

UiT

THE ARCTIC
UNIVERSITY
OF NORWAY

Is it possible to create computer tools for indigenous languages?



Norwegian Ministry
of Local Government
and Modernisation



SÁMEDIGGI
SÁMI PARLIAMENT

✓ ÁBČ
Divvun

Giellatekno



2019 | INTERNATIONAL YEAR OF
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)**



Romsa / Tromsø



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
- 

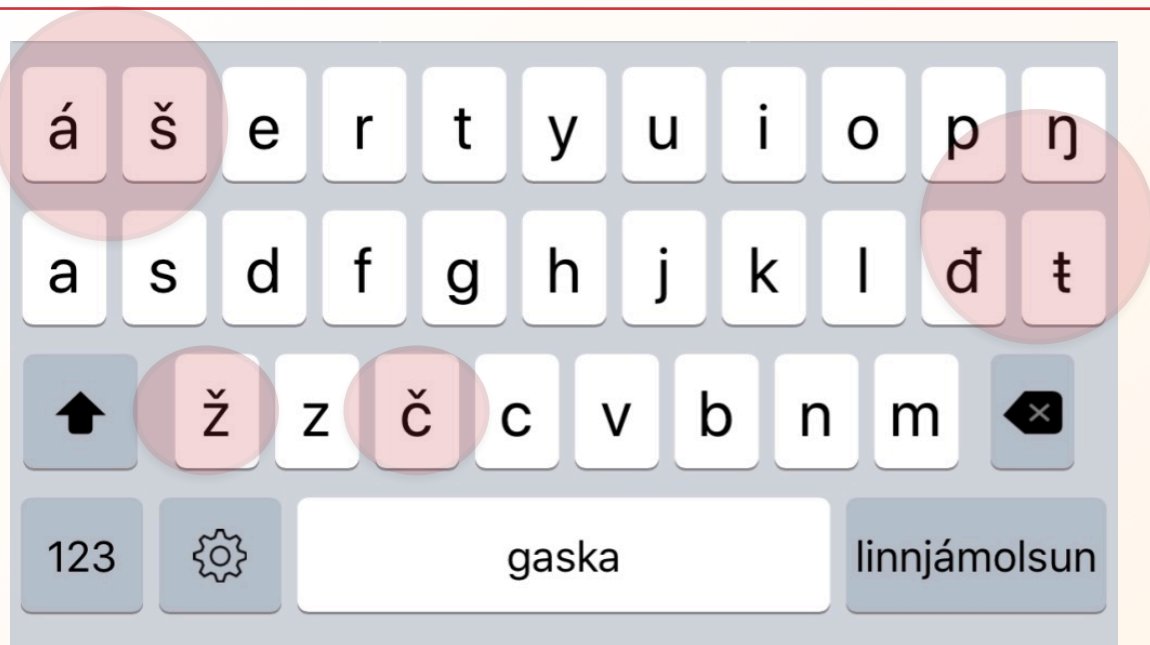
Computer keyboards

- 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



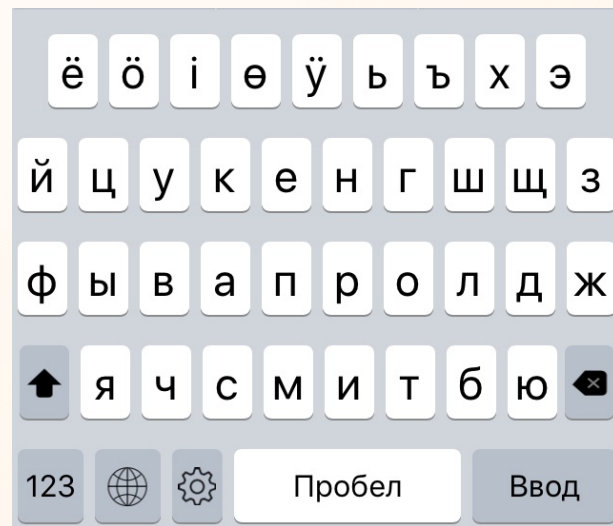
North Sámi cell phone keyboard

... 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:

