

Wikipedia som indikator og motor for språkleg revitalisering

Trond Trosterud

Giellatekno — Senter for samisk språkteknologi

Institutt for språkvitskap

Tromsø, 24. april 2015

Innhald

Innleiing

Wikipedia som indikator for språkleg revitalisering

Språksamfunna

WP-versjonane

Skribentane

Wikipedia som motor for språkleg revitalisering

Wikipedia i språkteknologisk forskning

Uralsk Wikipedia som korpus

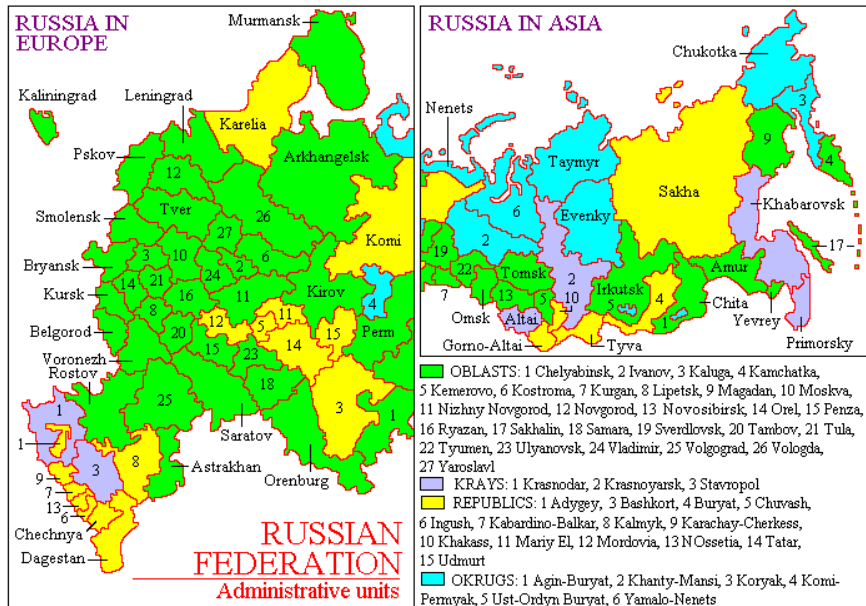
Konklusjon

Innleiing

- ▶ Kva faktorar fører til fungerande WP-versjonar?
- ▶ Kva konsekvensar har desse WP-versjonane?

Uralske språk i Russland

- ▶ Mordvinske språk
 - ▶ Erzja, Moksja
- ▶ Mariske språk
 - ▶ Vestmarisk, austmarisk
- ▶ Permiske språk
 - ▶ Udmurtisk, komi



Autonomous Areas in Russia



— Autonomous republic

- - - - Autonomous okrug

Birobijan is the only autonomous oblast in Russia.

^a *Chukotka Autonomous Okrug may now be independent of Magadan Oblast.*

Boundary representation is not necessarily authoritative.

Wikipedia som indikator og motor for språkleg revitalisering

└ Wikipedia som indikator for språkleg revitalisering

└ Språksamfunna



Dei ulike språksamfunna

- ▶ Moksja og erzja vart innlemma i Russland først
- ▶ Marisk og udmurtisk har vore under tyrkisk påverknad, mindre under russisk
- ▶ Austmarisk er det mest aktive språksamfunnet

Tabell: Wikipedia på uralske språk

Språk	WP	Talarar	WP-art.	Ord i WP	ord/talar	ord/art	Akt. skrb
Moksja	mdf	200000	1221	69888	0,35	57,2	2
Erzja	myv	400000	1863	70693	0,18	37,9	2
Vestmarisk	mrj	36822	7190	258188	7,01	35,9	1
Austmarisk	mhr	414000	6661	378925	0,92	56,9	2
Udmurtisk	udm	324000	3582	144685	0,45	40,4	4
Komi	kv	156000	4373	202852	1,30	46,4	16
Nordsamisk	se	25700	7166	174940	6,81	24,4	5
Færøysk	fo	66000	10929	1093696	16,57	100,1	29

Kommentarar

- ▶ (Nesten) alle WP-versjonane har bortimot dobbelt så lange artiklar som nordsamisk WP.
- ▶ Alle dei uralske WP-versjonane er små, men likevel store til minoritetsspråk å vere
- ▶ Austmarisk har dei lengste artiklane
 - ▶ ... og vestmarisk stikk seg ut

Wikipedia på uralske språk

- ▶ <http://mdf.wikipedia.org>
- ▶ <http://myv.wikipedia.org>
- ▶ <http://mrj.wikipedia.org>
- ▶ <http://mhr.wikipedia.org>
- ▶ <http://udm.wikipedia.org>
- ▶ <http://kv.wikipedia.org>

Skribentane

- ▶ Gjestearbeidarane:
`https://mrj.wikipedia.org/wiki/User:Kmoksy`
- ▶ Emigrantane:
`https://mrj.wikipedia.org/wiki/User:Tsikma`
- ▶ Dei innfødde

Gjestearbeidaren: Kmoksy

- ▶ <https://en.wikipedia.org/wiki/Special:CentralAuth/Kmoksy>
- ▶ Kmoksys metode:
 - ▶ For an unknown language: Copy text from model articles + Dictionary + Grammar = Simple sentences and Definitions. My methodology is this.
- ▶ Kmoksy si helsing til oss:
 - ▶ <http://tinyurl.com/pj93949>

Emigranten: Tsikma

- ▶ Arbeider i lag med Kmoksy
- ▶ Bur i Finland, og er svært aktiv
- ▶ `http://stats.wikimedia.org/NN/TablesWikipediaMHR.htm`

Dei innfødde: Sergej Ivanov

- ▶ Bur i Mari El, skriv aktivt
- ▶ skriv berre på ein Wikipedia
- ▶ `http://stats.wikimedia.org/NN/TablesWikipediaMHR.htm`

Skilnaden mellom Russland og Vest-Europa

- ▶ I vest har minoritetane i vår generasjon ikkje lært å skrive
- ▶ I Russland hadde alle minoritetane morsmålet sitt på skolen
- ▶ og alle er dermed potensielle skribentar

Wikipedia i språkteknologisk forskning

- ▶ <http://www.aclweb.org/anthology/>
- ▶ 34000 artiklar
 - ▶ Ref til Wikipedia: 3830
 - ▶ French Wikipedia: 900
 - ▶ Norwegian Wikipedia 137

Andre bruksområde

- ▶ Testmateriale for stavekontrollar
- ▶ Interwikilenkjer for leksikalske ressursar
- ▶ Testmateriale for grammatisk utvikling

Uralsk Wikipedia som korpus

- ▶ muter.oahpa.no
- ▶ valks.oahpa.no
- ▶ kyv.oahpa.no

Konklusjon

- ▶ Wikipediaversjonane i Russland er langt større og betre enn for minoritetsspråk på andre kontinent
- ▶ Med så små skribentsamfunn er fordelinga av gode eller dårlege WP-versjonar til ein viss grad tilfeldig
- ▶ Gjestearbeidarane har gjort ein betre jobb her enn mange andre stader
- ▶ Ei svært lita gruppe skribentar kan gjere ein stor skilnad for eit språksamfunn