

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

Trond Trosterud & Berit Nystad Eskonsipo

Keywords: *North Sami, minority language, language planning, vocabulary planning.*

Abstract

The topic of this investigation is the set of Norwegian words discussed on a North Sami translator's mailing list during one year, altogether 313 words. The words were grouped according to text domain, and to what extent existing dictionaries were able to meet the translators' needs. Most of the words discussed on the list were missing in relevant reference works. Two reasons for this are the paucity of North Sami text and the fact that Norwegian to North Sami lexicography has had North Sami dictionaries and word lists as their basis. The main finding of the article is that the words put under scrutiny by the mailing list belong to common, everyday language. The translator list thus may function as a roadmap for future North Sami lexicography.

1. Introduction

North Sami is a Finno-Ugric minority language spoken by between 25000 and 30000 indigenous Sami in Norway, Sweden and Finland (Sami statistics), where it has official status in seven, three and four municipalities, respectively. North Sami translators have for the past four years upheld a mailing list, giella@list.uit.no (from now on *the giella list*), for discussing issues related to North Sami translation, mostly into North Sami. The goal of this paper is to use the *giella list* as indicator of the status and further needs of the North Sami vocabulary, and thereby also lexicography. Lexicographical and terminological work does not take place in isolation, but is influenced by the linguistic ecology of the language in question. The article thus starts out with an assessment of status quo for the North Sami literacy. Section 3 presents the *giella list* discussions, and sections 4 through 6 the analysis. Finally comes a conclusion. The main focus on the *giella list* is on the situation in Norway, and our paper will have the same focus.

2. Background: North Sami literacy

The use of North Sami as a written language dates back to the 18th century, but due to 19th and especially 20th century assimilation policy the literary language was mainly restricted to the religious domain. The orthographic reform of 1979 marked a clear break in the literary tradition, and Sami born before the mid-seventies have not learned to read and write their mother tongue in school.

During the last 30 years, there has been a change in policy, and North Sami is now promoted to a certain extent. It is taught in the primary school, and 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 with an urgent need for teaching material include social sciences, North Sami, natural sciences, religion and mathematics (Hegg et al. 2008). In the North Sami core areas in Norway's Inner Finnmark, primary and secondary school is conducted in North Sami. In Karasjok, 76% of the pupils at higher secondary school had North Sami as first language or higher-level second language, the corresponding number for Kautokeino was 96%, reflecting the position of North Sami in these two municipalities

(Slaastad op.cit. table 40). North Sami is also heard in the radio channel *NRK Sápmi* every morning and afternoon five hours a day, in addition to digital radio and a local radio station in Kautokeino.

The amount of North Sami text visible to potential readers is far smaller. The main source of North Sami text is the daily 16-page newspaper *Ávvir*. In 2011, *Ávvir* had a circulation of 1088 copies (Mediebedriftenes landsforening). In percent of the North Sami speech community this corresponds approximately to the circulation of the leading Norwegian daily *Aftenposten*, but since *Ávvir* is the only North Sami newspaper it follows that the impact of North Sami newspaper text in the North Sami society as a whole is in fact quite limited. The media situation for Sami can thus be illustrated by comparing it to a situation where Norwegians would have access to only one newspaper in Norwegian, with the circulation of *Aftenposten* but only five 16-page editions a week. The Norwegian Sami broadcasting channel NRK Sápmi mostly avoids written North Sami, 92.5% of its web service is in Norwegian, and its daily 15 minute television news broadcast has Norwegian subtitles but North Sami dubbing. North Sami has a vivid national literature, but the number of titles is not large. The Norwegian library database *bibsys.no* reveals 654 works within the Dewey class 800 (Literature), including re-editions, since 1979, and 192 works the last decade (2002-2011). These numbers add up to two nice piles of books, but when we keep in mind that the corresponding Norwegian numbers are 65660 and 29341, respectively, we see that it becomes hard to even compare the Norwegian and Sami literary landscapes. Official information is available in North Sami to a variable extent. Municipality web pages in the North Sami administrative area are in principle bilingual, but in practice the Norwegian pages are more up to date. In Kautokeino, approximately half of the official forms to fill in by citizens were available in North Sami and Norwegian; the remaining ones were available only in Norwegian. The number of governmental official web pages is increasing (a recent harvest from Norway's official web portal *government.no* gave 2.5 million words in North Sami), but it is still nowhere near representing a functional alternative to the information available in Norwegian. Web pages with administrative information and fill-in forms are still only marginally available in North Sami, although the amount is increasing. In the religious domain North Sami text has kept its strong position. The old 19th century Bible has been replaced with a still not completed modern edition, and church services are conducted in North Sami, or with interpreters.

