

# Karjalan kieliteknologiaa

Flammie A Pirinen `flammie.pirinen@uit.no`, `@ffflammie`

helmikuussa 2023, UEF.fi / online



# Tromssassa kehitetty kieliteknologia

<https://divvun.org>

- ▶ Alkujaan saamelaiskielten tukemiseen tehtyjä kieliteknologiasovelluksia ja -ratkaisuja
- ▶ Avointa lähdekoodia ja vapaita ratkaisuja joita voi tarjota muille
- ▶ Kielenkäyttäjien tueksi: näppäimistöjä, oikaisulukimia, "älykkäitä" sanakirjoja, kieliopin tarkistimia, konekäännöstä, puheteknologiaa
- ▶ Myös tutkijoille: Korpuksia, annotoituja kieliaineistoja ja käyttöliittymiä yms.
- ▶ Enemmän kieliteknologeille vielä: morfologisia ja syntaktisia analysaattoreita ja koneluettavia sanakirjoja yms.
- ▶ Karjalan kielten sovellukset ovat enimmäkseen varhaisia testausversioita



# Karjalan kielen kieliteknologian tilanne Tromssan kieliteknologiajärjestelmissä

<https://giellalt.github.io>

- ▶ Kieliteknologiasovelluksemme perustuvat koneluettaviin sanakirjoihin ja käsin kirjoitettuihin morfologiaa, syntaksia ja jne. koskeviin sääntöihin
- ▶ Esim. pohjoissaamen sanakirjassa on 164 000 sanaa ja livvin 58 000
- ▶ Näitä on kehitetty kursseilla, vapaa-ajalla ja jonkin verran esim. Googlen kesäkoodissa

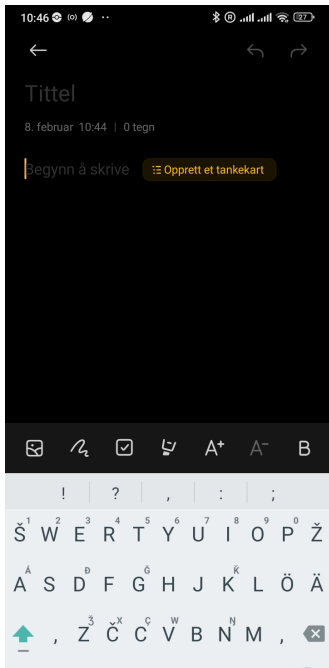
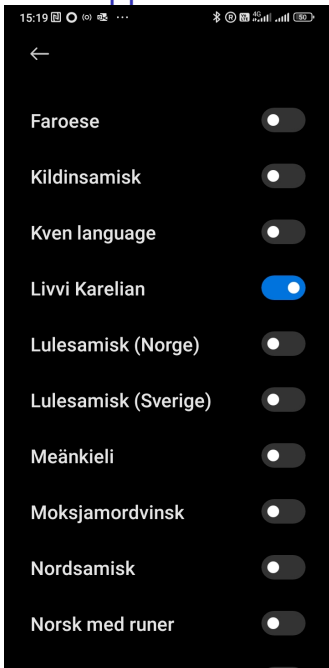


# Mobiilinäppämistöt

- ▶ **Divvun Dev Keyboards**, dev on testiversio jossa on jo karjalan näppäimistöasetteluja
- ▶ Androidille: <https://play.google.com/store/apps/details?id=no.uit.giella.keyboards.dev> (löytyy playstoresta)
- ▶ iPhoneille: <https://apps.apple.com/gb/app/divvun-dev-keyboards/id1518256662>



# Android-näppäimistö



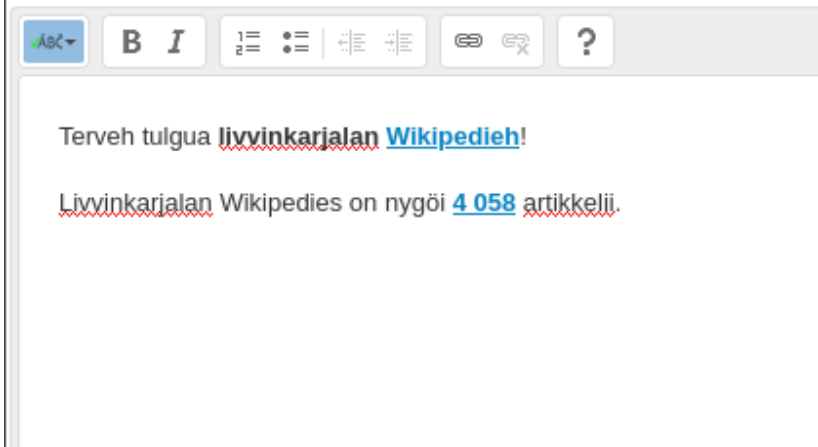
# Oikaisulukimet toimistosovelluksissa ja tietokoneella

- ▶ <https://divvun.org/proofing/proofing.html>:
- ▶ Toimivat Windowsin Officessa tai Libreofficessa kaikilla järjestelmillä
- ▶ Verkossa voi myös testata:  
<https://divvun.org/proofing/online-speller.html>
- ▶ (tulevaisuudessa Divvun Manager -sovellus korvanee käsin asennettavia lisäosia myös karjalan kielille)



# Test our spellers

Type or paste your text in the editor window:



The screenshot shows a web editor interface. At the top, there is a toolbar with various icons: a language dropdown showing 'ABC', bold (B) and italic (I) buttons, numbered list (1=, 2=) and bulleted list (:=) buttons, a link (chain) and unlink (chain with X) button, and a help (?) button. Below the toolbar is a text input area containing the following text:

Terveh tulgua livvinkarjalan [Wikipedieh!](#)

Liivinkarjalan Wikipedies on nygöi [4 058](#) artikkelii.

