
Samisk språkteknologi
ved
Universitetet i Tromsø

To integrerte prosjekt

Samisk tekstanalyse (Univ. i Tromsø)

- Lene Antonsen
- Saara Huhmarniemi
- (Marit Julien)
- Ilona Kivinen
- Trond Trosterud
- Linda Wiechetek

http://giellatekno.uit.no/

Samisk stavekontroll (Sámediggi)

- Børre Gaup
- Sjur Nørstebø Moshagen
- Thomas Omma
- Maaren Palismaa
- Tomi Pieski

http://www.divvun.no/

Disposisjon

- Grunnlaget vi bygde på
- Målsetjing for dei samiske språkteknologiprosjekta
- Resultat
- Kva kan resultata brukast til?
- Perspektiv framover

Disposisjon

Grunnlaget vi bygde på

Målsetjing for dei samiske språkteknologiprosjekta

Resultat

Kva kan resultata brukast til?

Perspektiv framover

Tidlegare arbeid med samisk

Nordsamisk

- Pekka Sammallahti, Kimmo Koskenniemi: Utkast til nordsamisk morfologisk analysator (substantiv, verb)
- Pekka Sammallahti: Nordsamisk-finsk ordbok (samisk del)

Lulesamisk

- Anders Kintel: Lulesamisk-norsk ordbok (lulesamisk del)

Inspirasjonskjelder

Inspirasjon frå Helsingforsmiljøet:

- Kimmo Koskenniemi: Metodar for grammatisk analyse
- Fred Karlsson: Metodar for syntaktisk analyse
- Lingsoft OY: Oppsett for språkteknologisk infrastruktur

Programvare:

- Xerox: Basisverkty for kompilering av grammatiske analysatorar
- VISL, Odense: Programvare for syntaktisk analyse

Vevgrensesnitt:

- Tekstlaboratoriet, Oslo: Oppsett for vevgrensesnitt

Arbeid i Tromsø før KUNSTI

2000

- Sjur Moshagen, Trond Trosterud: Analysator for sørsamiske substantiv

2001-2003

- Trond Trosterud: Nordsamisk analysator, prototype for lulesamisk analysator (med minimalt leksikon)
- Trond Trosterud: Inventering av enare- og skoltesamisk

Disposisjon

Grunnlaget vi bygde på

Målsetjing for dei samiske språkteknologiprosjekta

Resultat

Kva kan resultata brukast til?

Perspektiv framover

Tekstanalyseprosjektet

Lage program for syntaktisk analyse av nord- og lulesamisk

Analysere eit større tekstkorpus

Gjere resultata tilgjengelege via eit vevgrensesnitt

... meir ambisiøse mål for nord- enn for lule-samisk, pga. eit betre grunnlag for nordsamisk

Sametinget sitt prosjekt

Lage ordrette- og orddelingsprogram for nord- og lulesamisk

Gjere dei tilgjengeleg for relevante plattformar

- (Windows, Mac, Linux)

... og dei viktigaste brukarprogramma

- (kontor-, redigerings-, kommunikasjons- og publiseringsprogram)

Felles mål, overlapping

Begge prosjekta var avhengige av å

- fylle ut og perfeksjonere dei grammatiske analysatorane for nord- og lulesamisk
- arbeid med innsamling av og infrastruktur for tekst (korpus)
- dokumentere arbeidet

Skilnad prosjekta imellom

Ikkje same endeleg mål:

- syntaktisk analysator vs. retteprogram

Ikkje same innfallsvinkel:

- deskriptiv vs. normativ

Bieffekt: Eit godt samarbeid

Felles kjeldefiler i felles versjonskontrollsystem

Fellesmøter

Felles dokumentasjon

Arbeid på felles problemstillingar

- framfor alt knytt til den morfologiske analysatoren og korpuset

«Bytte» av arbeid:

- folk frå det eine prosjektet har til ein viss grad vore involvert i arbeid relevant for det andre

Disposisjon

Grunnlaget vi bygde på

Målsetjing for dei samiske språkteknologiprosjekta

Resultat

Kva kan resultata brukast til?

Perspektiv framover

Fellesresultat

Morfologiske analysatorar

- analyserer og genererer ei kvar ordform

Korpusinnsamling

- einspråkleg korpus
- tospråkleg parallellkorpus (til no: nordsamisk og bokmål)
- grammatisk analyse av desse tekstane

Dokumentasjon

- tilgjengeleg på internett

Retteprogramprosjektet

- Betaversjon til testing på Microsoft Office blir gjort tilgjengeleg no i februar
- Prosjektet blir avslutta i 2007
- Det blir ein liten etterbruksorganisasjon som skal halde programma ved like

Tekstanalyseprosjektet

... har laga

- syntaktiske analyseprogram og
- søkbare korpus

... og alt dette på internett

Bures boahtin sámi giellateknologiija prošektii

PDF

Introdukšuvdna

Dát lea Sámi giellateknologiijaprošeavtta ruovttosiidu.

