Sámegielaid modelleren – huksen ja heiveheapmi duohta giellamáilbmái

Lene Antonsen
Nákkosgirji buktojuvvon Philosophiae Doctor grádii
Guovvamánu 2018
Sámegeielaid modelleren
– huksen ja heiveheapmi duohta
giellamáilbmái

Lene Antonsen

Nákkosgirji geigejuvvon čuovvovaš gráda ovdii:
Philosophiae Doctor (PhD)

Humaniora, servodatdiehtaga ja oahpaeaddjioahpu fakultehta
UiT Norgga árktalaš universitehta
njukčamánu 1. b. 2018
3.2.2 Sátnegirjiid riggodahttit morfologijain (Art. IV ja V) 46
3.3 Giellamodealla gáržźideapmi 54
3.3.1 Eai buot sánit leat geavahusas 55
3.3.2 Giellamodealla gáržźideapmi norpma ektui 57
3.3.3 Badjelmearálaš genererema ráddjen 58
3.3.4 Giellamodealla gáržźideami čoahkkáigeassu 71
3.4 Giellamodealla viiddideapmi 72
3.4.1 Grammatihkkahárjehusaid genereren (Art. VI) 73
3.4.2 Oahpahalli morfologalaš gaskagiela giedahallan (Art. VII) 76
3.4.3 Oahpahalli syntávssalaš gaskagiela giedahallan (Art. VIII) 78
3.4.4 Giellamodealla viiddideami čoahkkáigeassu 80
3.5 Giellamodealla, dan geavahus ja duohta
  giellamáilbmí 81
4 FST addá vejolašvuodaid eamiálbmotgielaide 83
  4.1 Anárašgiella ovdamearkan 83
  4.2 Gielat Sámi oiggobealde 85
5 Konklušuvdna 89
Ovdasánit


Lean ovttasbargan lingvistajoavkkuin University of Albertas, ja lea leamaš hirmat miellagidevaš oahpásmuvvat Kanáda eamiálbmotgielalade, ja Antti Arppe ja Jordan Lachler leaba guovddášolbmo dán ovttasbarggus.

Erenoamážit háliidan giitit Ciprian Gerstenbergera, gii rähkada skriptaid maid dárbbásan ja veahkehan oaažit korpusteavstaid olamuddui, ja Chiara Argesa, gii dá lea joatkimin giellaoahppanreaidduid programmerenbarggu maid Sara Huhmarniemi, Heli Uibo ja Ryan Johnson ovdal leat bargan ja de maiddái Marja-Liisa Olthuisa, gii lea rahpan anarášgiela giellamáilmmi munne.

Giitu maiddái Linda Wiecheteekii gii lea veahkehan mu LaTex-prográmmmain mainna lean nákkosgirjii čállán. Eanemus giitu Trond Trosterudii, gii lea leamaš mu bagadalli. Vaikko sus lea ollu bargu ja jámma lea jodus, de in leat goassege ballan das ahnte sus i livčče dilli lohkat ja kommenteret mu barggu.


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

Romsa njukčamánu 1.b. 2018
Artihkallistu


Čoahkkáigeassu


Čájehan mo sáhttit ávkkástallat sámegielaid grámmatihkalaš giellamodeallain geavaheddjiidprogrammain. Grammatihkalaš gilkorat dahket vejolažžan addit metalingvist-talaš dieduidd ja máhcåhaga geavaheddjiide. FST ferte heivehit nu ahté dat buorebut speadjalastá duohta giela. FST lahkonanvuo gigjumla lea badjelmearálaš sátnehámiiid genereren, ja dan dihte lea dehálaš gáržžidit giellamodealla. Go ráhkkada giellateknologalaš reaidduid giellaahpahallami, de lea ávkkálaš dasa siskkildit giellaahpahalli gaskagiela. Suokkardalan muhtun vugiid mo sáhttá gáržžidit ja viiddidit giellamodealla vai buorebut sulastahttá duohta giellamáilmmi.

Čájehan ahte FST huksen vástida mángga davvimáilmmi eamiálbomotgielaš dárbbuide oazžut giellateknologalaš reaidduid. FST lahkonanvuohtg oktan buiriin vuoddostruktuurain mainna ávkkáställá vuoddobarggus mii lea dahkkon sámejgelaide, dahkká alladásat geavaheddjiidprogrammaid olámuddui vehádatgielaide main lea rikkes morfologiija, vátña teakstakorporpu 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

<table>
<thead>
<tr>
<th>Numero</th>
<th>Laite</th>
<th>Sivu</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Automáhta (FSA)</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>Grammatiikalaš transduser (FST)</td>
<td>13</td>
</tr>
<tr>
<td>2.3</td>
<td>Substantiivaleksikon LEXC-fiillas.</td>
<td>13</td>
</tr>
<tr>
<td>2.4</td>
<td>Joatkaleksikon LEXC-fiillas.</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>Leksikontransduser ja morfologalaš transduser.</td>
<td>16</td>
</tr>
<tr>
<td>2.6</td>
<td>Leksikontransdusera ja morfologalaš transdusera kompileren.</td>
<td>16</td>
</tr>
<tr>
<td>2.7</td>
<td>YAML testenfiila.</td>
<td>17</td>
</tr>
<tr>
<td>2.8</td>
<td>TWOLC njuolggadusat.</td>
<td>18</td>
</tr>
<tr>
<td>2.9</td>
<td>Suomagiela analysáhtor Giella-fin: Buot morfonologalaš rievddadeamit dálhpáhuvvet LEXC-prográmmas.</td>
<td>22</td>
</tr>
<tr>
<td>2.10</td>
<td>Kásusii juohkáseapmi SIKORis.</td>
<td>23</td>
</tr>
<tr>
<td>2.11</td>
<td>Giella-sama syntávssalaš analysáhtora njuolggadusaid beaktilvuolta.</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>Divvunprográmma evttohusat boastut čállon sániide.</td>
<td>46</td>
</tr>
<tr>
<td>3.2</td>
<td>Vuosttaš digisániid lohanveahkki.</td>
<td>47</td>
</tr>
<tr>
<td>3.3</td>
<td>Vuosttaš digisániid váldosátneartihkal.</td>
<td>47</td>
</tr>
<tr>
<td>3.4</td>
<td>Neahttagisániid lohanveahkki sáde sáni analysáhtorii.</td>
<td>51</td>
</tr>
<tr>
<td>3.5</td>
<td>Neahttagisániin beassá njuolga SIKORii.</td>
<td>51</td>
</tr>
<tr>
<td>3.6</td>
<td>Stávenložžen.</td>
<td>52</td>
</tr>
<tr>
<td>3.7</td>
<td>Neahttagisániit čilgejit infinihtta vearbba funkšuvnna.</td>
<td>53</td>
</tr>
<tr>
<td>3.8</td>
<td>FST: Dynámaša goallosteapmi.</td>
<td>56</td>
</tr>
<tr>
<td>3.9</td>
<td>FST: Bálgás lea diakrihtalaš leavga.</td>
<td>61</td>
</tr>
<tr>
<td>3.10</td>
<td>ICALL: Morfologalaš mähcat.</td>
<td>73</td>
</tr>
<tr>
<td>3.11</td>
<td>ICALL: QA-hárjehusa templáhta.</td>
<td>74</td>
</tr>
<tr>
<td>3.12</td>
<td>Divvun-prográmma ii máhte álo veahkehit giellaohpahalli.</td>
<td>76</td>
</tr>
<tr>
<td>3.13</td>
<td>L2-FST sisttisdoallá gaskagiela gehčosiid ja morfonologitija.</td>
<td>77</td>
</tr>
<tr>
<td>3.14</td>
<td>ICALL: QA-hárjehus mas lea friddja čállin.</td>
<td>79</td>
</tr>
<tr>
<td>3.15</td>
<td>ICALL: QA-hárjehusa analysa.</td>
<td>80</td>
</tr>
<tr>
<td>4.1</td>
<td>FST lea vuoddun ollu geahveddiidprográmmaide.</td>
<td>86</td>
</tr>
<tr>
<td>4.2</td>
<td>Giellateknologalaš reáiddut eamiálbmoogiid gielaide.</td>
<td>87</td>
</tr>
</tbody>
</table>
Tabeallalistu

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

Oanádusat čilgejuvvojit dadistaga teavsttas. Dás leat oanádusat mat leat geavahuvvon eanet siidduiun. Čuovvovaš siiddus leat čilgejuvvon analysagilkorat, variábelat ja báhca-hagat.

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 hullisámegiela analysáhtor
Giella-sme davvisámegiela analysáhtor
Giella-smj julevsámegiela analysáhtor
Giella-smn anárašgiela analysáhtor
Giella-sms nuortalašgiela analysáhtor
Giella-smX buot sámegielaaid analysáhtorat
ICALL eng. ‘Intelligent Computer-assisted Language Learning’ = giellaoahppan jierb-más dihtorvehkiin
L2 eng. ‘Language 2’ = giella maid olmmoš oahppá manŋel go lea oahppan vuosttaš gielas (= L1)
LEXC eng. ‘Lexicon Compiler’ = leksikona kompilerenprográmma, muhto maiddái prográmmenergiella mainna čatná oktii grammatihkalaš sáni ja sáni máddaga oktan affivssaiguin.
NDS Neahttadigisánit-sátnegirji
QA eng. ‘Question Answering’ = jearaldat oktan vástádusain
SIKOR Sámi Internationála Korpus, UiT Norgga arktalaš universitehta ja Norgga Sáme-dikki sámi teakstačoakkáldat
TWOLC eng. ‘Two Level Compiler’ = guovttedásat ovttastanprográmma. Programmerengiella mainna rievadit vuollálas hámiid rivtteshámat bajošhápmín ortográfalaš ja morphofonologalaš njuolggadusaid mielde
VD Vuosttaš digisánit-sátnegirji
eng. engelasgiella
Analysagilkorat, variábelat ja báhcahagat:

*Cns konsonánta
*Vow vokála
*CLEN konsonántaguovddáža guhkkun
*CSH konsonántaguovddáža oatnun
*RLEN vokálanguovddáža guhkkun
*RC konsonántaguovddáža arkifonema
*RV vokálanguovddáža arkifonema
*RVSH vokálanguovddáža oatnun
*QLD jeardatrádji
*WG geahnohis dássi

+A adjektiiva
+Acc akkusatiiva
+Adv advearba
+Arab arábalas lớnku
+Attr atribuhtta
+CLB cealkkarádji
+Cmp goallosteapmi
+Cmp/SgGen goallosteami mearusoassi lea ovttaidlogu genitiivvas
+Cmp/SgNom goallosteami mearusoassi lea ovttaidlogu nominatiivvas
+Cond konditionála (eaktovuohki)
+ConNeg yearbba biehttalanhápmi
+CS suhjkusuvdna
+Der/ suorggádus
+Dial/ suopman
+Du1 guvttidlogu vuosttaš persovdna
+Du2 guvttidlogu nuppi persovdna
+Du3 guvttidlogu goalmmát persovdna
+Err/Lex sótnemeattáhus
+Err/Orth cállinmeattáhus
+Ess essiva
+G3 lemmas lea dássemolsuma goalmmát gráda mii ii mekerjuvvo cállingielas
+Gen genitiiva
+Hom1, Hom2 homonymat 1 ja 2
+Hum olbmonamahus (dál geavahuvvo +Sem/Hum)
+III illatiiva
+Imprt imperatiiva
+Ind indikatiiva
+Inf infinitiiva
+Interr lahčanpronomen
+IV intransitiiva
+Loc lokatiiva
+N substantiiva
+Neg biehttalanvearba
+Nom nominatiiva
+NomAg dahkkinomen
+Num løku dahje løhkosúni
+Org organisääuvdna (dál geavahuvvo +Sem/Org)
+Par partitiiva
+Pers persovdppronomena
+Pron pronom
+Prs presansa
+Prt preterihtta
+Pl mánggaidlohku
+Pl1 mánggaidlogu vuosttaš persovdna
+Pl2 mánggaidlogu nuppi persovdna
+Pl3 mánggaidlogu goalmmát persovdna
+Po postposišuvdna
+Pr preposišuvdna
+Prop namma (propra)
+PrfPrc perfavtta partisihppa
+Prs presansa
+Prt preterihtta
+Px oamastangeažus
+Sem/Ani calli namma
+Sem/Fem nissona ovdanamma
+Sem/Hum olbmonamahus
+Sem/Plc báikenamahus
+Sem/Sur goargu/sohkanamma
+Sg ovttaidlohku
+Sg1 ovttaidlogu vuosttaš persovdna
+Sg2 ovttaidlogu nuppi persovdna
+Sg3 ovttaidlogu goalmmát persovdna
+Sup supiidna
+TV transitíiva
+V vearba
+VAbess vearbaabessiiva
+VGen vearbagenitiiva
Kapihtal 1

Láidehus

1.1 Nákkosgirjji áššečuolmmat


NDS-sátnegirji, man kapihtal 3.2.2 giedahallá, lea giellateknologalaš reaidu, ja jus juohká NDS geavahenlogu davvisámegielagiid loguin, de juohke davvisámegielat geavahii NDS olles 60 geardde 2017:s\(^1\), vaikko gávdno nubbi stuora davvisámi sátnegirji interneahitas\(^2\). Jus juohká Bokmålsordboka/Nynorskordboka (= Girjedárogiela sátnegirji/Oddadárogiela sátnegirji)\(^3\) geavahanlogu Norgga åssiid loguin, de juohke åssi geavahii sátnegirjji 8 geardde. Njealját kapihttalís cájehan ahfe unnit sámegielaíd geavahanlogut gorálaččat leat vel stuoribut go davvisámegiela geavahanlogut.

Dán nákkosgirjji giellateknologiija vuodduduvvo grammatihkkii ja lingvisttalaš njuollgadusaide, ja bajitäsi áššečuolbmma lea:

Leago grammatihkalaš giellamodealla vástádus sámegielaid ja eará davvimáilmmin eamiálbmotgielaíd giellateknologalaš dárrbuido?

