

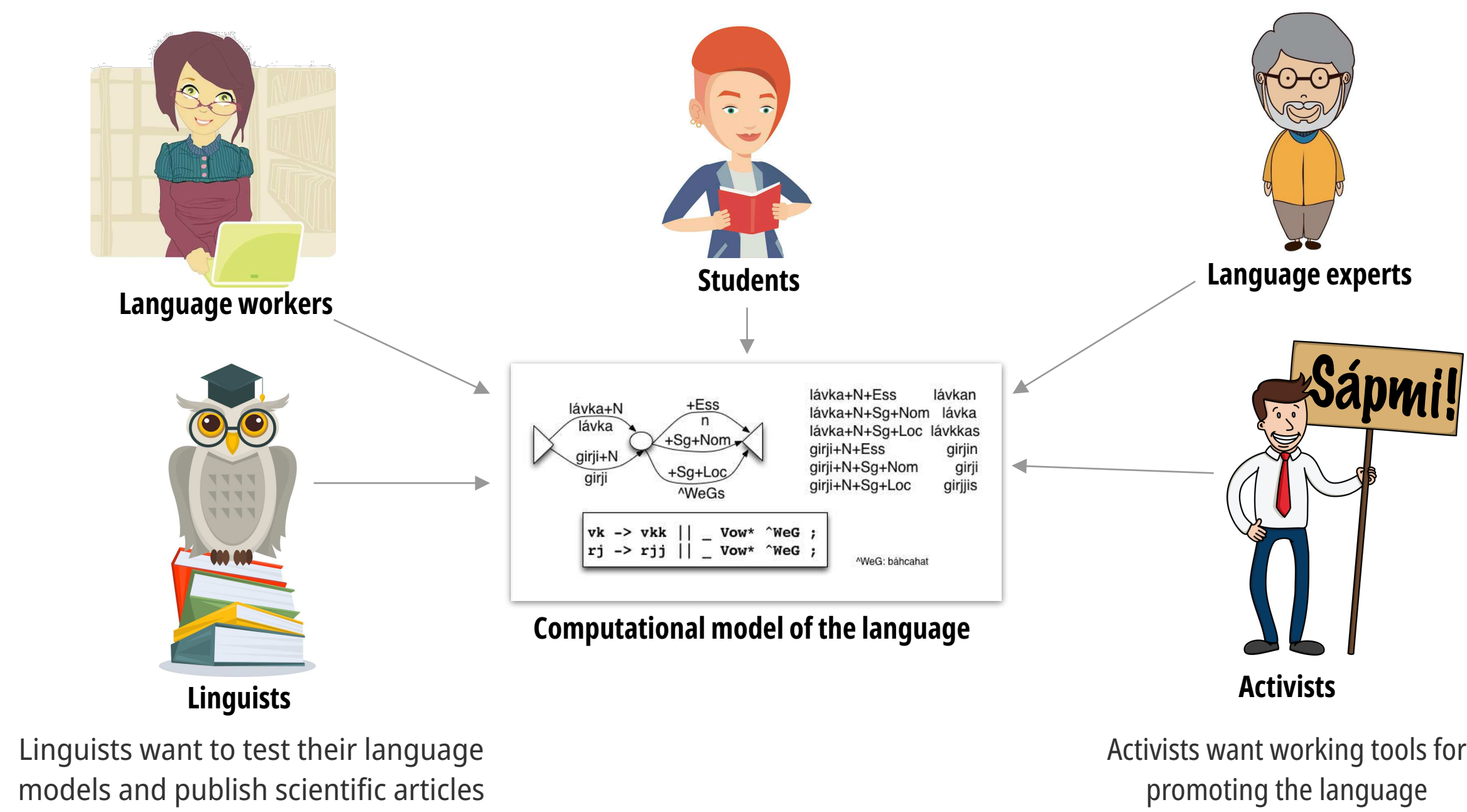
LANGUAGE TECHNOLOGY FOR INDIGENOUS LANGUAGES: ACHIEVEMENTS AND CHALLENGES

Sjur N. Moshagen, Trond Trosterud, Lene Antonsen and Børre Gaup
Sámi language technology at the University of Tromsø – Divvun and Giellatekno



We work together with communities to build language tools.

The language community builds and owns the computational model of the language, which can be used for different purposes. This is how our stakeholders cooperate:



Challenges: Technical blockers

We have achieved a lot, but there's still quite a way to go due to limitations imposed by the dominant vendors.