Romssa universitehta Sámi instituhtas lea Sámi giellateknologiija prošeakta (2001-) jođus. Prošeavtta ulbmil lea ráhkadit morfologalaš analyseren- ja disambiguerenprográmma davvisámegillii. Dás sáhtát lohkat eambbo sihke prošeavtta eará ulbmiliid ja prošeavtta vuođđoteknologiija ja lingvisttalaš filosofiija birra.

Interaktiivalaš prográmmat

Davvisámegiela analyseren	Davvisámegiela genereren
Julevsámegiela analyseren	Julevsámegiela genereren
Lullisámegiela analyseren	Lullisámegiela genereren
	Lohkosániid genereren davvi-, julev- ja lullisamgillii

Interaktiivalaš teavsttat

Dál lea vejolaš ohcat giellaohpalaš konstrukšuvnnaid min interaktiivalaš teavsttas.

Davvisámegiela interaktiivalaš korpus (čilgehusat eŋgelasgillii)
Leksikálalaš ressursat: frekveanssalisttat, a tergo-listtat, jna. (čilgehusat eŋgelasgillii)

Eará gielat

Færsulluid giella (čilgehusat eŋgelasgillii)
Ruonáeatnanlaš giella (čilgehusat eŋgelasgillii)

Davvisámegiella
- Álgosiidu
- Giellapolitihkka
- Plána
- Mo reaiddut ráhkaduvvojit
- Máttasámegiella
- Rievdadusat (engelasgillii)
- Preassadieđáhusat
- Sáddes fiillaid midjiide

built with Apache Forrest

Divvun - sámi korrektuvrareaiddut

Font size: Reset -a +a

PDF

Mii Divvun lea?

Divvun lea prošeakta, man Norgga Sámediggi lea álggahan. Ulbmil lea ráhkadit sámegielat korrektuvraprográmmaid. Dál vuos ráhkaduvvo korrektuvraprográmma ja sátnejuohkin davvisámegiela várás, ja korrektuvraprográmma julevsámegiela várás. Dán prioriterema birra logat eanet dáppe. Reaiddut galget doaibmat dábáleamos kantuvraprográmmain Windows:as, Linux:as ja Mac:as

Duogáš

Prošeakta vuos plánejuvvui, ja de biddjui johtui Sámedikki sávaldaga mielde addit sámegielagiidda ja earáide, geat hálidit sámegiela čállit, seamma reaidduid go dárogielagiin leat. Erenoamážit dát lea dehálaš dannego ollu sámegielagat eai leat oahppan sámegiela čállit. Prošeakta lea danne oassi strategiijas nannet sámegiela, ja buoridit sámegiela eavttuid čállingiellan.

Sámediggi ⇨, Gielda- ja guovlodepartemeanta ⇨, Oahpahus- ja dutkandepartemeanta ⇨ ja Kultur- ja girkodepartemeanta ⇨ leat ruhtadan prošeavtta.

Mii áigut ráhkadit korrektuvraprográmmaid, mat leat heivehuvvon sámegillii nugo lea Norggas, Ruotas ja Suomas. Dannego prošeakta čađahuvvo oalát Norgga prošeaktan, de soaitá válmmas prográmmain dát dihttot, muhto mii geahččalit ráhkadit daid dainna lágiin ahte juohkehaš sáhttá geavahit daid.

