

Mo riggodahttit sátnegirjji morfologiiain vai šaddá buorin lohkkisátnegirjin?

Lene Antonsen, Trond Trosterud ja Ryan Johnson
Giellatekno
<http://giellatekno.uit.no/>



Sámi leksikografija hástalusat

- ▶ x-sámi sátnegirjjit
- ▶ terminologijja sátnegirjjit
- ▶ sámi-sámi sátnegirjjit
- ▶ digitálalaš sátnegirjjit
 - ▶ Vuosttaš digisánit, Giellatekno 2008 <http://giellatekno.uit.no/words/dicts/index.eng.html>
 - ▶ Vuosttaš digibaakoeh, Giellatekno 2010 <http://giellatekno.uit.no/words/dicts/index-sma.eng.html>
 - ▶ davvisámi-engelasiel-davvisámi, Figueiredo <http://www.freelang.net/online/sami.php>
 - ▶ davvisámi-dáru-davvisámi, Davvi Girji http://533.davvi.no/ordbok_samnor.php?lang=sam

Girjjálašvuohta

- ▶ Larsson, L.-G. 1997: **Prästen och ordet. ur den samiska lexikografins historia.** LexicoNordica, 4:101—117.
- ▶ Trosterud, T. 2000: **Kåven, Brita E. (red) 2000: Stor norsk-samisk ordbok [book review].** LexicoNordica, 8:283—306.
- ▶ Antonsen, L., Trosterud, T., Gerstenberger, C.-V., ja Moshagen, S. N. 2009: **Ei intelligent ordbok for samisk.** LexicoNordica, 16:271—283.
- ▶ Magga, O. H. 2012: **Lexicography and indigenous languages.** — Fjeld, R. V. ja Torjusen, J. M., doaimm., Proceedings of the 15th EURALEX International Congress, s. 3—18, Oslo, Norway. Department of Linguistics and Scandinavian Studies, University of Oslo.
- ▶ Trosterud, T. ja Eskonsipo, B. N. 2012: **A north sami translator's mailing list seen as a key to minority language lexicography.** — Fjeld, R. V. ja Torjusen, J. M., doaimm., Proceedings of the 15th EURALEX International Congress, s. 250—256, Oslo, Norway. Department of Linguistics and Scandinavian Studies, University of Oslo.

Lemmasátnegirjji

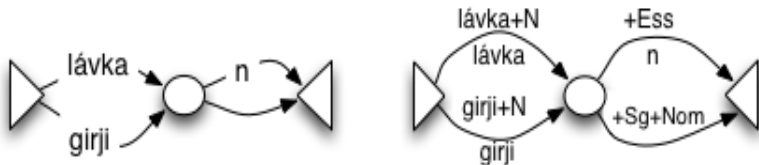
	Davvisámegiella	Suomagiella	Dárogiella
Galle sáni teavsttain	252 461	45 144	64 994
Galle lemma sátnelisttus	99 071	94 111	38 983
Gokčá	7,9 %	10,0 %	30,5 %

(Antonsen ja earát 2009)

Sátnegirji dárbbáša morfologiija

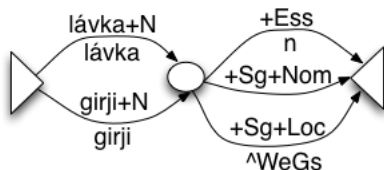
- ▶ sojaheapmi:
 - ▶ *niidii: nieida*
 - ▶ *eahtsa: iehtsedh, gáatan: gâetie*
- ▶ suorggideapmi:
 - ▶ *juhkkjuvvui: juhkat, juohkit*
 - ▶ *nööremes: noere*
- ▶ goallostapmi:
 - ▶ *bargojoavku: bargu + joavku*
 - ▶ *bigkemegierkine: bigkedh + gierkie*
 - ▶ 1.1 mill. sánis: 26.7 % substantiivvain leat goallossánit, dahje 7.0 % buot sániin
- ▶ partihkkalat:
 - ▶ *bođiige: boahtit*

Automáhtat – FST (loahpalaš dilletransduserat) – lexc



girji+N:gir'ji GOAHTI ;
lávka+N:láv'ka GOAHTI ;

Maidái morfologiiija — twolc



lávka+N+Ess	lávkan
lávka+N+Sg+Nom	lávka
lávka+N+Sg+Loc	lávkkas
girji+N+Ess	girjin
girji+N+Sg+Nom	girji
girji+N+Sg+Loc	girjjis

vk	->	vkk	 	-	Vow*	^WeG	;
rj	->	rjj	 	-	Vow*	^WeG	;

