

Suomalais-ugrilaiset kielet ja kieliteknologia

Trond Trosterud
Giellatekno, Saamen kieliteknologian keskus
<http://giellatekno.uit.no/>



June 14, 2019

Katsaus

Johdanto

Historiallinen tausta

Kieliopillinen analyysi

Sovellukset

- Korjausohjelmat

- Puhe

- e-sanakirjat

- Konekäännös

Maailmanlaajuinen yhteistyö

Yhteenvedo

Norjalaisia tällä puhekorokkeella ennen minua?

- ▶ Just Qvigstad 1899: Uebersicht der geschichte der lappischen sprachforschung
- ▶ Konrad Nielsen: Lectio praecursoria väitöstilaisuudessa 13. p. joulukuuta 1902

Lappalaisten murteiden tutkimisesta. *Suomalais-ugrilaisen seuran aikakauskirja* 1903 4:1-12

Nielsen 1903 s. 11f

Mutta täytyykö siis kaiken työn tarkoituksena tällä alalla olla yksinomaa keinojen hankkiminen kielen *menneisyyden* selittämiseksi? Eikö kielen elävällä *nykyisyydellä* ole omaa itsenäistä merkitystään? (...)

Monet merkit viittavat siihen, että lapinkielellä tai ainakin muutamilla sen murteilla ei ole enää pitkää elinaikaa jäljellä. Eikö tämä velvoita niitä kansoja, jotka – osittain ehdoin tahdoin, oikein suunnittelun mukaisesti – vaikuttavat oman kielen häviämiseen lapin kansan huuilta, – eikö se velvoita heitä pitämään huolta siitä, että jälkimaailma ainakin saa niin selvän kuvan kuin mahdollista tuosta voitetusta kielestä moninaisine murrevivahduksineen?

Mikä on siis fennougriistiikan tarkoitus?

- ▶ keinojen hankkiminen kielen menneisyyden selittämiseksi
- ▶ kielen elävän nykyisyyden kuvaaminen
 - ▶ Tämän päivän fennougriistiikka on pitkälti 27-vuotiaan väittelijän puolella (!)
- ▶ Kysymys on mutkikas, ja palaan asiaan, mutta ensin *päivän teemaan*

Ihmiset

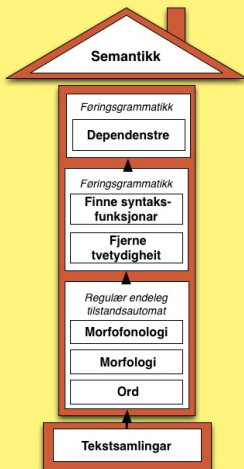
- ▶ **Giellatekno:** perustettiin 2001 (UiT).
Saamelaiskielten ja muiden pohjoisten kielen kieliteknologinen tutkimusyksikkö
Kieliopillinen mallintaminen, e-sanakirjat, e-opetus, korpusanalyysi, konekäännös, ...
 - ▶ Trond Trosterud, Lene Antonsen, Ciprian Gerstenberger, *Chiara Argese*
- ▶ **Divvun:** Perustettiin 2005 (UiT < Sisäasiainministeriö).
Infrastruktuuri, korjausohjelmat, synt puhe, terminologia
 - ▶ Sjur Moshagen, Thomas Omma, Maja Kappfjell, Børre Gaup, Linda Wiechetek, *Elena Paulsen*
- ▶ **AKU-projekti 2013-14:** Kieliopilliset mallit: Vuorimari, liivi, livvi, mokša, nenetsi, ...
 - ▶ Jack Rueter, Renate Blumberga, Julia Kuprina, Timo Rantakaulio, Sven-Erik Soosaar, ...

Miten tähän on päästy?

- ▶ Helsinki 90-luvun alussa
 - ▶ Lingvistiikan laitos Helsingissä (~ Budapestissa, Ankarassa)
 - ▶ Fennougristiikan laitokset Helsingissä ja Turussa
- ▶ Helsingissä, Tehtaankadulla: Lingsoft (1986-)
- ▶ Takaisin Norjassa, saamelaiskäräjät halusivat korjausohjelmia (2003)
- ▶ ... ja Sjur tarvitsi sopivasti töitä
- ▶ Norjan punavihreä hallitus laati uuden ulkopolitiikan (2010)
 - ▶ ja me teimme infrastruktuurin
- ▶ Tänäpä, 245 henkilöä on hakenut, ja saanut, salasanan palvelimeen

Mitä me teimme?

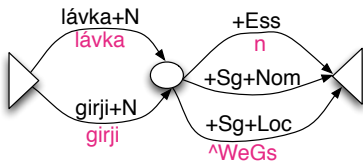
Grammatisk analyse



Leksikko

vielppis+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:viel'pis FALIS ;
 voitoфанas+Sem/Veh:voito#fatnas MALIS ;
 voitočuurvas+Sem/Dummytag:voito#čuur'vas MALIS ;
 voššanbierrgas+Sem/Tool:voššan#bier'gas MALIS ;
 voššanguolli+Sem/Food:voššan#guolli AIGI ;
 voššanlihtti+Sem/Ctain:voššan#lihtti AIGI ;
 vuovdevánddis+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:vuovde#ván'dis FALIS ;
 vuovdevárri+Sem/Plc-elevate:vuovde#várri AIGI ;
 vuovdečoažžu+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:vuovde#čoaž'žu BOAZU ;
 vuovdi+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Hum:vuovdi ACTOR ;
 vuovdi+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Plc:vuov'di AIGI ;