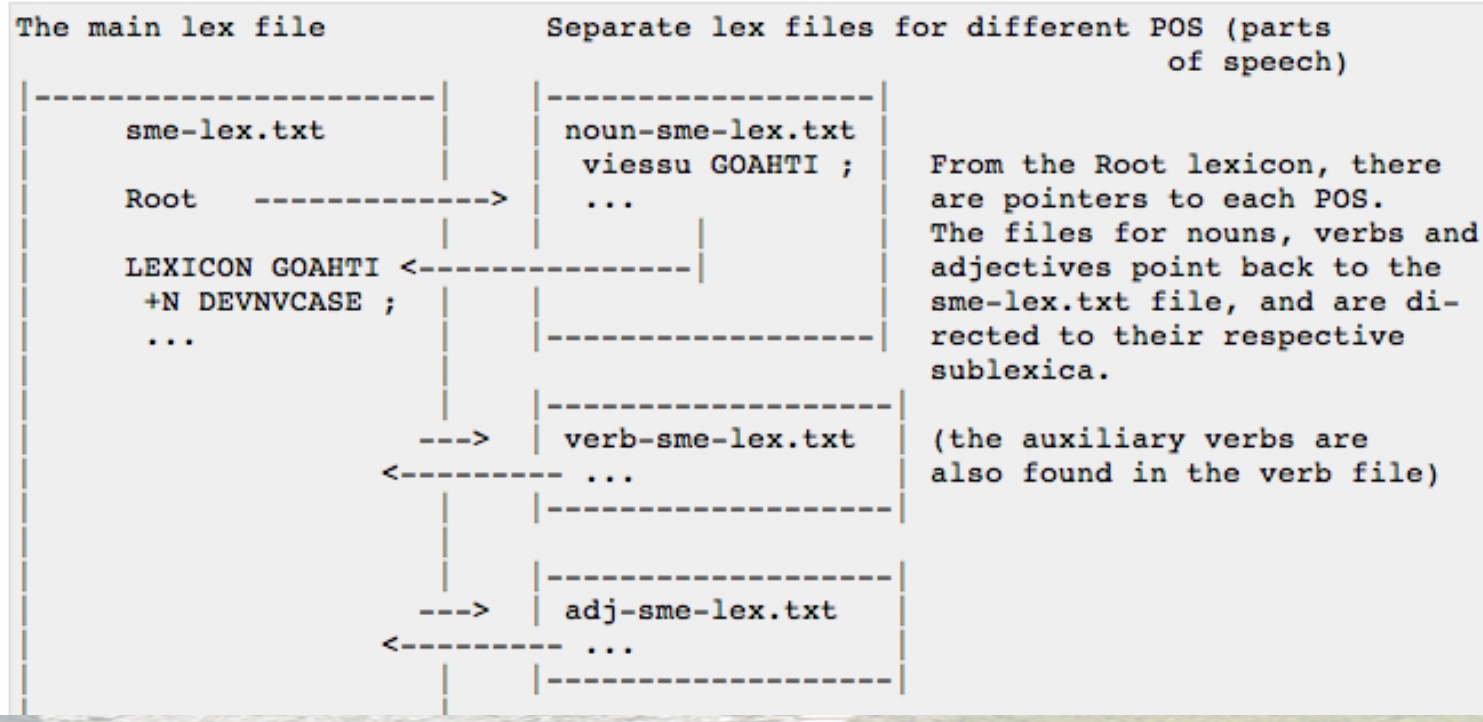
Organiseren

Prošeakta gullá Sámediggái. Das lea iežas prošeaktastivra mas leat miellahtut Sámedikkis, departemeanttain ja Romssa universiteahtas.

A flowchart over the sme files for morphological parsing

A flowchart over the sme files for morphological parsing

This flowchart gives an overview of how the sme sourcefiles are related. In principle, the other lg files are arranged in the same



Northern Sámi analyzer and disambiguator

PDF

- Here you may analyse and disambiguate Northern Sámi words

Here you may analyse and disambiguate Northern Sámi words

If you don't have a Sámi keyboard for Linux, Mac OSX or Windos XP, the letter "á" can be inserted from the keyboard. The special characters "č, đ, ŋ, š, ŧ, ž" may be written as c1, d1, n1, etc.

Write or paste the words in the window, and press the "Analyze" or "Reset" button. Grammatical tags are explained here.
The options "Analyze" gives all possible analyses, whereas the (default) option "Disambiguate" gives only the analyses appropriate for the given sentence. The program may also hyphenate the text for you, or add syllable boundaries, in which case it will not give a grammatical analysis. Choose "Hyphenate".

Type word forms:

```
Kárášjoga olbmuid heagga ja dearvvašvuohta lea
heajut dilis, go ambulánssa radio ii leat gielddas
doaibman badjel njealji vahkkui.
– Dát leat dohkketmeahttumis dilli, dadjá
ambulánsavuoddji Raymon Sakshaug.
```

