

A North Sami translator's mailing list seen as a key to minority language lexicography

Berit Merete Nystad Eskonsipo
Trond Trosterud
Centre for Sami Language Technology
<http://giellatekno.uit.no>



Contents

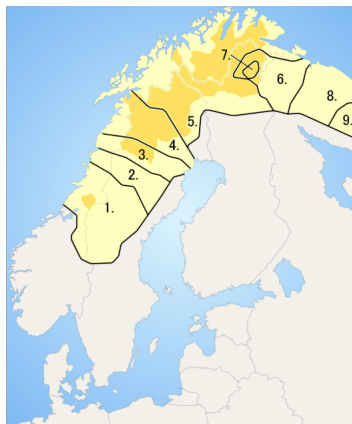
Introduction

- ▶ What is status quo for the North Sami modern vocabulary?
 - ▶ What concepts need established North Sami equivalents?
 - ▶ How are these concepts distributed over different text domains?

Introduction

- ▶ What is status quo for the North Sami modern vocabulary?
 - ▶ What concepts need established North Sami equivalents?
 - ▶ How are these concepts distributed over different text domains?
- ▶ To what extent can existing translation tools answer these needs?
 - ▶ Are the concepts found in dictionaries, term banks and corpora?
 - ▶ Do they get an adequate treatment?

The Sami language area



- 1. South Sami
- 2. Ume Sami
- 3. Pite Sami
- 4. Lule Sami
- 5. North Sami
- 6. Skolt Sami
- 7. Inari Sami
- 8. Kildin Sami
- 9. Ter Sami

Darkened area
represents
municipalities that
recognize Sami as
an official language.

Figure: The Sami language area – all together approx. 30,000 speakers

Milestones for the North Sami literary language

- ▶ Early period of North Sami literacy
 - ▶ Written language since the 18th century
 - ▶ Phonemic orthography 1832, New Testament 1840, Bible 1895

Milestones for the North Sami literary language

- ▶ Early period of North Sami literacy
 - ▶ Written language since the 18th century
 - ▶ Phonemic orthography 1832, New Testament 1840, Bible 1895
- ▶ Assimilation policy during late 19th and most of 20th century
 - ▶ Scientific work, especially in Finland
 - ▶ Common North Sami orthography for Norway & Sweden 1948, in little use due to the the prevailing language policy

Milestones for the North Sami literary language

- ▶ Early period of North Sami literacy
 - ▶ Written language since the 18th century
 - ▶ Phonemic orthography 1832, New Testament 1840, Bible 1895
- ▶ Assimilation policy during late 19th and most of 20th century
 - ▶ Scientific work, especially in Finland
 - ▶ Common North Sami orthography for Norway & Sweden 1948, in little use due to the the prevailing language policy
- ▶ A new language policy during the last 3 decades
 - ▶ New, unified orthography for North Sami (including Finland) from 1979
 - ▶ North Sami introduced in most areas of modern society

North Sami in the school system

- ▶ Taught in the primary school
 - ▶ approx. 1000 pupils a year have North Sami as their first language (Slaastad 2012).

North Sami in the school system

- ▶ Taught in the primary school
 - ▶ approx. 1000 pupils a year have North Sami as their first language (Slaastad 2012).
- ▶ Not all school subjects have access to teaching material in North Sami.
 - ▶ School subjects lacking teaching material: *social sciences, North Sami, natural sciences, religion and mathematics*

North Sami in the school system

- ▶ Taught in the primary school
 - ▶ approx. 1000 pupils a year have North Sami as their first language (Slaastad 2012).
- ▶ Not all school subjects have access to teaching material in North Sami.
 - ▶ School subjects lacking teaching material: *social sciences, North Sami, natural sciences, religion and mathematics*
- ▶ Pupils in higher secondary school:
 - ▶ In the core areas, 96% (Kautokeino) and 76% (Karasjok) of the pupils at higher secondary school had North Sami as first language or higher-level second language