Kimmo Koskeniemen kaksitasomalli



lávka+N+Ess	lávkan
lávka+N+Sg+Nom	lávka
lávka+N+Sg+Loc	lávkkas
girji+N+Ess	girjin
girji+N+Sg+Nom	girji
girji+N+Sg+Loc	girjjis

vk	->	vk			-	Vow*	^WeG	;
rj	->	rj			-	Vow*	^WeG	;

^WeG: triggersymbol – Weak Grade

"<Mii>"

"mii" Pron Interr Sg Nom

"mii" Pron Rel Sg Nom

"mun" Pron Pers Pl1 Nom

"<eat>"

"ii" V IV Neg Ind Pl1

"<leat>"

"leat" V IV Ind Prs Pl3

"leat" V IV Ind Prs Sg2

"leat" V IV Inf

"leat" V IV Ind Prs ConNeg

"leat" V IV Ind Prs Pl1

"<dan>"

"dat" Pron Pers Sg3 Acc

"dat" Pron Dem Sg Acc

"dat" Pron Pers Sg3 Gen

"dat" Pron Dem Sg Gen

"<mutalan>"

"mutalit" V* TV Der2 Der/eapmi N Sg Gen

"mutalit" V TV Actio Acc

"mutalit" V TV Actio Gen

"mutalit" V TV Ind Prs Sg1

"mutalit" V TV PrfPrc

"mutalit" V TV Actio Nom

"<->"

"," CLB

*V Pl1 follows, hence
Pron Pers Pl1.*

*Preceding V Neg, hence
ConNeg.*

*Not Dem (no A or N following).
Not Gen (not following Pr or
Num, nor precedes N, A, Po).
Hence Pron Acc.*

*PrfPrc, since it is complement to
a Neg verb.*

```
"<Mii>"  
  "mun" Pron Pers Pl1 Nom @SUBJ> #1->2  
"<eat>"  
  "ii" <aux> V IV Neg Ind Pl1 @FAUX #2->0  
"<leat>"  
  "leat" <aux> V IV Ind Prs ConNeg @IAUX #3->2  
"<dan>"  
  "dat" Pron Dem Sg Acc @OBJ> #4->5  
"<muitalan>"  
  "muitalit" <mv> V TV PrfPrc @IMV #5->3  
"<.>"  
  "." CLB #6->6
```

Lyhyt, opastettu kierros 1

- ▶ `http://giellatekno.uit.no/index.fin.html`
- ▶ `http://gtweb.uit.no/korp`
- ▶ `http://gtweb.uit.no/u_korp`
- ▶ `http://gtweb.uit.no/f_korp`
- ▶ `http://gtweb.uit.no/tyv_korp`

(Fennougristille kätevä paketti: tekstianalyysi, annotoidut tekstikorpuukset, ...)

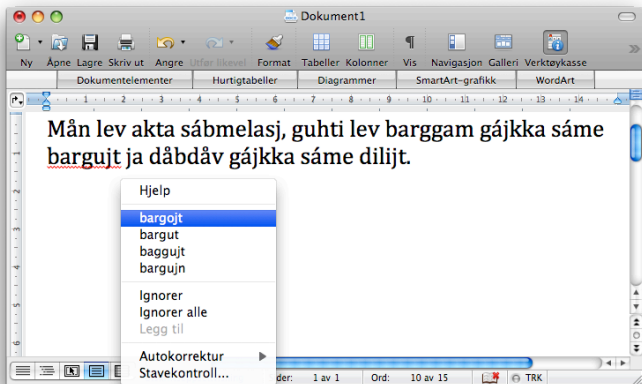
Lyhyt, opastettu kierros 2

- ▶ `http://divvun.no`
- ▶ `http://divvun.org`
- ▶ `http://dicts.uit.no`
- ▶ `http://oahpa.no`

(Kielten puhujille kätevä paketti: Korjausohjelma, e-sanakirja, e-opetusohjelmia)

Sovellukset

Korjausohjelmat saamen kielille — ✓ÁBC̥



How to build a speller:

1. collect word list from corpus
2. build a lexicon & morphology
3. make an error model
4. choose speller type:
 - Microsoft Office
 - Linux/Unix (aspell, hunspell, etc)
5. build a speller engine
6. integrate into applications
7. test the coverage
8. test the error model
9. test the speller engine
10. test the application integration

The **red-coloured areas** are automated in our infrastructure, partly as a direct result of the way it is built, and partly by leveraging development work done by other groups, mainly Voikko[4] and Hfst[5]. What remains is purely linguistic work.

The linguistic nature of the work is further emphasised by our focus on reuse and morphologic (and syntactic) analysis – the lexicon and morphology are directly usable for both morphological analysis and generation. This property of the infrastructure and the tools we build are further utilized in the automated testing, to ensure that all and only the expected word forms are accepted or generated.