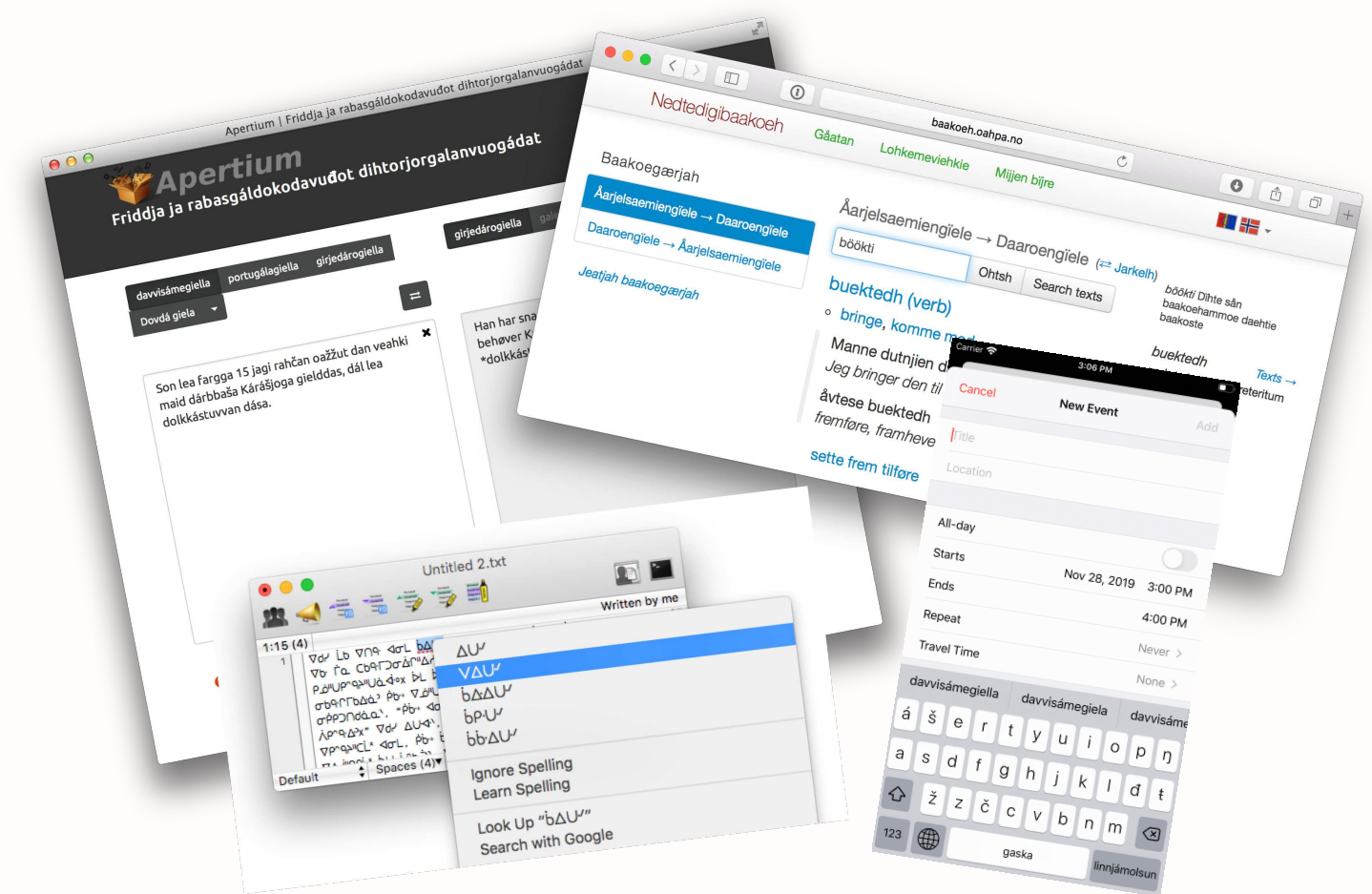
- **Localisation** of most operating systems is impossible, software localisation may also be blocked
- Frameworks for adding **dictionary content** to the system are not available to third parties
- **Spell checker integration** is blocked in Chrome OS, Google Docs and Microsoft Office 365
- **Combining diacritics** are often **unreadable**
- All major operating systems (save for macOS) are closed for **speech services**
- Likewise, **speech assistants** are just as closed