In the bottom right corner, there is a logo with the characters 'ÁBČ' above a stylized 'S' symbol. At the very bottom of the page, there are small navigation icons for back, forward, and search.

# Sanakirjoja

- ▶ <https://sanat.oahpa.no>
- ▶ perussanakirja, ymmärtää taivutettuja muotoja ja osaa taivuttaa sanoja
- ▶ sisältää selaimeen asennettavan skriptin jolla voi kääntää yksittäisiä sanoja verkkosivuilta





olo - fin ▾

siä

Haje

siä oon mahdollinen haamu sanasta ...

## siä (s.)

- (s.) sää , (s.) ilma
- (s.) hyvä sää , (s.) kaunis ilma
- (s.) huono sää , (s.) rajuilma , (s.) myrsky

s. sg. par.	siädy
s. sg. Apr	siälllyö
s. sg. gen.	siäñ
s. pl. par.	siälöi
s. pl. gen.	siälöin

siä *(Korpusta ei ole saatavilla)*  
siä s. sg. nom.

ädy

ällyö

än

älöi

älöin

siä

*(Korpusta ei ole saatavilla)*

siälöin

*siä* (s.) — 1. sää, 2. ilma, 3. hyvä sää, 4. kaunis ilma, 5. huono sää, 6. rajuilma, 7. myrsky

# Konekääntimiä

- ▶ Apertiumilla, verkossa on kokeiltavana beeta-versio:  
<https://beta.apertium.org>
- ▶ melko varhaiset versiot useimmilla kielipareilla

The screenshot displays the Apertium Beta web interface. At the top, the logo "Apertium Beta" is shown with a globe icon and the text "A free/open-source machine translation platform". The top right corner features a language dropdown menu set to "English" and a globe icon. Below the header, the interface is split into two main sections. The left section has a language dropdown menu with options: "Finnish", "Norwegian Bokmål", "Afrikaans", and "Detect Language". Below this is a text input area containing the Finnish sentence "Tänään on hyvä sää." and a close button (X). Below the input area are two buttons: "Translate a document" and "Translate a webpage". The right section has a language dropdown menu with options: "Livvi-Karelian", "Karelian", and "Danish". Below this is a text output area containing the Livvi-Karelian translation "Tänäpäi on #hyvä sää." and a copy icon. To the right of the output area are three checkboxes: "Mark unknown words" (unchecked), "Instant translation" (checked), and "Multi-step translation (experimental)" (unchecked). A blue "Translate" button is located at the top right of the right section.

# Korpuksia

- ▶ Korp-rajapinta: [https://gtweb.uit.no/f\\_korp/](https://gtweb.uit.no/f_korp/)
- ▶ Raakaversioita saatavilla lähiaikoina githubista
- ▶ lähinnä wikipediaa?

Muualta:

- ▶ VepKar-korpus: <https://dictorpus.krc.karelia.ru/en>
- ▶ dependenssiuppankkeja:  
<https://universaldependencies.org>

# vielä puuttuvaa kielitekniologiaa

- ▶ kieliopin tarkistusta
- ▶ puhetekniologiaa
- ▶ ...

# Lopuksi: ChatGPT-tekoälyä



kirjoita lyhyt johdanto karjalan kielen kieliteknologian tilasta ja tulevaisuudesta



Karjalan kieli on perinteinen itäinen suomalainen kieli, joka puhutaan Venäjän Karjalan alueella. Kieliteknologian kehitys on tuonut uusia mahdollisuuksia kehittää ja ylläpitää karjalan kielen digitaalisia resursseja ja sovelluksia. Tällä hetkellä karjalan kielen kieliteknologian tilassa on paljon potentiaalia, mutta myös haasteita, kuten riittävän resurssoinnin ja tietoteknisten ratkaisujen kehittämisen puute. Tulevaisuudessa odotetaan merkittävää edistystä kieliteknologian sovelluksissa karjalan kielen parissa, mikä parantaa kielen elinvoimaisuutta ja ylläpitämistä.



# Lopuksi: ChatGPT-tekoälyä



kirjoitapa sama karjalan kielellä



Karjalan kiil' on vanhahko itäsuomalaine kiil', kudai puhutah Karjalan alovehel Ven'an valdkes. Kiiliteknoloigan kehitämästä on tulnu uuzii mahdollisuuksii kehitädä da pidädä digitalisii resurssei da sovelluksii karjalan kiilen parah. Nygyääl kiiliteknoloigan tilah karjalan kiilen parah on lugui mahdollisuuksii, a tääl iän myös haasteit, kuten riittäväen resurssuinnin da teknoloigan ratkaijuiden kehitämisen puutteh. Tulevaisuudes odotetah merkittävää edistystä kiiliteknoloigan sovelluksissah karjalan kiilen parah, midä parandau kiilen elinvoimaisuutta da pidämistä.



🔄 Regenerate response





# Kysymyksiä? Palautetta

jälkikäteen palautetta voi lähettää osoitteeseen  
palaute@divvun.no tai suoraan minulle  
flammie.pirinen@uit.no.

Muita linkkejä:

- ▶ Käyttäjille: <https://divvun.org>
- ▶ <https://giellalt.github.io>
- ▶ Kieliteknologeille: <https://github.com/giellalt/>