URL's:

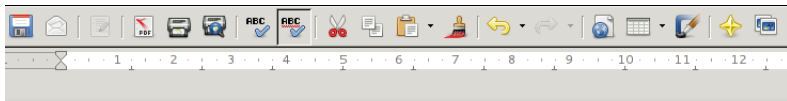
[4]: <http://voikko.sourceforge.net>

[5]: <http://hfst.sourceforge.net>

Kielioppitarkistin

- ▶ Seuraava askel korjausohjelmien jälkeen
- ▶ Ohjelma lukee lauseen ja löytää kieliopilliset virheet
 - ▶ kongruenssivirheet, sijamuotovirheet, sanajärjestysvirheet, ...
 - ▶ kirjoitusvirheet, joiden tulos on olemassa oleva sana

Alfaversio



Ealáhus- ja Luonddubellodat lea vuosttaš gearddi searván Sámedikkegižžui .

real word error (vuosttaš)

Explicacions...

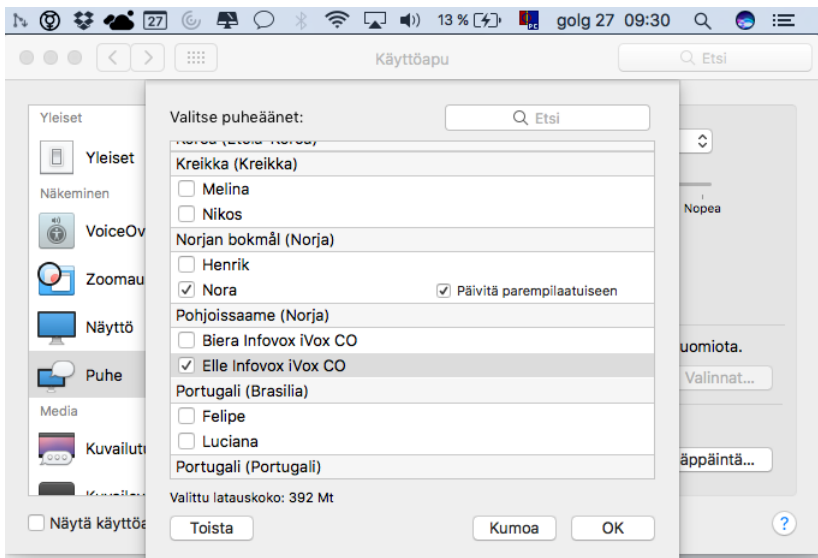
Ignora

Ortografia i gramàtica

Defineix la llengua de la selecció ▶

Defineix la llengua del paràgraf ▶

Synteettinen puhe – jos itse hakee



E-sanakirja taivuttaa sanat



– Norgga stuorimus gilvvut

Badjelaš vahku geahčen álgá Sámi vahku Romssas, ja dat *loahpahuvo* heargegilvuiin. Okta dain buoremus buoraevuiddiin aalaá gilvalit doppe, gilvvus

loahpahuvo

imus heargegilvun....

loahpahit (v.) — 1. avrunde, 2.

slutføre



Mále skeaŋkka gonagassii

Sámedikke presideanta Egil Olli oktan buot ráđđelahtuiguin mále ovtasráđiid álbmogiin gova riegdánbeaivveskeaŋka Gonagas Haraldii. ... **Olles ášši**

... ja tunnistaa huolimattoman kirjoitustavan

Mun vuohun mana. Hui harve vuohun duoji. Na jus dat **manna**
 lea gietkama siste, de diedusge sattan vuohttut duoii maid



Vuohttu go sámii e
 márkaniis boahhte
 Ođđasat.se | Sveri
 sverigesradio.se

Dán vahku álgogea
 duojárat ruotabeal
 bojkottet Johkamo

manna

mánná (subst.) — barn

manjá (prep.) — etter

manjá (adv.) — seinere

manjá (postp.) — etter

mannat (verb) — 1. gå, 2. dra,
 3. reise

Tykkää ·
 Kommentoi

Neahttadigisánit kirjoitusta varten: etsi käännös

The screenshot shows the website of the Kommunal- og moderniseringsdepartementet (Municipal and Modernization Department) of the Norwegian government. The page title is "Kommunal- og moderniseringsdepartementet". A search bar on the right contains the text "Søk ho: moderr".

The main content area features a large article with the heading "Ekspertutvalg skal se på oppgaveløsningen" (Expert committee to look at task resolution). A search tooltip is visible over the word "oppgaveløsningen", showing the following results:

- oppgaveløsningen
- oppgaveløsning (subst.) — 1. dahkamuščoavdin, 2. doaibmačoavdin

Below the article, there is a video player with the title "Video: Nyttårshilsen fra Jan Tore Sanner". The main article text includes the heading "Skal lete etter og fjerne tidstyver i hele staten" (Will look for and remove time thieves in the whole country) and the text: "Regjeringen pålegger alle statlige virksomheter å fjerne tidstyver i egen virksomhet neste år. I tillegg skal de foreslå forenkling av regelverk og arbeidsrutiner fastsatt av andre statlige myndigheter som skaper tidstyver for virksomheten." (The government orders all state-owned enterprises to remove time thieves in their own enterprise next year. In addition, they must propose simplification of regulations and work routines established by other state-owned enterprises that create time thieves for the enterprise.)

At the bottom right of the page, there is a photo of a woman in a blue hoodie with a white star on the sleeve, looking at a computer screen. The caption below the photo reads "Møtet med offentlig s..." (Meeting with public s...).

... saat käännöksen

Ordbøker

Nordsamisk → Norsk

Norsk → Nordsamisk

Nordsamisk → Finsk

Finsk → Nordsamisk

Andre ordbøker

Norsk → Nordsamisk (⇒ Snu)

oppgaveløsning

Søk

Søk i tekst

oppgaveløsning er en mulig form av ...

løsning (subst.)

løsning

subst. entall ubest.