The net outcome of this is that there is a discrepancy between North Sami speech and North Sami writing. In the core areas, North Sami is used as an everyday oral language in private life and most work settings (work communities without monolingual Norwegian speakers). Despite its 200-year literary tradition North Sami still faces a marginal position as a written language, though. The amount of North Sami text available to the average reader is limited and concentrated to relatively few domains.

3. The giella list and the material under investigation

The giella list is a mailing list for discussing issues related to North Sami translation, mostly into North Sami. The members are North Sami translators and philologists, mostly from Norway, but also from Finland and Sweden. The list contains discussions on neologisms, lexicalized compounds, most of them covering common concepts both for Norwegian and North Sami. The most common question on the giella list is *What is X in North Sami?*

We read through one year of the giella list (March 2011 to February 2012), and collected all Norwegian words presented to the list, 264 words and 49 phrases. For each

phrase we picked the most salient word, mostly the headword, so altogether we investigated 313 words. In the period under investigation, there were posted approximately 2100 posts, on the average 5 posts a day, and 5.3 posts for each query (including other languages than Norwegian).

Common to the Norwegian words discussed on the giella list was that the participants had not found adequate North Sami equivalents in the existing dictionaries, and hence saw it fit to turn to the common community for help. The 313 words, including the phrasal headwords, were distributed according to part of speech as follows: 240 nouns, 29 proper nouns, 31 adjectives, 12 verbs and 1 conjunction.

The verbs were either parts of phrases, like *slå på radioen* 'turn on the radio' or compounds like *skumlese* 'skim through', literally 'skim read'. Several of the words were confined to specific domains, like *kvalifikasjonsrammeverk* 'qualification framework' (education) and *intimkonsert* 'small concert' (culture), whereas other belonged to everyday vocabulary, like *respektfull* 'respectful', *høydepunkt* 'peak'. Specialized terms, or words with a meaning not known to a general Norwegian audience, were not discussed on the list.

We then checked whether the 313 words under investigation were listed in the largest Norwegian-North Sami dictionary (from now on *NSO*) (Kåven et al. 2000), in the normative Sami Parliament term base *risten.no* (from now on *Risten*), and whether they already had been translated to North Sami in the Tromsø Norwegian-North Sami corpus (from now on *TC*), a sentence-aligned parallel corpus of non-fiction compiled at the University of Tromsø, containing 1.6 million North Sami and 2.1 million Norwegian Bokmål words. We then looked at whether the lemma articles in NSO were able to answer the original list query. The words were classified according to textual domain, in order to see what part of the vocabulary caused problems to the translators. The textual domains covered are thus able to tell something about the written culture as well.

4. Coverage

The 313 words under investigation give a picture of what challenges North Sami translators face in their work. In order to see to what extent the existing lexicography, terminology work, and actual usage are able to address these challenges, we investigated how many of the words posted on giella list are found in NSO, in Risten, and in TC. We also checked how many of the compounds and how many of the simplexes are found in NSO, in Risten, and in TC. The results are presented in Table 1.