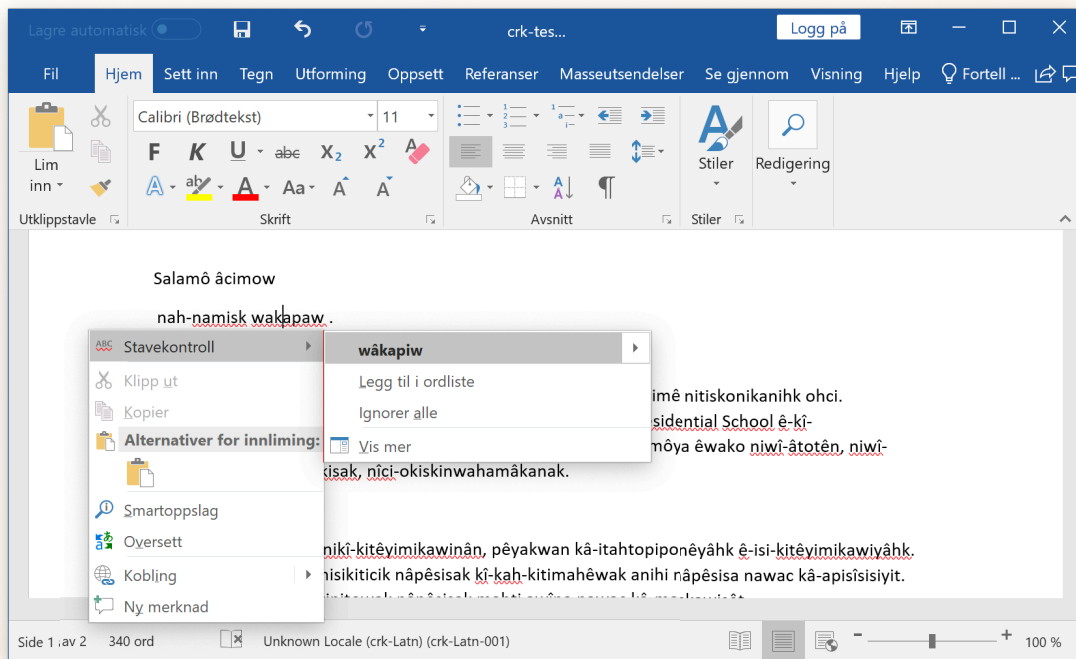
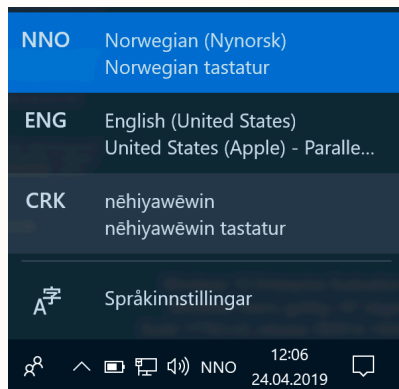
- MS Office
- LibreOffice
- Web apps
- ... and more