- čovdojupmi, čoaivddus

oppgave (subst.)

oppgave

subst.

1. dahkamuš, bargu, doaibma, doaibmamuš, bargodoaibma, bargamuš (arbeid)
2. bargobihtá (problem å løse, spørsmål å besvare)

oppgaveløsning (subst.)

oppgaveløsning

subst. entall ubest.

- dahkamuščoaivdin, doaibmačoaivdin

... ja näet sen kontekstissa

norwegian ▾

word ▾ is ▾

<any word> Aa

or

+

Fler språk Search

KWIC: hits per page: 25 ▾ sort within corpora: not sorted ▾

KWIC

Results: 3

Show context

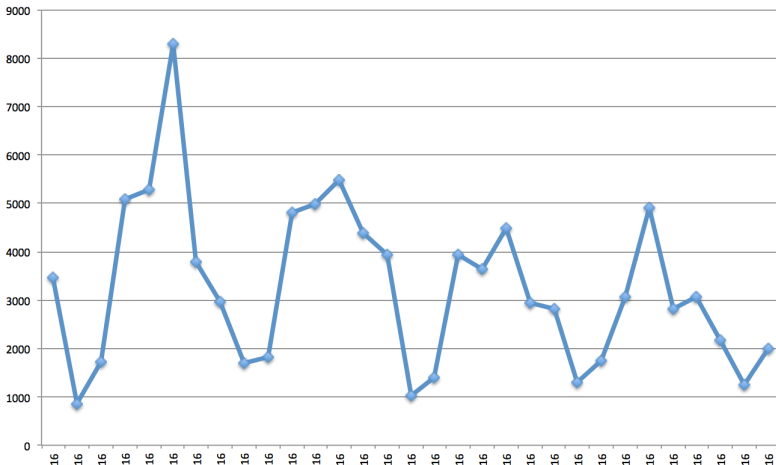
NORSK-NORDSAMISK (does not support extended context)

nom samordning og samarbeid om **oppgaveløsning** mellom sektormyndigheter og mellom statlige, regionale og kommunale orga
 id go suorgeeisevålddiid gaskasaš, ståhta , regionála ja gieldda orgánaid , priváhta organisašuvnnaid ja ásahasaid , ja almmolašvuoda
 nom samarbeid oppná effektivitet i **oppgaveløsning** ved at ressurser samordnes.
 rráduš Árvaluvvon fylkkaplána meannuduvvui fylkkadikkis geassemánu 11. b. 2003 .
 justisvesenets **oppgaveløsning** og service til publikum
 stisadoaimmahaga bargguid bargat ja gozihit báivlusaid álbmogii

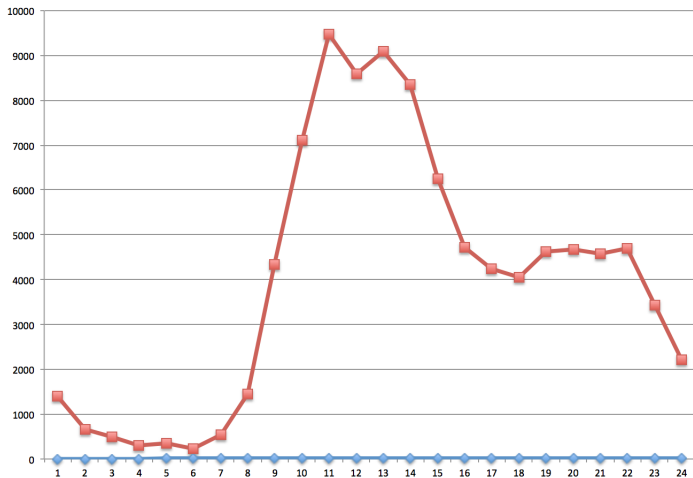
`http://gtweb.uit.no/all_nds.html`

Sanakirjakäyttö helmikuussa 2016: Päivästä toiseen

Searches



Sanakirjakäyttö helmikuussa 2016: Tunti tunnilta



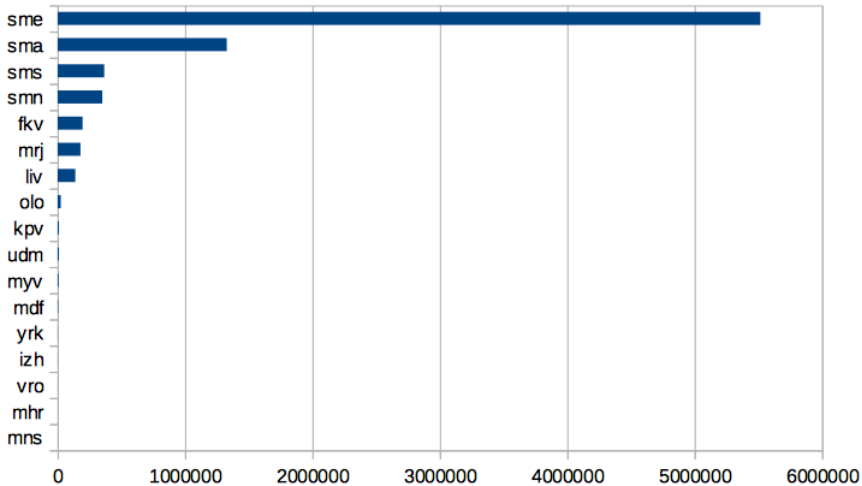
Sanakirjakäyttö eilen (torstaina, 20.9. 2018)