In sáhte dán dútakamušas vástidit dáidda gažaldagaide dievaslaččat, muhto mun lean välljen guorahallat nyeallje hástalusa mat leat bajitäsi áššečuolmmma eaktun. Nákkosgirjji gávcči artihkkala gusket dán nyealji hástalousii, ja lean välljen erenoamázít deattuhit ja kommenteret daid osiid artihkkaliin mat leat hástalusaide relevánttat. Guokte artihkkala giedahallet giellamodealla dahje analysáhtora huksema, sikhke morfologalaš ja syntávssalaš osiid, ja dan oktavuožas 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 geavaheddjüidprográmmaide.

Prográmmaide lea dehálaš makkar sátneanlysa giellamodealla addá sániide, ja makkar


\(^3\) http://ordbok.uib.no (24.02.2018)
hámiid dat generere iešgudet lemmai, ja dán oktavuodas lea nákkosgirjjí áššečulbmma mo oaažut giellamodealla sulastahtit giela nugo dat geavahuvvo duohta giellamáilmmis. Hástalusat leat:

3. Mo gárźžidit giellamodealla amas genereret badjelmearálaš hámiid?

4. Mo viiddidit giellamodealla vai dat maiddái máhhtá analyseret giellaoahpahalli gas-kagiela?

Njealját kapihttalis máhcan bajitdási áššečulbmii, go geahčan mo dákkár vuohki huk-set giellamodealla lea ávkkálaš davvimáilmmi eamiálbmotgielaide. Ii oktage nákkosgirjji artihkkaliin gula dán kapihttalii, muhto lean leamaš miede huksemin giellamodealla ja giellateknologolaš reaiddudd anárašgiela váráš ja guovtti kanádalaš eamiálbmotgiela, Plains Cree ja davvihaidagiela, váráš, ja maiddái miede almmuheamen artihkkaliid bohtosiid birra. Kapihtal čájeha maiddái ovddeš kapihttaliiid relevánsa davvimáilmmi eará eamiálbmotgielaide.

Suokkardalansámegielaidgrammatihka giellateknologolaš perspektiivvas, muhto liikká nu, ahte lea vejolaš lohkat vaikkó ii dovdda giellateknologija, jos beare lea máhttu ja beroštupmi sámegiela grammatihkas. Jus grammatihkkavudot giellateknologija galgá doaibmat bures, de lea dárbbašlaš gulahallat giellačehpiiguin ja lingvisttaiguin geat leat min giellateknologijafaseaskára olggobealde. Dán čállosis giellateknologija leage eanet giella go teknologija.

1.2 Duogáš


Norgga Sámédiggi ásahii jagi 2004 Divvun-joavkku man ulbmil lei ráhkadit sátnedárkkistanprográmmmaid golmma sámegiela váráš. Joavkku lea álgu rájes ovttasbargan


1.3 Nákkosgirjji artihkkalat ja mu buvtta

Mun lean erenoamážit leamaš mielde huksemin davvisáme- ja anárašgiela morfologalaš analysáhtoriiid, ja maiddåi buot sámeigielaid syntávssalaš analysáhtoriiid. Sântedårkkistanprográmmma vuodduuvvá morfologalaš analysáhtoriiidda, muihó analysáhtoriiid rádd-djejupmi norpma mielde, man birra čálan kapihttalis 3.3.2, gullá Divvun-joavkkku bargui. Dättetge lean geahččaladdan ja implementeren oamastangeażushámiid ráddjema, mii lea kapihttala 3.3.3.3 fáddán, danne go das lea mearkkašúpmi maiddåi teakstakorpuse analysii. Seamma kapihttalis čálan maiddåi oanehis namaid goallosteami ráddjemis, mainna lean bargan danne go lea dehálaš sântegirjjiid analysii, mii fas lea artihkkaliid IV ja V fáddán.

4Divvun-joavkkku lingvistan leat bargan Thomas Omna (davvisámegiela), Maja Kappfjell (lullisámegiela) ja Per-Eric Kuoljok, Inga Lill Sigga Mikkelsen ja Sandra Ráhka Nystø (julevsámegiela). Giellatekno lingvisttat barget fas sîhke morfologalaš ja syntávssalaš analysesáhtoriigun, ja oktibnuot buot sámeigielaignuin: Trond Trosterud ja mun, ja ovddeš bargit Marit Julien, Ilona Rauhala ja Linda Wiechetek.
Gáveci artihkkala leat miele dan nákkosdutkamušas:


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 čá llen stuorimus oasi Constraint grammar njuolładusain, ja Trond Trosterud dagai fas stuorimus oasi evaluenerbarggus. Artihkal giedahallá syntávssalaš analysa, muhto dat vuoddoduvvá fas morfologalaš analysii, man luksenbarggus Divvun-joavkku lullisámi lingvista lea leamaš gnovddášolmomoš.


Artihkal IV (*Ei intelligent elektronisk ordbok for samisk*) vál ddahallá vuogádaga mii riggodahttá sátnegirjji analysáhtora vehkiin, ja artihkkalis leat njeallje čálli. Sjur Moshagen dat fuomášii geavahit MacOS sátnelissttu Macdict prográmmma min sátnegirjái, ja son ráhhkadii sátnegirjji prototiippa. Ciprian Gerstenberger fuolahii sátnehámiid gene rerema sátnegirjji formáhtii ja buoridii vuogádaga nu ahte juohke sátnehápmái genererere juvvo sátnegirjearthikal. Trond Trosterud mearridii gáldofiillaid leksikográfalaš struktur-

---

5Divvun-joavkku Børre Gaup lea ráhhkadan teavsttaid konverterenvuogádaga vai lea vejošas dait analy seret. Giellatekno programerár ja lingvista Ciprian Gerstenberger lea fas heivehan SIKORa geavahuvvo sámgiegiala teavsttaide.

6Teknihkalaš osiid leat bargan Sjur N. Moshagen, Børre Gaup ja Tomi Pieski.
erema, ja geahččalattai sátnegirjji eanet programmaiguin, ja son lea dasa lassin fuolahan sátnegirjji evaluerema. Mu oassi lei lingvisttalaš bargu: heivehit analysáhtora sátnegirjái, ja sátnegirjji analysáhtorii, nugo homonymijaid giedahallan. Lean maiddái hábmen ja ráddjen sáníid sojahana paradigmaid presentašuvnna gielloahpahallama geahččanguovllus, ja bargen maiddái ollu sátnegirjji gáldofiillaiguin.


Artihkal VI (Generating modular grammar exercises with finite-state transducers) muitala mo analysáhtorii sáhtńa huksut vuogádaga morfologijaoahpahallamii, ja artihkkalis leat njeallje čálli. Ryan Johnson ja Heli Uibo leaba implementeren vuogádaga teknihkalaččat. Trond Trosterud lea erenoomážit buohtastahttánu vuogádaga eará seammasullasaš vuogádagaide. Mun lean fas hábmen hárjehusgenererema ja máhca-hatvuogádaga, 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 geavhit giellamodealla gielloahppanprogrammaide, ja daid artihkkaliid lingvisttalaš ja pedagogalaš implementeren lea vuostažettiin mu bargu, muhto många fuomášumi ja čoavddusevttohusa leat bohcciidan go lean ságastallan Trond Trosterudain.
1.4 Čállosa huksehus

Kapihtal 2

Giellamodealla huksen

2.1 Sámegielaid modelleren transduseriiguin

Teaksta lea sátnečoakkádat, mii lea ordnejuvvon dainna lágiin ahte sánit ovttas dahket cealkagiid main lea sisdoallu. Juohke sátni i leat dušše bustávaráidu mas lea dihto mearkkašupmi, muhto sátni ovddasta dihto leksema grammaticalaš sáni, ja ollu sánit gullet dihto sojahanparadigmii.


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

\footnote{Geavahan dás Jurafsky ja Martin (2008, 120) definišuvnna mas ‘types’ leat iešgužetlágan sátnehamit, ja ‘tokens’ leat buot sátnehamit.}
ollu unnit sánnettippat go sámegielain. Engelasgiela varbalemmas leat guokte dajhe golb-
ma iešgudetlágan finihtta varbahámi (earret to be, mas leat vihtta) ja íi okta dain leat hui hárve. Jus dán buohtastahttit davvisámegielain, de oaidnit ahhe bárrastávval varbbain leat 58 iešgudetlágan finihtta hámi, mat gokčet 45 paradigmasealla, go buot normatiiva variánttatt leat mielde. Seammás leat engelasgielas stuurra teakstačoakkkáldagat, nugo Corpus of Historical American English (COHA) mas leat 400 miljovnna sáni, ja Google Books mas leat 155 miljárdda sáni. Nu lea sánnettippaid ja sániid gorálašvuoughta ollu stuorit engelasgielas go sámegielain, ja engelasgiela heive bures giedahallat statistihkalaš bargovugiguin.

Antonsen ja Trosterud (2010) ozaiga guovtti dábálaš varbba sojahanhámiid engelas-
giela, girjedárogiela ja davvisámegiela odda testameanttain. Sudno guorahallan čájeha ahhe teavsttain ledje buot engelasgiela say ja serve ja dárogiela si ja tjene hámi, muhto dušše 44 % dadjat-varbba hámiin ja 31 % bálvalit-hámiin. SIKORa (= Sámi Interna-
tionála Korpus, UiT Norgga árktalaš universitehta ja Norgga Sámedikki sámi teaksta-
čoakkkálpat9) davvisámegiela oasis leat 25 miljovnna sáni, muhto dattetge váilot earret eará ollu dábálaš varbbaíd indikiivva guvttiidlogu hámi, nugo varbahámit dajaide ja boraide, ja dát čájeha ahhe íi leat vejolaš viiežžat buot sátnehámiid teakstačoakkkál-
dagas. Dan dihte sámi giellateknologija geavaha giellamodealla, mas leat automáhtat mat huksejitu buot sátnehámiid.

Teakstagiedahallamii mii dárbbášit maiddái sáni analysa, ja grammatihkalaš automáh-
ta muitala sáni lemma ja morfologija. Molsaevttolaš vuoiki livččii máddagastin (= eng. ‘stemming’, geahča ovdamearkka dihte Jurafsky ja Martin (2008, 102)), mii íi heive bures sámegielaiide, gielaid siskkáldas morfofonologalaš proseassaid dihte (Antonsen ja Troste-
rud, 2010).

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

---

2.2 SÁTNEHÁMIID HUKSEN


AKU-joavku (Avointa kieliteknologiaa uralilaisille vähemmistökielille)11 lea huksen nuortalašgiela analysáhtora (Giella-sms), ja The Freiburg Research Group in Saami Studies12 lea huksemin bit-onsámegiela analysáhtora (Giella-sje). Manit analysáhtora huksen álggahuvvui bitonsámi dokumentašuvnaprošeavttas Frán kust til kyst – Áhpegáttest áhpegáddáý13 mas Giellatekno lei mielde (Sjaggo ja Trosterud, 2015). Buot dát analysáh
torat geavahit Giellatekno- ja Divvun-joavkkuid Giella-vuodostruktuurrva, mii álkiadahttá huksenbarggu ja buvttada sátnedárkkissanprógramma (Moshagen ja earát, 2013). Oaná-

Analysáhtoriid morfologalaš oassi lea transduser, man sáhttá govahallat dego stuurra bálggesvuogádat mii manná giela buot vejolaš máddagiid ja geažuskombinašuvnnaid čada, 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 kompilerenprográamma) programmeriengiela

12http://site.uit.no/pite/ (14.02.2018)
KAPIHTAL 2. GIELLAMODEALLA HUKSEN

lasihat affivssaid sáni máddagii ja čatná dan oktii grammatihkalaš sániin. TWOLC (eng. ‘Two Level Compiler’ = guovttedásat kompilerenprogúmman) programmerengiela rievda-
da vuollálas hámiid rivtteshámat bajošhápmín (= eng. ‘surface form’) ortográfalaš ja
morfofonologalaš njuolggadusaid mielde. LEXC ja TWOLC fiillaid kompilerejuvojité14
bineara fiilan HFST-progúmmain (eng. ‘Helsinki Finite State Transducer Technology’ =
Helssega loahpalaš dilletransdusera teknologiija15) (Lindén ja earát, 2011) dahje XFST-
progúmmain (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 sáhttá HFST heivehit ja ovttaiduhttit eará programmagaun. Dehálaš lea maiddái
ahte HFST máhttá rievdadit bajit dási gilkoriid guovttedásat diakrihtalaš leavgan, ja dát
ii doaimma nugo galggašii XFST-progúmmas. Dákkár gilkoriid analysáhtorat geavahit
sáni goallosteamii stivremii, ja dan čilgen kapihttalis 3.3.2.

2.2.1 Geažusmorfologiija huksen

Govus 2.1: Automáhta (FSA) mii hukse sátnehámiid lávka, lávkan, girji, girjin. (Artikh-
kala VII govus 1.)

Geažusmorfologiija huksen sáhttá 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 sánii-
guin lávka, girji, ja das leat guokte bálgá loahpalaš dillái, nuppis lea geažus n ja nuppis
ii leat. Boadusin leat hámit lávka, girji, lávkan, girjin.

Go bargá grammatihkain, de lea ávvkálša diehtit ahte sánit lávkan ja girjin leat
essiiva-hámít, ja dan dihte mii geavahit grammatihkalaš transdusera (Beesley ja Kart-
tunen, 2003), geahča govvos 2.2. Dát lea dihto automáhtatiipa mas juohke sátnehámis

14Čállosis geavahan vearbba kompileret dušše go oaivvildan gáldofiillaid ovttastahttit binára fiilan.
2.2. SÁTNEHÁMIID HUKSEN


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

Govus 2.3: Govvosisleaoasášsubstantiivaleksikonas. Juohke linnjás lea okta mearkkuš, ja merkošis leat golbma oasi: 1) lemma (ruoná) ja gilkorat, nugo +CmpN/SgN (goallosteami várs) ja +Sem/Ani, mii muitala semantihkas (ruoksadat). Duppalčuoggá olgeš bealde lea 2) LEXC-mátta masa gehčosat goallostuvvojít (alit) ja 3) joatkaleksikona namma (ruoná). Mearkkuš loahpahuvvo beallečuoggáin.