Character coding: (•) utf-8 ( ) latin1

( ) Analyse
(•) Disambiguate
( ) Hyphenate

Analyze Reset

Sámi instituhtta, Romssa universitehta

Kárášjoga
	Kárášjohka N Prop Plc Sg Acc @OBJ

olbmuid
"olmmoš" N Pl Gen @GN>

heagga
	heagga N Sg Nom @SUBJ

ja
	ja CC @CC-NP

dearvvašvuohta
	dearvvašvuohta N Sg Nom @SUBJ

lea
	leat V IV Ind Prs Sg3 @+FAUXV

heajut
"headju" A Comp Attr @AN>

dilis
	dilli N Sg Loc @ADVL

,
	, CLB

go
	go CS @CS-VP

ambulánssa
"ambulánsa" N Sg Gen @GN>

radio
	radio N Sg Nom @SUBJ

ii
	ii V IV Neg Ind Sg3 @+FAUXV

leat
	leat V IV Ind Prs ConNeg @+FAUXV

gielddas
	gielda N Sg Loc @ADVL

doaibman
	doaibmat V IV PrfPrc @-FMAINV

badjel
	badjel Pr @ADVL

njealji
	njeallje Num Sg Gen @GP<

vahkkui
	vahkku N Sg Ill @ADVL

.
	. CLB <<<

North Sámi generator

PDF

- Here you may generate Northern Sámi word forms

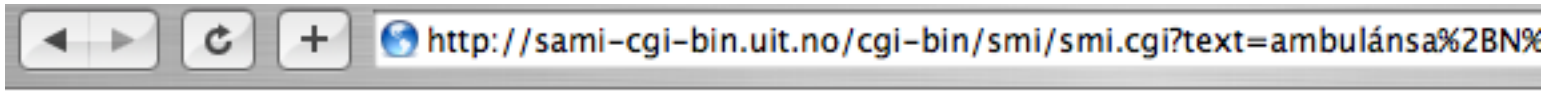
Here you may generate Northern Sámi word forms

Write The Sámi special letters as usual (in Unicode). If you don't have a modern computer, or otherwise don't have access to the Sámi letters, you may write like this: "á" as á, but the other six letters with basic letter + the digit 1, like this: "c1, d1, n1, s1, t1, z1", and choose **Latin 1** below. Write the word and grammatical specification in the window, and press the "Generate" or "Reset" (= "Sihko") button. Grammatical tags are explaine here. In order to get e.g. *viesus*, write **viessu+N+Sg+Loc**.

Type word form: ambulánsa+N+Sg+Com

Character coding: (•) utf-8 ( ) latin1

Generate Reset



Sámi instituhtta, Romssa universitehta

---

ambulánsa+N+Sg+Com => ambulánssain

Korpus

Einspråklege korpus:

- Nordsamisk: 4 202 318 ord
- Lulesamisk: 121 640 ord (ikkje konvertert til xml)
- Sørsamisk: 172 510 ord (ikkje konvertert til xml)

Parallelkorpus

- Bokmål → nordsamisk
  - Sametingsprotokollar, offentlege utgreiingar
- Bibeltekstar
  - Nordsamisk: NT, 1Mos, Salmane
  - Lulesamisk: NT
  - Sørsamisk: ca. 3 bøker frå NT og 5 frå GT

Korpusstruktur

Strukturerte tekstkorpus

- originalane er urørt – *fil*.***doc***, *fil*.***html***, *fil*.***pdf***, ...
- manuelt redigert metafil – *fil.doc*.***xsl***, *fil.html*.***xsl***, *fil.pdf*.***xsl***
  - store «æøå-problem», språkattkjenning, merking av feil
- konvertert fil – *fil.doc*.***xml***, *fil.html*.***xml***, *fil.pdf*.***xml***

dtd – definisjon for tekststruktur

- metadatainformasjon
- hierarkisk tekststruktur
- ulike teksttypar (overskrifter, brødtekst, lister, tabellar, ...)

Korpusa kan når som helst bli generert på nytt

Vevgrensesnitt

Tilgjengeleg på nett via eit grensesnitt frå Tekstlaboratoriet i Oslo (takk framforalt til Lars Nygård)

- *http://giellatekno.uit.no/text.en.html*
- Brukarnamn: **sami**
- Passord: **giella**
- (Vi vil dele korpuset i ein open og ein lukka del, pga. opphavsrett)

Både ein- og fleirspråklege korpus er søkbare via lemmaform og grammatisk analyse

Northern Saami interactive text corpus

PDF

Text search

Here you may search for text and grammatical functions in a Saami text corpus. The corpus will eventually be password protected, but so far it only contains public texts. There is thus a dummy user name, **sami**, and a dummy password, **giella**.

