

Proofing tools as the key factor for Faroese and Greenlandic linguistic sovereignty

Trond Trosterud

Department of Language and Culture, UiT The Arctic University of Norway

At a glance

- 1 In the current geopolitical situation, both Faroese and (especially) Greenlandic will play a more important role
- 2 At the moment, the proofing tools and linguistic infrastructure for Faroese and Greenlandic are not ready to meet the new challenges
- 3 An increased effort is needed in order to improve the situation

The Faroese spellchecker *Rættstavarin*

- Launched in 2019 as a part of the GiellaLT infrastructure
- Kept technically up-to-date since launched
- Based upon the vocabulary in Jóhan Henrik W. Poulsen et. al. *Foroyisk orðabók* (1998)

Challenges facing *Rættstavarin*:

- The dictionary was made to describe the traditional Faroese vocabulary, not for covering modern Faroese texts.
- Before launching the vocabulary was widened to include modern words, but since then it has not been updated.
- The spellchecker is updated technically but its content is not, making it vulnerable both to changes in the vocabulary and to new normative decisions.

The Greenlandic spellchecker *Kukkuniiaat*

- Launched in 2004, language model in the GiellaLT infrastructure, independent speller
- lexicon continuously updated by language secretariat in Nuuk

Challenges facing *Kukkuniiaat*:

- Technically, the speller has not been updated to be integrated in Microsoft Word (It no longer shows red lines under errors in Microsoft Word, but must be applied via a remote server)
- The grammatical component needs a thorough revision, to prevent overgeneration and make it easier to add new words.

Spellcheckers in society

During the last decades, Faroese and Greenlandic have strengthened their position online. Governmental whitepapers, university webpages are either translated from Danish or written in Faroese or Greenlandic from the outset.

- The present political situation will only strengthen this tendency.

