

U i T

N O R G G A
Á R K T A L A Š
U N I V E R S I T E H T A

Humaniora, servodatdiehtaga ja oahpaheaddjioahpu fakultehta

Sámegielaide modelleren – huksen ja heiveheapmi duohta giellamáilbmái

Lene Antonsen

*Nákkosgirji buktojuvvon Philosophiae Doctor grádii
Guovvamánu 2018*

Eai buot sánit čázi jeage

Vuosttaš Digisánit

subst. → čáhci

vann, vatn

Analyser: sg. gen. *el.* sg. akk.



**Sámegielaide modelleren
– huksen ja heiveheapmi duohta
giellamáilbmái**

Lene Antonsen

Nákkosgirji geigejuvvon čuovvovaš gráda ovdii:
Philosophiae Doctor (PhD)



Humaniora, servodatdiehtaga ja oahpaheaddjioahpu fakultehta
UiT Norgga árkálaš universitehta
njukčamánu 1. b. 2018

Sisdoallu

Ovdasánit	vii
Artihkallistu	viii
Čoahkkáigeassu	ix
Summary in English	x
Govuslistu	xi
Tabeallalistu	xi
Oanádusat	xiii
1 Láidehus	1
1.1 Nákkosgirjji áššečuolmmat	1
1.2 Duogáš	3
1.3 Nákkosgirjji artihkkalát ja mu buvttá	4
1.4 Čállosa huksehus	7
2 Giellamodealla huksen	9
2.1 Sámegielaid modelleren transduseriiguin	9
2.2 Sátnehámiid huksen	11
2.2.1 Geažusmorfologiija huksen	12
2.2.2 Morfofonologalaš rievddademiid modelleren	15
2.2.3 Molssaevttolaš vuogit	20
2.3 Buot sániide addit analysa	23
2.3.1 Sáni olggobeale ja siskkobeale hápmi (Art. I)	23
2.3.2 Sámegielaid morfologalaš analysáhtoriid buohtastahttin	25
2.3.3 Eahpesihkkaris norbma ja ollu variánttat	29
2.4 Rivttes analysa válljen	32
2.4.1 Disambigueren ráddjehusgrammatihkain, Constraint Grammariin	32
2.4.2 Sániid lemmatiseren (Art. II)	36
2.4.3 Sámegielaid syntávssalaš analysáhtorat	39
2.5 Čoahkkáigeassu	40
3 Giellamodealla heiveheapmi duohta giellamáilbmái	43
3.1 Giellamodealla geavaheapmi geavaheddjiidprográmmain	43
3.2 Sátnehámis lemmai ja lemmas sátnehámiide	44
3.2.1 Sátnedárkkistanprográmmii genereret sátnehámiid (Art. III)	44

3.2.2	Sátnegirjjiid riggodahttit morfologijain (Art. IV ja V)	46
3.3	Giellamodealla gáržžideapmi	54
3.3.1	Eai buot sánit leat geavahas	55
3.3.2	Giellamodealla gáržžideapmi norpma ektui	57
3.3.3	Badjelmearálaš genererema ráddjen	58
3.3.4	Giellamodealla gáržžideami čoahkkáigeassu	71
3.4	Giellamodealla viiddideapmi	72
3.4.1	Grammatihkkahárjehusaid genereren (Art. VI)	73
3.4.2	Oahpahalli morfologalaš gaskagiela gieđahallan (Art. VII)	76
3.4.3	Oahpahalli syntávssalaš gaskagiela gieđahallan (Art. VIII)	78
3.4.4	Giellamodealla viiddideami čoahkkáigeassu	80
3.5	Giellamodealla, dan geavahus ja duohta giellamáilbmi	81
4	FST addá vejolašvuodaid eamiálbmotgielaide	83
4.1	Anárašgiella ovdamearkan	83
4.2	Gielat Sámi olggobealde	85
5	Konklusuvdna	89

Ovdasánit

Bargagohten Giellateknos sámegeleaid giellateknologiijain čavčča 2006:s (earret golbma vahku geassebargin 2002:s). Dalle ledjen 15 jagi oahpahan sámegelela sihke mánáide ja rávesolbmuide, manjelgo ieš ledjen oahppan sámegelela. Dan dihte munnje šattai váibmoáššin geavahit giellateknologiija ávkin giellageavaheddjiide ja erenoamážit oahpahallide.

Giellateknobarggus lei ollu oahppan láhkai, ja leange oahppan juoidá juohke olbmos Divvun-joavkkus. Erenoamážit lea Sjur Moshagen vástidan mu jearaldagaide min prógrámmaid birra, ja Børre Gaup lea veahkehan mu teknihkalaš áššiiguin. Dasa lassin leat Marja-Lisa Kappfjell, Biret Ánne Bals Baal, Ritva Nystad, Elisabeth Scheller ja Hanna Outakoski leamaš dehálaš ságastallanustibat relevánta áššiin, ja muduige mielde ráhkadeamen buori birrasa min feaskáris.

Lean ovttasbargan lingvistajoavkkuin University of Albertas, ja lea leamaš hirkmat miellagiddevaš oahpásmuvvat Kanáda eamiálbmotgelelaide, ja Antti Arppe ja Jordan Lachler leaba guovddášolbmot dán ovttasbarggus.

Erenoamážit háliidan giitit Ciprian Gerstenbergera, gii ráhkada skriptaidd maid dárbašan ja veahkehan oazžut korpusteavsttaid olamuddui, ja Chiara Argesa, gii dál lea joatkimin giellaoahppanreaidduid programmerenbarggu maid Sara Huhmarniemi, Heli Uibo ja Ryan Johnson ovdal leat bargan ja de maddái Marja-Liisa Olthuisa, gii lea rahpan anárašgelela giellamáilmmi munnje.

Giitu maddái Linda Wiechetekii gii lea veahkehan mu LaTeX-prógrámmain mainna lean nákkosgirjji čállán. Eanemus giitu Trond Trosterudii, gii lea leamaš mu bagadalli. Vaikko sus lea ollu bargu ja jámma lea jođus, de in leat goassege ballan das ahte sus ii livčče dilli lohkat ja kommenteret mu barggu.

Maddái Nils Øivind Helander ja Laura Janda leaba kommenteren mu čálloša. Giitu maddái Berit Merete Nystad Eskonsipo gii lea lohkan mu čálloša ja fuomášuhtán sihke gelelalaš meattáhusaid ja váilevaš čilgehusaid.

Loahpas giittán iežan guoimmi, Sauli, gii gierdavaččat lea guldalan mu jurdagiid birra ja vástidan buot lágan gelelalaš jearaldagaide.

Romsa njukčamánu 1.b. 2018

Artihkallistu

1. **Artihkal I:** Antonsen, Lene ja Trond Trosterud. 2017. Ord sett innafra og utafra – en datalingvistisk analyse av nordsamisk. *Norsk lingvistisk tidsskrift* 35(1):153–185.
2. **Artihkal II:** Antonsen, Lene ja Trond Trosterud. 2011. Next to nothing – a cheap South Saami disambiguator. Girjjiis: *Proceedings of the NoDaLiDa 2011 Workshop Constraint Grammar Applications*. Čanas 14:10 NEALT Proceedings Series, siiddut 1–7. Tartu: NEALT.
3. **Artihkal III:** Antonsen, Lene. 2013. Čállinmeattáhusaid guorran. [English summary: Tracking misspellings]. *Sámi diedalaš áigečála* 2/2013: 7–32.
4. **Artihkal IV:** Antonsen, Lene, Ciprian-Virgil Gerstenberger, Sjur Nørstebø Moshagen ja Trond Trosterud 2009: Ei intelligent elektronisk ordbok for samisk. *Lexico-Nordica* 16:271–283.
5. **Artihkal V:** Johnson, Ryan, Lene Antonsen ja Trond Trosterud. 2013. Using finite state transducers for making efficient reading comprehension dictionaries. Girjjiis: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa 2013)*, Čanas 16 NEALT Proceedings Series, siiddut 59–71. Stroudsburg, Pennsylvania: ACL.
6. **Artihkal VI:** Antonsen, Lene, Ryan Johnson, Trond Trosterud ja Heli Uiibo. 2013. Generating modular grammar exercises with finite-state transducers. Girjjiis: *Proceedings of the second workshop on NLP for computer-assisted language learning at NoDaLiDa 2013*, Čanas 17 NEALT Proceedings Series, siiddut 27–38. Linköping: Linköping Electronic Conference Proceedings.
7. **Artihkal VII:** Antonsen, Lene. 2012. Improving feedback on L2 misspellings – an FST approach. Girjjiis: *Proceedings of the SLTC 2012 workshop on NLP for CALL, Lund 25th October 2012*, Čanas 80, siiddut 1–10. Linköping: LiU Electronic Press.
8. **Artihkal VIII:** Antonsen, Lene. 2013. Constraints in Free-input Question-Answering Drills. Girjjiis: *Proceedings of the second workshop on NLP for computer-assisted language learning at NoDaLiDa 2013*, Čanas 17 NEALT Proceedings Series, siiddut 11–26. Linköping: Linköping Electronic Conference Proceedings.

Čoahkkáigeassu

Nákkosgirjjis guorahalan leago grammatihkalaš giellamodealla vástáduš sámegeielaid ja eará davvimáilmmi eamiálbmotgeielaid giellateknologalaš dárbbuide? Grámmatihkalaš giellamodealla lea huksejuvvon *Finite state transducerin* (FST:n). Suokkardalan hástalusaid mat čuožžilit go galgá hukset ja heivehit sámegeielaid grammatihkalaš giellamodeallaid duohta giellamáilbmái.

FST:in sáhtta modelleret morfofonologalaš rievddademiid ja analyseret teakstakorpusiid vaikko teavsttain lea ollu lingvisttalaš variašuvdna. FST:in sáhtta maiddái genereret sátnehámiid mat eai gávdno korpusis, muhto gávdnojit gielas. Syntávssalaš analysáhtor mii lea huksejuvvon Constraint Grammariin, ii gáibit stuorra teakstačoakkáldaga, ja lea gierdilis vuohki disambigueret morfologalaš analysaid gaskkas.

Čájehan mo sáhttit ávkkástallat sámegeielaid grámmatihkalaš giellamodeallain geavaheddjiidprográmmain. Grammatihkalaš gilkorat dahket vejolažžan addit metalingvisttalaš dieđuid ja máhcahaga geavaheddjiide. FST ferte heivehit nu ahte dat buorebut speadjalastá duohta giela. FST lahkonanvuogi čuolbma lea badjelmearálaš sátnehámiid genereren, ja dan dihte lea dehálaš gáržžidit giellamodealla. Go ráhkada giellateknologalaš reaidduid giellaoahpahallamii, de lea ávkkálaš dasa siskkildit giellaoahpahalli gaskagiela. Suokkardalan muhtun vugiid mo sáhtta gáržžidit ja viiddidit giellamodealla vai buorebut sulastahtta duohta giellamáilmmi.

Čájehan ahte FST huksen vástida mángga davvimáilmmi eamiálbmotgiela dárbbuide oažžut giellateknologalaš reaidduid. FST lahkonanvuohki oktan buriin vuodđostruktuvrain mainna ávkkástallá vuodđobarggus mii lea dahkkon sámegeielaide, dahká alladásat geavaheddjiidprográmmaid olámuddui vehádatgielaide main lea rikkes morfologiija, vátna teakstakorpus ja unnán hállit.

Summary in English

In this thesis I investigate whether grammatical language modeling is an appropriate response to the need for language technology for the Saami languages and other circumpolar indigenous languages. The grammatical language models under investigation are built as Finite state transducers (FST). I examine the challenges of building such grammatical models for Saami languages and adapting them to real-world linguistic issues.

A finite state transducer (FST) makes it possible to model morphophonological alternations and analyse text in a corpus, even when there is considerable linguistic variation in the text. One can also generate word forms that are not found in the corpus, but exist in the language. A syntactic analyser based on Constraint Grammar does not require a large text corpus, and does robust disambiguation of multiple morphological analyses.

I show how Saami grammatical language models can be implemented in various user programs. Grammatical tags make it possible to provide both metalinguistic information and immediate feedback to the user's input. It is necessary to adapt the FSTs to real language usage. The FST approach causes overgeneration, which is why it is important to limit the language model. Including the learner's interlanguage is also useful for language learning tools based on language technology. I have examined a number of ways both to limit and to expand the language models.

I show that the construction of an FST is the key answer to the need for language technology tools for circumpolar indigenous languages. With the appropriate infrastructure available, which also makes it possible to port results achieved for the Saami language to other languages as well, the FST approach places advanced user applications with the reach of minority languages with complex morphology, meagre corpus resources, and few speakers.

Govuslistu

2.1	Automáhta (FSA).	12
2.2	Grammatihkalaš transduser (FST).	13
2.3	Substantiivaleksikon LEXC-fiillas.	13
2.4	Joatkkaleksikon LEXC-fiillas.	14
2.5	Leksikontransduser ja morfologalaš transduser.	16
2.6	Leksikontransdusera ja morfologalaš transdusera kompileren.	16
2.7	YAML testenfiila.	17
2.8	TWOLC njuolggadusat.	18
2.9	Suomagiela analysáhtor Giella-fin: Buot morfofonologalaš rievddadeamit dáhpáhuvvet LEXC-prográmmas.	22
2.10	Kásusiid juohkáseapmi SIKORis.	23
2.11	Giella-sma syntávssalaš analysáhtora njuolggadusaid beaktilvuohta.	38
3.1	Divvunprográmma evttohusat boastut čállon sániide.	46
3.2	Vuosttaš digisániid lohkanveahkki.	47
3.3	Vuosttaš digisániid váldosátneartihkal.	47
3.4	Neahttadigisániid lohkanveahkki sádde sáni analysáhtorii.	51
3.5	Neahttadigisániin beassá njuolga SIKORii.	51
3.6	Stávenložžen.	52
3.7	Neahttadigisánit čilgejit infinihtta vearbba funkšuvnna.	53
3.8	FST: Dynámalaš goallosteapmi.	56
3.9	FST: Bálgás lea diakrihtalaš leavga.	61
3.10	ICALL: Morfologalaš máhcahat.	73
3.11	ICALL: QA-hárjehusa templáhta.	74
3.12	Divvun-prográmma ii máhte álo veahkehit giellaoahpahalli.	76
3.13	L2-FST sisttisoallá gaskagiela gehčosiid ja morfofonologiija.	77
3.14	ICALL: QA-hárjehus mas lea friddja čállin.	79
3.15	ICALL: QA-hárjehusa analysa.	80
4.1	FST lea vuodđun ollu geavaheddjiidprográmmaide.	86
4.2	Giellateknologalaš reaiddut eamiálbmogiid gielaide.	87

Tabeallalistu

2.1	Anárašgiela transdusera báhcahagat ja daid morfofonologalaš rievddadeamit.	19
2.2	Giella-smX morfofonologalaš transduseriid buohtastahttin.	26
2.3	Giella-smX leksikontransduseriid buohtastahttin.	27
2.4	Giella-smX analysáhtoriid beaktilvuohta.	30
2.5	Giella-smX analysáhtoriid modulaid interakšuvdna.	39
3.1	VD ja NDS sátnegirjjiid iešvuođaid buohtastahttin.	54
3.2	Substantiivvaid leksikonat stivrejit oamastangehčosiid lasiheami.	62
3.3	Vearbalemmat main leat finihtta veabahámit SIKORis.	68
3.4	Vuohki ráddjet veabahámiid genererema.	69
3.5	Giella-sme: Vearbba dábáleamos homonymijjat.	70

Oanádusat

Oanádusat čilgejuvvojit dadistaga teavsttas. Dás leat oanádusat mat leat geavahuvvon eanet siidduin. Čuovvovaš siiddu leat čilgejuvvon analysagilkorat, variábelat ja báhcahagat.

CG eng. ‘Constraint Grammar’ = ráddjehusgrammatihkka

FSA eng. ‘Finite-State Automaton’ = loahpalaš dilleautomáhta

FST eng. ‘Finite-State Transducer’ = loahpalaš dilletransduser

Giella-fin suomagiela analysáhtor

Giella-sje bitonsámegiela analysáhtor

Giella-sma lullisámegiela analysáhtor

Giella-sme davvisámegiela analysáhtor

Giella-smj julevsámegiela analysáhtor

Giella-smn anárašgiela analysáhtor

Giella-sms nuortalašgiela analysáhtor

Giella-smX buot sámegielaid analysáhtorat

ICALL eng. ‘Intelligent Computer-assisted Language Learning’ = giellaoahppan jierbmás dihtorvehkiin

L2 eng. ‘Language 2’ = giella maid olmmoš oahppá mannel go lea oahppan vuosttaš gielas (= L1)

LEXC eng. ‘Lexicon Compiler’ = leksikona kompilierenprográmma, muhto maddái programmerengiella mainna čatná oktii grammatihkalaš sáni ja sáni máddaga oktan affivsaiguin.

NDS Neahttadigisánit-sátnegirji

QA eng. ‘Question Answering’ = jearaldat oktan vástádusain

SIKOR Sámi Internationála Korpus, UiT Norgga ártkalaš universitehta ja Norgga Sámedikki sámi teakstačoakkáldat

TWOLC eng. ‘Two Level Compiler’ = guovttedásat ovttastanprográmma. Programmerengiella mainna rievdadit vuollálas hámiid rivtteshámat bajošhápmin ortográfalaš ja morfofonologalaš njuolggadusaid mielde

VD Vuosttaš digisánit-sátnegirji

eng. engelasgiella

Analysagilkorat, variábelat ja báhcagat:

Cns konsonánta	+ Inf infinitiiva
Vow vokála	+ Interr gahčanpronomen
^CLEN konsonántaguovddáža guhkkun	+ IV intransitiiva
^CSH konsonántaguovddáža oatnun	+ Loc lokatiiva
^RLEN vokálaguovddáža guhkkun	+ N substantiiva
^RC konsonántaguovddáža arkifonema	+ Neg biehttanvearba
^RV vokálaguovddáža arkifonema	+ Nom nominatiiva
^RVSH vokálaguovddáža oatnun	+ NomAg dahkkinomen
^QLD jearaldatrádji	+ Num lohku dahje lohkosátni
^WG geahnohis dássi	+ Org organisašuvdna (dál geavahuvvo +Sem/Org)
+ A adjektiiva	+ Par partitiiva
+ Acc akkusatiiva	+ Pers persovdnapronomena
+ Adv advearba	+ Pron pronomen
+ Arab arábalaš lohku	+ Prs preseansa
+ Attr attribuhtta	+ Prt preterihtta
+ CLB cealkkarádji	+ Pl mánggaidlohku
+ Cmp goallosteapmi	+ Pl1 mánggaidlogu vuosttaš persovdna
+ Cmp/SgGen goallosteami mearusoassi lea ovttaidlogu genitiivvas	+ Pl2 mánggaidlogu nuppi persovdna
+ Cmp/SgNom goallosteami mearusoassi lea ovttaidlogu nominatiivvas	+ Pl3 mánggaidlogu goalmát persovdna
+ Cond konditionála (eaktovuohki)	+ Po postposišuvdna
+ ConNeg vearba biehttanhápmi	+ Pr preposišuvdna
+ CS subjunksuvdna	+ Prop namma (propra)
+ Der/ suorggádus	+ PrfPrc perfeavtta partisihppa
+ Dial/ suopman	+ Prs preseansa
+ Du1 guvttiidlogu vuosttaš persovdna	+ Prt preterihtta
+ Du2 guvttiidlogu nuppi persovdna	+ Px oamastangeažus
+ Du3 guvttiidlogu goalmát persovdna	+ Sem/Ani ealli namma
+ Err/Lex sátnemeattáhus	+ Sem/Fem nissona ovdanamma
+ Err/Orth čállinmeattáhus	+ Sem/Hum olbmonamahus
+ Ess essiiva	+ Sem/Plc báikenamahus
+ G3 lemmas lea dássemolsuma goalmát gráda mii ii merkejuvvo čállingielas	+ Sem/Sur goargu/sohkanamma
+ Gen genitiiva	+ Sg ovttaidlohku
+ Hom1, Hom2 homonymat 1 ja 2	+ Sg1 ovttaidlogu vuosttaš persovdna
+ Hum olbmonamahus (dál geavahuvvo +Sem/Hum)	+ Sg2 ovttaidlogu nuppi persovdna
+ Ill illatiiva	+ Sg3 ovttaidlogu goalmát persovdna
+ Imprt imperatiiva	+ Sup supiidna
+ Ind indikatiiva	+ TV transitiiva
	+ V vearba
	+ VAbess vearbaabessiiva
	+ VGen vearbagenitiiva

Kapihtal 1

Láidehus

1.1 Nákkosgirjji áššečuolmmat

Giellamodealla huksen lea vuohki guorahallat giela, sihke morfologalaččat ja syntávs-salaččat. Ieš giellamodealla sáhttá jurddašit leat hypotesan das makkár giella lea. Grammatihkkavuđot giellamodealla huksen vuodđuduvvo giela dokumentašuvdnii, nugo sátnelisttuide ja grammatihkkagirjjiide, giellačehpiid máhttui ja lingvisttaid intuitiivduvni. Gelddolaš ja dehálaš oassi barggus lea hypotesa testen. Korpusteavsttaid analysa muitala giellamodealla gokčanmeari ja leatgo rivttes analysat. Korpusanalysa mielddisbukta maiddá ahte giellamodealla ferte gokčat sániid mat eai leat sátnelisttuin, ja gielalaš ráhkadusaid maid grammatihkkagirji ii giedahala. Sátnehámiid genereren čájeha fas man deaivilit giellamodealla máhttá genereret dušše hámiid maid giellačeahppi dohkkehivččii gillii gullevaš sátnin.

Giellateknologalaš reaidut leat buot gielaide eaktun dasa ahte sáhttet ceavzit geavahangiellan odđáigásaš servodagas go ollu teknologalaš čovdosat vuodđuduvvet giellateknologijii. Máilmmi váldogielaid giellateknologalaš čovdosat eai atte dákkár reaidduid davvimáilmmi eamiálbmotgielaide main leat rikkes morfologijja, vánis teakstakorpus ja unnán hállit. Giellaservodagain dábálaččat eai leat nu galle aktiivvalaš čállit, ja lea váddásat eatnigillii go eanetlogugillii čállit go olmmoš ii beasa lohkat eatnigielas doarvái, iige čállit nu olu ahte čállin automatiserejuvvo, nugo Outakoski (2015) dutkamuš čájeha davvisámegeiela ektui. Helander (2016, 15–16) čujuha eanet dutkamušaide mat čájehit

ahte sámegeiella čállingiellan lea rašis juolggi alde. Dát dilli mielddisbuktá ahte giellateknologalaš reaiddut sáhttet leat ávkkálaččabut čállinproseassas vehádatgiela čállide go váldogiela čállide. Dása gullat maiddáa ahte stuora oassi giellageavaheddjiin leat nuppegielagat geat dávjá dárbbasit giellateknologalaš reaidduid veahkkin go galget buvttadit sáni rivttes sojahaemi, ja maiddáa geavahit sáni idiomáhtalaččat.

NDS-sátnegirji, man kapihtal 3.2.2 giedahallá, lea giellateknologalaš reaidu, ja jus juohká NDS geavahanlogu davvisámegeielagiid loguin, de juohke davvisámegeielat geavahii NDS olles 60 geardde 2017:s¹, vaikko gávdno nubbi stuora davvisámi sátnegirji interneahatas². Jus juohká Bokmålsordboka/Nynorskordboka (= Girjedárogiepla sátnegirji/Ođđadárogiepla sátnegirji)³ geavahanlogu Norgga ássiid loguin, de juohke ássi geavahii sátnegirjii 8 geardde. Njealját kapihttalis čájehan ahte unnit sámegeielaid geavahanlogut gorálaččat leat vel stuoribut go davvisámegeiela geavahanlogut.

Dán nákkosgirjii giellateknologiiija vuodđuduvvo grammatihkkii ja lingvisttalaš njuolgadusaide, ja bajitdási áššečuolbma lea:

Leago grammatihkalaš giellamodealla vástádus sámegeielaid ja eará davvimáilmmi eamiálbmotgielaid giellateknologalaš dárbbuide?

In sáhte dán dutkamušas vástidit dáidda gažaldagaide dievaslaččat, muhto mun lean välljen guorahallat njeallje hástalusa mat leat bajitdási áššečuolmma eaktun. Nákkosgirjii gávcei artihkkala gusket dán njealji hástalussii, ja lean välljen erenoamážit deattuhit ja kommenteret daid osiid artihkkaliin mat leat hástalusaide relevánttat. Guokte artihkkala giedahallet giellamodealla dahje analysáhtora huksema, sihke morfologalaš ja syntávssalaš osiid, ja dan oktavuodas suokkardalan guokte hástalusa:

1. Mo modelleret morfofonologalaš rievddademiid ja addit analysa buot sániide?
2. Mo välljet rivttes analysa konteavstta ektui?

Guhtta artihkkala giedahallet mo heivehit giellamodealla geavaheddjiidprográmmaide.

Prográmmaide lea dehálaš makkár sátneanalysa giellamodealla addá sániide, ja makkár

¹Davvisámegeielas: NDS:s leat 1,55 miljovna geavaheami ja máilmmis gávdnojit 25 700 davvisámegeielaga (<https://www.ethnologue.com/language/sme> (24.02.2018)). Dárogieplas: Bokmålsordboka/Nynorskordbokas leat 42,14 miljovna geavaheami (Språkrådet: Årsrapport for 2017) ja Norggas leat 5,3 miljovna ássi (<https://www.ssb.no/befolkning/nokkeltall/befolkning> (24.02.2018)).

²Davvi Girji, <http://533.davvi.no> (24.02.2018)

³<http://ordbok.uib.no> (24.02.2018)

hámiid dat generere iešgudet lemmai, ja dán oktavuodas lea nákkosgirjji áššečuolbma mo oažžut giellamodealla sulastahttit giela nugo dat geavahuvvo duohta giellamáilmmis. Hástalusat leat:

3. Mo gáržžidit giellamodealla amas genereret badjelmearálaš hámiid?
4. Mo viiddidit giellamodealla vai dat maddái máhtta analyseret giellaoahpahalli gaskagiela?

Njealját kapihttalis máhcan bajitdási áššečulbmii, go geahčan mo dákkár vuohki hukset giellamodealla lea ávkkálaš davvimáilmmi eamiálbmotgielaide. Ii oktage nákkosgirjji artihkkaliin gula dán kapihttali, muhto lean leamaš mielde huksemin giellamodealla ja giellateknologalaš reaidduid anárašgiela várás ja guovtti kanádalas eamiálbmotgiela, Plains Cree ja davvihaidagiela, várás, ja maddái mielde almmuheamen artihkkaliid bohtosiid birra. Kapihtal čájeha maddái ovddeš kapihttaliid relevánssa davvimáilmmi eará eamiálbmotgielaide.

Suokkardalan sámegeleaid grammatihka giellateknologalaš perspektiivvas, muhto liikká nu, ahte lea vejolaš lohkat vaikko ii dovdda giellateknologiiija, jos beare lea máhttu ja beroštupmi sámegeleaid grammatihkas. Jus grammatihkkavuđot giellateknologiiija galgá doaibmat bures, de lea dárbbášlaš gulahallat giellačehpiiguin ja lingvisttaiguin geat leat min giellateknologiiijafeaskára olggobealde. Dán čálloš giellateknologiiija leage eanet giella go teknologiija.

1.2 Duogáš

Sámegeleaid giellamodealla man dán nákkosgirjjis govvidan, lea UiT Norgga árktaš universitehta (UiT) hálddus, ja buot sámi giellateknologiiijabarggut dáhpáhuvet UiT vuoddostruktuurras. Bargu ásaht sámegeleaid giellateknologiiija álggahuvvui jagi 2001, ja das lei dušše okta virgi gitta 2004:i. Jagi 2008 bargu rievddai proševttas fásta dutkanjoavkun, mii gohčoduvvo Giellateknon.

Norgga Sámediggi ásaht jagi 2004 Divvun-joavkku man ulbmil lei ráhkadit sátnedárkkistanprográmmaid golmma sámegeleaid várás. Joavkku lea álggu rájes ovttasbargan

UiT:in, ja dat sirdojuvvui universitehta vuollái 2011:s. Giellatekno ja Divvun-joavku leaba ovttas huksen oktasaš giellateknologalaš vuodđoresurssaid, maidda gullet davvi-, julev- ja lullisámegielaideid morfologalaš modeallat. Giellatekno lea maiddá huksen syntávssalaš analysáhtoriid daidda gielaide.⁴ Ovttasbargui gullá maiddá Giella-nammasaš vuodđostruktuvra mas sáhtta hukset ja testet morfologalaš giellamodeallaid ja daid vuodul ráhkadit sátnedárkkistanprográmmaid (Moshagen ja earát, 2013). Joavkkuid erohus lea ahte Giellateknos lea fokus analysaprográmmaid deskriptiiva beliide, ja Divvun-joavku fas ráhkada daid normatiiva veršuvnmaid.

Sámi giellateknologiija bargu vuodđuduvvá hui ollu suomagiela giellateknologiija dutkamii ja čovdosiidda, erenoamážit Helsset universitehta dutkiid bargui. Suomagiella lea sámegielaideid lagaš fuolkegiella, ja erenoamáš relevánta giellateknologiijii leat dan guovtti giela oktasaš iešvuodát nugo máškidis sátnortnet, mearehis ollu sátnegoallosteapmi, ja hui rikkes sojahan- ja suorggádusvuogádagat. Stáhtaguoddi giellan leat suomagielas ollu eanet resurssat go sámegielain, ja dan dihte sámegiela dutkanbiras viežžá máhtu ja teknologiija Suomas (Rehm ja Uszkoreit (2012) addiba suoma giellateknologiija bajilgova).

1.3 Nákkosgirjji artihkkalat ja mu buvttá

Mun lean erenoamážit leamaš mielde huksemin davvisáme- ja anárašgiela morfologalaš analysáhtoriid, ja maiddá buot sámegielaideid syntávssalaš analysáhtoriid. Sátnedárkkistanprográmma vuodđuduvvá morfologalaš analysáhtoriidda, muhto analysáhtoriid ráddjejuvppi norpma mielde, man birra čálán kapihttalas 3.3.2, gullá Divvun-joavkku bargui. Dattetge lean geahččaladdan ja implementeren oamastangeažushámiid ráddjema, mii lea kapihttala 3.3.3.3 fáddán, danne go das lea mearkkašupmi maiddá teakstakorpusa analysii. Seamma kapihttalas čálán maiddá oanehis namaid goallosteami ráddjemis, mainna lean bargan danne go lea dehálaš sátnegirjjiid analysii, mii fas lea artihkkaliid IV ja V fáddán.

⁴Divvun-joavkku lingvistan leat bargan Thomas Omma (davvisámegielain), Maja Kappfjell (lullisámegielain) ja Per-Eric Kuoljok, Inga Lill Sigga Mikkelsen ja Sandra Ráhka Nystø (julevsámegielain). Giellatekno lingvisttat barget fas sihke morfologalaš ja syntávssalaš analysáhtoriiguin, ja oktiibuot buot sámegielaiguin: Trond Trosterud ja mun, ja ovddeš bargit Marit Julien, Ilona Rauhala ja Linda Wiechetek.

Gávcci artihkkala leat mielde dan nákkosdutkanmušas:

Artihkkala I (*Ord sett innafra og utafra – en datalingvistisk analyse av nordsamisk*) váldofáddá lea maid grammatihkalaš analysáhtor sáhtttá muitalit davvisámi grammatihkas go geavahuvvo analyseret stuora davvisámi teakstakorpusa⁵. Artihkkalis leat guokte čálli, ja mu oassi artihkalbarggus lei čilget ja árvvoštallat grammatihkalaš analysa, nugo kásus- ja vearbahámiid, teakstakorpusa iešgudet šáŋŋeriid ektui. Trond Trosteruda lea erenoamážit árvvoštallan iešgudet sátneráhkadanproseassaid produktivitehta. Giella-sme huksenbargu álggahuvvui 2001:s, ja máŋggas leat leamaš mielde huksenbarggus, earret artihkkala čállit.

Artihkkala II (*Next to nothing – a cheap South Saami disambiguator*) fáddá lea lullisámi syntávssalaš analysáhtora ráhkadeapmi Constraint Grammariin. Artihkkalis leat guokte čálli, ja letne ovttasbargan buot artihkkala osiin, muhto mun čállen stuorimus oasi Constraint grammar njuolggadusain, ja Trond Trosterud dagai fas stuorimus oasi evaluerenbarggus. Artihkal giedahallá syntávssalaš analysa, muhto dat vuodđoduvvá fas morfologalaš analysii, man huksenbarggus Divvun-joavkku lullisámi lingvista lea leamaš guovddášolmmoš.

Artihkkala III (*Čállinmeattáhusaid guorran*) lean okto čállán. Artihkkalis leat guokte áššečuolmma: Makkár váttisvuodát čállin leat čállinvuogi ektui, ja man muddui davvisámegiela Divvun-prográmma máhtta veahkehit čálli. Divvun-prográmmii geavahuvvo Giella-sme analysáhtor, maid lean leamaš mielde huksemin, muhto analysáhtora heiveheapmi sátnedárkkistanprográmmii, prográmma evttohanvuogádat ja teknihkalaš čovdosat leat Divvun-joavkku bargu⁶.

Artihkal IV (*Ei intelligent elektronisk ordbok for samisk*) válddahallá vuogádaga mii riggodahtta sátnegirjji analysáhtora vehkiin, ja artihkkalis leat njeallje čálli. Sjur Moshagen dat fuomášii geavahit MacOS sátnelisttu Macdict prográmma min sátnegirjái, ja son ráhkadii sátnegirjji prototiippa. Ciprian Gerstenberger fuolahii sátnehámiid genererema sátnegirjji formáhtii ja buoridii vuogádaga nu ahte juohke sátnehápmái genererjuvvo sátnegirjeartihkal. Trond Trosterud mearridii gáldofiillaid leksikografalaš struktur-

⁵Divvun-joavkku Børre Gaup lea ráhkadan teavsttaid konverterenvuogádaga vai lea vejolaš daid analyseret. Giellatekno programmerár ja lingvista Ciprian Gerstenberger lea fas heivehan SIKORa geavahanlavtta sámegiela teavsttaide.

⁶Teknihkalaš osiid leat bargan Sjur N. Moshagen, Børre Gaup ja Tomi Pieski.

erema, ja geahččalattai sátnegirjji eanet prográmmaiguin, ja son lea dasa lassin fuolahan sátnegirjji evaluerema. Mu oassi lei lingvisttalaš bargu: heivehit analysáhtora sátnegirjái, ja sátnegirjji analysáhtorii, nugo homonymiijaid gieđahallan. Lean maiddái hábmen ja ráddjen sániid sojahanparadigmaid presentašuvnna giellaoahpahallama geahččanguovllus, ja barga maiddái ollu sátnegirjji gáldofillaiguin.

Artihkal V (*Using finite state transducers for making efficient reading comprehension dictionaries*) gieđahallá mo dahkat artihkkala IV sátnegirjevuogádaga dynámalažžan, ja artihkkalis leat golbma čáli. Ryan Johnsona oassi lei hábmet ja implementeret Neahttadigisániid teknihkalaččat. Trond Trosterud oassi lei erenoamážit geahččaladdat riehtačállinložžema, ja dasa lassin odđadárogiela ja girjedárogiela heiveheapmi lea su bargu. Artihkal hukse artihkkala IV bargui, ja mu oassi lei joatkit barggu oažžut analysáhtora ja sátnegirjefiillaid heivet oktii, dán artihkkalis dat guoská maiddái lullisámi analysáhtorii ja sátnegirjefiillaide, ja fuolahit daid gielaide lingvisttalaš presentašuvnna sátnegirjji geavahanlavttas (= eng. ‘user interface’). Lean maiddái čadahan sámegeleaid sátnegirjjiid evaluerema mii ovdanbuktojuvvo artihkkalis. Neahttadigisániid ollislašvuoda buot čállit leat hábmen ovttas.

Artihkal VI (*Generating modular grammar exercises with finite-state transducers*) muitala mo analysáhtoriin sáhtta hukset vuogádaga morfologijaoahpahallamii, ja artihkkalis leat njeallje čáli. Ryan Johnson ja Heli Uiibo leaba implementeren vuogádaga teknihkalaččat. Trond Trosterud lea erenoamážit buohtastahttán vuogádaga eará seammasullasaš vuogádagaide. Mun lean fas hábmen hárjehusgenererema ja máhcahatvuogádaga, ja jodihan evaluerema.