Click here to search the corpus ⇨

The search interface

The first search field is in the box above the text options ». For each word you fill inn, you may or may not specify its grammatical properties (if you don't, but ask for the lemma form, you will get all inflected forms of the word in question. If you have several search words, you may specify the minimum and/or maximum number of words you allow between them. You may also leave the text field empty, and search for, say, any verb in the past tense indicative followed by a locative noun. The search interface makes it possible to search for sentences with specific words, grammatical categories, and (indirectly) grammatical constructions.

Syntaktiske taggar

Basisfunksjonar:

- @SUBJ, @OBJ, @+FMAINV, @ADVL, ..

Komplement:

- @GP> (genitivkomplement til P på høgresida)
- @GP< (genitivkomplement til P på venstresida)
- ...

Taggar for ikkje-lingvistisk definerte tekstelement

- @HNOUN, @TITLE, @APP, ...

http://foni.uio.no/CE2/html/index.php?corpus=sami

cvs

Glossa | Corpus of Modern Sami

leat

options »

lemma form

interval:

min

max

options »

accusative

interval:

min

max

options »

verb

+ -

++ --

Regular expressions: ☑

Search within: s

Hits per page: 20

Max results : 200

Randomize ☐

Context:

◉ sentence ○ word

0 left 0 right

Search corpus

Reset form

corpus-id +

Sami Corpus

Developed by Sami language technology project,
in cooperation with The Text Laboratory.

vokála#guovddáš:vokála#guovddáž LEXDIMINC;
giella#guovddáš:giella#guovddáž LEXDIMINC;
skuvla#guovddáš:skuvla#guovddáž LEXDIMINC;
eaome#guovddáš:eaome#guovddáž LEXDIMINC;

Sámi giellateknologiija

tekstlab.

Results from search in

CWB expression: "([(**lemma="leat" %c**)][(**case="Acc"**)][(**pos="V"**)]) ;"

Action:

Hits found: **76**

Results pages: **1** 2 3 4

104 Sámediggeráđđi **lea áigodagas doallan** 5 čoahkkima , dain 3 telefončoahkkima , ja gieđahallan 27 ášši .

616 Sámediggi **lea Sámediggeplánas ovddidan** dárbbu sierra sámi dutkan ja alit oahpu stuorradiggedieđáhussii .

823 Áigeguovdilis vejolašvuohta maid fertešii čielggadit **lea rájá rahpan** Supmii , nu ahte luossabivdit sáhttet Supmii vuovdit luosaid , ja ahte Suoma beale luossaoastiide addojuvvo vejolašvuohta oastit luosaid Norgga bealde .

957 Sámediggi oaidná ahte kulturrádđi **lea ruđaid prioriteren** biergasiidda mat mánáidgárddiin galget geavahuvvot sámi giela ja kultuvrra gaskkusteapmái .

1017 Giellaráđi bargun **lea maid suokkardit** usiid .

1223 Sámediggi oaidná ahte kulturrádđi **lea ru** n biergasiidda mat mánáidgárddiin galget geavahuvvot sámi giela ja kultuvrra gaskkusteapmái .

lemma: ruhta
pos: N
syn: @OBJ
number: Pl
case: Acc

1283 Giellaráđi bargun **lea maid suokkardit** oahppogirjemanusiid .

1575 Sámediggi **lea sámi álbmotválljen** orgána mii bargá sámi álbmoga ovddas .

2484 Dás lea maid sáhka olmmošvuoigatvuođas maid eamiálbmot lea massán , stáhta morálalaš ja politihkalaš geatnegasvuođas fasttášit dan mo **lea sápmelaččaid vealahan** , ja geavahit vejolašvuođa soabadeami , árvvusatnima ja ovttaárvosašvuođa vuođul ásahit ođđa vuođu sápmelaččaid ja Norgga stáhta gaskka.Dán vuođul dáhttu Sámediggi Stuorradikki dahkat vejolaššan njulget dan vearrivuođa mii lea dahkkojuvvon .

CWB expression: "([(**lemma="leat" %c**)][(**case="Acc"**)][(**pos="V"**)]) ;"

Action:

Hits found: **76**

Results pages: **1** 2 3 4

104 Sámediggeráđđi **lea áigodagas doallan** 5 čoahkkima , dain 3 telefončoahkkima , ja gieđahallan 27 ášši .

616 Sámediggi **lea Sámediggeplánas ovddidan** dárbbu sierra sámi dutkan ja alit oahpu stuorradiggedieđáhussii .

823 Áigeguovdilis vejolašvu **lea rájá rahpan** Supmii , nu ahte luossabivdit sáhttet Supmii vuovdit luosaid , ja ahte Suoma b rvo vejolašvuohta oastit luosaid Norgga bealde .