Gielas-smX morfologalaš transduserat leat ráhkaduvvon LEXC-prográmmmerengielain, ja gohčoduvvojít leksikontransduserin ja dain leat mánga leksikona. Juohke leksikonas lea sierra namma, ov damearkka dihte gávdno vearbaleksikon man namma lea LEXICON VERB. Juohke sátneluohká várás lea unnimustá okta leksikon, muhto dávjá leat ovttá sátneluohká lemmat juhkkojuvvon eanet leksikonaide vai álkit sáhttá stivret sáníid goal-

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

Ovtta leksikonas leat diito meari merkošat (= eng. ‘entries’), ja juohke mearkkuš lea cáollojuvvon ovttta limnjái. Dábálaččat lea okta mearkkuš juohke lemma várás, muhto ovttta lemmas sáhttet leat eanet merkošat, jus leat variánttat. Juohke mearkkuš sisttis-doallá golbma oasi: lemma oktan gikoriiguin ja LEXC-mátta, sirrejuvvon duppalčuoggáín, dasto boahtá sátnegaska ja de joatkkaileksikona namma. Loahpas lea sátnegaska ja beallečuokkis, nugo govvoisin 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 báhcahagat (= eng. ‘triggers’), mat gullet ieš sátnehápmái, ja dán oasi gohčodit fas leksikontransdusera vuolit dássin. Gilkorat leat máŋggaláganan: morfologalaččat, semantihkalaččat ja dasa lassin gilkorat mat stivrejit goallosteami (geahča kapihtalla 3.3.3.1 ja 3.3.2).

Sániit mat čuvvot seammá sojahanminstara mannet seammá joatkkaileksikonii. Dавви-sámegielas leat substantiivvain golbma máttatiippa: bárrastávval, bárahisstávval ja kontrakšuvdanáddagat,17 muhto maiddái ovttta máttatiippa siskkobealde leat erohusat sániid sojahanminstariid gaskkas. Govvoisin 2.3 njeallje bárrastávval substantiivva mannet AIGI-joatkkaileksikonii, golbma bárahisstávval substantiivva mannet MALIS-joatkkaileksi-koniin ja kontrakšuvdanabármennyvat mannet juogo FALES dahje BOAZU joatkkaileksikonai de. Leksikonaid namat veahkehit lingvistta välljet rivttes joatkkaileksikon go lasiha

17Eará tearpmat leat vokálamáttta, konsonántamáttta ja kontrakšuvdanamáttta (Nickel ja Sammallahti, 2011).
2.2. SÁTNEHÁMIID HUKSEN

odda sáníid fiillaide. Davvisámeigela 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


Leksikontransdusera bajit dássi ovddasta grammatihkalaš sáni, vuovdi+N+Sg+Loc, ja vuolit dásis lea sreanga mas lea mätta ja affivssat ja dasa lassin báhcahagat mat morfonologalaš transduseris báhcahit morfonologalaš rievddademiid. Govvosis 2.3 leat guokte vuovdi-lemma. Nuppis lea semantiikalaš gilik +Sem/Plc ja nuppis lea +Sem/Hum. Dát guokte lemma mannet goabbatlágan joatkkaleksikonaide danne go dat gullet goabbatlágan sojahaminstrariidda. Lemma vuovdi:vuovdi’i mammá vidáseappot AIGI-joatkkaleksikonii kos lasihuvvo ee. +N+Sg+Loc:”WGs, 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, mat fas leat attan morfonologalaš transduserii, nugo govvo-siin 2.5 ja 2.6. FST buvttusin leat sátnéhamit vuovdis (vuovdi+N+Sem/Hum+Sg+Loc) ja vuovddis (vuovdi+N+Sem/Plc+Sg+Loc).

Govus 2.5: Transduser ráhkada sátnehámiid lávka ja vuovdi. Leksikontransduseris addojuvvo ^WG-báhcahat, mii morfofonologalaš transduseris báhcaha konsonánttaguovddáža rievddademiid vk:vvk ja vd:vdd. (Artihkkala I govus 1, heivehuvvon.)

Govus 2.6: Leksikontransdusera bajit dássi lea grammahkalaš sátäi, ja vuolit dásis lea sreanga, mii fas lea morfológhalaš transdusera atta (LEXC-mátta, geaužus ja báhcahat). Buvttus lea sátnehápmi vuovddis. Ruoná oassi jávká dalle go dát guokte transdusera kompilerejuvvojet oktan transduserin. (Artihkkala I govus 2, heivehuvvon.)
2.2. SÁTNEHÁMIID HUKSEN


Test 196: Verb – tuárššud (Lexical/Generation)

| 1/9 | PASS | tuárššud+V+Inf == tuárššud |
| 2/9 | PASS | tuárššud+V+Ind+Pres+Sg1 == tuáršum |
| 3/9 | PASS | tuárššud+V+Ind+Pres+Sg3 == tuáršu |
| 4/9 | PASS | tuárššud+V+Ind+Pres+Du1 == tuáršsoon |
| 5/9 | PASS | tuárššud+V+Ind+Pres+Pl3 == tuárššuh |
| 6/9 | PASS | tuárššud+V+Ind+Pres+ConNeg == tuáršu |
| 7/9 | FAIL | tuárššud+V+Ind+Prt+Sg1 == Missing results: tuórššum |
| 7/9 | FAIL | tuárššud+V+Ind+Prt+Sg1 == Unexpected results: tuórššum |
| 8/9 | PASS | tuárššud+V+Ind+Pres+Sg3 == tuáršui |
| 9/9 | PASS | tuárššud+V+Ind+Prt+Pl3 == tuórššuuv |

Govus 2.7: YAML testenfiillain iskamin anárašgiela yearbbba tuárššud sojahanhámiid genererema. torššum sajis generejuvvo tuorššum.

Anárašgiela FST huksen lea buorrin ovdamearkan das mo sáhttá bargat dalle go ii leat doarvái dokumentašuvdna ovdalgo álggaha barggu. Anárašgiela lingvista lei guovddáš ohmmoš dán barggus, ja huksiimet YAML formáhtas testenfiillaid main leat oktiibuoat 22 435 grammatihkalaš sátni:sátnehápmi-pára (eng. ‘YAML Ain’t Markup Language’
KAPIHTAL 2. GIELLAMODEALLA HUKSEN


"Consonant shortening and gradation kk:k and kk4:k"
{k:0 <=> Vow: _ [k:k4:] Vow: [WS:1%H:CSH:] ([WS:1%I:RVH:])] %> ;

"Gradation for kk4"
k4:h <=> Vow: k: - Vow: (Cns:) %WS: ;

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

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


Anárašgiela dássemolsumis muhtun gëmináltain leat guokte vejolaš rievddadeami, nugo kk:h ja kk:k, päkkį:päähi (= båhkka) ja kukkan:kukkan (= lieddi). Dákkár dáhpahusain ovdavålljejuvvon rievddadeapmi lea merkekeahntá, ja erenoamáš rievddadeapmi merke-
juvvo arkifoneman, dás k4, ja nu oazžut guóke vejolaš rievddadeami, kk4:h ja kk:k.
TWOLC-njuolggadusat leat čilgejuvven govvosis 2.8.
Anárašgiela transduseris leat geavahan 16 báhcahaga. Tabealla 2.1 lean čilgen ja ad-
dán ovdamearkkaid das mo báhcahagat rievdidit vokálaid ja konsonánttaid guhkkodaga ja kvalitehta.

<table>
<thead>
<tr>
<th>Báhcahat</th>
<th>Posišuvdna</th>
<th>Makkár rievdan</th>
<th>Ovdamearka</th>
</tr>
</thead>
<tbody>
<tr>
<td>°RLEN</td>
<td>vokálaguovddáš</td>
<td>guhkkun</td>
<td>a &gt; aa, alge:aalgan ‘bárdni’</td>
</tr>
<tr>
<td>°RVSH</td>
<td>vokálaguovddáš</td>
<td>oatnun</td>
<td>aa &gt; a, vaará:varáttáá ‘vára’</td>
</tr>
<tr>
<td>°VBACK</td>
<td>vokálaguovddáš</td>
<td>kvaliteahtta</td>
<td>ā &gt; a, väzzid:vazzim ‘vázzi’</td>
</tr>
<tr>
<td>°VHIGH</td>
<td>vokálaguovddáš</td>
<td>kvaliteahtta</td>
<td>ā &gt; ā, väzzid:vázáččij</td>
</tr>
<tr>
<td>i2</td>
<td>vokálaguovddáš</td>
<td>kvaliteahtta</td>
<td>iā &gt; e, siámmoo:semmp ‘hoavrristit’</td>
</tr>
<tr>
<td>°CLEN</td>
<td>kons.guovddáš</td>
<td>guhkkun</td>
<td>h &gt; hh, päähi:pahhijn ‘báhkkí’</td>
</tr>
<tr>
<td>°WG</td>
<td>kons.guovddáš</td>
<td>dássésmolun</td>
<td>tt &gt; d, ruttt:ruudá ‘ruhta’</td>
</tr>
<tr>
<td>°SLEN</td>
<td>soggi</td>
<td>guhkkun</td>
<td>e &gt; ee, alge:algeest</td>
</tr>
<tr>
<td>°SVSH</td>
<td>soggi</td>
<td>oatnun</td>
<td>ee &gt; e, eemeed:emedán ‘eamit’</td>
</tr>
<tr>
<td>°SVLOW</td>
<td>soggi</td>
<td>kvaliteahtta</td>
<td>u &gt; o, áldu:áldoin ‘áldu’</td>
</tr>
<tr>
<td>°ÁI</td>
<td>soggi</td>
<td>kvaliteahtta</td>
<td>oktan °WG:in á &gt; i, kuáhtás:kyevtis ‘guovttis’</td>
</tr>
<tr>
<td>°ÁE</td>
<td>soggi</td>
<td>kvaliteahtta</td>
<td>oktan °WG:in á &gt; e, luámmán:lyeme ‘luomi’</td>
</tr>
<tr>
<td>u2</td>
<td>soggi</td>
<td>kvaliteahtta</td>
<td>u &gt; o ja</td>
</tr>
<tr>
<td>°CSH</td>
<td>kons.guovddáš</td>
<td>kvaliteahtta</td>
<td>ſáá &gt; á, čäälli:čallud:čállojjí ‘cállit’</td>
</tr>
<tr>
<td>°EA</td>
<td>vokálaguovddáš</td>
<td>oatnun</td>
<td>aa &gt; a, aarrün:aránnän ‘árran’</td>
</tr>
<tr>
<td>°FCD</td>
<td>soggi</td>
<td>kvaliteahtta</td>
<td>e &gt; ſá ja</td>
</tr>
<tr>
<td></td>
<td>loahppakons.</td>
<td>sihkkun</td>
<td>siihkkut n, luámmán:lyeme</td>
</tr>
</tbody>
</table>

Tabealla 2.1: Anárašgiela transdusera báhcahagat ja daid morfonologalaš rievd-
dadeamit. i2 ja u2 leat sokki arkifonemat, mat dasa lassin báhcahit vokálaguovddáža rievdadeami.

Morottaja (2016) válddáhallá Giella-smn leksikona- ja morfonologalaš transduseriid vuorrováikkulusa. Son evttoha moissaaevttolaš luksenvuogi mas lea eanet morfonologo-
galaš diedut LEXC-máddagis. Bárrastávval máddagiid soggevokála merkejuvvo arkifone-


2.2.3 Molssaevttolaš vuogit

2.2.3.1 Morfonologalaš ráidonjuolggadusat


Buhttennjuolggadusaiguin morfonologiiju giedahalloyuvvo vugin mii sulastahttá generatiiva fonologiia. Njuolggadusaid ortnet lea dehálaš danne go juohke njuolggadusa buvttus lea čuovvovaš njuolggadusa atta, ja na njuolggadusat doibmet ráidun. Dáinna lágiin bustávva sáhttá rievdat oddasit, go das eai leat dušše guoke dási nugo TWOLC-prográmmas, mas buot njuolggadusat doibmet bálddalasat.
2.2. SÁTNEHÁMIID HUKSEN


2.2.3.2 Buot morfonologijja lea leksikontransduseris


Suomagielas lea vokálaharmoniija, ja giela gávcci vokálas, leat guda vokálas morfonologalaš ráddjejumit, ja dat leat golbma ovdavokála *y, ö, ä* ja golbma manŋevokála *u, o, a*. Vokálaharmoniija guoská sání máddagii: Sánis ii sáhte leat sihhe ovdavokála ja manŋevokála, muhto dat guoská maiddái morfologiijii. Jus máddagis lea ovdavokála, de gehčosis ii sáhte leat manŋevokála, ja jus máddagis leat manŋevokálat, de gehčosis ii sáhte leat ovdavokála. Gehčosisin leat dan dihte guokte variántta, ovdavokálain dahje manŋevokálain.

Vokálaharmoniija livččii vejolaš čoavdit TWOLC-prográmmmain, ja ráhkadit njuolggadusa mii rievdadivččii gehčosa ovdavokálaid manŋevokálan *ä:a*, *y:u*, *ä:o* jus máddagis
2.3 Buot sániide addit analysa

2.3.1 Sáni olggobeale ja siskkobeale hápmi (Art. I)

Artihkal I (Ord sett innfra og utfra – en datalingvistisk analyse av nordsamisk) suokkardállá man bures Giella-sme analysere SIKORa davvisámi teavsttaid. Sáhhtágo analysa muitalit sátneráhkadeami produktiviteahta birra? Maid analysa muitala substantiivvaid kásusiid juohkáseamis ja vearbbaid finihtta sojahanhámiid juohkáseamis?

