

Verdd - A Simplified Online Interface for Skolt Sami Dictionary Editing

Lexicographic resources for endangered languages are often fragmented into several different paper publications that have been digitized by different conventions and during different eras. The severely endangered Skolt Sami is no exception. For one part, a massive open-source dictionary for Skolt Sami has been made available on the MediaWiki-based Akusanat (Hämäläinen & Rueter, 2018) and the Giellatekno infrastructure (Moshagen et al., 2013). For the other, valuable resources such as Sammallahti and Moshnikoff materials (Sammallahti & Moshnikoff, 1991) are not formatted in the same structure and their inclusion in the aforementioned systems poses a challenge.

In order to make all three of the lexicographic resources for Skolt Sami available for researchers and non-academic dictionary users alike, we propose a new extension to the MediaWiki based Akusanat dictionary that makes intercomparison of the different materials possible together with a facilitated semantic search functionality. The main goal is to alleviate the workload of dictionary editors when bringing lexicographic data available in the unified system. As community involvement in editing online dictionaries has previously been identified as a viable way of extending the lexicographic data (cf Everson et al., 2019), we propose a simplified interface for non-technical community members to actively participate in the dictionary editing process.

Our proposed system allows the involvement of Skolt Sami speakers by giving them the ability to interact with the system to find existing translations, modify them and verify the correctness of information presented in it. The system has an extensive search interface for querying lexemes based on their parts-of-speech, ContLex, assonance, consonance, Inflex type. All this has been developed in an intuitive fashion so that non-technical people can easily work with the complexities of XML data.

References

1. Everson, R., Honoré, W., & Grimm, S. (2019). An Online Platform for Community-Based Language Description and Documentation. In *Proceedings of ComputEL* (Vol. 1).
2. Hämäläinen, M., & Rueter, J. M. (2018). Advances in synchronized XML-MediaWiki dictionary development in the context of endangered Uralic languages. In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (pp. 967-978).
3. Moshagen, S. N., Pirinen, T., & Trosterud, T. (2013). Building an open-source development infrastructure for language technology projects. In *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)* (pp. 343-352).
4. Sammallahti, P., & Moshnikoff J. (1991). *Suomi-koltansaame sanakirja. Lää'dd-sää'm sää'nnke'rjj*. Girjegiisá Oy. Ohcejohka.