957 Sámediggi oaidná ahte k **oriteren** biergasiidda mat mánáidgárddiin galget geavahuvvot sámi giela ja kultuvrra gaskkusteapmá

1017 Giellaráđi bargun **lea maid suokkardit** oahppogirjemanusiid .

1223 Sámediggi oaidná ahte kulturráđđi **lea ruđaid prioriteren** biergasiidda mat mánáidgárddiin galget geavahuvvot sámi giela ja kultuvrra gaskkusteapmái .

1283 Giellaráđi bargun **lea maid suokkardit** oahppogirjemanusiid .

1575 Sámediggi **lea sámi álbmotválljen** orgána mii bargá sámi álbmoga ovddas .

2484 Dás lea maid sáhka olmmošvuoigatvuođas maid eamiálbmot lea massán , stáhta morálalaš ja politihkalaš geatnegasvuođas fasttášit dan mo **lea sápmelaččaid vealahan** , ja geavahit vejolašvuođa soabadeami , árvvusatnima ja ovttaárvosašvuođa vuođul ásahit ođđa vuođu sápmelaččaid ja Norgga stáhta gaskka.Dán vuođul dáhttu Sámediggi Stuorradikki dahkat vejolaššan njulget dan vearrivuođa mii lea dahkkojuvvon .