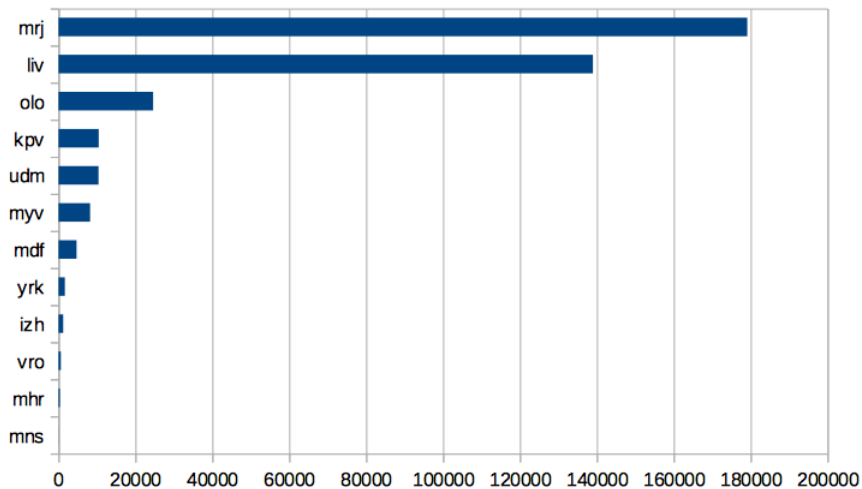
Sanakirja	Haut
Pohjoissaame	9577
Koltansaame	2015
Eteläsaame	1829
Inarinsaame	1796
Länsimari	891
Liivi	505
Kveeni	287
Livvi	176

Table:

Neahttadigisánit 27 kielelle

- ▶ Pohjoissaame: <http://sanit.oahpa.no>
- ▶ Eteläsaame: <http://baakoeh.oahpa.no>
- ▶ Inarinsaame: <http://saanih.oahpa.no>
- ▶ Koltansaame: <http://saan.oahpa.no>
- ▶ Pohjoiset itämerensuomalaiset kielet: <http://sanat.oahpa.no>
- ▶ Eteläiset itämerensuomalaiset kielet: <http://sonad.oahpa.no>
- ▶ Mordvalaiskielet: <http://valks.oahpa.no>
- ▶ Nenetsi ja mansi: <http://vada.oahpa.no>
- ▶ Marilaiset kielet: <http://muter.oahpa.no>
- ▶ Permiläiset kielet: <http://kyv.oahpa.no>
- ▶ Preerie-cree: <http://itwewina.oahpa.no>
- ▶ Pohjoishaida: <http://guusaaw.oahpa.no>





Konekäännös

Konekäännöksen kaksi **tavoitetta**

1. Tekstin ymmärtäminen
2. Tekstin kirjoittaminen
 - ▶ Kone kääntää tekstejä kapeista domeeneista: Esim. ohjelmavalikot/lokalisointi, käyttöohjeet, säätiedotukset .. ja ihminen oikolukee

Konekäännöksen kaksi **menetelmää**

1. Koneoppiminen
2. Sääntöpohjainen menetelmä

Koneoppiminen

- ▶ Käännöskone oppii rinnakkaiskorpuksen perusteella.
- ▶ Mitä enemmän tekstiä, sitä parempi tulos.
- ▶ Rikas taivutusmorfologia on tälle menetelmälle hankala.

Google suomi-englanti: Sujuva englanti

- ▶ Alkuperäinen teksti: Heikki Ylikangas *Kalevala*-lehdessä 2011
 - ▶ Jonasta tyydyttää suomalaistutkimuksen päälinja. Hän ottaa kantaa vain joihinkin tulkintoihin.
 - ▶ Yksi näistä harvoista on allekirjoittaneen käsitys siitä, että Suomi pani toivonsa luovutettavien alueiden palauttamisessa Saksaan jo talvisodan eikä vasta välirauhan aikana.
- ▶ Google Translate
 - ▶ **Tina** is satisfied with the main line of Finnish research. He only takes a stand on some interpretations.
 - ▶ One of these few has been **a sign of the fact** that Finland has expressed hope for the return of the divested areas to Germany as early as the Winter War and not just during **the Middle Ages**.
- ▶ Suomi bijai doaivvus **luovutettavien** guovlluid **palauttamisessa** Duiskii juo dálvesoađi ii easka gaska ráfi áigin.

Sääntöpohjainen konekäännös

Lähdekielen lingvistinen analyysi

```
"<Mus>"  
  "mun" Pron Pers Sg1 Loc <hab> @ADVL>  
"<lea>"  
  "leat" V IV Ind Prs Sg3 @+FMAINV  
"<bárdni>"  
  "bárdni" N Sem/Hum Sg Nom <ext> @<SUBJ  
"<.>"  
  "." CLB
```

```
"<Šiljus>"  
  "šillju" N Sem/Plc Sg Loc @ADVL-ine>  
"<lea>"  
  "leat" V IV Ind Prs Sg3 @+FMAINV  
"<bárdni>"  
  "bárdni" N Sem/Hum Sg Nom <ext> @<SUBJ  
"<.>"  
  "." CLB
```

Lähdekielen lingvistinen analyysi

"<Mus>"