Artihkkaliid VII (*Improving feedback on L2 misspellings – an FST approach*) ja **VIII (*Constraints in Free-input Question-Answering Drills*)** lean čállán okto. Munnje lea leamaš váibmoáššin geavahit giellamodealla giellaoahppanprográmmaide, ja daid artihkkaliid lingvisttalaš ja pedagogalaš implementeren lea vuosttažettiin mu bargu, muhto mánga fuomášumi ja čoavddusevttohusa leat bohcciidan go lean ságastallan Trond Trosterudain.

1.4 Čállosa huksehus

Dán nákkosgirjjis leat vihtta kapihttala. Juohke artihkal presenterejuvvo sierra vuollekapihttalis juogo nuppi dahje goalmmát kapihttalis, muhto čujuhan maiddá artihkkaliidda muđui dalle go lea relevánta. Nákkosgirjjis leat maiddá eará vuollekapihttalat mat dievasmahttet artihkkaliid: Kapihttalis 2 čálán analysáhtora huksema ja ovdal namuhuvvon 1. ja 2. hástalusaid birra. Stuora oasi dán kapihttala sisdoalus gokčá artihkal I, muhto kapihttalis in vuosttažettiin govvit davvisámegeiela analysáhtora, nugo dagan artihkkalis. Dát geahččanguovlu boahtá erenoamážit ovdan kapihttalis 2.3.2 go buoh-tastahtán sámegeielaidda analysáhtoriid. Kapihttala 3 fáddá lea mo heivehit analysáhtora geavaheddjiidprográmmaide, ja dasa gullet 3. ja 4. hástalusat. Nákkosgirjji áššečuolmma hástalus badjelmearálaš genererema birra lean unnán giedahallan artihkkaliin, main leat eará áššečuolmmat, ja dan dihte suokkardalan dán beali sierra kapihttalis 3.3. Kapihttalis 4 čálán mo giellamodealla mii lea dán barggu fáddán, lea vástádus davvimáilmmi eamiálbmotgielaid dárbbuide. Loahpas boahtá konklusuvdna 5. kapihttalis.

Kapihtal 2

Giellamodealla huksen

2.1 Sámegielaid modelleren transduseriiguin

Teaksta lea sátnecoakkáldat, mii lea ordnejuvvon dainna lágiin ahte sánit ovttas dahket cealkagiid main lea sisdoallu. Juohke sátni ii leat dušše bustávvaráidu mas lea dihto mearkkašupmi, muhto sátni ovddasta dihto leksema grammatihkalaš sáni, ja ollu sánit gullet dihto sojahanparadigmii.

Dihtor lea buorre veahkkeneavvu teavsttaid giedahallamii, go dihtor bargá johtilit ja dárkilit. Jus galgá sáhttit čállit sámegiela sániid dihtoriin, de dihtor ferte maiddá hálddašit sámegielaid erenoamáš bustávid (loga dan birra Antonsen ja Trosterud (2010); Trosterud (2012)). Gávdnojit mánggalágan giellateknologiiijat sátneanalysa várás – juohke teknologiiija heive dihto geavahussii ja dihto gielaide. Sámegielain lea ollu morfologiiija – ja dan dihte sámegielain leat ollu iešguđetlágan sátnehámit maid dán oktavuodas gohčodan sátnetiipan (= eng. ‘types’⁷). Sámegielain leat unnán digitálalaš teavsttat nu ahte giellateknologalaš giedahallamii leat olámuttos unnán sánit (= eng. ‘token’). Dan dihte eat gávna buot giela sátnehámiid teavsttain, eatge gávna buotlágan gielalaš ráhkadusaid.

Nubbi čuovvumuš dás lea ahte sátnetiippaid ja sániid gorálašvuohta (= eng. ‘type-token ratio’) šaddá unni. Gorálašvuohta ii berreše leat menddo unni, go geavaha statistihkalaš bargovugiid, mat eaktudit ahte leat mánga dáhpáhusa seammá gielalaš ráhkadusas. Muhtun eará gielain, ovdamearkka dihte engelasgielas, lea unnán morfologiiija ja dan dihte

⁷Geavahan dás Jurafsky ja Martin (2008, 120) definišuvnna mas ‘types’ leat iešguđetlágan sátnehámit, ja ‘tokens’ leat buot sátnehámit.

ollu unnit sátnetiipat go sámegielain. Engelasgiela vearbalemmas leat guokte dahje golbma iešgudetlágan finihtta veabahámi (earret *to be*, mas leat vihtta) ja ii okta dain leat hui hárve. Jus dán buohtastahttit davvisámegielain, de oaidnit ahte bárrastávval vearbain leat 58 iešgudetlágan finihtta hámi, mat gokčet 45 paradigmasealla, go buot normatiiva variánttat leat mielde. Seammás leat engelasgielas stuorra teakstačoakkáldagat, nugo *Corpus of Historical American English (COHA)* mas leat 400 miljovna sáni, ja *Google Books* mas leat 155 miljárdda sáni⁸. Nu lea sátnetiippaid ja sániid gorálašvuohta ollu stuorit engelasgielas go sámegielain, ja engelasgiela heive bures giedahallat statistihkalaš bargovugiiguin.

Antonsen ja Trosterud (2010) ozaiga guovtti dábálaš vearbba sojahanhámiid engelasgiela, girjedárogiela ja davvisámegiela odđa testameanttain. Sudno guorahallan čájeha ahte teavsttain ledje buot engelasgiela *say* ja *serve* ja dárogiela *si* ja *tjene* hámit, muhto dušše 44 % *dadjat*-vearbba hámiin ja 31 % *bálvalit*-hámiin. SIKORa (= Sámi Internationála Korpus, UiT Norgga árkálaš universitehta ja Norgga Sámedikki sámi teakstačoakkáldat⁹) davvisámegiela oasis leat 25 miljovna sáni, muhto dattetge váilot earret eará ollu dábálaš vearbaid indikatiivva guvttiidlogu hámit, nugo veabahámit *dajaide* ja *boraide*, ja dát čájeha ahte ii leat vejolaš viežžat buot sátnehámiid teakstačoakkáldagas. Dan dihte sámi giellateknologiija geavaha giellamodealla, mas leat automáhtat mat huksejit buot sátnehámiid.

Teakstagiedahallamii mii dárbbasit maiddá sáni analysa, ja grammatihkalaš automáhta muitala sáni lemma ja morfologiija. Molssaevttolaš vuohki livččii máddagastin (= eng. ‘stemming’, geahča ovdamearkka dihte Jurafsky ja Martin (2008, 102)), mii ii heive bures sámegielaide, gielaid siskkáldas morfofonologalaš proseassaid dihte (Antonsen ja Trosterud, 2010).

Go dihtor giedahallá sámegiela teavstta ja sániid analysaid, iige geavat statistihkalaš lahknanvuogi, de dihtoris eai leat eará árvvoštallamat go njuolggadusat maid olmmoš lasiha dihtorprográmmii. Njuolggadus sáhtá čujuhit dihto sátnái, nugo *nissonis* dahje dan lemmai *nisu* dahje *nisson*, muhto vel buoret lea generaliseret ja čujuhit sániid mor-

⁸<https://googlebooks.byu.edu/compare-googleBooks.asp> (14.02.2018)

⁹Geavahanlakta: <http://gtweb.uit.no/korp/> Buot SIKOR-čujuhusat gusket 22.06.2017-veršuvdnii jus eai leat eará dieđut teavsttas. Artihkkalis I leat boarrásat analysat.

fologalaš dahje semantihkalaš iešvuodaide, go dalle njuolggadus guoská buot sániide main leat dát iešvuodat. Daid sániid dihtor sáhtta dovdat go sátneanalysas leat gilkorat, ovdamearkka dihte +N muitala ahte sátni lea substantiiva, +Sem/Hum ahte sátni lea olbmonamahus, ja +Loc muitala ahte sátni lea lokatiivvas. (Gilkorat leat čilgejuvvon oanáduslisttus.)

2.2 Sátnehámiid huksen

Giellatekno- ja Divvun-joavkkut leaba ovttas huksen analysáhtora golmma sámegillii. Analysáhtorat leat Giella-sme (davvisámegillii), Giella-sma (lullisámegillii) ja Giella-smj (julevsámegillii). Giellatekno lea maiddái ráhkadan Giella-smn anárašgiela várás, ovttasbarggus anárašgiela lingvisttain ja dán barggus lei maiddái *Anarâškielâ servi* mielde¹⁰.

AKU-joavku (Avointa kieliteknologiaa uralilaisille vähemmistökielille)¹¹ lea huksen nuortalašgiela analysáhtora (Giella-sms), ja *The Freiburg Research Group in Saami Studies*¹² lea huksemin bitonsámegiela analysáhtora (Giella-sje). Manjit analysáhtora huksen álggahuvvui bitonsámi dokumentašuvdnaproševttas *Från kust til kyst – Áhpegáttest áhpegáddáj*¹³ mas Giellatekno lei mielde (Sjaggo ja Trosterud, 2015). Buot dát analysáhtorat geavahit Giellatekno- ja Divvun-joavkkuid Giella-vuođdostruktuvrra, mii álkidahtta huksenbarggu ja buvttada sátnedárkkistanprográmma (Moshagen ja earát, 2013). Oanádusat *sje, sma, sme, smj, smn, sms* leat gielaid ISO-kodat 639-3. Teavsttas geavahan *Giella-smX* go čujuhan buot sámegiela analysáhtoriidda.

Analysáhtoriid morfologalaš oassi lea transduser, man sáhtta govahallat dego stuorra bálggesvuogádat mii manná giela buot vejolaš máddagiid ja geažuskombinašuvnnaid čađa, ja čatná oktii juohke sátnehámi rivttes lemmai oktan analysain. Dievaslaš transduseris lea bálggis buot giela sániide, ja transduser hilgu sániid mat eai gula gillii.

Dán kapihttalis čilgen mo dákkár transduser huksejuvvo. Dan huksenbarggus geavahit LEXC (eng. ‘Lexicon Compiler’ = leksikona kompilerinprográmma) programmerengiela

¹⁰Marja-Liisa Olthuis lea leamaš váldolingvista, ja leksikontransdusera huksenbarggus leigga maiddái mielde Suoma sámedikki bargi Miina Seurujärvi ja Anárašgiela searvvi prošektibargi Erika Sarivaara.

¹¹<http://www.ling.helsinki.fi/~rueter/aku-index.shtml> (14.02.2018)

¹²<http://saami.uni-freiburg.de/psdp/syntax/> (14.02.2018)

¹³<http://site.uit.no/pite/> (14.02.2018)

lasihit affivssaid sáni máddagii ja čatná dan oktii grammatihkalaš sániin. TWOLC (eng. ‘Two Level Compiler’ = guovttedásat kompilierenprográmma) programmerengiela rievda-da vuollálas hámiid rivtteshámat bajošhápmiin (= eng. ‘surface form’) ortográfalaš ja morfofonologalaš njuolggadusaid mielde. LEXC ja TWOLC fiillaid kompilerejuvvojit¹⁴ bineara fiilan HFST-prográmmain (eng. ‘Helsinki Finite State Transducer Technology’ = Helssega loahpalaš dilletransdusera teknologijja¹⁵) (Lindén ja earát, 2011) dahje XFST-prográmmain (eng. ‘Xerox Finite State Tool’ = Xerox loahpalaš dili reaidu¹⁶) (Beesley ja Karttunen, 2003), mii bargá johtileappot go HFST, mii fas lea rabas gáldokoda. Dan dihte sáhtta HFST heivehit ja ovttaiduhttit eará prográmmaiguin. Dehálaš lea maiddáa ahte HFST máhtta rievdadit bajit dási gilkoriid guovttedásat diakrihtalaš leavgan, ja dát ii doaimma nugo galggašii XFST-prográmmas. Dákkár gilkoriid analysáhtorat geavahit sáni goallosteami stivremii, ja dan čilgen kapihttalas 3.3.2.

2.2.1 Geážusmorfologijja huksen



Govus 2.1: Automáhta (FSA) mii hukse sátnehámiid *lávka*, *lávkan*, *girji*, *girjin*. (Artihkala VII govus 1.)

Geážusmorfologijja huksemii sáhtta geavahit automáhtaid, FSA (eng. ‘Finite-State Automaton’ = loahpalaš dilleautomáhta). Automáhtas leat bálgát mat ráhkadit buot vejolaš mátta- ja affiksakombinašuvnnaid gielas. Govvosis 2.1 lea unna automáhtaš sániiguin *lávka*, *girji*, ja das leat guokte bálgá loahpalaš dillái, nuppis lea geážus *n* ja nuppis ii leat. Boađusin leat hámit *lávka*, *girji*, *lávkan*, *girjin*.

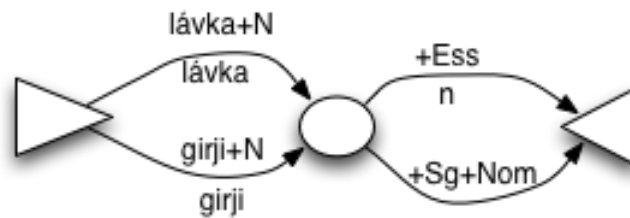
Go bargá grammatihkain, de lea ávkkálaš diehtit ahte sánit *lávkan* ja *girjin* leat essiiva-hámit, ja dan dihte mii geavahit grammatihkalaš transdusera (Beesley ja Karttunen, 2003), geahča govvosa 2.2. Dát lea dihto automáhtatiipa mas juohke sátnehámis

¹⁴Čállošis geavahan vearbba kompileret dušše go oaivvildan gáldofillaid ovttastahttit binára fiilan.

¹⁵<http://hfst.github.io> (14.02.2018)

¹⁶<http://fsmbook.com> (14.02.2018)

leat guokte ovddasteami: sátnehápmi *lávkan* ja *lávka+N+Ess*, mii lea lemma oktan grammatihkalaš gilkoriiguin, man gohčodit grammatihkalaš sátnin. Lemma lea hápmi mii ovddasta leksema analysáhtoris, dahje dan vuollemáddaga jus lea suorggádus. Transduser manná goappaš guvlui. Jus atta lea *lávkan*, de *lávka+N+Ess* lea buvttus, dahje nuppe guvlui. Dákkár transduser gohčoduvvo FST:n (eng. ‘Finite-State Transducer’ = loahpalaš dilletransduser).



Govus 2.2: Grammatihkalaš transduser (FST). Juohke sánis leat guokte ovddasteami: grammatihkalaš sátni ja sátnehápmi, ovdamearkka dihte *lávka+N+Ess* ja *lávkan*. (Artihkala VII govus 2.)

```

vielppis+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:viel'pis FALIS ;
vuoitofanas+Sem/Veh:vuoito#fatnas MALIS ;
vuoitočuorvvas+Sem/Dummytag:vuoito#čuor'vas MALIS ;
vuoššanbierggas+Sem/Tool:vuoššan#bier'gas MALIS ;
vuoššanguolli+Sem/Food:vuoššan#guolli AIGI ;
vuoššanlihti+Sem/Ctain:vuoššan#lihti AIGI ;
vuovdevánddis+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:vuovde#ván'dis FALIS ;
vuovdevárri+Sem/Plc-elevate:vuovde#várri AIGI ;
vuovdečoavžžu+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Ani:vuovde#čoav'žu BOAZU ;
vuovdi+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Hum:vuovdi ACTOR ;
vuovdi+CmpN/SgN+CmpN/SgG+CmpN/PlG+Sem/Plc:vuov'di AIGI ;

```

Govus 2.3: Govvosis lea oasáš substantiivaleksikonas. Juohke linnjás lea okta mearkkuš, ja merkošis leat golbma oasi: 1) lemma (ruoná) ja gilkorat, nugo +CmpN/SgN (goallosteami várás) ja +Sem/Ani, mii muitala semantihkas (ruoksadat). Duppálčuoggá olgeš bealde lea 2) LEXC-mátta masa gehčosat goallostuvvojit (alit) ja 3) joatkkaleksikona namma (ruoná). Mearkkuš loahpahuvo beallečuoggáin.

Giella-smX morfologalaš transduserat leat ráhkaduvvon LEXC-prográmmereingelain, ja gohčoduvvojit *leksikontransduserin* ja dain leat márga *leksikona*. Juohke leksikonas lea sierra namma, ovdamearkka dihte gávdno vearbaleksikon man namma lea LEXICON VERB. Juohke sátneluohká várás lea unnimustá okta leksikon, muhto dávjá leat ovtta sátneluohká lemmat juhkkovuvvon eanet leksikonaide vai álkit sáhtttá stivret sániid goal-

LEXICON AIGI !Bisyll. Shc
 +N: GOAHTICMP ;
 +N+Sg+Nom: K ;
 +N+Sg+Gen:X4 K ;
 +N+Sg+Acc:X4 K ;
 +N+Sg+Loc:X4s K ;
 +N+Sg+Loc:X4st NPxC ;

Govus 2.4: Joatkkaleksikonas leat golbma oasi: 1) grammatihkalaš sáni gilkorat (ruok-sadat), 2) báhcayat, X4, ja geažus, *s* ja *st* (alihat). X4 báhcaya earret eará konsonán-taguovddáža geahnohis dási. De bohtá 3) odđa joatkkaleksikona namma (ruoná). K-joatkkaleksikon lasiha vejolaš partihkala, ja NPxC-joatkkaleksikon lasiha oamastan-gehčosiid substantiivva lokatiivahápmái, ja GOAHTICMP lea joatkkaleksikonaid oktasaš goallostanleksikon. Juohke mearkkuš loahpahuvvo beallečuoggáin.

losteami dahje morfologalaš gáržžidusaid, maid giedahalan kapihttalis 3.3.2. Leksikonain galggašedje leat listejuvvon nu ollu lemmat ahte lea vejolaš genereret buot giela sániid.

Ovttá leksikonas leat dihto meari *merkošat* (= eng. ‘entries’), ja juohke mearkkuš lea čállojuvvon ovttá linnjái. Dábálaččat lea okta mearkkuš juohke lemma várás, muhto ovttá lemmas sáhttet leat eanet merkošat, jus leat variánttat. Juohke mearkkuš sisttisdoallá golbma oasi: lemma oktan gilkoriiguin ja LEXC-mátta, sirrejuvvon duppalčuoggáin, dasto bohtá sátnegaska ja de joatkkaleksikona namma. Loahpas lea sátnegaska ja beallečuokkis, nugo govvosiin 2.3. Duppalčuoggá gurut bealde leat lemma ja gilkorat, mat gullet grammatihkalaš sátnái, ja mii gohčodit dan leksikontransdusera bajit dássin. Duppalčuoggá olgeš bealde lea LEXC-mátta ja báhcagat (= eng. ‘triggers’), mat gullet ieš sátnehápmái, ja dán oasi gohčodit fas leksikontransdusera vuolit dássin. Gilkorat leat mánggalárganat: morfologalaččat, semantihkalaččat ja dasa lassin gilkorat mat stivrejit goallosteami (geahča kapihttala 3.3.3.1 ja 3.3.2).

Sánit mat čuvvot seammá sojahanminstara mannet seammá joatkkaleksikonii. Davvisámegeielas leat substantiivvain golbma máttatiippa: bárrastávval, bárahisstávval ja kontrakšuvdnamáddagat¹⁷, muhto maiddá ovtta máttatiippa siskkobealde leat erohusat sániid sojahanminstariid gaskkas. Govvosiin 2.3 njeallje bárrastávval substantiivva mannet AIGI-joatkkaleksikonii, golbma bárahisstávval substantiivva mannet MALIS-joatkkaleksikonii ja kontrakšuvdnasubstantiivvat mannet juogo FALIS dahje BOAZU joatkkaleksikonaide. Leksikonaid namat veahkehit lingvistta välljet rivttes joatkkaleksikona go lasiha

¹⁷Eará tearpmat leat vokálamátta, konsonántamátta ja kontrakšuvdnamáttá (Nickel ja Sammallahti, 2011).

odđa sániid fillaide. Davvisámegiela substantiivaleksikonain leat 90 000 merkoša mat mannet 156 joatkkaleksikonii. Lemmaleksikonas ja joatkkaleksikonas lea ovttalágan huksehus (govvosat 2.3 ja 2.4).

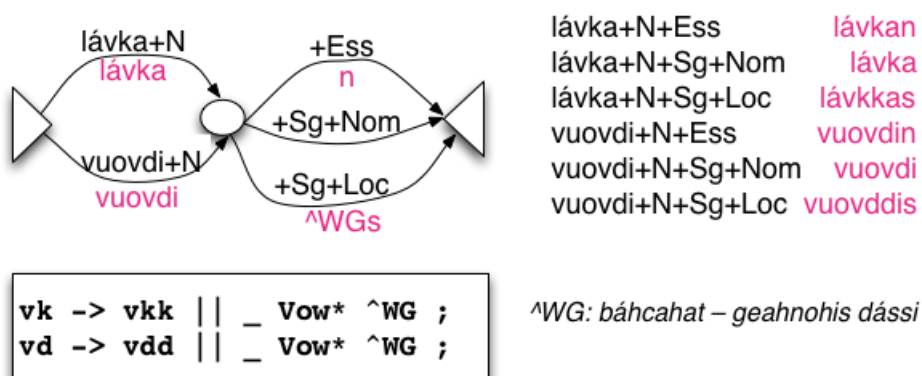
2.2.2 Morfofonologalaš rievddademiid modelleren

Leksema lea grammatihkalaš ovttat dat mas lea sihke hápmi ja sisdoallu ja mii albmana sátnehápmiin. Leksemii gullet buot sátnehámit mat ovddastit leksema iešguđet sojahemiid. Go *vuovdi*-sátnái lasihit *s* ja *n*, de ráhkadit VUOVDI-leksema lokatiiva- ja essiivahámiid go VUOVDI lea dahkkinomen: *vuovdis*, *vuovdin*. Muhto sámegielas ii leat dušše geažusmorfologiiija, muhto maiddá siskkáldas morfologalaš proseassat, ja sániin leat dávjá rievddademiid maiddá máddagis. Gávdno nubbi VUOVDI-leksema man lokatiivahápmi lea ovttaidlogus *vuovddis* (konsonántaguovddáš lea geahnohis dásis, go rievddada *vd:vdd*) ja mánggaidlogus *vuovddiin* (maiddá vokálaguovddáš rievddada, *uo:u*).

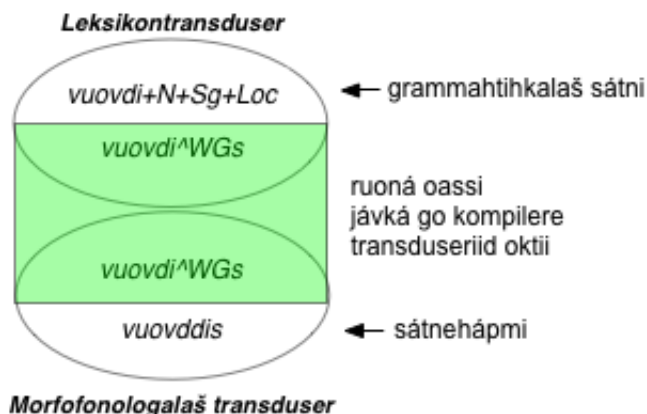
Siskkáldas morfologalaš proseassaid dihte buot Giella-smX analysáhtoriin leat guokte transdusera mat doibmet oktanaga: Nubbi hukse leksikona oktan geažusmorfologiiijain (leksikontransduser nugo čilgejuvvo kapihttalas 2.2.1), ja nubbi fuolaha siskkáldas morfologalaš rievddademiid. Dát sierra morfologalaš transduser lea čállojuvvo TWOLC-programmain (Koskenniemi, 1983; Karttunen, 1994, 2001).

Leksikontransdusera bajit dássi ovddasta grammatihkalaš sáni, *vuovdi+N+Sg+Loc*, ja vuolit dásis lea sreanga mas lea mátta ja affivssat ja dasa lassin báhcagat mat morfologalaš transduseris báhcáhit morfologalaš rievddademiid. Govvosis 2.3 leat guokte *vuovdi*-lemma. Nuppis lea semantihkalaš gilkor +Sem/Plc ja nuppis lea +Sem/Hum. Dát guokte lemma mannet goabatlágan joatkkaleksikonaide danne go dat gullet goabatlágan sojahanminstariidda. Lemma *vuovdi:vuovd'i* manná viidáseappot AIGI-joatkkaleksikonii gos lasihuvvo ee. +N+Sg+Loc: ^WG_s, mas lea geažus *s* ja symbola ^WG (eng. ‘Weak Grade’ = geahnohis dássi). Dan dihte leat leksikontransduseris guokte grammatihkalaš sáni *vuovdi+N+Sg+Loc*, main leat goappatlágan vuolit dási sreanggat: *vuovdis* ja *vuovdi^WG_s*, mat fas leat attan morfologalaš transduserii, nugo govvo-siin 2.5 ja 2.6. FST buvttusin leat sátnehámit *vuovdis* (*vuovdi+N+Sem/Hum+Sg+Loc*) ja *vuovddis* (*vuovdi+N+Sem/Plc+Sg+Loc*).

Bajábeale ovdamearkkas báhcahat lei $\hat{W}G$. Morfofonologaš rievddademiid sáht-
 tá maiddái dahkat báhcahagaid haga, ja luohttit dušše fonologaš kontekstii. Ovdamearkkat leat TWOLC-njuolggadusat mat gusket sáni lohppii. Sámegeiela fonotákssa ráddje makkár konsonánttat sáhttet leat sáni loahpas. Davvisámegeiela leksikontransduseris leat buot substantiivvat ja adjektiivvat main lea geažus *-laš*, merkejuvvon *-lašš* LEXC-máddagis, ovdamearkka dihte *sápmelaš:sápmelašš*. Morfofonologaš transdusera njuolggadus geavaha $\hat{W}G$ báhcahit *žž:čč*-rievddadeami, muhto sáni loahppa lea doarvái konteaksta rievdadit *žž:š*.



Govus 2.5: Transduser ráhkada sátnehámiid *lávka* ja *vuovdi*. Leksikontransduseris addojuvvo $\hat{W}G$ -báhcahat, mii morfofonologaš transduseris báhcaha konsonántaguovddáža rievddademiid *vk:vkk* ja *vd:vdd*. (Artihkkala I govus 1, heivehuvvon.)



Govus 2.6: Leksikontransdusera bajit dássi lea grammatihkalaš sátni, ja vuolit dásis lea sreanga, mii fas lea morfofologaš transdusera atta (LEXC-mátta, geažus ja báhcahat). Buvttus lea sátnehápmi *vuovdis*. Ruoná oassi jávká dalle go dát guokte transdusera kompilerejuvvojit oktan transduserin. (Artihkkala I govus 2, heivehuvvon.)

Anárašgiela morfofonologiija hábmen transduserin lea stuorit hástalus go davvisáme-giela morfofonologiija hábmen, earret eará danne go anárašgiela čállingiella sisttisdoallá eanet vokálarievddademiid go davvisámi čállingiella. (Davvisámegiela morfofonologiija huksema giedahallet Moshagen ja earát (2005).) Anárašgielas lea regressiiva assimilašuvdna nugo eará sáme-gielain, muhto konsonántaguovddáža dássemolsun mielddisbuk-tá vokálarievddademiid maiddái čállingiella, nugo vokálaguhkkun čuovvu konsonántta oatnuma. Anárašgielas lea maiddái konsonántaguhkkun go oanehis konsonántaguovd-dáš guhkku guhkes soggevokála ovddabealde. Konsonántaguhkkun dávjá neutralisere dássemolsuma, nugo *kissá*-sánis (= bussá): *kissá*, *kisá*, *kissáin* (nominatiiva, genitiiva, komitatiiva), muhto ii álo, nugo *päkki*-sánis (= bahkka): *päkki*, *päähi*, *pähijn*, man konsonántaguovddáš rievddada *kk:h:hh*.

Anárašgiela morfofonologiija ii lean ollislaččat dokumenterejuvvon dalle go bargu hukset Giella-smn álggahuvvui 2014 čavčča. Muhtun osiid morfofonologiijas lea čilgen Morrottaja (2007), guovddáš sojahanparadigmaid lea giedahallan Olthuis (2000) ja stuorimus sátnegirjji ja anárašgiela grammatihka leat čállán Itkonen ja earát (1988). Giellamo-dealla huksen gáibida ahte galgá modelleret buot sániid gielas, ii dušše prototiippalaš ovdamearkkaid, ja dasa dokumentašuvdna ii lean doarvái.

Test 196: Verb – tuárššuođ (Lexical/Generation)

```
[1/9] [PASS] tuárššuođ+V+Inf => tuárššuođ
[2/9] [PASS] tuárššuođ+V+Ind+Prs+Sg1 => tuáršum
[3/9] [PASS] tuárššuođ+V+Ind+Prs+Sg3 => tuárššu
[4/9] [PASS] tuárššuođ+V+Ind+Prs+Du1 => tuárššoon
[5/9] [PASS] tuárššuođ+V+Ind+Prs+Pl3 => tuárššuh
[6/9] [PASS] tuárššuođ+V+Ind+Prs+ConNeg => tuáršu
[7/9] [FAIL] tuárššuođ+V+Ind+Prt+Sg1 => Missing results: torššum
[7/9] [FAIL] tuárššuođ+V+Ind+Prt+Sg1 => Unexpected results: tuorššum
[8/9] [PASS] tuárššuođ+V+Ind+Prt+Sg3 => tuáršui
[9/9] [PASS] tuárššuođ+V+Ind+Prt+Pl3 => tuorššuu
```

Govus 2.7: YAML testenfillain iskamin anárašgiela vearbba *tuárššuođ* sojahanhámiid genererema. *torššum* sajis generejuvvo *tuorššum*.

Anárašgiela FST huksen lea buorrin ovdamearkan das mo sáhttá bargat dalle go ii leat doarvái dokumentašuvdna ovdalگو álggaha barggu. Anárašgiela lingvista lei guovd-dáš olmmoš dán barggus, ja huksiimet YAML formáhtas testenfillaid main leat oktiibuot 22 435 grammatihkalaš *sátni:sátnehápmi*-pára (eng. ‘YAML Ain’t Markup Language’

= YAML ii leat markerengiella). YAML-fiillat ledje veahkkin go ráhkadeimmet njuolggadusaid, muhto maiddá dárbbaslaččat regrešuvdnatestemii (govus 2.7). Go leat ollu morfofonologalaš rievddadeamit, de lea regrešuvdnatesten ávkkálaš juohke TWOLC-fiilla rievdadusa maŋjel. (YAML-testen lasihuvvui Giella-vuoddostrukturuvrii easka 2013:s, nu ahte Giella-sme, Giella-smj ja Giella-sma analysáhtorat huksejuvvojedje YAML-testen-fiillaid haga.)

Várrogasvuoda dihte diktit morfofonologalaš transdusera lasihit konsonántta dahje vokála dušše jus lea mearka mii čájeha ahte dása sáhtta boahit bustávva, arkifonema. Dábálaččat arkifonema rievdá kránnjábustáva kopijjan, muhto ii álo. Substantiivva *saje* (= sadji) LEXC-mátta lea dákkár: *sa[^]RVj[^]RCe* mas leat arkifonemat [^]RV (eng. ‘Root Vowel’ = vokálaguovddáš) ja [^]RC (eng. ‘Root Consonant’ = konsonántaguovddáš). Njuolggadusaid bokte arkifonemat sáhttet rievdat bustávvan, ja njuolggadusat vuodđuduvvet kontekstii, ja nu analysáhtor generere hámiid nugo *saijeen*, *soojij*. Anárašgiela transduseris morfofonologalaš njuolggadus ii sáhte lasihit odđa segmeantta nollas, ja dainna eastada lasiheames bustáva gokko ii galgga. Davvisámegiela analysáhtoris lea okta ja julevsámegiela analysáhtoris leat vihtta njuolggadusa mat lasihit odđa segmeantta nollas.

```
"Consonant shortening and gradation kk:k and kk4:k"
k:0 <=> Vow: _ [k:l k4:] Vow: [%^WG:|^CSH:] ([%^RLEN:|^RVSH:]) %> ;

"Gradation for kk4"
k4:h <=> Vow: k: _ Vow: (Cns:) %^WG: ;

"kk:hh gradation for kk4"
k:h <=> Vow: _ k4:h Vow: (Cns:) %^WG: %^CLEN: ;

"Root vowel u lengthening"
%^RV:u <=> u _ Cns:+ Vow: (Vow:) (Cns:+) Triggers:* %^RLEN: ;
```

Govus 2.8: Sihke [^]WG ja [^]CSH báhcáhit *kk:k* rievddadeami, muhto dušše [^]WG *kk4:h* rievddadeami. [^]CLEN guhkida geahnohis konsonántta *h:hh* ja nu ožžot sánit njeallje iešguđetlágan konsonántaguovddáža: *päkki:páákán:päähi:pahhijn*. [^]RLEN guhkida vokálaguovddáža, nugo *kukká:kuukán*, go LEXC-mátta lea *ku[^]RVkká*.

Anárašgiela dássemolsumis muhtun gemináhtain leat guokte vejolaš rievddadeami, nugo *kk:h* ja *kk:k*, *päkki:päähi* (= báhkka) ja *kukká:kuká* (= lieddi). Dákkár dáhpáhusain ovdaválljejuvvon rievddadeapmi lea merkekeahhtá, ja erenoamáš rievddadeapmi merke-

juvvo arkifoneman, dás *kʰ*, ja nu oažžut guokte vejolaš rievddadeami, *kkʰ:h* ja *kk:k*. TWOLC-njuolggadusat leat čilgejuvvon govvošis 2.8.

Anárašgiela transduseris leat geavahan 16 báhcahaga. Tabeallas 2.1 lean čilgen ja adán ovdamearkkaid das mo báhcahagat rievdadit vokálaid ja konsonánttaid guhkkodaga ja kvalitehta.

Báhcahat	Posišuvdna	Makkár rievdan	Ovdamearka
^RLEN	vokálaguovddáš	guhkkun	<i>a > aa</i> , <i>alge:aalgan</i> ‘bárdni’
^RVSH	vokálaguovddáš	oatnun	<i>aa > a</i> , <i>vaarâ:varâtáá</i> ‘várri’
^VBACK	vokálaguovddáš	kvaliteahtta	<i>ä > a</i> , <i>vâzziđ:vâzzim</i> ‘vâzzit’
^VHIGH	vokálaguovddáš	kvaliteahtta	<i>ä > á</i> , <i>vâzziđ:vázáččij</i>
i2	vokálaguovddáš	kvaliteahtta	<i>iä > e</i> , <i>siämmoo:semmup</i> ‘hoavrristit’
^CLEN	kons.guovddáš	guhkkun	<i>h > hh</i> , <i>pâahi:pahhijn</i> ‘báhkki’
^WG	kons.guovddáš	dássemolsun	<i>tt > đ</i> , <i>ruttâ:ruudâ</i> ‘ruhta’
^SLEN	soggi	guhkkun	<i>e > ee</i> , <i>alge:algeest</i>
^SVSH	soggi	oatnun	<i>ee > e</i> , <i>eemed:emedân</i> ‘eamit’
^SVLOW	soggi	kvaliteahtta	<i>u > o</i> , <i>áldu:áldoin</i> ‘áldu’
^ÁI	soggi	kvaliteahtta	oktan ^WG:in <i>á > i</i> , <i>kuáhtás:kyevtis</i> ‘guovttis’
^ÁE	soggi	kvaliteahtta	oktan ^WG:in <i>á > e</i> , <i>luámmán:lyeme</i> ‘luomi’
u2	soggi	kvaliteahtta	<i>u > o</i> ja
	vokálaguovddáš	kvaliteahtta	<i>ää > á</i> , <i>čäällid:čallud:čallojii</i> ‘čállit’
^CSH	kons.guovddáš	oatnun	<i>rr > r</i> ja
	vokálaguovddáš	oatnun	<i>aa > a</i> , <i>aarrân:arânân</i> ‘árran’
^EA	vokálaguovddáš	kvaliteahtta	<i>e > iä</i> ja
	soggi	kvaliteahtta	<i>i > á</i> , <i>mecci:miácán</i> ‘meahcci’
^FCD	loahppakons.	sihkkun	sihkkut <i>n</i> , <i>luámmán:lyeme</i>

Tabella 2.1: Anárašgiela transdusera báhcahagat ja daid morfofonologalaš rievddadeamit. i2 ja u2 leat sokki arkifonemat, mat dasa lassin báhcahit vokálaguovddáža rievddadeami.

Morottaja (2016) válddáhalla Giella-smn leksikona- ja morfofonologalaš transduseriid vuorrováikkuhusa. Son evttoha molssaevttolaš huksenvuogi mas lea eanet morfofonologalaš dieđut LEXC-máddagis. Bárrastávval máddagiid soggevokála merkejuvvo arkifone-

man, vejolaččat oktan báhcáhagain, ja TWOLC-njuolggadusat rievdadit daid bajošhámat vokálan. Soggevokála rievddadeapmi váikkuhivččii maiddá vokálaguovddáža kvalitehtii. Boadusin TWOLC-njuolggadusat govvidivčče anárašgiela morfofonologiiija čielgaseappot, ja leksikontransduseris livčče vátnásat joatkkaleksikonat go dálá analysáhtoris.