Proofing tools in the GiellaLT infrastructure

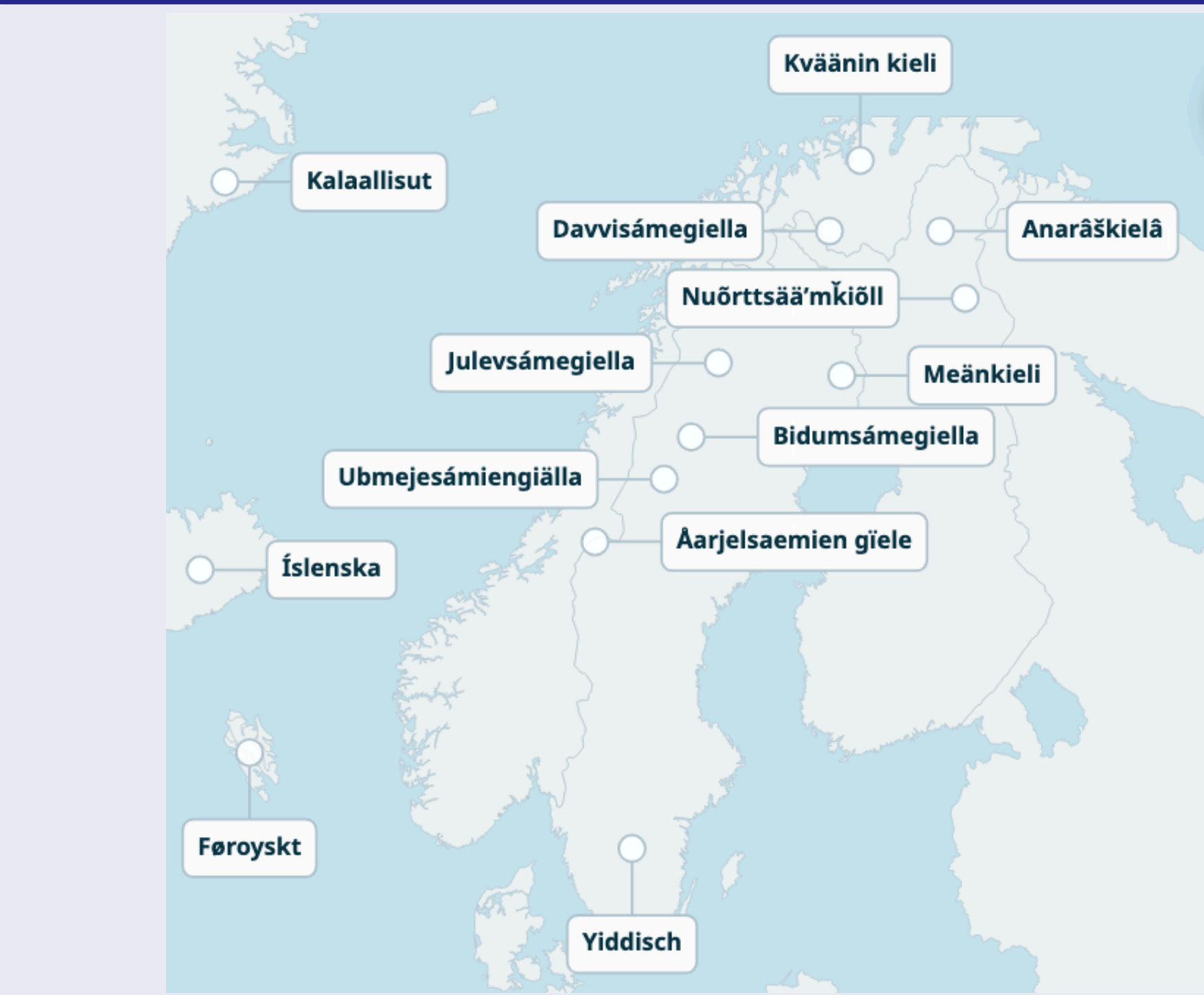


Figure: GiellaLT proofing tools for Nordic languages, <https://borealum.org>

Proofing tools for Faroese and Greenlandic

The Faroese spellchecker is integrated into the text, the Greenlandic one is a plug-in outside the document.