^WeG: báhcahat

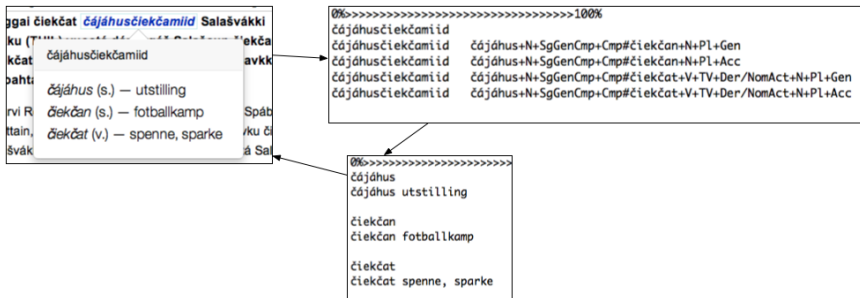
girji+N: gir'ji GOAHTI ;
lávka+N: láv'ka GOAHTI ;

Sátnegirjjet main leat sátnehámit doibmet offline

- ▶ Vuosttaš digisánit: 9999 lemma → 110 000 sátnehámi
- ▶ Voestes digibaakoeh: 10 657 lemma → 85 000 sátnehámi

Neahttadigisánit doibmet online

FST-analysa ja jorgalus



Neahttadigisánit addet sojahanhámiid

sma→nob ▾ noere Oza

noere (adj.)

○ ung

adj. Sg Nom	noere
adj. Comp Sg Nom	nuerebe
adj. Superl Sg Nom	nööremes

Dárbašlaš heiveheamit

- ▶ Lasihit mánggaidlogu sániid analysáhtorii:
gaskabeaivi – gaskabeaivvit
gávdni – gávnnit
→ gaskabeaivvit+N+Pl, gávnnit+N+Pl
- ▶ Čoavdit homonymiijaid analysáhtoris:
vuovdi:vuovddi – vuovdi:vuovdi
beassi:beassi – beassi:beasi
→ vuovdi+N+NomAg, beassi+N+G3
- ▶ Čorget sojahanparadigmaid go leat variánttat:
fievrridit:fievrridin – fievrredit:fievrredin
→ fievrridit+v1+V, fievrridit+v2+V

Neahttadigisánit – FST sátnegirjjit

`http://sanit.oahpa.no/`

`http://baakoeh.oahpa.no/`

Mobiilatelefovnas

nordsamisk → norsk (⇌ Swap)

váccát

Søk

vázzit (v.) –

1. gå, spásere

Odne mun váccán johtilit bargui, go mus lea hoahppu.

I dag går jeg fort til jobb, fordi jeg har det travelt.

váccát is a possible form of ...

vázzit

v. ind. pres. 2.p.ent.

Davisamegiella: Sátnegirjjiid buohtastahttin

FST-sátnegirji ja sátnehápmesátnegirji: 9999 lemma.
 Lemmasátnegirji mas livčče 99 000 lemma, govččasii 7,9 %

	Jorgalus		Oasse- jorgalus		li leat jorgalus		100 %
SH-sátneg: buot sánit	36 231	81.9 %			8025	18.1 %	44 256
FST-sátneg.: buot sánit	39 920	90.2 %	549	1.2 %	3787	8.6 %	44 256
SH-sátneg: earálágan sánit	7874	57.8 %			5758	42.2 %	13 632
FST-sátneg: earálágan sánit	10 862	80.0 %	448	3.3 %	2322	17.0 %	13 632

Lullisámeigiella: Sátnegirjjiid buohtastahttin

FST-sátnegirji ja sátnehápmesátnegirji: 10 657 lemma

	Jorgalus		Oasse- jorgalus		li leat jorgalus		100 %
SH-sátneg: buot sánit	44 989	72.7 %			15 268	25.3 %	60 037
FST-sátneg.: buot sánit	53 295	88.8 %	475	0.8 %	6266	10.4 %	60 037
SH-sátneg: earálágan sánit	5015	41.4 %			7100	58.6 %	12 092
FST-sátneg.: earálágan sánit	8039	67.0 %	308	2.6 %	3660	30.5 %	12 007

#SoMe-sátnegirji (Sosiála mediat)

#SoMe-sátnegirji (Sosiála mediat)

Jua, livzze buorre jus nu ain birgese dan aigge, ahte eamidat maid alo siiddas go manat leat unnit, nu ahte manat harjanit ju unnin leahket doppe ja besset de oahppat. Muhto oluhat darbbasit ...