"mun" Pron Pers Sg1 Loc <hab> @ADVL>

"<lea>"

"leat" V IV Ind Prs Sg3 @+FMAINV

"<bárdni>"

"bárdni" N Sem/Hum Sg Nom <ext> @<SUBJ

"<.>"

"." CLB

"<Šiljus>"

"šillju" N Sem/Plc Sg Loc @ADVL-ine>

"<lea>"

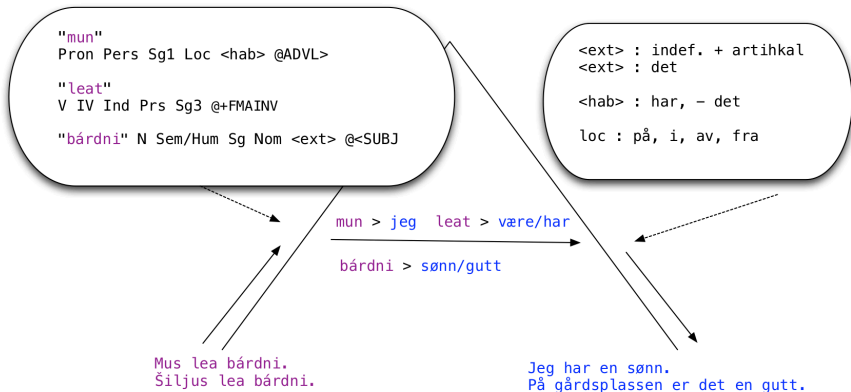
"leat" V IV Ind Prs Sg3 @+FMAINV

"<bárdni>"

"bárdni" N Sem/Hum Sg Nom <ext> @<SUBJ

"<.>"

"." CLB

Sääntöpohjainen konekäännös, <http://jorgal.uit.no/>

Saamen korkeakoulun nettisivu



Sámi allaskuvlla
Sámi University of Applied Sciences

Sámegiili | Norsk (beta) | English

RUOKTOT

OAHPUT

STUDEANTTAIDE

GIRJERÁJUS

DUTKAN

SA BIRRA



DUTKIIDSEARVI

ORGANISAŠUVDNA



STIVRRAT JA LÁVDEGOTTIT



BARGIT

LATNJAKÁRTA

SA birra

**Sámi oahpahas – Sámi dutkan – Sámi ásahas
1989 - 2019**

Sámi allaskuvlla almmolaš rahpan dáhpáhuvai skábmamánu 1. beaivvi 1989. Nu ollašuvai sámiid gaskkas guhkit áiggi sávdalat fállat sámii oahpaheaddjioahpu - ja das marjil eará ge alit dási oahpuid –

... norjaksi



ORGANISASJON +

STYRER OG KOMITEER +

FAGFORENINGER +

ANSATTE

ROMKART

om SH

Samisk undervisning – Samisk forskning – Samisk institusjon 1989 - 2017

Samisk høgskoles offentlig åpning skjedde den 1. november 1989. Slik gjennomførtes det mellom samer lengre tid ønske å tilby samisk lærerutdanning - og etter dette annen også høyere nivå

Konekäännöksen paradoksi

- ▶ Koneoppimismenetelmä tuottaa sujuvaa kieltä, sopii parhaiten posteditoimiseen.
 - ▶ Silti sitä käytetään ymmärtämistä varten, vaikka se ei välttämättä käännä lähdekieltä
- ▶ Säätöpohjainen menetelmä tuottaa kömpelöä kieltä
 - ▶ Silti se on enemmän käytössä tekstituotantoa varten

Konekäännös vähemmistökielille?

- ▶ **Tekstin tuotanto:** Konekäännös tehokkuuden parantajana
 - ▶ Norjan valtio tuskin hyväksyy, että saamenkielisten virallisten sivujen käännöskustannukset ovat isompia kun englanninkielisten
 - ▶ Tehokkaammalla käännösmenetelmällä voidaan tuottaa lisää vähemmistökielisiä tekstejä
- ▶ **Konekäännös tekstin ymmärtämistä varten:**
 - ▶ ... tarjoaa samalla *vähemmistölle vapauden* käyttää vähemmistökieltä (vrt. Saamelainen korkeakoulu)

Monikielisyyden arkipäivä

- ▶ Kaarasjoen ala-asteen viikkosuunnitelma:
<http://bit.ly/2eWcyEr>

Kolmas käyttötarkoitus: Konekäännösohjelma kirjoitusapuna

Esimerkkilause:

Lea vejolaš geavahit dihtorjorgaleami riektečállinprogramman.
Maŋgii prográmma lea maid gávnnan giellaoahpalaš feillaid.