Spelling checker for South Sámi in MS Word

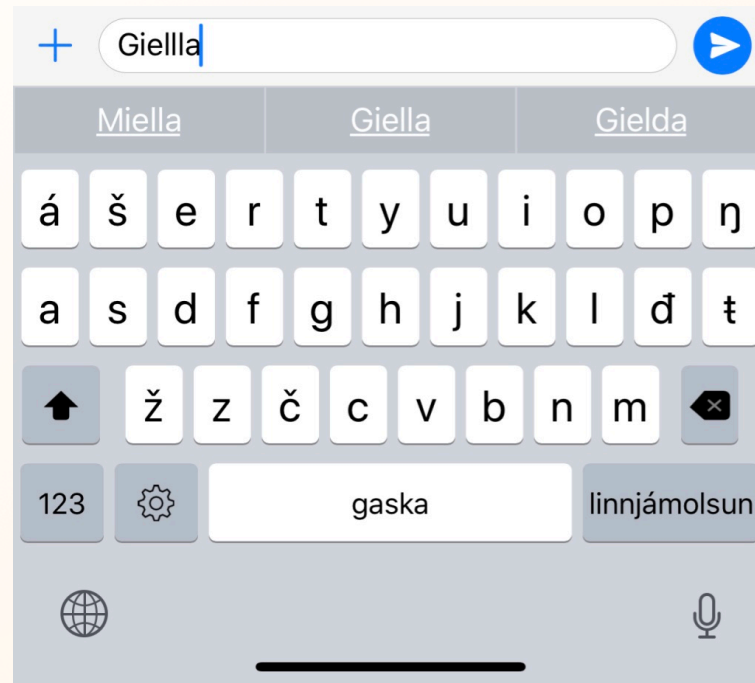
... for all languages

- 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

sma→nob ▾	eahtsa	Ohtsh
iehtsedh (verb)	PRESENS (daan biejjien)	PRETERITUM (jååktan)
◦ elske , være glad i	1.p.ent. (manne) eahtsam	iehtsiejim eehtsim
Manne datnem eahtsam. <i>Jeg elsker deg.</i>	2.p.ent. (datne) eahtsah	iehtsiejih eehtsih
◦ engste seg for noen	3.p.ent. (dih̄te) eahtsa	iehtsieji eehtsi
Datneste eahtsam. <i>Jeg er engstelig for deg.</i>	1.p.tot. (månnoeh) iehtsien	iehtsiejimen eehtsimen
	2.p.tot. (dåtnoeh) iehtseden (dåtnoeh) iehtsiejidien	iehtsiejiden eehtsiden
	3.p.tot. (dah guaktah) iehtsiejægan	iehtsiejigan eehtsigan
	1.p.ftt. (mijjieh) iehtsebe	iehtsiejimh

South Sámi dictionary

Language learning tools

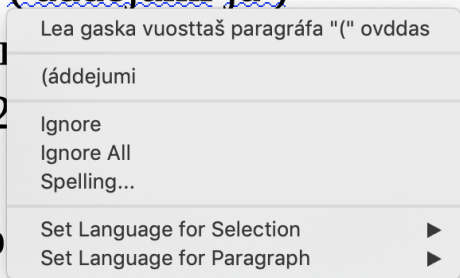
- Using our analysers to give better feedback
- ... and to create dynamic exercises from real texts

The screenshot shows the OAHPA! webapp interface. At the top, the title "OAHPA!" is displayed in a stylized font. To its right are icons for MORFA-C, MORFA-S, VASTA, SAHKA, LEKSA, and NUMRA. On the left side, there is a cartoon character wearing a green hat and a blue shirt, with the text "VASTA" below it. The main content area includes a "Dássi" (Lesson) section with a "First level" dropdown menu. Below this is a button labeled "Ođđa bargobihát" (New exercise). The exercise prompt is "Maid mii oaidnit?" (What do I see?). The user's input is "Dii oaidnibehtet nieida." (The girl sees the girl.), which is marked as incorrect with a red 'X'. Below the input is a button labeled "Iskka vástádusaid" (Check answers). The feedback message is "Nominative doesn't go with a transitive verb." and the score is "Du čuoggát: 0/1". On the right side, there is a "Veahkkegiella" (Foreign language) dropdown menu set to "English" and a "Grammar explanations" button. A text box on the right contains the sentence: "Vástit olles cealkagiin. Fuobmá ahte jus jearaldagas lea moai/mii, de don vástidat doai/dii." (Answer all questions. Remember that if the subject is someone, you answer with doai/dii.) At the bottom, there is a "Resurssat" (Resources) section with links to "Bagadus" (Dictionary), "Neahttasátnegirji" (Online dictionary), and "Grammatihkka" (Grammar). The footer includes the copyright notice "Copyright 2012 Romssa universitehta" and "Contact oahpa@hum.uit.no", a link to "Linja dán hárrjehussii" (This line is for the teacher), and a "HELP" button.