- ▶ Ártegis ságat: 20 % čállojuvvon davvisámegiell boallobeavddi haga.

#SoMe-sátnegirji (Sosiála mediat)

Jua, livzze buorre jus nu ain birgese dan aigge, ahte eamidat maid alo siiddas go manat leat unnit, nu ahte manat harjanit ju unnin leahket doppe ja besset de oahppat. Muhto oluhat darbbasit ...

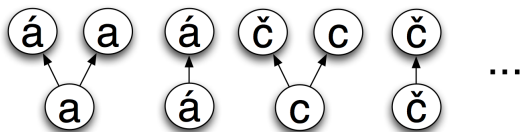
- ▶ Ártegis ságat: 20 % čállojuvvon davvisámegiell boallobeavddi haga.

Munno vástadus:

Davvisámegiell FST-sátnegirji gos teavsttat leat analyserejuvvon dábálaš FST:in ja dasa lassin FST:in mii lea heivehuvvon #SoMe čállinvuohkái.

#SoMe-FST

Jos teavsttas lea c, de sánis sáhtta maid leat ...



#SoMe sátnegirji

Boađus dábálaš FST:in ja #SoMe-FST:in

	Jorgalus		Oasse- jorgalus		li leat jorgalus		100 %
Dáb. FST buot sánit:	50 263	76.3 %	250	0.4 %	15 364	23.3 %	65 877
#SoMe-FST buot sánit:	54 197	80.6 %	315	0.5 %	12 753	19.0 %	67 265
Dáb. FST earálágan sánit:	8813	50.8 %	224	1.3 %	8326	48.0 %	17 363
#SoMe-FST earálágan sánit:	10 596	59.8 %	286	1.6 %	6825	38.5 %	17 707

Ođđadárogiella dáru-sámi-sátnegirjjis

Ođđadárogiella dáru-sámi-sátnegirjjis

Manne ođđadárogiella lea dehálaš?

Ođđadárogiella dáru-sámi-sátnegirjjis

Manne ođđadárogiella lea dehálaš?

- ▶ Don ipmirdat ahte *–ar* lea seamma go *–er*, muhto dihtor ii
- ▶ Dážat čállet sámi áššiid birra:
 - ▶ Tove Bull, Inger Johansen, Svein Lund, Sjur Moshagen, Jon Todal, Trond Trosterud, Ketil Zachariassen, ...
- ▶ Virggálaš teavsttat sáhttet leat ođđadárogillii čállojuvvon
- ▶ Mánngga sápmelaččas ii leat leamaš ođđadárogiella skuvllas
- ▶ Suomabeal sápmelažžii sáhtta leat váttis diehtit leago ođđadárogiel vai girjedárogiel teaksta

Viiddiduvvon dárogiela sátnegirji

Girjedárogiella-davvisámegiella FST-sátnegirji ođđadárogiel teavsttas (Wikipedia)

		FST gokčá		Sátnegirji gokčá	
Girjedáru teaksta	Buot girjedáru variánttat	703 950	93.36 %	1 644 286	84.50 %
Ođđadáru teaksta	Konservatiiva girjedáru	2 191 428	79.34 %	3 504 733	66.96 %
	Buot girjedáru variánttat	1 849 654	82.56 %	3 206 796	69.77 %
	Girjedáru oktan ođđadáruin	1 101 116	89.62 %	2 530 995	76.14 %

Mo viiddiduvvon sátnegirji lea dahkkon

- ▶ FST jorgala dábálaš sániid ođđadarogielas girjedárogillii:
 - ▶ *kva, korleis, kjem, sjølv, ikkje, truleg, betre, ...*

Mo viiddiduvvon sátnegirji lea dahkkon

- ▶ FST jorgala dábálaš sániid ođđadarogielas girjedárogillii:
 - ▶ *kva, korleis, kjem, sjølv, ikkje, truleg, betre, ...*
- ▶ Girjedárogie la FST:i masa lea lasihuvvon ođđadárogiel morfologijja

LEXICON m1 ! dag, hane

+N+Msc+Sg: 0-en ;

+N+Msc+Pl: er-ene ;

+N+Msc+Pl+Nynorsk: ar-ane ;