lemma: Sámediggeplána
pos: N
syn: @OBJ
number: Sg
case: Acc

Litt meir om analysatorane

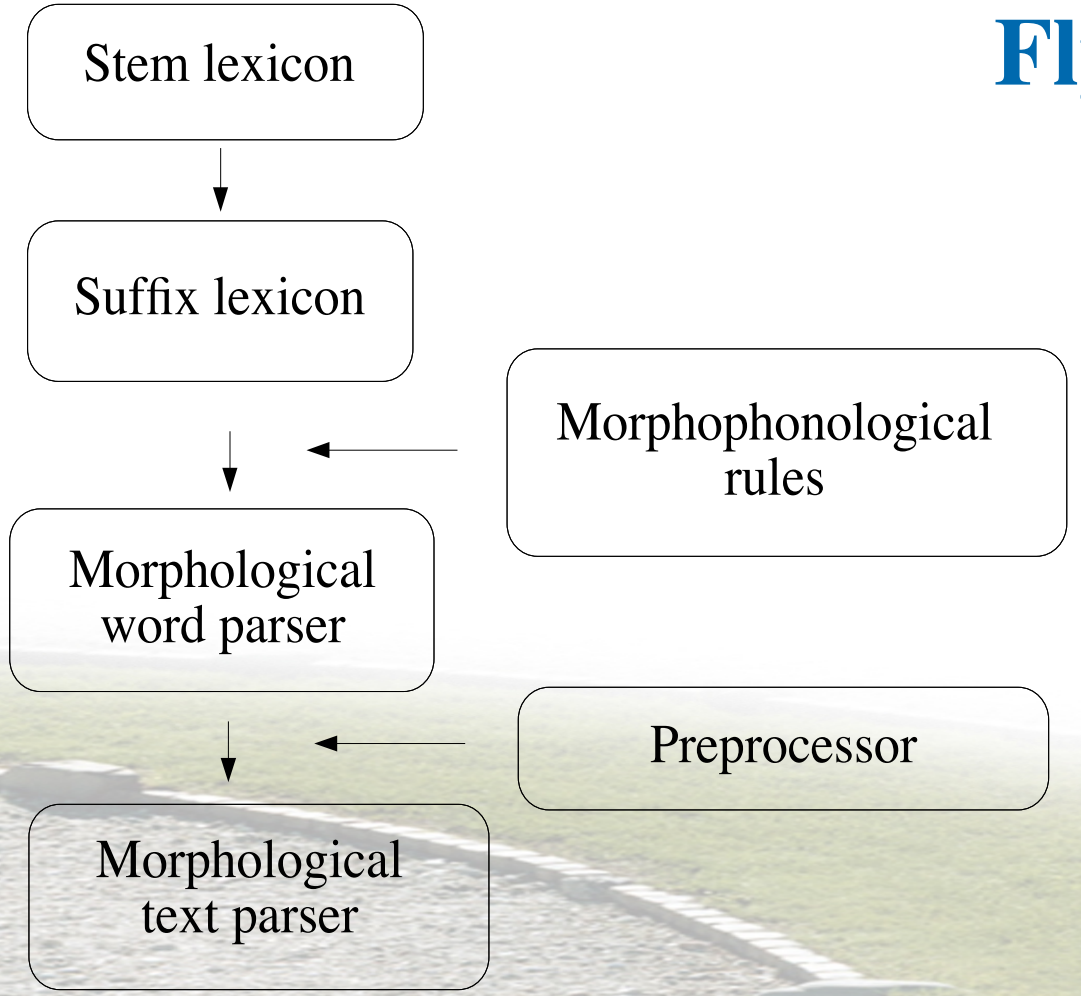
Oversyn

Grammatisk analysator

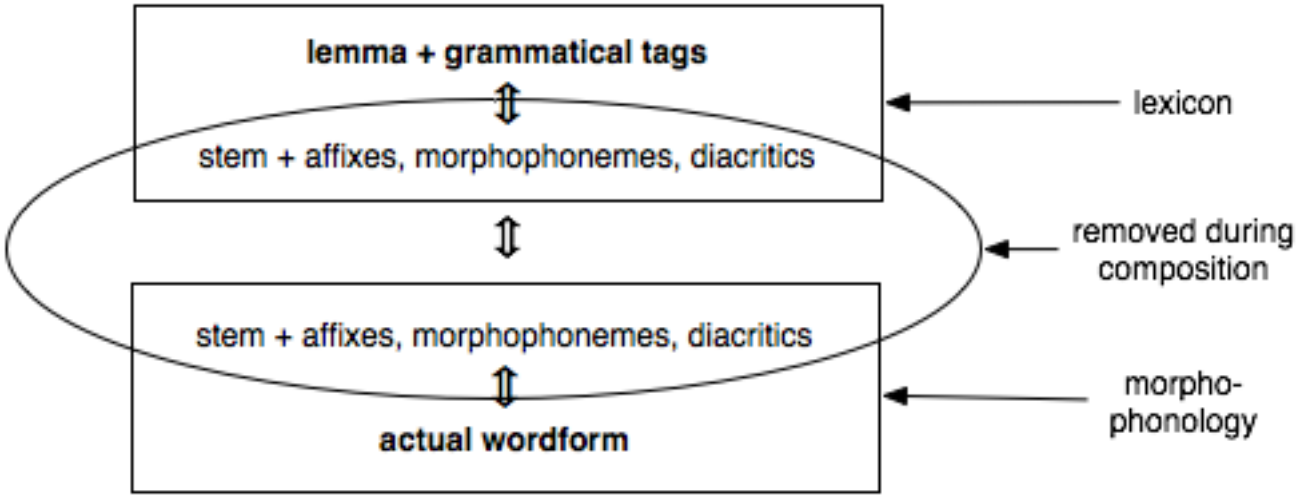
- Fonologi
- Morfologi og leksikon

Disambiguering og syntaktisk analyse

Flytskjema


Stem lexicon
Suffix lexicon
Morphophonological rules
Morphological word parser
Preprocessor
Morphological text parser

Serielle transdusarar



Disambiguering

```
"<Mii>"
        "mun" Pron Pers Pl1 Nom
        "mii" Pron Interr Sg Nom
"<eat>"
        "ii" V Neg Ind Pl1
"<leat>"
        "leat" V Ind Prs Pl1
        "leat" V Ind Prs Pl3
        "leat" V Ind Prs Sg2
        "leat" V Inf
        "leat" V Ind Prs ConNeg
"<dan>"
        "dat" Pron Dem Sg Acc
        "dat" Pron Dem Sg Gen
        "dat" Pron Pers Sg3 Acc
        "dat" Pron Pers Sg3 Gen
"<muitalan>"
        "muitalit" V PrfPrc
        "muitalit" V Actio Gen
        "muitalit" V Actio Nom
        "muitalit" V Ind Prs Sg1
"<.>"
        "." CLB
```

=>

```
"<Mii>"
        "mun" Pron Pers Pl1 Nom
"<eat>"
        "ii" V Neg Ind Pl1
"<leat>"
        "leat" V Ind Prs ConNeg
"<dan>"
        "dat" Pron Pers Sg3 Acc
"<muitalan>"
        "muitalit" V PrfPrc
"<.>"
        "." CLB <<<
```

Reglane attom disambigueringa

```
"<Mii>"
        "mun" Pron Pers Pl1 Nom
        "mii" Pron Interr Sg Nom
"<eat>"
        "ii" V Neg Ind Pl1
"<leat>"
        "leat" V Ind Prs Pl1
        "leat" V Ind Prs Pl3
        "leat" V Ind Prs Sg2
        "leat" V Inf
        "leat" V Ind Prs ConNeg
"<dan>"
        "dat" Pron Dem Sg Acc
        "dat" Pron Dem Sg Gen
        "dat" Pron Pers Sg3 Acc
        "dat" Pron Pers Sg3 Gen
"<muitalan>"
        "muitalit" V PrfPrc
        "muitalit" V Actio Gen
        "muitalit" V Actio Nom
        "muitalit" V Ind Prs Sg1
"<.>"
        "." CLB
```

