data description, an overview.

Feature	Value
Audio files	1 000
Total duration	12.7 hours
Avg. length	46 s
Narrators	99
Female voices	51
Female readings	575
Male voices	48
Male readings	425

Future work

This is the initial release of the corpus, with 1 000 Swedish recordings between 2011 and 2018. There are possibilities the corpus will be extended with recordings after 2018, and also to add an English version of the corpus. The authors are currently working on initial segmental and prosodic analyses of the recordings.

Acknowledgements

This work is made widely accessible through the Swedish Research Council funded national infrastructure Språkbanken Tal (2023-00161_VR).

References

- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. <https://doi.org/10.48550/ARXIV.2212.04356>
- Sveriges Riksdag. (2025). *Förordning (2010:769) med instruktion för Myndigheten för tillgängliga medier*. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/forordning-2010769-med-instruktion-for_sfs-2010-769/
- Tännander, C., O’Regan, J., & Edlund, J. (2026). A shoal of voices: Parallel read speech from professional Swedish narrators. *Proc. of LREC 2026*. Language Resources and Evaluation Conference (LREC) 2026.