Go hukse TWOLC-filla, de sáhtá čállit njuolggadusaid goitge guovttelágan prinsihpaid mielde, juogo listet buot bustávvarašuvnnaid, ja lasihit konteavsttaid daidda, dahje listet relevánta konteavsttaid ja lasihit bustávvarašuvnnaid daidda. Bargovuogi válljen váikkuha sihke njuolggadusaid lohku ja man johtilit transduser kompilerejuvvo leksikontransduseriin. Bustávvarašuvdnalisten gáibida eanet konteakstanjuolggadusaid, ja dan dihte šaddá unnit bargu prográmmii buohtastahttit bustávvarašuvnnaid konteavsttaid, ja kompileren manná johtileappot go konteakstalistenvugiin. (Trosterud ja Uibo, 2005).

2.2.3 Molssaevttolaš vuogit

2.2.3.1 Morfofonologalaš ráidonjuolggadusat

Molssaevttolaš lahkonanvuohkin livččii čállit morfofonologalaš rievddadannjuolggadusaid leksikontransduseris, nu gohčoduvvon buhttenjuolggadusaid (= eng. ‘rewrite rules’) (Beesley ja Karttunen, 2003, 478–526). Dan sajis go defineret guovttedásat bustávvarašuvnnaid, de čállit segmeanttaid buhttenjuolggadusaid, mat leat vuodđuduvvon olgeš ja gurutbeale konteavsttaide. Buorre bealli das lea vejolašvuolta buhttet guokte dahje eanet bustáva ovttain njuolggadusain, go guovttedásat transduser gáibida ovtta njuolggadusa juohke bustáva relašuvnna namas. Ovdamearkka dihte lea anárašgiela *mecci*-substantiivva (= meahcci) illatiivahápmi *miäcán*. Leksikontransdusera mátta lea $me \hat{R}Vcci$, ja guovttedásat transduser dárbbáša guokte njuolggadusa rievddadit $e \hat{R}V:iä$, muhto ráidovuogádagas sáhtá dahkat dan ovttain buhttenjuolggadusain.

Buhttenjuolggadusaiguin morfofonologiiija giedahallojuvvo vugiin mii sulastahtá generatiiva fonologiiija. Njuolggadusaid ortnet lea dehálaš danne go juohke njuolggadusa buvttus lea čuovvovaš njuolggadusa atta, ja nu njuolggadusat doibmet ráidun. Dáinna lágiin bustávva sáhtá rievdat odđasit, go das eai leat dušše guokte dási nugo TWOLC-prográmmas, mas buot njuolggadusat doibmet bálddalasat.

Go morfofonologijja lea kompleaksa, de lea hástalus čállit ráidonjuolggadusaid rivttes ortnega mielde. Davvisámegielas lea ovdamearkka dihte konsonántačoahkki mii gáibida *ht:ð*-njuolggadusa, mii ii galgga váikkuhit jus konsonántačoahkki gullá *htt:ht*-rievddadeapmái. Dávjá dáhpáhuvvet guokte dahje golbma morfofonologalaš proseassa oktana, nugo *boahtit:bođii* main lea sihke konsonántaguovddáža dássemolsun ja regressiiva assimilašuvdna, go vokálaguovddáža diftonjja njuolgá sojahangehčosa *ii* dihte. Kompilierenleavttu ja segmeanttaid buhttenjuolggadusaid dihte Trosterud ja Uibo (2005) evttoheaba geavahit sihke buhttenjuolggadusaid ja guovttedásat relašuvdnanjuolggadusaid seammá analysáhtoris. Muđui lea relevánta ahte kompilierenprográmma Foma (Hulden, 2009), mii kompilere johtileappot go HFST, ii kompilere TWOLC-njuolggadusaid, muhto buhttenjuolggadusaid. Foma lea dál mielde Giella-vuoddostruktuvrras.

2.2.3.2 Buot morfofonologijja lea leksikontransduseris

Giella-vuoddostruktuvrras lea maiddá suomagielá FST. Giella-fin lei álgoálggus suomagielá Omorfi-analysáhtora (Pirinen, 2015) kopijja, muhto dasa leat earret eará lasihuvvon ollu sánit ja dynámalaš goallosteapmi, ja gilkorat leat heivehuvvon Giella-vuoddostruktuvrii. Morfologalaš struktuvra lea dattetge bisson. Buot morfofonologalaš rievddadeamit dáhpáhuvvet LEXC-prográmmas. Substantiiva *vaate* (= bivttas) sojahettiin sihke konsonántaguovddáš ja soggevokála sáhttet guhkkut: *vaatetta*, *vaatteen*. Sáni mátta transdusera leksikonfiillas lea *vaat* ja nu sáhttet sihke *t:tt* ja *e:ee* rievddamat dáhpáhuvvat LEXC:s, joatkkaleksikonaid bokte.

Suomagielas lea vokálaharmonijja, ja giela gávcci vokálas, leat guđa vokálas morfofonologalaš ráddjejumit, ja dat leat golbma ovdavokála *y*, *ö*, *ä* ja golbma manjevokála *u*, *o*, *a*. Vokálaharmonijja guoská sáni máddagii: Sánis ii sáhte leat sihke ovdavokála ja manjevokála, muhto dat guoská maiddá morfologijjii. Jus máddagis lea ovdavokála, de gehčosis ii sáhte leat manjevokála, ja jus máddagis leat manjevokálat, de gehčosis ii sáhte leat ovdavokála. Gehčosiin leat dan dihte guokte variántta, ovdavokálain dahje manjevokálain.

Vokálaharmonijja livččii vejolaš čoavdit TWOLC-prográmmain, ja ráhkadit njuolggadusa mii rievdadivččii gehčosa ovdavokálaid manjevokálan *ä:a*, *y:u*, *ä:o* jus máddagis

```

vaate+N:vaat  NOUN_LAITE  ;
käsité+N:käsit  NOUN_KÄSITE  ;

LEXICON NOUN_LAITE
O:e  NOUN_BACK_NOMINATIVE  ;
O:tee  NOUN_BACK_SINGULARS  ;
O:te  NOUN_BACK_PLURALS  ;
O:et  NOUN_BACK_PARTITIVE_TA  ;
...

LEXICON NOUN_KÄSITE
O:e  NOUN_FRONT_NOMINATIVE  ;
O:tee  NOUN_FRONT_SINGULARS  ;
O:te  NOUN_FRONT_PLURALS  ;
O:et  NOUN_FRONT_PARTITIVE_TÄ  ;
...

LEXICON NOUN_BACK_PARTITIVE_TA
+Sg+Par:%>ta  NOUN_BACK_POSS_AN_OPT  ;

LEXICON NOUN_FRONT_PARTITIVE_TÄ
+Sg+Par:%>tä  NOUN_FRONT_POSS_ÄN_OPT  ;

```

Govus 2.9: Suomagiela analysáhtor Giella-fin: Buot morfofonologalaš rievddadeamit dáhpáhuvvet LEXC-prográmmas, ja dat mielddisbuktá buohtalas joatkkaleksikonaid mat mannet juogo *a-* dahje *ä-*geážusleksikonaide.

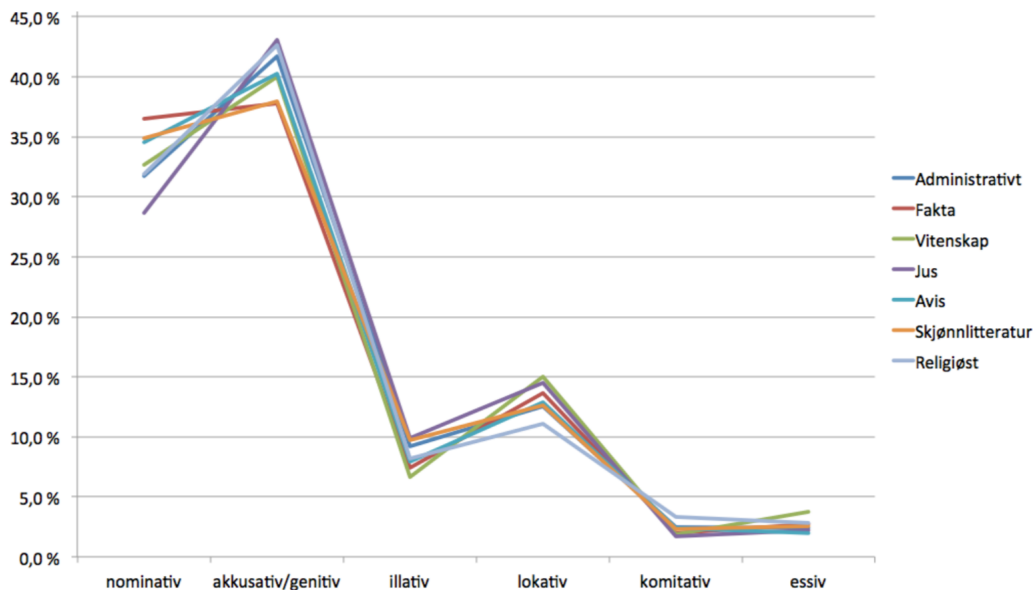
lea mañjevokála *u*, *o*, *a*. Muhto Giella-fin analysáhtoris vokálaharmoniija lea dahkkon TWOLC-prográmma haga: Morfologalaš joatkkaleksikonain manná bálggis juogo *a-* dahje *ä-*geážusleksikonaide. Nu ahte NOUN_LAITE-joatkkaleksikonas manná bálggis NOUN_BACK_PARTITIVE_TA-leksikonii ja boadus lea *vaate+N+Sg+Par:vaatetta*. NOUN_KÄSITE-joatkkaleksikonas fas manná bálggis NOUN_FRONT_PARTITIVE_TÄ-leksikonii mii fas addá *käsité+N+Sg+Par:käsitettä* (= doaba). Dáinna lágiin čovdojuvvojit buot morfofonologalaš rievdamat, ja TWOLC-fiillas lea dušše okta njuolggadus, mii lasiha goallossázu vokálaid gaskii sániid goallosteami oktavuodas. Muhto dát lahknanvuohki addá bálddalas joatkkaleksikonaid: nubbi mañjevokálaid várás, ja nubbi eará vokálaid várás (govus 2.9).

2.3 Buot sániide addit analysa

2.3.1 Sáni olggobeale ja siskkobeale hápmi (Art. I)

Artihkal I (*Ord sett innafra og utafra – en datalingvistisk analyse av nordsamisk*) suokkar-dallá man bures Giella-sme analysere SIKORa davvisámi teavsttaid. Sáhttago analysa muitalit sátneráhkadeami produktiviteahta birra? Maid analysa muitala substantiivvaid kásusiid juohkáseamis ja vearbbaid finihitta sojahanhámiid juohkáseamis?

Artihkkalis bohtá ovdan ahte Giella-sme addá morfologalaš analysa 97,8 % korpusa sániide. Dalle eai leat mielde dovdameahttun sánit maid dárogiebla analysáhtor dovdá dárú sitáhtan, eaige bajilčállagat mat leat čállojuvvon versálaiguin, danne go analysáhtor ii leat heivehuvvon daidda. Bargu dahkat vejolažžan analyseret sániid čállojuvvon versálaiguin ja earuhit daid inišialaoanádusain ja namain, ii leat vuoruhuvvon danne go dát čovdojuvvo geavaheddjiidprográmmain.



Govus 2.10: Kásusiid juohkáseapmi SIKORa šánjeriid mielde mat leat hálddahus-, fákta, diedalaš-, láchka-, aviisa-, čáppagirjjálašvuoda- ja oskuteavsttat. Juohke sárggis čájeha ovtta šánjera kásusjuohkáseami (oktiibuot 100 %). (Artihkkala I govus 7).

Listu mas leat 2,2 % dovdameahttun sánit, sisttisoallá eará gielaid sitáhtaid (mat eai leat dárú sitáhtat), ja dasa lassin čállinhámiid mat eai leat norpma mielde čállojuvvon, eaige lasihuvvon deskriptiivva analysáhtorii. Listtus leat muhtun dovdameahttun

namat ja muđuige dohkálaš sámegeiela sánit mat váilot analysáhtoris, muhto das leat maiddá sánit mat leat originálafillaidda heajos konverterema bohtosat. Dávjá guoská dát iešgudet vugiide čállit sámegeiela erenoamáš bustávaid ovdalگو Unicode-standárda¹⁸ šattai dábálažžan, ja váttisvuodaide OCR-lohkamis (eng. ‘Optical Character Recognition’ = optihkalaš mearkaáican) ja pdf-fillaidda konverteremis teakstan. Váttisvuodát čovdojuvvojit muhtun muddui go lasihit njuolggadusaid juohke dokumeantta konverterenfilii, dan sajis go divvut origináladokumeanttas.

Vai analysa sáhtta mitalit giela grammatihka birra, de Giella-sme vällje rivttes morfologalaš analysa konteavstta mielde (eanet dan birra kapihttalis 2.4). SIKOR-teavsttat leat juhkkovuvvon čiežan šáŋŋerin, ja artihkal giedahallá man muddui leat grammatihkalaš erohusat šáŋŋeriid gaskkas. Korpusa šáŋŋerat leat aviisa-, osku-, fákta-, hálddahus-, láhka-, diedalaš- ja čáppagirjjálašvuodateavsttat. Artihkkala guorahallan čájeha ahte kásusiid juohkáseamis ii leat nu stuora variašuvdna šáŋŋeriid gaskkas, ja dainna lágiin olles korpus addá buori muddui rivttes gova kásusgeavaheamis, šáŋŋeriin beroškeahhtá (govus 2.10). Juridihkalaš teavsttat spiehkastit eanemusat eará šáŋŋeriin, dain lea unnit oassi nominatiivahámit ja stuořit oassi akkusatiiva-genitiiva-hámit go eará šáŋŋeriin. Vearbbaid áigejuohkáseamis ja persovvna-logu sojahaemis leat fas ollu stuořit erohusat šáŋŋeriid gaskkas. Go buohtastahtta vearbbaid dávjodaga, de oaidná ahte čáppagirjjálašvuodas geavahuvvojit eanet vearbbat go eará šáŋŋeriin, ja čáppagirjjálašvuodas leat maiddá čielgasit eanemus iešgudetlágan vearbbat.

Artihkal giedahallá maiddá man muddui Giella-sme máhtta čilget sátneráhkadeami, mat leat sániid goallosteammi ja suorggideammi. Ollu goallossáni ja suorggideammit leat leksikaliserejuvvon, mii mearkaša ahte sátnái lea sierra mearkkuš leksikonfillas, vaikko sátni maiddá oazžu dynámalaš analysa. Sátni oazžu dynámalaš analysa go analysáhtor ieš hukse goallossáni go lea goallostanbálggis sáni mearus oasis vuoddoassái, dahje analysáhtor suorggidahtta sáni go lasiha suorgása sáni máddagii. Dainna lágiin sáhttet sáni analysas leat guokte dahje golbma lemma, nugo *čuovvuleapmi* suorggádusa analysas leat lemmat *čuovvuleapmi*, *čuovvulit*, *čuovvut*. Dušše buot produktiivvamus suorggádusat leat dynámalaččat.

¹⁸<https://www.unicode.org/standard/standard.html> (14.02.2018)

10,7 % korpusa substantiivvain ožžo dynámalaš goallostananalyisa. Guorahallamis boahtá ovdan ahte sáhtášii leat stuorit lohku, danne go dasa lassin 0,6 % substantiivvain leat goallossánit maid mearusoasi ja vuodđooasi analysáhtor dovdá, muhto leksikontransduseris ii leat goallostanbálggis daid gaskka.

Artihkkalis letne maiddái árvvoštallan muhtun sátneráhkadanproseassaid produktiviteahta, ja goallosteampi lea ollu eanet produktiiva go suorggideampi.

2.3.2 Sámegeielaid morfologalaš analysáhtoriid buohtastahttin

Buot Giella-smX analysáhtoriin lea sihke leksikontransduser ja morfofonologalaš transduser. Muhto dannego čállingielain eai leat seammá ollu morfofonologalaš proseassat, de morfofonologalaš transduseriin eai leat seammá ollu njuolggadusat. Maiddái morfofonologalaš njuolggadusaid organiseren váikkuha daid lohku, nugo čilgejuvvon kapihttalas 2.2.2, ja ollu rievddademiid sáhtta maiddái čoavdit eanet joatkkaleksikonaid bokte, nugo suomagiela Giella-fin čájeha.

Tabeallas 2.2 lean buohtastahttán Giella-smX morfofonologalaš transduseriid. Lullisámi transduseris (Giella-sma) leat sihke unnimus njuolggadusat ja unnimus konteavsttat. Lullisámegeielas lea regressiiva vokálaassimilašuvdna, muhto ii leat konsonántaguovddáža dássemolsun. Njuolggaduslogut čájehit muđui ahte davvi- ja julevsámi čállingielain leat unnit morfofonologalaš rievddadeamit go anáraš- ja nuortalašgiela čállingielain.

Miellagiddevaš lea oaidnit ahte nuortalašgiela analysáhtoris (Giella-sms) leat eanet joatkkaleksikonat go eará sámegeielaid analysáhtoriin, muhto maiddái eanet kompleksiteahtta morfofonologalaš transduseris (TWOLC): Njuolggadusat, konteavsttat, arki fonemat ja báhcagat leat ollu eanet go eará gielaid fiillain. Dát speadjalastá nuortalašgiela kompleksa čállingiela, mii geavaha 35 bustáva ja golbma dárkaga čállit árbevirolaš sániid, ja dasa lassin vihtta bustáva erenoamážit loatnasániide. Čállingielas leat 12 diftonja, ja buot logi vokála sáhtta čállit sihke oanehažžan ja guhkin. Morfofonologalaš rievddadeamit leat konsonánttaid dássemolsun ja vokálaguovddáža guhkkun ja oatnun. (Sammallahti ja Mosnikoff, 1991; Feist, 2010).

Gielaid morfofonologalaš kompleksiteahta ektui lean tabellii 2.2 bidjan davvi-, julev- ja bitonsámi analysáhtoriid oktan joavkun, ja anáraš- ja nuortalašgiela analysáhtoriid

	TWOLC				LEXC
	Njuolggadusat (konteavsttat)	Variá- belat	Arkifonemat/ báhcahagat	Kompileren- áigi	Subst. joatkkaleks.
I					
Davvisámegiela					
Giella-sme	112 (114)	98	35	0m 14,7s	156
Julevsámegiela					
Giella-smj	65 (262)	11	26	0m 5,5s	117
Bitonsámegiela					
Giella-sje	27 (36)	13	16	0m 2,5s	17
II					
Anárašgiela					
Giella-smn	90 (250)	37	34	1m 33,1s	153
Nuortalašgiela					
Giella-sms	115 (419)	3	59	0m 3,2s	318
III					
Lullisámegiela					
Giella-sma	50 (80)	28	23	0m 0,4s	127

Tabealla 2.2: Giella-smX morfofonologalaš transduseriid buohtastahttin. Transduserat leat juhkkojuvvon golbman joavkun čállingiela morfofonologalaš rievddademiid mielde. I: Konsonánttaid dássemolsun ja unnán vokálaid regressiiva assimilašuvdna II: Konsonánttaid dássemolsun ja vokálaid regressiiva assimilašuvdna III: Vokálaid regressiiva assimilašuvdna. Variábellogus leat mielde dušše bustávvelašuvnnaid variábelat, eai konteavsttaid variábelat. Kompileren lea dahkkon MacBook-Pro mašiinnas mas lea 3,5 GHz Intel Core i7 ja 16 GB RAM. Olgeš ceakkočuolddas leat substantiivvaid joatkkaleksikonaid logut nugo leat leksikontransduseriin. (Analysáhtorat 09.01.2018.)

nubbin joavkun. Julevsámi ja nuortalašgiela analysáhtoriin leat listejuvvon bustávvelašuvnnaid, ja daidda lasihuvvon konteavsttat. Davvisámegiela, anárašgiela ja bitonsámegiela analysáhtorat čuvvot fas eará vuogi: Konteavsttat leat listejuvvon, ja daidda leat variábeliid bokte lasihuvvon bustávvelašuvnnaid. Tabellat 2.2 oaidná ahte Giella-smj ja Giella-sms TWOLC-njuolggadusat, main lea bustávvelašuvdnalisten, kompilerejuvvojit ollu johtileappot go transduseriin main lea konteakstalisten, nugo árvaluvvon kapihttalit 2.2.2. Nuortalašgiela analysáhtor ii geavat variábeliid bustávvelašuvnnaid, muhto addá njuolggadusaid juohke bustávvelašuvdnii. Dát dahká ahte TWOLC-programmii šaddá unnán bargu njuolggadusaid viidodaga buohtastahttit, ja filla njuolggadusat kompilerejuvvojit hirbmat johtilit, dušše 3,5 sekunddas. Jus buohtastahtta anárašgielain, de filla kompileren ádjána olles 1 minuhta 33,1 sekunda. Anárašgiela analysáhtoris leat njuolg-

gadusat main leat erenoamáš ollu vejolaš bustávvat listejuvvon variábeliidda. Njuolggadus mas leat eanemus vejolašvuodat variábeliid ektui, dárbbáša 41 sekundda kompilerejuvvot.

Garvin dihte morfofonologalaš njuolggadusaid, de sáhtta ráhkadit eanet joatkkaleksikonaid. Tabealla 2.2 olgeš ceakkočuoldda logut mitalit galle iešgudetlágan joatkkaleksikonii substantiivvat mannet, čájehan dihte huksehusaid erohusaid. Davvisámegiela analysáhtoris leat eanet joatkkaleksikonat go julevsámegiela analysáhtoris, vaikko gielain leat sullii seammá ollu morfofonologalaš proseassat. Dát stuorra joatkkaleksikonlohku boahká Divvun-joavku barggus davvisámegielain hehtten dihte badjelmearálaš genererema ja garvin dihte sátnehámiid mat gokčēt čállinmeattáhusaid (dan birra čálán kapihtalis 3.3.3).

Analysáhtor	Substantiivvat		Vearbbat		Adjektiivvat		Advearbbat		
	dynámalaš		dynámalaš		dynámalaš		dynámalaš		sátnegihput
	-	+	-	+	-	+	-	+	
Davvisámegiela									
Giella-sme	17 635	75 112	11 970	3350	5523	3896	3604	305	368
Anárašgiela									
Giella-smn	8523	6761	4805	891	2101	403	1132	92	16
Lullisámegiela									
Giella-sma	12 972	4495	8945	2067	3649	545	1990	78	37
Julevsámegiela									
Giella-smj	13 754	1220	6458	1045	2175	564	1353	226	26
Nuortalašgiela									
Giella-sms	8100	13 625	3869	829	1962	1512	1324	18	375
Bišonsámegiela									
Giella-sje	2440	0	1648	0	138	0	212	0	0

Tabealla 2.3: Giella-smX leksikontransduseriid buohtastahttin: Galle iešgudetlágan lemma leat dihto leksikonain. Olgeš ceakkočuoldda logus (+) leat dušše sánit mat maiddá ožžot dynámalaš analysa, gurut logus (-) dat eai leat mielde. Advearbbain leat sátnegihput sierra lohkun. Bišonsámi analysáhtoris ii leat vuos dynámalaš sátnehuksen. (Analysáhtorat 09.01.2018.)

Tabealla 2.3 buohtastahtta Giella-smX leksikontransduseriid substantiiva-, vearba-, adjektiiv- ja advearbafillaid sturrodagaid. Substantiivvaid ja adjektiivvaid olgeš ceakko-

čuolddas leat sánit mat leat leksikaliserejuvvon vaikko dat ožžot dynámalaš analysa, juogo goallosteapmin dahje suorggádussan, dahje maiddá i vearbaabessiivan jus lea advearba. Goallosteapmit nugo *bajásšaddat* ja *lávžegeahčen*, eai oaččo dynámalaš goallostananalysa, danne go ii leat goallostanbálggis advearbaleksikona ja eará leksikonaid gaska.

Davvisámegiela analysáhtoriin lea bargojuvvon eanemusat, ja tabeallas 2.3 dan analysáhtoris leat eanemus merkošat juohke ceakkočuolddas. Dušše sátnegihput mat doibmet advearban, nugo *ovddos guvlvi*, leat veaháš eanet nuortalašgiela analysáhtoris go davvisámegiela analysáhtoris. Stuurimus analysáhtoriid gaskasaš erohus lea substantiivvaid loguin, davvisámi substantiivafillas leat measta 93 000 merkoša, sullii guđa geardde eambo go eará analysáhtoriin, earret nuortalašgielas mas leat measta 22 000 merkoša.

Davvisámegielas leat erenoamáš ollu leksikaliseremat ja váldosivva lea heiveheapmi geavaheddjiidprográmmaide. Davvisámegiela sierra sátnedárkkistanprográmma mii lea heivehuvvon Ávvir-aviissa bordinprográmmii, doaibmá dynámalaš goallosteami haga, ja dan dihte leat korpusteavsttaid vuodul lasihuvvon ollu leksikaliserejuvvon goallosteapmit erenoamážit substantiivafillii ja adjektiivafillii. Dihtorjorgaleapmái lea dárbu leksikaliseret sániid oazžun dihte idiomáhtalaš ulbmilgiela, ja jorgalanvuogádagain lea davvisámegiella gáldogiellan.

Eanemusat davvisámi leksikonfiillaide leat lasihuvvon sánit mat eai leat duodaštuvvon geavahusas teavsttain, muhto leat tearbmalisttuin mat bohtet ovddeš Giellalávdegottis ja dálá Sámi Giellagáldus. Ollu sánit mat leat ráhkaduvvon eanetlogugielaid tearbmalisttuid vuodul, eai leat vuos bohtán geavahussii sámeigiela teavsttain. Davvisámegiela fiillaide leat maiddá lasihuvvon automáhtalaččat genererejuvvon vierissánit, nugo *konskriberet*, *bramaisma* maid vuoddu leat dárur vierissánit *konskribere*, *bramaisme*. Go iskan korpusanalysa lemmaid ektui, de bohtá ovdan ahte 30 % analysáhtora davvisámi substantiiva- ja vearbalemmain eai gávdno SIKORis. Dása mun máhcan kapihttalis 3.

Bitonsámi analysáhtor lea unnimus, eaige dasa leat lasihuvvon dynámalaš goallostan- dahje suorggidanbálgát (tabealla 2.3). Leksikontransdusera lemmat vižžojuvvojit dadista- ga bitonsámi leksikálalaš diehtovuodus, ja bargu hukset teakstakorpusa jietnafillaid ja his- torjjálaš teavsttaid vuodul, lea easka álggahuvvon (Wilbur, 2018). Go SIKORis eai leat vuos teavsttat maiguin sáhtášin iskat analysáhtora beaktivuoda, de lea váttis Giella-sje

buohtastahttit eará Giella-smX analysáhtoriidda, ja dan dihte Giella-sje ii leat mielde tabeallas 2.4.

Anárašgiela ja nuortalaš giela analysáhtoriin leat muđui unnimusat sánit mat eai leat goallosteamat dahje suorggádusat, ja dat leatge Giella-smX oddaseamos analysáhtora gaskkas. Tabealla 2.3 substantiivalogus eai leat namat (proprat). Juohke analysáhtoris leat sullii 50 000 nama, ja dain leat 32 000 nama sierra oktasaš fiillas mii genererejuvvo buot sámegeielaide. Oktasaš fiillas leat namat nugo *Sjur* ja *Colorado*, mat sáhttet geavahuvvot buot sámegeielaid teavsttain.

Tabeallas 2.4 lean buohtastahttán man bures Giella-smX analysáhtorat máhttet analyseret SIKOR-teavsttaid. Davvisámegeiela analysáhtoris lea čielgasit buoremus gokčanmearri, 97,6 %. Anárašgiela analysáhtora gokčanmearri lea buoret go julevsámegeiela ja lullisámegeiela analysáhtoriid gokčanmearri, vaikko anárašgiela analysáhtoris leat unnit lemmat (tabeallas 2.3). Anárašgiela korpusis dáidá leat unnit variašuvdna go eará korpusiin go stuorra oasi lea ovtta čáli bargu (Olthuis ja earát, 2016). Nuortalašgiela analysáhtor ii máhte analyseret go 84,1 % korpusis, ja sivva sáhttá leat ahte Giella-smX analysáhtoris leat unnit lemmat go eará Giella-smX analysáhtoriin. Muđui lea čállingielas ollu variašuvdna ja mángii lea váttis giela dárkagiid konverteret.

Sániin lea dábálaččat eanet go okta analysa, ja disambigueren, rivttes analysa válljen konteavstta ektui, čilgejuvvo kapihttalas 2.4. Vaikko davvisámegeiela korpusa analysas eai leat go 2,6 % dovdameahttun sánit, de dát sánit gusket ollu cealkagiidda. Artihkkala I analysii geavaheimme dušše cealkagiid main juohke sánis lea analysa, ja dat dahke dušše 76,7 % korpusis. Dan dihte leage dehálaš ahte syntávssalaš analysáhtor lea gierdil ja máhttá addit analysa vaikko cealkagis leat dovdameahttun sánit.

2.3.3 Eahpesihkkaris norbma ja ollu variánttat

Davvisámegeiela dálá čállinvuohki lea nuorra, ja lea leamaš fámus easka jagi 1979 rájes, ja dasa bohte vel stuora rievdadusat jagi 1985. Odđa čállinvuohki buvttii odđa oktavuodaid jietnadaga ja bustáva gaskkas, go lei Ruoŋa/Norgga ja Suoma čállingielaid ovtastuhttin (Artihkal III). Maiddái eará Giella-smX gielaide čállinvuogit leat nuorat. Dálá lullisámi čállinvuohki lea leamaš fámus 1974 rájes, julevsámi čállinvuohki bodii 1983:s,

	Analyserejuvvon sánit korpusis	Dovdameahttun sánit korpusis	Buot sánit korpusis
Davvisámegiela			
Giella-sme	23 224 604 97,6 %	573 821 2,4 %	23 798 425 100 %
Anárašgiela			
Giella-smn	1 276 586 95,2 %	63 742 4,8 %	1 340 328 100 %
Lullisámegiela			
Giella-sma	1 008 662 94,0 %	64 804 6,0 %	1 073 466 100 %
Julevsámegiela			
Giella-smj	922 922 92,4 %	76 397 7,6 %	999 319 100 %
Nuortalašgiela			
Giella-sms	135 901 84,1 %	25 703 15,9 %	161 604 100 %

Tabealla 2.4: Giella-smX analysáhtoriid beaktivuotta: Proseantalogut muitalit man stuora oasi SIKORis analysáhtorat máhttet analyseret (12.01.2018). Eai cealkkamearokat eai sánit čállojuvvon versálaiguin leat mielde. Anárašgiela ja nuorttalašgiela korpusa sánit mat leat ožžon analysa suomagiela analysáhtoriin (Giella-fin), ja eará sámegielaid korpusa sánit mat leat ožžon analysa girjedároggiela analysáhtoriin (Giella-nob), adnojuvvojit sitáhtan, eai dovdameahhtumin.

nuortalašgiela čállinvuohki 1991:s, anárašgiela čállinvuohki 1996:s ja bitonsámegiela dálá standárda čállinvuohki almmuhuvvui easka 2016:s Wilbura (2016) sátnegirjji bokte.

Divvun-prográmma válbmanii davvisámegiela ja julevsámegiela várás jagi 2007, ja jagi 2010 gárvánii maiddái lullisámegiela veršuvdna (Antonsen ja Trosterud, 2010). Jagi 2017 rájes anárašgiela betaveršuvdna lea leamaš geavahusas, vaikko dat ii leat almmolaččat almmuhuvvon. Divvun-prográmma merke rukses sázuin buot sániid mat eai leat čállojuvvon norpma mielde, ja norbman leat čállinvuogi prinsihpaid generaliseremat ja sátnegirjjiid ja grammatihkkagirjjiid čállinhámit. Divvun-prográmma sátnehámit divvojuvvojit dallego bohtet odđa normerenmearrádusat (Artihkal III).

Davvisámi čállingiela norbma dohkkeha veaháš variašuvnna, ja erenoamážit muhtun loatnasániid čállinhámiin lea variašuvdna sátnegirjjiis nubbái (Artihkal III). Maiddái sojahanparadigmmain sáhttet leat bálddalashámit, erenoamážit adjektiivvaid kompareremis. Adjektiiva *njálggat/njálggis* sodjá sihke bárahisstávvaladjektiivvan ja kontrakšuvdnaadjektiivvan (Nickel ja Sammallahti, 2011, 643), ja Divvun-prográmma dohkkeha mánga kom-

paratiivahámi: *njálgát*, *njálgásut*, *njálgásit*, *njálgáset*, *njálgáseabbo*, *njálgásat*, *njálgásabbo*.

Artihkkala III dutkanmateriálas leat 4 % sániin čállojuvvon norpma vuostá, ja dan logus leat maiddái duohtasátne-čállinmeattáhusat mielde (sánit mat gávdnojit gielas eará mearkkašumiin dahje sojahemiin). Artihkkala ulbmil lea oažžut muhtun muddui homogena dutkanmateriála, ja mun gudden teavsttaid main vuhttui ahte čállis ii lean sámegiella vuosttašgiellan dahje ii lean hárijánan čállit sámegillii. SIKORis leat buot lágan teavsttat ja maiddái teavsttat mat leat čállojuvvon ovdalgo dihto normerenmearrádusat bohte jagis 2005. Sániin mat ožžot analysa davvisámi SIKORis, ožžot 4,8 % analysa dušše deskriptiiva analysáhtoriin, ja dat mearkkaša ahte normatiiva analysáhtor ii dovdda daid. Dasa lassin bohtet sánit mat ožžot boasttuanalyssa. Ovdamearkka dihte leat korpusa analysas substantiivahámit *vuosttáš* ja *vuosttážin* 404 geardde, muhto juohke dáhpáhus lea rievtti mielde ortnetlohku *vuosttáš*, essiivvas *vuosttážin*, mii ii leat čállojuvvon norpma mielde.

Lullisámegiela čállingiela alfabehtas leat bustávát *æ* ja *ö*, ja bustávát *ä* ja *ø* leat norpma vuostá. Mánja čálii välljejit dattetge geavahit juogo *ä*, *ö*, mat heivejit ruoŋagiela boallobevdái, dahje *æ*, *ø*, mat fas heivejit dárogiera boallobevdái. Lullisámegiela alfabehtas lea sihke *i* ja *ï*, muhto ollu čállit geavahit čadat dušše *i*. Vai sáhtta analyseret lullisámi teavsttaid, de deskriptiiva analysáhtor ferte dohkkehít *ä*, *ø*, *i* bustávaid *æ*, *ö*, *ï* sajiin, ja dat gohčoduvvo stávenložžemin (= eng. ‘spell relax’). Olles lullisámi korpusis ožžot 6,6 % sániin analysa dušše deskriptiiva analysáhtoriin – eaige normatiiva analysáhtoriin. 10,9 % dáiin sániin sisttisdoallet *ä* dahje *ø*. Dát duodašta man dárbbášlaš sierra boallobevdi lea. Juo 1980-logu rájes lea gávdnon davvisámegiela boallobevdi dihtora várás, muhto easkka jagi 2017 rájes lea leamaš vejolaš viežžat sierra boallobevddi dihtorii maiddái julev- ja lullisámegiela várás¹⁹.

Ollu variašuvdna čállingielas dahká hástaleaddjin ráhkadit analysáhtora go šaddet eanet iešguđetlágan sátnehámit, ja variašuvdna dahká vel dehálaččabun testet analysáhtora teakstakorpusiin ja lasihit sániid ja sojahemiid maid analysáhtor ii dovdda. Variánttat mat leat duohtasátne-čállinmeattáhusat, dahket váddáseabbon oažžut rivttes syntávssalaš analysa. Syntávssalaš njuolgadusaid čállin lea čuovvovaš kapihttala fáddá.

¹⁹<http://divvun.no/keyboards/desktopindex.html> (26.9.2017)

2.4 Rivttes analysa válljen

Jus háliida diehtit kásusiid juohkáseami teakstakorpusis, dahje geavahit analysa vuodđun dihtorjorgaleapmái, de ii leat doarvái diehtit sátnehámi vejolaš morfologalaš analysaid, muhto lea dárbbaslaš diehtit sáni morfologalaš analysa konteavstta ektui. Dat mearkkaša ahte lea dárbu disambigueret analysaid gaskkas (disambigueret = jávkadit guovtte- ja mánggačilggolašvuoda).

2.4.1 Disambigueren ráddjehusgrammatihkain, Constraint Grammariin

Morfologalaš transduser (FST) máhttá sihke analyseret ja genereret sániid. Syntávssalaš analysáhtor máhttá dušše analyseret cealkagiid, iige genereret daid.

