

Interactive Pedagogical Programs Based on Constraint Grammar

Lene Antonsen, Saara Huhmarniemi, Trond Trosterud
Centre for Sámi Language Technology



Parser-based CALL programs

Parser-based CALL programs for learners of North Sámi based on pre-existing LT resources developed at the University of Tromsø:

- ▶ finite state morphological analyser/generator (fst)
- ▶ constraint grammar (CG) parser
- ▶ number word generator (xfst)

The morphological analyser/generator is implemented with fst and compiled with the Xerox compilers twolc and lexc.

The morphological disambiguator is implemented in the CG-framework.

Previous accounts on parser-based CALL

Very few parser-based CALL (Computer Assisted Language Learning) programs are available for actual use online. We have looked at

- ▶ **e-tutor**, a program for teaching German to foreigners at <http://e-tutor.org/> with Head-driven Phrase Structure Grammar (HPSG). e-tutor gives very good feedback to student's errors, but the possible input is very restricted.
- ▶ **VISL**-suite of games for teaching grammatical analysis at <http://visl.sdu.dk/> with vislcg3. One of the programs accepts free user input. The input is analysed or changed into grammar exercises.

The screenshot shows a web browser window with the URL <http://oahpa.uit.no/oahpa/> in the address bar. The main content area has a textured green background and features the title "OAHPA!" in large, black, hand-drawn letters. Below the title is a paragraph: "Oahpa is a collection of games that are developed for North Sami learners. From the right, select the instruction language, and whether you prefer Eastern or Western dialect. Then select the game." The page is organized into a grid of game options, each with a colorful icon and a description. On the right side, there is a sidebar with two dropdown menus for "Interface language" (set to "English") and "North Sami dialect" (set to "western"). Below these are links for "Teacher's and student's guide", "VISL-games for North Sami", "About OAHPA!", and "Feedback". At the bottom left, there is a copyright notice: "Copyright 2009 The University of Tromsø. Contact oahpa@hum.uit.no".

OAHPA!

Oahpa is a collection of games that are developed for North Sami learners. From the right, select the instruction language, and whether you prefer Eastern or Western dialect. Then select the game.

 MORFA-S Practise morphology	 VASTA Answer to questions	 LEKSA Words and translations
 MORFA-C Morphology in sentential context	 SAHKA Dialogue game	 NUMRA Practise numerals

Interface language: English

North Sami dialect: western

Teacher's and student's guide

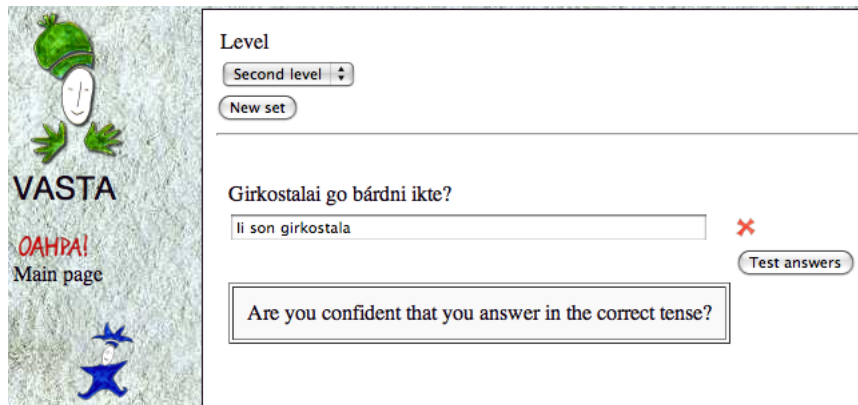
VISL-games for North Sami

About OAHPA!

Feedback

Copyright 2009 The University of Tromsø. Contact oahpa@hum.uit.no

Vasta – QA drill, questions generated from templates



Level

Second level ▾

New set

Girkostalai go bárdni ikte?

li son girkostala

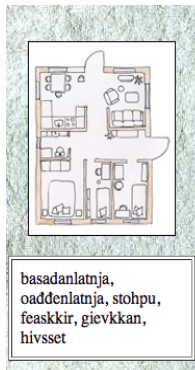
✗

Test answers

Are you confident that you answer in the correct tense?