North Sami media

- ▶ Radio / TV
 - ▶ *NRK Sápmi* radio five hours a day,
 - ▶ and a local radio station in Kautokeino.
 - ▶ *NRK Sápmi* 15 min TV news daily (with dubbing rather than subtitles)
- ▶ *Ávvir*, a daily newspaper
- ▶ Several quarterly magazines

The quadrolingual central government

- ▶ *<http://ráđđehus.no>*
(<http://www.regjeringen.no/se.html?id=4>)
- ▶ <http://gonagasviessu.no>

A North Sami translator's mailing list seen as a key to minority language lexicography

└ Background

└ North Sami literacy

Ráddehus.no – regjeringen.no

http://www.regjeringen.no/se.html?id=4

Bokmål | Nynorsk | Sámegeallat | English

ráddehus.no Stolttenberga nubbi ráddehus Departemeanttat

Fáddá A-Å Neahttakarta Hjeip Oktavuohta

Oza oles ráddehusa.no:es

Oza

Čámmhápmi a a a

Norgga ekonomija ovdána bures

- Norgga ekonomija lea ceavzán bures ruhtadanroasu váttisvuodaid márgg. Árvorshkadeapmi lea dál lessánan ovcci jahkenjealljásasa márggálapaid. Barggolašvuotta lea buoret go 2008 buoremus áiggi, ja bargguhisvuotta bissu vuollegaš dásis. Ráddehusa reviderejuvvon bušehta árvalusaln mii láhčit dilálašvuodaid dasa, ahte norgga ekonomija ovdána bures ain viidáseappot, ruhtadanministtar Sigbjørn Johnsen dadjala. **(Ruhtadandepartemeanta, 15.5.2012)**

Vearo- ja divatrivevdasut >

Elektrovnnalaš vearoakoarttat >

EDAG álkidahtá ealáhusaid barggu >

Norga áigu dahkat loatnesoahpamuša IMF:in >

2012 dárrkistuvvon nátiálaabušehta

Galledii Sámedikkii

Sáhtadállii Eii Blakstad galledii otna Sámediggeráđii Vibeke Larsena. Čoaikkima fáddán lei plánejuvvon eálfápmoinnjá gaskal Ofovuotta, Báhccavuona ja

Ráddehuskvartála boahhteáigi

Ráddehus lea mearridan čohkket departemeanttaid lanka ja birra ošá ráddehuskvartála. Mo dáinna áššiin gaigá bargat viidáseappot, čieigá márggii go ráddehus lea

Ráddehusa bargu

Ráddehus-sátni geavahuvo bealválaččat guovtti áđejuimis; oles ráddehus čoaikkis Gonagasa Jodiheami vuodie stáhtaráđdečoahkkimis, ja oles ráddehus

The giella list

- ▶ <https://list.uit.no/sympa/info/giella>
 - ▶ approx. 100 members
 - ▶ Translators and philologists from Norway, Finland and Sweden
- ▶ Issues related to North Sami translation
 - ▶ The prototypical question is *What is X in North Sami?*

Our material

- ▶ Giella list discussions
- ▶ March 2011 to February 2012
- ▶ Norwegian words (N=313)

Procedure

1. Check coverage and treatment of the candidate words in these sources:
 - ▶ largest Norwegian-North Sami dictionary (*NSO*, Kåven et al. 2000)
 - ▶ normative Sami Parliament term base (*Risten*, <http://risten.no>)
 - ▶ Tromsø Norwegian-North Sami corpus (*TC*, <http://giellatekno.uit.no/text.en.html>)
2. See which text domains the words belong to
3. Evaluate the suggestions on the Giella list

Coverage

The coverage of the material in three sources: NSO, Risten and TC, in absolute values and in percentage of total for each category (N)

Table: Coverage in dictionary, termbank and corpus

Words	NSO		Risten		TC		NSO / Risten		All	
All (N=313)	101	32%	62	20%	167	53%	128	41%	204	65%
Compounds only (n=219)	51	23%	28	13%	89	41%	70	32%	120	55%
Simplex only (n=94)	50	53%	34	36%	78	83%	58	62%	84	89%

Common words not in dictionary

- ▶ 212 of 313 words were not found in NSO
 - ▶ 27 of these words were among the 10000 most common words in Norwegian (source: NoWaC corpus).
 - ▶ *bane* n. 'field', *streng* a. 'strict', *tilbakemelding* 'feedback', *bygg* 'building', *innspill* 'suggestion', *henvendelse* 'inquiry', *satsing* 'prioritizing', *nestleder* 'vice president', *bærbar* 'portable', *høydepunkt* 'climax'.