Syntávssalaš analysáhtor sisttisoallá grammatihka mii formulere ráddjehusaid das guđemuš analysa heive kontekstii. Ráddjehusaid lea olmmoš čállán. Formalisma gohčoduvvo *Constraint Grammarin* (Karlsson, 1990), ja dan kompiláhtora vuosttaš veršuvnnaid ráhkadedje Karlsson (1990); Karlsson ja earát (1995).

Manjel leat Tapanainen (1996) ja Eckhard Bick ja VISL-group (2008) ráhkadan odđa, buoriduvvon veršuvnnaid. Dálá Giella-smX-analysáhtoriid njuolggadusat ovtastuvvojit VISLCG3-prográmmain.

Oktiláš teavsttas lea ollu homonymiija. Davvisámegielas leat eanet homonymat go ovdamearkka dihte julevsámegielas mii lea konservatiivvalaččat go davvisámegiella, ja gehčosat eai leat gahččan oktii seammá ollu go davvisámegielas. Davvisámegielas lea measta olles homonymiija akkusatiivva ja genitiivva gaskkas, ja jus substantiivvas ii leat dássemolsun, de ovttaidlogus lea homonymiija maiddái nominatiivahámiin. Bárahisstávvalvearbain leat eanet homonyma hámit go bárrastávvalvearbain, ja *leat*-hámis leat olles vihtta analysa: infinitiiva, preseanssa biehttalanhápmi, ja golbma preseansahámi indikatiivvas: +Sg2, +Pl1 ja +Pl3.

Homonymiijat leat mánggaláganat, ja analysalohku lea dan duohken man ollu sánit leat leksikontransduseris, ja man ollu morfologalaš, syntávssalaš ja semantihkalaš dieđuid

FST addá. Homonymat sáhttet leat guokte sátnehámi seammá lemmas, nugo *nieiddain*, mii lea sihke ovttaidlogu komitatiiva ja mánggaidlogu lokatiiva. Homonymat sáhttet leat sátnehámit guovtti lemmas, nugo *bohte* man lemma sáhttá leat sihke *boahit* ja *bohtat*, dahje *maid* mii sáhttá leat advearba *maid* ja relatiiva- ja gahčanpronomena *mii* hápmi. Gávdnojit maiddá semantihkalaš homonymiijat sátneluohká siskkobealde: Giella-smX analysáhtoriin leat sierra gilkorat namaid refereanttaid ektui, ja *Trosterud* sáhttá leat sihke goargu +Sem/Sur ja báikenamma +Sem/Plc. Namma *Sara* sáhttá lea sihke goargu +Sem/Sur ja nissona namma +Sem/Fem, ja *Sara* oažžu maiddá gilhora +Attr muitalan dihte ahte nomengihpus namma sáhttá leat nuppi sáni attribuhtta, iige soja kásusiid mielde, nugo cealkagis *Sara Olsenis lea biila*. Syntávssalaš homonymiija lea maiddá arábalaš loguin, mat čállingielas ožžot gehčosa dušše dihto kásusiin, ja go ii leat kásusgeažus, nugo *12*, de lohku oažžu olles vihtta analysa, go lasihuvvo kásusgilkor buot vejolaš analysaid ektui: +Sg+Nom, +Sg+Acc, +Sg+Gen ja dasa lassin illatiiva ja lokatiiva attribuhttan, +Ill+Attr ja +Loc+Attr, seamma láchkai go lohkosániid geahnohis sojahheapmi (Nickel ja Sammallahti, 2011, 92).

Trosterud ja Wiechetek (2007) čáliiga ahte davvisámegiel sátnehámiin leat gaskamearálaččat 2,6 vejolaš morfologalaš analysa, ja julevsámegiel sátnehámiin fas 2,0 vejolaš analysa. Giella-smX analysalohku lea lassánan 2007 rájes, ja davvisámi teakstačoakkáldagas mas leat 35 000 sáni, maid Giella-sme dovdá, lea dál gaskamearálaš analysalohku 3,3²⁰.

Syntávssalaš analysáhtor vállje analysa mii lea riehta cealkaga ektui, go dat sihkku analysaid mat eai heive – dahje eará sániiguin: Analysáhtor *disambiguere*. Made eanet semantihkalaš ja syntávssalaš gilkoriid morfologalaš analysa addá, dađe eanet analysaid lea dárbu disambigueret. Sámegielaid disambigueret grammatihkalaš analysáhtorin. Statistihkalaš analysáhtora huksen gáibidivččii stuorit teakstakorpusa juohke sámegillii, go dat mii sámegielain lea.

Grammatihkalaš analysáhtor sáhttá bargat badjin vulos (= eng. ‘top down’), ja dalle dat geahččaladdá hypotesaid das makkár cealkkaráhkadus heive cealkagii, ja hypotesat leat vuodđun disambigueremii. Muhto lunddolaš giella ii álo čuovo syntávssalaš njuolggadusaid; čállii sáhttá rievddadit oaivila gaskan cealkaga, son sáhttá lasihit guhkes čilgehu-

²⁰<https://victorio.uit.no/biggies/trunk/gt/sme/corp/sme-goldcorpus.txt> (14.02.2018)

said, ja cealkka šaddá njulgestaga ilá kompleaksan. Badjin vulos analysáhtorat gáibidit maiddáahte buot sánit ožžot morfologalaš analysa. Sámegeielaid analysáhtorat barget vuollin bajás (= eng. ‘bottom up’) vuogi mielde, ja dan dihte analysáhtor álo addá analysa, maiddáahte cealkafragmeanttaide ja guhkes moalkecealkagiidda. Lullisámi analysáhtora evalueremis dat nákke oalle bures addit rivttes lemma oktan sátneluohká-analysain vaikko testenkorpusis leat ollu dovdameahhtun sánit, geahča kapihtal 2.4.2.

Ovdamearkan das mo ráddjehusgrammatihkka doaibmá, lea ambiposišuvnnaid disambigueren. Davvisámegeielas leat 28 ambiposišuvnna, mat leat adposišuvnnaid maid sáhtá geavahit sihke preposišuvdnan ja postposišuvdnan (Antonsen ja earát, 2012). Morfologalaš analysáhtor lasiha daidda sihke +Pr ja +Po, ja *rastá* oažžu analysan sihke RASTÁ+PR ja RASTÁ+PO. Konteavsttas dábálaččat lohkkái lea čielggas goabbá analysa lea riektá. Go genitiivakomplemeanta lea adposišuvnna olgeš bealde, de rivttes analysa lea RASTÁ+PR, nugo cealkagis (1). Go genitiivakomplemeanta lea adposišuvnna gurut bealde, dalle lea fas RASTÁ+PO rivttes analysa, nugo cealkagis (2).

- (1) *Dat lea goitge eanas áigge nu ahte dearvan besset rastá joga.* (SIKOR: Johan Turi, Muitalus sámiid birra)
- (2) *Ja go leat beassan joga rastá dearvan, (...)* (SIKOR: Johan Turi, Muitalus sámiid birra)
- (3) *Sutnje ii leat skuvlaovttasbargu rájiid rastá amas ášši.* (SIKOR: Ávvir)

Maiddáahte *joga*-sánis lea guovttečilggolašvuotta, danne go analysas leat sihke *johka+N+Sg+Acc* ja *johka+N+Sg+Gen*. Cealkagis (1) *joga* lea genitiivahámis danne go *rastá* lea preposišuvdna, ja cealkagis (2) fas danne go *rastá* lea postposišuvdna.

Cealkka (3) sisttisoallá vejolaš genitiivakomplemeantta ambiposišuvnna goappáge bealde: *rájiid (rastá)* ja *(rastá) amas ášši*. Vuolábealde leat cealkaga sániid morfologalaš analysat, ja konteavstta rivttes analysa lea merkejuvvon njuolain ja loguin. Syntávssalaš-analysáhtor sihkku buot eará analysaid, ja analysa vuolábealde leat Constraint Grammar-analysáhtora njuolggadusat (CG-njuolggadusat) čilgejuvvon.

Sutnje	son+Pron+Pers+Sg3+Ill	<= 1
Sutnje	suotnjat+V+IV+Ind+Prs+Du1	
Sutnje	suotnjat+V+IV+Ind+Prt+Pl3	
ii	ii+V+IV+Neg+Ind+Sg3	
leat	leat+V+IV+Ind+Prs+Pl1	
leat	leat+V+IV+Ind+Prs+Pl3	
leat	leat+V+IV+Ind+Prs+Sg2	
leat	leat+V+IV+Ind+Prs+ConNeg	<= 2
leat	leat+V+IV+Inf	
skuvlaovttasbargu	skuvlaovttasbargu+N+Sg+Nom	
rájiid	ráđji+N+Pl+Gen	<= 3
rájiid	ráđji+N+Pl+Acc	
rastá	rastá+Pr	
rastá	rastá+Po	<= 4
rastá	rastá+Adv	
amas	ii+V+IV+Neg+Sup+Sg3	
amas	amas+A+Sg+Nom	
amas	amas+A+Attr	<= 5
amas	amas+CS	
amas	amastit+V+TV+VGen	
amas	amastit+V+TV+Imprt+ConNeg	
amas	amastit+V+TV+Imprt+Sg2	
amas	amastit+V+TV+Ind+Prs+ConNeg	
amas	amasmuvvat+V+IV+VGen	
amas	amasmuvvat+V+IV+Imprt+ConNeg	
amas	amasmuvvat+V+IV+Imprt+Sg2	
amas	amasmuvvat+V+IV+Ind+Prs+ConNeg	
ášši	ášši+N+G3+Sg+Gen	
ášši	ášši+N+G3+Sg+Acc	
ášši	ášši+N+G3+Sg+Nom	<= 6
.	.+CLB	

1. Finihtta vearba ii heive ovddabeale ráiddu ‘biehttanvearba + *leat* + substantiiva nominatiivvas’ jus ii leat jietnadanvearba, dahje jus ii leat subjunkšuvdna dahje relatiivpronomen vearbba gurut bealde: Jávkkat finihtta vearbaanalysaid +Prs+Du1, +Prt+Pl3.
2. Gurut beale sátni lea biehttanvearba: Vállje vearbba biehttanhámi +ConNeg.
3. Olgeš beale sátni lea postposišuvdna, +Po, ja sátni ii leat transitiivva vearbba objektaposišuvnnas: Vállje genitiivva +Gen.

4. Gurut beale sátni lea genitiivvas: Vállje postposišuvnna +Po.
5. Olgeš beale sátni lea substantiiva, iige boade vearba dan maŋjel: Vállje adjektiivva attribuhttahámi +A+Attr.
6. Nominatiiva heive kontekstii danne go váldovearba lea *leat*: Vállje nominatiivva +Nom.

2.4.2 Sániid lemmatiseren (Art. II)

Lemmatiseren mearkkaša dihto teavsttas gávndat sániid rivttes lemma ja sátneluohká konteavstta ektui. Cealkaga (3) morfologalaš analysas oaidnit ahte *amas*-sáni máŋgga analysas leat vihtta iešgudetlágan lemma: *ii*, *amastit*, *amasmuvvat* (vearbbat) ja *amas* sihke subjunkšuvdnan ja adjektiivvan. Syntávssalaš analysáhtor vállje *amas*-adjektiivva rivttes lemman. Lemmatiseremis lea vejolaš stuorra teakstačoakkáldagas ráhkadit listtu main lemmat leat ordnejuvvon dávjodaga mielde, ja ovdamearkka dihte oaidnit ahte *amas* adjektiivvan lea ollu dávjjit go *amas* subjunkšuvdnan²¹.

Njuolggadusaiguin mat geavahit cealkagiid syntávssa veahkin, sáhtta viehka álki disambigueret lemmaid gaskkas mat gullet iešgudetlágan sátneluohkáide. Cealkagis (3) *amas*-sátni ii sáhte leat subjunkšuvdna iige biehttanvearba danne go čuovvovaš sátni lea substantiiva, eaige leat sánit substantiivva olgeš bealde. Ollu váddásat lea disambigueret lemmaid gaskkas mat gullet seammá sátneluohkkái, nugo vearbaid *amastit* ja *amasmuvvat*. Dás sáhtta geavahit transitivitehta veahkin. Danne go *amastit* lea transitiviiva vearba, de dat válljejuvvo jus cealkagis lea vejolaš objekta, *son ii amas ođđa ránnjá*.

Čuovvovaš cealkagiin ii leat doarvá luohhtit cealkaga syntáksii, muhto lea dárbu maid dái geavahit dan maid olmmoš diehtá olbmuid, sáhpániid ja luosaid birra, go disambiguere *vuodjit* ja *vuodjat*-vearbaid gaskkas:

- (4) *Muđui vudjet olbmot alla leahtuin dáikko, lohká Gerd Pedersen.* (SIKOR: Ávvir)

²¹SIKORis leat 2327 *amas*-sáni mat leat adjektiivvat, ja 290 *amas*-sáni mat leat subjunkšuvnmat, http://giellatekno.uit.no/lists/sme/sme_lemma.freq (22.11.2017)

- (5) *Manne sáhpánat vudje rastá joga?* (Dán cealkaga parafrása: *Manne sáhpán vuojai rastá joga?* SIKOR: Áššu)
- (6) *Garra rávnnji dihtii vudjet luosat gátti lahka ja čoahkkánit goatnili.* (SIKOR: Min Áigi)

Sáhtta ráhkadit njuolggadusa mii vällje vearbalemma subjeavtta mielde, ja nu sáhtta välljet *vuodjat* go subjeavttas lea +Sem/Ani-gilkor, mii muitala ahte lea ealli. Muhto lea váddásat ráhkadit njuolggadusaid mat välljejit riehta, dalle go subjeakta lea olmmoš, danne go olmmoš dábálaččat máhtta sihke *vuodjit* ja *vuodjat*.

Lemmatiseren lea ávkkálaš go áigu teakstakorpuseis ohcat odđa sániid sátnegirjái, ja jus galgá jorgalit teavstta nuppi gillii. Lemmatiserema bokte sáhtta sátnegirji jorgalit sáni, vaikko ohcansátni lea sojahuvvon hámis, eanet das lea čilgejuvvon kapihttalas 3.2.2. Muhto maiddái go bargá sátnebálddalastimiin (= eng. ‘word alignment’) statistihkalaš lahknanvugiin, de lea álkit gávdnat dan guovtti giela sátnebáraid jus vuos sáhtta lemmitiseret teavsttaid, go sámegeielas leksemii gullet ollu iešgudetlágan sátnehámit. SIKORis 14 davvisámi vearbbas leat eanet go 100 iešgudetlágan sátnehámi²². Go lemmitisere teavstta (= buhte juohke sátnehámi lemmain oktan sátneluohká gilkoriin), de juohke sáni dávjodat lassána sakka.

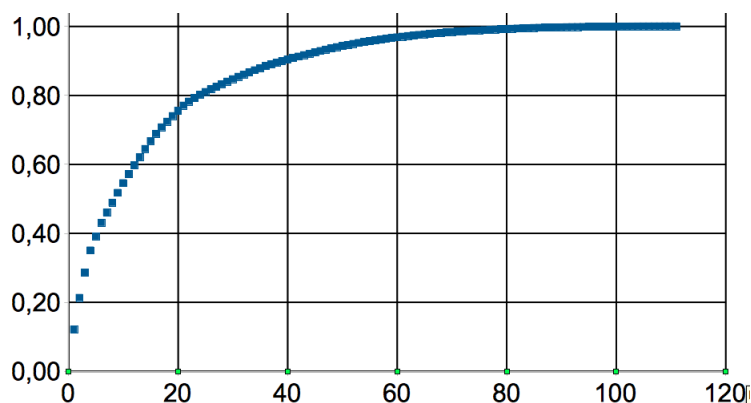
Giella-sme syntávssalaš analysáhtoris leat 1835 njuolggadusa mat juogo välljejit dahje hilgot analysaid (Artihkal I). Eanaš njuolggadusat välljejit rivttes morfologalaš analysa, ovdamearkka dihte juogo ovttaidlogu komitatiivva dahje mánggaidlogu lokatiivva substantiivvain nugo *biillain*, *skuvllain*. Lemmatiseremii birgešii ollu unnit njuolggadusai-guin. Giella-sme syntávssalaš analysáhtoris dušše 250 njuolggadusa välljejit dahje hilgot sátneluohká.

Artihkkalis II (*Next to nothing – a cheap South Saami disambiguator*)²³ lea jearaldat sáhttago unna CG-njuolggadusseahtain lemmitiseret lullisámi teavstta Constraint Grammariin. Ulbmil lea lemmitiseret korpusteavsttaid, ja artihkkalis bohtá ovdan ahte dušše 115:in disambiguerennjuolggadusain lea vejolaš oažžut viehka buori bohtosa lem-

²²Logus leat maiddái vearbba sojahanhámit oktan laktapartihkkaliiguin. Buot eanemus sátnehámit leat *leat*-vearbbas: 183 iešgudetlágan hámi. Artihkkala IV mielde leat juohke vearbbas 93 sátnehámi, muhto dan logus eai leat laktapartihkkalat mielde.

²³Errata: Artihkkala 2. siiddu ‘Figure ??’ galgá leat ‘Figure 2’.

matiseremis. Korpusteavsttain juohke sánis leat gaskamearálaččat 1,6 analysa go maddái leat mielde dovdameahttun sánit (FST gokčanmearri lei dušše 91,3 %), ja 1,8 analysa go leat mielde dušše cealkagat main lea ollislaš morfologalaš analysa. Disambigueren manai seammá bures buot cealkagiin, iige dušše daiguin main lea ollislaš morfologalaš analysa. Juohke sánis lea gaskamearálaččat 1,06 analysa disambiguerema manjel.



Govus 2.11: Giella-sma syntávssalaš analysáhtora njuolggadusaid beaktivuohta. Láskkoávssis = galle njuolggadusa. Ceakkoávssis = man stuorra oassi homonymiijas lea disambiguerenjvovon. (Artihkkala II govus 3.)

Govvosa 2.11 geavli govvida mo lullisámi njuolggadusat doibmet. Guoktelogi beaktileamos njuolggadusa (17,4 %) dahket olles 80 % disambigueremis. Manjimuš 5 % disambiguerenbargui dárbbášuvvojit 55 njuolggadusa.

Bargu duodašta ahte ráddjehusgrammatihkalaš lahknanvuohki lea gierdil maddái dalle go leat 8,7 % dovdameahttun sánit teavsttain. Artihkkalis evaluerejetne analysáhtora analysa gollecorpusa (= teaksta man analysa lea manuálalaččat divvojuvvon) ektui. Lemmatiserema recall (= buktosa rivttes analysat juhkkujuvvon gollecorpusa analysaide) lea buorre, 98,0 %, muhto deaivilvuohka (= buktosa rivttes analysat juhkkujuvvon buktosa buot analysaide) ii leat seammá buorre: 87,6–88,6 %. Dat mearkkaša ahte 98 % rivttes analysain leat mielde analysabohtosis, muhto guovtte- ja mánggačilggolašvuohka lea báhcán. Bargu buoridit bohtosa deaivilvuoda 88 %:s ovdamearkka dihte 95 %:i, gáibida ollu eanet barggu go beassat 88 %:i, geahča govvosa 2.11 geavlli.

2.4.3 Sámegeielaid syntávssalaš analysáhtorat

Nugo namuhuvvon kapihttalis 2.2, de leat guhtta Giella-smX analysáhtora. Buot morfologalaš analysáhtorat leat huksejuvvon FST:n, ja syntávssalaš analysáhtorat geavahit Constraint Grammara. Tabeallas 2.5 oaidnit mo Giella-smX lea modulariserejuvvon. Morfologalaš ja morfofonologalaš transduseriid ferte hukset juohke gillii sierra. Lea váttis sirdit osiid transduseris nubbái maiddái danne go sámegeielaid čállingielat leat nu iešgudetláganat.

Morfologalaš analysáhtora buktosis leat generaliseremat: sátneluohkáid ja eará morfologalaš iešvuodaidda gilkorat, ja semantihkalaš gilkorat. Ráddjehusnjuolggadusat čujuhit sihke lemmaid ja gilkoriidda, ja njuolggadusat mat čujuhit gilkoriidda eai leat čadnon dušše ovtta gillii.

Analysáhtor	Giella-sme	Giella-smj	Giella-sma	Giella-smn	Giella-sms
Leksikon ja morfologiija	Davvi-sámi FST	Julev-sámi FST	Lulli-sámi FST	Anáraš-giela FST	Nuortalaš-giela FST
Disambiguáhtor	Davvisámi CG	Julevsámi CG	Lullisámi CG	Heivehuvvon davvisámi CG	Heivehuvvon davvisámi CG
Syntávssalaš funkšuvnnaid lasiheapmi	Oktasaš sámi CG-analysáhtor				
Dependeansa lasiheapmi	Oktasaš sámi CG-analysáhtor				

Tabealla 2.5: Giella-smX analysáhtoriid modulaid interakšuvdna. CG = Constraint Grammar. Giella-smn ja -sms geavahit heivehuvvon davvisámi disambiguáhtora danne go eai leat vuos ráhkaduvvon sierra disambiguáhtorat.

Vaikko juohke sámegeiella dárbbáša sierra disambiguáhtora, de muhtun sámegeielain leat seammalágan homonymiijat mat dahket vejolažžan kopieret disambiguerennjuolggadusaid sekveanssaid analysáhtoris nubbái. Ovdamearkka dihte davvisámegeielas leat substantiivvaid ovttaidlogu komitatiivahámit ja mánggaidlogu lokatiivahámit homonymat. Julevsámegeielas lea seammalágan homonymiija ovttaidlogu komitatiivahámiid ja mánggaidlogu inessiivahámiid gaskkas. Davvisámegeielas leat substantiivvain akkusatiiva ja

genitiiva hámiin olles homonymiija, anárašgielas lea seammá homonymiija, muhto dušše ovttaidlogus. Anárašgiela ja nuortalašgiela disambiguáhtorin lea davvisámegiela disambiguáhtor, ja bargu heivehit ja divvut njuolggadusaid daidda gielaide lea easkka álggahuvvon.

Davvisámi disambiguáhtoriin lea bargojuvvon eanemusat danne go davvisámegiela analysa geavahuvvo máŋgga prográmmas. Dihtorjorgalanprográmmaid gáldogiella lea davvisámegiella, juoga mii gáibida buori morfologalaš ja syntávssalaš analysa (Antonsen ja earát, 2017). Giellaoahppanprográmma *Oahpa* davvisámegiela veršuvnnas leat maiddá giellahárjehusat mat analyserejit oahppi addaga syntávssalaččat, geahča kapihtala 3.4.3. Evalueremis mii dahkkui jagi 2009 teavsttaiguin dovdameahhtun sániid haga, oáččui Giella-sme F-score (mii rehkenasto sihke presišuvnna ja recall vuodul) 0,99 lematiseremis ja 0,94 olles morfologalaš disambigueremis, ja Giella-smj gis oáččui 0,95 ja 0,88 (Antonsen ja earát, 2010).

Funkšuvdnagilkoriid lasiheapmi ii leat čadnon ovttá gillii seamma ládje go homonymiija čoavdin. Syntávssalaš analysáhtor lasiha funkšuvdnagilkoriid cealkaga sániide, ja njuolggadusat čujuhit vuosttažettiin morfologalaš ja semantihkalaš gilkoriidda ja hárve lemmaide. Dependeanjuolggadusat čujuhit fas funkšuvdnagilkoriidda. Dan dihte lea vejolaš geavahit seamma dependeansaanalysáhtora buot sámegielaide (eanet detáljat dás: Antonsen ja earát (2010)).

2.5 Čoahkkáigeassu

Sámegielaidda analysáhtorat leat ráhkaduvvon giellamodeallan mas buot sátnehámit leat huksejuvvon transduserin LEXC-prográmmaid, ja boadus lea FST mii modellere morfologiija. Sámegielaidda leat hástaleaddji morfofonologalaš rievddadeamit, maid mii čoavdit guovttedásat bustávvelašuvnnaiguin TWOLC-prográmmas.

Giella-smX analysáhtoriid oktan ulbmiliin lea addit analysa buot sániide stuorra teakstakorpusis. Artihkal I giedahallá stuorimus analysáhtora, Giella-sme, ja artihkkalis boahdá ovdan ahte davvisámegiela analysáhtor gokčá 97,8 % SIKOR-teavsttaid sániin.

Juohke sátni sáhtta oážžut eanet go ovttá analysa, ja danne lea dárbbalaš disam-

bigueret analysaid gaskkas. Artihkal II duodašta ahte Constraint Grammar (ráddjehusgrammatihkka) lea gierdilis vuohki disambigueret teakstakorpusa maiddái dalle go leat dovdameahttun sáni, jus ulbmil lea lemmatiseret teavstta. Dát lea dehálaš danne go vaikko Giella-sme analysáhtoris lea dan mađe stuora gokčanmearri, de dattetge ii báze go 76,7 % korpusis syntávssalaš analysii jus guoddá eret cealkagiid mat sisttisdollet unni-musat ovtta dovdameahttun sáni. Artihkal čájeha maiddái ahte Constraint Grammariin lea vejolaš viehka unnán bargguin oažžut buriid bohtosiid lemmatiseremis. Eanaš guovtte-čilggolašvuhta mii báhcá, leat seamma lemma iešgudetlágan analysat.

Kapihtal 3

Giellamodealla heiveheapmi duohta giellamáilbmái

3.1 Giellamodealla geavaheapmi geavaheddjiidprográmmain

Mannan kapihttalis lei sáhka mo Giella-smX analysáhtoriid morfologalaš oassi lea huksejuvvon FST:n ja syntávssalaš oassi Constraint Grammar-njuolggadusaiguin. Dán kapihttalis leat fáddán mo sáhtta geavahit Giella-smX analysáhtoriid sámegeiela čálliid ja oahpahalliid geavaheddjiidprográmmain. Analysáhtorat leat ávkkálaččat giellateknologalaš prográmmain danne go dat addet sátnehámi grammatihkalaš sáni, grammatihkalaš sáni albmonemiid ja dat hilgot sániid mat eai gula gillii.

Kapihttalis 2 lei vuosttažettiin sáhka teavsttaid analyseremis ja mo oažžut analysáhtora dovdat buot sátnehámiid teakstačoakkáldagas. Dán kapihttalis lea sátnehámiid genereren dehálaš, ja geavaheddjiidprográmmaide leat sihke analysáhtora badjelmearálaš sátnehámiid genereren ja analysáhtora badjelmearálaš analysat čuolbman. Ulbmil lea oažžut giellamodealla sulastahttit duohta giela.

Sátnedárkkistanprográmma (kapihttalis 3.2.1) vuodđuduvvá dasa ahte analysáhtor dohkkeha buot giela sániid ja hilgu sániid mat eai leat čállojuvvon riektáčállima mielde. Go sátni lea dovdameahttun analysáhtorii, de analysáhtor generere relevánta evttohusaid dasa maid čálli lei oaivvildan čállit. Sátnegirji (kapihttalis 3.2.2) analysere addaga vai

gávdná lemma mii fas jorgaluvvo nuppi gillii, ja sátnegirjeartihkkalii analysáhtor generere sojahanparadigma mii sisttisdoallá lemma grammatihkalaš sániid albmonemiid. Maiddá giellahárjehusat (kapihttalis 3.4.1) geavahit grammatihkalaš sáni ja dan albmonemiid.

Buot namuhuvvon prográmmaide lea dárbu gáržžidit giellamodealla, ja ráddjet genererema. Go analysáhtora ráddjema birra in leat čállán nu ollu nákkosgirjji artihkkaliin, de lean lasihan kapihttala 3.3, man fáddán lea mo eastadit genereremis hámiid mat eai geavahuvvo gielas ja maiddá analysáhtora gáržžideapmi norpma ektui. Dasa lassin digaštalan dárbbu ráddjet genererema ja analysaid pedagogalaš sivaiddihte.

Maŋimuš kapihttaliin, 3.4.2 ja 3.4.3, giedahalan giellamodealla viiddideami. Ulbmil lea oážžut analysáhtora dovdat maiddá sátnehámiid mat dábálaččat eai adnojuvvo gillii gullevaš sátnin, muhto maid giellaoahpahalli sáhtta čállit.

3.2 Sátnehámis lemmai ja lemmas sátnehámiide

3.2.1 Sátnedárkkistanprográmmii genereret sátnehámiid (Art. III)

Artihkkalis III (*Čállinmeattáhusaid guorran*) leat guokte oasis. Vuosttaš oasis lean guorahallan davvisámi teavsttaid oaidnin dihte makkár váttisvuodát čáliin leat čállinvuogi ektui. Dárbbasitgo eatnigielat čállit máhttit eksplisihhta grammatihka (nugo olmmoš oahppá skuvllas dahje grammatihkkagirjjiid vehkiin) vai sáhttet čállit norpma mielde? Dutkanmateriálas 4 % sániin eai leat čállojuvvon norpma mielde. Stuorra oasis čállinmeattáhusain lea konsonántaguovddázis, ja dasa lassin lea čáliide váttis čállingielas earuhit vokálaidda *a* ja *á*. Analysabohtosat čájehit ahte čálii ii sáhte dušše luohttit iežas dadjamii, muhto dárbbasa diehtit makkár buncarággát iežas suopmanis leat čállingiela ektui ja dalle geavahit eksplisihhta grammatihkkamáhtu veahkkin. Nubbi čuolbma lea ahte norbma ii álo leat nu čielggas, ja muhtun sánit leat čállojuvvon iešguđetge ládje sátnegirjjiin.

Maiddá iniálianoanádusaid ja loguid kásusmerken lea dábalaš meattáhustiipa, ja daidda riektačállinnjuolggadusat sáhttet leat njulgestaga snugormastingeadgin čállái. Loatna-

sániid heiveheapmi čuovvu njuolggadusaid, muhto muhtun sániin leat dattetge nu máŋga vejolaš čállinhámi ahte dat eai sajáiduva čállingillii, geahča maiddá kapihttala 2.3.3.

Artihkkala nubbi oassi guorahallá man muddui Divvun-programma Word-veršuvnna sáhtta veahkehit čálii. Divvun váldá vuhtii dušše ovttaskas sáni iige konteavstta, ja dan dihte programma ii sáhte evttohit divvut sáni mii gávdno gielas eará mearkkašumiin dahje sojahemiin (duohtasátnemeattáhus). Dasa dárbbášuvvo grammatihkkadárkkistan-programma mii váldá konteavstta vuhtii, ja dákkár programmain lea bargu álggahuvvon, geahča Wiechetek (2012).

Dan dihte Divvun-programma potentiála lea gávdnat ii-sátni-meattáhusaid, ja dat dahketge stuorimus oasi čáliid meattáhusain. Evaluerema boadus lea ahte programma fuomášuhtta čállái 78 % meattáhusain, ja 82 %:s dain dáhpáhusain programma fállá čállái relevánta čállinhámi kontekstafáhus, mii sisttisdoallá eanemusat vihtta evttohusa. Artihkal almmostahtta maiddá ahte programmas leat buoridanmunit, nugo váilevaš sániid lasiheapmi ja evttohusaid válljema algoritmma buorideapmi.

Kapihttalis 2.3.1 analysáhtora huksema ulbmil lei analyseret buot teakstačoakkáldaga sániid. Sátnedárkkistan-programma fas galgá dohkkehit dušše sániid mat leat čállojuvvon norpma mielde. Analysáhtora ráddjen norpma ektui giedahalan kapihttalis 3.3.2.

Sátnedárkkistan-programma stuorra hástalus lea evttohit relevánta riektačállon sáni čállái. Programma algoritma vuodđuduvvá dasa, ahte galgá divvut boastut čállon sáni nu unnán go vejolaš vai beassá riektačállon sátnái. Programma logahallá divvumušaid, mat sáhttet leat bustáva lasiheapmi, sihkkun, buhtten dahje guovtti bálddalas bustáva lonuheapmi, ja daid submi gohčoduvvo *divvungaskan*. Jus ferte dahkat guokte dákkár divvumuša, de lea divvungaska 2. (Damerau, 1964; Levenshtein, 1965).

Divvun-programma vuoruha evttohusaid main lea unnimus divvungaska, muhto algoritmii gullá maiddá fonologalaš komponeanta, mainna sáhtta vuoruhit evttohusaid. Dáinna programma vuoruha muhtun evttohusaid mat gullet meattáhusaide maid eatnigielagat dávjá dahket, ovdamearkka dihte $ie > ea$, mas lea divvungaska 2, ovddabeale evttohusat main leat divvungaska 1, nugo govvošis 3.1 mas Divvun evttoha *filkkas* hápmái *fylkkas* easkka guđádin. Programmas ii gávdno njuolggadus vuoruhit $i > y$ -rievdadusa, ja dan dihte dat ii evttohuvoe ovdal eará vejolašvuodaid.

dierpmi > *dearpmi* ⁽²⁾, *dearbmi* ⁽³⁾, *dierpmá* ⁽¹⁾, *fierpmi* ⁽¹⁾, *jierpmi* ⁽¹⁾

filkkas > *fikkas* ⁽¹⁾, *Hilkkas* ⁽¹⁾, *bilkkas* ⁽¹⁾, *fiŋkkas* ⁽¹⁾, *fiškkas* ⁽¹⁾, *fyllkkas* ⁽¹⁾, ...

Govus 3.1: Divvunprográmma evttohusat boastut čállon sániide. Juohke evttohussii ad-
dojuvvo divvungaska. (Artihkkala III ovdamearkkat 10 ja 11.)

Artihkkalis geavahuvvo Divvun-veršuvdna mii lea oaivvilduvvon Microsoft Word-prográmmii. Odđa Divvun-prográmma, mii lea leamaš viežžan láhkai 2013 rájes LibreOffice-prográmma várás, geavaha HFST ja eará evttohus-teknologiiija (Voikko), mii dahká vejolažžan váldit vuhtii maiddái eará sáni iešvuodaidd: sátnehámi dahje lemma dávjodaga teakstakorpusis, ja maiddái morfologalaš hámi dávjodaga.

3.2.2 Sátnegirjjiid riggodahttit morfologiiain (Art. IV ja V)

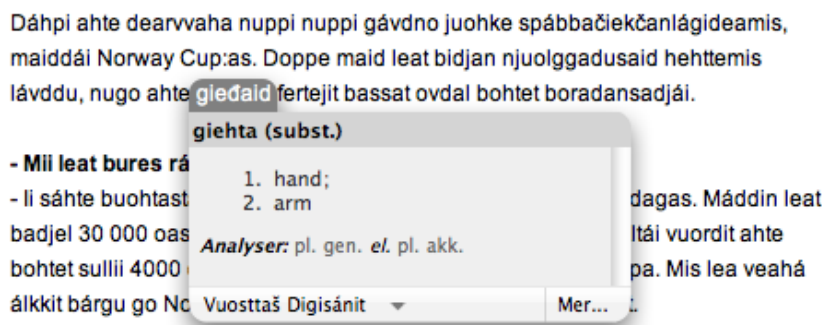
Dán kapihttalis giedahalan mo morfologalaš analysáhtoriin sáhttá riggodahttit sátnegirjji nu ahte geavaheaddji beassá sojahuvvon sánis rivttes lemma-artihkkalii, ja mo sátnegirjeartihkkalii genereret sojahanparadigmaid.

Artihkkala IV (*Ei intelligent elektronisk ordbok for samisk*) fáddá lea *Vuosttaš digisánit-sátnegirjji* (VD) vuogádat mii laktá morfologiiija sátnegirjái. Sátnegirjevuogá-dahkii leat laktojuvvon sátnegirjji gáldofiillat, morfologalaš generáhtor ja sátnegirjegeava-hanlakta mainna sáhttá juogo čállit ohcansáni dahje coahkkalit sáni teavsttas. Dákkár modulii morfologalaš analysa lea vealtameahttun.

VD lahknanvuohki lea sátnegirjái FST:in genereret sátnegirjji lemmaid sátnehámiid, čoavddahámiid ja giellaoahpahalli konteavsttaid. Dieđut vurkojuvvojit diehtovuđđui man lingvisttalaš sisdoallu lea stáhtalaš, muhto geavahanlakta dahká sisdoalu dynámalažžan geavaheaddjái.