+N: R ;

Ovttá sátnegirjjis bohtet mánga sátnegirjji

1. Dábálaš sátnegirji
 - ▶ Jos sátni lea leksikaliserejuvvon, de mii čájehit dušše dan jorgalusa
2. Studeantasátnegirji
 - ▶ Sátnegirji jorgala maid goallossániid osiid
3. #SoMe sátnegirji
 - ▶ #SoMe-sátnegirji dohkkeha maid *acdntz*

Eará sámegielat

1. Bihtánsámegiella: rgg/rrg ↔ rg
2. Nuortalašgiella ja anársámegiella: bustávat
3. Gielddasámegiella: bustávat + guokte ortografiija



Тёсст-Википедьяйя баяс

Тэдта лй кйллт сáмь кйл тёсст-Википедья шўррмусс оаллк лысст.

Ныдтѣ цѣлдѣкэ, шѣннтэв лыстэ сѣрья кйллт сáмь кйллэсѣт тёсст-Википедья ([Wikimedia Incubator](#)) лынкэм гуэйкэ. Тин вѣнкѣ кйллт сáмь кйл Википедья лынкэм гуэйкэ ля́ннч шўрр цыстэсѣт. Куэссѣ ля́ннч тэннэ тёсст лыстэ́нѣ ё́ннэ сáмас кырьхэ, танна тáдт проект ля́ннч э́ллм кйллт сáмь кйл Википедья шѣннтмэнч. Куэссѣ сѣ́нн ля́ннч э́ллмса шѣннтмэнч, танна пугк сáмас кырьха текстэ лй́ннченѣ Википедья-Энциклопедьяе пыйймэнч.

Eará gielat

1. Bustávaváttisvuodát:
 - ▶ románalašgiella, buot Ruošša unnitlogugielat
2. #SoMe-fenomena
 - ▶ slávlaš ja románalaš gielat

Re: račun za struju

Postao/la **motorola01** dana 06/02/2013 14:56

“ sahan je napisao/la:

“ kakavdanakneiskustvu je napisao/la:

došao mi račun za struju za januar, tačnije za obrač. period od čini mi se 29.12. do 28.01., iznos računa za pfening manji nego prošli račun, od 19.1. do 1.2. niko nije bio u stanu, supruga i ja bili na godišnjem, prethodni računi, za novembar, oktobar, za marku dvije veći nego za januar kad nas deset dana nije bilo u kući, je li imao ko slično iskustvo?

Ovo ti je klasični primjer "daljinskog očitavanja", male plate u EP 😊 pa **mrsko onome sto treba da dodje da ocita da ide na lice mjesta, stavi cifru od proslog mjeseca.** U sustini ti mozes da odes do EP da oni to isprave a mozes i da platis pa da slijedeci mjesec budes u nekoj maloj pretplati, jer svejedno placas kilovate.

Ako si recimo platio od 0 - 150 kw pa slijedeci mjesec ti izracunali da si potrosio od 150 - 300 a ti u sustini potrosio od 150 - 200 (sto pokazuje tvoje brojilo staro-novo stanje) ako platis bices slijedeci mjesec u pretplati, tako da oni nikad gube 😊

Grijesis! Ljudi koji dobiju spisak kupaca koje danas trebaju očitati, dobiju "prazan list" tako da nemaju prosjek potrosnje od proslog, niti bilo kojeg mjeseca. Trebao je otici do brojila i provjeriti stanje i uporediti ga sa stanjem na racunu. Ako opet sumnja na potrosnju, odnosno ispravnost brojila, onda iskljuci sve potrosace u stanu i opet do brojila i vidi da li se okrece. Dalje znas i sam, sta i kako.

Konklusuvdna

FST-sátnegirjjis leat ovdamunit sátnehápmesátnegirjji ektui:

- ▶ FST-sátnegirji gierdá sihke gielalaš ja ortográfalaš variašuvnnaid
- ▶ FST-sátnegirji máhttá eanet morfologijja: davvisámegiela 57,8 % → 83.3 % , ja lullisámegiela 41.4 % → 69.6 %
- ▶ Mii sáhttit geavahit resurssaid mat juo gávdnojit
- ▶ lea álkit programmereri:
 - ▶ FST-sátnegirji lea modulára
- ▶ lea álkit geavaheddjái:
 - ▶ ii dárbaš viežžat ja sajáiduhhtit
 - ▶ geavaha álo ođđaseamos veršuvnna