The text domains

Table: The material distributed according to text domains

Admin	Education	Health	Sports	Other	General
82 26%	70 22%	34 11%	27 9%	54 21%	46 18%
valgportal 'election website'	lærestoff 'syllabus'	stivkrampe 'tetanus'	midtbane 'center field'	musling 'molluscs'	innspill 'suggestion'

The text domains

- ▶ The domains discussed on the list reflect the priorities reported by translators:
 - ▶ Administrative texts for official bodies
 - ▶ School textbooks and scientific text
 - ▶ Bilingual signs on hospitals and health information aimed at the public

The text domains

- ▶ The domains discussed on the list reflect the priorities reported by translators:
 - ▶ Administrative texts for official bodies
 - ▶ School textbooks and scientific text
 - ▶ Bilingual signs on hospitals and health information aimed at the public
- ▶ Other parts of modern vocabulary are not subject to vocabulary planning to the same extent
- ▶ ... and the result is lexical vacillation within the norm

The shopping mall effect

- ▶ **kaffetrakter** (drip coffeemaker)
- ▶ How come we know this term in Norwegian?
 - ▶ When we buy them, the store refers to e.g.
 - ▶ *Bosch kaffetrakter TKA1411V*
 - ▶ Drip coffeemakers seldom occur in texts

The shopping mall effect

- ▶ **kaffetrakter** (drip coffeemaker)
- ▶ How come we know this term in Norwegian?
 - ▶ When we buy them, the store refers to e.g.
 - ▶ *Bosch kaffetrakter TKA1411V*
 - ▶ Drip coffeemakers seldom occur in texts
- ▶ There is no common arena to learn the term in North Sami
 - ▶ People either use the Norwegian term
 - ▶ ... or invent private neologisms, possibly varying from family to family

The shopping mall effect

- ▶ **kaffetrakter** (drip coffeemaker)
- ▶ How come we know this term in Norwegian?
 - ▶ When we buy them, the store refers to e.g.
 - ▶ *Bosch kaffetrakter TKA1411V*
 - ▶ Drip coffeemakers seldom occur in texts
- ▶ There is no common arena to learn the term in North Sami
 - ▶ People either use the Norwegian term
 - ▶ ... or invent private neologisms, possibly varying from family to family
- ▶ Written North Sami avoids such direct loans, and is thus in need of a generally accepted term

List discussions

Example: A spreadsheet term for *rad* 'row'

- > Hei,
- > Movt dii dáid lávebehtet gohčodit:
- >
- > Regneark **lean gohčodan rehkenastinárkan**
- > Arbeidsbok – lean dál bidjan vuos bargogirjin, muhto gávdno go eará
- > sátni?
- > Celle **Lean matematihkkagirjái aiddo bidjan "latnja"**
- > Rad
- > Kolonne **Lean matematihkkagirjjis aiddo bidjan "ráidu" (doppe in dárbbahan "rad" jorgalit...)**
- > Formel **evttohan: hápmu**
- > Fane **evttohan: leavga**
- > Funksjon : **evttohan doaibma**

The dictionaries cannot help

- ▶ NSO:
 - ▶ **rad** s. *gurgadus*, *-ss-:s*, *ráidu*, *-idd-* (rekke, linje, kolonne, kjede, serie)

- └ Analysis
- └ List discussions

... neither can *risten.no*

Search

Help

Search for:

Search method:

Search field:

Dictionary:

Search the following language(s):

Show hits

Hits

The term db of Sámediggi

Hits 1–2 of 2

nor **rad** FRITID OG SPORT

swe **rad** FRITID OG SPORT

The search took 0.25 second(s).