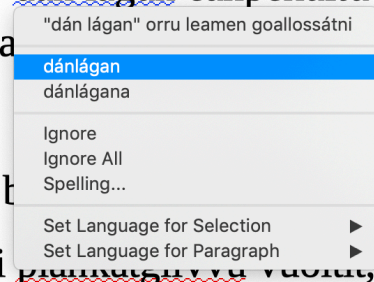
North Sámi language learning webapp

Grammar checkers

birrasiin. – Go olbmot rahpasit
ktá buoret (áddejumi ja)
mustahtti
ahii jagis 2
mofiillaid
illat lea jo
sohkabeallerollat, bearašoktavuođat,



– li oktage galgga oažžut lobi geavahit veah
nannen dihte iežas vuoigatvuodaid. Lea bu
bearráigeahččat ahte dán lágan eahpekultuv
boahhteáiggis, loahpa
Preassadieđáhus
Ustitvuodain vuoitá
VUOITI: Sámedikki



North Sámi grammar checker

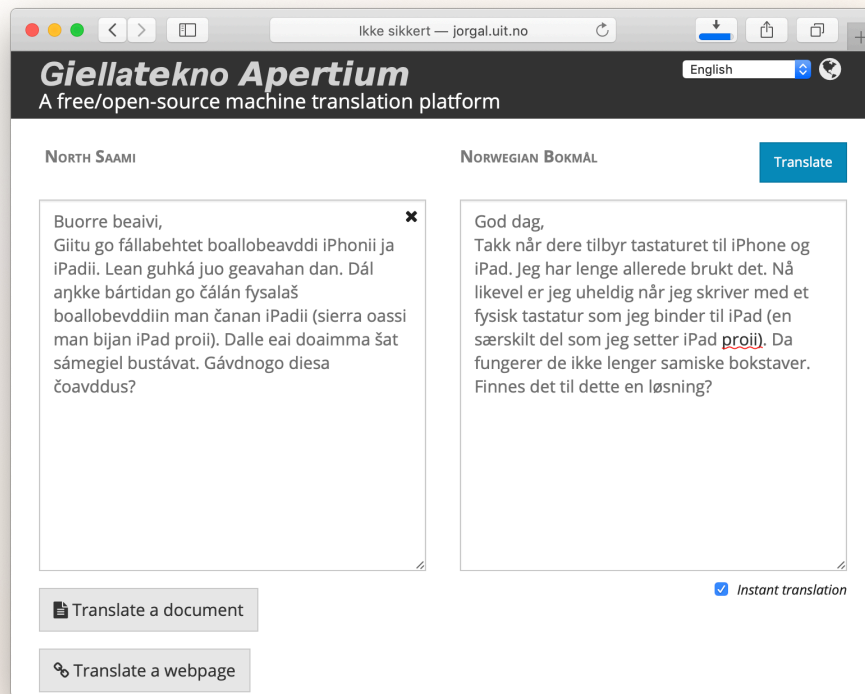
Machine translation

Translate:

- Text as you type
- Documents
- Web pages

Main point:

