

Giellatekno and ICALL

Lemmas are rare in running text

	North Saami	Finnish	Norwegian
Number of words in the text	252 461	45 144	64 994
Number of lemmas in the dictionary	99 071	94 111	38 983
Coverage	7,9 %	10,0 %	30,5 %

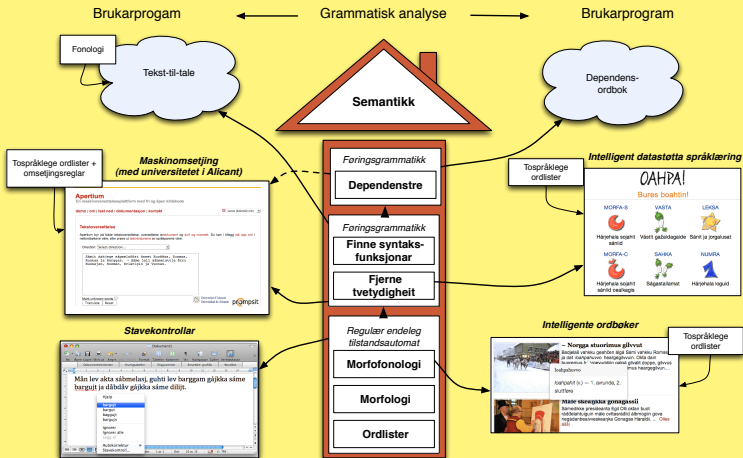
(Antonsen et al 2009)

The morphology is rich, and not always concatenative (North and South Saami examples)

- ▶ Inflection:
 - ▶ Case: *niidii* → *nieida* girl, *gáatan* → *gáetie* house
 - ▶ Person-tense: *eahtsa* → *iehtsedh* love, *bođii* → *boahhtit* come
- ▶ Derivation:
 - ▶ Passivisation: *juhkkojuvvui* → *juhkat* drink, *juohkit* share
 - ▶ Superlative: *nööremes* → *noere* young
- ▶ Compounding:
 - ▶ *bargojoavku* → *bargu* + *joavku* work + group
 - ▶ *bigkemegierkine* → *bigkedh* + *gierkie* building + stone
 - ▶ In a 1.1 mill word North Saami corpus, 26.7 % of the nouns are compounds, or 7.0 % of all words
- ▶ Cliticisation:
 - ▶ *bodiiige* → *boahhtit* + *ge* came + too



Samisk språkteknologi ved Universitetet i Tromsø – Giellatekno og Divvun-gruppa:
Lingvistikk i praksis

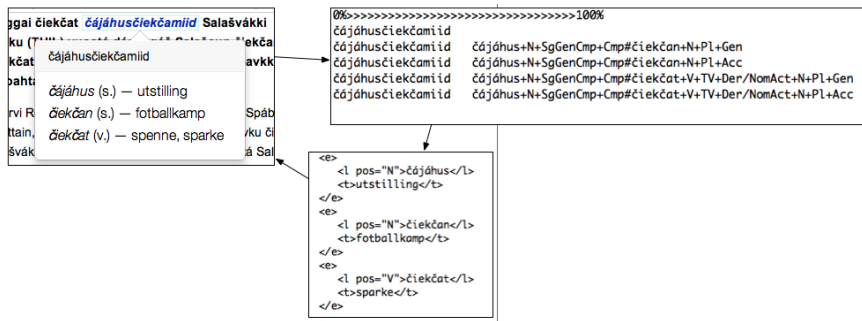


ICALL tools – main site: <http://oahpa.no/>

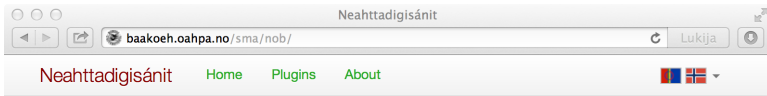
- ▶ VISL
http://oahpa.no/sme/usevisl/VISL_spealut.eng.html
- ▶ Oahpa <http://oahpa.no/davvi/>
- ▶ Teaksta (Werti) <http://gtlab.uit.no:8080/teaksta/>
- ▶ Korpusdict <http://gtweb.uit.no/korp/>
- ▶ Neahttadigisánit (FST-dict) <http://sanit.oahpa.no/>
- ▶ Our ICALL-documentation
<http://giellatekno.uit.no/ped/index.html>

Dynamic FST dictionaries

FST-analysis and translation via a web service



The dictionary understands inflected forms



South Sámi → Norwegian

Norwegian → South Sámi

Other dictionaries

South Sámi → Norwegian (⇌ Swap)

eahtsa

Search

iehtsedh (v., i)

1. elske, være glad i

Manne datnem eahtsam.

Jeg elsker deg.

2. engste seg for noen

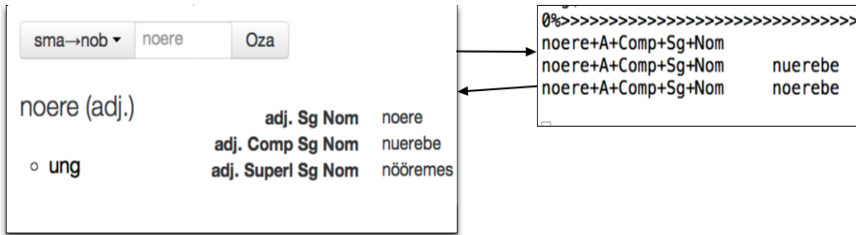
Datneste eahtsam.

Jeg er engstelig for deg.