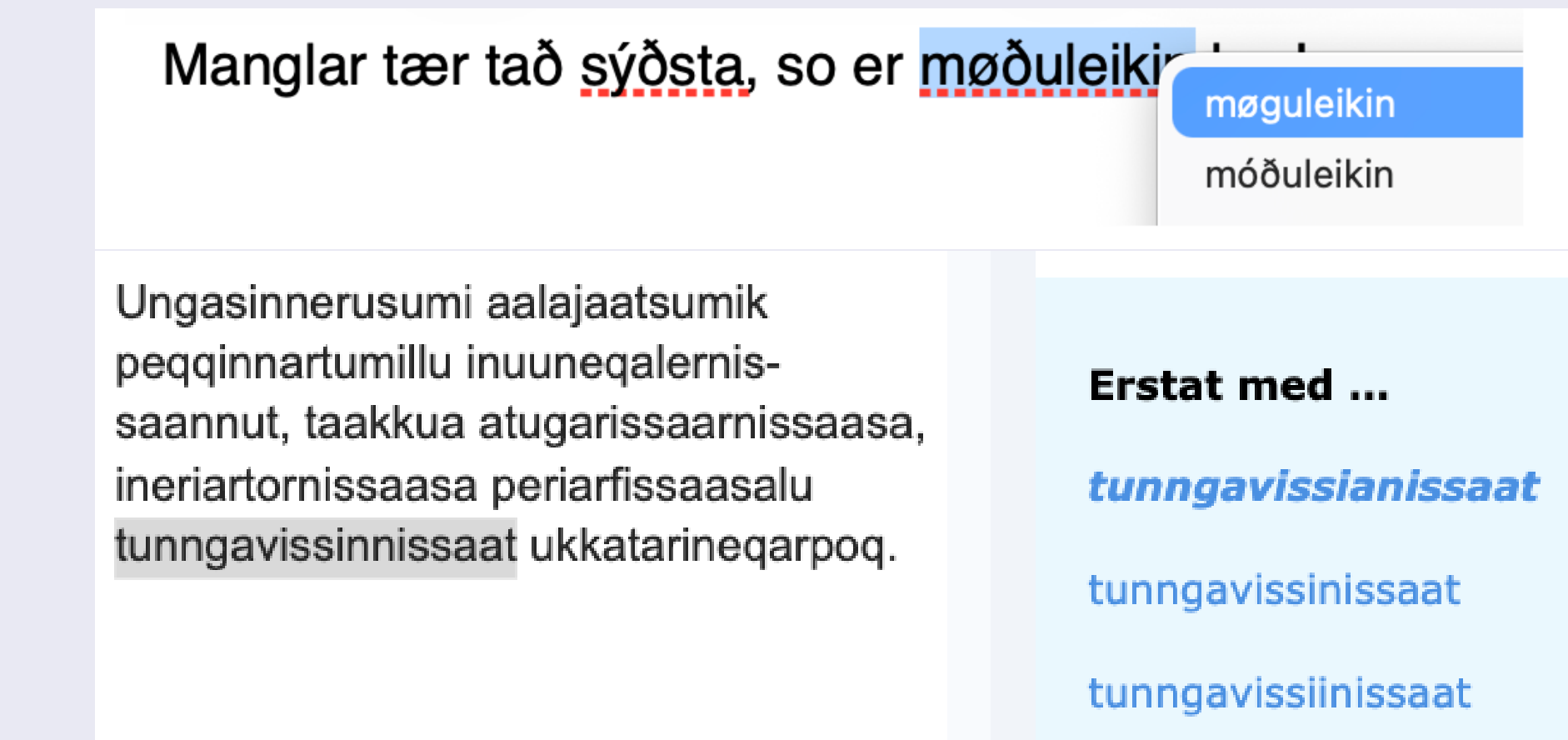


Figure: Spellcheckers for Faroese and Greenlandic

Available corpus resources (2024)

Faroese and Greenlandic have relatively small amounts of text available, only 1-2 % of what is available for e.g. Icelandic.

	Corpus Estimated	
Faroese	65 M	400 M
Greenlandic	29 M	100 M
Icelandic	2,4000 M	12,000 M

Why not just leave it to AI?

Both Faroese and Greenlandic have a relatively short history as administrative languages. Central language planning issues, especially vocabulary issues, are not settled. Proofing tools based on existing corpora will face several problems:

- Available corpora are too small to cover the whole vocabulary
- There is no guarantee that correct forms will outnumber unwanted ones
- Corpus-based proofing will deprive the Faroese and Greenlandic language councils their normative function

The grammatical language models

The language models are manually written and give both a descriptive and a normative view of what the language is and what it should be. They contain the lemma stock for each language, as well as inflectional and derivational suffixes and phonological rules. North Sámi is shown as reference.

	Lemmas	Coverage
Faroese	106,000	93.3 %
Greenlandic	64,000	72.0 %
North Sámi	132,000	92.1 %

North Sámi and Greenlandic are morphologically rich languages.

The language models are used both to lemmatise and identify hitherto unknown words in the vocabulary and as a basis for the proofing tools.

Suggestion mechanisms

Despite problems with vocabulary and integration, the spellchecker mechanism itself is in place for both languages.

	Error corpus	Sugg Top-5	Sugg 1st
Faroese	335	96.7 %	85.1 %
Greenlandic	566	93.8 %	86.9 %
North Sámi	10,700	92.6 %	81.5 %

The spellcheckers are tested against a representative set of typographical errors. **Top-5** shows how often the intended form is given as one of five suggestions and **Sugg 1st** shows how often it is given as the top suggestion.

Take-home message

The Faroese and Greenlandic proofing tools should be continuously developed to make it possible to uphold a native literacy.

- The Faroese language secretariat needs increased funding to collect text collection, monitor the vocabulary and add it to the proofing tool
- The Greenlandic language secretariat (Oqaaseliriffik) needs increased funding to do a thorough revision of the language model
- In addition to continuous funding, a large effort is needed to bring spellers up to date.
- Possible funding could come from **national**, **Nordic** and **ircumpolar** sources.

Ensuring functional proofing tools should be seen as a central part of the national infrastructure and treated accordingly.

References

- Pirinen, Flammie A. and Sjur Nørstebø Moshagen. 2025. Divvunspell—Finite-State Spell-Checking and Correction on Modern Platforms. In *Proceedings of the 9th Workshop on Constraint Grammar and Finite State NLP*, pages 59–63, Tallinn, Estonia. University of Tartu Library.
- Pirinen, Flammie A., Sjur Moshagen and Katri Hiovain-Asikainen. 2023. GiellaLT — a stable infrastructure for Nordic minority languages and beyond. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 643–649, Tórshavn, Faroe Islands. University of Tartu Library.
- Steingrímsson, Steinþór, Iben Nyholm Debess, Kimmo Granqvist, Per Langgård and Trond Trosterud 2024. Language Technology for Less-Resourced Languages in the Nordics Current Developments and Collaborative Opportunities. Stjórnarráð Íslands

giellalt.github.io