Lemmaid sátnehámit genererejuvvojit lemmalisttu gilkorlisttuin mas leat ovddalgihtii definerejuvvon hámit maid háliidit sátnegirjái, ja daid gaskkas lea ovdamearkka dihte substantiivva mánggaidlogu akkusatiiva- ja genitiivahámit (+N+Pl+Acc, +N+Pl+Gen). Juohke sátnehápmái genererejuvvo sátnegirjeartihkal mas leat dieđut morfologalaš hámi birra ja jorgalus (govus 3.2). Dát sátnegirjeartihkal lea fas linkejuvvon sáni váldoartihkkalii, mas leat čoavddahámit ja vejolaččat ovdamearkacealkagat (govus 3.3). Čoavd-

dahámit addet sáni morfofonologalaš rievddademiid. Pronomeniidda háliideimmet ovdamearkacealkagiid juohke sátnehápmái, eat dušše lemmai, ja maiddái addit eará lemma go FST lemma, vai sáhtta juohkit sojahanparadigmaid persovvna-logu mielde amaset sojahanparadigmat šaddat menddo stuorrát ja kompleksat giellaohpahallái. Dan dihte čáliimet daidda sátnegirjeartihkkaliid, dan sajis go genereret daid. Seamma láhkai dagaimet *leat-* ja biehttalanvearbba sojahanhámiide.



Govus 3.2: VD lohkanveahkki: Geavaheaddji coahkkala sáni teavsttas, *giedaid*, masa lea genererejuvvon sátnegirjeartihkal mas lea jorgalus ja morfologalaš analysa: *pl. gen. dahje pl. akk.* Sátnegirjeartihkkalis lea *Mer...*-liŋka (= eambo) sáni vuodđohápmái *giehta*. (Artihkkala IV govus 1.)



Govus 3.3: VD: *giehta*-lemma válđoartihkal, mas leat jorgalusat, substantiivva čovvdahámit ja ovdamearkacealkka. (Artihkkala IV govus 2.)

Sátnegirji ja FST fertejit heivet oktii dego bátnejuvllat. Lemma ja sátneluohkká fertejit leat seammát vai lea vejolaš genereret sojahanhámiid. Vai sátnegirji galgá addit rivttes sátnegirjeartihkkala geavaheaddjái, de lei dárbu heivehit analysáhtora, sihke homonymaid ja mánggaidlogu lemmaid dihte.

Muhtun sátnegirjeartihkkaliin leat homonyma vuodđohámit, muhto dain leat goappatlágan sojahanparadigmat ja jorgalusat. Artihkal IV giedahallá davvisámi-dáru sátnegirji, ja das leat čuovvovaš ovdamearkkat: Jus geavaheaddji ohcá *lohkkis*, man vuodđohápmi lea *lohkki*, de sátnegirji galggašii addit sátnegirjeartihkkala LOHKKI+N = *leser* (dahkkinomen), iige LOHKKI+N = *lokk* (mainna gokčá lihti). Seamma láhkai galgá sátni *beasis* addit sátnegirjeartihkkala BEASSI+N = *reir* (lotti huksehus) iige BEASSI+N = *never* (soagi bárku). Dasa geavahit morfologalaš gilkoriid analysáhtoris ja sátnegirjis. Buot dahkkinomenat merkejuvvojit +NomAg²⁴ (láhten ‘Nomen Agentis’ = dahkkinomen) ja substantiivvat main lea dássemolsuma goalmát gráda mii ii leat merkejuvvon čállingielas, merkejuvvojit +G3 (eng. ‘Grade 3’ = goalmát gráda). Dáinna lágiin earuhit LOHKKI+N ja LOHKKI+N+NOMAG, ja maiddái BEASSI+N ja BEASSI+N+G3. Namuhuvvon gilkorat sihkkarastet maiddái ahte vuogádat generere rivttes čovddahámiid, ovdamearkka dihte LOHKKI+N+NOMAG addá hámiid *lohkki*, *lohkkái*, *lohkkiide* ja LOHKKI+N gis addá *lohkki*, *lohkkái*, *lohkkiide*.

Mii ráhkadeimmet manjel maiddái lullisámi VD-veršuvnna ja välljiimet geavahit homonymiijagilkoriid +Hom1 ja +Hom2 dalle go lemmain ii leat seammá jorgalus ja dasa lassin čuvvot goappatlágan sojahanminstariid, nugo GOVLEDH+N+HOM1 = *å høre* (= gullat) ja GOVLEDH+N+HOM2 = *å høres* (= gullot). Justa dán guovtti vearbba homonymiija livččii leamaš vejolaš čovdit transitivitehtagilkoriin, danne go nubbi lea intransitiiva (+V+IV) ja nubbi lea transitiiva (+V+TV).

Sátnegirjiin leat maiddái sánit maid mánggaidloguhámis lea eará mearkkašupmi go ovttaidloguhámis lea, ja dan dihte lea sierra sátnegirjeartihkal daidda. Ovdamearkkat leat GÁVDNI+N = *nytte*, *gagn* ja GÁVNNIT+N = *sengetøy* ja GÁFFE+N = *kaffe* ja GÁFET+N = *kaffebønner*; *ukokt kaffe*. Dákkár lemmat gáibidit ahte maiddái FST:s leat sihke ovttaidlogu- ja mánggaidlogulemmat.

²⁴Artihkkalis IV lea geavahuvvon gilkora boarrásat veršuvdna: +Actor.

Dahkan dihte čovddahámiid ipmirdahttin geavaheaddjái, de leat lasihan konteavstta grammatihkalaš tearpmaide lassin. Áigeadvearbbat válldahit vearbbaid áigesojaheami, persovdnapronomenat válldahit vearbbaid persovnna-logu-sojaheami ja biehttalanvearba válldaha vearbba biehttalanhámi. Ovdamearkka dihte *borren*-hápmi ovdanbuktojuvvo ná: (*ikte mun*) *borren*, ja seamma láhkai lullisámegillii (*daan biejjien manne*) *byöpmem*. Vearbbat main dábálaččat ii leat olmmoš subjeaktan, eai oaččo persovdnapronomen, muhto baicce čujuheaddji pronomena *dat*, lullisámegillii *dihte*. Sátnegirji čuovvu maid dái Jernslettena (1988) vuogi addit adjektiivvaid attribuhttahámi oktan substantiivvain. Vai substantiiva heive adjektiivii, de lasihit heivvolaš substantiivva sátnegirjji gáldofiillii. Čájehan dihte mánggaidlogu lohkosátnegihpu siskkáldas kongruenessa, de leat addán substantiivva *gáma*-lemma konteakstan lohkosániide. Gáldofiilla lemmaid vuodul FST generere sátnegirjeartihkkalii čovddahámiid oktan konteavsttain, nugo *rukkes bivttas*, *suhkkes vuovdi*, *guovttit gápmagat*, *guvttiid gápmagiid*.

Davvisámi sátnegirjgis maid artihkal IV válldahallá, ledje 5192 lemma ja sátnegirji oaččui dainna vugiin oktiibuot 505 480 sátnehápmeartihkkala. Artihkkala IV evalueren čájehii ahte dovdameahttun sániid gaskkas dahke goallossánit stuorimus oasi, muhto dan čuolmma eat sáhttán čovdit VD geavahanlavttas. Sátnegirjái eai genererejuvvon dynámalaš suorggádusat, ja dovdameahttun suorggádusaid gaskkas ledje passiiva vearbahámit deháleamos suorggádusat. Mañjelgo artihkal IV čállojuvvui, lasiheimmet evaluere ma vuodul transitiivva vearbbaid oktan passiiva hámiiguin genererenfiillii, ja dasa lassin finihitta vearbahámiid oktan jearranpartihkkaliin. Amas sátnegirjefiila šaddat mendo stuorisin, válldiimet eret substantiivvaid ja adjektiivvaid hámiid main lei oamastangeažus, dainna ákkain ahte dat eai leat nu ávkkálaččat geavaheaddjái go dákkár sátnehámit eai lean duhát dábálaččamus sátnehámiid gaskkas frekveansalisttus ráhkaduvvon korpusa vuodul 2009:s. VD odđasat veršuvnnas, mii lea geavahuvvon artihkkala V evalueremii, leat 9999 lemma, maiguin genereriimet oktiibuot 252 787 sátnehápmeartihkkala.

Go árvoštallá dán čovdosa olles SIKORii, iige dušše ráddjejuvvon frekveansalistui, de oaidná ahte adjektiivvat oamastangehčosiin leat davvisámegielas hui hárve, ja eanaš dain sáhttet maid dái oazžut substantiivaanalysa: *buriiidis*, *boarrásiiddiset*, *nuoraidasamet*, *ipmilbalolaččaidis*, *ráhkkásiiddán*, *vártnuhemiidat*. SIKORis leat maid dái moadde kompa-

rerejuvvon hámi oktan oamastangehčosiin, muhto dušše ovtta lemmas: *buoremusaset*, *buoremusaideaset*.

Substantiivvat oamastangehčosiin leat ollu dávjjibut ja gávdnojit maiddá čáppagirj-
jálašvuoda teavsttain mat leat geavahusas sámegiela álgoahpahas²⁵. Teavsttain leat
erenomážit fuolkevuodasánit oamastangehčosiin, muhto maiddá rumašlahtut, orrunsajit
ja eallit. Olles oamastangeažusparadigma dagašii 117 hámi²⁶ juohke substantiivii, muhto
muhtun hámit leat oalle marginála geavahusas, ovdamearkka dihte essiivahámit (Anton-
sen ja Janda, 2015, 29). Dan sajis go genereret buot hámiid buot substantiivvaide, de
sáhtášii árvvoštallat genereret ovdamearkka dihte 50 dábáleamos hámi substantiivvaide
mat dávjjimusat geavahuvvojit oamastangehčosiin.

VD-vuogádagas leat buot sojahanhámit ovddalgihtii genererejuvvon diehtovuđđui ja
sojahanhámiid artihkkalat leat linkejuvvon lemma sátnegirjeartihkkalii. **Artihkkala V**
(*Using finite state transducers for making efficient reading comprehension dictionaries*)
giedahallá dynámalaš lahknanvuogi: Geavaheadji atta sáddejuvvo Giella-smX:i mii dan
lemmatisere (lemmatiseren čilgejuvvo kapihtalis 2.4.2), ja analysa lemma, dahje lem-
mat jus leat eanet analysat, sáddejuvvojit sátnegirjevuogádahkii. Artihkal V válddahallá
davvisámi *Neahttadigisániid* ja lullisámi *Nedtedigibaakoeh* (NDS).

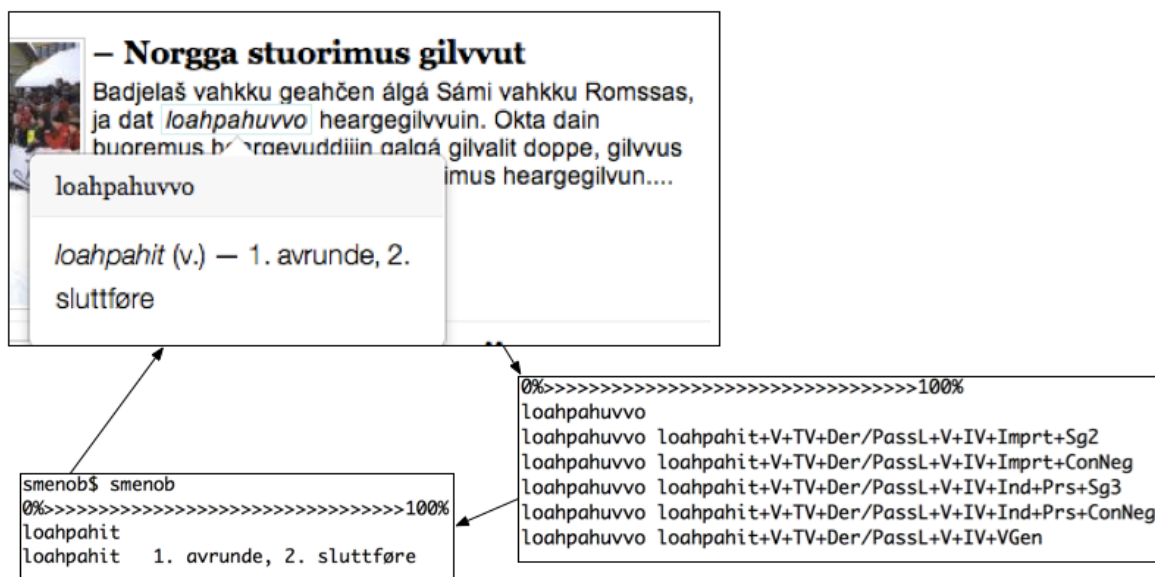
NDS-sátnegirjjis lea sierra modulan neahttalohkkái lakti girjemearka ja geavaheadji
sáhttá seamma láhkai go VD:in, logadettiin teavsttaid coahkkalit sáni mii sáddejuvvo
sátnegirjái, mii fas fállá sátnehámi lemma ja jorgalusa (govus 3.4). Muhto NDS geava-
hanlavttas lea boallu mii sádde ohcansáni lemma SIKORa geavahanlaktii, ja geavahead-
djái fálojuvvojit buot korpusa cealkagat main lea ohccojuvvon lemma. Govvosa 3.5 sátnegirjeartihkkalis lea okta cealkkaovdamearka *loahpahit*-lemmain, muhto *Oza buot hámiid teavsttain*-boaluin geavaheadji oažžu 7402 cealkaga dasa lassin²⁷.

Deháleamos lea go NDS-lahkonanvugiin sáhttá jorgalit sániid mat eai leat sátnegirj-

²⁵Ovdamearkka dihte Čábbámus iditguovssu-románas (Elle Márjá Vars 2002, Idut), mii lea nu gohčoduvvon álkeslohkanteaksta, leat 23 iešgudetlágan sátnehámi mat leat substantiiva oamastangehčosiin: *áhčistis, baksamiiddis, bussázan, čalmmiidis, čiegunvuodaidan, dáidagiinnis, eadnát, eatnis, lanjat, latnjasis, liđiidat, liđiidis, Linázan, lottázan, mánázan, mielastis, namas, nieiddažan, olbmásis, sallasis, salastis, sojjidat, vuovttaidis*

²⁶Antonsen ja Janda (2015) mielde leat 78 iešgudetlágan hámi, muhto sátnegirji addá maiddá morfologalaš analysa, ja earuha ovdamearkka dihte akkusatiiva- ja genitiivahámiid vaikko ieš sátnehámit leat homonymat.

²⁷SIKOR Veršuvdna 08.12.2016.



Govus 3.4: NDS: Geavaheaddji coahkkala sáni *loahpáhuvvo* maid analysáhtor lemmatisere. Boadus lea *loahpahit* maid sátnegirji jorgala dárogillii.

ji gáldofiillas, nugo goallossániid, suorggádusaid ja sániid main lea laktapartihkal, ja nu leat čovdojuvvon stuorimus váttisvuodát mat bohtet ovdan artihkkala IV evalueremis. Analysáhtora bokte sátnegirji maiddá gokčá buot oamastangeažushámiid.

Jus atta lea goallossátmi mii ii gávdno sátnegirjjiis, de analysáhtor addá sáni máddagiid, maid sátnegirji fas jorgala nuppi gillii, nugo *bargojoavku* šaddá *bargu* ja *joavku*. Nu guhká go goallossátmi lea čadačuovgi, de jorgalus addá buori gova goallossáni mearkkašumis. Jus suorggádus ii gávdno sátnegirjjiis, de analysáhtor addá sáni máddaga maid sátnegirji jorgala nuppi gillii.

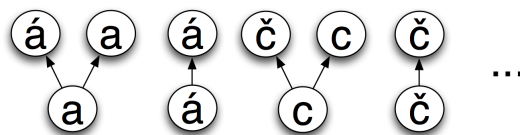
The screenshot shows the NDS interface for the word *loahpahit*. It includes a search bar with the word *loahpáhuvvo*, a dropdown menu showing the selected form *loahpahit (verb)*, and a list of grammatical features: *(verb) avrunde*, *(verb) slutføre*. There is also a section for *Oza* (conjugation) with *Oza teavsttain*. The interface also shows the word *loahpahit* and its grammatical information: *loahpahit Oza buot hámiid teavsttain → loahpahit verb passiv verb indikativ presens nektingsform loahpahit verb passiv verb indikativ presens 3.p.ent.*

Govus 3.5: NDS fállá ovdamearkacealkaga *loahpahit*-vearbain, muhto *Oza buot hámiid teavsttain*-boaluin geavaheaddji beassá maiddá njuolga SIKORa geavahanlaktii gos oažžu eanet cealkagiid mat sisttisdollet seamma lemma.

Sátnegirji geavaha deskriptiivlaš analysáhtora ja sáhtá dáinna lágiin dovdat maid-
dái sátnehámiid mat eai leat čállojuvvon norpma mielde. Leksikontransduseris dákkár
hámit leat lasihuvvon lemmaid normatiiva hápmái. Jus ohcá *kántuvrras*, de lemmatiseren
addá normatiiva lemma *kantuvra*: **kántuvrras kantuvra+Err/Orth-a-á+N+Sg+Loc**.

Artihkkala V evalueremis bohtá ovdan ahte NDS lahknanvuohki²⁸ buorida sátnegirj-
ji gokčanmeari hui ollu VD-lahkonanvuogi²⁹ ektui. Davvisámegiela NDS áidnageardásaš
sániid (= eng. ‘unique words’) gokčanmearri lea 79,7 %, ja VD gokčanmearri 57,8 %. Lulli-
sámegielas leat gokčanmearit 67,0 % NDS:in ja 41,1 % VD:in. Lahkonanvugiid iešvuodat
lean buohtastahtán tabeallas 3.1.

Artihkkalis V válddahallat maiddái NDS davvisámi veršuvnna mii lea heivehuvvon
boallobeavdái mas eai leat sierra sámi bustávat. Dat geavaha sierra FST³⁰ masa lea lasi-
huvvon oppalaš stávenložžen, mii mearkkaša ahte dihto bustávat ožžot eará mearkkašumi,
dahje lassemearkkašumi, nugo davvisámegielas: *a:a/á*, *c:c/č*, *d:d/ď*, *n:n/ŋ*, *s:s/š*, *t:t/ť*,
z:z/ž (govus 3.6) ja lullisámegielas: *i:i/ï*, *ø:ø/ö*, *ä:ä/æ*. (Lullisámi stávenložžen lei maid-
dái implementerejuvvon VD:i.) Stávenložžemiin FST dulko davvisámi sáni *manna* golb-
man vejolaš davvisámi sátnin: *manná*, *mánná*, *maŋŋá*. Dát lea dehálaš earenoamážit jus
áigu geavahit sátnegirjji sosiála mediain, nugo Facebookas, gos lea dábálaš čállit távvaldihtoriin
dahje mobiilatelefovnain. Easkka maŋimuš áigge lea leamaš vejolaš sajáiduhttit
sámi bustávaid daidda reaidduide, eaige buot sámegiela čállit dan dieđe. Artihkkala
evaluerema mielde stávenložžen buorida NDS Facebook-teavsttaid gokčanmeari 50,8 %:s
59,8 %:i.



Govus 3.6: Stávenložžen lea sierra FST mii konkatinerejuvvo morfologalaš FST:in. Dát
oppalaš stávenložžen addá dihto bustávaide eanet mearkkašumiid: *a* dulkojuvvo sihke *a:n*
ja *á:n* ja *c* fas *c:n* ja *č:n*. Bustávva *á* dulkojuvvo dušše *á:n*.

Maŋjelgo artihkal V čállojuvvui, de leat davvisámi NDS:i lasihuvvon ollu sátnebárat

²⁸Artihkkalis dat gohčoduvvo ‘FST-dictionary’.

²⁹Artihkkalis dat gohčoduvvo ‘wordform-dictionary’.

³⁰`analyser-dict-gt-desc-mobile.hfst`

nu ahte sátnegirjjiis leat dál 30 000 sámi ohcanáni³¹. NDS geavahanlavttas lea leamaš buoridanmunni suorggádusaid ovdanbuktimis, ja sátnegirjái leat lasihuvvon suorggádusgilkorat ja muhtun infinihttavearbbaid gilkorat³². Ovdamearkka dihte sátni *oađesta* analysa lea lemma *oađdit* oktan gilkoriiguin +V+IV+Der/st+V+Ind+Prs+Sg3. Lemma jorgaluvvo dárogilli *sove* ja suorggádusgilkor +Der/st jorgaluvvo *å X litt* (= eng. ‘to X a little’) ja dasa lassin čilgejuvvo suorggádusa funkšuvdna: *diminutiiva/subitiiva*. FST analysa gilkorat dahket vejolažžan genereret dárkilis dieđuid suorggádusa semantihkas. Seamma láhkai lea dahkkon ovdamearkka dihte vearbaabessiivvain (govus 3.7). Sáni *oađekeahtá* analysas lea *oađdit+V+IV+VAbess*, ja gilkor +VAbess jorgaluvvo *uten å X* (= eng. ‘without X-ing’).

Davvisámegiella (Dábálaš) → Dárogiegla (⇔ Molsso)

oađekeahtá lea vejolaččat dán sáni hápmi:

oađekeahtá

oađdit (verb)

- (verb) *sove*

Mánná oađđá eatni báiddas.
Barnet sover ved siden av mora.

oađdit *Oza buot hámiid*
teavsttain →
 oađdit verb verbabess. = (uten å)

+ verbabessiv

- *uten å X*

Govus 3.7: NDS čilge geavaheaddjái vearbaabessiiva *oađekeahtá* mearkkašumi.

Dán kapihttala fáddá lea mo FST riggohtahtá sátnegirjjiid, go FST:in sáhtá lemma-tiseret addaga gávdnan dihte sátnegirjeartihkkala, maidái dalle go atta lea goallossátni dahje suorggádus mii ii leat sátnegirjjiis. Grammatihkkagilkorat addet geavaheaddjái lassedieđuid grammatihkalaš hámiid birra, muhto lea dárbu jorgalit daid nu ahte +VAbess šaddá *uten å X*. FST:in sáhtá genereret lemmaid sojahanparadigmaid geavaheaddjái veahkin ja daidda hámiide maidái genereret giellaoahppanpedagogalaš konteavsttaid

³¹Giellatekno programmerár Ciprian Gerstenberger lea leamaš guovddáš olmmoš sátnegirjjiid gáldo-fiillaid gieđahallamis.

³²Giellatekno programmerár Chiara Argese lea implementeren buoridemiid.

vai lea álkit geavaheaddjái diehtit mo galgá geavahit juohke hámi. FST lemmatisere teakstakorpusa, ja dat dahká vejolažžan gávdnat cealkkaovdamearkkaid sátnehámi lemma bokte.

Ješvuohta	Vuosttaš digisánit VD	Neahttadigisánit NDS
Mo dovdá ohcansániid Morfologalaš veahkki Sátnehuksena čilgen	FST genereren čoavddahámit –	FST analysa olles paradigmát ✓
Dovdá oamastangehčosiid dynámalaš goallossániid dynámalaš suorggádusaid hámiid mat eai čuovo čállinnorpma	– – – –	✓ ✓ ✓ ✓
Lullisámi stávenložžen Davvisámi stávenložžen	✓ –	✓ ✓
'Coahkkal teavsttas'-funkšuvdna Liŋkejuvvon SIKORii	✓ eanaš prográmmain –	✓ html online ✓

Tabella 3.1: VD ja NDS sátnegirjjiid iešvuođaid buohtastahttin.

3.3 Giellamodealla gáržžideapmi

Go ráhkadit analysáhtora dainna vugiin ahte viežžá sátnehámiid teakstakorpuse, de oažžut dušše sániid mat gávdnjit teavsttain, ja nu mii diehtit ahte analysáhtora sánit duođaid leat geavahusas. Muhto dát bargovuohki eaktuda hui stuorra teakstakorpusa mii gokčá buot šáŋjeriid, ja erenoamáš stuora korpusa jus gielas leat ollu iešguđetlágan sojahanhámit. Sámeielagiid gielas leat eanet iešguđetlágan sátnehámit go sámeielat teakstakorpusein leat. Dat mearkkaša ahte dákkár analysáhtor ii dovddaše buot sániid maid sámeielat olmmoš geavahivččii, juoga mii ovdamearkka dihte dagašii heajos sátnedivvunprográmma. Sátnehámit mat eai galggaše leat apmasat eatnigielagiidda, nugo vearbahámit *dajaide* ja *boraide*, eai gávdno SIKORis, geahča kapihttalas 2.1. Dan dihte mii leat välljen hukset transdusera (FST) ja dainna hukset sátnehámiid. Muhto FST

čuolbma lea ahte dat generere maiddái sátnehámiid mat eai leat sámegielagiid gielas. Dán kapihttalis giedahalan dán čuolmma.

3.3.1 Eai buot sánit leat geavahusas

Kapihttalis 2.3.2 muitalin ahte 30 % davvisámi Giella-sme substantiiva- ja vearbalemmain eai gávdno SIKORis. Dát lemmat leat dattetge duodaštuvvon leat giellamáilmmis dainna lágiin ahte dat gávdnojit dohkkehuvvon tearbmalisttuin, dahje daid vuodđu lea mo normerenmearrádusaid mielde heivehivččii dárú sáni dahje nama sámegiela loatnasátnin. Ollu dain sániin bohtet domeanain mat eai leat mielde SIKORis.

Giella-smX analysáhtoriin lea oktasaš nammaleksikon mas leat 30 000 nama. Leksikonas leat ollu namat mat eai leat vuos geavahuvvon buot sámegielain. Ovdamearkka dihte ii leat nu stuorra vejolašvuohta gávdnat *Kampen*³³-nama kásussojahemiid anárašgiela teavsttain, iige *Heinola*³⁴-nama inessiivahámi *Heinolisnie* lullisámi teavsttain. Muhto dalle go muhtun čállá lullisámegillii mátkereivve Suomas, de Divvun-prográmma veahkeha su čállit nama *Heinolisnie* norpma mielde.

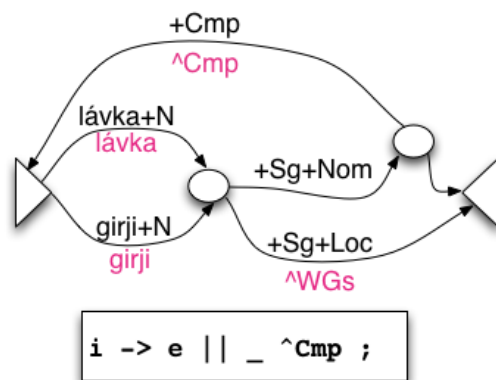
Sámegielaide gullet maiddái goallossánit, ja mearusoassin sáhttet leat máŋga sátneluohká sánit. Ii leat vejolaš leksikaliseret buot goallossániid go gielaide jámma ihtet ođđa goallosteamit. Analysáhtoris lea dan dihte dynámalaš goallosteapmi go analysáhtor ieš hukse goallossáni go lea goallostanbálggis sáni mearusoasis vuodđooassái, muhto dát ráhkada maiddái sániid mat eai leat geavahusas.

Dynámalaš goallostemiin FST hukse buot vejolaš substantiiva-substantiiva-goallostemiid go manná bálggis buot substantiivvaid joatkkaleksikonain. Bálgái lasihuvvojit gilkor +Cmp ja báhcahat $\hat{C}MP$ ja de bálggis manná substantiivaleksikonii (govus 3.8). Morfofonologalaš njuolggadusat mat addet allegrohámi (oanidit vokála), čujuhit $\hat{C}MP:i$. Mearusoassi oažžu juogo allegrohámi dahje largohámi dahje goappašagaid, substantiivvaid iešgudetlágan joatkkaleksikonain.

Maiddá suorggiduvvon substantiivvat sáhttet leat mielde dynámalaš goallosteamis. Dan dihte manná goallostanbálggis vearbaleksikona, adjektiivaleksikona ja substantiivalek-

³³Oslo gávpotoassi.

³⁴Unna gávpogaš Lulli-Suomas.



Govus 3.8: Dynámalaš goallosteapmi. +Sg+Nom-dilis mannet guokte bálgá, nubbi loahpalaš dillái ja nubbi ruovttoluotta vuosttaš dillái. Manjit bálgás lea gilkor +Cmp ja báhcahat ^Cmp mii báhcaha vokála rievdadeami *i.e.* Transduser addá sátnehámiid *girjelávka lávka girjji*. Muhto transduser hukse maiddái hámiid nugo *girjegirji lávkalávkkas girjelávka girjji*. Dát transduser ii dohkket goallosteami man mearusoassi lea lokatiivvas, nugo **girjjislávka*. (Artihkkala I govus 3.)

sikona gaska. Diakrihtalaš leavggaid bokte (čilgejuvvo kapihttalis 3.3.3.3) addojuvvo gáržžidus das ahte goallosteapmi dáhpáhuvvá dušše jus sihke mearusoassi ja vuoddoassi leat substantiivvat, ležžet dal lemmat dahje suorggádušat, nugo *bivdinvejolašvuohta*:

bivdinvejolašvuohta

bivdit+V+TV+Der/NomAct+N+Cmp/SgNom+Cmp#vejolaš+A+Attr+Der/vuota+N+Sg+Nom

Goallostanbálgát mannet maiddái lohkosániin dihto substantiivvaide ja adjektiivvaide, nugo *guovttejuolggat*:

guovttejuolggat guokte+Num+Cmp/SgGen+Cmp#juolgi+N+Der/t+A+Sg+Nom

Dynámalaš goallosteapmi ráhkada maiddái sániid maid ii oktage geavahivčče semantihkalaš sivaide geažil, nugo govvošis 3.8 mas *girjelávka* lea vuogas sátni, muhto *girjegirji* vejolaččat ii oro heivemin masage. FST dohkkeha maiddái sániid main leat vaikko man ollu mearusoasit, ovdamearkka dihte *girjelávka girjelávka*. Vaikko FST goallosta sániid beroškeahtta semantihkas, de leat morfologalaš hámit ráddjejuvvon daidda mat gielas leat geavahusas. Deskriptiiva FST diktá buot mearusosiid goallostit ovttaidlogu nominatiivvas ja genitiivvas, ja mánggaidlogu genitiivvas.

3.3.2 Giellamodealla gáržžideapmi norpma ektui

Go FST galgá máhttit analyseret buot sániid teakstakorpusis, de dat ferte maiddáid dovdat sániid mat eai leat čállojuvvon norpma mielde. Sánit ja bálgát mat ráhkadit hámiid mat eai leat riektáčállinnorpma mielde³⁵, merkejuvvojit normerengilkoriin, nugo +Err/Orth (eng. ‘Error Orthography’ = boasttučállinhápmi). Muhtun normeren-gilkorat karakteriserejit čállinmeattáhusa, nugo +Err/Orth-a-á muitala ahte lea *a* dan sajis go *á*, dahje nuppe ládje. Gilkora sáhtta lasihit eaŋkilhápmái, nugo *dássašii*-adverbii man normerejuvvon čállinhápmi lea *dássažii*. Jus lasiha gilkora lemmai, de dat oažžu sojahanhámiid joatkkaleksikona bokte, nugo *kántuvra* (*kantuvra*), ja dien láhkai genererejuvvo olles paradigma man hámit eai čuovo norpma: *kántuvrra*, *kántuvrii*, *kántuvrras*...

Maiddáid morfologalaš bálgái lea vejolaš lasihit +Err/Orth-gilkora, ovdamearkka dihte bálgái mii addá *:as* lokatiiva geažusin buot iniálaoanádusaide ja arábalaš loguide, dan sajis go norpma mielde *:s*. Dáinna lágiin analysáhtor dovda hámiid *NRK:as* ja *2015:as* vaikko sánit eai leat čállojuvvon norpma mielde, ja analysas lea +Err/Orth-gilkor amas boahit mielde normatiiva geavahussii.

Gilkor +Err/Lex (eng. ‘Error Lexical’ = boasttulemma) lasihuvvo lemmai mii lea geavahusas, muhto dattetge lea norpma olggobealde, nugo loatnasániide mat eai adnojuvvo dohkálažžan, ovdamearkka dihte *fallskejarpma*, ja sániide mat eai čuovo dábalaš suorggideami, nugo *falášteaddji* (buoret livččii *faláštalli*).

Bálgát main leat +Err-gilkorat, leat mielde deskriptiivvalaš FST:s³⁶, muhto váldojuvvojit eret dalle go ovttasta normatiivvalaš FST³⁷, mainna generere normatiivvalaš sátnehámiid sátnedárkkistanprográmmii ja sátnegirjeartihkkaliid sojahanparadigmaide. Nugo namuhin kapihttalas 2.3.3, de ožžot 4,8 % SIKOR sániin dušše deskriptiiva +Err/Orth-analyssa, muhto mángga sánis sáhtta leat sihke normatiiva ja +Err/Orth-analyssa, ovdamearkka dihte *lohka*-sátni oažžu analysa sihke substantiivan (*lohka+N+Sg+Nom* ja

³⁵Norbma vuodđuduvvo ovddeš Sámi giellarádi ja Sámi giellalávdegotti ja dálá Sámi Giellagáldu mearrádusaide, čállingiela prinsihpaid generaliseremii, prentejuvvon sátnegirjjiide ja grammatihkaide (njálm-málaš diehtu Thomas Ommas).

³⁶`analyser-gt-desc.xfst` ja `analyser-gt-desc.hfst`

³⁷`analyser-gt-norm.xfst` ja `analyser-gt-norm.hfst`

lohkka+N+Sg+Acc) ja +Err/Orth-analyisa vearban (lohkat+V+TV+Ind+Prs+Sg3+Err/Orth, man rivttes čállinhápmi livččii *lohká*).

Sátneárkkistanprográmma ja eará čállindoarjjaprográmmaid dihte lea dárbu gáržžidit sániid dynámalaš goallosteami mearusosiid hámiid norpma ektui. Normatiiva FST ovdaválljen³⁸ lea ahte mearusosiid álo lea ovttaidlogu nominatiivvas. Dasa lassin sáhtá gilkoriid bokte addit eanet bálgáid. Leksikonii animáhta mearusosiid merkejuvvo +CmpN/PIG-gilkoriin (N = enj. ‘Normative’) danne go sáhtá leat mánggaidlogu genitiivvas, nugo sánis *mánáidskuola*. Seamma láhkai lasihuvvo gilkor +CmpN/SgG daidda substantiivvaide mat sáhttet goallostuvvot nuppi substantiivvain go leat ovttaidlogu genitiivvas, ovdamearkka dihte *gusamielki*. Muhto normatiiva analysáhtor dohkkeha maid dái *gussamielki*, danne go *gussa* sáhtá lea ovttaidlogu nominatiivvas eará goallosániin, nugo *gusanávet*. Seamma láhkai normatiiva analysáhtor dohkkeha *gusanávet*, danne go *gusamielki* lea vejolaš. Vaikko Divvun-prográmma dohkkeha *gusanávet* ja *gussamielki*, de odda Divvun-prográmma evttohusvuogádat vuoruhá leksikaliserejuvvon sániid, nu ahte ii evttot daid sániid riektáčállon hápmi.

Sáhtá maid dáid stivret mearusosiid hámi vuoddoosi gilkorá bokte, ovdamearkka dihte +CmpN/SgPIGenLeft gáibida ahte mearusosiid galgá leat mánggaidlogu genitiivvas.

3.3.3 Badjelmearálaš genererema ráddjen

Kapihttalis 3.3.2 lei sáhka mo leksikontransduseris sáhtá earuhit deskriptiiva ja normatiiva hámiid vai sáhtá ráhkadit sihke deskriptiiva ja normatiiva FST. Dán kapihttalis lea sáhka mo garvit sátnehámiid mat rievtti mielde leat riektá čállojuvvon morfologalaš njuolggadusaid mielde, muhto sátnehámiid eai geavahuvvo sámegeielas, ja dahket geava-heddjiidprográmmaid fuonibun.

Normatiiva FST vuodul sátneárkkistanprográmma merke sátnehámiid mat eai leat čállojuvvon norpma mielde. Muhtun čállinmeattáhusaid prográmma ii gávna danne go sátnehápmi oázžu eará normatiiva analysa – ja lea duohtasátneattáhus. Jus FST generere sátnehámiid mat eai geavahuvvo geielas, de dát hámiid sáhttet dahkat eanet čállin-

³⁸Goallosteami posišuvdna- ja hápmegilkorat doibmet dušše HFST kompileremis, danne go XFST ii máhte rievdadit bajit dási gilkoriid guovttedásat diakrihtalaš leavgan.

meattáhusaid duohtasátnemeattáhussan go dat mii livččii riehta, ja dat bilida prográmma beaktivuoda. Sáhtta dadjat ahte dákkár hámit *čihket* ii-sátne-meattáhusaid.

Artihkkalis III bodii ovdan ahte Divvun-prográmma gávdná 78 % dutkanmateriála čállinmeattáhusein, ja buot čállinmeattáhusat leat ii-sátne-meattáhusat. Dat mearkkaša ahte 22 % čállinmeattáhusein leat duohtasátnemeattáhusat, maid Divvun-prográmma ii sáhte gávdnat. Jus sáhtášii unnidit badjelmearálaš genererema, de eanet čállinmeattáhusat šattašedje ii-sátne-meattáhussan. 82 % dáhpáhusain main Divvun-prográmma gávdná čállinmeattáhusa, prográmma máhtta evttohit relevánta čállinhámi. Maiddái dán logu sáhtášii buoridit jus ráddješii badjelmearálaš genererema.