Table 1. The coverage of the material in three sources: NSO, Risten and TC, in absolute values and in percentage of total for each category (N).

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%

Table 1 shows the coverage of the investigated giella list words in three sources; dictionary, term base and corpus. As we see, the dictionary NSO covers 32 % of the 313 words posted on giella list. The term base covers 20 % of the posted words. The dictionary and term base together cover 41 % of the words posted on the list. The corpus has the highest coverage, it

covers 53 % of all the words posted on the list, even though it is quite limited both with respect to size (2.1 million words only) and genre (it represents text from official web pages and whitepapers). This is a remarkable result in itself, and it reveals the weaknesses of especially NSO. The corpus is still not available to users in a convenient interface, but it is evidently much needed.

If we look at the coverage in all three sources together, dictionary, term base and corpus, we see that they cover 65 % of all the words posted on the list. That means that 35 % of the words posted on the list are not found neither in the dictionary, term base nor corpus.

Lacunae in NSO are to be expected, since it was made by converting a North Sami-Norwegian dictionary, rather than taking the Norwegian lexicon as a starting point. Whatever not found in the North Sami primary material will at the outset lack in the NSO, although 5000 new words were added to NSO during the compilation process (Kåven et al. 2000: III).

NSO contains 25000 lemmata and Risten 14000, the difference in coverage for the words from the giella list is of approximately the same proportion, reflecting that both the giella list members and probably even the Risten terminologists have had a focus upon everyday language. Table 1 also shows that the simplex words are better represented in all three sources than compounds.

Many of the lacunae in NSO corresponded to quite common words. Of the 212 words posted on the giella list and not found in NSO, 27 of these words were among the 10000 most common words in Norwegian. The words from the giella list not found in NSO were measured against lemmatized frequency data from the Norwegian NoWaC corpus (NoWaC). Examples of common Norwegian words include *bane* n. 'field', *streng* a. 'strict', *tilbakemelding* 'feedback', *bygg* 'building', *innspill* 'suggestion', *henvendelse* 'inquiry', *satsing* 'prioritizing', *nestleder* 'vice president', *bærbar* 'portable', *høydepunkt* 'climax'.

A further reason for the dictionary lacunae might be the paucity of North Sami texts, and hence the lack of North Sami equivalent to the Norwegian words. These concepts are of course central for Sami as well. When there exist North Sami equivalents, they are known mainly from oral language rather than from written text, or there exists in any case no consensus as to what the correct translation is. In other cases, like for example *stivkrampe* 'tetanus' and *midtbane* 'center field', our impression is that North Sami speakers use direct loans in speech, but for the purpose of writing the Norwegian words are felt inappropriate. The net outcome of this is that many North Sami ad hoc neologisms are in use and they are differing from area to area, or even family to family.

Of the 101 words discussed on the list and found in the dictionary, the most common words were *mål* 'goal', *tiltak* 'action', *aktivitet* 'activity', *funksjon* 'function', *rad* 'row', *spiss* 'odd', *eiendom* 'real estate' (measured against NoWaC). For most of these words, the dictionary left semantic distinctions open. One case in point is the Norwegian word *rad* 'row', which was given in the context of computer spreadsheet localization. NSO offered the following article:

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

As can be seen from the Norwegian words in parentheses, the dictionary user was not able to find North Sami words differentiating between *rad* 'row' on the one side, and *kolonne* 'column' on the other. The discussion on the list subsequently focused on the usage of North Sami *ráidu* within different domains (traditional handicraft, convoy driving, mathematics, computer localization). Also Risten offered *gurgadus* as the North Sami equivalent of the Norwegian word *rad*, but it classified the word within the domain leisure/sport, and did not give any help on usage as a computer localization term, except for explicitly stating that it

was a sport term.

In some cases, the intended meaning of the word was lacking in NSO. An example of this is the Norwegian word *orientering*. The list posting made it clear that the intended meaning was 'survey', but the dictionary offered the sports term 'orienteering' only.

