

Fellesprosjekt for vedlikehald og sørsamiske verkty

Eit framhald av Divvun-prosjektet

Det har tidlegare vorte laga separate planar for sørsamiske verkty, og for den vidare drifta og vedlikehaldet av dei eksisterande verktya. Denne planen er ein revisjon av dei to tidlegare, der alt arbeidet blir sett under eitt. Som eit arbeidsnamn på prosjektet kan ein kalla det Divvun 2.

Table of contents

1 Mål og oppgåver.....	2
2 Planlagt arbeid for sørsamisk.....	3
2.1 Lingvistisk arbeid:.....	3
2.2 Datalingvistisk arbeid:.....	3
2.3 Sørsamisk ordbok.....	3
2.4 Korpusinnsamling.....	4
2.5 Stavekontrollspesifikt arbeid.....	4
2.6 Rettskrivingsspørsmål for sørsamisk.....	4
3 Stillingar.....	4
4 Timeplan.....	5
5 Budsjett.....	5

1 Mål og oppgåver

Å laga eit sett med sørsamiske stavekontrollar, med same oppsett som for nord- og lulesamisk, og halda dei nord- og lulesamiske verktya ved like. Dette betyr for den sørsamiske delen:

- stavekontroll for MS Office for Windows og Mac
- stavekontroll for OpenOffice.org basert på open kjeldekode
- stavekontroll for InDesign/InCopy
- orddeling i MS Office, Win og Mac
- orddeling for InDesign/InCopy

Jf elles planane for det noverande Divvun-prosjektet, på www.divvun.no. Når det gjeld å halda ved like dei eksisterande verktya, krev det:

- språkleg vedlikehald av leksikon (ordlista) og andre lister
- språklege oppdateringar av dei eksisterande verktya
- halda kontakt med leverandørar av eksternt utvikla verkty, forhandla fram nye versjonar når det trengst (t.d. på grunn av nye versjonar av vertsprogramma)
- kontakt med brukarane
- ansvar for tekniske problem i verktya som byggjer på open kjeldekode
- ansvar for å halda ved like og byggja ut det eksisterande korpuset, inkl å samla inn meir tekst, sjå over rutinar for kontinuerleg innsamling (avistekster, tekster frå Sámediggi, osb)

Det bør òg høyra til oppgåvene å vurdera, planleggja og starta prosjekt for nye språk, nye verkty og nye målgrupper.

Eit slikt prosjektområde er arbeid med terminologi og fleirspråklege leksika. Fleire prosjekt arbeider allereie med dette, og vi er avhengige av å integrera dette i systemet vårt for å dra nytte av desse resultatane i praktisk språkarbeid:

- maskinomsetjing, både frå nord- til lulesamisk, frå majoritetsspråka til nordsamisk, og mellom andre språkpar
- terminologisk arbeid, framfor alt innanfor risten.no, men også som forarbeid til risten.no innafor ulike fagområde. Fleire miljø er opptekne av dette (m.a. det samiske juristmiljøet og dei samiske språkkonsulentane), og for å kunne ta i bruk terminologiane sine i integrerte applikasjonar treng dei støtte av eit fungerande språkteknologisk miljø for nord- og lulesamisk

Fungerande to- og fleirspråklege moderne samfunn kviler, og vil i framtida i langt større grad kvile på avanserte språkteknologiske løysingar. Skal alle dei ressursane som blir lagt ned i språkleg revitalisering, språkopplæring og -utdanning for det samiske folket bli brukt i praksis, trengst det språkteknologisk kompetanse og eit breitt spekter språkteknologiske løysingar innafor ulike felt.

Alt dette arbeidet kan og bør integrerast med dei morfologiske automatane våre for å kunne vera tilgjengelege i ulike appliasjonar til nytte for ulike brukargrupper. Det å gjera dette arbeidet vil vera ein del av arbeidet til dei lule- og sørsamiske lingvistane.

Dei lule- og nordsamiske lingvistane vil også ha ei rolle å spela i arbeidet med det sørsamiske retteprogrammet. Det å ha berre ei lingviststilling, særleg der det inneber så mykje pionerarbeid som med eit sørsamisk retteprogram, er ikkje bra, og det å ha andre lingvistar, med erfaring frå liknande arbeid med nord- og lulesamisk, å stø seg til, vil vera ein stor fordel.