<http://tiny.cc/6k7vyy>

Kieliteknologia saamen kielille tänään

- ▶ Toimivat sovellukset
 - ▶ Korjausohjelmat, e-sanakirjat ovat olemassa, ja käytössä
 - ▶ Synteettinen puhe on olemassa (ei niin paljon käytössä)
 - ▶ Konekäännös on aika aikaisessa vaiheessa
 - ▶ Kielentutkijoille on tarjolla annotoituja korpuksia
- ▶ Puuttuvat / tulevat (?) sovellukset
 - ▶ Kielioppitarkistimet ovat tulossa
 - ▶ Kieliteknologia kännyköille on lastenkengissä
 - ▶ jne.
- ▶ Muiden saamen kielten sovellukset eivät ole pohjoissaamen tasolla

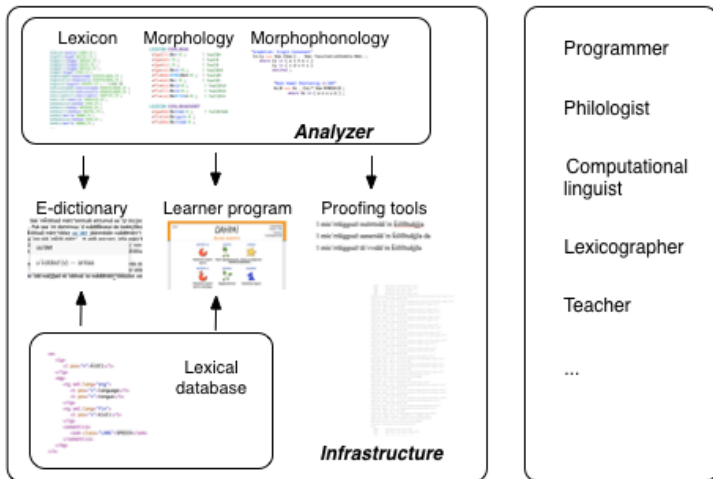
UiT: Kansainvälinen yhteistyö

Tutkimusryhmät	Kielet	(Kielet)
Giellatekno + Divvun (UiT)	11	8
Tutkimusryhmät, jotka käyttävät UiT-palvelimia:		
Pyssyjoki, Norja ▶ Kainun institutti	1	
Helsinki, Suomi ▶ AKU	8	6
Tartto, Viro ▶ SAMEST	2	1
Syktvykar, Venäjä ▶ FU-lab	1	
Ulan-Ude, Venäjä ▶ BSU	1	5
Edmonton, Kanada ▶ UofAlberta	2	4
Nuuk, Grönlanti ▶ Oq	1	
Yhteensä 51 kieltä:	27	24

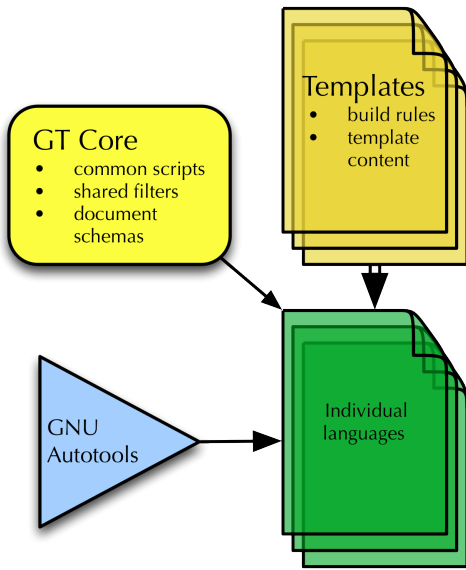
Nämä kielet käyttävät UiT:n infrastruktuuria



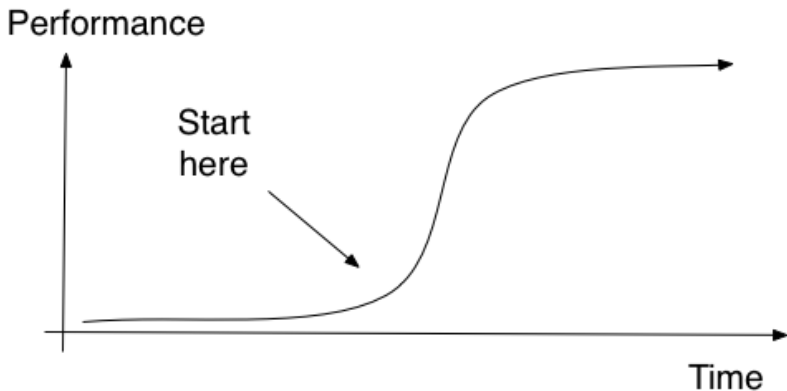
UiT:n Giella-infrastruktuuri (huom! työryhmä!)



Tuotantolinja on sama kaikille kielille



Giella-infrastruktuuri tarjoaa mahdollisuuden tulla mukaan S-käyrän keskellä



Venäjälläkin?

Vertailun vuoksi: Miten kieliteknologia auttaa saamelaisia kieliyhteisöjä?

- ▶ Se muuttaa kirjoituskulttuuria.
(esimerkkeinä päivälehti *Ávvir*)
- ▶ Se auttaa kielenelvyttämistä...
Missä määrin se auttaa jätän sosiolingvisteille tutkimustehtäväksi

Onko Venäjä erilainen?