("Did the boy go to church yesterday?"
"No, he does not.")

Sahka – dialogue program with precomposed questions



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

Buorre beavi! Bures bohtin 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. Gude lanjas TV lea du orrunsajis?

Dat lea stobus.

Gude latnjii moai bidje mu TV?

Moai bidje TV hivssegis.

✘ The answer should contain an illative.

Answer

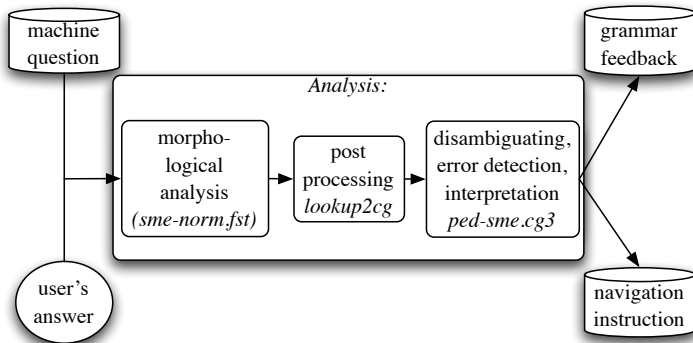
(Question: "In which room should we place the TV?"
Answer: "We should place it in the toilet (Loc).")

The Constraint Grammar parser *vislcg3*

- ▶ consists of manually written, context dependent rules which *add*, *remove*, *select* or *replace* readings containing tags or sets of grammatical tags in a given sentential context.
- ▶ Context conditions may be linked to any tag or tag set of any word anywhere in the sentence, either locally (in a fixed subdomain of the context) or globally (in the whole context).
- ▶ Context conditions in the same rule may be linked, i.e. conditioned upon each other, negated or blocked by interfering baseforms, wordforms or tags.

In 2007, we had a F-measure for Sámi at 95.9, and more developed CG grammars all achieve F-measure for precision and recall in the upper nineties.

Schematical view of the whole process



Morphological analysis

"<Gude>"

"guhte" Pron Interr Sg Acc

"guhte" Pron Rel Sg Gen

"guhte" Pron Rel Sg Acc

"guhte" Pron Interr Sg Gen

"<latnjii>"

"latnja" N Sg Ill

"<moai>"

"mun" Pron Pers Du1 Nom

"<bidje>"

"bidjat" V TV Ind Prs Du1

"bidjat" V TV Ind Prt Pl3

"<mu>"

"mun" Pron Pers Sg1 Gen

"mun" Pron Pers Sg1 Acc

"<TV>"

"TV" N ACR Sg Acc

"TV" N ACR Sg Nom

"TV" N ACR Sg Gen

"<^qst>"

"^sahka" QDL gosa_bidjat_TV

"<Moai>"

"Mo" N Prop Plc Sg Ill

"mun" Pron Pers Du1 Nom

"Moa" N Prop Plc Sg Ill

"<bidje>"

"bidjat" V TV Ind Prs Du1

"bidjat" V TV Ind Prt Pl3

"<TV>"

"TV" N ACR Sg Acc

"TV" N ACR Sg Nom

"TV" N ACR Sg Gen

"<hivssegis>"

"hivsset" N Sg Nom PxSg3

"hivsset" N Sg Loc

"hivsset" N Sg Gen PxSg3

"hivsset" N Sg Acc PxSg3

"<. >"

"," CLB

Assignment of navigation tag

```
"<Gude>"  
  "guhte" Pron Interr Sg Gen &grm-missing-Ill  
"<latnjii>"  
  "latnja" N Sg Ill  
"<moai>"  
  "mun" Pron Pers Du1 Nom  
"<bidje>"  
  "bidjat" V TV Ind Prs Du1  
"<mu>"  
  "mun" Pron Pers Sg1 Gen  
"<TV>"  
  "TV" N ACR Sg Acc  
"<^qst>"  
  "^sahka" QDL gosa_bidjat_TV &dia-hivsset  
"<Moai>"  
  "mun" Pron Pers Du1 Nom  
"<bidje>"  
  "bidjat" V TV Ind Prs Du1  
"<TV>"  
  "TV" N ACR Sg Gen  
"<hivssegis>"  
  "hivsset" N Sg Loc  
"<.>"  
  "." CLB
```

```
MAP (&dia-hivsset) TARGET QDL IF (Ø (gosa_bidjat_TV))  
(*1 ("hivsset") BARRIER ROOMS OR Neg) ;
```

