

# Learner corpus from a free-form student response system

Lene Antonsen  
Centre for Sami Language Technology  
<http://giellatekno.uit.no/>

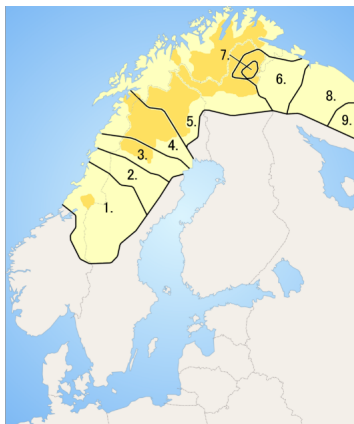


## Learner corpora

We have a free-form student response system

- ▶ It gives metalinguistic feedback instead of the correct form
- ▶ We can add different kinds of exercises and follow the students response to the feedback
- ▶ We get a learner corpus with students with Norwegian and Finnish as their L1
  - ▶ The corpora are parallel, we can look at the same features

## The Saami language area



- 1. South Sami
- 2. Ume Sami
- 3. Pite Sami
- 4. Lule Sami
- 5. North Sami
- 6. Skolt Sami
- 7. Inari Sami
- 8. Kildin Sami
- 9. Ter Sami

Darkened area  
represents  
municipalities that  
recognize Sami as  
an official language.

Figure: The Saami language area – all together approx. 25,000 speakers

## Second Language Acquisition

L1 influence on the L2 acquisition

- ▶ Norwegian vs. Finnish as the students' L1
- ▶ Germanic vs. Finno-Ugric language

## Norwegian and English verbforms

Verb: 'kjøre' – 'drive':

	Norwegian: 6 wordforms	English: 5 wordforms
<i>Inf.</i>	kjøre	drive
<i>Prs.</i>	<b>kjører</b>	<b>drives, drive</b>
<i>Prt.</i>	<b>kjørte</b>	<b>drove</b>
<i>Imprt.</i>	<b>kjør</b>	<b>drive</b>
<i>Prf.Prc.</i>	kjørt	driven
<i>Progr.</i>		driving
<i>Prs.Prc.</i>	kjørende	driving

Finite forms in **red**

## Finnish verbforms

**Verb: 'ajaa' ('drive') :**

ajaa, ajan, ajat, ajaa, ajamme, ajatte, ajavat, ajoin, ajoit, ajoi, ajoimme, ajoitte, ajoivat, ajaisin, ajaisit, ajaisi, ajaisimme, ajaisitte, ajaisivat, ajanen, ajanet, ajanee, ajanemme, ajanette, ajanevat, ajakaami, aja, ajakoon, ajakaamme, ajakaa, ajakoot, ajaen, ajaessa, ajamaan, ajamassa, ajamasta, ajamalla, ajamatta, ajanut, ajaneen, ajanutta, ajaneeseen, ajaneessa, ajaneesta, ajaneelle, ajaneella, ajaneelta, ajaneeksi, ajaneena, ajaneet, ajaneiden, ajaneitten, ajaneita, ajaneihin, ajaneisiin, ajaneissa, ajaneista, ajaneille, ajaneilla, ajaneilta, ajaneiksi, ajaneina, ajanein, ajava, ajavan, ajavaa, ajavaan, ajavassa, ajavasta, ajavalle, ajavalla, ajavalta, ajavaksi, ajavana, ajavien, ajavain, ajavia, ajaviin, ajavissa, ajavista, ajaville, ajavilla, ajavilta, ajaviksi, ajavina, ajavin...

Finite forms in red

## North Saami verbforms

**North Saami verb: vuodjit ('drive') – 60 different word forms:**

vudje, vudjen, vudjet, vudjon, vudjos, vudjoset, vudjoska, vudjot,  
 vuoján, vuoját, vuojáše, vuojášedje, vuojášeidde, vuojášeiddet,  
 vuojášeigga, vuojášeimme, vuojášeimmet, vuojášii, vuojášin,  
 vuojášit, vuješ, vuježa, vuježan, vuježat, vuježeaba, vuježeahppi,  
 vuježehpet, vuježetne, vuježit, vujii, vujiide, vujiidet, vujiiga,  
 vujiime, vujiimet, vuoddji, vuoddjit, vuoddju, vuoddjiba,  
 vuoddjibeahhti, vuoddjibehtet, vuodjá, vuodjime, vuodjimis, vuodjit,  
 vuodján, vuoje, vuoji, vuoddji, vuojedettiin, vuojedettiinan,  
 vuojedettiinat, vuojedettiinename, vuojedettiineamet,  
 vuojedettiinaset, vuojedettiinaskka, vuojedettiinatte,  
 vuojedettiinattet, vuojedettiinis

Finite forms in **red**: 4 modi: Indicative, Imperative, Conditional, Potential. 9 person/number

## Derivations

Every derivation has the same amount of word forms

- ▶ vuddjojuvvot ('be driven')
- ▶ vuojáhallat ('be driven over')
- ▶ vuodjilit ('drive off')
- ▶ vuoddját ('start driving')
- ▶ vuojášit ('drive around')
- ▶ vuojihit ('get someone to drive / transport')
- ▶ ...