Today, minority languages are locked out by major companies

ACHIEVEMENTS: TOOLS AND RESOURCES MADE FOR SAAMI BY DIVVUN/GIELLATEKNO

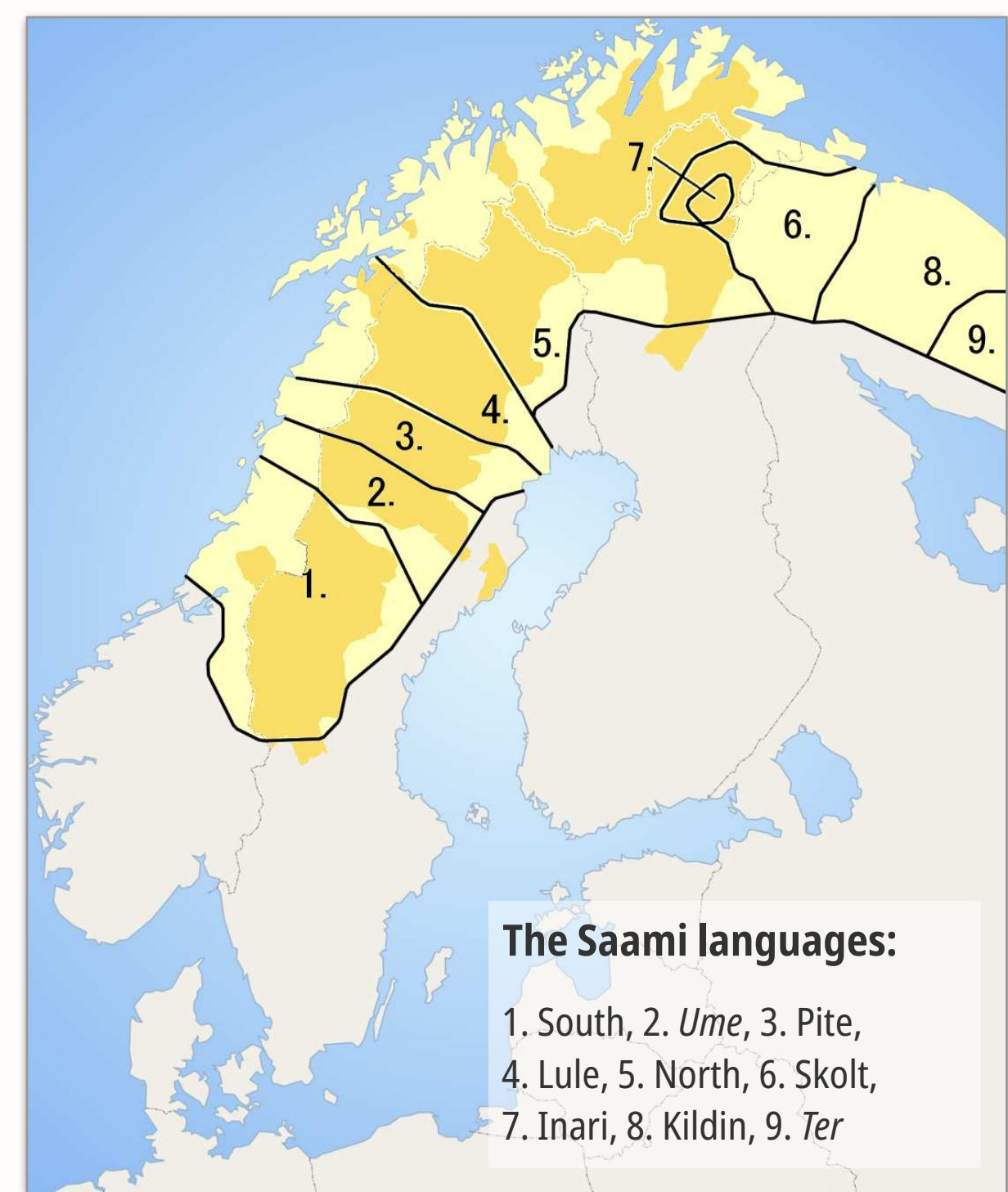
- Spell checkers
- Intelligent Computed-Assisted Language Learning (ICALL) programs
- Text-to-speech tools
- Corpus of analysed texts
- Dictionaries with grammar
- Rule-based machine translation
- Mobile phone keyboards with spell checker (iOS and Android)
- Computer keyboards, with locale registration (macOS, Windows, Linux)



Nearly all our tools are language independent and open source!
Find out more at indigenous-langtech.uit.no



THE SAAMI LANGUAGES AT A GLANCE



The Saami languages have both rich suffixation, stem-alternating processes and dynamic compounding. The languages have small monolingual text corpora and even smaller parallel ones (Finnish, Norwegian).
The 25 000 speakers of North Saami are 90% of all Saami speakers.

Solution: a Manifesto for Open Language Technology

We propose that the following principles be upheld by all vendors, producers and distributors in the software and digital services industries:

- **Open localisation:** all software should be localisable independently of the producer of the software.
- **Open interfaces:** all language-related programming interfaces should be open by default.
- **Open resources:** all language resources should be open and accessible for everyone, given the permission of the language community.
- **Accessible standards:** language-related international standards should be respected, fully implemented and implementations should be updated regularly.

We believe adherence to the above principles would lead to a better world for indigenous and minority languages, and agree that it is of utmost importance to push ahead with enabling indigenous self-determination in the digital realm.

IT'S TIME FOR INDIGENOUS SELF-DETERMINATION IN THE DIGITAL REALM!

SIGN THE MANIFESTO FOR OPEN LANGUAGE TECHNOLOGY AT: OPENLT.ORG

