# Interactive pedagogical programs based on Constraint Grammar



Lene Antonsen, Saara Huhmarniemi, Trond Trosterud University of Tromsø, Norway, {lene.antonsen; trond.trosterud}@uit.no; saara.huhmarniemi@helsinki,fi

#### Abstract

We have made a set of interactive parser-based CALL programs for North Sámi. The programs are based on a finite state morphological analyser and a constraint grammar parser which is used for syntactic analysis and navigating in the dialogues. The CG-parser provides effective and reliable handling of a wide variety of user input. In addition, relaxation of the grammatical analysis of the user input enables locating grammatical errors and reacting to the errors with appropriate feedback messages.



The Oahpa programs are freely available at http://oahpa.uit.no. The programs include a basic morphological exercises (Morfa-S), a question-answer (OA) drill (Vasta), word quiz (Leksa), morphological exercises in a sentential frame (Morfa-C), a dialogue program (Sahka) and a numeral quiz (Numra).

### Pedagogical lexicon

The OAHPA! programs share a set of common resources: a pedagogical lexicon and a morphological generator that is used for generating the different word forms that appear in the programs. The dialectal variation is taken into account in the lexicon as well as in the morphology. The semantic class is used in the sentence generator for Vasta and Morfa-C. The lexical entry for monni "egg" is given to the right.

```
stranglettones
      eth anti-targe "noblineggeders
eth anti-targe" film anung of tra-
  committees
com class="F000 Occum"/s
near class="Fold Occopy" As

otherwise class="to-plot of distribute" no

graphic="yes soggith" rise="2"/s-

sizatest class="00" bit/s-

staress

which near="50"/s-

shock near="sore /s-
```

## Morphological feedback



If the user does not inflect the lemma correctly in the morphological exercises, she can ask for hints about the inflection, and try once more, instead of getting the correct answer straight away.

The detailed feedback messages are determined by the combination of morphological features in the lexicon and the inflection task at hand. The morphological specification below gives a rule stating that there is a vowel change in illative singular for bisyllabic nouns that end with the vowel i. The corresponding feedback message instructs the user to remember the vowel change

```
when close-fitted latter engine the 
engine and "HII" number - "by but whap-
wheness
 message lámijálnítovel change i má.
                                            20 halp
 "mount" and experiently you and could have
```

The system-internal representation of monni states it is a bisyllabic i-stem, which triggers i > á change in illative.

The user types the errouneous monnii, and gets feedback from the machine.

A correct answer gets green colour as feedback.

#### Background and pedagogical motivation

The pedagogical programs in OAHPA! are based upon three pre-existing language technology resources developed at the University of Tromsø: a morphological analyser/generator, a CG parser for North Sámi and a number word generator compiled with the Xerox compiler xfst.

The main goal of the development of OAHPA! was to develop a language tutoring system going beyond simple multiple-choice questions or string matching algorithms, with free-form dialogues and sophisticated error analysis. Immediate error feedback and advice about morphology and grammar were seen as important requirements for the program.

The sentence generator in Morfa-C and Vasta is able to generate a virtually unlimited number of different tasks, and allows the student to use the programs over and over again.

### Constraint Grammar (CG)

Constraint grammar is a syntactic framework for choosing correct grammatical analysis of a given wordform, based upon the context it occurs within. Each rule removes or selects readings, or adds or removes a syntactic tag. Inappropriate analyses are removed, but the last analysis is never removed. CG thus always gives an analysis, and is therefore a very robust framework, well fit to handle potentially erroneous input.

```
"destate" from Sature 1, the
health' from Sature from
Sature Sature from
"All negligibes" is 54 sec
"all negligibes" is 54 sec
        State Peop Box Carden
State Peop Box Lyl Ace
State Peop Box Carden
State Peop Box Carden
State State
   The House' with the fee fee fee 
feel failed with the fee fee fee 
feel call.
"Abbition" Thee December 850 854 854 2

"Indicate" Free December 850 854 854 2

"Little Red 45 Ft. Acc 8003-80-44

"east Store Paris SQL Hos 45403-90-44
```

schildren was a te and the against a serve

The morphological analyser gives the words in Makkár láihegályvuid don háliidat? "What kind of bread do you want?" all possible grammatical analyses

The CG grammar then picks the correct analysis, and adds grammatical function and dependency structure.

#### Evaluation

The overall evaluation shows that the students answer correctly slightly half of the time. By far the most popular program is the basic morphological drill (but the interactve programs have been logged for a couple of days only).