5. The text genres

We also classified the words discussed on the giella list according to the text genre within which the translators work. Table 2 gives an overview, the category *Other* contains minor genres, and the category *General* contains words hard to place in any specific domain.

Table 2. The material distributed according to text genre.

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

The distribution corresponds to the main priority reported by translators: Administrative texts loom large, as do also translation of school textbooks and scientific texts (included in the category *Education*). Bilingual signs on hospitals as well as health information aimed at the public explain the category *Health*. The words belong to clearly defined domains, and are established as terms within the respective domains, but they still form part of everyday language use and cannot be considered specialist language. The consequences for a language society not having a clear policy on terms like this is vacillation within the norm. When fluent speakers hesitate on how to translate these terms to their mother tongue it is in any case an indication of diglossia, or at least of literary diglossia.

6. The giella list suggestions

If we look at the giella list words that were not found in NSO or Risten, altogether 185 words, we see that the giella list suggested North Sami equivalents to 158 of these and only 27 Norwegian words did not get any suggested equivalents. For some questions the list actually gave clear answers, like for the Norwegian word *kontaktinformasjon*. This word got the North Sami equivalent *oktavuođadieđut*, from the genitive of the already established words *oktavuohta* 'contact' (literally 'one-ness') combined with a plural form of the noun *diehtu* 'information'. For other words, the person who originally posed the question was left to decide between the different suggestions given by the list. A case in point is the Norwegian word *elveforbygning* 'river wall', which got the following North Sami proposals: *johkagáddeduvdda* 'river shore support', *mielleduvdda* 'river bank support', *miellefierrancaggi* 'river bank erosion prevention'. The compounds were formed by different combinations of the North Sami nouns *johka* 'river', *gáddi* 'shore', *duvdda* 'support, fence', *mielli* 'river bank', *fierran* 'rotation', and *caggi* 'support, fence', with the first parts of the compound in nominative singular, eventually with the usual phonological reduction *i, u -> e, o* (*mielli -> mielle-*). The person who posted the question then decided to use a variety of the first proposal, *johkaduvdda*, because he found it used in texts. List recommendations and actual usage thus strengthen each other.

Some of the North Sami equivalent suggestions that were not found in NSO or Risten,

but had been previously suggested in other contexts, were not accepted by the linguistic experts on the list. A case in point is the North Sami *fuolastusdiehtu* for Norwegian *bekymringsmelding* 'note of concern', which was found on a list of terms from the social sector. The term was rejected by the giella list because of an error in the derivation pattern used. The first part of the compound, *fuolastus-*, is an action noun derived from a diminutive stem *fuolast-* of the base verb *fuollat*, which is a homonym for two verbs meaning 'care about' and 'whittle', respectively. A diminutive derivation is possible only for the second meaning of the word, and the term thus misleadingly gave the absurd meaning '*whittle note'. The term was thus rejected, and *fuolastuvvandiedáhus* was offered as a replacement. Here, the first part of the compound is the action noun *fuolastuvvan* from *fuolastuvvat* 'be concerned', a lexicalized inchoative derivation of *fuollat*.

An other controversial and previously used term suggestion was North Sami *gáfesilon* for the Norwegian word *kaffetrakter* 'coffee machine'. The problem turned out to be an error in the base word from which the noun *silon* was derived. Since *silon* has a short *l* it must be (erroneously) derived from the North Sami verb *sillut* 'become thin', and not from the intended verb *sil'let* 'to filter'. Instead of *silon* 'tool to become thin with', it was suggested to use *sillehas* 'a filter'. The role of the list was in these cases to improve the quality of the resulting terminology by ensuring that it was morphologically well formed.