Artihkkalis boahtá ovdan ahte Giella-sme addá morfologalaš analysa 97,8 % korpusa sániide. Dalle eai leat miede dovdameahttun sánit maid dárogiela analysáhtor dovdá dáru sitáhtan, eague bajilčállagat mat leat čállojuvvon versálaigvin, danne go analysáhtor ii leat heivehuvvon daidda. Bargu dahkat vejolažžan analyseret sániid čállojuvvon versálaigvin ja earuhit daid initiálaaoánádusain ja namain, ii leat vuoruhuvvon danne go dát čovdojuvvo geavaheddiidprográmmmain. 

Govus 2.10: Kásusiid juohkáseapmi SIKORa šáŋŋeriid miede mat leat hálldahus-, fákta, diedalaš-, láhka-, aviisa-, čáppagirjjálašvuoda- ja oskuteavsttát. Juohke sársggis čájeha ovtta šánjera kásusjuohkáseami (oktiibuet 100 %). (Artihkkala I govus 7).

Listu mas leat 2,2 % dovdameahttun sánit, sisttisdoallá eará gielaid sitáhtaid (mat eai leat dáru sitáhtat), ja dasa lassin čállinhámidi mat eai leat norpma miede čállojuvvon, eague lasihuvvon descriptiivva analysáhtorii. Listtus leat muhtun dovdameahttun
namat ja muuduge dólkaša sámegiela sáñit mat váïlot analysáhtoris, muhto das leat maiddái sáñit mat leat originálfíillaid heajos konverterema bohtosat. Dávjá guoská dáti iešgudet vugide čällit sámegiela erenomáš bustávaid ovdalgo Unicode-standárda18 šattai dábaalažžan, ja váttisvuodaide OCR-lohkamis (eng. ‘Optical Character Recognition’ = optihiššaš mearkaáicin) ja pdf-fiillaid konverteremis teakstan. Váttisvuodat čovdojuvvojit muhtun muuddui go lasiht njuoleggadusaid juohe dokumeantta konverterenfiilii, dan sajis go divvut origináladokumeanttas.


2.3. BUOT SÁNIIDE ADDIT ANALYSA

10,7 % korpusua substantiivvain ožžo dynamalaš goallostanalysa. Guorahallamis boahtá ovdan ahte sáhtášii leat stuorit lohku, danne go dasa lassin 0,6 % substantiivvain leat goallossánit maid mearusoasi ja vuoodooasi analysáhtor dovdá, muhto leksikontransduseris ii leat goallostanbálgeois daid gaskka.

Artihkkalis letne maiddái árvvoštallan muhtun sátneråhkadanproseassaid produktivi-teahta, ja goallosteapmi lea ollu eanet produktiona go suorggideapmi.

2.3.2 Sámegielaid morfologalaš analysáhtoriid buohtastahttin

Buot Giella-smX analysáhtoriin lea sihke leksikontransduser ja morfotonologalaš transduser. Muhto dannego cállingielain eai leat seammá ollu morfotonologalaš proseassat, de morfotonologalaš transduserin eai leat seammá ollu njuolggadusat. Maiddái morfotonologalaš njuolggadusaid organiseren váikkuha daid lohku, nugo čilgejuvon kapihttalis 2.2.2, ja ollu rievddademiid sáhttá maiddái çovvdi eanet joatkaleksikonaid bokte, nugo suomagiela Giella-fin čájeha.


Gielaš morfotonologalaš kompleksitehta ektui lean tabellii 2.2 bidjan davvi-, julevja bitonsámie analysáhtoriid oktan joavkun, ja anáraš- ja nuortalašgiela analysáhtoriid
<table>
<thead>
<tr>
<th></th>
<th>TWOLC</th>
<th>LEXC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Njuolggadusat (konteavsttat)</td>
<td>Variá-belat</td>
</tr>
<tr>
<td>I</td>
<td>Davvisámegiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-smé</td>
<td>112 (114)</td>
</tr>
<tr>
<td></td>
<td>Julevsámegiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-smj</td>
<td>65 (262)</td>
</tr>
<tr>
<td></td>
<td>Bitonsámegiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-sje</td>
<td>27 (36)</td>
</tr>
<tr>
<td>II</td>
<td>Anárašgiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-smn</td>
<td>90 (250)</td>
</tr>
<tr>
<td></td>
<td>Nuortalašgiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-sms</td>
<td>115 (419)</td>
</tr>
<tr>
<td>III</td>
<td>Lullisámegiela</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Giella-sma</td>
<td>50 (80)</td>
</tr>
</tbody>
</table>

gadusat main leat erenoamáš ollu vejolaš bustárvvat listejuvvon variábeliidda. Njuolggadus
mas leat eanemus vejolašvuodat variábeliid ektui, dárbbása 41 sekundda kompilerejuvoit.

Garvin dihte morfonologalaš njuolggadusaid, de sáhttá ráhkadit eanet joatkkaleks-
ikonaid. Tabealla 2.2 olgeš ceakkočuolddla logut muitalit galle iešgužetlágan joatkka-
leksikonii substantiiivvat mannet, cájehan dihte huksehusaid erohusaid. Davvisámegiela
analysáhtoris leat eanet joatkkaleksikonat go julevsámegiela analysáhtoris, vaikko gielain
leat sullii seammá ollu morfonologalaš proseassat. Dát stuurra joatkkaleksikonlohku
boahtá Divvun-joavku barggus davvisámegielain hehtten dihte badjelmearáláš genere-
rema ja garvin dihte sátnehámíid mat gökcet cállimmeattáhusaid (dan birra cállan kapih-
talis 3.3.3).

<table>
<thead>
<tr>
<th>Analysáhtor</th>
<th>Substantiiivvat</th>
<th>Vearbbat</th>
<th>Adjektiivvat</th>
<th>Advearbbat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dynamálaš</td>
<td>dynamálaš</td>
<td>dynamálaš</td>
<td>sátnegihput</td>
</tr>
<tr>
<td>Davvisámegiela</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giella-sme</td>
<td>17 635</td>
<td>75 112</td>
<td>11 970</td>
<td>3350</td>
</tr>
<tr>
<td>Anárásgiela</td>
<td>8523</td>
<td>6761</td>
<td>4805</td>
<td>891</td>
</tr>
<tr>
<td>Giella-smn</td>
<td>12 972</td>
<td>4495</td>
<td>8945</td>
<td>2067</td>
</tr>
<tr>
<td>Lullisámegiela</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giella-sma</td>
<td>13 754</td>
<td>1220</td>
<td>6458</td>
<td>1045</td>
</tr>
<tr>
<td>Julevsáme-giela</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giella-smj</td>
<td>13 754</td>
<td>1220</td>
<td>6458</td>
<td>1045</td>
</tr>
<tr>
<td>Nuortalašgiela</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giella-sms</td>
<td>13 754</td>
<td>1220</td>
<td>6458</td>
<td>1045</td>
</tr>
<tr>
<td>Bitonsámegiela</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giella-sje</td>
<td>2440</td>
<td>0</td>
<td>1648</td>
<td>0</td>
</tr>
</tbody>
</table>

Tabealla 2.3: Giella-smX leksikontransduseriit buohtastahttin: Galle iešgužetlágan lem-
ma leat dihto leksikonain. Olgeš ceakkočuolddla logus (+) leat dušše sánit mat maiđdái
ožžot dynamálaš analysa, gurut logus (−) dat eai leat mielde. Advearbbaivn leat sá-
negihput sierra lohkun. Bitonsámi analysáhtoris ii leat vuos dynamálaš sátnehuksen.
(Analysáhtorat 09.01.2018.)

Tabealla 2.3 buohtastahttá Giella-smX leksikontransduseriit substantiiva-, vearba-
adjektiiva- ja advearbafiillaid sturrodagaid. Substantiiivvaid ja adjektiivvaid olgeš ceakko-
从小段的文本中可以看出，它讨论了两种语言的分析方法，尤其是分析了两种语言的词汇和语法。文本中提到了一种叫做“bajássaddat”的分析方法，以及另一种叫做“lávžegeahčen”的方法。这两种分析方法都被用于对这两种语言的词汇和语法进行详细的研究。文本还提到了一种叫做“ovddos guvltui”的方法，它被用于对语言的词汇和语法进行详细的研究。
2.3. BUOT SÁNIIDE ADDIT ANALYSA

Anárašgiela ja nuortaláš giela analyysáhtoriin leat mudui unnimusat sánit mat eai leat goallosteamit dahje suorrgádusat, ja dat leatge Giella-smX oddaseamos analyysáhtora gaskkas. Tabealla 2.3 substantiivalogus eai leat namat (proprat). Juohke analyysáhtoris leat sullii 50 000 nama, ja dain leat 32 000 nama sierra oktasaš fiillas mii genererejuvvo buot sámegielaide. Oktasaš fiillas leat namat nugo Sjur ja Colorado, mat sáhttet geavahuuvot buot sámegielaide teavsttain.

Tabeallas 2.4 lean buohtastahttán man bures Giella-smX analyysáhtorat máhttet analyseret SIKOR-teavsttaid. Davvisámeigiela analyysáhtoris lea čielgasit buoremus gökcan-mearri, 97,6 %. Anárašgiela analyysáhtora gökcanmearri lea buoret go julevsámeigiela ja lullisámeigiela analyysáhtoriid gökcanmearri, vaikko anárašgiela analyysáhtoris leat un-nit lemmat (tabeallas 2.3). Anárašgiela korpusis dáidá leat un-nit varisašuvdna go eará korpusiin go stuorra oasi lea ovtta čállí bargu (Olthuis ja earát, 2016). Nuortalášgiela analyysáhttor ii máhte analyseret go 84,1 % korpusis, ja sivva sáhttá leat ahte Giella-sms analyysáhtoris leat un-nit lemmat go eará Giella-smX analyysáhtoriin. Mudui lea čáljegielas ollu varisašuvdna ja mágii lea váttis giela dárkagiid konverteret.

Sániin lea dáblelaččat eanet go okta analysa, ja disambigueren, rivttes analysa váljen konteavstta ektui, čilgejuvvo kapihttalis 2.4. Vaikko davvisámeigiela 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 dehálāš ahte syntávssalaš analyysáhtor lea gierdil ja máhttá addit analysa vaikko cealkagis leat dovdameahttun sánit.

2.3.3 Eahpesihkkaris norbma ja ollu variánttat


Davvisámi čállingiela norbma dohkkehá variašuvnna, ja erenoamážit muhtun loatnasániid čállinhámiiin lea variašuvdna sátnegirjjis nubbái (Artihkal III). Maidáái sojahanparadigmain sáhttet leat bálddalashámít, erenoamážit adjektiivvaid kompareremis. Adjekiivia *njállgat/*njállgis sodjá sihke bárahisstávaladjektiivan ja kontrakšuvdnaadjektiivan (Nickel ja Sammallahti, 2011, 643), ja Divvun-prográmma dohkkehá mánga kom-
2.3. BUOT SÁNIIDE ADDIT ANALYSA

paratiivahámi: njálgát, njálgásut, njálgásit, njálgáset, njálgáseabbo, njálgásat, njálgásabo.

Artihkkala III dutkanmateriálas leat 4 % sániin čállojuvvon norpma vuostá, ja dan logus leat maiddái duohtasátne-čállinmeattáhusat miedle (sániit mat gávdnojít gielas eará mearkkašumiin dahje sojahemiin). Artihkkala ulbmiil lea oažžut muhtun muddui homoge-
na dutkanmateriála, ja mun gudden teavsttaid main vuhttuui ahte čállis ii lean sámegiillu
vuosttašgiillan dahje ii lean hártjánan čállit sámegillii. SIKORis leat buot lágan teavsttaid ja maiddái teavsttaid mat leat čállojuvvon ovdalgo dihío normerenmearrádusat bohte jágis
2005. Sániin mat ožžot analysa davvisámi SIKORis, ožžot 4,8 % analysa dušše deskripti-
iva analysáhtoriin, ja dat mearkkása ahte normatiiva analysáhtor ii dovdda dait. Dasa
lassin bohtet sániit mat ožžot boasttuanalyasa. Ovdamearkka dihte leat korpusa analysas
substantiiivahámít vuosttáš ja vuosttážin 404 geardde, muhto juohke dáhpáhus lea rievt
miedle ortnetlohku vuosttáš, essiivvas vuosttážin, mii ii leat čállojuvvon norpma miedle.

Lullisámegiela čállingiela alfabehtas leat bustávat æ ja ö, ja bustávat å ja ø leat norp-
ma vuostá. Mänɡa čálli valljejít dattetge geavahit juogo å, ö, mat heivejít ruotagiela
boallobeavdái, dahje æ, ø, mat fas heivejít dárogiela boallobeavdái. Lullisámegiela al-
fabehetas lea sihke i ja i, muhto ollu čállit geavahit čadat dušše i. Vai sáhttá analyseret
lullisámi teavsttaid, de deskriptiiva analysáhtor ferte dohhkehit å, ø, i bustávaid æ, ö, i
sajiin, ja dat gohčoduvvo stávenložžemin (= eng. ‘spell relax’). Olles lullisámi korpusis
ožžot 6,6 % sániin analysa dušše deskriptiiva analysáhtoriin – eaige normatiiva analysá-
htorii. 10,9 % dáin sániin sisttisdollet å dahje ø. Dát duodašta ma dárbbašlaš sierra
boallobeavdi lea. Juo 1980-logu rájes lea gávdnon davvisámegiela boallobeavdi dihtora
várás, muhto easkka jagi 2017 rájes lea leamaš vejolaš viežžat sierra boallobeavddi dihtorii
maiddái julev- ja lullisámegiela várás19.