Fleirtalspronomen fordi verbet står i fleirtal

leat står etter nektignsverbet, det er altså ikkje Inf eller personbøygd form

Ikkje demonstrativ, sidan det ikkje står foran adjektiv eller substantiv
- Ikkje genitiv, sidan det ikkje står etter talord eller preposisjon, og ikkje foran substantiv, adjektiv eller postposisjon

Partisipp, sidan det står som komplement til nektingsverb

Disposisjon

Grunnlaget vi bygde på

Målsetjing for dei samiske språkteknologiprosjekta

Resultat

Kva kan resultata brukast til?

Perspektiv framover

Ordretteprogramma

Dei samiske skriftspråka er unge (1979, 1983)

- Ingen fødd før 1980 har lært å lese og skrive på skolen
- Mesteparten av den teksten samar les, er på norsk, svensk eller finsk

Manglande kjennskap til ortografien er ei hindring for å skrive samisk

- Svært få samar fødd før 1980 er skriveføre
- Samisk tekst (anna enn nyskrivne avisartiklar og skjønnlitteratur) er som oftast omsetjingar frå norsk, svensk eller finsk

Retteprogram for samisk vil gjere det mogleg for langt fleire å skrive samisk

Analyseprogramma og korpusa

Blant språkforskarar er det stor interesse for samisk, men data er vanskeleg tilgjengeleg

- analyserte korpora vil utvide det empiriske grunnlaget til forskarane

For første gong er parallellkorpus tilgjengeleg

- det blir mogleg å drive med kontrastive studier av samisk
- det vil gjere samisk terminologiarbeid langt meir effektivt

Med ordfrekvensdata får vi vite korleis det samiske ordforrådet oppfører seg

Bruksområde som treng programma våre

Leksikografi

Informasjonssøk

Maskinomsetjing

Tekst-til-tale

...

Overføring til andre prosjekt

Språkteknologiprosjekt fell som regel innafor ein av desse tre kategoriane:

1. *Kommersielle prosjekt* — lukka for innsyn
2. *Akademiske prosjekt* — opne, men ikkje innretta på praktisk bruk, eller ikkje skalert opp til full dekning
3. *Open kjeldekode-prosjekt* — opne, praktiske siktemål, men med gammal eller for enkel teknologi til å vere brukbare for andre språk enn dei med svært enkle bøyingsmønster (engelsk, indonesisk, maori, ...)

Overføring til andre prosjekt

Språkteknologiprosjekt fell som regel innafor ein av desse tre kategoriane:

1. *Kommersielle prosjekt* — lukka for innsyn
2. *Akademiske prosjekt* — opne, men ikkje innretta på praktisk bruk, eller ikkje skalert opp til full dekning
3. *Open kjeldekode-prosjekt* — opne, praktiske siktemål, men med gammal eller for enkel teknologi til å vere brukbare for andre språk enn dei med svært enkle bøyingsmønster (engelsk, indonesisk, maori, ...)

Prosjektet vårt er *både* ope *og* laga for praktisk bruk, og derfor relevant for andre språksamfunn

Døme på overføring så langt

Grønlandsk

- ordretteprogram
- neste steg: syntaktisk analyse

Komi

- integrert ordbok (komi – finsk/engelsk) og analysator for komi
- moglege neste steg:
  - utvikle analysatoren betre
  - koble den finsk/engelske sida til finske/engelske analysatorar

Disposisjon

Grunnlaget vi bygde på

Målsetjing for dei samiske språkteknologiprosjekta

Resultat

Kva kan resultata brukast til?

Perspektiv framover

Konkrete framtidsplanar

Lage pedagogisk programvare
Vidareutvikle dei analyseprogramma vi har
Utvide til analysatorar for andre samiske språk
Integrere analysatorane vår med ordbøker
Grønlandsk: syntaktisk analyse

Langsiktige planar

Informasjonssøk
Tekst-til-tale
Maskinomsetjing
...

Oppsummering

Vi arbeider i eit hundreårsperspektiv

Dette og det førre NFR-finansierte prosjektet har lagt grunnlaget for arbeid med samisk språkteknologi det neste hundreåret