

Sørsamisk stavekontroll

Eit framhald av Divvun-prosjektet

Oppdatert versjon, juni 2007

Table of contents

1 Mål.....	2
2 Status pr september 2005.....	2
3 Planlagt arbeid.....	2
3.1 Lingvistisk arbeid:.....	2
3.2 Datalingvistisk arbeid:.....	2
3.3 Sørsamisk ordbok.....	3
3.4 Korpusinnsamling.....	3
3.5 Stavekontrollspesifikt arbeid.....	3
3.6 Rettskrivingsspørsmål for sørsamisk.....	3
4 Ferdige modular frå Divvun-prosjektet.....	4
5 Anna arbeid.....	4
6 Stillingar.....	4
7 Timeplan.....	4
8 Budsjett.....	5

1 Mål

Å laga eit sett med sørsamiske stavekontrollar, med same oppsett som for lulesamisk. Dette betyr:

- stavekontroll for MS Office for Windows
- stavekontroll for MS Office for Mac
- stavekontroll for OpenOffice.org
- minst ein stavekontroll basert på open kjeldekode
- orddeling i MS Office, Win og Mac

Jf elles planane for det noverande Divvun-prosjektet, på www.divvun.no.

2 Status pr september 2005

Det er gjort ein del lingvistisk arbeid, fyrst og fremst med substantivbøying og omlydsreglar for substantiv. Alle usamansette substantiv er skrive inn. Ei relativt komplett liste over det eksisterande arbeidet er:

- substantivleksikon med alle usamansette substantiv frå Sydsamisk-norsk ordbok (kring 4300 ord)
- namn (kring 24 000)
- bøyingsmorfologi for substantiv
- morfologi for pronomen og andre lukka ordklasser
- tonivåreglar for nominal bøyning

Dette utgangspunktet er vesentleg dårlegare enn det tilsvarande var for lulesamisk då Divvun-prosjektet starta mot slutten av 2004. Der var den morfologiske modellen meir eller mindre komplett for alle opne ordklasser, og tonivåreglane ganske heildekkjande. Det som mangla for lulesamisk var ein skikkeleg gjennomgang av ein morsmålstalar for å komplettera og retta feil.

3 Planlagt arbeid

3.1 Lingvistisk arbeid:

- Verbmorfologi (hovudmønsteret er klårt, og stammeklasse er merka av i ordboka, det kan hende det er variasjon frå verb til verb når det gjeld kor mykje omlyd verba har, dette må i tilfelle sjekkast manuelt)
- Adjektivmorfologi (her trengst det grunnforskning)
- Samansetjing (her treng vi undersøkingar av forma til forleddet)
- Derivasjon (data frå referansegrammatikken må systematiserast og komplementerast)

3.2 Datalingvistisk arbeid:

- morfologifonologi for verb (arbeidet er starta, og vi har i prinsippet eit system for å kombinere verb- og omlydsklasse, men til no har det ikkje fungert.)

- morfofonologi for adjektiv (her har vi ikkje gjort noko; dette er eit problematisk punkt, i og med at referansegrammatikken har så lite hjelp å gje.)
- morfofonologi for avleiing/samansetjing (tonivåmorfologien for substantiva burde funka, men det er ikkje sjekka)

3.3 Sørsamisk ordbok

Det vil vera ei uvurderleg hjelp for eit slikt prosjekt å få tilgang til Sydsamisk-norsk ordbok (Bergsland & Mattsson Magga, 1993) i elektronisk form. — Det er no komne signal om at boka vil bli tilgjengeleg, ev. etter nokre rettingar. Budsjettet og planane byggjer på denne føresetnaden.

3.4 Korpusinnsamling

Det må setjast av ein del ressursar på å samla inn så mykje sørsamisk tekst som det berre er mogleg. Det vil vera rimeleg å bruka dei same kjeldene som ved det noverande Divvun-prosjektet bruker, dvs Sametings-administrasjonen, kommunane (Snåsa), departementa, Bibelforlaget/Kyrkja, forlag og forfattarar, og privatpersonar som sjølve vil gje tekstane sine til prosjektet.