## Derivations

- ▶ vuddjojuvvot ('be driven')
- ▶ vuojáhallat ('be driven over')
- ▶ vuodjilit ('drive off')
- ▶ vuoddját ('start driving')
- ▶ vuojášit ('drive around')
- ▶ vuojihit ('get someone to drive / transport')
- ▶ ...

## Second Language Acquisition

Norwegian vs. Finnish as L1 – what difference does it make?

- ▶ The system makes it possible to compare students with Norwegian and Finnish as their L1:
  - ▶ interlanguage (syntax and morphology)
  - ▶ response to metalinguistic feedback

## Inflectional morphology

Slabakova 2009: What Is Easy and What Is Hard to Acquire in a Second Language?. Proceedings of GASLA 2009, 280–294.

1. Inflectional morphology reflects syntactic and semantic differences between languages;
2. Narrow syntactic operations and meaning calculation are universal;
3. In order to acquire syntax and meaning in a second language, the learner has to go through the inflectional morphology;
4. Hence, morphology is the bottleneck of acquisition

## Inflectional morphology

- ▶ Focus on Form
- ▶ Focus on FormS

M. H. Long 1991: Focus on Form: A Design Feature in Language Teaching. Foreign Language Research in Cross-cultural Perspective.

Norris and Ortega 2000: Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. Language Learning 50 (3).

# Frequency

Schmidt 1990: The role of consciousness in second language learning. *Applied Linguistics* 11, 129-158.

## Frequency

- ▶ number of word forms vs. input
  - ▶ ex. Greenlandic: **angerlarniaraluarpunga** 'I was planning to go home though'
  - ▶ ex. Saami: **vuojašeažžaba** 'eventually the two persons drive around'
- ▶ the learner has so learn how to analyse a wordform, and how to generate a wordform

## ICALL programs – <http://oahpa.no/davvi/>

HELP

# OAHPA!

Bures boahhtin!

Veahkkegiella  
English

Suopman  
Guovdageaidnu

<b>MORFA-S</b>  Hárjehala sojahit sániid	<b>VASTA</b>  Vástit gažaldagaide Answer to questions	<b>LEKSA</b>  Sánit ja jorgalusat
<b>MORFA-C</b>  Hárjehala sojahit sániid cealkagis	<b>SAHKA</b>  Ságastallamat	<b>NUMRA</b>  Hárjehala loguid

OAHPA lea interneahttaprogámma nuoraide ja rávesolbmuide geat leat oahpahallame davvisámegiela. Progámma sáhtát heivehit fáttáid ja dási mielde, ja ođđa bargobihtát ráhkaduvvojit automáhtalaččat.

## Vasta-F – a QA-drill with free input

Maid rádddealmmái áigu oastit?

Son áigu oastit stuora viessu ✘

Iskka vástádusaid

Your answer should contain an **accusative** .

Du čuoggát: 0/1

Case for the object, e.g. girjji, girjjiid. Interrogates are: maid, man, gean, geaid. Click to read more.

Copyright 2012 Romssa universitehta  
Contact: romssa@kuvvut.fi

'What is the mayor going to buy? He is going to buy a big house. Nom.'

## Vasta-S – QA-drill with given lemmas

Mo studeanttat vulget Guovdageidnui? Sii **vuodjit njeallje biila** .

Sii vulget njeallje biillain



Iskka vástádusaid

Remember **agreement between numeral and noun**.

Du čuoggát

If the numeral is singular, the noun should be singular. The case of the numeral should agree with the case of the noun. Click to read more.

Copyright 2012  
Contact oahpa@

'How are the students going to Kautokeino? They are going with four.SgNom cars.SgCom.'



## Sahka – QA-drill, a tailored dialogue 1 - NEW

Answer to the questions with full sentences. Remember big initial letter in placenames.

---

Buorre beaivi! Bures boahtin mu geahčái!

Mun lean aiddo fáren sisa iežan odđa orrunsadjái. Mus leat lossa viessogálvvut dáppe feaskáris. Gillešit go veahkehit mu?

De gillen.

Mus lea TV dás. Guđe lanjas TV lea du orrunsajis?

Dat lea stobus.

Guđe latnjii moai bidje mu TV?

Moai bidje TV hivssegis.

✘ The answer should contain an illative.

Answer

'In which room should we place the TV? We should place it in the bathroom. Loc.'

## The system

The basic grammatical analysis of the student's input is done with pre-existing language technology resources developed at the University of Tromsø

- ▶ a finite state morphological analyser/generator (FST)
- ▶ a constraint grammar (CG) parser – adjusted

Antonsen, L., Huhmarniemi, S., and Trosterud, T. (2009). Interactive pedagogical programs based on constraint grammar. In Proceedings of the 17th Nordic Conference of Computational Linguistics, Nealt Proceedings Series 4.