Ollu variašuvdna čállingielas dahká hástaleaddjin ráhkadit analysáhtora go šaddet
eanet iešgudetlágan sátnehámít, ja variašuvdna dahká vel dehálačcabun testet analysáht-
tora teakstakorpusin ja lasiit sániid ja sojahemiid maid analysáhtor ii dovdda. Variánt-
tat mat leat duohtasátne-čállinmeattáhusat, dahket váddéseabbon oazžüt rivttes syntáv-
salaš analyxa. Syntávssalaš njuolggadusaid čállin lea čuvvvošaš kapihttala fáddá.

2.4 Rivttes analysa válljen

Jus hâliida diehtit kâsusiيد juohkáseami teakstakorpusis, dahje geavahit analysa vuoddun dihotorjorgaleapmái, de ii leat doarvái diehtit sátnéhámí vejolaә morfologalaә analysaid, 
muhto lea dárbbashaә diehtit sâni morfologalaә analysa konteavstta ektui. Dat mearkkaә 
alte lea dårbu disambigueret analysaid gaskkas (disambigueret = jávkadit guovtte- ja 
månggaәilggolaәsvuoda).

2.4.1 Disambigueren ráddjehusgrammatikhkain,

.Constraint Grammariin

Morfologalaә transduzer (FST) máhattá sihke analyseret ja genereret sâniid. Syntávssalaә 
alanalysâhtor máhattá duәse analyseret cealkagiid, iige genereret daid.

Syntávssalaә analysâhtor sisttisdoallá grammatihka mii formulere ráddjehusaid das 
gudemuә analysa heive kontekstii. Ráddjehusaid lea ohmmoә cállán. Formalisma gohco-
duvvo *Constraint Grammarin* (Karlsson, 1990), ja dan kompiláhtora vuostaә verþuvnmaid 

Maŋŋel leat Tapanainen (1996) ja Eckhard Bick ja VISL-group (2008) nâhkadan odda, 
buoriduvvon verþuvnmaid. Dálá Giella-smX-analysâhtoriid njuolggadusat ovttastuvvojít 
VISLCG3-prográmmain.

Oktilaә teavsttas lea ollu homonymiija. Davvisámegielas leat eanet homonymat go 
ovdamearkka dihte julevsámegielas mii lea konservatiivvalač️ät go davvisámegiella, ja 
gehčosat eai leat gahčan oktii seammá ollu go davvisámegielas. Davvisámegielas lea 
measta olles homonymija akkusatiivva ja genitiivva gaskkas, ja jus substantiivvas ii leat 
dássemolsun, de ovttaidlogus lea homonymija maiddái nominatiivahámiin. Bárahisstå-
valvearbbain leat eanet homonyma hâmit go bárrståvvalvearbbain, ja *leat*-hámis leat 
olles vihtta analysa: infinitiiva, preseanssa biehttalanhápmi, ja golbma preseansahámi 
indikatiivvas: +Sg2, +Pl1 ja +Pl3.

Homonymiijat leat månggaláganat, ja analosalohku lea dan duohken man ollu sání 
leat leksikontransduzeris, ja man ollu morfologalaә, syntávssalaә ja semantihkalaә dieduid


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

said, ja cealkka šaddá njulgestaga ilá kompleaksan. Badjin vulos analysáhtorat gáibidit maiddái ahce buot sánit ožžot morfologalaš analysa. Sámegielaid analysáhtorat barget vuollin bajás (= eng. ‘bottom up’) vuogi mielde, ja dan dihte analysáhtor álo addá analysa, maiddái cealkkafragmeanttaide ja guhkes moalkecealkagiidda. Lullisámi analysáhtorat evaluere mis dat nákce oalle bures addit rivottes lemma oktan sátneluohká-analysain vaikko testenkorporus leat ollu dovdameeahttun sánit, geahča kapihtal 2.4.2.


(1) Dat lea goîtge canas áigge nu ahce dearvan besset rastá joga. (SIKOR: Johan Turi, Muitalus sámíid birra)

(2) Ja go leat beassan joga rastá dearvan, (...) (SIKOR: Johan Turi, Muitalus sámíid birra)

(3) Sutnje i leat skuvlaovtasbargu rájiid rastá amas ášši. (SIKOR: Ávvír)

Maiddái joga-sánis lea guovttečilggolašvuohta, 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.

1. Finihtta vearba ii heive ovddabeale ráiddu ‘riebtalanvearba + leat + substantiiva nominatiivvas’ jus ii leat jietnadanvearba, dahje jus ii leat subjunkšuvdna dahje relatiiivapronomen vearbba gurut bealde: Jávkkat finihtta vearbaanalysaid +Prs+Du1, +Prt+Pl3.

2. Gurut beale sátni lea riebttalanvearba: Vállje vearbba riebttalanhámi +ConNeg.


5. Olgeš beale sátni lea substantiiva, iige boade vearba dan manŋel: Vállje adjektiivva attribuhttalámi +A+Attr.


2.4.2 Sániid lemmatiseren (Art. II)


Njuolggadusaiguin mat geavahit cealkagiid syntávsssa veahkkin, sáhttá viehka álki disambiguereet lemmaid gaskkas mat gullet iešgudetlágan sátneluohkahó. Cealkagis (3) amas-sátni ii sáhtle leat subjunkšuvdnia iige biehttalanvearba danne go čuovvovaš sátni lea substantiiva, eage leat sánit substantiivva olgeš bealde. Ollu váddásat lea disambiguereet lemmaid gaskkas mat gullet seammá sátneluohkkái, nugo vearbbaid amastit ja amasmuvvat. Dás sáhttá geavahit transitivitehta veahkkin. Danne go amastit lea transitiiva vearba, de dat válljejuvvo jus cealkagis lea vejolaš objeakta, son ii amas odda ránnjá.

Čuovvovaš cealkagiin ii leat doarvái luohhtit cealkaga syntáksii, muhto lea dárbu maid-dái geavahit dan maid olmmos diehtá olbnuid, sáhpániid ja luosaid birra, go disambiguere vuodjit ja vuodjat-vearbbaid gaskkas:

(4) Mudui vudjet olbnot alla leahtuin dákko, lohká Gerd Pedersen. (SIKOR: Ávvir)

2.4. RIVTTE ANALYSA VÁLLJEN

(5) Manne sáhpánat vudje rastá joga? (Dán cealkaga parafrása: Manne sáhpánat vuojai rastá joga? SIKOR: Äššu)

(6) Garra ravnjni dihtii vudjet luosat gáttu lahka ja čoahkkánit goatníjii. (SIKOR: Min Áigi)

Sáhttá ráhkadit njuolggadusa mii vállje vearbalemma subjeavtta mielde, ja nu sáhttá válljet *vuodjat* go subjeavttas lea +Sem/Ani-gílkor, mii muitalat ahte lea ealli. Muhto lea váddásat ráhkadit njuolggadusaid mat válljejit riekta, dalle go subjeakta lea olmmoš, danne go olmmoš dábałačcat máhttá sille *vuodjit* ja *vuodjat*.

Lemmatiseren lea ávkkálas go áigu teakstakorporusis ohec odda sániiid sántnegirjái, ja jus galgá jorgalit teavstta nuppi gillii. Lemmatiserema bokte sáhttá sántnegirji jorgalit sáni, vaikko ohcansátni lea sojahuvvon hámis, eanet das lea čilgejuvvon kapihttalís 3.2.2. Muhto maiddái go bargá sátnebålndalastimiiin ( = eng. ‘word alignment’) statistihkalaš lahkovanvugiin, de lea álkit gávdnat dan guovtti giela sátnebáraid jus vuos sáhttá lemmatiseret teavsttaid, go sámegielas lexkemii gullet ollu iešgužetlágan sátnehámit. SIKORis 14 davvisámi vearbbas leat eanet go 100 iešgužetlágan sátnehámi^22. Go lemmatisere teavstta (= buhtte juohke sántnehámi lemmmain oktan sántneluohká gilkoriin), de juohke sáni dávjodat lassána sakka.


Artihkkalis II (*Next to nothing – a cheap South Saami disambiguator*)^23 lea jearaldat sáhttágo unna CG-njuolggadusasæhtain lemmatiseret lullisámi teavstta Constraint Grammariin. Ulbmil lea lemmatiseret korpussteavsttaid, ja artihkkalis boahtá ovdan ahte dušše 115:in disambiguerenjuolggadusain lea vejolaš aazžut viehka buori bohtosa lem-

---


^23Errata: Artihkkala 2. siiddu ‘Figure ??’ galgá leat ‘Figure 2’.
matiseremis. Korpusteavsttain juokhe sánis leat gaskamearállačcat 1,6 analysa go maiddái leat mielde dovdameahttuń sánit (FST gokčanmearri lei duše 91,3 %), ja 1,8 analysa go leat mielde duše cealkagat main lea ollislaš morfologalaš analysa. Disambigueren manai seammá bures buot cealkagin, iige duše daiguin main lea ollislaš morfologalaš analysa. Juokhe sánis lea gaskamearállačcat 1,06 analysa disambiguerema manjel.

Govus 2.11: Giella-sma syntávssalaš analysáhtora njuolggadusaíd beaktilvuohta. Láskkoávssis = galle njuolggadusa. Ceakkoávssis = man stuorra oassi homonymiijas lea disambiguerejuvvon. (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 dárbbasuvvojit 55 njuolggadusa.

Bargu duodašta ahle ráddjehusgrammatihkalaš láhkonanvuohki lea gierdil maiddái dalle go leat 8,7 % dovdameahttuń sánit teavsttain. Artihkkalis evaluerejetne analysáhtora analysa gollekorpusa (= teaksta man analysa lea manuálalaččat divvojuvvon) ektui. Lemmatiserema recall (= buktosa rivttes analysat juhkkojuvvon gollekorpusa analysaide) lea buorre, 98,0 %, muhto deaivilvuohta (= buktosa rivttes analysat juhkkojuvvon buktosa buot analysaide) ii leat seammá buorre: 87,6-88,6 %. Dat mearkkaša ahle 98 % rivttes analysain leat mielde analysaboothsis, muhto guovtte- ja mänggačilggolašvuohta 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 Sámegielaid syntávssalaš analysáhtorat

Nugo namuhuvvon kapihttalis 2.2, de leat guhtta Giella-smX analysáhtora. Buot morfolo-
galaš analysáhtorat leat huksejuvvon FST:n, ja syntávssalaš analysáhtorat geavahit Con-
straint Grammara. Tabeallas 2.5 oaidnit mo Giella-smX lea modulariserejuvvon. Morfolo-
galaš ja morfofonologalaš transduseriid ferte huksit juohke gillii sierra. Lea váttis sirdit
osiid transduseris nubbái maiddái danne go sámegielaid čállingielat leat nu iešgudetlá-
ganat.

Morfologalaš analysáhtora buktosis leat generaliseremat: sátneluohkáid ja eará mor-
fologalaš iešvuodaid gilkorat, ja semantihkalaš gilkorat. Ráddjehusnuolggadusat čujuhit
sihke lemmaide ja gilkoridda, ja nuolggadusat mat čujuhit gilkoridda eai leat čadnon
dušše ovttá gillii.

<table>
<thead>
<tr>
<th>Analysáhtor</th>
<th>Giella-sme</th>
<th>Giella-smj</th>
<th>Giella-sma</th>
<th>Giella-smn</th>
<th>Giella-sms</th>
</tr>
</thead>
</table>
| Leksikon ja | Davvisámi  | Julevsámi  | Lullisámi  | Anáraš-giela | Nuortalaš-
| morfologiija| FST       | FST       | FST       | FST       | giela FST |
| Disam-
|        | Davvisámi | Julevsámi | Lullisámi | Heivehuvvon | Heivehuvvon |
| biguáhtor | CG         | CG         | CG         | davvisámi  | davvisámi |
| Syntávssalaš|           |            |            | CG         | CG         |
| funkšuvnnaid|            |            |            |            |            |
| lasíheapmi|            |            |            |            |            |
| Dependeanssa|            |            |            |            |            |
| lasíheapmi|            |            |            |            |            |
| 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 sámegiella dárbbaša sierra disambiguáhtora, de muhtun sámegielain leat
seammalágan homonymiijat mat dahkan vejolažžan kopieret disambiguerenjuolggadu-
said sekveanssaid analysáhtoris nubbái. Ovdamarkka dihte davvisámeigielas leat sub-
stantiiivvaid otvtaidlogo komitatiivahámit ja mánggaidlogo lokatiivahámit homonymat.
Julevsámeigielas lea seammalágan homonymiija otvtaidlogo komitatiivahámiid ja máng-
gaidlogo inessiivahámiid gaskkas. Davvisámeigielas leat substantiiivvain akkusatiiva ja
genitiiva hámii olles homonymiija, anárašgielas lea seammá homonymiija, muhto dušše ovttaidlogus. Anárašgiela ja nuortalašgiela disambiguáhtorin lea davisámeğiela disambiguáhtor, ja bargu heivehit ja divvut njuolggadusaid daidda gielaide lea easkka álgghahuuvon.

Davvisámi disambiguáhtorin lea bargojuuvvon eanemusat danne go davisámeğiela analysa geavałuuvvvo máŋgga prögrammas. Dítörtjorgalanprögrammaid gáldogiella lea davisámeğiella, juoga mii gáibida buori morfologalaš ja syntávssalaš anałysa (Antonsen ja earát, 2017). Giellaahppanprögrammaa Oahpa davisámeğiela veršuvnnnas leat maiddái giellahárjehusat mat anałyserejít oahppi addaga syntávssalačcat, geahča kapiht-tala 3.4.3. Evaluieremis mii dahkkui jagi 2009 teavsttaiguin dovđameahttun sániid haga, oacčui Giella-sm F-score (mii rehkenasto sihke presišuvnnja ja recall vuodul) 0,99 lemmatiseremis ja 0,94 olles morfologalaš disambiguéremis, ja Giella-smj gis oacčui 0,95 ja 0,88 (Antonsen ja earát, 2010).