Čuovvovaš vuollekapihttaliin attán muhtun davvisámegeiela ovdamearkkaid mat leat áigequodilat sátnedárkkistanprográmma ektui, danne go dát badjelmearálaš hámit sulaštattet eará morfologalaš hámiid. Vuos geahččat mo sáhtta ráddjet dynámalaš goallosteami posišuvdna-gilkoriiguin. Ja de čilgen mo leksikontransduseris sáhtta ráddjet oamastangehčosiid lasiheami amaset čiehkát lokatiivahámiid ii-sátne-meattáhusaid, ja kapihttaliin dan manjel attán guokte ovdamearkka das ahte maiddái vearbagererema ráddjemiin sáhtášii buoridit Divvun-prográmma beaktivuoda.

Kapihttaliin 3.3.3.6 ja 3.3.3.7 suokkardalan finihhta veorbahámiid ráddjema, in dušše sátnedárkkistanprográmma dihte, muhto maiddái prográmmaide mat fáallet sojahanparadigmaid geavaheddjiide, nugo NDS.

3.3.3.1 Substantiivaid posišuvnnat goallosteamis

Lea giksi go normatiiva FST dohkkeha dábálaš čállinmeattáhusa dynámalaš goallossátnin. Sátne *gaskkavahkku* sáhtta leat *gaskavahkku*-sáni boasttuhápmi, muhto maiddái sániid *gaska* ja *vahkku* goallosteami. Semantihkalaš sivaid geažil dát goallosteami ii leat vuorddehahti, ja mii háliidit baicce ahte sátnedárkkistanprográmma galgá hilgut sáni *gaskkavahkku* ja evttohit *gaskavahkku* dan sajis.

Dan sáhtta dahkat go stivre substantiivaid posišuvnna goallosteamis. Dasa geavahit +CmpNP-gilkoriid, mas P mearkkaša posišuvnna: +CmpNP/None (lemma ii sáhte goallostetit), +CmpNP/First (lemma sáhtta dušše goallostetit mearusoassin) ja +CmpNP/Last (lemma sáhtta dušše goallostetit vuodđooassin). Go lasiha +CmpNP/Last-gilkora *gask-*

ka-lemmai, de normatiiva FST³⁹ hilgu buot dynámalaš goallostemiid main lea *gaskka* mearusoassin, earret eará *gaskkavahkku*, ja dalle lea dárbu leksikaliseret vejolaš sániid main lea *gaskka* mearusoassin.

3.3.3.2 Oanehis namaid goallosteapmi

NDS geavaha deskriptiivvalaš FST, mas leat unnit gáržžideamit goallosteapmi go normatiiva FST:s. Muhtun analysat mat eai daga váttisvuodaid korpusanalysii danne go dat jávket disambigueremis, leat unohasat sátnegirjji oktavuodas. Ovdamearkan dás lea oanehis namaid goallosteapmi. Davvisámegielas eai leat ovtastávval substantiivvat, earret moadde odđa loatnasáni, muhto Giella-smX oktasaš nammafiillas leat ollu ovtastávval namat mat sáhttet addit sávakeahtes analysaid. Ohcansátni *Buolbmát* sáhtta oazžut analysa mas lea *Bu* mearusoassin ja *olbmát* vuodđooassin, nugo vuolábeale deskriptiiva analysas:

Buolbmát Bu+N+Prop+Sem/Plc+Cmp/SgNom+Cmp#olmmái+N+Pl+Nom

Buolbmát Buolbmát+N+Prop+Sem/Plc+Sg+Nom

Gielaláččat goallostananalyisa livččii vejolaš, muhto oalle marginála hápmi, ja čáli dáid-dásii lasihit goallossázu, ja čállit *Bu-olbmát*, sáni guovttečilggolašvuoda dihte. Sátnegirjái dákkár analysat eai leat sávahahttit. Čoavddus lei bidjat buot ovtastávval namaid siera nammaleksikonii man bálgás lea diakrihtalaš leavga mii ii divtte sáni goallostit eará sániide goallossázu haga. Diakrihtalaš leavga geavahus čilgejuvvo kapihttalas 3.3.3.3.

3.3.3.3 Substantiivvat ja adjektiivvat oamastangehčosiiguin

Leksikontransduseris joatkkaleksikon mearrida sáni morfologijja. Lemmat *eadni* ja *rudni* mannet seammá joatkkaleksikonii, ja ožžot dan dihtege seamma sojahanparadigma, suorggideamiid ja goallostemiid. Muhto sámegeielat olmmoš ii dáidde geavahit buot dán guovtti leksema sátnehámiid. Vejolaččat huksejuvvojit sátnehámit maid dušše dihtor dovdá.

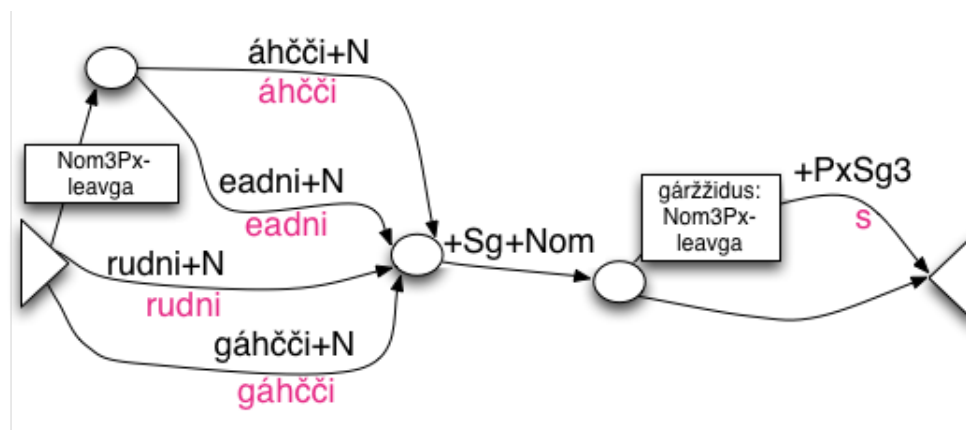
Go guorahallá SIKORa, de oaidná ovdamearkka dihte ahte buot substantiivvaide ja adjektiivvaide eai geavahuvvo buot oamastangehčosat. Sg3-oamastangeažus lasihuvvo

³⁹+CmpNP-gilkorat doibmet dušše `analyser-gt-norm.hfst:s`

substantiivva nominatiiva-hápmái measta dušše go lea fuolkevuodanamahus dahje sulasaš⁴⁰: *eadnis* lea dohkálaš, nugo cealkagis (1), muhto *rudnis* ii livčče lunddolaš geavahit makkárga cealkagis (Antonsen ja Janda, 2015).

- (1) *Sii hálidedje gohčodit su Sakarjan áhčis mielde, muhto eadnis dajai ...* (SIKOR: Ođđa testameanta)

Genererema sáhttá ráddjet joatkkaleksikonaidd bokte: Livččii vejolaš ráhkadit sierra joatkkaleksikona *rudni*-lemmai, vai dán lemma bálgás ii leat +Nom+PxSg3, eaige eará nominatiiva goalmmát persovdna-hámit (Px = enj. ‘Possessive suffix’ = oamastangeažus). Buoret čoavddus lea geavahit diakrihtalaš leavggaid, mat leat vuogas reaidun diekkár gáržžidemiide danne go dat eai leat čadnojuvvon dihto joatkkaleksikonii ja dan sojahanmállii, muhto leavggaid bokte sáhttá ráddjet vaikko makkár lemma leksikonas. Diakrihtalaš leavggat eai oidno analysas, muhto daiguin addit muhtun árvvu dihto lemmaide dahje bálgáide, ja de eará sajis fas bidjat gáržžidusa mii hilgu dahje luoitá sániid mas lea diet árvu (govus 3.9).



Govus 3.9: Transduser mas lea diakrihtalaš leavga. NOMSG3-leavga lasihuvvo leksikonii mas leat lemmat *eadni* ja *áhčči*. Bálgás mii addá oamastangehčosa nominatiivahápmái, lea gáržžidus mii gáibida NOMSG3-leavga, ja dan dihte bálggis ii doaimma buot lemmaide. Transduser generere sániid *eadni*, *rudni*, *áhčči*, *gáhčči*, *eadnis*, *áhččis*, muhto ii sániid *rudnis*, *gáhččis*.

Artihkkalis I čállen ahte badjelmearálaš genereren ii daga váttisvuodaid danne go dat bohtá oidnosii dušše jus muhtun čállá sáni mas lea sáhka. Muhto jus FST dohkkeha sáni *rudnis*, de sátnedárkkistanprográmma ii hilggo sáni, vaikko sáni normerejuvvon jovs-

⁴⁰Sullasaš substantiiva lea Ođđa testameanttas *oahpaheaddji*, ovdamearkka dihte cealkagis *Máhttájeaddji ii leat stuorit go oahpaheaddjis, ii ge bálvleaddji stuorit go isidis*.

sushápmi livččii *rutnis*, nugo *rudni+N+Sg+Loc:rutnis*. SIKORis lea okta ovdamearka das:

- (2) *Juohkeolbmovuoigatvuodain sáhtta oaggut **rudnis** [rutnis] dahje geasset.* (SIKOR: Lundui.fi)

Danne lea dehálaš ráddjet genererema. Min geavtlaš čoavddus lea ahte substantiivaleksikon juhkkovuoigatvuodain dan mielde man ollu oamastangeažushámit galget genererejuvvot (tabealla 3.2). Buot lemmat maidda galgá NOMSG3-leavga, čohkkejuvvot sierra substantiivaleksikonii, man namma lea NOUNPXKIN, ja leavga biddjojuvvot bálgái mii manná dán leksikonii, seamma láhkai go govvois 3.9. Dainna lágiin leksikontransduseris leat maiddá sánit mat eai obanassiige oáččo oamastangehčosiid, NOUNNOPX, go bálgáin mii manná oamastangeažusleksikonaide gáibiduvvo PX-leavga. Stuurimus oassi substantiivaleksikona lemmain gullet NOUNPX:i, mii lea ovdaválljejuvvon leksikon odđa lemmain. Muhto lingvistica geavaha iežas gielladovddu ja dasa lassin SIKORa veahkin, go sirdá lemmaid NOUNNOPX:i, ja son dahká dan erenoamážit dalle go hámit main lea oamastangeažus, sáhttet čiehkát dábálaš čállinmeattáhusaid. Adjektiivaleksikonas lea fas nuppelágan dilli, go dušše guoktenuppelot adjektiivalemma gullet LEXICON ADJECTIVEPX:i, namalassii adjektiivvat⁴¹ maid leat gávdnan SIKORis ja eará teavsttain oktan oamastangehčosiin.

Substantiivaleksikonat	Oamastangehčosiid lasiheapmi
LEXICON NOUNPXKIN	buot hámit
LEXICON NOUNPX	buot hámit earret +Nom+PxSg3/PxDu3/PxP13
LEXICON NOUNNOPX	eai lasihuvvo hámit oamastangehčosiiguin

Tabealla 3.2: Substantiivaid leksikonat stivrejit oamastangehčosiid lasiheami.

Báhcá čuolbman ahte go lea vejolaš čállit +Nom+PxSg3-hámi, nugo *eadnis*, de dattetge dát čállinhápmi dávjá lea čállinmeattáhus. SIKORis leat 21 *eadnis*-hámi, muhto dušše čieža dain leat nominatiivahápmi oktan oamastangehčosiin.

Substantiivvat oktan oamastangehčosiin dahket stuurra paradigma mas leat 105 sealla

⁴¹*nuorra, buorre, uhcci, unni, boaris, ipmilbalolaš, ráhkis, vistelágaš, láhkásaš, vártnuheapme, ovddeš, suohtas*

ja 79 iešgudetlágan hámi. Dán kapihttalas lean geahččan +Nom+PxSg3-hámiid, muhto maiddá eará sojajeamat eai leat geavahusas buot substantiivvaiguin. Oamastangeažus-geavahusa berresii guorahallat eanet, ja dan vuodul lasihit gáržžidusaid FST:i.

3.3.3.4 Vearbbaid imperatiivva biehttalanhápmi

Davvisámegiela lea vejolaš ráhkadit guovttelágan biehttalanhámiid, main nubbi geavahuvvo dušše goalmát persovnnaid imperatiivva biehttaleami váldovearban. Imperatiivva biehttalanhápmi lea geavahuvvon boares oskuteavsttain, nugo biibbalis (Nickel ja Sammallahti, 2011, s. 61). Sojahanhámi sáhttá árvoštallat leat jávkan gielas, go ođđaseamos davvisámegiela grammatihkas dákkár hápmi ii namuhuvvo imperatiivva biehttalanhámiid gaskkas (Svonni, 2015, 107). Dattetge imperatiivva biehttalanhápmi gávdno SIKORis Ođđa testameanttas, gos hápmi lea geavahuvvon oktiibuot logi vearbalemmas. Go dasa lassin geavahan davvisámegiela biibbalveršuvnna mii lea interneahas⁴², de gávnnan oktiibuot 30 vearbalemma. Dain leat 18 bárrastávvalvearbbat⁴³, nugo cealkagiin (3) ja (4). Muđui leat bárahisstávvalvearbbat⁴⁴ ja golbma kontrakšuvdnavearbbat⁴⁵.

(3) *Allos oktage **atno** mu jallan.* (SIKOR: Ođđa testameanta)

(4) *Allos oktage **behtto** iežas!* (SIKOR: Ođđa testameanta)

FST generere dákkár imperatiivva biehttalanhámiid eanaš bárrastávvalvearbbain, ja hámit merkejuvvojit +ConNegII-gilkoriin. Imperatiivva biehttalanhámis lea gievrras dássi ja soggevokála *o*, ja dábálaččat dát leat sátnehámit mat maiddá ožžot analysa finihtta vearban, nugo *ožžo*, *borro*, *ođdo*. Muhto dalle go passiivavearba gáibida buot gievrramus dási, nugo *dahkko*, *biddjo*, de imperatiivva biehttalanhápmi čiehká dábálaš čállinmeattáhusa go passiivavearbbas lea gievrras dássi, nugo sánit čállojuvvon buoiddes bustávaiguin cealkagiin (5) ja (6). Normatiivva hámit leat roahkkeruoduid siste. Cealkagis (7) hápmi fas čiehká eará čállinmeattáhusa, go *-hallat*-vearbbas, mii lea passiiva, čállojuvvo *-ot*-passiivahápmi. Normatiivva hápmi lea *vuottáhalla*.

⁴²<https://www.bibel.no/Nettbibelen> (23.11.2017)

⁴³*addit, atnit, beahhtit, boahhtit, báhcit, dahkat, goargnut, gullat, heahpanaddat, jearrat, jáhkkit, oaidnit, sivahallat, sárdnut, vahágahttit, vuolgit, ássat, šaddat*

⁴⁴*árpmihit, čájehit, dorvvastit, illudit, leat, rámidit, suorganit, váivvidit, veahkehit*

⁴⁵*gillát, loktet, rápmot*

- (5) *Bargu **dahko** [dahkko] Sámedikkiin ovttasráđiin ja gárvána jáhkkimis cuoŋománu 2010.* (SIKOR: Ávvir)
- (6) *Mánnaí **bidjo** [biddjo] dakkaviđe go riegáda dovdanboalu juolgái.* (SIKOR: Ávvir)
- (7) *Unnitlohku **vuottáhallo** [vuottáhallá].* (SIKOR: Ávvir)

Vaikko grammatihkkagirjjiid mielde imperatiiva biehttalanhámit leat boarásmuvvan hámit, de hámit leat dattetge geavahas biibaljorgalanbarggus mii lea jođus dál. Go hámis lea nu ráddjejuvvon syntávssalaš geavahas, de livččii grammatihkkadárkkistanprográmmii álki fuomášuhttit čállái gii áiggukeahtta geavaha dan hámi. Muhto Divvunprográmma lea álkit sajáiduhhtit buot čállin- ja redigerenprográmmaide, ja nu livččii vuogas ráddjet dáid marginála hámiid sátnedárkkistanprográmmas.

Go analysere davvisámi SIKOR normatiiva analysáhtoriin⁴⁶, de imperatiivva biehttanláhpmi lea 366 sáni áidna analysa, ja dain leat fas 341 duohtasátnemeattáhusa. Sátnedárkkistanprográmma buorránivččii jus leksikontransduseris válldášii eret dán bálgá.

Čoavddus sáhtášii leat leksikaliseret dan golbmalogi hámi mat leat geavahuvvon Ođđa testameanttas ja addit daidda gilgora nu ahte hámit eai leat mielde dábálaš Divvunprográmmas, ja vejolačcat dušše váldit daid mielde sierra Divvunprográmmii oavvilduvvon biibal- ja oskuteavsttaid bargui. Dát dagašii Divvunprográmma beaktileabbon go gávnašii dán áigge dábálaš čállinmeattáhusaid, nugo bajábeale ovdamearkacealkagiin (5)–(7).

3.3.3.5 Vearbagenitiiva

Vearbagenitiiva lea vearbba infinihhta hápmi, mii sáhtta almmuhit vuogi, mainna lágiin juoga dáhpáhuvvá, nugo *suga cealkagis - Dál bohtet suga gáddái.* (SIKOR: Áššu). Vearbagenitiiva lea dasa lassin dihto advearbalágan dajaldagain preposišuvnnain ovttas, nugo *lahka jámi.* Vearbagenitiiva ii leat nu produktiiva, ja davvisámegielas lea tendeansa eanet geavahit eará infinihhta vearbahámiid, nugo instrumentála konvearba *-miin* (Ylikoski, 2009, 80–81) ja (Ylikoski, 2002, 88–97).

Vearbagenitiiva sáhtta maiddá geavahuvvot reduplikatiiva vearbagenitiivan, mas guokte ideanttalaš vearbagenitiivva leat cealkagis manjálaga. Dakkár geavaheamis leat seman-

⁴⁶analyser-gt-norm.hfst

tihkalaččat guokte iešguđetlágan mearkkašumi, proksimatiiva vearbagenitiiva, ahte juoga lea measta ollašuvvan, mii ii leat produktiiva, ja geavahuvvo dušše mottiin vearbain, nugo *boadi boadi*, *geargga geargga*. Nubbi mearkkašupmi lea intensiiva vearbagenitiiva, mii ráhkaduvvo lihkanvearbain (guhtta) ja jietnadanvearbain (oktanuppelohkái). Jompanen (2015).

Bargadettiin Giella-sme disambiguáhtoriin lean SIKORis ja eará teavsttain, maiddái grammatihkain ja bajábealde namuhuvvon artihkkaliin, gávdnan oktiibuot 79 vearbba main lea vearbagenitiivahápmi, buot bárrastávvalvearbbat⁴⁷ earret guokte bárahisstávvalvearbbat⁴⁸ ja guokte kontrakšuvdnavearbbat⁴⁹.

Vearbagenitiivahápmi lea homonyma vearbba preseansa biehttalanhámiin earret bárrastávvalvearbain main lea *-ut* dahje *-it* loahpas infinitiivvas, ja nu sáhtta leat duohtasátne-meattáhusaid áidna normatiiva analysa.

Čuovvovaš golmma cealkkaovdamearkkas lea biehttalanhápmi dahje finihttahápmi mii ii leat čállojuvvon norpma mielde, ja áidna normatiiva analysa lea vearbagenitiiva, mii dainna lágiin čiehká ii-sátne-meattáhusa. Cealkagis (8) lea čállojuvvon vearbagenitiiva *oačču* dan sajis go biehttalanhámi normatiiva hápmi *oaččo*. Cealkagiin (9), (10) ja (11) lea čállojuvvon *oačču*, *sihku*, *sáhti* finihtta vearbbahámiid sajis.

(8) *Ii oktage **oačču** [oaččo] vuolil 7850 ruvdnosaš lasáhusa.* (SIKOR: Ávvir)

(9) *Ja juste das **oačču** [oažžu] ge Sámediggi, erenoamážit Finnmárkkus, olu spiža.*
(SIKOR: Ávvir)

(10) *... dušše dan dihte go dát lea dat maŋimuš háve go dát čájálmas čájehuvvo, muitala Siri Broch Johansen dan botta go **sihku** [sihkku] gatnjaliid.* (SIKOR: Ávvir)

(11) *Riddu Riddu festivála jodiheaddji, Ranghild Dalheim Eriksen, lea duhtavaš fes-*

⁴⁷ *bassit bárdnat bealkit biškut boahit borrat bossut cealkit civkit čárvut čállit čeargut čierrut čuoigat čuorvut doapmat duorrat fáiput galkat gáhččat gállit geargat geavzut gievvut gilljut girdit gullat guoddit heahpanaddat hoigat holvut huikit hupmat jápmit johtit juorbut jorgut juoigat lávbut luoibmat murdit njáhkát njoammut njurgut oaddit oahkut oaidnit ohcat orrut ráhtat reaškit rieħčut rievdat riidet rohkut ruohttat ruossut sáđđat sihkelastit skeaikit skierbmut soabbut speažžut steamput suhkat suoibut šlívut šloahtat šnjirgut vázzit viehkat vuodjat vuodjit vuohčut vuoššat*

⁴⁸ *doarggistit garrudit*

⁴⁹ *čohkkát láddjet*

*tiválain, vaikko balai ahte čoaska dálki **sáhti** [sáhtii] muddet veahá festivála.*
(SIKOR: Ávvir)

Vearbagenitiivahámit leat sátnedárkkisteami giksin danne go dain ii leat sierra geažus, ja *ut-* ja *it-*vearbbain dušše soggevokála earuha daid biehttalanhámis. Vearbagenitiiva sáhttá maiddái čiehkát ahte konsonántaguovddáš lea čállojuvvon geahnohis dásis dan sajis go gievrras dásis, nugo cealkagiin (9) ja (10). *it-*vearbbaid preterihtta Sg3-hámis lea seammá sullasaš čuolbma, go galgá čállojuvvot *ii* loahpas, muhto oarjesuopmaniid assimilašuvnna dihte lea dábálaš čállit dušše *i*. Vearbagenitiivahápmi sáhttá čiehkát maiddái dákkár *ii-sátni-meattáhusa*.

Vearbagenitiiva lea áidna normatiiva analysa 981 sánis SIKORis, ja 450 dáhpáhusas dan verbii ii leat dokumenterejuvvon vearbagenitiivageavahus, ja buot dát dáhpáhusat leat rievtti mielde eará sánit mat leat boastut čállojuvvon. Dát guoská 100 iešgudetlágan verbii. Gáržžidus vearbagenitiivabálgáin buoridivččii sihke sátnedárkkistanprográmma ja korpusanalysa.

3.3.3.6 Finihtta vearbahámit

NDS sátnegirji fállá sojahanparadigmaid geavaheaddjái. Jearaldat lea makkár sojahanparadigma galggašii fállat. Sátnegirji standárda sojahanparadigma vearbbaide sisttisdoallá buot finihhtahámiid, ja indikatiiva modusis leat ovcci persovdna-logu mat buot leat guovtti áiggis. Áiggun dás geahččat golbma vearbajoavkku maid ii vealtameahttumit leat nu lunddolaš presenteret buot dáiguin hámiiguin.

Vuosttaš joavku leat vearbbat mat geavahuvvojit dušše dálkki birra. Dálkevearbbat spiehkastit dábálaš vearbbaid sojahanparadigmas, ovdamearkka dihte vearba *arvit*. Giella-sme leksikontransduseris diekkár vearbbat mannet sierra joatkkaleksikonaide mat dušše lasihit +Sg3 finihhtta hápmin: *arvá*, *arvvii*. FST generere buot eará vearbbaid ovcci persovnma-logu hámiid guovtti áiggis.

Muhto lea nubbi vearbajoavku man ii leat nu álki jurddašit ahte galggašii leat vejolaš geavahit vuosttaš ja nuppi persovnmaid sojaheamis, nugo *dáhpáhuvvat*: *?dáhpáhuvan*, *gánnáhit*: *?gánnáhin*, *látnjat*: *?lánjan*. Dán čállošis gohčodan dáid vearbbaid dáhpáhuvvat-vearban. Jus diekkár hámit eai leat geavahusas, de sáhtášii daid genererema ráddjet

analysáhtoris, ja sátnedárkkistanprográmma ii livčče daid dohkkehan iige evttohan daid, ja dien láhkai sáhtášii garvit muhtun duohtasátne-čuolmmaid.

Goalmmát joavkkus leat vearbbat mat mitalit erenoamážit elliid láhttemiid birra, nugo *goarjut*, ja dan vuosttaš ja nuppi persovvna hámit eai gávdno teakstakorpuseris. Dáid vearbaid gohčodan dán teavsttas elliid láhttenvearban. Dákkár vearbbat sáhttet leat várálaččat ráddjet FST:s, danne go čáli sáhtášii geavahit dakkár hámiid máidnasis, jus mun-persovvna lea ealli. Ovdamearkan dasa lea go állat mitala iežas fearániid birra cealkagis (12) ja geavaha *civken*-sáni. Seamma veorbain karakteriserejuvvo olbmuid jietnadeapmi humoristtalaččat cealkagis (13). Dáid ovdamearkkaid vuodul in sihkoše elliid láhttenvearbaid vuosttaš ja nuppi persovvna hámiid giellamodeallas, vaikko eai gávdno SIKORis.

(12) *Šattai issoras riedja, mun **civken** baluin olles čoddagiin ja gazzebiehtár civkkii vel garraseappot.* (SIKOR: Ávvir)

(13) *Duollet dálle gullojit maiddái Sámedikki áirasat **civkimin**.* (SIKOR: Áššu)

Dálkevearbaid morfologalaš genereren lea ráddjejuvvon Giella-sme FST:s, ja dušše guokte finihtta hámi leat mielde sojahanparadigmas maid NDS-sátnegirji presentere geavaheddjiide. Muhto lea jearaldat berrešiigo sátnegirji eksponeret giellaálggahalliid dáhpáhuvat-vearbaid hámiide mat eai leat geavahusas ja elliid-láhttenvearbaid hámiide mat illá leat geavahusas?

Lean välljen ahte NDS ii berreše eksponeret geavaheddjiid dáhpáhuvat- ja elliid-láhttenvearbaid vuosttaš ja nuppi persovvna hámiide. Čoavddus dasa lea ráddjet genererema sátnegirji gáldofiillas, nu ahte dákkár veorbain leat dušše goalmmát persovvna hámit sojahanparadigmas mii addojuvvo geavaheaddjái. Ráddjen dahkko sátnegirji gáldofiillas attribuhtain `context="dat"`. Sojahanparadigmas attribuhtta addá *dat*-pronomena konteakstan, dan sajis go *son: dat goarju*. Maiddá giellaoahppanprográmma genereremis, geahča kapihttala 3.4.1, lea gáldofiillain ráddjenvejolašvuohta, nugo čuovvovaš attribuhttaárvu oaivvilduvvon veorbaid main galget genererejuvvot dušše goalmmát-

persovdnahámit ja daidda lassin moadde infinihttahámi:

`gen_only="+Inf,+Sg3,+Pl3,+ConNeg,+PrfPrc"`.⁵⁰

3.3.3.7 Finihtta vearbajoahanparadigmat eará geahččanguovllus

Giella-sme leksikontransduseris dálkevearbbat mannet joatkkaleksikonaide maid sojahanbálgát leat ráddjejuvvon. Muđui buot vearbalemmat ožžot buot sojahanhámiid mat leat grammatihkkagirjjiid tabeallain, main leat +Sg1-ja +Sg2-hámit bajimusas ja buot ovcci persovvna-logu-hámi orrot leamen seamm dehálaččat. SIKOR davvisámi teavsttat addet nuppelágan gova finihitta vearbba sojahanparadigmat. Juohke šánjeris +Sg3-hámit dominerejit, 48,8 % rájes (oskuteavsttain) gitta 69,6 % rádjai (juridihkalaš teavsttain).

1. ja 2. persovvnaid hámit		3. persovvna hámit	
Sihkkaris hámit	Maid dát hámit (buot hámit)	Sihkkaris hámit	Maid dát hámit (buot hámit)
+Prs+Sg2, +Prt+Pl1	+Prt+Sg2	+Prs+Sg3, +Prt+Sg3	+Prs+Pl3
+Prt+Pl2, +Prs+Pl2	+Prs+Pl1	+Prs+Du3, +Prt+Du3	+Prt+Pl3
+Prt+Du1, +Prt+Du2	+Prs+Du1		
+Prs+Du2			
3321 = 47 %	4192 = 59 %	5992 = 84 %	6694 = 94 %

Tabealla 3.3: Vearbalemmat main leat finihitta veabahámit SIKORis, N = 7120. 94 % dain vearbalemmain leat SIKORis 3. persovdnahámit.

Teavsttaid sisdoallu diedusge váikkuha vearbbaid sojaheapmái. Čáppagirjjálašvuoda teavsttain lea 67,8 % finihitta vearbbain +Sg3-hámit, go teavsttaid čállit leat válljen čállit goalmát persovvnas. Buoret gova oazžut go geahččat vearbalemmaid juohkáseami (tabealla 3.3). SIKOR teavsttain leat 7120 iešguđetlágan vearbalemma finihitta hámis, ja dain dušše 47 % leat 1. ja 2. persovvnaid sihkkaris hámit. Eahpesihkkaris hámit leat dákkárat mat leat homonymat juogo 3. persovvna hámiin dahje infinitiivvain, geahča tabealla 3.5. Go válldán daid hámiid mielde, de 59 % vearbalemmain leat 1. ja 2. persovvnaid hámit.

⁵⁰Maiddái substantiivvain lea sátnegirjji gáldofillas ráddjenvejolašvuotta. Substantiivvat mat dábbalaččat eai geavahuvvo mánggaidlogu hámis, eai genererejuvvo mánggaidlogus, vaikko hápmi gávdno analysáhtoris. Dát stivrejuvvo sátnegirjji gáldofillas attribuhtain seamma vuogi mielde go vearbbain: `gen_only="Sg"`.

Tabeallas 3.3 84 % vearbalemmain leat +Sg3 finihtta hámit. Olles 94 % vearbalemmain leat 3. persovdnahámit go válddán mielde +Pl3-hámiid, mat eai leat sihkkaris hámiid gaskkas. Vaikko maiddái dáid loguid báidná teavsttaid sisdoallu, de dat doarju gova das ahte buot vearbbat geavahuvvojit 3. persovnnas, muhto dušše oassi vearbain geavahuvvojit 1. dahje 2. persovnnas. Dán juohkáseami vuodul sáhtta vearbbaide jurddašit earálágan paradigmastruktuvrra, mii vuoruha 3. persovnna veabahámiid, ja dušše oassi vearbain ožžot 1. ja 2. persovdnahámiid.

Dát boadus heive bures Janda ja Nessel (2010) evttohussii govvidit sojahanparadigmaid radiála kategoriijan mas lea guovddáš prototiipa mas lea gullevašvuohta lahtuide mat leat seamma kategoriijas muhto ravddabut. Ruoššagiela vuodul soai ráhkadeaba dákkár vurdojuvvon huksehusa, mas leat prototiippalaččamus lahttu gurut bealde:

3Sg > 3Pl > 1/2.persovnna hámit > imperatiiva > gerunda/partisihppa

Gielaidrasttildeaddji guorahallamat čájehit ahte typologalaččat nolla-affivssat leat mealgat dábálaččabut goalmmát persovnnas go vuosttaš ja nuppi persovnnain (Bybee, 1985, 52). Maiddái dát heive davvisámegillii go bárrastávval- ja kontrakšuvdnavearbbaide +Sg3 veabahámit leat gehčosa haga. Bárahisstávvalvearbbaide +Sg3-hámis lea geažus *a*, muhto veabahápmi lea geavahusas maiddái gehčosa haga: *muitala* – *muital* (Nielsen, 1979 [1926–29], 160).

Vearbaleksikonat	Finihtta hámit
LEXICON WEATHER-VERB	+Sg3
LEXICON HUMAN-SUBJECT-VERB	+Sg1, +Sg2, +Sg3 +Du1, +Du2, +Du3 +P11, +P12, +P13
LEXICON DEFAULT-VERB	+Sg3, +P13

Tabealla 3.4: Leksikontransduseris: Dáinna lágidemiin dušše dihto vearbbat oččošedje buot finihtta sojahanhámiid.

Sátnegirjjiid ektui čovden veabahámiid badjelmearálaš genererema go lasihin gáržžidusaid sátnegirjjiid gáldofilii. Muhto go dát badjelmearálaš genereren ii leat sávahahtti muđuige, de livččii buoret addit gáržžidusaid FST:i. Sáhtášii lágidit vearbaleksikona seam-

má vuogi mielde go substantiivaleksikon lágiduvvo oamastangehčosiid ektui. Livččii vejo-
 laš juohkit vearbalemmaid golmma leksikonii, nugo tabeallas 3.4 lea dahkkon, ja diakrihta-
 laš leavggaiguin gáržžidit vearbaid paradigmmaid. Vuosttaš leksikonas, WEATHER-VERB,
 leat dálkevearbbat mat ožžot dušše +Sg3-hámiid, ja nuppi leksikonas, HUMAN-SUBJECT-
 VERB, leat vearbbat main sáhhtá leat olmmoš subjeaktan ja daidda genererejuvvojit buot
 ovcci persovnna-logu-hámi. Goalmmát leksikonas, DEFAULT-VERB, leat vearbbat mat
 ožžot dušše +Sg3 ja +Pl3 hámiid. DEFAULT-VERB lea leksikon buot odđa vearbbaide.
 Dušše jus korpus dahje sámegeala lingvista lea duodaštan ahte vearbba subjeakta sáht-
 tá leat vuosttaš dahje nuppi persovnnas, de vearba sirdojuvvo HUMAN-SUBJECT-VERB
 leksikonii.

	Vearbahámiid homonymiijat			
	Ovda- mearka	1. ja 2. persovnnaid hámit	Eará finihitta hámit	Infinihitta hámit
Bst	<i>dáhpáhuva</i> <i>dáhpáhuvvet</i> <i>dáhpáhuvve</i>	+Imprt+Sg2 +Prt+Sg2, +Imprt+Pl2 +Prs+Du1	+Prs+Pl3 +Prt+Pl3	+Prs+ConNeg
Bhst	<i>gánnát</i> <i>?gánnáhat</i> <i>gánnáhan</i> <i>gánnáhit</i>	+Imprt+Sg2 +Prs+Pl1, +Prs+Sg2 +Prs+Sg1 +Prs+Pl1, +Prt+Sg2	+Prs+Pl3	+Prs+ConNeg +PrfPrc, +Prt+ConNeg +Inf
Ktr	<i>savvo</i> <i>savvot</i>	+Imprt+Sg2 +Prs+Sg2, +Prs+Pl1	+Prs+Sg3	+VGen, +Prs+ConNeg +Inf

Tabealla 3.5: Giella-sme: Vearbba dábáleamos homonymiijat. Tabeallas leat giellamo-
 dealla analysat, ja 1. ja 2. persovnnaid hámit eai gávdno SIKORis. Sátnehápmi *gánnáhat*
 gávdno oktii SIKORis, čállinmeattáhussan *gánnáha* sajis cealkagis: ... *das leat nu máŋ-
 ga boasttuvuoda ahte gánnáhat [gánnáha] geahččat...* (Ávvir). Bst = bárrastávvalvearba,
 Bhst = bárahisstávvalvearba, Ktr = kontrákta vearba.

Bajábealde lean čujuhan sátnegirjji dárbbuide, go eat háliit giellaoahpahalliide čájehit
 sojahanhámiid mat eai leat geavahusas. Muhto tabealla 3.4 čoavddus álkidahtášii vear-
 bahámiid disambiguerema ollu. Ollu vearbaid homonymiijat leat 1. ja 2. persovnnaid
 hámiid ja eará finihitta dahje infinihitta hámiid gaskkas. Jus ii generereše tabealla 3.5
 ovdamearkavearbaid 1. ja 2. persovnnaid hámiid, de dušše duppalsárgá olgeš beale

hámit livčče morfologalaš analysan. Vearbba biehttalanhámi analysa lea álki välljet dahje sihkkut danne go dat gáibida ahte biehttalanvearba lea hámi gurut bealde. Dušše pre-seanssa +Pl3 ja infinitiivahámiid lea váttis disambigueret nugo bajilčállagiin, čuoggáidlist-tuin ja cealkagiin main leat eanet oalgecealkagat.

3.3.4 Giellamodealla gáržžideami čoahkkáigeassu

Áidna realisttalaš vuohki sámegielaide ráhkadit analysáhtora, lea hukset FST, mii lea hui fámolaš reaidu sihke analyseret ja genereret sátnehámiid. Teakstakorpuseret eat gávnnáše buot sániid, muhto FST:s lea fas badjelmearálaš genereren čuolbman: Mii genereret sániid mat eai gávdno.

