

Samisk språkteknologi og Samisk skolehistorie – samarbeidet til nå og videre framover

Berit Merete Nystad Eskonsipo
Senter for samisk språkteknologi
<http://giellatekno.uit.no>



Introduksjon

- ▶ Samisk språkteknologi ved Universitetet i Tromsø
 - ▶ **Divvun**: lage og vedlikeholde ord- og grammatikkretteprogram for samisk, korpusinnsamling
 - ▶ **Giellatekno**: lage grammatikk- og syntaktisk basert språkteknologi for samisk (f.eks Oahpa, digitale ordbøker, oversettelsesprogram) + forskning

- ▶ Hvorfor er Samisk skolehistorie viktig i arbeidet med samisk språkteknologi?

Korpus

- ▶ Elektronisk samling av tekster
 - ▶ systematisk sammensatt tekstsamling
 - ▶ store mengder løpende tekst
 - ▶ kan søkes i og analyseres ved hjelp av dataprogram
- ▶ Typer av korpus:
 - ▶ enspråklig korpus
 - ▶ parallellkorpus (originale tekster og oversettelser)

Korpus med samiske tekster

- ▶ Divvun har siden 2005 samlet inn samiske tekster for Sametinget og bygget opp et samisk elektronisk korpus (flestep tekster på nordsamisk, men også tekster på lule- og sørsamisk)
 - ▶ Bundet korpus - krever innlogging og kontrakt med Sametinget (avistekster, skjønnlitteratur, Det nye testamentet)
 - ▶ Frikorpus - kan brukes og lastes ned fritt (Samisk skolehistorie, faktatekster, administrative tekster, lovtekster)

Hvor stort er parallellkorpuset vårt?

- ▶ Bokmål - nordsamisk
 - ▶ Bokmål: 2,4 mill. ord
 - ▶ Nordsamisk: 2,1 mill. ord
 - ▶ 156 500 setningspar
- ▶ Jf. andre parallellkorpus
 - ▶ *Europarl (EU x EU)*: 50 millioner ord per språk
 - ▶ *OPUS*: 100 - 500 millioner ord per språk, for ulike språkpar
 - ▶ *Acquis Communautaire*: 20 - 60 millioner ord / språk
(Norsk: 14 millioner)

Samisk skolehistorie som korpus

- ▶ Samisk skolehistorie bind 1-5 - del av frikorpuset
- ▶ Alle artiklene fins på minst to språk, hvorav minst et samisk språk
- ▶ Består både av reine faktatekster og fortellinger som språklig sett ligger nærmere opp mot skjønnlitteratur enn resten av parallellkorpuset (administrative tekster, lovtekster)

Samisk skolehistorie som korpus

Tabell: Ord i Samisk skolehistorie 1-5

Nordsamisk	362 347
Sørsamisk	14 332
Lulesamisk	8 545
Bokmål	317 540
Nynorsk	145 525
Engelsk	49 094
Dansk	1 092

faget « flyttsamekunnskap » ble innført i	framhaldsskolen	i Karasjok med et 108 timers kurs
Das lei álgu 1956:s go « johttisámemáhtu » (flyttsamekunnskap) álggahuvvui joatkkaskuvllas (framhaldsskuvllas) Kárášjogas . Dan kursas lei 108 diimmu oahpahas , mii juhkkovjuvui 18 vahkkui , ja lei 6 diimmu vahkkui oahpahas .		
på 18 uker á 6 timer .	Framhaldsskolen	for flyttsamer var 2-årig , og det
Joatkkaskuvla lei 2 jagi johttisámiide , ja lei eaktun ahte nuppi jagi galget dát diimmut adnot servodatohppui ja dearvvašvuodaohppui .		
skolehistorie 1 .) Med lov om	framhaldsskoler	av 1946 blei alle kommuner pålagt å
realskuvla ja álfbmotallaskuvla (ovdal gohčoduvvon nuoraidskuvlan) .		
. I Finnmark hadde det ikke vært	framhaldsskoler	før krigen , men nå blei dette
Go 1946:s bodii joatkkaskuvlaláhka (lov om framhaldsskoler) , geatnegahttojedje buot suohkanat fállat joatkkaskuvlla .		
nå blei dette skoleslaget raskt utbygd .	Framhaldsskolen	blei blant annet brukt til etterskoling av ungdom
Finnmárkkus eai lean leamaš joatkkaskuvllat ovdal soadi , muhto dál huksejuvvojedje dákkár skuvllat jođánit .		
videre utdanning . Her ser vi på	framhaldsskolen	gjennom dokumenter fra 1950-tallet og et tilbakeblikk
Dás mii geahčadit joatkkaskuvlla muhtin 1950-logu dokumeantaid vuođul ja muittašallamiiguin 50 jagi maŋŋil .		
et tilbakeblikk 50 år etter . Mens	framhaldsskolen	fantes i de fleste samiske kommuner og
Joatkkaskuvla han lei eanaš sámi suohkaniin , ja das lei maiddái muhtin muddui oahpahas sámi fáttáin ja ealáhusain . Eai lean nu galle suohkana main lei realskuvlafálaldat , ja realskuvllain ii oro leamaš makkárge sámi sisoallu , ii ge sámegeiella fágan .		
blei faget « flyttsamekunnskap » innført i	framhaldsskolen	i Karasjok Kárášjohka i 1956 og
Muhto nuppe dáfus álggahuvvuige fága « johttisámemáhtu » Kárášjoga joatkkaskuvllii 1956:s ja guokte jagi dan maŋŋil álggahuvvui Kárášjoga nuoraidskuvllas boazodoallosuorgi .		

Figur: Parallellkorpus i bruk, søkeord **framhaldsskole**

Nob-sme oversetting av framhaldsskole og omgangsskole

Tabell: Framhaldsskole

Joatkkaskuvla	9
Framhaldsskuvla	(2) 1
Stuoraskuvla	1
Til sammen	11

Tabell: Omgangsskole

Johttiskuvla	7
Vuorroskuvla	1
Til sammen	8

Hvorfor er korpus viktig og hva kan det brukes til?

- ▶ Forbedre kildefiler
- ▶ Teste språkteknologiske analyse- og grammatikkverktøy (Divvun, syntaktisk analyse)
- ▶ Finne ordpar og eksempler til ordbøker
- ▶ Undersøke bruken av ord
- ▶ Undersøke hvordan enkelte ord blir oversatt (Mer variasjon enn i ordbøker)
- ▶ lage grammatikkoppgaver for studenter
- ▶ Utarbeide oversettelsesminne
- ▶ Utnytte disse ressursene i språkteknologisk forskning og utvikling

Konklusjon

- ▶ Samisk skolehistorie er viktig for samisk språk teknologi fordi:
 - ▶ den fins fritt tilgjengelig i elektronisk format
 - ▶ den utgjør en stor del av samisk - norsk parallellkorpus
 - ▶ den består av tekster som ligger nærmere skjønnlitteratur enn resten av parallellkorpuset

Takk til mine kolleger ved Giellatekno og Divvun
for fruktbare diskusjoner

Giitu! :-)