Fönkšuvdnagilkiirid lasiheapmi ii leat čadnon ovttia gillii seamma ládje go homonymiija čoavdít. Syntávssalaš anałysáhtor lasiha fönkšuvdnagilkiirid cealkaga sániide, ja njuolggadusat čujuhit vuosttažettiin morfologalaš ja semantihkalaš gilkoriidđa ja hárve lemmaide. Dependeeansanjuolggadusat čujuhit fas fönkšuvdnagilkiiridđa. Dan dihte lea vejolaš geavait seamma dependeansaanalysáhtora buot sámeğielaide (eanet detáljat dás: Antonsen ja earát (2010)).

2.5 Čoahkkáigeassu

Sámeğielaide anałysáhtorat leat ráhkaduvvon giellamodeallan mas buot sátnénámit leat huksejuuvvon transduserin LEXC-prögrammain, ja boadus lea FST mii modellere morfologiija. Sámeğielaide hámii hástaleaddji morfophonologalaš rievdadéamit, maid mii čoavdít guovttedásat bustávvalašuvnnnaiguin TWOLC-prögrammas.

Giella-smX anałysáhtořiit oktan ulbmii lea addit anałysa buot sániide stuorra teaka-tokorpusis. Ärtihkal I giedahallá stuorimus anałysáhtora, Giella-sm, ja ärtihkkalis boahtá ovdan ahte davisámeğiela anałysáhtor gokčá 97,8 % SIKOR-teavsttaid sániin.

Juohke sátni sahttä oazžut eanet go ovttia anałysa, ja danne lea dárbbašlaš disam-
Kapihttal 3

Giellamodealla heiveheapmi duohta giellamáilbmái

3.1 Giellamodealla geavaheapmi geavaheddjiidprográmmmain

Mannan kapihttalis lei sáhka mo Giella-smX analysáhtoriid morfologalaš oassi lea hukse-
juvvon FST:n ja syntávssalaš oassi Constraint Grammar-njuolggadusaiguin. Dán kapiht-
talis leat fáddán mo sáhttá geavahit Giella-smX analysáhtoriid sámegiela cálliid ja oah-
pahalliid geavaheddjiidprográmmmain. Analysáhtorat leat ávkkálaččat giellateknologalaš
prográmmmain 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 oazžut analy-
sáhtora dovdat buot sátnehámiid teakstačoakkáldagas. Dán kapihttalis lea sátnehámiid
genereren dehálaš, ja geavaheddjiiidprográmmmaide leat sihke analysáhtora badjelmearálaš
sátnehámiid genereren ja analysáhtora badjelmearálaš analysat čuolbman. Ulbmil lea
oazžut giellamodealla sulastahtit duohta giela.

Sátnedárkkistanprográmmma (kapihttalis 3.2.1) vuodduduvvá dasa ahte analysáhtor
dohkkeha buot giela sániid ja hilgu sániid mat eai leat čallojuvvon riektačállima miede.
Go sátni lea dov dameahttun analysáhtoriid, de analysáhtor generere relevánta evttohusaid
dasa maíd cálli lei oaiivvildan cállit. Sátnegirji (kapihttalis 3.2.2) analysere addaga vai
gávdná lemma mii fas jorgaluvvo nuppi gillii, ja sátnegirjeartihkkaliin analysáhtor generere sojahanparadigma mii sisttisdoallá lemma grammatihkalaš sániiid albmonemiid. Maiddái giellahárjehusat (kapihttalis 3.4.1) geavahit grammatihkalaš sání ja dan albmonemiid.

Buot namuhuvvon prögrammaide lea dárbu gáržžidit giellamodealla, ja réddjet generererema. Go analysáhtora réddjema birra in leat čállán nu ollu nákkosgirjji artihkkaliin, de lean lasihan kapihttala 3.3, man fáddán lea mo eastradit genereremis hámiid mat eai geavahuvvo gielas ja maiddái analysáhtora gáržžideapmi norpma ektui. Dasa lassin digaštalan dárrbú réddjét genererema ja analysaid pedagogalaš sivaid dihte.

Manjimuš kapihttašiin, 3.4.2 ja 3.4.3, giedahalan giellamodealla viiddideami. Ulbmiil lea oazzžut analysáhtora dovdat maiddái sátnehámiid mat dábałaččat eai adnojuvvo gillii gullevaš sátnin, muhto maid giellaoahpahalli sáhttá čállit.

3.2 Sátnehámis lemmai ja lemmas sátnehámiide

3.2.1 Sátnedárkkistanprögrámmiin genereret sátnehámiid (Art. III)


Maiddái iništálaoanádusaid ja loguid kásusmerken lea dábałaš meattáhustiipa, ja daid riektačállinnjuolggadusat sáhttet leat njulgestaga snugormastingeadgin cállái. Loatna-
sániid heiveheapmi čuovvu njuolggadusaid, muhto muhtun sániin leat dattetge nu mânga vejolaš čállinhámi ahnte dat eai sajáiduva čállingillii, geahča maiddái kapitltala 2.3.3.


Dan dihte Divvun-prográmma potentiálá lea gávñat ni-sátni-meattáhusaid, ja dat dahketge stuorimus oasi čálliid meattáhusain. Evaluerema boadus lea ahnte prográmma fuomášuhttá čállái 78 % meattáhusain, ja 82 %:s dan dáhpáhusain prográmma fállá čállái relevánta čállinhámi konteakstafálus, mii sísttisdoallá eanemusat vihta evttolusa. Ar-thkal almmostahttá maiddái ahnte prôgraßmmas leat buoridanmunit, nugo váilevaš sániid lasiheapmi ja evttolusaid válljema algoritma buorideapmi.

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


Divvun-prográmma vuorhua evttolusaid main lea unnimus divvungaska, muhto algoritmii gullá maiddái fonologalaš komponeanta, mainna sáhttá vuoruhit evttolusaid. Dáinnna prográmma vuorhua muhtun evttolusaid mat gullet meattáhusaide maid eatnigielagat dávjá dahket, ovdamearkka dihte $ie > ea$, mas lea divvungaska 2, ovddabeale evttolusat main leat divvungaska 1, nugo govvosis 3.1 mas Divvun evttoha filkkas hápmái fylkkas easkka guádín. Prográmmas ii gávdno njuolggadus vuoruhit $i > y$-rievdadusa, ja dan dihte dat ii evttoluvvõge ovdal eará vejolašvuodaid.
Govus 3.1: Divvunproggramma evttousat boastut čällon sániiide. Juohke evttouhussii ad-dojuvvo divvungaska. (Artihkkala III ovdamearkkat 10 ja 11.)


3.2.2 Sátnegirjjiiid riggodahttit morfologiijain (Art. IV ja V)

Dán kapihttalis giedahalan mo morfologalaš analysáhtorin sáhttá riggodahttit sátnegirji nu ahte geavaheddj beassá sojahuvvon sánis rivttes lemma-artihkkalii, ja mo sátnegirjeartihkkalii genereret sojahanparadigmaid.


VD lahkonanvuohki lea sátnegirjái FST:in genereret sátnegirjii lemmaid sátnehámiid, čoavddahámiid ja giellaoahpahalli konteavsttaid. Diedut wurkojuvvojit diehtovuddui man lingvisttalasi sisoallu lea stáhttalasi, muhto geavahanlakta dahká sisdoalu dynámalažžan geavaheddjái.

Lemmaid sátnehámit genererejuvvojit lemmalisttu gilkorlisttuin mas leat ovddalgihtii definerejuvvon hámit maid hálidit sátnegirjái, ja daid gaskkas lea ovdamearkka dihëte substantiiivva mängaidlogu akkusatiiva- ja genitiivahámitt (+N+Pl+Acc, +N+Pl+Gen). Juohke sátnehápmái genererejuvvo sátnegirjeartihkal mas leat diedut morfologalaš hámi birra ja jorgalus (govus 3.2). Dát sátnegirjeartihkal lea fas lijkejuvvon sání váldoartihkkalii, mas leat čoavddahámit ja vejolaččat ovdamearkacealkagat (govus 3.3). Čoavd-
3.2. SÁTNEHÁMIS LEMMAI JA LEMMAS SÁTNEHÁMIIDE

dahámít addet sáni morfonologalaš rievddademiid. Pronomeniidda hálidadeimmet ovdam-
mearkacealkagiid juohke sátnehápmái, eat duše lemmai, ja maiddái addit eará lemma
go FST lemma, vai sáhttá juohkit sojahanparadigmaid persovnna-logu mielde amaset
sojahanparadigmat šaddat menddo stoorrát ja kompleaksat giellaoahpahallái. Dan dihte
čáliimet daidda sátnegirjearthikkaliid, dan sajís go genereret daid. Seamma láhkai da-
gaimet *leat-* ja biختtalanvearbbba sojahanhámiide.

Govus 3.2: VD lohkanveahkkki: Geavaheaddji coahkkala sáni teavsttas, *giedaid*, masa lea
genererejuvon sátnegirjearthikal mas lea jorgalus ja morfologalaš analysa: pl. gen. dahje
pl. akk. Sátnegirjearthikkalis lea *Mer...*-liŋka (= eambbo) sáni vuoddohápmái *giehta.*
(Artihkkala IV govus 1.)

Govus 3.3: VD: *giehta*-lemma váldoartihkal, mas leat jorgalusat, substantiivva čoaad-
dahámít ja ovdamemerkacealkka. (Artihkkala IV govus 2.)
Såtnegirji och FST fertejte heivet oktii dego bátnejuvllat. Lemma ja satneeluohkká fertejte leat seammát vai lea vejolaš genereret sojahanhámiid. Vai såtnegirji galgá addit rivttes såtnegirjeartihkkala geavaheaddjái, de lei dárbu heivehit analysáhtora, siihe homonymaid ja mánggaidlogu leemaid dihte.


Mii rálkhadeimmet manjel maiddái lullisámi VD-veršuvmnja och válljíjìmet geavahit homonymijagilkoriid +Hom1 ja +Hom2 dalle go lemmain ii leat seammá jorgalus och dasa lassin çuvvot goappatlágan sojahananminstarid, nugo GOVLEDH+N+Hom1 = á hore (= gullat) och GOVLEDH+N+Hom2 = á hores (= gullot). Justa dán guovtti yearba homonymijia livččii leamash vejolaš çoavdit transitivitehtagilkoriin, danne go nubbi lea intransitiiva (+V+IV) ja nubbi lea transitiiva (+V+TV).


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


Go árvvoštallá dán čovdosa olles SIKORii, iige dušše ráddjejuvven frekevalsisttui, de oaidná ahnte adjektiivvät oamastangehčosiin leat davvisámegjelas hui hárve, ja eanaš dain sáhttet maiddái oazžut substantiiivaanalysa: buriidis, boarrássidset, nuoraidasamet, ipmilbalolaččaidis, ráhkásiiddán, vártnuhemiidat. SIKORis leat maiddái moadde kompa-

49
rerejuvvon hámi oktan oamastangehčosiin, muhto dušhe ovttia lemmas: buoremusaset, buoremusaideaset.


NDS-sátnegirjíis lea sierra modulan neahttalohkkái lakti girjemearka ja geavaheaddji sáhttá seamma láhkai go VD:in, logadettiin teavsttaid coahkkalit sáni mii sád-dejuvvo sátnegirjáí, mii fas fállá sátnehámi lemma ja jorgalusa (govus 3.4). Muhto NDS geavahanlavttas lea boallu mii sádé ohcansáni lemma SIKORa geavahanlaktii, ja geavaheadjái fálloujuvvjút buot korpusa cealkagat main lea ohccojuvvvon lemma. Govvsosa 3.5 sátnegirjeartihkkaliis lea okta cealkkaovdamearka loahpahít-lemmain, muhto Oza buot hámiid teavsttain-boaluin geavaheaddji oañju 7402 cealkaga dasa lassin.

Deháleamos lea go NDS-lahkonanvugiiin sáhttá jorgalit sáñiid mat eai leat sátnegirj-

---


Govus 3.4: NDS: Geavaheaddjii coahkkala sáni loahpahuvvo maid analysáhtor lemmatisere. Boadus lea loahpahit maid sátnegirji jorgala dárogillii.

ji gáldofilllas, nugo goallossániid, suorggádusaid ja sániid main lea laktapartihkal, ja nu leat cõvdojuvvon stuorimus váttisuovdat mat bohtet ovdan artihkkala IV evalueremis. Analysáhtora bokte sátnegirji maiddái gokčá buot oamastangeażushámiid.

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

Govus 3.5: NDS fallá ovdamearkacealkaka loahpahit-yearbbain, muhto Oza buot hámiid teavsttain-boaluin geavaheaddjii beassá maiddái njuolga SIKORa geavahanlaktii gos oazžu eanet cealkagiid mat sisttisdollet seamma lemma.
Sätnegirji geavaha deskriptiivvalaš analyzáhtora ja sáhttá dánin lágiin dovat maiddái sätnehámiid mat eai leat čállojuvvon norpma mielde. Leksikontransduseris dákkár hámit leat lasihuvvon lenne leet nortiiva hápmái. Yus ohcá kántuvrras, de lemmatiserin addá nortiiva lemma kantuva: kántuvrras kantuva+Err/Orth-a-á+N+Sg+Loc.

Artihkkala V evalueremis boahánt ovdan ahnte NDS lahkonanvuohki28 buorida sätnegirjii gokčanmearri hui ollu VD-lahkonanvuogi29 ektui. Davvisámegiela NDS áidnageardásaš sániid (= eng. ‘unique words’) gokčanmearri lea 79,7 %, ja VD gokčanmearri 57,8 %. Lullisámegielas leat gokčanmearit 67,0 % NDS:in ja 41,1 % VD:in. Lahkonanvugid iešvuodat lean buohtastahttán tabeallas 3.1.