Term Entry

CLASSIFICATION:

FRITID OG SPORT
IDRETT OG FRITID
GENERELT

SWEDISH

rad – *S*
SYNONYMS
serie

NORTH SAMÍ

gurgadus – *S, IVm*
INFLECTS LIKE: gurgadus – gurgadussii – gurgadusaide

NORWEGIAN

rad – *S*

FINNISH

rivi – *S*

... nor our parallel corpus TC

- ▶ *rad*: No hits
- ▶ *kolonne*: 5 hits
 - ▶ *kolonne - ruktáráidu*
 - ▶ Forskriften §§ 36 , 39 : Opplysninger om hvordan velgeren kan endre rekkefølgen av kandidatene , samt en **kolonne** med ruter til endring av rekkefølgen , skal være trykket på stemmesedlene
 - ▶ Láhkaásahusa §§ 36 , 39 Dieđut das mo jienasteaddji sáhhtá nuppástuhttit eavttohasaid vuoruid , ja vel **ruktáráidu** vuoruid nuppástuhttima várás , galget leat prentejuvvon jienastanlihpuide .
 - ▶ *kolonne - kolonna*
 - ▶ Endringene framkommer under hvert kapittel i egen **kolonne** :
 - ▶ Rievdadusat bohtet ovdan sierra **kolonnas** kapihttaliid buohta :

... but the giella list has suggestions

Dát leamaS anus: Cuolda ja gurgadas.

Beahkká

Quoting Else Turi <else.turi@nrk.no>:

Evttohus:

Ceakkoráidu (kolonne) (vrd. Gávttis leat dilit ja ceakkodilit)

Ja

Ráidu (rad)

Else

... but the giella list has suggestions

Dát leamaS anus: Cuolda ja gurgadas.

Beahkká

Quoting Else Turi <else.turi@nrk.no>:

Evttohus:

Ceakkoráidu (kolonne) (vrd. Gávttis leat dilit ja ceakkodilit)

Ja

Ráidu (rad)

Else

(unfortunately it has more than one)

List suggestions to the queries

- ▶ 158 of 185 words not found in NSO or Risten got list suggestions
- ▶ Clear answers:
 - ▶ kontaktinformasjon 'contact information': *oktavuođadieđut*
- ▶ Different suggestions:
 - ▶ elveforbygning 'river wall': *johkagáddeduvdda*, *mielleduvdda*, *miellefierrancaggi* > *johkaduvdda*

List acceptance

- ▶ Previously suggested equivalents not accepted by linguistic experts on the list:
 - ▶ bekymringsmelding 'note of concern': *fuolastusdiehtu* > *fuolastuvvandieđáhus*
 - ▶ kaffetrakter 'drip coffeemaker': *gáfesilon* > *gáfesillehas*

The role of the giella list

- ▶ Fill the gap between the dictionaries and the reality
- ▶ Get confirmation from other North Sami speakers
- ▶ Improve the quality of the resulting terminology by ensuring that it was morphologically well formed.

Conclusion 1

- ▶ The *questions* to the giella list reveal lacunas in the North Sami vocabulary
- ▶ The *answers* picture a lack of consensus even for core parts of the vocabulary

Conclusion 1

- ▶ The *questions* to the giella list reveal lacunas in the North Sami vocabulary
- ▶ The *answers* picture a lack of consensus even for core parts of the vocabulary
- ▶ ... but the discussion also shows that minority language experts are able to build such a consensus in an efficient way

Conclusion 2

- ▶ We see the giella list as an indicator of the state of the art
- ▶ Central parts of the vocabulary is missing from key lexicographical resources, especially
 - ▶ everyday language not found in texts
 - ▶ new concepts
 - ▶ concepts linked to the modern society
- ▶ The most important text domains for professional translators are *administrative texts*, *education* and *health*, in that order

Conclusion 3

- ▶ Norwegian - Sami lexicography must take the Norwegian vocabulary into account to a greater extent than before
- ▶ Norwegian - Sami dictionaries must take Norwegian as a starting point and ensure that the basic vocabulary is covered

Thanks to our colleagues at Giellatekno for fruitful discussions

Thanks to our colleagues at Giellatekno for fruitful discussions
Thank you for listening.
Any questions?