Badjelmearálaš genereren lea guovttelágan. FST generere morfologalaččat dohkálaš sániid mat eai leat dán rádjai geavahuvvon korpuseret. Hárve báikenamaid, vierissániid dahje odda dohkkehuvvon tearpmaid sátnehámit leat gergosat geavahuvvot, go olmmoš čállá teavsttaid odda domeanain.

Nuppelágan badjelmearálaš genereren lea vearrát, nammalassii badjelmearálaš sojahheapmi ja suorggideapmi. Dat leat hámit maid giellačeahppi ii dohkkehivčče. Dát badjelmearálaš genereren lea FST lahknanvuogi boadusin. Go hukse giellamodealla FST:n, de deháleamos gáldut leat álggus sátnegirjjit ja grammatihkkagirjjiid sojahanparadigmat. Huksenbarggu ulbmil lea vuosttažettiin genereret lemmaid buot vejolaš sojahanhámiid, suorggádusaid ja goallostemiid, ja dušše suorggádusat ja goallosteamit mat eai leat produktiivvat ja gullet dušše soames lemmaid, biddjojuvvojit njuolga leksikonfilii.

Grammatihkkagirjjiid sojahanparadigmat addet sátneluohká maksimála paradigma, ja huksetettiin FST dát sojahheapmit addojuvvojit dábálaččat buot sátneluohká lemmaid, go mihttomearrin lea gokčat teakstakorpuseret sániid nu bures go vejolaš.

Lean dán kapihttalas válddahallan golbma vuogi gáržžidit giellamodealla. Gilkoriid lasiheami bokte sáhtta välljet guđe lemmit dahje bálgát galget leat mielde iešgudetge FST kompileremis, ja nu sáhtta ráhkadit normatiiva FST, muhto maidái sierra FST ovdamearkka dihte oskuteavsttaid várás. Nubbi vuohki lea hukset eanet joatkkaleksikonaid, ja dien láhkai ásahtit eanet molssaevttolaš morfologalaš bálgáid. Goalmmát vuohki lea diakrihtalaš leavggaiguin addit muhtun morfologalaš hámiid dušše dihto lemmaid.

Ovdamearkka dihte sáhttit addit vearbagenitiivva dušše vearbalemmaide maidda hápmi lea duođastuvvon korpusis, ja 1. ja 2. persovnnaid finihttahámiid dušše vearbalemmaide maidda gávdno olmmošlaš subjeakta korpusis, dahje lingvistta gielladovddu mielde dat livččii vejolaš.

3.4 Giellamodealla viiddideapmi

Árbevirolaš CALL (eng. ‘Computer-Assisted Language Learning’ = giellaoahppan dihtora vehkiin) vuogádagat fáallet hárjehusaid nugo mánggaválljema (= eng. ‘multiple choice’), sániid dahje cealkagiid oktiibidjama, sániid dahje govaid coahkkaleami ja eaŋkilsániid deavdima. Máhcahat muitala dábálaččat dušše leago vástádus seammá go čoavddavástádus. ICALL (eng. ‘Intelligent CALL’ = jierbmás CALL) geavaha giellateknologalaš analysáhtora, ja dainna lágiin máhtta sihke genereret bargobihtáid ja analyseret geavaheaddji vástádusa ja nu dakkaviđe addit máhcahaga das maid geavaheaddji lea čállán.

Dán kapihttalas lea sáhka heivehit FST nu ahte dat sáhtta addit metalingvisttalaš máhcahaga giellaoahpahallái, sihke morfologalaš ja syntávssalaš máhcahaga. Kapihttalas 3.4.1 válddahalan vuogádaga mii generere morfologalaš hárjehusaid geavaheaddjái, FST:in oktan gáldofiillaiguin mat leat XML-formáhtas. Dát vuollekapihtal oktan artihkkaliin VI lea mielde nákkosgirjjis maiddái danne go dat čilge vuogádaga mii lea vuodđun čuovvovaš guovtti vuollekapihttalii oktan artihkkaliiguin VII ja VIII. Vuogádagas geavaheaddji oažžu metalingvisttalaš meattáhusmáhcahaga, mat leat vuodđuduvvon sátnehámiid morfologalaš proseassaide. Kapihttalas 3.4.2 čájehan mo sáhtta viiddidit giellamodealla nu ahte lea vejolaš addit metalingvisttalaš máhcahaga maiddái dasa maid geavaheaddji lea čállán, ja kapihttalas 3.4.3 viiddidan vuogádaga maiddái addit metalingvisttalaš máhcahaga giellaoahpahalli syntávssas.

Čuovvovaš golmma vuollekapihttala reaiduin ii leat ulbmil ahte buot sánit gielas galget leat mielde, nugo ovddit kapihttaliin, muhto reaiduide lea välljejuvvon sátnevuorká ja morfologiija mat leat relevánttat giellaoahpahallái. Kapihttalas 3.4.1 leat heiveheamit vuosttažettiin vuogádaga gáldofiillain mat leat XML-formáhtas, kapihttalas 3.4.2

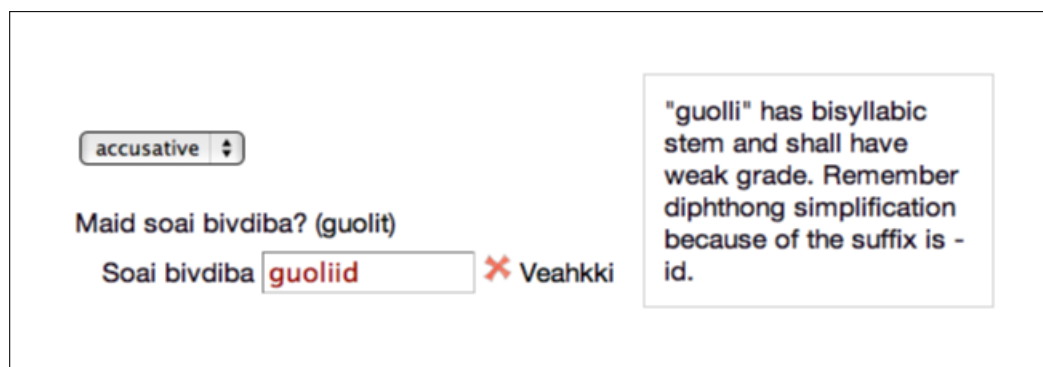
leat heiveheamit dasa lassin FST:s ja kapihttalis 3.4.3 leat heiveheamit maiddái syntávssalaš analysáhtoris.

3.4.1 Grammatihkkahárjehusaid genereren (Art. VI)

Artihkal VI (*Generating modular grammar exercises with finite-state transducers*) čilge mo FST lea geavahuvvon hukset guokte ICALL vuogádaga morfologijaoahpahallamii. Go olmmoš oahpahallá sámegeala, de lea dárbbaslaš hárjehallat ollu grammatihka, ja prográmmat ledje jurddašuvvon luohkkálanja oahpahussii lassin.

Nubbi vuogádat lea buhtes sojahanhárjehallan, ja dat addá geavaheaddjái lemma maid son galgá sojahit konteavstta haga. Spiehkastat lea vearbasojaheapmi, masa geavaheaddji oážžu pronomena ja áigeadvearbba veahkkin ipmirdit guđe hámis lea sáhka. Nubbi vuogádat lea kontekstuálalaš sojahanhárjehallan, mii addá jearaldaga ja vástádusa konteakstan hárjehussii (= QA-hárjehusat). Dán kapihttalis geavahan mañit vuogádaga ovdamearkan mo sáhttá hukset XML-fiillaid mat gulahallet FST:in.

Geavaheaddji oážžu jearaldaga, ja vástádušas lea rabas gieddi masa geavaheaddji galgá čállit lemma rivttes sojahanhámi. Lemma addojuvvo geavaheaddjái juogo jearaldaga bokte, dahje addojuvvo sierra (govus 3.10).



Govus 3.10: Rivttes vástáduš lea mánggaidlogu akkusatiivahápmi *guliid*. Máhcahat addá dieduid sáni morfologalaš proseassaid birra mat gusket máddagii, dássemolsumii, diftonjanjuolgamii ja gehčosii: "*guolli* lea bárrastávval sátni mii galgá leat geahnohis dásis. Muite ahte diftonja njuolgá *id*-gehčosa dihte." (Artihkkala VI govus 3.)

Vuogádagas leat golbma dehálaš oasi: leksikon-, hárjehus- ja máhcahatfiillat, mat buot leat XML-formáhtas. Leksikonfiillain leat válljejuvvon lemmat (1200 substantiivva, 750 vearbba, 300 adjektiivva ja dasa lassin muhtun pronomenat, lohkosánit ja advearb-

```

<question>
  <text>Maid SUBJ MAINV luomus</text>
  <element id="SUBJ">
    <grammar pos="Pron"/>
  </element>
  <element id="MAINV">
    <id>bargat</id>
    <grammar tag="V+Cond+Prs+Person-Number"/>
  </element>
</question>
<answer>
  <text>Luomus SUBJ V-COND</text>
  <element game="morfa" id="V-COND" task="yes">
    <sem class="ACTIVITY"/>
    <grammar tag="V+Cond+Prs+Person-Number"/>
    <agreement id="MAINV"/>
  </element>
</answer>

```

Govus 3.11: QA-hárjehusa templáhta. (Govus boahtá artihkkalis VI.)

bat). Lemmaiguin oktan morfologalaš gilkoriiguin ja FST:in genererejuvvot sullii 80 000 sátnehámi.

Hárjehusminstariid elemeanttat leat variábelat (čállojuvvon versálaiguin) ja konstánttat (sátnehámit). Variábel definerejuvvo elemeantan mas lea juogo dihto lemma, sátneluohkká dahje semantihkalaš seahtta mas lemma galgá välljejuvvot. Dasa lassin lea välljejuvvon syntávssalaš funkšuvdna dahje morfologalaš ráddjejuvit. Govvosa 3.11 ovdamearkkas lea variábel SUBJ definerejuvvon pronomenin, MAINV lea definerejuvvon *bargat*-lemman mii galgá leat konditionála modusis. Vástádusas lea variábel V-COND, mii galgá boahit semantihkalaš seahtas ACTIVITY ja leat konditionála modusis. Sierra modula fuolaha kongrueanssa pronomena ja vearbaid gaskkas, ja maiddá jearaldaga ja vástádusa gaskkas, nu ahte jus jearaldaga subjeakta ja verbála lea +Du2, de vástádusa subjeakta ja verbála lea +Du1. Hárjehusovdamearkkas V-COND-variábel presenterejuvvo rabas gieddin geavaheaddjái, ja lemma mii lea gessojuvvon ACTIVITY-seahtas, addojuvvo vuodđohámis.

Govvosa 3.11 hárjehusminsttar generere QA-hárjehusaid nugo "*Maid doai barggašeidde luomus? Luomus moai ... [vearbba infinitiivahápmi].*" ACTIVITY-seahtas leat 87 vearbba maid lean meroštallan heivvolažžan, ja ovttas iešgudetlágan pronomeniiguin, hárjehusminsttar genererešii oktiibuot 783 iešgudetlágan hárjehusa. Hárjehusaid evalueremis (340 soaittáhagas välljejuvvon QA-bára) adde evaluerejeaddjit juohke hárjehussii

gaskamearálaččat 2,9 čuoggá rivttes grammatihkas, 2,8 dohkálaš semantihkas ja 2,9 pedagogalaš heivvolašvuodas (3 lea buoremus čuokkeslohku).

Vuogádat buohtastahtá geavaheaddji vástádusa sátnehámiin maid FST lea genereren grammatihkalaš sánis mii lea definerejuvvon rivttes vástádussan. Lea eanet go okta rivttes vástádus, jus FST lea genereren variánttaid. Vuogádat sáhtá dohkkehit čállinhámiid main eai leat sáme giela sierra bustávat jus geavaha stávenložžema (govus 3.6).

Vuogádaga meattáhusmáhcahagas ii leat alladásat vuorrováikkuhus, go vuogádat registrere dušše ahte geavaheaddji lea čállán dohkálaš vástádusa dahje boasttuvástádusa. Dattetge geavaheaddji oažžu relevánta meattáhusmáhcahaga jus vástádus ii dohkkehuvvo. Meattáhusmáhcahat lea huksejuvvon dainna ahte buot morfologalaš proseassat leat listejuvvon, ja dat fas kombinerejuvvojit guvttiin paramehteriin, mat leat hárjehustiipa ja lemma iešvuodát mat leat relevánttat máhcahahkii. Lemma iešvuodát leat máttatiipa, diftonŋa, konsonánta, vokálaráddji ja soggevokála, ja dieđut leat vurkejuvvon leksikonfiillas XML-elemeantta attribuhttan, nugo *guolli*-substantiivvas:

```
<l diphthong="yes" gradation="yes" pos="n" finis="0" stemvowel="i" stem="2syll1">guolli</l>
```

Govvosa 3.10 máhcahat sisttisoallá njeallje modula mat leat mánggaidlogu akkusatiivva definišuvnna paramehterat: Jus máttatiipa lea bárrastávval (eng. ‘bisyllabic stem’) (1), de sátnehápmi galgá leat geahnohis dásis (eng. ‘weak grade’) (2), ja diftonŋa galgá njuolgat (eng. ‘diphthong simplification’), jus lemma XML-elemeanttas lea merkejuvvon diftonŋa ja soggevokála *i* (3). Mánggaidlogu akkusatiivva geažus (eng. ‘suffix’) lea *id* (4).

Dát vuogádat lea beaktil ráhkadit morfologalaš hárjehusaid gielaide main lea rikkes morfologiija, muhto unnán čálalaš resurssat. Sáhtá geavahit FST mii lea juo huksejuvvon sátnedárkkistanprográmma várás, dahje hukset sierra FST mas leat lemmat ja morfologiija mat leat relevánttat giellaoahpahussii, ja dainna genereret ollu hárjehusaid automáhtalaččat. Sihke addaga ja buktaga lea vejolaš prosesseret mánggain paramehteriin. Dákkár paramehter sáhtá ovdamearkka dihte leat dihto suopmana hámit ja sojaheamit. Hámiide dahje lemmaid mat gullet dihto suopmanii lasihuvvo +Dial/-gilkor (eng. ‘Dialect’ = suopman), ja sáhtá ráhkadit sierra suopman-FST:id go kompilerenproseassas hilgu bálgáid mat leat merkejuvvon eará suopmaniid +Dial/-gilkoriin. Seamamá ládje lea vejolaš genereret grammatihkkahárjehusaid dihto čállinvariánttain, jus gielas

lea eanet go okta čállinvuohki geavahas, nugo eamiálbmotgielain mángii lea, geahča kapihttala 4 ovdamearkkaid.

3.4.2 Oahpahalli morfologalaš gaskagiela gieđahallan (Art. VII)

Divvun-prográmma lea heivehuvvon eatnigielčállái dainna lágiin ahte vurdo ahte čállí diehtá mo sátni jietnaduvvo, muhto ii vealtameahttumit dieđe sáni rivttes čállinhámi. Giellaoahpahalli gielas, dás gohčoduvvon L2 (eng. ‘Language 2’ = giella maid olmmoš oahppá maŋjel go lea oahppan vuosttaš gielas (= L1)), leat dávjá eará meattáhusat go eatnigielaga gielas. Giellaoahpahalli sáhtta lasihit boasttugehčosa sátnái dahje ii muitte váldit vuhtii buot morfofonologalaš beliid. Boađus lea sátni man jietnadeapmi ii su- lastahte jovssushámi. Govvosis 3.12 oaidnit ahte giellaoahpahalli lea lasihan *i*-gehčosa *eadni*-substantiivii, muhto ii leat muitán soggevokála rievddadeami *i:á*, ja čállá *eadnii* illa- tiivahápmiin, dan sajis go *eadnái*. Divvun-prográmma evttohanalgoritma ii bastte addit rivttes sátnehámi čállái, vaikko divvungaska ii leat go okta, danne go prográmma ii vuorut *i:á* meattáhusa.

Mun attán girjji eadnii.

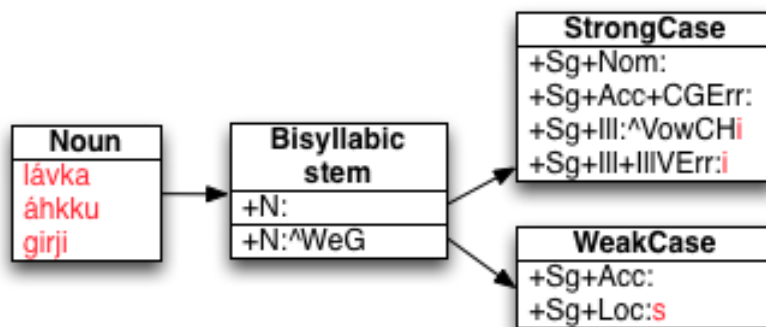
eadni
eatni
adnii-
ednii

Govus 3.12: Divvun-prográmma ii máhte dás addit relevánta evttohusa mii livččii *eadnái*.

Artihkkalis VII (*Improving feedback on L2 misspellings – an FST approach*) lean huksen L2-FST masa lasihin giellaoahpahalli dábálaš boasttuhámiid oktan iešgudetlágan meattáhusgilkoriiguin vai lea vejolaš addit metalingvisttalaš máhcahaga. Govvosa 3.12 ovdamearkkas Giella-sme analysáhtor ii atte makkárge analysa *eadnii*-sátnái, muhto L2-FST:in sátni oččošii dákkár analysa: *eadni+N+Sg+Ill+IllVErr* (eng. ‘Illative Vowel Error’ = illatiivva vokálameattáhus). *+IllVErr*-gilkor sáhtta báhcahit metalingvisttalaš máhcahaga: ‘*Muitte soggevokála rievddadeami illatiivvas.*’.

Dábálaš giellaoahpahalli boasttučállin lea maiddái čállit *a* gokko galggašii leat *á*. Kapihttala 3.2.2 čoavddus lea stávenložžen, mainna sáhtášii sihke gávdnat guđe lemma

oahppi lea čállán, ja maiddái hámi morfologalaš dieđuid. Muhto ii boade ovdan makkár boasttučállimis lea sáhka, ja nu ii leat vejolaš addit metalingvisttalaš máhcahaga oahpahallái. Dasa dárbbáša sierra meattáhusgilkora. Hástalus lea addit sihke stávenložžema ja meattáhusgilkora analysáhtorii.



Govus 3.13: L2-FST sisttisdoallá gaskagiela gehčosiid ja morfofonologiiija: Transduser addá gehčosiid *s* ja *i* ja báhcahagaid \wedge WEG dássemolsuma ja \wedge VOWCH vokálarievdadeami várás. Lassebálgát dáid báhcahagaid haga leat gilkorastojuvvon +IllVErr ja +CGErr (enjg. ‘Consonant Gradation Error’ = dássemolsunmeattáhus). (Artihkkala VII govus 3.)

Boasttuhámiid lasihin L2-transduserii njeljiin vugiin, álo ovttas +Err-gilkoriin mii muitala makkár boasttuhámis lea sáhka.

1. Leksikontransduseris adden lassebálgáid gehčosiidda mat leat norpma olggobealde, ovdamearkka dihte bárahisstávvalsubstantiivvaid illatiivagehčosa *i*, mii addá *hivset*-substantiivva illatiivahámi *hivsegi* dan sajis go čállingiela *hivsegi*.
2. Leksikontransduseris adden bálgáid sihke TWOLC-báhcahagain ja TWOLC-báhcahaga haga, ja transduser generere ovdamearkka dihte akkusatiivahámi sihke dássemolsumiin ja dássemolsuma haga, nugo *áhkku*, *áhkku* govvosis 3.13.
3. Morfofonologalaš transduseris adden globála njuolggadusaid, ovdamearkka dihte ahte juohke *á* galgá rievdat *a:n*. Leksikontransduseris bálggis lea merkejuvvon +Err-gilkoriin sihke bajit ja vuolit dásis, ja TWOLC-njuolggadus jávkada gilgora dušše jus rievdan ii dáhpáhuva.
4. Konkatinerejin nuppiin transduseriin mas leat buot báikenamat unna álgobustávain ja gilkor +LowercaseErr mii karakterisere čállinmeattáhusa mas báikenamma

lea čállojuvvon unna álgobustávain. Dát lea dábálaš čállinmeattáhus buot čállin, ii dušše L2-gielas, ja FST-viiddideapmi dahká álkibun analysáhtorii dulkot geavaheaddji addaga.

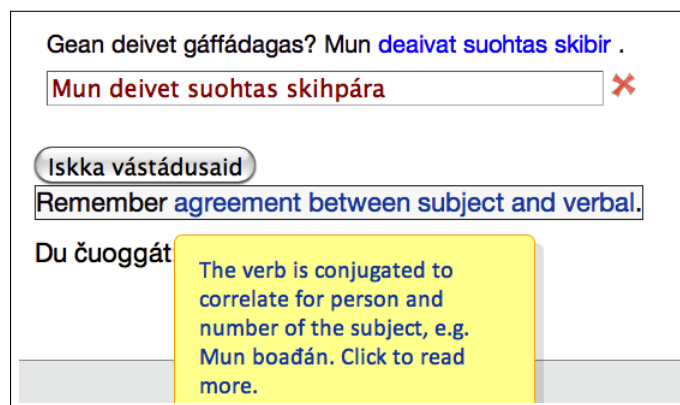
Boasttuhámiid gilkorastin dahká vejolažžan gávdat ja kommenteret systemáhtalaš čállinmeattáhusaid. Ii-sátni-meattáhusat ožžot analysa, ja duohtasátnemeattáhusat ožžot liigeanalysa, man vuodul sáhttá evttohit relevánta hámi maiddá dalle go lea stuorra divvungaska. Go analysáhtor identifisere grammatihkalaš sáni beroškeahtta čállinmeattáhusas, de lea maiddá vejolaš vuoruhit addit máhcahaga dušše syntávssas ja badjelgeahččat čállinmeattáhusaid.

Evaluerejin L2-FST beaktivuoda ICALL-vuogádagain mii presenterejuvvo artihkalis VIII. Geavahin vuogádaga geavahanlogga. Dábálaš deskriptiiva FST dovddai dušše 8,1 % geavahanlogga čállinmeattáhusain, ja L2-FST dovddai 44 % čállinmeattáhusain, ja attii maiddá metalingvisttalaš dieđu čállinmeattáhusain. Vaikko analysáhtor addá ollu eanet analysaid juohke sátnái, de dat ii dattetge bilit disambiguerema ráddjejuvvon ICALL-vuogádagas, mainna evalueren lea dahkkon. Čuolbma lea go buot liigebálgát dahket analysáhtora nu ollu stuoribun, ja ádjána guhkit áiggi oažžut analysa go dábálaš deskriptiiva FST:in.

3.4.3 Oahpahalli syntávssalaš gaskagiela gieđahallan (Art. VIII)

Artihkal VIII (*Constraints in Free-input Question-Answering Drills*) válddahallá guokte QA-hárjehusvuogádaga (jearaldat-vástádus) giellaoahppama várás, main geavaheaddji čállá friddja vástádusa jearaldahkii maid vuogádat addá. Nubbi QA-hárjehus sisttisdoallá dialogaid main leat ovddalgihtii čállojuvvon jearaldagat, ja nubbi geavaha jearaldatgeneráhtora mii lea čilgehuvvon kapihttalis 3.4.1. Goappašagat addet metalingvisttalaš máhcahaga geavaheaddjái, ja dialogavuogádagas bohtá dasa lassin čuovvulanjearaldat go vástádus lea dohkkehuvvon, ja nu dat orru ahte dihtor lea ipmirdan geavaheaddji vástádusa. Dainna lágiin dat sulastahtta duohtamáilmmi ságastallama.

Vuogádaga čuolbma lei ahte giellaoahpahalli garvá váttis giellahuksehusaid iežas vástádusas. Čoavddus lei addit eavttuid vástádussii: geavaheaddji ferte vástádusastis geavahit



Govus 3.14: QA-hárjehus. Lemmat maid geavaheaddji galgá sojahit, leat alit ivdnái. Geavaheaddji oazžu metalingvisttalaš máhcahaga, mii lea čilgejuvvon lagabut tool-tipas (fiskes ivdni), mas lea linka grammatihkkačilgehusaide: "Muitte kongrueanssa subjeavtta ja verbála gaskkas. Vearba sojahuvvo subjeavtta persovvna ja logu mielde, omd. *Mun boadán*. Coahkkal jus háliidat lohkat eanet." (Artihkkala VIII govus 2.)

dihto lemmaid mat leat vižžojuvvon stuorra lemmasehtain (govus 3.14). QA-hárjehusaid evalueremis, maid dahken autenttalaš geavahanloggain, bohtá ovdan ahte QA-hárjehusat main leat eavttut, geasuhit studeanttaid eanemusat danne go dat buorebut čájeha ahte lea morfologalaš hárjehallamis sáhka, iige rievtti mielde duohtamáilmmi ságastallamis.

Vástádus oktan jearaldagain analyserejuvvon CG-njuolggadusaiguin. Jearaldat ja vástádus sáddejuvvojit analysáhtorii oktan sreangan. Addojuvvon lemmat leat QDL (eng. ‘Question Delimiter’ = jearaldatrádji) vuolde. Sreanga disambiguerejuvvon muhtun muddui ja CG-njuolggadus lasiha meattáhusgilkora sihke syntávssalaš boasttuvuođaid dihte (govus 3.15), muhto maddái jus lea čállinmeattáhus mas lea +Err-gilkor nugo kapihttalis 3.4.2, dahje jus addojuvvon lemmat eai leat geavahuvvon vástádusas.

Constraint Grammar-fiillas REMOVE- ja SELECT-njuolggadusat muhtun muddui disambiguerejit sisdoalu, muhto ii leat dárbu disambigueret buot, nugo govvošis 3.15 gos finihtta vearbbas leat velge guokte analysa. MAP-njuolggadusaiguin vuogádat gilkorastá vástádusa semantihkalaš sisdoalu (ovdamearkka dihte leago positiivvalaš vai negatiivvalaš vástádus, ja merke sáni mii lea jearransáni vástádus). Vuogádat maddái lasiha gilkorastá jus eai buot addojuvvon lemmat leat mielde vástádusas, ja jus gávdná grammatihkalaš meattáhusa. Grammatihkalaš meattáhustiippaid leat meroštallan ovddalgihtii.

```

"<Gean>"
  "gii" Pron Interr Sg Acc
"<deivet>"
  "deaivat" V TV Imprt P12
  "deaivat" V TV Ind Prt Sg2
"<gáffádagas>"
  "gáffádat" Org N Sg Loc
"<^avastas>"
  "^avastas" QDL
  "deaivat" V
  "suohtas" A
  "skibir" N
"<Mun>"
  "mun" Pron Pers Sg1 Nom
"<deivet>"
  "deaivat" V TV Ind Prt Sg2 &grm-non-agr-subj-v
"<suohtas>"
  "suohtas" A Attr
"<skihpára>"
  "skibir" Hum N Sg Acc
"<.>"

```

Govus 3.15: Jearaldat ja vástáduš addojuvvojit analysáhtorii oktan sreangan. Addojuvvon lemmat leat QDL vuolde. CG-njuolggadus lasiha &grm-non-agr-subj-v verbii danne go geavaheaddji lea vástidan seammá vearbahámiin go jearaldagas lea: *Mun deivet...* dan sajis go *Mun deiven....* (Artihkkala VIII govus 4.)

3.4.4 Giellamodealla viiddideami čoahkkáigeassu

Dán kapihttala fáddá lei mo sáhtta viiddidit giellamodealla nu ahte gokčá maiddá giellaoahpahalli gaskagiela. Morfoloġalaččat dat čovdojuvvo go FST:i lasiha bálgáid mat addet systemáhtalaš boasttuhámiid, main leat iešguđetlágan meattáhusgilkorat mat dahket vejolažžan sihke addit rivttes analysa ja dasa lassin addit metalingvisttalaš máhcahaga.

Giella-sme syntávssalaš analysáhtora disambiguerennjuolggadusat vuodđduvvet dasa ahte cealkagat čuvvot eatnigiela syntávssa, go ulbmil lea teakstakorpusa disambigueren (kapihtal 2.4). Dán kapihttalis lean čájehan mo syntávssalaš analysáhtora njuolggadusaid sáhtta heivehit giellaoahpahalli syntáksii. Analysáhtor máhtta teavstta disambigueret, vaikko cealkagis váilu kongrueansa dakko gokko galggašii leat, dahje objekta lea čállojuvvon nominatiivahápmiin. Analysáhtor maiddá lasiha dákkár dáhpáhusaide gilgora mii sáhtta báhcahit metalingvisttalaš dieđuid čállái. Áibbas friddja čállojuvvon teavsttain dat livččii lean hástaleaddji, muhto giellaoahpahalli attá lea vástáduš ICALL-vuogádaga jearaldahkii, ja jearaldat eaktuda vástádušas dihto vearbasojaheami ja kásusa, juoga mii dahká olu álkibun addit máhcahaga.

Morfoloġalaš viiddideapmi dahká ahte dárbbáša unnit syntávssalaš viiddideami go

juohke sánis leat eanet vejolaš analysat, ja syntávssalaš viiddideapmi fas dahká ahte dárbbasuvvo unnit morfologalaš viiddideapmi.

3.5 Giellamodealla, dan geavahus ja duohta giellamáilbmi

Dán kapihttala fáddá lei giellamodealla heiveheapmi. Artihkkalat III–VIII giedahallet iešgudet vugiid mo geavahit giellamodealla geavaheddjiidprográmmain. Morfologalaš analysáhtor ja geavaheddjiidprográmma galget heivet oktii dego bátnejuvllat, ja lea dárbu analysáhtorii lasihit gilkoriid, mat addet dárbbaslaš dieđuid vai geavaheddjiidprográmma máhtta válljet rivttes lemma dahje addit rivttes sátnehámi. Dáinna lágiin gilkorat addet prográmmaide jierpmi – daid bokte sáhtta generaliseret ja dat doibmet njuolggadusaid ja algoritmmaid referánsan.

Ovttá leksikontransduseris sáhtta kompileret mánga iešgudetlágan FST. Gilkoriiguin sáhtta hehttet dihto lemmaid dahje bálgáid boahtimis mielde analysáhtorii normatiiva giela várás, dahje dihto suopmana várás. Diakrihtalaš leavvgaiguin sáhtta addit gáržžidusa sáni posišuvdnii ja hápmái go gálllostuvvo eará sániide mearusoassin.

Lean erenoamážit giedahallan FST giellamodealla stuorra hástalus genereret dušše sátnehámiid mat geavahuvvojit gielas. Dát mearkkaša hámiid mat gávdnojit teakstadahje jietnakorpusis, dahje maid eatnigielat dohkkehivččii hápmin maid son sáhtášii geavahit. Geavaheddjiidprográmmaid gáldofillain lea vejolaš ráddjet genererema prográmmii mas lea sáhka, ja nu sáhtta ovdamearkka dihte ráddjet giellaoahppama pedagogalaš sivaidda geažil, nugo lean dahkan sátnegirjái ja ICALL-prográmmii, mii lea čilgejuvvon kapihttaliin 3.2.2 ja 3.4.1.

Muhto mii háliidit giellamodealla leat nu lahka duohta sámegiela go vejolaš. Dan dihte lea dárbu gáržžidit giellamodealla sihke analysa ja genererema dáfus, ja dan leat Giella-smX-analysáhtoriin dahkan guovtti láhkai: Mii leat čujuhan lemmaid joatkkaleksikonaide main leat iešgudetlágan morfologalaš bálgát, ja diakrihtalaš leavvgaid bokte stivret guđiin lemmaid bálgát mannet dihto morfologalaš sojahemiide, nugo oamastangehčosiidda. Gáržžidusat mat leat dahkkon ovdamearkka dihte Giella-sme analysáhtorii,

eai leat ollislaččat, ja sáhtášii árvoštallat eanet gáržžidusaid, nugo vearbaid finihtta ja infinihtta hámiin. Badjelmearálaš genereren lea čuolbma, ja morfologalaš analysáhtora gáržžideapmi boahtá buot prográmmaide buorrin.

Kapihttalis 3.3.3 lean erenoamážit geahččan guokte čuolmma mat gusket infinihtta veabahámiid genereremii. Imperatiivva biehttalanhápmi livččii hui ávkkálaš sihkkut normatiiva giellamodeallas, muhto hámit leat dattetge geavahusas, vaikko hui marginála, ja dat geavahus gáibidivččii dalle sierra Divvun-prográmma. Nubbi ovdamearka lea veabagenitiiva, mii ii leat nu produktiiva, ja veabagenitiivva badjelmearálaš hámit čihket dábálaš čállinmeattáhusaid.

Lean maiddái geahččan finihtta veabahámiid. Juohke modusis ja áiggis vearbbaide genererejuvvojit ovcci finihtta hámi. Buot vearbbat earret dálkevearbbat ožžot dálá analysáhtoris buot ovcci persovnna-logu-hámi juohke modusis. Mu evttohus lea hukset vearbaleksikona eará láhkai go dál, vai vearbbat maidda ii leat duodaštuvvon olmmošlaš subjeakta, ožžot dušše +Sg3 ja +P13 hámiid.

Giellamodealla ulbmil lea máhttit analyseret buot sániid teavsttain. Jus teavsttaid čállit leat giellaoahpahallit, de lea dárbu viiddidit giellamodealla fátmastit maiddái sin gaskagiela hámiid mat spiehkastit jovssushámiin. Artihkkala VII fáddá lea mo sáhtta lasihit boasttubálgáid oktan gilkoriin, mii maiddái dahká vejolažžan addit metalingvistatlaš máhcahaga čállái. Maiddái syntávssalaš analysáhtora njuolggadusaid sáhtta heivehit giellaoahpahalli syntáksii, ja artihkal VIII čájeha ahte ICALL gáržžiduvvon vuogádaga siskkobealde, lea vejolaš addit máhcahaga čállái maiddái syntávssalaš meattáhusaide.

Kapihtal 4

FST addá vejolašvuodaid eamiálbmotgielaide

4.1 Anárašgiella ovdamearkan

Giella-smX analysáhtorat leat huksejuvvon gielaide main leat unnán hállit ja unnán teavsttat digitála hámis. Giella-smX gielaid gaskkas davvisámegiella lea stuorimus giella 25 700 hálliin⁵¹. Eará sámegielain leat eanemusat 1000 hálli, ja daid gielaid teakstakorpusiid sátnelogut leat 1,3 miljovna dahje unnit, gč. kapihtala 2.3.2. Goalmmát kapihtala artihkkalat III–VIII giedahallet erenoamážit davvisámegiella ja lullisámegiella geavaheaddjiprográmmaid. Dán kapihtalis geahčan dán stuorit geahčanguovllus: Mo FST addá dárbbaslaš reaiduid davvimáilmmi eamiálbmot- ja veahádatgielaide. Vuosttaš ovdamearka lea go FST oanehis áiggis lea šaddan ávkálažžan vel ovtta sámegillii, namalassii anárašgillii, man FST-bargu álggahuvvui gieskat. Dan maŋjel ákkastalan mo FST huksen vástida maiddái eará eamiálbmotgielaid dárbbuide.

Anárašgiella FST, Giella-smn, lea ođđaseamos analysáhtor Giella-smX joavkkus, ja analysáhtor lea huksejuvvon Giella-vuođdostruktuvrra siskkobealde mii álkidahtta huksenbarggu ja buvttada sátnedárkkistanprográmma (Moshagen ja earát, 2013). Anárašgiella giellaservodat lea dan dihte johtilit ožžon dárbbaslaš geavaheddjiidprográmmaid. Giella-smn huksen čájeha makkár rolla analysáhtoris sáhtta leat giella ealáskahttinbarggus.

⁵¹<https://www.ethnologue.com/language/sme> (24.02.2018)

1980-logus anárašgiella lei measta jávkan hállangiellan, ja dušše guovtti bearrašis giella sirdojuvvui mánáide. Vuosttaš oassi barggus ealáskahttit giella álggii 1997:s ja dalle ása-hedje giellabeasi gos mánát ohppe anárašgiella. Dát bargu lihkostuvai nu bures ahte logi jagi maŋjel šattai dárbbaslažžan nannet rávesolbmuid giellamáhtu oažžun dihte doarvái oahpaheddjiid váldit vuostá anárašgielat mánáid geat álge skuvlii. Giellaservodat dárbbasii maiddá anárašgielagiid eará dehálaš virggiide nugo almmolaš hálddahusain, medias ja girkus. Giellaaktivisttat bidje johtui *CASLE*-prošeavtta⁵² 2009–2010 mas njálm-málaš giellamáhttu lei guovddázis (Olthuis ja earát, 2013, 7–11). Giellaealáskahttinbarggu boadusin leat sullii 450 anárašgiella hálli, muhto dušše moattilogis čállet giella bures, ja geavatlaččat okta olmmoš lea atnán fuola anárašgiella girjjálaš buvttadeamis maŋimuš 25 jagi áigge (Olthuis ja earát, 2016).