According to Einar Haugen's model of language planning (Haugen 1987, cf. Table 3), the process can be depicted in a four-fold matrix along the dichotomies society vs. language and norm vs. function. The selection procedure is often done by formal bodies. In the Sami vocabulary context this is in principle done by official term committees and subsequently published in the Risten database. Since only 20% of the words discussed on the list were found in Risten, the formal process clearly is not comprehensive enough, and the selection is carried out by the translators themselves. Codification of the norm has mainly been done already, but in some cases the formal properties of the word-formation processes were cast into doubt, it was then carried out as a dialogue between translators and linguists on the list. Implementation and elaboration were conducted outside of the list, but, notably, by the same persons that participated in the discussion.

Table 3. Einar Haugen's model of language planning.

	Norm	Function
Society	(1) Selection	(3) Implementation
Language	(2) Codification	(4) Elaboration

Of the 212 words not found in NSO, 31 did not receive any suggestions at all. The largest group of neglected words (13) came from the health domain; all other text genres were small. One reason for this may be that the amount of text produced within this genre is small, compared to the other domains, so that fewer translators work with those terms.

7. Conclusion

The most baffling result of our investigation is that we reveal the need for asking about North Sami equivalents to very common Norwegian words, words among the 10000 most common words in Norwegian. The paucity of texts, and the narrow base of lexicographic resources, leaves lacunas in the North Sami vocabulary of modern words. Thus, the translators feel a need of confirmation from other North Sami speakers. For languages where a modern vocabulary has been both codified and implemented, such issues have already been settled.

The situation for North Sami may be characterized in the following way: To find a North Sami equivalent of a Norwegian word, be it in the dictionary, or in actual usage, is in itself not enough for the translators. Explicit suggestions from colleagues may be either misleading or paper constructs lacking acceptance by the speech community. Examples of actual usage, on the other hand, may lack formal approval, or even represent idiosyncratic use. What the translators need is a combination of the two: Hard evidence that the North Sami translation has actually been used, as well as an approval from the translator community (in this case: the giella list). For lexicography dealing with dictionaries from majority to minority languages the present investigation shows the importance of both normative and descriptive work: When in other speech communities terminological work is confined to specialist domains, minority language communities must do vocabulary planning also on the modern core vocabulary. In order to establish themselves, terms must both be formally approved and attested in real use.

Notes

¹ Source: Mediebedriftenes landsforening, <http://www.mediebedriftene.no/>

² Norwegian Web as a corpus, <http://www.tekstlab.uio.no/nowac/>.

³ Divvun: divvun.no, giellatekno: giellatekno.uit.no.

References

A. Dictionaries

Kåven, B., J. H. Eira, J. Jernsletten, I. Nordal and A. Solbakk 2000. *Stor norsk-samisk ordbok. Dáru-sámi sátnegirji*. Karasjok: Davvi Girji.

B. Other literature

Haugen, E. 1987. *Blessings of Babel. Bilingualism and Language Planning. Problems and Pleasures*. Berlin, New York, Amsterdam: Mouton de Gruyter.

Hegg, L., O. H. Brinchmann-Hansen, J. Aslaksen and O.-M. Kleven 2008. *Samiske læremidler. En kartlegging av eksisterende læremidler og behov for læremidler per oktober 2008*. Finnut Consult.

http://www.udir.no/Upload/laremidler/5/Sluttrapport_Finnut_endelig.pdf.

Mediebedriftenes landsforening. 16 February 2012.

<http://www.mediebedriftene.no/novus/upload/file/Opplagstall/Avisopplag%202011.xl>
sx.

NoWaC. February 2012. <http://www.tekstlab.uio.no/nowac/>.

Risten. March 2012. <http://www.risten.no/>.

Sami statistics. 9 March 2005. <http://www.sami-statistics.info/default.asp?nc=257&id=46>.

Slaastad, T. I. 2012. *Samisk statistikk 2012 - Sámi statistihkka 2012*.

http://www.ssb.no/emner/00/00/20/rapp_samer/rapp_201203/rapp_201203.pdf.