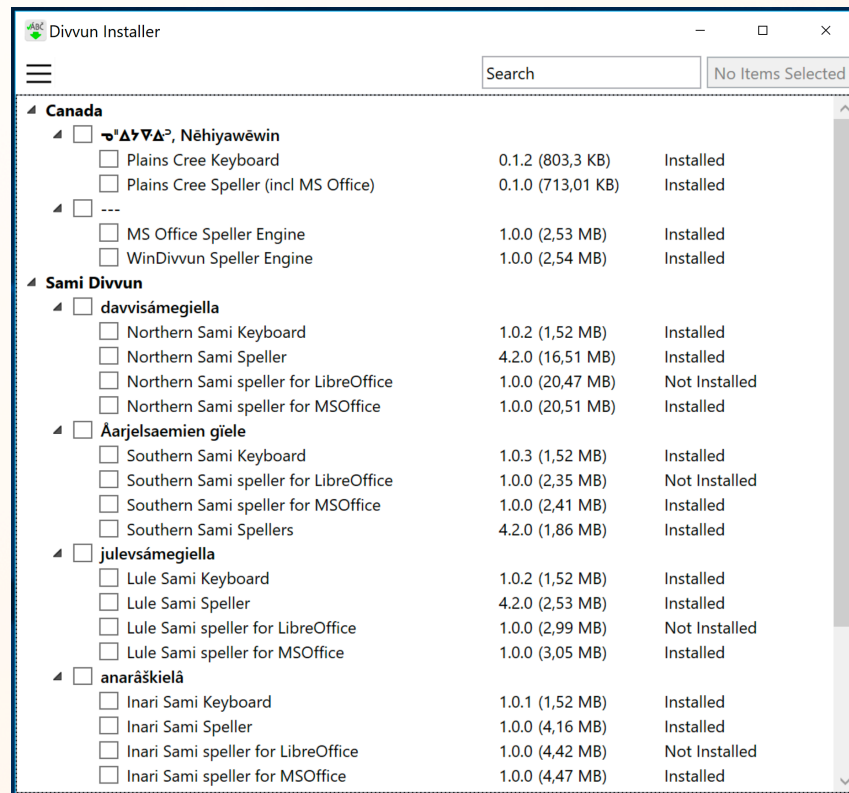
- Frees the indigenous community from having to use the majority language



North Sámi to Norwegian machine translation

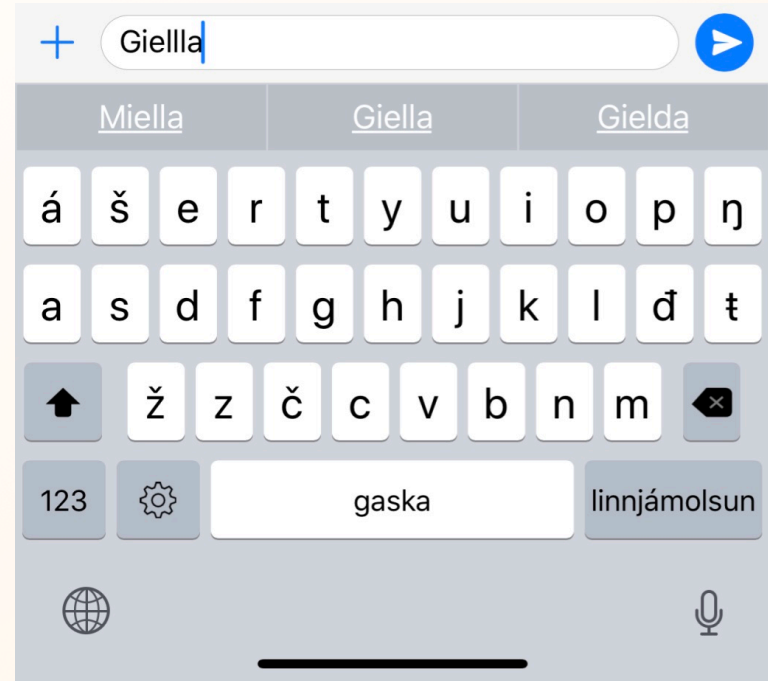
- Windows & macOS
- A tool to install all writing aids
- Keep all tools up-to-date

- 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 ...

```
modes:  
  mobile:  
    default: |  
      > J ɔ ▷ ↵ d ɔ ɾ . / \  
      < ɔ C ◁ ↵ L b ɔ ɔ ɔ ɔ ɔ  
      ^ ɾ ɾ Δ ɾ ɾ ɾ ɾ ɾ ɾ ɾ  
      V ɾ U ɾ ɾ ɾ ɾ ɾ ɾ ɾ
```

