

## **Between Descriptive and Normative, Reclaiming a Skolt Sami Domain**

### **Introduction**

This paper will present our work with a morphological generator and analyzer for the modern Skolt Sami language. The pluricentric nature of Skolt Sami poses challenges in solving this task. Skolt Sami pluricentricity is observed in three areas: (1) the geographic location and political history; (2) extensive fieldwork in dialects not included in the literary standard (less extensive fieldwork in the literary standard dialects), and (3) the establishment of a normative body for the development of a language standard with a more Finnish orientation.

### **Background**

Living in three countries, Finland, Norway and Russia, the Skolt Sami are exposed to pluricentric language domains. The speakers themselves represent two main dialects: Paatsjoki and Suõ'nn'jel. The Paatsjoki dialect is divided into the variants of Njauddâm (extinct), Paaččjokk, Peäccam, and Mue'tkķ, and the Suõ'nn'jel dialect is divided into the variants of Suõ'nn'jel, Njuõ'ttjäu'rr, Sââ'rvesjäu'rr, see [1] and [2]. The dictionary of Easterns Sami language [3] provides extensive information only on Paatsjoki dialect. In the aftermath of World War II, resettlement and restructuring of reindeer husbandry has meant the establishment of a few heterogeneous language centers and the loss of any community with a homogeneous language form.

The morphological description of the language includes alphabet extensions to cover the differences between the Finnish and Norwegian as well as the neighboring Inari and North Sami languages, i.e., written materials may automatically include person and place names written with letters not native to Skolt Sami. (At present, a standard work flow has yet to be established for Cyrillic to Latin transliteration.)

### **Morphological development and testing**

Our tool includes two separate levels for facilitating linguistic variation: the two-level phonological description (twolc), and the concatenation level (lexc). Whereas the twolc description declares a finite alphabet for the individual language, including “funny letters”, the lexc description allows for variation in concatenational morphemes and specific dialect or usage tags. These dialect and usage tags may also be applied to lexical items and are important in tackling the pluricentricity of the language.

One important phase in morphological description development is testing. This is done based on a gold standard for the normative analyzer and corpora for the descriptive analyzer. The normative standard is derived from minutes to meetings of the normative body Ķiõlljuâggtõs, which provide spelling, inflectional and lexical norms for the language. Corpora at Giellatekno provide nearly twenty years of administrative and legal texts.

### **Conclusion**

In this abstract, we have presented the development of our open-source tools for Skolt Sami morphology. We have also presented the socio-linguistic background relevant for conducting this research.

## **Bibliography**

- [1] P. Sammallahti, *The Saami Languages: An Introduction.*, Karasjok, Norway: Davvi Girji, 1998.
- [2] T. Fiest, *A Grammar of Skolt Saami. Mémoires de la Société Finno-Ougrienne*, 273. Helsinki:, Helsinki: Suomalais-Ugrilainen Seura., 2015.
- [3] T. I. Itkonen, *Koltan- ja Kuolanlapin sanakirja I-II [Wörterbuch des Kolta- und Kolalappischen]. Lexica Societatis Fenno-Ugricae XV.*, Helsinki: Suomalais-Ugrilainen Seura, 1958.