Program	District	"The same	Test	- 6
March 5	6/202	A 177	1772	10.0
Leben	1000	4240	9000	107.4
The same	109.0	2.010	3.356	99.4
March 21	150.0	1670	3030	48.8
Section	100	200	444	19.5
See la	1.6		10.0	13.7
Table	1700	18 179	1000	95,44

http://oahpa.uit.no

The 322 logged Sahka errors are distributed along the following

Allertic byggs	-	And Support	
and the second	100	management (France)	$\mathbf{I}$
series are a	23	are assessful Name	
manage 2000 ages	46	manage branch	
and the limited by		and professions	
weavy Assistant	24	wrong mod	7

For Sahka we test precision (correctly identified errors/all diagnostised errors) recall (correctly identified errors/all errors) and accuracy (correct judgements/cases)

Precision = 0.8; Recall = 0.68; Accuracy = 0.82 (N=584)

#### Conclusion

By using the syntactical analyser for North Sámi, combined with a set of error-detection rules, we have been able to build a flexible CALL resource. The programs are modular, and the modules may be improved by adding more materials words, tasks, dialogues, levels, words from textbooks. The CG parser framework was originally chosen as parser framework for Sámi due to its extraordinary results for free-text parsing. The present project has shown that CG is well fit for making pedagogical dialogue systems as well

#### Article version of this poster:

Antonsen, Lene, Saara Huhmarniemi and Trond Trosterud 2009: Interactive pedagogical programs based on constraint grammar. Proceedings of the 17th Nordic Conference of Computational Linguistics. Nealt Proceedings Series 4. http://dspace.utlib.ee/dspace/ handle/10062/9206

# CG-parser in live analysis in the interactive programs Vasta and Sahka

The programs are based upon free-form interaction: Within certain limits, the student may formulate her own

We use constraint grammar to disambiguate the student's input only to a certain extent, because there will probably be grammatical and orthographic errors. The manually written, context dependent rules are mainly used for selecting the correct analysis in case of homonymy. The last part of our grammar consists of rules for giving feedback to the student's grammatical errors, and rules for navigating to the correct next question of in the dialogue, due to the student's answer.

The system question and student answer are merged and analysed together, delimited by the boundary marker ^qst ODL. They first get a morphological analysis, and are then disambiguated, and, if possible, assigned an error tag or a navigation tag.

# America of a greates with the matrices. Namentes by miles there is weather. 100 E-10

2011/01/25/25 (2010)

Above is a part of a dialogue in Sahka on furnituring a flat.

Below is the analysis of the third question and answer pair.

The morphological analysis, is disambiguated and there are

assigned a grammar-error-tag (&grm-missing-Ill) and a

ericinalistaje, induse 1 maio, portant

navigation-tag (&dia-hivsset):

"mdja." • IV and ere but.

"Man" Provi Fors Sgs for "ATO"

"mo" Proc Fors the few

with a wind sector

THE RESERVE

"Attraingle-"
"Yoursel" Hillie Loc

"IN" IN MAKE SO MAKE

Tourist From Entern Sg Act Spre-missing-Still Special error undern Sg See Spre-missing-still "eletricits" is Sg Still.

"entre" qui gara biojet... y écie-francet

Service Confidence in .

Jacks South Laws South In Jacks 1 And have been as Code length areas stelly over TVP. New York and State and St.

Man have ablabe Obere who believe with recommobile Man. حوافعت بالمشاهدة مهوشة الحجاة ومهدم يتمول لمح Name (on THE Administration TA) in the comment of The part of should contain provide to

## Grammar feedback

The system may give feedback to grammatical errors. In the question which analysis is to the left, the systems asks "Where do we put the TV set?" The student answers Moai bidje TV hivssega ("We put the TV in the WC"), with accusative hivssega rather than the correct illative

change id-form-nighting-310.5 The or men-dicate contain an interpretation ages

grammar-error-tag (&grm-missing-Ill) to the input sentence (the analyse to the left). In the grammar feedback library, the tag in question looks up a message in the appropriate interface language (in this example, English), and the user is presented with the feedback The answer should contain an illative, as shown above

The CG-rules disambiguate the input, and the rule above adds a

### Navigation



Navigating inside the dialogue is implemented in CGrules. The user input is tagged during analysis with information on whether the answer is interpreted as affirmative or negative. In addition, a special tag indicates whether the sentence contains some information that should be stored for the following questions or utterances. The program is thus able to store simple information such as the student's name, place where she lives and for example the type of her car and use this information in tailored utterances.

In the example to the left the question is "In which room do we put the TV?" One of the alternatives for the navigation is due to the target tag being assigned because of the lemma hivsset ("WC"). The answer will be "That is not a good idea. Make a new try." The CG-rule is:

```
SAF Chair History DANCE (N. II (6 OpensubsdjetsTV))
(*) ( himset ) Bakking Fales )
```

There are alternative links in the dialogue, one of them is due to the &dia-hivsset tag:

```
said: hyper/figurables/* reconfigure of djob (Vf),
should faith table just man indige on PV/V back
odd, transfer from the transfer of the reconstitution of
creatilities get 11 between Excellent schools of
and to with target "definabl" their "pera budget beautiful" chief. Acet. matte the cetter their, of their cetter.
```

Every question has its own unique id, which is used in navigating between questions. In addition, the CG-rules may be tailored for specific questions, like in the rule above.

Age-tags are assigned with help of regex-rules to the answer to the question "How old are you?". Due to the tags the system choose the correct link for moving to the next dialogue branch tailored to the student's age.

# Schematical view of the process