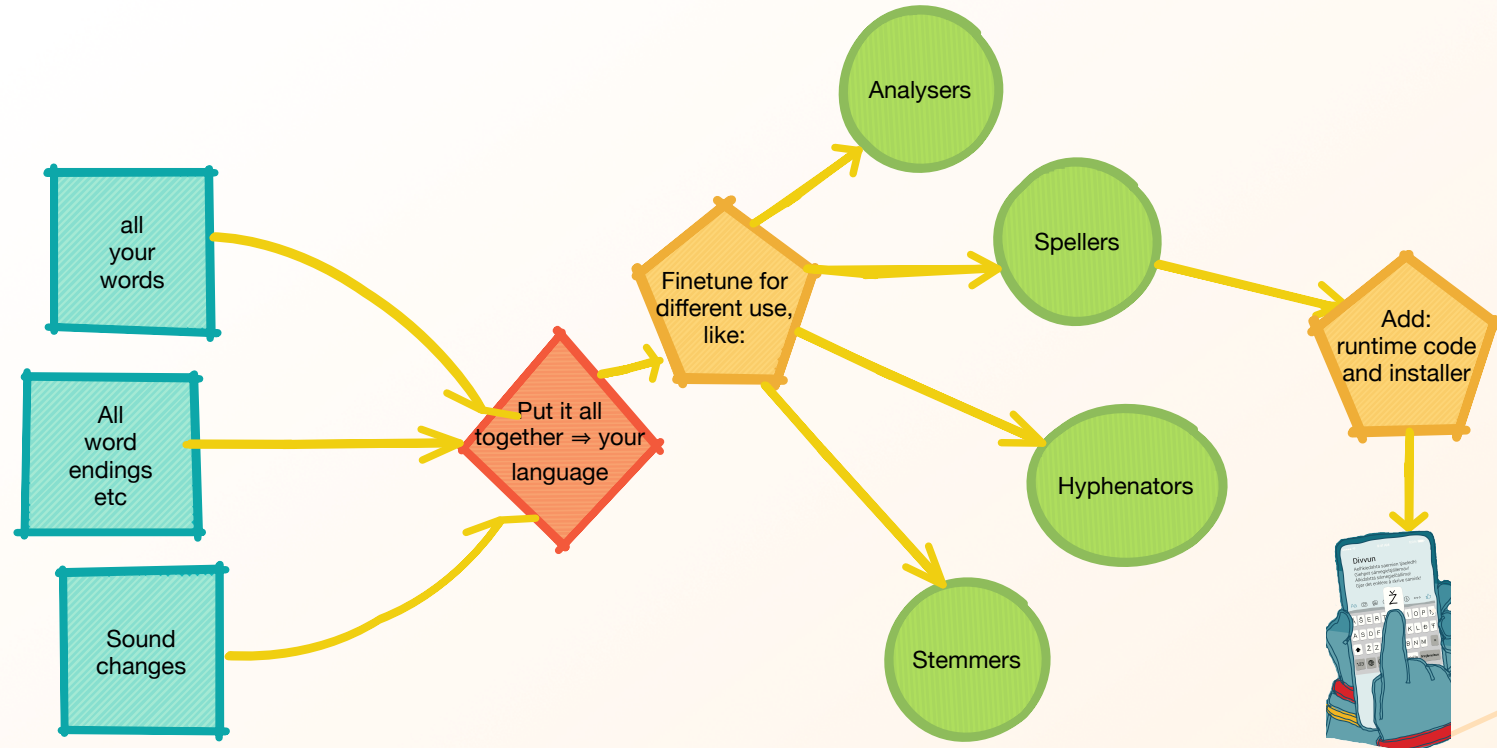


- ... on your phone:



Instant cell phone keyboard! Try hands-on after the talk!

How did we do it?



