Using authentic texts for grammar exercises for a minority language

Lene Antonsen, Chiara Argese

http://giellatekno.uit.no/

UiT / THE ARCTIC UNIVERSITY OF NORWAY

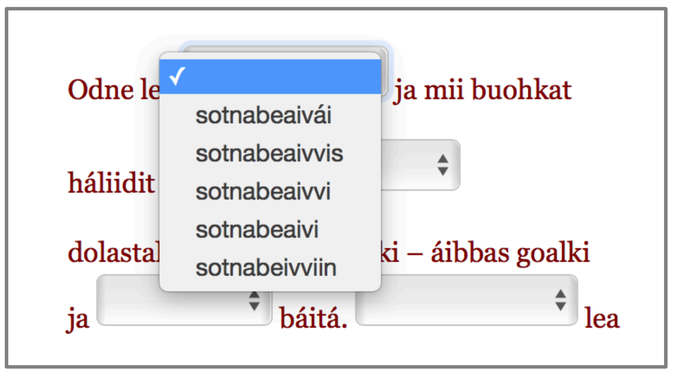
Introduction

We present an ATICALL (AuthenticText ICALL) system with automatic visual input enhancement activities for training complex inflection systems in North Saami. We have adapted the freely available VIEW (Visual Input Enhancement of the Web) system.

Our system, called **Konteaksta**, is based on finite state transducers (FST) and Constraint Grammar, originally built for other purposes.

North Saami language

North Saami is a morphology-rich language, with nominal inflection for two numbers, six cases, and possession. Nouns have paradigms both with and without possessive declension indicated. Verbs have 45 finite forms including three persons for singular, dual and plural, in four modi, and two tenses for indicative.
The verbs are also inflected for ten different non-finite forms. Nouns, adjectives and verbs may be divided into groups according to stem type, each type having different paradigms. Suffixation is accompanied by phonological alternations.



Odne lea **sotnabeaivi** ja mii buohkat háliidit vuolgit **mearragáddái** dolastallat. Lea čáppa **dálki** – áibbas goalki ja **beaivváš** báitá. **Beaivvádagas** lea oalle liekkas.

Odne lea sotnabeaivi ja mii buohkat
háliidit vuolgit mearragáddái dolastallat.
Lea čáppa _____ (dálki) – áibbas goalki
ja _____ (beaivváš) báitá. Beaivvádagas
lea oalle liekkas.

http://oahpa.no/konteaksta

Activities and target words

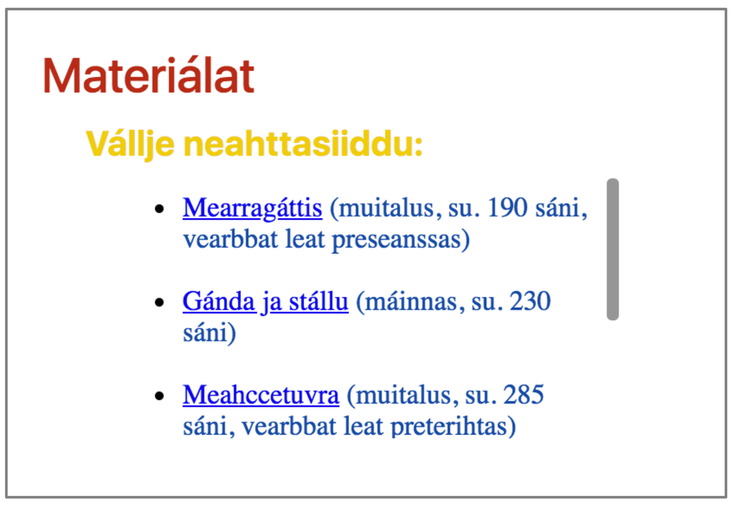
The VIEW system includes four different types of activities:
Highlight, adds colour to target wordforms, to make the learner aware of the morphology
Click, the learner clicks on target wordforms
Multiple-choice, the user selects correct word-form from a list
Cloze, the learner types in the correct wordform

Target words for North Saami are:
- Nouns
- Finite verbforms
- Non-finite verbforms
- Negation form of verbs

Generating key-answers

Instead of using the original text, we decided to generate all key-answers, based on the morpho-logical analysis of the target word. The reasons for this are:

- much variation in orthography.
- much misspellings.
- that it makes it possible to accept more morphological forms. This allows us to include also ambiguous target words.



