

How to make the computer manage natural
language choices in a language learning process?
Linguistic and pedagogic problems.

Lene Antonsen, Biret Ánne Bals Baal,
Saara Huhmarniemi, Trond Trosterud



<http://giellatekno.uit.no/oahpa/>

The Sámi parliament and the University of Tromsø have financed the work

For grammar learning:

Word classes, syntax

In order to learn Sámi:

- ▶ **Leksa:** Sátnequiz - Sámi/Norwegian and Norwegian/Sámi
- ▶ **Numra:** Exercise numerals
- ▶ **Morfa:** Exercise word inflection, also in context
- ▶ **Vasta:** Exercise question answering
- ▶ **Sahka:** Participate in dialogues on a given topic

Pedagogical programs

Pedagogical programs usually do not contain language technology, but rather

- ▶ multiple choice
- ▶ string matching, e.g. *viesus* = 6 marks: v i e s u s

Pedagogical programs

Pedagogical programs usually do not contain language technology, but rather

- ▶ multiple choice
- ▶ string matching, e.g. *viesus* = 6 marks: v i e s u s

Language technology:

- ▶ analysis, e.g. *viesus* = *viessu* N Sg Loc

Vision

The program should supervise the student in the same way as a teacher does.

Vasta – exercise question answering



OHJA!

[Morfa](#) [Morfa - cealkagat](#) **Vasta** [Leksa](#) [Numra](#)

Level

First level ▾

Odda sánit

Suopman (not used)

Oarjesuopman

Nuortasuopman

Geaid doai dovdabeahtti?

Dáppe moai dovde dušše iežame fuolkkit

Maid don lohket ikte?

Acceptable answers:

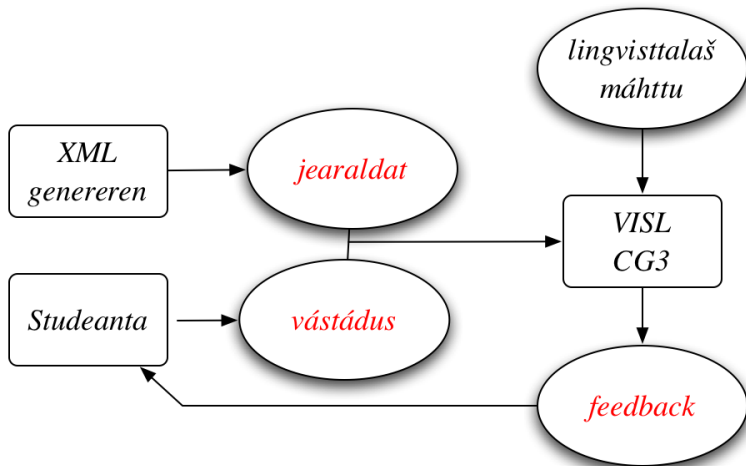
- ▶ Mun han lohken ollu áviissaid.
- ▶ Ikte mun gal lohken buori girjji.
- ▶ In lohkan maidege.
- ▶ Ikte in lohkan.

Maid don lohket ikte?

The Vasta-program gives feedback if the answer is not acceptable:

- ▶ Mun lohket ollu áviissaid.
→ Husk kongruens mellom subjekt og verbal.
- ▶ Mun lohken ollu áviissat.
→ Objektet skal være i akkusativ.
- ▶ Don lohket ollu áviissaid.
→ Er du sikker på at du svarer i riktig person?

Our system



Question generering

<text>Maid SUBJ MAINV ikte</text>

Linguistic knowledge

We use our knowledge about:

- ▶ Sámi syntax
- ▶ the learner's interlanguage

Sámi syntax

E.g. what a Sámi NP may contain:

- ▶ NP → Pron A N Num Adv A CC Adv A N
mu boares áhku guokte hui stuora ja hirmat váralaš beatnaga
- ▶ what kind of agreement inside the NP

Natural dialogue

D: Sidat go gáfe? G: In dieđe vuos.

Natural dialogue

D: Sidat go gáfe? G: In dieđe vuos.

→ Det er for enkelt å svare vet-ikke. Prøv igjen.

Natural dialogue

D: Siđat go gáfe? G: In dieđe vuos.

→ Det er for enkelt å svare vet-ikke. Prøv igjen.

D: Maid háliidat borrat? G: Láibbi.

Natural dialogue

D: Siđat go gáfe? G: In dieđe vuos.

→ Det er for enkelt å svare vet-ikke. Prøv igjen.

D: Maid háliidat borrat? G: Láibbi.

→ Svaret ditt må alltid inneholde et finitt verb.

Natural dialogue

D: Siđat go gáfe? G: In dieđe vuos.

→ Det er for enkelt å svare vet-ikke. Prøv igjen.

D: Maid háliidat borrat? G: Láibbi.

→ Svaret ditt må alltid inneholde et finitt verb.

D: Áiggut go vázzit bargui odne? G: Ale jeara nu olu.

Natural dialogue

D: Siđat go gáfe? G: In dieđe vuos.

→ Det er for enkelt å svare vet-ikke. Prøv igjen.

D: Maid háliidat borrat? G: Láibbi.

