

## Nokre av prosjekta våre

Trond Trosterud  
Senter for samisk språkteknologi  
Institutt for språkvitskap  
Universitetet i Tromsø



*Tromsø, 7. oktober 2013*

## Innhald

AKU - uralske språk

AltLab - Edmonton i Alberta

SamEst: Tromsø - Tarto

FAD2

Samiske dialektband

Nasjonalbibliotekets digitaliseringsprosjekt

∅

# AKU - uralske språk

## AKU - uralske språk

- ▶ Jack Rueter
- ▶ 5 andre arbeider på prosjektet
- ▶ Erzya, Komi, Moksha, Tundra Nenets, Western Mari,
- ▶ Ingrian, Livonian, Olonetsian, Veps
- ▶ Eastern Mari, Udmurt
- ▶ Khanty, Komi, Nganasan

## AKU i Neahttadigisánit

- ▶ `sanat.oahpa.no`
- ▶ `valks.oahpa.no`
- ▶ `muter.oahpa.no`
- ▶ `kyv.oahpa.no`
- ▶ `vada.oahpa.no`

# AltLab - Edmonton i Alberta

## AltLab - Edmonton i Alberta

- ▶ Antti Arppe, Jordan Lachler, Conor Snoek, Dorothy Thunder, Juhani Järvikivi
- ▶ Præriecree
- ▶ Nord-Haida
- ▶ <http://altlab.artsrn.ualberta.ca/>

## Nordamerikanske ordbøker i Neahttadigsánit

- ▶ `pikiskwewina.oahpa.no`
- ▶ `guusaaw.oahpa.no`



## Plains cree stavekontroll x 2

- ▶ «Rumensk»: latinskbasert præriecree
- ▶ «Tsjekkisk»: stavingsalfabetsbasert præriecree

## Søknader, AltLab

- ▶ Det kanadiske forskingsrådet (100 siders søknad!)
- ▶ Mobilitetsstipend (Lene)

# SamEst: Tromsø - Tarto

## SamEst: Tromsø - Tartu

- ▶ Estisk: Heiki-Jaan Kaalep, Heli Uiho Jaak Pruulmann
- ▶ Võru: Sulev Iva
- ▶ <http://giellatekno.uit.no/ped/samest/Samest.html>

## SamEst: Tre prosjekt

- ▶ Estisk Oahpa for russarar og utlendingar
- ▶ Võru-oahpa: revitaliseringsprosjekt
- ▶ Maskinomsetjing: Finsk til estisk / nordsamisk

# FAD2

## FAD2

- ▶ Ein del hundre tusen frå KMD for å arbeide med sør- og lulesamisk
- ▶ Korpusinnsamling, parallellkorpus
- ▶ Termekstrahering ???

# Samiske dialektband



## Samiske dialektband

- ▶ LIA - digitaliseringsprosjekt i Oslo:
- ▶ Nasjonalbiblioteket digitaliserer, vi transkriberer
- ▶ UiT: 3 mnd til Arnstein Johnskareng

# Nasjonalbibliotekets digitaliseringsprosjekt

## Nasjonalbibliotekets digitaliseringsprosjekt

- ▶ Nasjonalbiblioteket digitaliserer *alt* som er publisert i Noreg
- ▶ ... til no utan samiske bokstavar
- ▶ Men frå og med i går også med samiske bokstavar

## Evaluering og framtidsutsikt

- ▶ I dag ca. 1% feil
- ▶ Det er alt for mykje, men kan ssv. bli redusert til 0.1%
- ▶ Med det vil vi kunne få eit *mykje* større korpus
- ▶ ... og vi vil kunne få eldre tekstar



## 0

- ▶ Stor Ð har tre små: đ, ð,
- ▶ Liten ɳ har to store: Ɔ og Ɔ (!)
- ▶ N-aktig stor ɳ: samisk
- ▶ n-aktig stor ɳ: afrikanske språk

## no og i framtida

- ▶ Times: N-aktig
- ▶ Times New Roman: n-aktig
- ▶ Status quo eller splitt?
- ▶ Eit potensielt politisk problem for Noreg