From the list under under Materiálat (“Materials”) the user can choose to work with recommended web-texts, mainly from textbooks published on the web. Each link has information about genre, length and features like tense.

Alternatives to using “any” web-text

For a minority language there are challenges in finding suitable web-texts. Literacy among Saamis is not very strong, and Saami web-texts tend to be short, and with many misspellings. Even if the generation of key-answers is a solution for misspellings, the high rate of misspellings makes the analysis less reliable. The ATICALL approach becomes less pedagogical for learners, because they are exposed for the misspellings.

The solution is using proofread materials:
- using texts from textbooks published on the web, and giving links to these texts as “recommended texts”.
- implementing the possibility for teachers to upload proofread material or their own texts. They may then send the URL for each activity to their students.

Even if we find good texts in North Saami on the web, they often contain fragments of the majority language (Norwegian, Finnish, Swedish), like a menu, or a dateline.

Cooperation and feedback from users

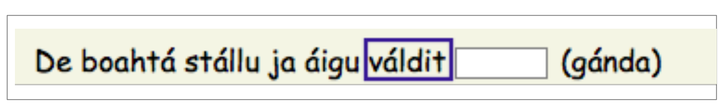
A group of teachers at an upper secondary school are using Konteaksta, and we have adjusted the design according to their recommendations.

Also a teacher and her students at our university have given user feedback, and we are also planning a “usage session” with some students to see how they understand and use the the options in the program.

Despite there being further improvements we are aware of and not yet have implemented, Konteaksta is welcomed by students and teachers in both schools and universities, especially because of the sparseness of learning materials for North Saami.

```
"<De>"
    "de" Adv @ADVL>
"<boahtá>"
    "boahtit" V IV Ind Prs Sg3 @+FMAINV
"<stállu>"
    "stállu" N Sem/Hum Sg Nom @<SUBJ
"<ja>"
    "ja" CC @CVP
"<áigu>"
    "áigut" V IV Ind Prs Sg3 @+FAUXV
"<váldit>"
    "váldit" V TV Inf @-FMAINV
"<gándda>"
    "gánda" N Sem/Hum Sg Acc @<OBJ
"<.>"
    "." CLB
```

Future work: more sophisticated feedback for the user




```
De   boahtá stállu    ja  áigu váldit ..... (gánda)
Then comes  the troll and will take   ..... (the boy)
```

Now the feedback for the user is only that the answer turns red if it is wrong, and green if it is correct. We want to give more sophisticated feedback, based on the grammatical analysis of the sentence. For many of the targets, one can as a first feedback to an incorrect answer, highlight in blue a word as a hint for choosing the correct form. In this example the verb *váldit*, which is a transitive verb in infinitive, V TV Inf, is highlighted because it triggers the accusative case for the target word, which is the object in this sentence.

REFERENCES

Meurers, Ziai, Amaral, Boyd, Dimitros, Metcalf, and Ott. 2011. Enhancing authentic web pages for language learners. In Proceedings for the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-5) at NAACL-HLT 2010, pages 10–18, Los Angeles.
Reynold, Schaf, and Meurers 2014. A VIEW of Russian: Visual input enhancement and adaptive feedback. In Proceedings of the 3rd workshop on NLP for computer-assisted language learning at SLTC 2014. Uppsala University, pages 98–112. Linköping Electronic Conference Proceedings.

7th Workshop on NLP4CALL, SLTC 2018
Stockholm, Sweden, 7 November 2018