Giella ealáskahttinbargu lea dál bohtán dan muddui ahte lea dárbu bidjat anárašgiella čállima guovddázii. Sámi giellaláhka mii lea leamaš fámus 2004 rájes, geatnegahtta Anára gieldda almmuhit dieđuid golmma sámegillii, davvi-, nuortalaš- ja anárašgillii. Skuvllas dárbbasit anárašgiel skuvlagirjjiid, ja YLE Sámi rádio galgá juohke beaivve interneah-tas almmuhit váldoodđasiid maiddá anárašgillii. Anárašgielagat háliidit maiddá oažžut buotlágan čáppagirjjálašvuoda teavsttaid anárašgillii. (Olthuis ja Gerstenberger, boah-timin).

Anarâškielâ servi lea iskamin ruhtadit prošeavtta *Čyeti čälled anarâškielân* man ulbmil lea oažžut čuođi odđa anárašgiella čáli. Ovdaprošeavtta jearahallamis bohtá ovdan ahte odđa čállit dárbbasit sátnedárkkistanprográmma ja neahttasátnegirjji, muhto maiddá sojahanparadigmaid veahkin go eatnasat leat nuppegielagat (Olthuis ja Gerstenberger, boah-timin). Giellateknologiija lea vástádus sin dárbbuide.

Bargu hukset anárašgiella FST álggahuvvui albma láhkai jagis 2014. Giella-smn lea dál dan muttus ahte dat dovdá 95,2 % teakstakorpusa 1,3 miljovna sániin (tabeal-la 2.4). Giella-smX analysáhtoriiguin lea vejolaš hukset iešgudetlágan geavaheddjiid-prográmmaid (govus 4.1), ja Giella-smn lea juo laktojuvvon dihtorjorgalanprográmmii mas lea davvisámegiella gáldogiellan⁵³ (Antonsen ja earát, 2017). Dán prošeavtta ruhta-

⁵²Complementary Aanaar Saami Language Education for adults

⁵³<http://gtweb.uit.no/mt-testing/> (14.02.2018)

deapmi⁵⁴ dagaige vejolažžan hukset Giella-smn. Analysáhtor geavahuvvo maiddá Divvun-programma betaveršuvnnas⁵⁵ ja NDS (Nettidigisäänih) sátnegirjjiis⁵⁶. Anárašgiela NDS geavahanlogut čájehit ahte sátnegirji duodai lea lihkostuvvan giellateknologalaš veahkkeneavvun čálliide. Jus juohká áigodaga 01.01–24.02.2018 geavahanlogu anárašgielagiid loguin, de lea juohke anárašgielat geavahan NDS 116 geardde, mii lea hui stuorra lohku davvisámegiela NDS geavahusa ektui. Jus dan rehkenastá seamma láhkai, de juohke davvisámegielat geavahii NDS 11 geardde seamma áigodagas⁵⁷. Dát speadjalastá dili ahte ollu stuorit oassi anárašgielagiin go davvisámegielagiin leat nuppegielagat, muhto dat maiddá nanne gova das ahte anárašgielagiin lea garra dáhttu virkkosmahttit gielaset, ja ahte sii ávkkástallet giellateknologiiijas dán barggus.

4.2 Gielat Sámi olggobealde

Resursavátnivuohta lea dábálaš dilli maiddá eará eamiálbmot- ja vehádatgielain, nugo sámegielaid lagaš fuolkegielas kvenagielas, mas leat 2000–8000 hálli. Maiddá kvenagillii lea Giella-vuoddostruktuurras huksejuvvon FST, Giella-fkv, mii dál geavahuvvo sihke sátnedárkkistanprogrammas ja neahttasátnegirjjiis⁵⁸. (Trosterud ja earát, 2017).

FST huksen lea vástádus gielaide maiddá Sámi lagašguovllu olggobealde. Davvimáilmmi eamiálbmotgielain lea hui rikkes morfologiiija ja teakstakorpus lea menddo unni jus dihtor galgašii oahppat sojahanminstariid. Ruonáeatnangielas (kalaallisut) leat 50 000 hálli ja ruonáeatnangillii leat almmuhuvvon ollu teavsttat, go giella geavahuvvo sihke váldohálddahas, beaivválaš aviissain ja vuoddoskuvllain. Muhto ruonáeatnangiella lea polysyntehtalaš giella, mas leat unnán sátnemáddagat ja ollu suorggideamit, ja dát dahká sátnetiippaid ja sániid gorálašvuoda erenoamáš unnin. Vuosttaš sátnedárkkistanprogramma ráhkaduvvui sátnelisttus mas ledje 350 000 sáni, ja dattetge dat govčcai dušše 25 % sániin oktilaš aviisateavsttas. Lei čielggas ahte fertii iskat hukset sátnedárkkistanpro-

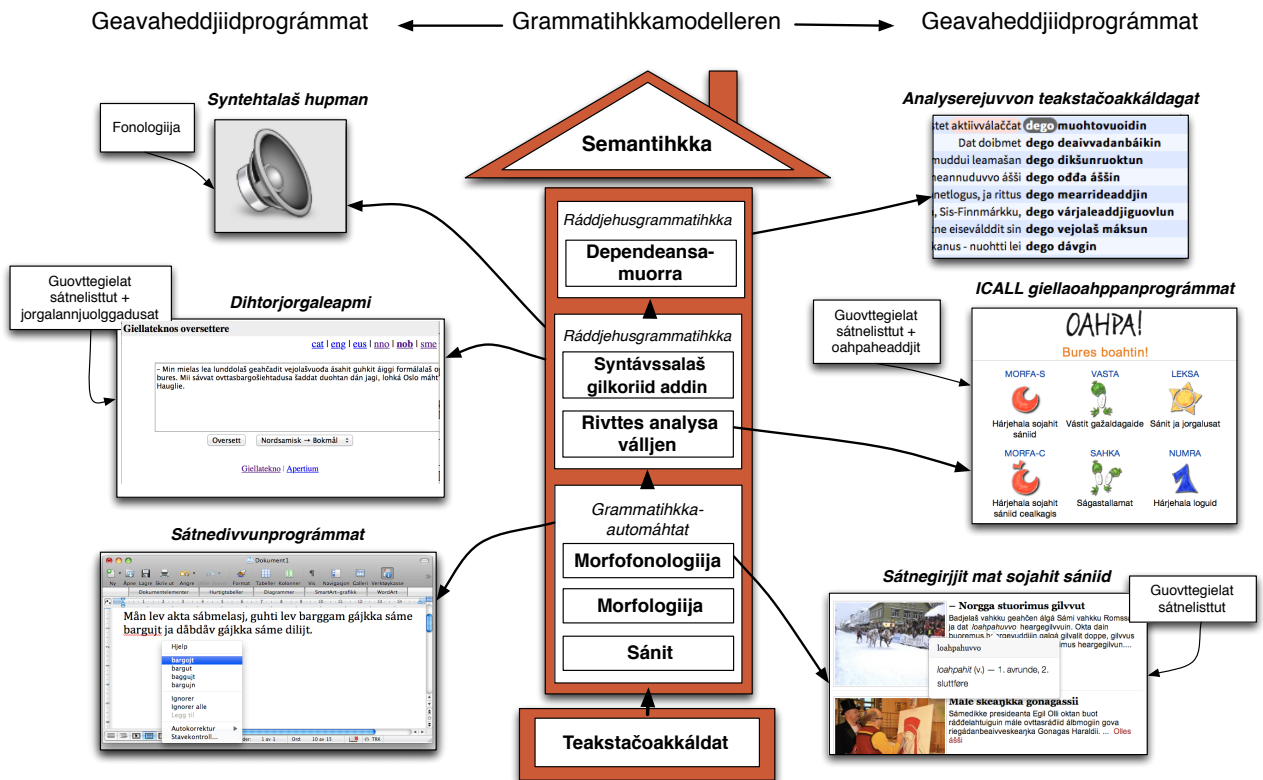
⁵⁴Norgga dutkanrádi juolludeapmi nr. 234299 ja Kone Foundation

⁵⁵<http://divvun.org/proofing/proofing.html>

⁵⁶<http://saanih.oahpa.no> (14.02.2018)

⁵⁷<https://www.ethnologue.com/language/smn> (24.02.2018) mielde anárašgielagiid lohku lea 300, muhto čujuhuvvo boares gáldui (2001 olmmošlohkamii). Giellageavaheddjiid lohku lassána johtilit, ja dan dihte lean válljen geavahit ođđasat logu: 450 anárašgielaga (Olthuis ja earát, 2016).

⁵⁸Giellateknologalaš reaidut leat fálus <http://giellatekno.uit.no/cgi/index.fkv.sme.html> (14.02.2018)



Govus 4.1: FST lea vuodđun ollu geavaheddjiidprogrammaide. Giella-smX joavkkus leat dušše davvisámegielas buot dát reaidnut. Lullisámegiela analysáhtor geavahuvvo buot reaidduide earret syntehtalaš hupmamii. Julevsáme-, anáraš- ja nuortalašgielaid analysáhtorat geavahuvvojit sátnedárkkistanprográmmii, dihortjorgaleapmái ja korpus-analysii. Anáraš- ja nuortalašgielaid analysáhtorat geavahuvvojit maidái sátnegirijis.

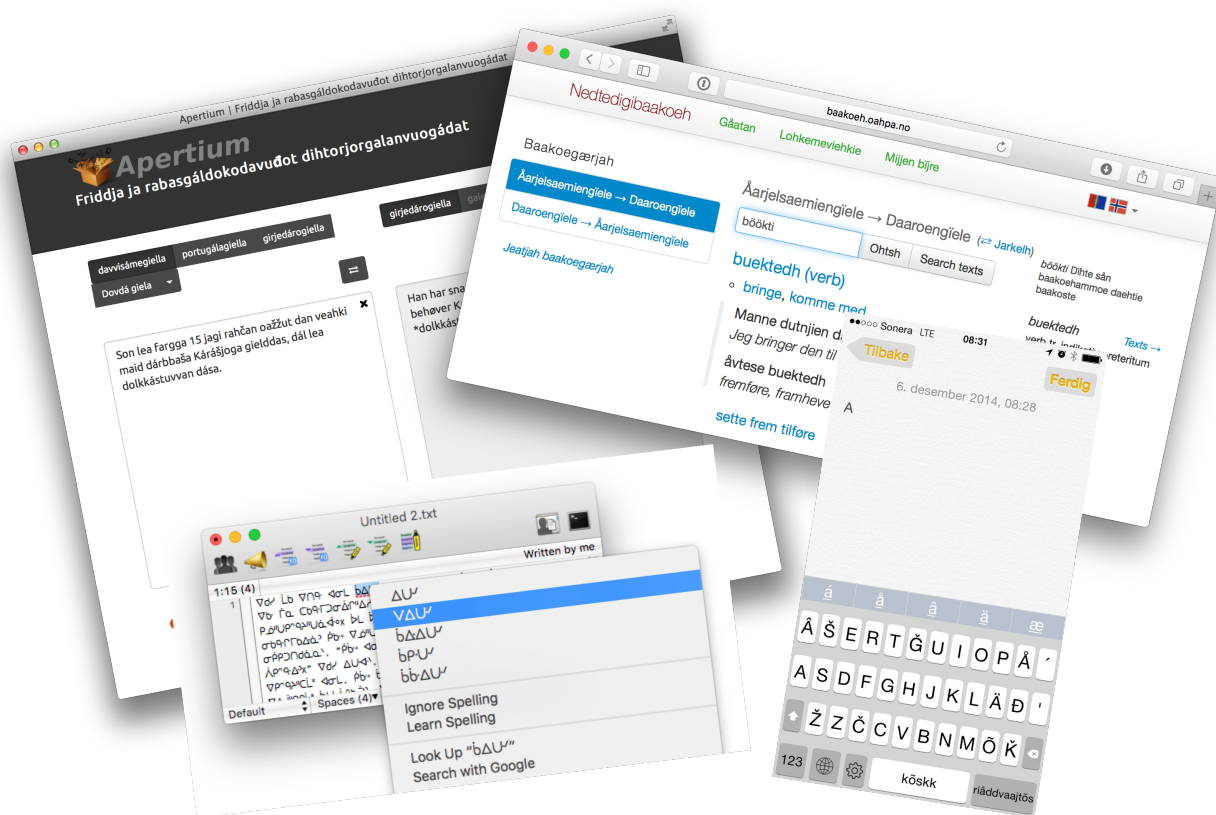
grámma eará lahkonanvugiin, ja jagi 2006 válbmanii ruonáeatnangiela sátnedárkkistanprográmma *Kukkuníiaat* mii lea genererejuvvo FST:in ja huksejuvvo Giellatekno- ja Divvun-joavkkuid dalá vuoddostrukturras (geahča detáljaid Heilmann ja earát (2009)). Juo jagi 2007 *Kukkuníiaat* govčcai 90 % sániin oktilaš teavsttas (Oqaaserpassualeriffik, dovdameahtun jahki), ja dan mañnel leat prográmmas bohtán odđa buoriduvvo veršuvnat.

Eanaš Davvi-Amerihká eamiálbmotgielain leat ollu unnit hállit go ruonáeatnangiela. Kanáda eamiálbmotgiella Plains Cree (nêhiyawêwin), gullá algonquian giellabearrašii, ja ethnologue.com mielde das leat 34 000 hálli⁵⁹, muhto rivttes lohku orru leamen ollu unnit (Harrigan ja earát, 2017). Maidái Plains Cree lea polysyntehtalaš giella, muhto ruonáeatnangiela ektui, de gielas leat ollu unnit resurssat. Giela teakstakorpusis leat vuollil

⁵⁹<https://www.ethnologue.com/language/crk> (29.12.2017)

miljovna sáni, ja teavsttain lea váilevaš standardiseren. Golbma čállinvuogi latiinnalaš bustávaiguin leat geavahas (guhkes vokálat merkejuvvojit makrovnnain dahje sirkumfleavssain, dahje guhkes vokálat eai merkejuvvo), ja daidda čállinvugiide lassin čállojuvvo Plains Cree maiddái stávvalalfabehtain. (Arppe ja earát, 2016).

Kanáda eanetlogugielaid, engelasgiela ja fránskkagiela, giellateknologiija ii doaimma guovllu eamiálbmotgielaide, ja dán rádjai ii leat gávdnon sátnedárkkistanprográmma ovttaga Davvi-Amerihká eamiálbmotgillii. Vástádus lei hukset FST, ja ovttas University of Albertain huksiimet Plains Cree FST (Snoek ja earát, 2014; Harrigan ja earát, 2017). Giellaservodat lea dien láhkai ožžon sátnedárkkistanprográmma sihke latiinnalaš čállinvugiiguin ja stávvalalfabehtain, nugo govvosis 4.2. (Prográmma lea velge betadásis, iige leat almmuhuvvon viidát.) Dasa lassin leat huksejuvvon ICALL-prográmma betaveršuvdna *nêhiyawêtan*, ja neahttasátnegirji *itwewina* NDS vuoddostruktuurras (Arppe ja earát, 2016; Bontogon ja earát, boahtimin).



Govus 4.2: Giellateknologalaš reaidut eamiálbmotgielaide. Gurut ravddas lea njuolggadusvuodot jorgalanprográmma davvisámegielas dárogillii, Plains Cree sátnedárkkistanprográmma stávvalalfabehtain, lullisámegiela sátnegirji mii dovdá morfologiija (NDS) ja nuortalašgiela boallobeavdi mobiilalefuvna várás.

Nubbi ovdamearka Kanádas lea davvihaidagiella, Xaad Kíl. Dan fuolkegiella lea lullihaidagiella, ja dát guokte giela dahket ovttas makrogiela haidagiella mii lea isoláhta. Davvihaida čielga hálliid lohku leat vuollil 12, muhto 4500 olbmo gullet čerdii, ja daid gaskkas lea stuora beroštupmi giela oahppat. Davvihaida dihtorlogahahtti sátnegirji (Lachler, 2010) mas leat morfologalaš dieđut, fievrriduvvui LEXC-formáhtii ja oanehis áig-gis huksiimet FST. Vaikko gielas leat nu unnán hállit, de Giella-vuodđostruktuvrra bokte lea davvihaidagielas sátnedárkkistanprográmma olámuttos, ja neahttasátnegirji *Gúusaaw*, mas lea morfologiija ja ráhkaduvvon NDS-formáhtas, lea juo almmuhuvvon interneahtas⁶⁰. (Lachler ja earát, boahtimin).

⁶⁰Giellateknologalaš reaiddut leat fáalus <http://giellatekno.uit.no/cgi/index.hdn.sme.html> (14.02.2018)

Kapihtal 5

Konklusuvdna

Dán nákkosgirjji váldoteavsttas ja gávcci almmuhuvvon artihkkalis lean suokkardallan mo sáhtta hukset giellamodealla ja heivehit dan duohta giellamáilbmái. Nuppi kapihtalis ákkastallen manne giellamodealla huksen FST hámis lea buorre vuohki ráhkadit analysáhtoriid sámegeielaide. FST:in sáhtta modelleret morfofonologalaš rievddademiid ja analyseret teakstakorpusiid vaikko teavsttain lea ollu variašuvdna, muhto maiddá genereret buot vejolaš sátnehámiid vaikko eai gávdno korpusis. Syntávssalaš analysáhtor mii lea huksejuvvon Constraint Grammariin, ii gáibit stuorra teakstačoakkáldaga, ja lea gierdilis vuohki disambigueret morfologalaš analysaid gaskkas, dalle go lea guovtte- dahje mánggačilggolašvuoha. Goalmmát kapihttala fádda lea mo geavahit FST geavaheadjiidprográmmain, ja manne lea dehálaš heivehit FST ja syntávssalaš analysáhtora vai buorebut speadjalastet duohta giela. Dát mielddisbuktá sihke badjelmearálaš genererema garvima ja giellaoahpahalli gaskagiela siskkildit analysáhtorii. Mus eai leat buot vástádušat, muhto lean suokkardallan muhtun vugiid ja daid čuovvumušaid.

Njealját kapihttalis lean čájehan mo FST huksen oktan Giella-vuoddostruktuvrrain vástida mángga davvimáilmmi eamiálbmotgiela dárbbuide oažžut giellateknologalaš reaiduid. Giella-vuoddostruktuvra maid sáhtta viežžat internehtas, lea dahkan álkibun hukset FST, ja maiddá geavahit FST njuolga sátnegirjjis ja sátnedárkkistanprográmmas. Buot vuoddostruktuvrra buorádušat mat leat dahkkon sámegeielaide várás, bohtet maiddá eará eamiálbmot- ja vehádatgielaide buorin. Dát vuoddostruktuvra ja FST lahkonan-

vuohki dahket alladásat geavaheddjiidprográmmaid olámuddui maiddá i eará vehádatgelaide main lea rikkes morfologiija, vátna teakstakorpus ja unnán hállit.

Girjjálašvuohta

- Antonsen, Lene, Ciprian Gerstenberger, Maja Kappfjell, Sandra Nystø Rahka, Marja-Liisa Olthuis, Trond Trosterud ja Francis M. Tyers. 2017. Machine translation with North Saami as a pivot language. Girjjiis: *Proceedings of the 21st Nordic Conference of Computational Linguistics (NoDaLiDa 2017)*, Čanas 29 *NEALT Proceedings Series*, siiddut 123–131. Linköping, Sweden: Linköping University Electronic Press. <<http://www.ep.liu.se/ecp/131/015/ecp17131015.pdf>> (14.02.2018).
- Antonsen, Lene ja Laura Alexis Janda. 2015. Oamastanráhkadusat davvisámi girjjálašvuođas. [English summary: Possessive constructions in North Saami prose]. *Diedut* 2/2015:9–43. <http://giellatekno.uit.no/background/Diedut202015-2_artihkkalcoakkaldat_antonsen_janda.pdf> (14.02.2018).
- Antonsen, Lene, Laura Alexis Janda ja Berit Anne Bals Baal. 2012. Njealji davvisámi adposišuvnna geavahus. [English summary: The use of four North Sami adpositions]. *Sámi diedalaš áigečála* 2/2012:7–38. <<http://site.uit.no/aigecala/sda-2-2012-antonsen-ja-janda-ja-baal/>> (14.02.2018).
- Antonsen, Lene ja Trond Trosterud. 2010. Manne dihtor galgá máhttit grammatihka? [English summary: Why the computer should know its Sami grammar]. *Sámi diedalaš áigečála* 1/2010:3–28. <<http://site.uit.no/aigecala/sda-1-2010-antonsen-ja-trosterud/>> (14.02.2018).
- Antonsen, Lene, Trond Trosterud ja Linda Wiechetek. 2010. Reusing grammatical resources for new languages. Girjjiis: *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, siiddut 2782–2789. The Association for Computational Linguistics, Stroudsburg: ELRA. <http://www.lrec-conf.org/proceedings/lrec2010/pdf/254/_Paper.pdf> (14.02.2018).
- Arppe, Antti, Jordan Lachler, Trond Trosterud, Lene Antonsen ja Sjur N. Moshagen. 2016. Basic Language Resource Kits for Endangered Languages: A Case Study of Plains Cree. Girjjiis: C. Soria, L. Pretorius, T. Declerck, J. Mariani, K. Scannell, ja E. Wandl-Vogt, doaimm., *Proceedings of Workshop Collaboration and Computing for Under-Resourced Languages – Towards an Alliance for Digital Language Diversity (CCURL 2016)*, siiddut 1–8. <http://altlab.artsrn.ualberta.ca/wp-content/uploads/2017/05/LREC_CCURL_Arppe_et_al_2016C.pdf> (14.02.2018).
- Beesley, Kenneth R. ja Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Studies in Computational Linguistics. Stanford, USA: CSLI publications.
- Bontogon, Megan, Antti Arppe, Lene Antonsen, Dorothy Thunder ja Jordan Lachler. boahtimin. Intelligent Computer Assisted Language Learning (ICALL) for nêhiyawêwin: An In-Depth User Experience Evaluation. *Canadian Modern Languages Review* .

- Bybee, Joan. 1985. *Morphology*. Amsterdam – Philadelphia: John Benjamins.
- Damerau, Frederick J. 1964. A technique for computer detection and correction of spelling errors. *Communications of the ACM* 7(3):171–176. <<https://dl.acm.org/citation.cfm?id=363994>> (14.02.2018).
- Feist, Timothy. 2010. *A grammar of Skolt Saami*. PhD-čáalus, University of Manchester. <<https://www.escholar.manchester.ac.uk/uk-ac-man-scw:123128>> (14.02.2018).
- Harrigan, Atticus G., Katherine Schmirler, Antti Arppe, Lene Antonsen, Trond Trosterud ja Arok Wolvengrey. 2017. Learning from the computational modelling of Plains Cree verbs. *Morphology* 27:565–598. <<https://www.escholar.manchester.ac.uk/uk-ac-man-scw:123128>> (14.02.2018).
- Heilmann, Beatrine, Per Langgård, Aviaq Tobiassen ja Trond Trosterud. 2009. Language Technology for Inuit Languages: The Nuuk-Tromsø Connection. Poster. Workshop: Arctic Languages: Syntax, Morphology, Lexicon. 24–26.9.2009 UiT, Tromsø. <http://giellatekno.uit.no/background/inuit.pdf>.
- Helander, Nils Øivind. 2016. *Ohppojuvvon ja sohppojuvvon giella. Gielladiidolašvuohta, čálamáhttu ja guovttegielatvuohta*. Dieđut 1/2016. Guovdageaidnu: Sámi Allaskuvla.
- Hulden, Mans. 2009. Foma: a finite-state compiler and library. Girjjis: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009) Demonstrations Session*, siiddut 29–32. <<http://wing.comp.nus.edu.sg/~antho/E/E09/E09-2008.pdf>>.
- Itkonen, Erkki, Raija Bartens ja Lea Laitinen. 1988. *Inarilappisches Wörterbuch, Čanas 1–4 Lexica Societatis Fenno-Ugricae*. Helsinki: Societé Finno-Ougrienne.
- Janda, Laura Alexis ja Tore Nettet. 2010. Paradigm structure: Evidence from Russian suffix shift. *Cognitive Linguistics* 21(4):699–725. <<https://www.degruyter.com/view/j/cogl.2010.21.issue-4/issue-files/cogl.2010.21.-issue-4.xml>> (14.02.2018).
- Jernsletten, Nils. 1988. *Álgošátnegirji. Samisk-norsk ordbok*. Oslo: Universitetsforlaget.
- Jomppanen, Marjatta. 2015. Davvisámegiela reduplikatiiva vearbagenetiiva boadi boadi ja bosu bosu – morfologijja, syntákssa ja semantihkka. [English summary: Reduplication of the non-finite form called the verb genitive in North Saami.]. *Sámi dieđalaš áigečála* 2/2015:33–59. <<http://site.uit.no/aigecala/sda-2-2015-jomppanen/>> (14.02.2018).
- Jurafsky, Daniel ja James H. Martin. 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Second Edition*. New Jersey: Prentice Hall.
- Karlsson, Fred. 1990. Constraint grammar as a framework for parsing running text. Girjjis: H. Karlgren, doaimm., *Papers Presented to the 13th International Conference on Computational Linguistics (COLING-90) on the Occasion of the 25th Anniversary of COLING and the 350th Anniversary of Helsinki University*, Čanas 3. Helsinki: Yliopistopaino. <<https://dl.acm.org/citation.cfm?id=991176>> (14.02.2018).

- Karlsson, Fred, Atro Voutilainen, Juha Heikkilä ja Arto Anttila. 1995. *Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text*. Berlin – New York: Mouton de Gruyter.
- Karttunen, Lauri. 1994. Constructing lexical transducers. Girjii: *Proceedings of the 15th International Conference on Computational Linguistics (COLING-94)*, siiddut 406–411. Kyoto. <<http://www.aclweb.org/anthology/C94-1066>> (14.02.2018).
- Karttunen, Lauri. 2001. A short history of two-level morphology. A paper presented at ESSLLI 2001. <<http://www.ling.helsinki.fi/~koskenni/essli-2001-karttunen/>> (14.02.2018).
- Koskenniemi, Kimmo. 1983. *Two-level Morphology. A General Computational Model for Word-forms Production and Generation*, Čanas 11 *Publications of the Department of General Linguistics*. Helsinki: University of Helsinki.
- Lachler, Jordan. 2010. *Dictionary of Alaskan Haida*. Juneau, Alaska: Sealaska Heritage Institute. <http://www.sealaskaheritage.org/sites/default/files/Haida_dictionary_web.pdf> (14.02.2018).
- Lachler, Jordan, Lene Antonsen, Trond Trosterud, Sjur N. Moshagen ja Antti Arppe. boahhtimin. Modeling Northern Haida Verb Morphology. Girjii: *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*.
- Levenshtein, V. I. 1965. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 10(8), 707–710, *jorgaluvvon publikašuvnnas Doklady Akademii Nauk SSSR* 163(4):845–848. <http://profs.sci.univr.it/~liptak/ALBioinfo/2011_2012/files/levenshtein66.pdf> (14.02.2018).
- Lindén, K., E. Axelson, S. Hardwick, T.A. Pirinen ja M. Silfverberg. 2011. HFST – framework for compiling and applying morphologies. Girjii: C. Mahlow ja M. Piotrowski, doaimm., *Systems and Frameworks for Computational Morphology (SFCM 2011)*, Čanas 100 *Communications in Computer and Information Science*. Heidelberg: Springer. <https://link.springer.com/chapter/10.1007/978-3-642-23138-4/_5> (14.02.2018).
- Morottaja, Matti. 2007. *Anarâškielâ raavuuh*. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Morottaja, Petter. 2016. *Anarâškielâ vokalij kvalitetnubástusâi myensteristem Kielâtekno morfologisii analysaattorist*. Masterčáalus, Giellagas-institut, Oulu ollâopâttâh, Oulu. <<http://jultika.oulu.fi/Record/nbnfioulu-201611103003>> (14.02.2018).
- Moshagen, Sjur, Pekka Sammallahti ja Trond Trosterud. 2005. Twol at work. Girjii: A. Arppe, L. Carlson, K. Lindén, J. Piitulainen, M. Suominen, M. Vainio, H. Westerlund, ja A. Yli-Jyrä, doaimm., *Inquiries into Words, Constraints and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday*, siiddut 94–105. Stanford, CA: CSLI. <<https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/10-moshagen-sammallahti-trosterud.pdf>> (14.02.2018).

- Moshagen, Sjur Nørstebø, Tommi A. Pirinen ja Trond Trosterud. 2013. Building an open-source development infrastructure for language technology projects. Girjijis: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa 2013)*, Čanas 16 *NEALT Proceedings Series*, siiddut 343–352. <<http://www.ep.liu.se/ecp/article.asp?issue=085&article=031>> (14.02.2018).
- Nickel, Klaus Peter ja Pekka Sammallahti. 2011. *Nordsamisk grammatikk*. Karasjok: Davvi Girji, 2. almmuh.
- Nielsen, Konrad. 1979 [1926–29]. *Lærebok i lappisk (samisk). 1. grammatikk. Lydlære, formllære, orddannelseslære og syntaks samt tillegg*. Oslo: Universitetsforlaget.
- Olthuis, Marja-Liisa. 2000. *Inarinsaamen kielen kielioppi*. Inari: Sämitigge.
- Olthuis, Marja-Liisa ja Ciprian-Virgil Gerstenberger. 2013. *Strengthening indigenous languages by means of language technology. The case of Aanaar Saami in Finland*. Girjijis: T. McCarty, S. Nicholas, ja G. Wigglesworth, doaimm., *World of Indigenous Languages: Politics, pedagogies and prospects for language revitalization and maintenance*. United Kingdom: Multilingual Matters Ltd.
- Olthuis, Marja-Liisa, Suvi Kivelä ja Tove Skutnabb-Kangas. 2013. *Revitalising Indigenous Languages. How to Recreate a Lost Generation*. United Kingdom: Multilingual Matters Ltd.
- Olthuis, Marja-Liisa, Erika Sarivaara, Ciprian Gerstenberger ja Trond Trosterud. 2016. Čyeti čälled anaráškielân ‘100 writers for Aanaar Saami’. Kirjälistemohjelm huksim uhkevuálásii kielâ várás. Sámegiela ja sámi girjjálašvuoda dutkan- ja bagadansymposia 2.–3.2.2016 Romssas Abstrávttat. <<http://site.uit.no/samesymposia/files/2015/10/Samesymposiaabstravttat160102.pdf>> (14.02.2018).
- Oqaaserpassualeriffik. dovdameahttun jahki. A bit of history. <<http://tech.oqaasileriffik.gl/a-bit-of-history/>> (03.01.2018).
- Outakoski, Hanna. 2015. *Multilingual literacy among young learners of North Sámi: Contexts, complexity and writing in Sápmi*. PhD-čáalus, Department of Language Studies, Umeå University, Umeå.
- Pirinen, Tommi A. 2015. Development and Use of Computational Morphology of Finnish in the Open Source and Open Science Era: Notes on Experiences with Omorfi Development. *SKY Journal of Linguistics* 28:381–393. <http://www.linguistics.fi/julkaisut/SKY2015/SKYJoL28_Pirinen.pdf> (14.02.2018).
- Rehm, Georg ja Hans Uszkoreit. 2012. Language technology support for Finnish. Girjijis: G. Rehm ja H. Uszkoreit, doaimm., *The Finnish Language in the Digital Age Language*, White Paper Series. Berlin – Heidelberg: Springer. <<https://link.springer.com/book/10.1007%2F978-3-642-27248-6>> (14.02.2018).
- Sammallahti, Pekka ja J. Mosnikoff. 1991. *Suomi-koltansaame sanakirja. Lää'dd-sää'm sää'nnke'rrj*. Ohcejohka: Girjegiisa.

- Sjaggo, Ann-Charlotte ja Trond Trosterud. 2015. Om pitesamiskt språk. Girjji: B. Evjen ja M. Myrvoll, doaimm., *Från kust til kyst = Áhpegáttest áhpegáddáj*, siiddut 223–231. Tromsø: Orkana Forlag.
- Snoek, Conor, Dorothy Thunder, Kaidi Lõo, Antti Arppe, Jordan Lachler, Sjur Moshagen ja Trond Trosterud. 2014. Modeling the noun morphology of Plains Cree. Girjji: *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL 2014)*, siiddut 34–42. <<http://acl2014.org/acl2014/W14-22/pdf/W14-2205.pdf>> (14.02.2018).
- Svonni, Mikael. 2015. *Davvisámegiella – sánit ja cealkagat. Láidehus sámi lingvistihkkii*. Giron: Ravda lágádus.
- Tapanainen, Pasi. 1996. The constraint grammar parser CG-2. Girjji: *Corpora and Cross-linguistic Research. Theory, Method, and Case Studies*, Čanas 27 *Publications of the Department of General Linguistics*. Helsinki: University of Helsinki.
- Trosterud, Sindre Reino, Trond Trosterud, Anna-Kaisa Räisänen, Leena Niiranen, Mervi Haavisto ja Kaisa Maliniemi. 2017. A morphological analyser for kven. Girjji: *Proceedings for the 3rd Workshop on Computational Linguistics for Uralic Languages (IWCLUL 2017)*, siiddut 76–88. <<http://aclweb.org/anthology/W17-0608>> (14.02.2018).
- Trosterud, Trond. 2012. A restricted freedom of choice: Linguistic diversity in the digital landscape. *Nordlyd* 39(2):89–104. <<http://septentrio.uit.no/index.php/nordlyd/article/view/2474>> (14.02.2018).
- Trosterud, Trond ja Heli Uiho. 2005. Consonant gradation in Estonian and Sami: Two-level solution. Girjji: A. Arppe, L. Carlson, K. Lindén, J. Piitulainen, M. Suominen, M. Vainio, H. Westerlund, ja A. Yli-Jyrä, doaimm., *Inquiries into Words, Constraints and Contexts. Festschrift for Kimmo Koskenniemi on his 60th Birthday*, siiddut 136–150. CSLI Studies in Computational Linguistics online. <<https://web.stanford.edu/group/cslipublications/cslipublications/koskenniemi-festschrift/14-trosterud-uiho.pdf>> (14.02.2018).
- Trosterud, Trond ja Linda Wiechetek. 2007. Disambiguering av homonymi i nord- og lulesamisk. Girjji: J. Ylikoski ja A. Aikio, doaimm., *Sámit, sánit, sátnehámit. Riepmočála Pekka Sammallahtii miessemánu 21. beaivve 2007*, Čanas 253 *Suomalais-Ugrilaisen Seuran Toimituksia = Mémoires de la Société Finno-Ougrienne*, siiddut 375–395. Helsinki: Suomalais-Ugrilainen Seura. <http://www.sgr.fi/sust/sust253/sust253_trosterudjawiechetek.pdf> (14.02.2018).
- VISL-group. 2008. Constraint grammar. <<http://beta.visl.sdu.dk/constraintgrammar.html>> (14.02.2018).
- Wiechetek, Linda. 2012. Constraint Grammar based Correction of Grammatical Errors for North Sámi. Girjji: G. D. Pauw, G.-M. Schryver, M. Forcada, K. Sarasola, F. Tyers, ja P. Wagacha, doaimm., *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages (SaLTMil 8 - AfLaT2012)*, siiddut 35–40. European Language Resources Association (ELRA), Istanbul, Turkey. <<http://tshwanedje.com/publications/SaLTMil8-AfLaT2012.pdf>> (14.02.2018).

- Wilbur, Joshua. 2016. *Pitesamisk ordbok samt stavningsregler*, Čanas 2 *Samica*. Freiburg: Albert-Ludwigs-Universität Freiburg.
- Wilbur, Joshua. 2018. Extracting inflectional class assignment in Pite Saami. Girjjis: T. A. Pirinen, M. Rießler, T. Trosterud, ja F. M. Tyers, doaimm., *Proceedings of the 4th International Workshop on Computational Linguistics for Uralic languages (IWCLUL 2018)*, ACL Anthology, siiddut 156–170. <<http://aclweb.org/anthology/W18-02>> (14.02.2018).
- Ylikoski, Jussi. 2002. Zu den adverbialen Nominalkonstruktionen im Nordlappischen. *Finnisch-Ugrische Forschungen* 57:68–166. <<http://cc.oulu.fi/~jylikosk/filer/zudenadv1.pdf>> (14.02.2018).
- Ylikoski, Jussi. 2009. *Non-finites in North Saami*, Čanas 257 *Suomalais-Ugrilaisen Seuran Toimituksia*. – *Mémoires de la Société Finno-Ougrienne*. Helsinki: Suomalais-Ugrilainen Seura. <<http://www.sgr.fi/sust/sust257/sust257.pdf>> (14.02.2018).

