

Building open-source Text-to-Speech synthesis for minority languages with ternary quantity: Lule Saami as an example

Katri Hiovain-Asikainen, Juraj Šimko and Sjur Moshagen
Paper presented at the 50th Poznań Linguistic Meeting

The current presentation will focus on two neighboring Saami (Uralic) languages: North and Lule Saami. Spoken in the Northernmost parts of Scandinavia (Norway, Sweden and Finland), the speakers are at least bilingual in Saami and one or more majority languages. By the amount of speakers, North and Lule Saami are different: while North Saami has by far the most speakers among the altogether nine Saami languages (approx. 25 000), Lule Saami (ca. 650 speakers; Morén-Duoljjá, 2013) is classified as “severely endangered” by the UNESCO atlas of endangered languages.

The languages are structurally similar, and after some training, they might be mutually intelligible to some extent. However, as part of language revitalization and preservation as well as accelerating digitalization, separate languages need separate language and speech technology tools to meet the needs of modern language users. At UiT The Arctic University of Norway in Tromsø, a variety of Saami language technology tools such as dictionaries, morphological analyzers, spell checkers and language learning tools etc. have been maintained and developed since 2001 by the Divvun and Giellatekno groups. As an important addition to the set of language tools, a Text-to-Speech synthesis (further: TTS) was developed for North Saami in 2015. Currently, a Lule Saami TTS based on open-source technologies is being developed. A TTS system is useful for a number of purposes and applications, not only aiding vision impaired individuals but also language learners, bringing more accessibility to many kinds of contents and utilities.

Some of the typological and phonetic features of for example North and Lule Saami are setting challenges in building a high quality TTS. One of these is the ternary quantity system in both of these languages. In Lule Saami, there are triplets that differ only by the quantity, especially of the length of the intervocalic consonant, for example: [ʔo̞ɑ̃:r:ɪɛ] ‘squirrel; Q3’ – [ʔoɑr:ɪɛ] ‘squirrels’ or ‘reason’; Q2 – [ʔoɑrɪɛ] ‘reasons’; Q1 (examples by Fangel-Gustavson et al., 2014). The orthography does not show difference between the Q3 and Q2 forms, both are written as oarre in the standard text. Our first experiments on building an open-sourced TTS have shown that a simple rule-based formant synthesis (such as Espeak) is not able to fully cover for this phonetic phenomenon without a separate syntactically disambiguating text-processing pipeline. More advanced, machine-learning (e.g. neural network) based techniques are building a duration model from training data which will presumably also model and address the quantity alternation of the language in an intelligible way.

Designing and recording the speech corpus and building the TTS system are equally important and challenging tasks in the case of developing TTS for endangered languages. A fair amount of speech data (minimum of 10 hours) is needed to build a TTS with good voice quality, and for addressing the quantity related features. In our presentation, we discuss these challenges, different TTS frameworks and also propose some solutions suitable for minority languages.

Keywords:

speech synthesis, minority languages, Lule Saami, Saami languages, speech technology, ternary quantity

References:

Fangel-Gustavson, N., Ridouane, R., & Morén-Duolljá, B. (2014, May). Quantity contrast in Lule Saami: A three-way system. In Proceedings of the 10th International Seminar on Speech production (pp. 106-109).

Morén-Duolljá, B. (2013.) Lule Saami Morphology: Back to Basics. Presentation. II workshop. University of Tromsø. 19 Feb. 2013.

Divvun website: <https://divvun.no/fi/>; Divvun Tacotron: <https://github.com/divvun/lang-sme-ml-speech>

Espeak GitHub: <https://github.com/espeak-ng>