- ▶ Venäjän uralilaiset kieliyhteisöt eivät toistaiseksi käyttydy kuten saamelaiset Pohjoismaissa
 - ▶ Älymystön uudet sukupolvet rekrytoidaan maalta, niiden lapset kasvaavat venäjänkielisiksi
- ▶ Kielenelvyttämisestä ei siis ole vielä (???) puhe

Venäjän suomalais-ugrilaiset tasavallat

- ▶ Tasavallat ja autonomiset alueet tarvitsevat kieliteknologiaa
 - ▶ ... niin kauan kuin saavat käyttää kieltään
 - ▶ <http://fu-lab.ru/conference/node/12>
(Glavnaja, dokumenty)

Komilaisten FU-lab

- ▶ <http://fu-lab.ru>
- ▶ Näppäimistöt, korpuukset,
Korjausohjelma hunspell-menetelmän mukaan
- ▶ Hunspellin edut: Komilaiset tekivät kaiken itse
(Norjan projekti ei koskaan saanut rahaa)
- ▶ Hunspellin haitat: Hunspell tarjoaa vain korjausohjelman, ei
muuta

Kieliteknologian asema fennougristiikan (ja saamen tutkimuksen) oppiaineessa

- ▶ ... on tällä hetkellä marginaalinen, jos sitäkään
 - ▶ samalla Giellatekno + Divvun = maailman suurin työpaikka saamen kielentutkijoilla
- ▶ Kieliyhteiskunta törmää jatkuvasti tietokoneongelmiin
 - ▶ Tavallisimmat työpaikat (kielten opetteja, sanakirja/terminologia-työ) käyttävät molemmat kieliteknologiaa
- ▶ ... jonka takia meidän pitäisi ehkä harkita asiaa

Entäs kielihistoria – miten sille kävi?

- ▶ Tutkimus on aina poliittista
 - ▶ Suomen fennougristiikan tehtävänä oli luoda suomalaisille loistavan menneisyyden
 - ▶ Norjassa jo suomen ja saamen kielten valinta tutkimusaiheena on antirasistinen työ
 - ▶ Sáogatissa kielihistoria käsitellään harva se päivä (“ketkä tulivat ensin?”)
 - ▶ ... saamelaisten vastustajille kielihistorian tehtävänä on ottaa saamelaisilta oikeudet pois
- ▶ Kielihistoria on siis yhtä poliittinen – ja tarpeellinen – kuin se oli Seuran perustamisvuonna 1883

Onko uralilainen kieliteknologia tuonut tutkimustuloksia?

1. Artikkelit kielimalleista
 - ▶ Oma konferenssisarja: <https://sisu.ut.ee/iwclul2019>
2. Artikkelit sovellusten yhteiskunnallisista vaikutuksista
3. Artikkelit, jotka käyttävät SIKOR-ressursseja hyväksi

Onko uralilainen kieliteknologia muuttanut maailmaa?

- ▶ Saamelaisten kieliyhteisöt kyllä:
 - ▶ Päivälehti joka päivä
 - ▶ Kirjoittajien määrä kasvaa
- ▶ Vielä on parantamiset varaa:
 - ▶ Ávvirissa on lukijan palsta korkeintaan kerran kuukaudessa, jos sitäkään
 - ▶ Ei yksikään saamelainen kirjoittaa Wikipediaan
- ▶ Entäs Venäjällä?
 - ▶ ... ei yhtä paljon.

Miten osallistua uralilaiskielten kieliteknologiaan?

- ▶ Kaikki ovat tervetulleita...
 - ▶ käyttämään tutkimusapuneuvoja (korpuksia)
 - ▶ tekemään vastaavia kielimalleja “omalle” kielelle, kehittämään sovelluksia
 - ▶ joko meidän palvelimella tai ottamalla meidän lähdekoodin mukaan
- ▶ Mikä on tilanne nyt?

Saamelaiskielten analysaattorien kattavuus

	Leksikko	Kattavuus
Pohjoissaame	152.332	96,9 %
Inarinsaame	47.849	93,4 %
Luulajansaame	48.977	92,5 %
Eteläsaame	61.485	92,3 %
Koltansaame	40.356	70,0 %

Table: Kattavuus: Lene Antonsen 2018: Eai buot sánit čázi jeage – giellamáille huksen ja heiveheapmi duohta giellamáilbmái. PhD diss. UiT

Venäjän uralilaiskielten analysaattorien kattavuus

	Leksikko	Kattavuus
Itämeri	68.917	95.94 %
Komi	181.422	87.55 %
Nenetsi	18.299	75.68 %
Udmurtti	51.829	71.96 %
Mokš	31.857	70.23 %
Ersä	118.173	69.47 %
Länsimari	23.315	69.27 %
Permjakki	46.183	68.87 %

Henkilöt? Rahaa?

1. Kieliteknologia on meille hyödyksi – me voidaan joka tehdä se työaikana
2. Kieliteknologia tarjoaa hakemuksen kirjoittajalle suuria hypetyksen mahdollisuuksia
3. Kieliteknologia tarjoaa myös *uudet* rahalähteet
 - ▶ Ei vain tutkimusta varten – mutta myös tekemistä varten
4. Tyypillinen työryhmä
 - ▶ Infrastruktuuri (Tromssa) + kieliteknologi + fennougristi(t) + puhuja(t)

Yhteenveto

- ▶ Kieliyhteisöille:
 - ▶ Sääntöpohjaiset kielioppimallit ja avoin lähdekoodi tarjoavat mahdollisuuden pärjätä digitaalisessa maailmassa myös kielille, joilla on runsas taivutusmorfologia
- ▶ Meille fennougristeille:
 - ▶ Kielioppimallit ovat kattavampia kuin tavalliset referenssikieliopit, ja sen takia kiinnostavia meille fennougristeille
 - ▶ Kieliteknologia ja annotoidut korpuukset avaavat uuden vaiheen kielentutkimuksessa
 - ▶ Sosiolingvistit ja kielensuunnittelijat saavat tehokkaita työkaluja

*<http://giellalt.uit.no>, <http://giellatekno.uit.no>
<http://divvun.no>, <http://divvun.org>*