Navigating in the dialogue – alternative links

```
MAP (&dia-hivsset) TARGET QDL IF (0 (gosa_bidjat_TV))
(*1 ("hivsset") BARRIER ROOMS OR Neg) ;
```

```
<utt type="question" name="gosa_bidjat_TV">
  <text>Gude latnjii moai bidje mu TV?</text>
  <alt target="hivsset" link="gosa_bidjat_TV">
    <text>Dat gal ii heive! Geahččal oddasit.</text>
  </alt>
  <alt target="default" link="gosa_bidjat_beavddi">
    <text>Moai gudde dan ovttas dohko.</text>
  </alt>
</utt>
```

Question: "In which room should we place the TV?"

Alt. toilet: "That is not a good idea. Make a new try."

Default: "We carry it there together."

Navigating in the dialogue – alternative branches

```
# Picking the age
MAP (&dia-adult) TARGET Num (*-1 QDL LINK 0 (Man_boaris_don_lear))
(0 ("([2-9][0-9])"r)) ; #
MAP (&dia-young) TARGET Num (*-1 QDL LINK 0 (Man_boaris_don_lear))
(0 ("([1][0-9])"r)) ; #
MAP (&dia-child) TARGET Num (*-1 QDL LINK 0 (Man_boaris_don_lear))
(0 ("([1-9])"r)) ;
```

```
<utt type="question" name="Man_boaris_don_lear">
  <text>Man boaris don lear?</text>
  <alt target="young" link="at_school_young"/>
  <alt target="child" link="begin_school_child"/>
  <alt target="adult" link="job_adult"/>
  <alt target="default" link="job_adult"/>
</utt>
</topic>
```

("How old are you?")

Disambiguation and assignment of grammar tag

```
"<Gude>"
  "guhte" Pron Interr Sg Gen &grm-missing-Ill
"<latnjii>"
  "latnja" N Sg Ill
"<moai>"
  "mun" Pron Pers Du1 Nom
"<bidje>"
  "bidjat" V TV Ind Prs Du1
"<mu>"
  "mun" Pron Pers Sg1 Gen
"<TV>"
  "TV" N ACR Sg Acc
"<^qst>"
  "^sahka" QDL gosa_bidjat_TV &dia-hivsset
"<Moai>"
  "mun" Pron Pers Du1 Nom
"<bidje>"
  "bidjat" V TV Ind Prs Du1
"<TV>"
  "TV" N ACR Sg Gen
"<hivssegis>"
  "hivsset" N Sg Loc
"<.>"
  "." CLB
```

```
MAP (&grm-missing-Ill) TARGET ("guhte") IF
(1 (N Ill) LINK *1 QDL LINK NOT *1 Ill OR
DOHKO OR Neg BARRIER S-BOUNDARY);
```

```
<message id="grm-missing-Ill">The answer
should contain an illative.</message>
```

The grammar errors we have rules for 1

Verbs and their arguments

- ▶ 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

The grammar errors we have rules for 2

Other

- ▶ agreement inside NP
- ▶ numeral expressions: case and number
- ▶ PP: case of noun, pp based upon the interrogative
- ▶ time expressions
- ▶ special adverbs
- ▶ particles according to word order
- ▶ comparison of adjectives

Meta comments

- ▶ "Answering *I-don't-know* is too simple. Try again."
- ▶ "Your answer must always contain a finite verb. Could there be a typo in the verbform?"
- ▶ "You must use one of the words in the wordlist in the left margin."
- ▶ "You have not used the correct adjective. Try again."
- ▶ The user can quit the dialogue in a proper way by using the verb "heaitit" (= to quit) – then the system navigates to the closing utterance of the dialogue (to be implemented)

Evaluation: The actual use of the system

- ▶ The programs are freely available at internet since Feb., 2009
- ▶ They get appr. 500 queries/day (not bad for a population of 20000 speakers)