Maŋŋelgo artihkal V čállojuvvui, de leat davvisámi NDS:i lasihuvvon ollu sätnebárat

28Artihkkalis dat ghóčuduvvo ‘FST-dictionary’.
29Artihkkalis dat ghóčuduvvo ‘wordform-dictionary’.
30analyser-dict-gt-desc-mobile hfst
3.2. SÁTNEHÁMIS LEMMAI JA LEMMAS SÁTNEHÁMIIDE

nu ahte sátnegirjjis leat dál 30 000 sámi ohcanání. NDS geavahanlavttas lea leamaš buoridanmunni suorggádusaid ovdanbuktimis, ja sátnegirjái leat lasihuvvon suorggádus-gilkorat ja muhtun infinihtiitavearbaid gilkorat. Ovdamearkka dihte sátni oadesta analyysa lea lemma oaddit oktan gilkoriguin +V+IV+Der/st+V+Ind+Prs+Sg3. Lemma jorgaluvvo dáregilleni sove ja suorggádusgilkor +Der/st jorgaluvvo à X litt (= eng. ‘to X a little’) ja dasa lassin ķilgejuvvo suorggádusa funkšuvdna: diminutiiva/subitiiva. FST analyysa gilkorat dahket vejolažžan genereret dárkilis dieduid suorggádusa semantihkas. Seamma láhkkäi lea dahkkon ovdamearkka dihte veearbaabessiivvain (govus 3.7). Sáni oadekeahttá analyses lea oaddit+V+IV+VAbess, ja gilkor +VAbess jorgaluvvo uten à X (= eng. ‘without X-ing’).

Govus 3.7: NDS ķilge geavaheaddjái veearbaabessiiva oadekeahttá mearkkašumi.

Dán kapihttala fáddá lea mo FST riggodahttá sátnegirjjid, go FST:in sáhttá lema-
tiseret addaga gávdan dihte sátnegirjeartihkkala, maiddái dalle go atta lea goallossát-
ni dahje suorggádus mii ii leat sátnegirjjis. Grammatihkkagilkorat addet geavaheaddjái lassedieduid grammatihkkalaš hámiid birra, muhto lea dárbu jorgalit daid nu ahte +VAbess šaddá uten à X. FST:in sáhttá genereret lemašaid jojanparadigmigned geavaheaddjái veahkin ja daidda hámiide maiddái genereret gielloaahppanpedagogalaš konteavsttaid

31Giellatekno programmerár Ciprian Gerstenberger lea leamaš guovddáš ohmoš sátnegirjjiid gáldofiillaid giežahallamis.
32Giellatekno programmerár Chiara Argese lea implementeren buoridemii.
vai lea álkit geavaheaddjái dichtit mo galgá geavahit juohke hámi. FST lemmatisere teakstkorpusa, ja dat dahká vejolažžan gávdnat cealkkaovdamearkkaid sátnehámi lema bokte.

<table>
<thead>
<tr>
<th>Iešvuohta</th>
<th>Vuosttaš digisánit</th>
<th>Neahttagisánit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mo dovída ohcansániid</td>
<td>FST genereren čaoavddahámít</td>
<td>FST analysa olles paradigmát</td>
</tr>
<tr>
<td>Morfologalaš veahkki</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Sátnehuksema čilgen</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Dovdá</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>oamastangehčosiid</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>dynamálaš goallossániid</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>dynamálaš suorggdádusaid</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>hámiid mat eai čuvo čállinnorpma</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Lullisámi stávenložžen</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Davvisámi stávenložžen</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>’Coahkkal teavsttas’-funkšuvdna</td>
<td>✓ eanaš programmain</td>
<td>✓ html online</td>
</tr>
<tr>
<td>Linkejuvvon SIKORii</td>
<td>–</td>
<td>✓</td>
</tr>
</tbody>
</table>

Tabealla 3.1: VD ja NDS sátnegirjiid iešvuodaiz buohtstahttin.

3.3 Giellamodealla gáržžideapmi

Go ráhkadit analysáhtora dainna vugin ahnte viežžá sátnehámiid teakstkorpusis, de oazžut dušše sániid mat gávdnojit teavsttain, ja nu mii dichtit ahnte analysáhtora sáníit duodaid leat geavahas. Muhto dát bargovuohki eaktuda hui stoorra teakstkorpusa mii gokčá buot šánperiid, ja erenoamáš stuora korpusa jus gielas leat ollu iešguđetlágan sojahanhámít. Sámegielagiid gielas leat eanet iešguđetlágan sátnehámít go sámegiela teakstkorpusiiin leat. Dat mearkkaša ahnte dákkár analysáhtor ii dovddaše buot sániid maid sámegielat olmmoš geavahivččii, juoga mii ovdamearkka dihte dagašii heajos sátne-divvuunprográmma. Sátnehamít mat eai galggaše leat apmasat eatnigielagiidda, nugo vearbahámít dajuđe ja boraide, eai gávdno SIKORis, geahča kapihttalis 2.1. Dan dihte mii leat välljen hukset transdusera (FST) ja dainna hukset sátnehámiid. Muhto FST
3.3.1 Eai buot sánit leat geavahusas

Kapihttalis 2.3.2 muitalin ahte 30 % davvisámi Giella-sme substantiiva- ja varbalemmain eai gávdno SIKORis. Dát lemmat leat dattetge duodaštuvvon leat giellamáilmmis dainna lágiin ahte dat gávdnojit dohkkehuvvon tearbmalisttuin, dahje daid vuoddu lea mo normerenmearrádusaid mielde heivehivččii dáru sáni dahje nama sámeğiela 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 i leat nu stuorra vejolašvuolhta gávdnat Kampen33-nama kásussojahemiid anárásgiela teavsttain, iige Heinola34-nama inessiivahám Heinolisnie lullisámi teavsttain. Muhto dalle go muhtun cállá lullisámegillii mátkereivve Suomas, de Divvun-prográmma veahkeha su cállit nama Heinolisnie norpma miede.

Sámegielaide gullet maiddái goallossánit, ja mearrusoassin sáhttet leat mánga sátneluohká sánit. Ii leat vejolaš leksikaliseret buot goallossániid go gielaide jámma ihtet odda goallosteamit. Analysáhtoris lea dan dihte dynámalaš goallosteapmi go analysáhtor ieš hukse goallossání go lea goallostanbálggis sáni mearrusoasis vuoddoassái, muhto dát ráhkada maiddái sániid mat eai leat geavahusas.


Maiddái suorggiduvvon substantiiivvat sáhttet leat miede dynámalaš goallosteamis. Dan dihte manná goallostanbálggis varbaaleksikonai, adjektiivaleksikon ja substantiiivalek-

33Oslo gávpopoassi.
34Unna gávpogaš Lulli-Suomas.

sikona gaskka. Diakrihtalaš leavggaid bokte (čilgejuvvo kapihttalis 3.3.3.3) addojuvvo gáržjidus das ahte goallosteapmi dálhpáhuuvvá dušše jus sikh meareusoassi ja vuoddoassi leat substantiivvat, ležžet dal lemmat dahlie suorggádusat, nugo bivdinvejolašvuolta:

bivdinvejolašvuolta

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

Goallostanbálgát mannet maiddái lohosánii dihto substantiivvaide ja adjektiivvaide, nugo guovttejuolggat:

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

Dynámalaš goallosteapmi ráhkada maiddái sáníid maid ií oktage geavahivčče semantihkalaš sivaid geažil, nugo govovsis 3.8 mas girjelávka lea vuogas sátni, muhto girjegirjii vejolaččat ií oro heivemin masage. FST dohkkeha maiddái sáníid main leat vaikko man ollu meareusoasit, ovdamearkka dihte girjelávkagirjelávka. Vaikko FST goallosta sáníid beroškealahdta semantihkas, de leat morfológalaš hámit ráddjejuvwon daidda mat gielas leat geavalusas. Deskriiptiiva FST diktá buot meareusoiid goallostit ovtaidlogu nominatiivvas ja genitiivvas, ja mánggaidlogu genitiivvas.
3.3.2 Giellamodealla gáržžideapmi norpma ektui


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

Bálgát main leat +Err-gilkorat, leat miele deskriptiivvalaš FST:s36, muhto váldojuvvojit eret dalle go ovttasta normatiivvalaš FST37, mainna generere normatiivvalaš sátnehámíid sátnérdkkistaprogámmii ja sátnegirjeartihkkiidi sojahannahadigmaide. Nugo namuhin kapihttalis 2.3.3, de ožžot 4,8 % SIKOR sániin dušše deskriptiiva +Err/Orth-analysa, muhto mángga sánis sáhttá leat siihe normatiiva ja +Err/Orth-analysa, ovdamearkka dihte lohka-sátni oaažžu analysa siihe substantiivan (lohka+N+Sg+Nom ja

35Norbma vuodduhuvvo ovdде Sámi giellärádi ja Sámi giellalávdegotti ja dálá Sámi Giellagáldu mearrádusaide, cáällingela prinsihpaide generaliseremii, prentejuvvon sátnegirjjide ja grammattikaide (njálmmálaš diehtu Thomas Ommas).
36analysger-gt-desc.xfstd ja analysger-gt-desc.hfst
37analysger-gt-norm.xfstd ja analysger-gt-norm.hfst

Sáhttá maiddái stivret mearusooasi hämi vuoddooasi gilkora bokte, ovdamearkkka dihte +CmpN/SgPIGenLeft gáibida ah te mearusooassi galgá leat mänggaidlogu genitiivvas.

### 3.3.3 Badjelmearálaš genererema ráddjen

Kapihttalis 3.3.2 lei sáhka mo leksikontransduseris sáhttá earuhit deskriptiiva ja normatiiva hámiid vai sáhttá ráhkadit sihke deskriptiiva ja normatiiva FST. Dán kapihttalis lea sáhka no garvit sátnehámiid mat rievtti mielde leat riekta čállojuvvon morfologalaš njualggadusaid mielde, muhto sátnehámít eai geavahuwvo sámegielas, ja dahket geavahedjiidprográmmaid fuonibun.

Normatiiva FST vuodul sátnedárkkistanprográmma merke sátnehámiid mat eai leat čállojuvvvon norpma mielde. Muhtun čállinmeattáhusaid prógramma ií gávmina danne go sátnehámí oaažžu eará normatiiva analyza – ja lea duohtasátnemeattáhus. Jus FST generere sátnehámiid mat eai geavahuwvo gielas, de dát hámit sáhttet dahkat eanet čállin-
meattáhuses duohtasátneattáhussan go dat mii livččii riekta, ja dat bilida prográmma beaktilvuoda. Sáhttá dadjat ah te dákkár hámit čihket ii-sátni-meattáhuses.

Artihkkalis III bodii ovdan ah te Divvun-prográmma gávdná 78 % dutkanmateriála čállinmeattáhusain, ja buot čállinmeattáhusat leat ii-sátni-meattáhusat. Dat mearkkaša ah te 22 % čállinmeattáhusain leat duohtasátneattáhussat, maid Divvun-prográmma ii sáhte gávdnat. Jus sáhttášii unnidit badjelmearálaš genererema, de eanet čállinmeattáhussat šattašedje ii-sátni-meattáhussan. 82 % dáhpáhusain main Divvun-prográmma gávdná čállinmeattáhusa, prográmma máhttá evtttohit relevánta čállinhámi. Maiddái dán logu sáhttášii buoridit jus ráddješii badjelmearálaš genererema.

Čuovvovaš vuollekapihttaliin attán muhtun davvisánegiela ovdamearkkaid mat leat áigeguovdilat sátnedárkkistanprográmma ektui, danne go dát badjelmearálaš hámit sulastahttet eará morfologalaš hámiid. Vuos geahččat mo sáhttá ráddjet dynamalaš goallosteami posíšuvdna-gilkoríi-guin. Ja de čilgen mo leksikontransduseris sáhttá ráddjet oamastangehčosiid lasiheami amaset čiehkot lokatiivahámiid ii-sátni-meattáhuses, ja kapsihtaliin dan manjel attán guokte ovdamearkka das ah te maiddái vearbagenererema ráddjemiin sáhttášii buoridit Divvun-prográmma beaktilvuoda.

Kapihtaliin 3.3.3.6 ja 3.3.3.7 suokkardalan finihtta vearbahámiid ráddjema, in dušše sátnedárkkistanprográmma dihte, muhto maiddái prográmmaide mat fället sojahanparadigmaid geavaheddjüide, nugo NDS.

### 3.3.3.1 Substantiivvaid posišuvnnat goallosteamis

Lea giki go normatiiva FST dohkkeha dábálaš čállinmeattáhusa dynamalaš goallossátnin. Sátni gaskkavakhku sáhttá leat gaskkavakhku-sáni boaštuhápmi, muhto maiddái sániid gaskka ja vakkku goallosteapmi. Semantihkalaš sivaid geažil dát goallosteapmi ii leat vuorddehahtti, ja mii háliidit baicce ah te sátnedárkkistanprográmma galgá hilgut sáni gaskkavakhku ja evtttohit gaskkavakhku dan sajis.

Dan sáhttá dahkat go stivre substantiivvaid posišuvnna goallosteamis. Dasa geavahit +CmpNP-gilkoriid, mas P mearkkaša posišuvnna: +CmpNP/None (lemma ii sáhte goallostit), +CmpNP/First (lemma sáhttá dušše goallostit mearusoassin) ja +CmpNP/Last (lemma sáhttá dušše goallostit vuoddoassin). Go lasiha +CmpNP/Last-gilkora gask-
ka-lemmait, de normatiiva FST\textsuperscript{39} hilgu buot dynámalaš goallonstemuid main lea gaskka mearusoassin, earret eará gaskkavahkku, ja dalle lea dárba leksikaliseret vejolaš sániid main lea gaskka mearusoassin.