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

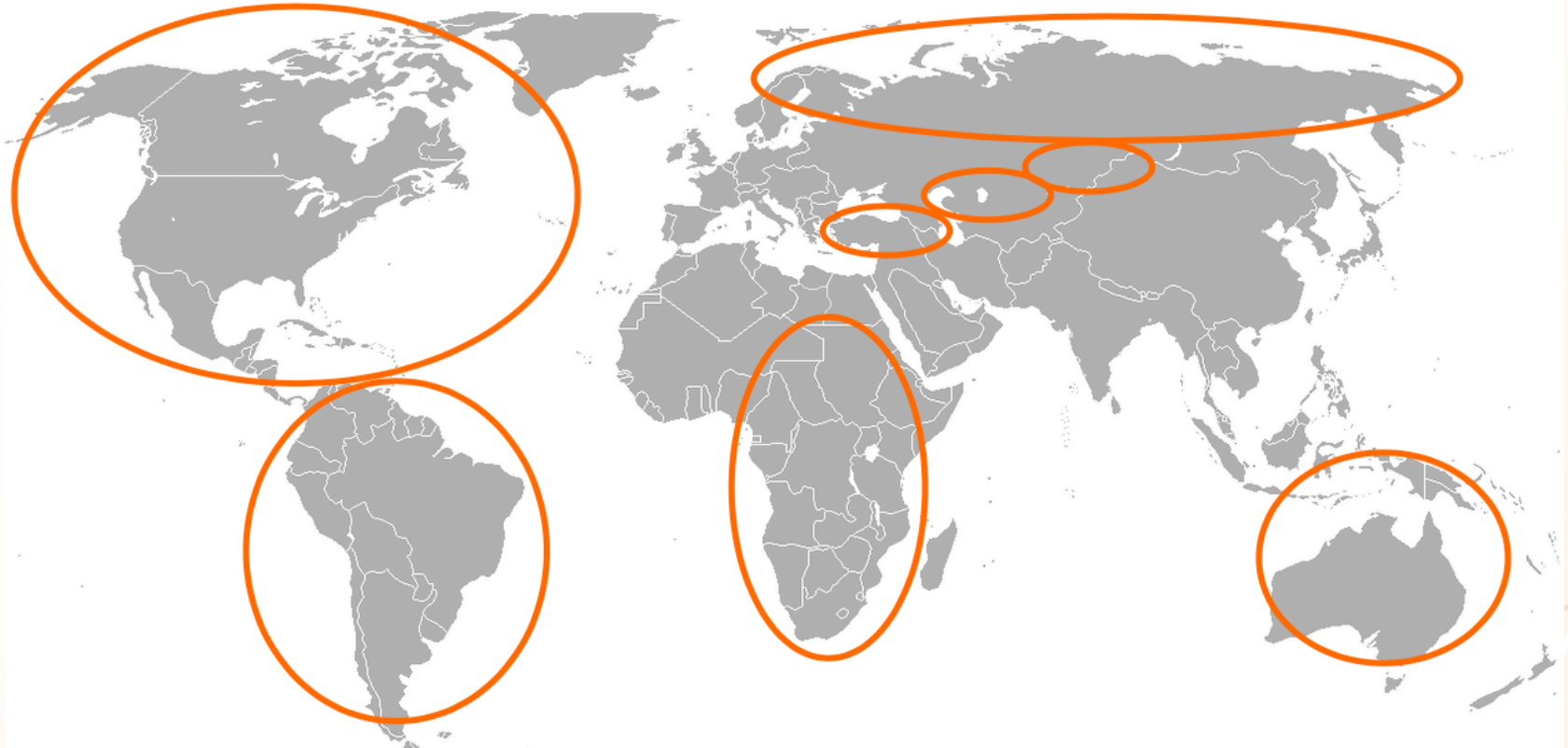
This holds for:

- *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** / Inuktitut
- *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



**Words
Language in use**

Grammar



***Computational model
of the language***

**Mobile: Keyboard with
proofing**

Proofing tools

Dictionary with grammar

Machine translation

E-learning

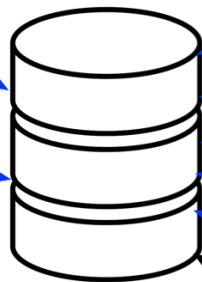
.....



**Words
Language in use**



Grammar



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

Mobile: Keyboard with proofing	<i>iPhone Android ...</i>
Proofing tools	<i>MS Office Libre O ...</i>
Dictionary with grammar	<i>E-dict- ionary</i>
Machine translation	<i>Apertium</i>
E-learning	<i>Oahpa View ...</i>
.....	...

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