## 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 valence
- ▶ 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

## Student-computer interaction 1

Geasa Máret lea liikostan?

'Who did Máret have a crush on?'

<p>Son lea liikostan duot bealjehis bártni 'She has a crush on that.Nom deaf boy.Acc'</p>	0	This verb wants an illative.
<p>Son lei liikostan duot bealjehis bárdnái 'She has a crush on that.Nom deaf boy.'</p>	0	Missing agreement between demonstr. pronoun and noun.
<p>Son lei liikostan duon bealjehis bárdnái 'She has a crush on that deaf boy.'</p>	1	

## Student-computer interaction 2

Lea go dus heasta?

'Do you have a horse?'

<p>mus in leat heasta 'I don't.Sg1 have a horse'</p>	0	Are you confident that you answer in correct person?
<p>mus in lean heasta 'I don't.Sg1 have.Sg1 a horse'</p>	0	Are you confident that you answer in correct time?
<p>mus in lea heasta 'I don't.Sg1 have.Sg3 a horse'</p>	0	The verb should have negation form.
<p>mus ii leat heasta</p>	1	

## Evaluation

Precision: 0.85 (correctly identified errors/all diagnosed errors)

Recall: 0.93 (correctly identified errors/all errors)

53% of the erroneous sentences contained misspellings.

Antonsen, L., Huhmarniemi, S., and Trosterud, T. (2009). Constraint grammar in dialogue systems. In Proceedings of the 17th Nordic Conference of Computational Linguistics, volume 8 of NEALT Proceeding Series

## Misspellings: Levels of errors

- ▶ Substance errors (errors in encoding/decoding)
  - ▶ a vs. á, special letters: š č ž đ ŋ t
- ▶ Text errors (usage)
  - ▶ suprasegmental processes like vowel harmony and consonant gradation
  - ▶ incorrect suffixing

James C. (1998). Errors in language learning and use: exploring error analysis. Longman. 129pp

## FST: Error tags for systematic misspellings

error tag	erroneous form	target form	
Lowercase	"<london>"	London	
AErr	"<man <b>na</b> >"	mánná	'child.SgNom'
AiErr	"<boah <b>ta</b> n>"	boah <span style="color: green;">tán</span>	'come.V.PrfPrc'
CGErr	"<skuv <b>l</b> as>"	skuv <span style="color: green;">llas</span>	'school.SgLoc'
DiphErr	"<vi <b>ess</b> ui>"	vissui	'house.SgIII'
III <sub>V</sub> Err	"<skuv <b>l</b> ai>"	skuv <span style="color: green;">lii</span>	'school.SgIII'
III <sub>E</sub> Err	"<hiv <b>ss</b> egi>"	hiv <span style="color: green;">ssegii</span>	'toilet.SgIII'

and also the combination of these:

"<fallej**oh**kas>" "Fállejohka" N Prop LowercaseErr CGErr Sg Loc  
AErr

Fállejogas placename.Loc

edit distance: 4



## Student-computer interaction 3

Makkár bargu sus lea doppe?

'What kind of work she has there?'

<p>Son <b>barga</b> rektorin</p>	<p>0</p>	<p>You might have forgotten to write á i "barga"</p> <p>"barga": When inflecting Sg3, the a in 2. syllable changes to á</p>
<p>Son bargá rektorin 'She works as a headmaster'</p>	<p>1</p>	

## Student-computer interaction 4

Gosa manat dál?

'Where do you go now'

Mun manan hoteallii	0	"hoteallii" misses diphthong simplification
Mun manan hotellii 'I go to the hotel'	1	

## Recognized misspellings

<b>Errors</b>	<b>Norm.FST.</b>		<b>Err.FST</b>	
Not recognised	871	91.9%	563	56.0%
Recognized	77	8.1%	443	44.0%
Total	948	100%	1006	100%

**Table:** Parsing 2705 QA-pairs. Comparing the normal FST with the error-FST. Some sentences have more than one misspelling.

## Feedback to input

N=1561		
Misspellings	804	33,5 %
Syntactic errors	1071	44,6 %
Comments on semantics	527	21,9 %
Altogether	2402	100 %

**Table:** Parsing 2705 qa-pairs. Some sentences have more than one error feedback. Prec=0.96 Rec=0.99

## Correct answers

Questions demanding certain case in answer	Total	Correct answers	
Locative	2928	1388	47.4 %
Accusative	780	241	30.9 %
Illative	854	228	26.7 %
Comitative	196	44	22.4 %

Table: Correct answers

## Work in progress

The work is twofold:

- ▶ Making a good system for learning North Saami
- ▶ Making a good learner language corpus for research
  - ▶ get Finnish students to use the system
  - ▶ track the students according to L1
  - ▶ track the students individually (also questionnaire)

## Work in progress

The work is twofold:

- ▶ Making a good system for learning North Saami
- ▶ Making a good learner language corpus for research
  - ▶ get Finnish students to use the system
  - ▶ track the students according to L1
  - ▶ track the students individually (also questionnaire)

Giitu – Kiitos – Takk – Thank you!