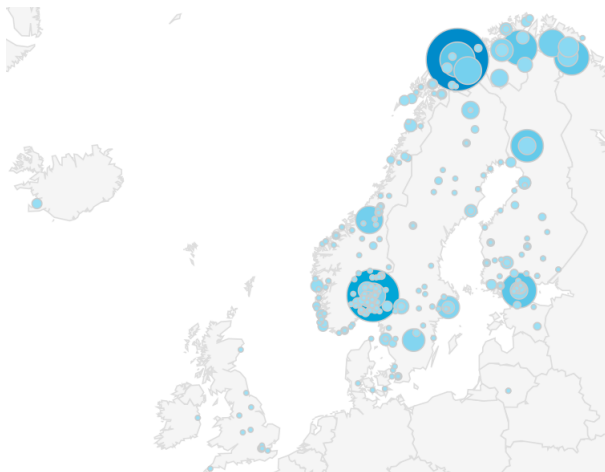
eahtsa is a possible form of ...

iehtsedh
v. TV Ind Prs Sg3

... and may generate word paradigms via FST



Logging User Activity – Google Analytics: North Saami Oahpa



Overview

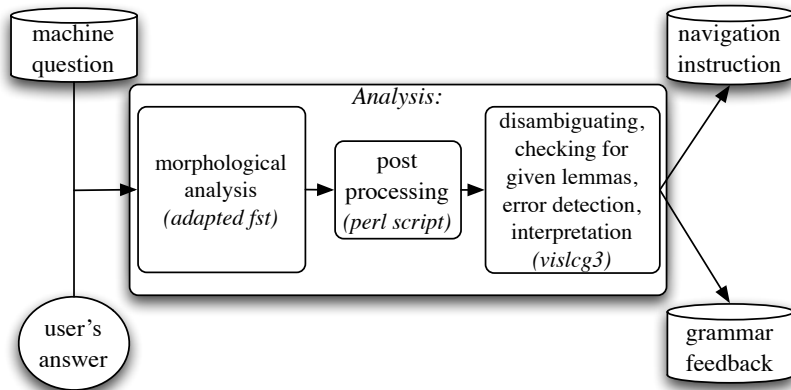
- ▶ Leksa – translating words
- ▶ Numra – numbers
- ▶ MorfaS, MorfaC – word inflection
- ▶ Vasta, Sahka – answer questions
- ▶ Teaksta (Werti) – make grammar tasks from internet texts
- ▶ Korpusdict – how words are used in authentic texts
- ▶ Neahttadigisánit – dictionary with morphology
- ▶ Divvun – spell checker program

Resources

South Saami and North Saami or only North Saami

	Lexicon	Trans- ducer	Morph. analyser	Synt. analyser	Corpus
Numra:		x			
Leksa:	biling.				
MorfaS, MorfaC:	x		x		
Vasta, Sahka:	x		x	x	
Teaksta:			x	x	
Korpusdict:			x	x	x
Neahtta- digisánit:	biling.		x		
Divvun:	full		full		x

Overview Vasta and Sahka



The grammatical errors we have rules for

- ▶ verbs: finite, infinite, negative form, correct person/tense according to the question
- ▶ case of argument based upon the interrogative
- ▶ case of argument based upon valency
- ▶ locative vs. illative based upon movement
- ▶ subject/verbal agreement
- ▶ agreement inside NP
- ▶ numeral expressions: case and number
- ▶ PP: case of noun and pp based upon the interrogative
- ▶ time expressions
- ▶ special adverbs
- ▶ particles according to word order
- ▶ comparison of adjectives

Misspelling: non-word errors

- ▶ “X is not in our lexicon. Could it be a typo?”

Error tags for systematical misspellings

"<londonis>" "London" N Prop **LowercaseErr** Plc Sg Loc
– target: "Londonis"

"<barru>" "bárru" N Sg Nom **AErr** – target: "bárru" 'wave'

"<viessui>" "viessu" N Sg Ill **DiphErr** – target: "vissui" 'house'

"<áhkku>" "áhkku" N Sg Nom

"<áhkku>" "áhkku" **CGErr** N Sg Acc – target: "áhku"
'grandmother'

Erroneous forms in the student's input.

```
"<Ahkku>"  
"Ahkku" ?  
"<manná>"  
"mannat" V IV Ind Prs Sg3|  
"<lundii>"  
"lundii" ?  
"<odne>"  
"odne" Adv
```

Figure 5: ‘Grandmother goes to Lund today.’ analysed with the regular FST.

Erroneous forms in the student's input.

```
"<Ahkku>"  
  "áhkku"  CGErr Sg Acc AErr  
  "áhkku"  CGErr Sg Gen AErr  
  "áhkku"  N Sg Nom AErr      <- correct  
"<manná>"  
  "mannat" V IV Ind Prs Sg3   <- correct  
  "mánná"  Hum N Sg Nom AErr  
  "mánná"  Hum N CGErr Sg Acc AErr  
  "mánná"  Hum N CGErr Sg Gen AErr  
"<lundii>"  
  "Lund"   N Prop LowercaseErr Plc Sg Ill  
"<odne>"  
  "odne"   Adv
```

Analysed with an FST enriched with erroneous forms.

Constraint grammar guesser for proper nouns

"<Gosa>"

"gosa" Adv

"<du>"

"don" Pron Pers Sg2 Gen

"<áhkku>"

"áhkku" N Sg Nom

"<vuolgá>"

"vuolgit" V IV Ind Prs Sg3

"<geasselupmui>"

"geasseluopmu" N Sg III

"<qdl>"

"qdl" QDL

"<Ahkku>"

"áhkku" CGErr Sg Acc AErr

"áhkku" CGErr Sg Gen AErr

→ "áhkku" N Sg Nom AErr ←

"<vuolgá>"

"vuolgit" V IV Ind Prs Sg3

"<Dudweilerii>"

"Dudweiler" N Prop Sg III Guess

'Where does your grandmother go for holiday? Grandmother goes to Dudweiler.'