Ved å slå i hop dei to tidlegare, separate planane for drift av divvun og for sørsamiske verkty, vil ein spare ein del ved å slå i hop relaterte stillingar til heile, og det vil framleis vera rom for andre oppgåver. Det vil derfor vera mogleg å leggja drifta av risten.no inn som ein del av arbeidsoppgåvene til dette nye Divvun 2-prosjektet, og samtidig koordinera og integrera relevant arbeid som ev. blir gjort ved universitetet i Tromsø (t.d. omkring semiautomatisk bygging av elektroniske ordbøker) med arbeidet med risten.no.

2 Planlagt arbeid for sørsamisk

2.1 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 systemiserast og komplementerast)
- Prinsipp for tilpassing og bøyning av framandord til sørsamisk (dette er eit spørsmål der sørsamiske språknormerarar framleis ikkje har gode svar, her trengst det altså innsats frå normative instansar)

2.2 Datalingvistisk arbeid:

- morfofonologi 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 slik det er tenkt.)
- 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)
- utviding av substantivmorfologien til også å gjelde evt. nye bøyingsmønster for framandord, testing av eksisterande modellar

2.3 Sørsamisk ordbok

Det vil vera ein føresetnad for eit slikt prosjekt å få tilgang til Sydsamisk-norsk ordbok (Bergsland & Mattsson Magga, 1993) i elektronisk form. — Det har no kome signal om

at boka vil bli tilgjengeleg, ev. etter nokre rettingar. Budsjettet og planane byggjer på denne føresetnaden.

2.4 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 det noverande Divvun-prosjektet bruker, dvs Sametings-administrasjonen, kommunane, departementa, Bibelforlaget/Kyrkja, forlag og forfattarar, og privatpersonar som sjølve vil gje tekstane sine til prosjektet.

2.5 Stavekontrollspesifikt arbeid

Sjølv om prosjektet byggjer på arbeid som alt er gjort i Divvun-prosjektet, trengst det 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

2.6 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 i Noreg og i Sverige er aktivt med i prosjektet.

- Kva er rett, kva er gale?
- korleis skal vi stava låneord frå norsk/svensk m.fl. i sørsamisk, og tilsvarande for latinske og greske ("internasjonale") låneord?
- datoformat
- forkortingar
- bøyning av framande namn

3 Stillingar

Ut i frå dei oppgåvene som er nemnde over, får vi desse stillingane:

- sørsamisk lingvist — 100 %
- lulesamisk lingvist — 50 %
- nordsamisk lingvist — 50 %
- programmerar/datalingvist — 100 %
- teknisk sekretær — 100 %
- prosjektleiar — 100 %

Det er til dels høgt spesialisert kunnskap som krevst i alle stillingane, og det har tatt (og tek) tid å byggja opp kompetansen og erfaringa som trengst. For å halda på denne kompetansen, bør stillingane helst vera heile, eller slike som kan kombinerast til heile stillingar.

For alle stillingane gjeld det at dei krev god forståing av formalismane som blir nytta (pr. i dag er dei basert på verkty frå Xerox, sjølv om dette kan endra seg vil dei datalingvistiske prinsippa forbli dei same), og ei grunnleggjande forståing av språkteknologi og korrekturprogram. Ein del av denne kunnskapen vil det vera rimeleg at ein tileignar seg i stillinga, men det bør krevjast avsluttande høgskuleeksamen i relevante fag for kvar av stillingane, eller praksis som dokumenterer tilsvarende kunnskap.

Alle stillingane krev òg nært samarbeid med Universitetet i Tromsø og det språkteknologiske miljøet der.

4 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 det noverande Divvun-prosjektet er avslutta, dvs 1. januar 2008, og varer til 31. desember 2010 med den føreslåtte timeplanen.

5 Budsjett

Alle tal er i 1000 NOK

	2008	2009	2010
Løn			
Sørsamisk lingvist	600	600	600
Lulesamisk lingvist	300	300	300
Nordsamisk lingvist	300	300	300
Datalingvist/programm.	600	600	600
Teknisk sekretær	600	600	600
Prosjektleiær	600	600	600
Eksterne tenester	100	100	100
Reiser, utstyr, m.m.	150	150	150
Totalt	3250	3250	3250
Total prosjektsum			9750

Posten "Eksterne tenester" er integrering mot MS Office for Windows og Mac, utført av eit eksternt IT-firma, på same vis som i det noverande Divvun-prosjektet.