

DEMO

Giellatekno Open-source click-in-text dictionaries for bringing closely related languages into contact.

Abstract

This is a demo of online dictionaries between minority languages with open-source FST morphology and lexical work. Open-source click-in-text dictionaries, NeahttaDigiSanit (NDS), have been developed in the Giellatekno infrastructure since 2012-2013. Originally, these dictionaries were implemented to provide a Sami-language reading tool to provide click-in-text wordform recognition and translation for majority-language readers by which they could access text content written in minority languages. From the very beginning of 2013 this same dictionary has been localized not only for Sami languages, but for work with Balto-Finnic, Mordvin, Mari, Permic, Ugrian and Samoyedic languages, as well. There are now a number of NDS projects with Finnish, Hungarian, Russian and German majority language hubs. There are also a few instances of closely related language pairs outside of the Sami and Balto-Finnic branch; these include the Permic languages and Mordvin languages.

As of late October 2016, there are three sets of two-directional dictionaries operating between closely related languages: Udmurt<->Komi Zyrian; Komi Permyak<->Komi Zyrian, and Erzya<->Moksha.

Permic language pairs	Source	Target	Lemmas
http://kyv.oahpa.no	Udmurt	Komi Zyrian	3427
	Komi Zyrian	Udmurt	5020
	Komi Permyak	Komi Zyrian	1252
	Komi Zyrian	Komi Permyak	1981
Mordvin language pairs	Erzya	Moksha	3677
http://valks.oahpa.no	Moksha	Erzya	3613

The lemmas and their glosses have been taken from the publications by Evgeni Tsypanov (Udmurt, Komi Permyak and Komi Zyrian), and Osip Polyakov (Moksha and Erzya). Open-source FST work comes directly projects in the Giellatekno infrastructure.

Resources: Dictionaries

О.Е. Поляков 2004: Мокшень и эрзянь кяльхненъ фкакс- и аф фкаксшисна. Синь валлукссна. Эрзянь ды мокшонь кельтненъ вейкекс- ды аволь вейкексист. Сынст валлувост. / О.Е. Поляков, J. Rueter; Н.П. Огарёвонь лемса Мордовская государственная университетсь. – Саранск: «Краснай Октябрь» типографиясь. – 264 с.
<https://victorio.uit.no/langtech/trunk/words/dicts/mdfmyv/>

<https://victorio.uit.no/langtech/trunk/words/dicts/myvmdf/>

Цыпанов, Е.А. 1999: ПЕРЫМ-КОМИ ГИЖӖД КЫВ. Сыктывкар — «Пролог»
небӖг ледзанін.

<https://victorio.uit.no/langtech/trunk/words/dicts/kpvkoi/>

<https://victorio.uit.no/langtech/trunk/words/dicts/koikpv/>

Цыпанов Й. 2007: Удмурт кыввор да лыддянкуд. Россияса наукаяс
академия Урал юкӖд Коми туялан шӖрин Кыв, литература да
история институт. — Сыктывкар.

<https://victorio.uit.no/langtech/trunk/words/dicts/udmkpv/>

<https://victorio.uit.no/langtech/trunk/words/dicts/kpvudm/>

Open-source FSTs:

<https://victorio.uit.no/langtech/trunk/langs/koi/>

<https://victorio.uit.no/langtech/trunk/langs/kpv/>

<https://victorio.uit.no/langtech/trunk/langs/mdf/>

<https://victorio.uit.no/langtech/trunk/langs/myv/>

<https://victorio.uit.no/langtech/trunk/langs/udm/>