→ Svaret ditt må alltid inneholde et finitt verb.

D: Áiggut go vázzit bargui odne? G: Ale jeara nu olu.

→ Er du sikker på at du svarer i riktig person?

Problems – 1

Didaktihkka versus pragmatihkka

The goal is to train morphology – Solution:

- ▶ No elipsis
- ▶ The finite verb is compulsory
- ▶ Answer with the same verb when it is natural to do it
- ▶ No inclusive 1st person dual and plural
- ▶ The answer *I do not know* is not accepted

Problems – 2

No finite verb in the clause

**Mun vuolggan ihttin.*

→ Svaret ditt må alltid inneholde et finitt verb.

Problems – 2

Possible solution:

**Mun vuolggan ihttin.*

→ Svaret ditt må alltid inneholde et finitt verb. Kan det være en skrivefeil?

Two finite verbs in the clause

**Mun áiggun vuolggán. versus Mun boran haman.*

In a finite-finite-construction:

- the verbs should have same inflection
- no adverb between

But this is not enough

Problems – 3

Possible solution:

Semantic set:

LIST INFV = *astat ádjánit áigut álgit beassat berret bivvat*

Rule : Not possible *šith* (INFV finite) + (VERB finite)

Problems – 4

Nominative versus accusative

We cannot base our conclusion upon word order, and the subject is not compulsory

- ▶ We can utilize the question - if it asks for an object (but it is still possible to answer without an object)

Problems – 4

Possible solution:

Define the verbs and make semantic sets, e.g.:

- ▶ verbs which have object as a compulsory argument (Strict Transitive Verbs)
- ▶ verbs which cannot have a HUMAN as object

Problems – 4

Possible solution:

Define the verbs and make semantic sets, e.g.:

- ▶ verbs which have object as a compulsory argument (Strict Transitive Verbs)
- ▶ verbs which cannot have a HUMAN as object
borrat - HUMAN can be subject, not object
lohkat - the same, but a N Prop can be object,
e.g. *Ikte mun lohken Fosse.*

Problems – 4

Possible solution:

Define the verbs and make semantic sets, e.g.:

- ▶ verbs which have object as a compulsory argument (Strict Transitive Verbs)
- ▶ verbs which cannot have a HUMAN as object
 - borrat* - HUMAN can be subject, not object
 - lohkat* - the same, but a N Prop can be object, e.g. *Ikte mun lohken Fosse*.
 - and the verb has another meaning: *Mun lohken mánáid*.

Spelling errors

1. the word does not exist:
→ *X finnes ikke i vårt leksikon. Kan det være en skrivefeil?*
2. unintended lemma (leksem)
3. correct lemma, but unintended word form

We add the case suffix to the Nom

Our pedlexicon has 1512 nouns

▶ LOCATIVE -s/-is :

57 % correct lemma - unintended word form (PxSg3 - e.g. *viessus*)

0,5 % unintended lemma

(e.g. *eanas Adv (eatnamis)* or

verb *-stit* – imperative, verbgenitive, negation – e.g. *čogus (čohkumis)*)

We add the case suffix to the Nom

- ▶ LOCATIVE -s/-is:
 - 57 % correct lemma - unintended word form
 - 0,5 % unintended lemma
- ▶ ILLATIVE -i/-ii:
 - 0 % correct lemma - unintended word form
 - 2,3 % unintended lemma
 - (mostly Verb past tense Sg3, e.g. *báddii (báddái)*)

Spelling error gives unintended inflexion

e.g. possessive suffixes

biilas N Sg Nom Px Sg3 versus *biillas* N Sg Loc

**Áhčči lea biilas.*

Possible solutions:

- ▶ Remove possessive suffixes, except from when it syntactically is quite clear that it could be.
- ▶ Make comment to the user:
→ Mener du lokativ? I så fall er det feil stadieveksling.

Spelling error gives unintended lemma

- ▶ *viessut*: *viessut* Inf or *viessat* Imprt
but the student probably meant *viesut* N Pl Nom.

Spelling error gives unintended lemma

- ▶ *viessut*: *viessut* Inf or *viessat* Imprt
but the student probably meant *viesut* N Pl Nom.
- ▶ *luomos*: A Attr
**Eadni lea luomos*.
→ Her skulle det ikke vært attributtform.

Spelling error gives unintended lemma

- ▶ *viessut*: *viessut* Inf or *viessat* Imprt
but the student probably meant *viesut* N Pl Nom.
- ▶ *luomos*: A Attr
**Eadni lea luomos*.
→ Her skulle det ikke vært attributtform.
Gos eadni lea?
→ Svaret burde inneholde en lokativ.

Possible solutions:

- ▶ Remove problematic lemmas and word forms
- ▶ Identify wordpairs and ask the user:
→ Mener du viessu = hus? I så fall er det feil stadieveksling.

Correct or not

Better that errors slip through than not accepting what is correct

- muhto duhtágo geavaheaddji dasa?

Evaluation and improvement

- ▶ Feedback from users
- ▶ Feedback from teachers
- ▶ Collect a question corpus (Vasta-Internet log)