Som prøve på faktisk språkbruk bør ein undersøkje om det er mogleg å bruka materiale frå dei nasjonale prøvene. Der finst det eit ganske stort materiale med sørsamiske elevtekstar.

3.5 Stavekontrollspesifikt arbeid

Sjølv om prosjektet byggjer på arbeid som alt er gjort i Divvun-prosjektet, vil det krevast tilpassingar til sørsamisk, og ein god del testing og språkspesifikt arbeid. Dette arbeidet vil m.a. omfatta:

- sjekka forslagsmekanismen og dei forslaga som blir gjeve
- testa dei ulike stavekontrollane

3.6 Rettskrivingsspørsmål for sørsamisk

Dette er eit ressurskrevjande punkt, i og med at sørsamisk til no i liten grad har vorte eksplisitt normert. Eit normerande verkty som ein stavekontroll krev at ein bestemmer seg for kva ein vil ha innanfor den offisielle norma, om det no er stor eller liten variasjon. Derfor er det viktig at dei normerande organa for sørsamisk både nasjonalt og nordisk er (inter)aktivt med i prosjektet.

- Kva er rett, kva er gale?
- korleis skal vi stave lånord frå norsk/svensk m.fl. i sørsamisk?
- datoformat
- forkortingar
- bøyning av framande namn

4 Ferdige modular frå Divvun-prosjektet

I og med at det sørsamiske prosjektet byggjer på det noverande prosjektet, er det sjølv sagt ein del arbeid som ikkje treng repeterast for sørsamisk. Spesifikt gjeld det følgjande punkt:

- infrastruktur for
 - språkteknologiutvikling og testing
 - korpshandtering og -bruk
 - stavekontrollutvikling
- tilpassing av stavekontrollmotor(ar)
- vi har også eit ferdig team av programmerarar og datalingvistar som ikkje treng opplæring, og som vil vera ei stor hjelp for nye prosjektmedarbeidarar

5 Anna arbeid

Ein del teknisk og anna arbeid må gjerast for sørsamisk uavhengig av kva som har vorte gjort før:

- brukardokumentasjon
- nettsider (for brukarane)
- installeringsprogram (lokalisering til sørsamisk)

6 Stillingar

- sørsamisk lingvist — 100%
- datalingvist - 100% fyrste året, 50% dei to andre
- programmerar - 25%
- prosjektleiar - 50%

7 Timeplan

Prosjektet bør gå over tre - 3 - år. Det har vist seg at det krevst mykje tid både til å samla inn korpus, og til det reint lingvistiske arbeidet. Og med tanke på at det trengst meir lingvistisk grunnarbeid enn for både nord- og lulesamisk, vil det derfor ikkje vera realistisk med eit prosjekt kortare enn tre år. Prosjektstart er rekna til januar 2008.

Tidsplanen vil vere avhengig av at vi får tak i ein kvalifisert sørsamisk lingvist. Sjølv med ein kvalifisert lingvist trengst det ein god del opplæring. Med ein prosjektperiode på tre år er det rom for slik opplæring.

Prosjektstart er tenkt rett etter at Divvun-prosjektet er avslutta, dvs 1. januar 2008, og varer til 31. desember 2010 med den føreslåtte timeplanen.

Prosjektet vil samarbeida med den permanente driftsorganisasjonen for dei samiske korrekturverktya.

8 Budsjett

Alle tal er i 1000 NOK

	2008	2009	2010
Løn			
Lingvist	600	600	600
Datalingvist	600	300	300
Programmerar	150	150	150
Prosjektleiar	300	300	300
Eksterne tenester	100	100	100
Reiser, utstyr, m.m.	100	100	100
Totalt	1850	1550	1550
Total prosjektsum			4950

Dei eksterne tenestene er integrering mot MS Office for Windows og Mac, på same vis som i det noverande Divvun-prosjektet.