The usage of the programs is not evenly distributed:

Morfa-S	Leksa	Sahka	Numra	Morfa-C	Vasta
41%	27%	13%	12%	5%	2%

Evaluation: How the rules have been working

If the system has identified an error in the user's input, its response is evaluated as follows:

Rule type	correct (tp)	wrong (fp)	corr. %
wrong tense	7	0	100,0
wrong V after neg	3	0	100,0
no infinite V	1	0	100,0
orth. error	44	2	95,7
wrong case for V-arg	26	4	86,7
no finite verb	19	4	82,6
wrong S-V agreement	17	8	68,0
wrong V choice	7	4	63,6
wrong word	4	4	50,0
wrong case after Num	1	1	50,0

The main problem for the users is *misspellings*

Evaluation: Some rules are (almost) not in use:

- ▶ agreement inside NP (except for numeral expressions)
- ▶ time expressions
- ▶ particles according to word order
- ▶ PP: case of noun, pp based of the interrogative

Possible reasons why they are not in use:

- ▶ The users do not give that elaborate answers
- ▶ Some of these errors (e.g. PP errors) are not that frequent

Evaluation: Precision and recall – today

Running the full log on the present system, we get the following results:

Error type	tp	fp	tn	fn	prec	rec.	acc.	F-ms.
Gramm. err.	641	234	769	7	0,73	0,99	0,85	0,84
Sem. err.	805	69	764	12	0,92	0,99	0,95	0,95
Orth. err	875	0	776	0	1	1	1	1
Other err.	695	180	751	25	0,79	0,97	0,88	0,87
	3016	483	3060	44	0,86	0,98	0,92	0,92

The high recall compared to the somewhat lower precision indicates that the system is a bit too critical towards the students:

- ▶ It almost never lets through a (targeted) mistake, with the price of flagging some correct answers as errors.

Conclusion

- ▶ By using a sloppy version of the syntactical CG analyser for North Sámi, combined with a set of error-detection rules, we have been able to build a flexible CALL resource.
- ▶ The present project has shown that CG is well fit for making pedagogical dialogue systems.
- ▶ The program suite is a novelty among pedagogical programs for Sámi, and indeed dialogue and open QA-programs are rare within the field of parser-based CALL.

Acknowledgments

Thanks to the faculty of Humanities at the University of Tromsø, and the Sámi Parliament in Norway, for funding the project.

Centre for Sámi Language Technology

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

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

References

- Kenneth R. Beesley and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI publications in Computational Linguistics. USA.
- Eckhard Bick. 2003. PaNoLa: Integrating Constraint Grammar and CALL applications for Nordic languages. Holmboe, Henrik (ed.): *Nordic Language Technology, Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*. 183–190, København: Museum Tusulanums Forlag.
- Eckhard Bick. 2005. Live use of Corpus data and Corpus annotation tools in CALL: Some new developments in VISL. Holmboe, Henrik (ed.): *Nordic Language Technology, Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*, 171–185. København: Museum Tusulanums Forlag.
- Johann Gampfer and Judith Knapp. 2001. A review of intelligent CALL systems. *Computer Assisted Language Learning* 15(4):329–342.
- Trude Heift. 2001. Intelligent Language Tutoring Systems for Grammar Practice. *Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online]* 6(2).
- Trude Heift and Devlan Nicholson. 2001. Web Delivery of Adaptive and Interactive Language Tutoring. *International Journal of Artificial Intelligence in Education* 12(4):310–325.
- Trude Heift and Mathias Schulze. 2007. *Errors and intelligence in computer-assisted language learning: parsers and pedagogues*. Routledge studies in computer-assisted language learning 2. New York : Routledge.
- Fred Karlsson and Atro Voutilainen and Juha Heikkilä and Arto Anttila. 1995. *Constraint grammar: a language-independent system for parsing unrestricted text*. Mouton de Gruyter.
- Trond Trosterud. 2007. *Language technology for endangered languages: Sámi as a case study*. <http://giellatekno.uit.no/background/rvik.pdf> University of Tromsø, Norway.
- VISL-group. 2008. *Constraint Grammar*. http://beta.visl.sdu.dk/constraint_grammar.html University of Southern Denmark.