Is it possible to create computer tools for indigenous languages?
The Divvun and Giellatekno groups at UiT

• **Divvun (tool development):**
  – Børre Gaup
  – Elena Paulsen
  – Linda Wiechetek
  – Sjur Moshagen

• **Giellatekno (academic research):**
  – Lene Antonsen
  – Trond Trosterud

• **More people at home (altogether 11)**
Tools

- Computer keyboards
- Cell phone keyboards
- Word analysers
- Dictionaries with grammar
- Spelling checkers
- Phone keyboards with spellers
- Word form generation
- Hyphenation
- Language learning tools
- Grammar checkers
- Machine translation
Computer keyboards

- On Windows the keyboard installer also registers the locale
- Makes Windows know your language
- Similarly on macOS

Computer keyboard for Skolt Sámi
Cell phone keyboards

North Sámi cell phone keyboard
… for any syllabic or alphabetic writing system

Prototype Cree Syllabic keyboard
In cooperation with AltLab at University of Alberta

Prerelease Komi Cyrillic keyboard
In cooperation with FU-lab in Syktyvkar
Spelling checkers

The spelling checkers work on:
• Linux, macOS, Windows

They are:
• open source & fast

They work with:
• MS Office
• LibreOffice
• Web apps
• … and more

Spelling checker for South Sámi in MS Word
for all languages

- Prerelease Cree speller
- Works in MS Word
- With the correct language code
Cell phone keyboards with spelling checkers

- Fast and responsive
- Available for:
  - Android
  - iPhone
  - Chromebook coming soon

North Sámi cell phone keyboard with spelling checker
Dictionaries with grammar

Includes:

- Analyser to look up any word form
- Generator to generate paradigm

South Sámi dictionary
Language learning tools

- Using our analysers to give better feedback
- ... and to create dynamic exercises from real texts

North Sámi language learning webapp
Grammar checkers

North Sámi grammar checker
Machine translation

Translate:
- Text as you type
- Documents
- Web pages

Main point:
- Frees the indigenous community from having to use the majority language

North Sámi to Norwegian machine translation
The Divvun Installer

- Windows & macOS
- A tool to install all writing aids
- Keep all tools up-to-date
Demo 1

• North Sámi keyboard with speller
Demo 2

• You write this … and get this … … on your phone:

modes:
  mobile:
    default: |
      > J C F Q W E R T Y U I O P | |
      < L C F V B N M " . , | |
                        A S D G F T Y R | |
                        Z X C V B N M - | |

Instant cell phone keyboard! Try hands-on after the talk!
How did we do it?

- All your words
- All word endings etc
- Sound changes

Put it all together → your language

Finetune for different use, like:

- Analysers
- Spellers
- Hyphenators
- Stemmers

Add: runtime code and installer
Our infrastructure

- Free and open
- Code publically available in Github and Subversion
- Encourages reuse
  - One source for all tools
- Rule based → any language
- … even without digital texts
Issues we do not control

• **Main obstacle:** Computer systems are closed
  • Most systems only recognise a couple of hundred languages, the vast majority of languages are left out
  • Language technology such as Siri, Amazon Alexa, Google Translate is important to the big companies, and they keep it to themselves
    • As a consequence, all indigenous languages are blocked from their systems
  • Example: machine translation
Solution

• Open up
• Clearly defined interfaces
  • => we can implement tools

• Mainstream language technology not fit for most indigenous languages

• We have working technology — but are often not allowed to use it

• Language communities must control their own language technology
Most languages need a computational model of their grammar

This holds for:

- *All indigenous languages* (except the ones in Polynesia)
- At least 2/3 of the world's 7000 languages
- For some languages a computational model may come in handy, but they may also do without

*Note:*
The smaller amount of text that is available for the language, the stronger is the need for a computational model of the grammar
Areas with highly inflecting languages
Indigenous languages with computational models

At least the following languages (those with released tools are boldfaced) with / without our infrastructure:

• Nordic countries: South, Pite, Lule, North, Inari and Skolt Sámi
• North-East: Tundra Nenets, Evenki, North Mansi / Chukchi
• North-West: North Slope Iñupiaq, Greenlandic / Inuktut
• North America: Plains Cree, Northern Haida, Odawa, Tsuut’ina / Navajo
• South America: / Aymara, Kichwa, Cuzco Quechua, Guaraní

… and several more. The number of grammar models is rising

Unfortunately, too many computational models are not turned into tools
How to do this in practice

Linguist

Language expert

Programmer
Computational model of the language

Words Language in use

Grammar

Mobile: Keyboard with proofing

Proofing tools

Dictionary with grammar

Machine translation

E-learning

.....
This is possible for every language:
We offer open source infrastructure and integration into end-user programs.
Conclusion

• Yes, it is possible to create computer tools for indigenous languages
• No quick fix — good tools require a lot of work
• … done by the language community with linguists and programmers
• GiellaLT infrastructure is one way of turning the computational model of the language into practical programs
• Provides working solutions used every day by indigenous communities

For more information:
We will be around here till Friday
indigenous-langtech.uit.no