

Sámi proofing tools

Main goals

- a set of proofing tools, freely available
- a reusable tool set
- increased knowledge
- better linguistic resources
- foundation for future tools (other Sámi languages, grammar checkers, more)

Limits

- project finished by 30.9.2007
- no more than 11,3 mill. NOK
- 4,5 persons in the project team
- grammars, corpora
- technical constraint: tools must be Unicode-based

Project language: English

- Goal: reusable tool set and framework
 - we want to help others, including those outside of Scandinavia
- Only common language for all team members
- English is only required for documentation, and communication between all of us

Deliverables

- Proofing tools for Sámi as follows:
 - Speller for North Sámi
 - Speller for Lule Sámi
 - Hyphenator for North Sámi
 - User documentation and installation programs, localised to the two Sámi languages and FIN, SVE, NOR (and maybe English)

Deliverables (2)

- Proofing tools for these OSes:
 - Windows 2000 and newer
 - MacOS X 10.3 and newer
 - Debian Linux with Sámi support (as pr Skolelinux)

Deliverables (3)

- Proofing tools for the following environments:
 - MS Office
 - OpenOffice
 - InDesign (hyphenator only)
 - Aspell or similar

Our tasks

- finalise the linguistic model
- extend and complete the lexicon
- test the lexicon and linguistic model against actual texts
- build a reasonable corpus base
- document, both technically and for the end users
- create and test inst. packages for Aspell & OpenOffice

Outsourced tasks

- Integrate with and compile against MS Office
- Integrate with InDesign
- Possibly create the MS Office installation packages, and prepare localisation of them
- Possibly integrate with OpenOffice
- Help developing conversion scripts from our internal format to theirs

Main milestones

- Oct. 1, 2004: Official start, project manager in place
- Jan. 1, 2005: Effective start, all project members in place
- Jul. 1, 2005: Deadline for invitation to tender; infrastructure ready, project organisation and communication working well

Main milestones (2)

- Jan. 1, 2006: Outsourcing contracts signed, cooperation has started; ALFA of North Sámi aSpell (sme-aSpell-ALFA).
- Jun. 1, 2006: ALFA of North and Lule Office spellers (MSOffice-ALFA, OOffice-ALFA), ALFA of North Sámi pro hyphenator (InDesign-ALFA); corpus building finished

Main milestones (3)

- Jan. 1, 2007: BETA of all proofing tools (aSpell-BETA, MSOffice-BETA, OOffice-BETA, InDesign-BETA).
- Jul. 1, 2007: All proofing tools ready, version 1.0 is publicly released (aSpell-FINAL, MSOffice-FINAL, OOffice-FINAL, InDesign-FINAL).
- Sep. 30, 2007: Project closing, all documentation and project info handed over to maintenance group; 1.0 bugs fixed

The Future

- More Sámi languages, other languages
- More tools:
 - grammar checkers
 - inflecting thesauri
 - inflecting dictionaries
 - speech technology

The Future in practise

- Project should be changed to a permanent maintenance organisation
- Will be smaller
- The maintainers will serve as the foundation for the future expansion into new tools and new languages