3.3.3.2 Oanehis namaid goallonsteapmi

NDS geavaha deskriptiivvalaš FST, mas leat unnit gáržžideamit goallonsteamis go normatiiva FST:s. Muhtun analysat mat eai daga vâttisvuoaid korpusanalysii danne go dat jávket disambigueremis, leat unohasat sätnegirjji oktavuodas. Ovdamearkan dás lea oanehis namaid goallonsteapmi. Davvisámegielas eai leat ovttastávval substantiivvat, earret moadde odda loatnasáni, muhto Giella-smX oktasaš mammafiillas leat ollu ovttastávval namat mat sáhttet addit sâvakeahtes analysaid. Ohcansátni Buolbmát sähttá oazžut analysa mas lea Bu mearusoassin ja olbmat vuodduoassin, mugo vuolábeale deskriptiiva analysas:

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

Gielalaččat goallonstananalysa livččii vejolaš, muhto oalle marginála hápmi, ja cállit dáid-dášii lasihit goalloSSázu, ja cállit Bu-olbmát, sání guovttečilggolašvuoda dihte. Sátnegirjái dákkár analysat eai leat sávahahttit. Ćoavddus lei bidjat buot ovttastávval namaid sierra nammaleksikonii man bálías lea diakrihtalaš leavga mií ii divtte sáni goallostit eará sániide goalloSSázu haga. Diakrihtalaš leavgga geavahus čilgejuvvo kapihttalis 3.3.3.3.

3.3.3.3 Substantiivvat ja adjektiivvat oamastangehčosiiguin


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

\textsuperscript{39}\textsuperscript{+}CmpNP-gilkorat doibmet dušše analyse-gt-norm.hfst:s
substantiivva nominatiiva-hápmái measta duše go lea fuolkevuoadanamahus dahje sul-

(1) Sii hálidedje gohčodit su Sakarjan āhčis mielde, muhto eadnis dajai ... (SIKOR: Odda testameanta)

Genererema sáhhtá ráddjet joatkkalesikonaid bokte: Livččii vejolaš ráhkadit sierra joatkkaleksikaonrudni-lemmai, vai dán lemma bálgás ii leat +Nom+PxSg3, eaige eará nom-
inatiiva goalmnát persovdna-hámít (Px = eng. ‘Possessive suffix’ = oamastangeažus).

Buoret čoavddus lea geavahit diakrihtalaš leavggaid, mat leat vuogas reaidun diekkár gáržžidemiide danne go dat eai leat čadnjovvon dihto joatkkalesikonii ja dan sojaham-
mállii, muhto leavggaid bokte sáhhtá ráddjet vaikko makkár lemma leksikonas. Diakri-
htalaš leavggaat eai oidno analysas, muhto daiguin addit muhtun árvvu dihto lemmamaide
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 leavgga. NOMSG3-leavgga lasihuvvo leksikonii
mas leat lemmat eadni ja āhčči. Bálgás mii addá oamastangehčosa nominatiivahápmái, lea
gáržžidus mii gáibida NOMSG3-leavgga, ja dan dihte bálggis ii doaimma buot lemmameide.
Transduser generere sániid eadni, rudni, āhčči, gáhčči, eadnis, āhččis, muhto ii sániid
rudnis, gáhččis.

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

⁴⁰Sullasaž substantiiva lea Odda testameanttas oahpaheaddjij, ovdamearkka dihte cealkagis Máhltájeadd-
djii leat stuurit go oahpaheaddjij, ii ge bálvaleaddjii stuurit go isidis.
sushápmi livččii *rutnis*, nugo *rudni+N+Sg+Loc:rutnis*. SIKORis lea okta ovdamearka das:

\(2\) *Juohkeolbmovuoigatvuodain* sáhttá oaggut *rutnis* [rutnis] dahje geasset. (SIKOR: Lundui.fi)

Danne lea dehálaš ráddjet genererema. Min geavatlaš čoavddus lea ahte substantiivivaleksikon juhkkojuvvo vuoleleksikonaide da mielde man ollu oamastangežushámit galget genererejuvvo (tabealla 3.2). Buot lemmat maidda galgá NomSG3-leavga, čohkkejuvvo-jit sierra substantiivivaleksikonii, man namma lea NOUNPXiN, ja leavga biddjojuvvo bál-gái mii manná dán leksikonii, semma láhkai go govvosis 3.9. Daíma lágiin leksikontransduseris leat maáddái sánt mat eai obanassiige oacčo oamastangehčosiid, NOUNNoPX, go bálgáin mii manná oamastangežusleksikonaide gáibiduvvo Px-leavga. Stuorimus oassi substantiivivaleksikonii lemmain gullet NOUNPx:i, mii lea ovdaválljuvvojuvvo leksikon odda leemmaide. Muhto lingvista geavaha iežas gielladovddu ja dasa lassin SIKORa veahkkin, go sirdá leemaid NOUNNoPx:i, ja son dakhá dan erenomážit dalle go hámít main lea oamastangežus, sáhttet čiehkat dábálaš čállinmeattáhusaid. Adjektivivaleksikonias lea fas nuppelágan dilli, go dušše guoktenuppelot adjektivalemma gullet Lexicon AdjectitivePx:i, namalassii adjektiivvat41 maid leat gávdnan SIKORis ja eará teavsttain oktan oamastangehčosiin.

<table>
<thead>
<tr>
<th>Substantiivivaleksikonat</th>
<th>Oamastangehčosiid lasiheapmi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexicon NOUNPxiN</td>
<td>buot hámít</td>
</tr>
<tr>
<td>Lexicon NOUNPx</td>
<td>buot hámít earret +Nom+PxSg3/PxDu3/PxPl3</td>
</tr>
<tr>
<td>Lexicon NOUNNoPx</td>
<td>eai lasihuvvo hámít oamastangehčosiiguiin</td>
</tr>
</tbody>
</table>

Tabealla 3.2: Substantiivvaid leksikonat stivrejit oamastangehčosiid lasiheami.

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

Substantiiviivvat oktan oamastangehčosiin dahket stuorra paradigma mas leat 105 sealla

---

41*nuorra, buorre, uhecci, unni, boaris, ùpnilbalolaš, rähkis, vistelágas, láhkásaš, vårtnuheapme, ovdečs, suohtas*
ja 79 iešgudetlāgān hámi. Dān kapihttalīs lean geahččan +Nom+PxSg3-hāmiid, muhto maiddāi eara sojaheamit eai leat geavahusas buot substantiivvaiɡuin. Oamastangeažus-geavahusa berrešii guorahallat eanet, ja dan vuodul lasihit gāržžidusaɪ FST:i.

3.3.3.4 Vearbbaid imperatiivva biehttalanhāpmi

Davvisámeğielas lea vejolaš rāhkidat guovttelāgān biehttalanhāmiid, main nubbi geavahuvvo dušše goalmmát persovnnaid imperatiiva biehttaleami váldovearban. Imperatiivva biehttalanhāpmi lea geavahuvvov boares oskuteavsttain, nugo biibbalis (Nickel ja Sama-mallahti, 2011, s. 61). Sojahanhāmi sahttá árvvoštallat leat jávkan gielas, go oddaseamos davvisámeğiela grammatihkas dākkár hápmi iī namuhuvvo imperatiiva biehttalanhāmiid gaskkas (Svomi, 2015, 107). Dattetge imperatiivva biehttalanhāpmi gāvdno SIKORis Odda testameanttas, los hápmi lea geavahuvvon oktiibuot logi vearbalemmas. Go dasa lassin geavahan davvisámeğiela biibbalveršuvnna mii lea interneahtas42, de gāvnnan oktiibuot 30 vearbalemma. Dain leat 18 bārrastávvalvearbbat43, nugo cealkagiin (3) ja (4).

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

(4) Allos oktage behtto iežas! (SIKOR: Odda testameanta)


42https://www.bibel.no/Nettbibelen (23.11.2017)
43addit, atnit, beahttit, boahtit, bāhcit, dahkat, goargŋut, gullat, heahpanaddat, jearrat, jāhkkit, oaidnit, sivahallat, sārdnut, rābāghahtit, vuolgit, ássat, sāddat
44ármihit, čājehit, dorvvastit, illudit, leat, rāmidit, suorganit, váivvidit, veahkehit
45gillát, loktet, rāpmot
3.3.3.5 Vearbagenitiiva


Vearbagenitiiva sáhhttá maiddái geavahuvvot reduplikatiiva vearbagenitiivan, mas guokte ideanttalaš vearbagenitiivva leat cealkagis maŋŋálaga. Dakkár geavaheamis leat seman-

46analyser-gt-norm.hfst
3.3. GIELLAMODEALLA GÁRŻŻIDEAPMI


Barggadettiin Giella-sme disambiguáhtoriin lean SIKORis ja eará teavsttain, maiddái grammatihkain ja bajábealde namuhuvvon artihkaiin, gávdnan oktiiboot 79 veerbba main lea vearbagenitiivahápmi, buot bárrastávvalvearbbat47 earret guoke ite rárahisstávvalvearbbba48 ja guote kontrakšuvdnavearbbba49.

Vearbagenitiivahápmi lea homonyma veerbba preesanssa biehttalanhámiin earret bárrastávvalvearbbain main lea -ut dahje -it loahpas infinitiivvas, ja nu sähttá leat duhohtsátne-meatáltahasaid áidna normatiiva analysa.

Čuovvovaš golemma cealkkaovdamearkkas lea biehttalanhámi dahje finihttahápmi mii ii leat cálllojuvvon norpma mielde, ja áidna normatiiva analysa lea vearbagenitiiva, mii dainna lágiin čiehká ii-sáti-meattáltahas. Cealkagis (8) lea cálllojuvvon vearbagenitiiva oacchu dan sajis go biehttalanhámi normatiiva hápmi oacch. Cealkagiin (9), (10) ja (11) lea cálllojuvvon oacchu, sikh, sáhti finihtta vearbbahámiid sajis.

(8)  Ii oktage oacchu [oacch] vuolil 7850 ruvdnosaš lasáhusa. (SIKOR: Ávvir)

(9)  Ja juste das oacch [oacch] ge Sámediggi, enanoamážit Finnmárkkus, olu spíža. (SIKOR: Ávvir)

(10)  ... dušše dan dihte go dát lea dat manjimuš háve go dát cájálmas cájehuvvo, muitala Siri Broch Johansen dan botta go sikh [sikk] gatnjaliid. (SIKOR: Ávvir)

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

---

47bassit bárdnat bealkiti biškut boahitt borrar bussat cealkit civkit cárvut cállit čírrput čírrput čióogit čúorvut doapmat doorvat jápput galát gáňččat gálit gaygrt gauvut gíját gíját gíját gúlát quoddit hêahpånnaddat hojat holvat huikt hupmat jápmít johtit juorbut jorjut juoogit láolvut luovmait murdit njákkat njoammút njårgut oädít oädkut oädnit ochat oorrut râhtat reášít rieččit rieèdèt rieèdèt rohkait ruohtatts roussat sàddat sîkkelašstí skeàkit skierbmot sobbat speàjìt stëmmput subbat suojbat sîlvat sëlahat šnåjgít vázztit vihcat vuodjat vuodjat vuõhčút vuõshat

48doarggístit garrudit

49čohkkát làddjet
tiválain, vaikko balai ah-te čoaska dálki sáhti [sáhtii] muddet veahá festivála.
(SIKOR: Ávvir)

Vearbagenitiivahámmit leat sátnedárkkisteami giksin danne go dain ii leat sierra geožus, ja ut- ja it-vearbbain dušše soggevokála earuha daid biehttalanhámis. Vearbagenitiiva sáhttá maiidáí čiehkhat ah-te konsonántanguovddáš lea čállojuvvon geahnohis dásis dan sajís go gievrras dásis, nugo cealkagiiin (9) ja (10). it-vearbbaid preterihtta Sg3-hámis lea seammá sullasaš čuolbma, go galgá čállojuvvot i loahpas, muhto oarjesuopmaniid assimilašuvnna dihte lea dábálaš čállit dušše i. Vearbagenitiivahápmi sáhttá čiehkhat maiidáí dákkrár ii-sántni-meeattáhusa.

Vearbagenitiiva lea áidna normatiiva analysa 981 sánís SIKORis, ja 450 dáhpáhusas dan verbii ii leat dokumenterejuvvon vearbagenitiivageavahuus, ja buot dát dáhpáhusat leat rievtti miede eará sánit mat leat boastut čállojuvvon. Dát guoská 100 iešgudetláagan verbii. Gáržžidus vearbagenitiivabálvíín buoridivččii sihke sátnedárkkistanprogramma ja korpusanalysa.

3.3.3.6 Finihtta vearbahámit


Vuosttaš joavku leat vearbbat mat geavahuvvojit dušše dálkki birra. Dálkevearbbat spiehkastit dábálaš vearbbaid sojahanparadigmias, ovdamearkka dihte vearba arvit. Giella-sme leksikontransduceris diekkár vearbbat mannet sierra joatkkaaleksikonaide mat dušše lasihit +Sg3 finihtta hápmin: arvá, arvvii. FST generere buot eará vearbbaid ovcci persovnna-logu hámiid guovtti áiggis.

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

Goalmmát joavkkus leat vearbbat mat mualit erenoamážit elliid láhttemiid birra, nugo goarjut, ja dan vuosttaš ja nuppi persovnna hámit eai gávdno teakstakorpusis. Dáid vearbbaid gohčodan dán teavsttas elliid láhtttenvearban. Dákkár vearbbat sáhttet leat várálaččat ráddjet FST:s, danne go čálli sáhtášii geavahit dakkár hámiid máidnasis, jus mun-persovdna lea eall. Ovdamearkan dasa lea go állat muitala iéžas fearániid birra cealkagis (12) ja geavaha civken-sáni. Seamma vearbbain karakteriserejuvvo olbmuid jietnadeapmi humoristtalaččat cealkagis (13). Dáid ovdamearkkaid vuodul in sihkoše elliid láhtttenvearbbaid vuosttaš ja nuppi persovnna 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)

Dálkevearbbaid morfologalaš genereren lea ráddjeuvvon Giella-sme FST:s, ja dušše guok-te finihtta hámi leat mickle sojahanparadigmas maid NDS-sátnegirji presentere geavahed-djiiide. Muho lea jearaldat berrešiigo sátnegirji ekspernet giellaálggahalliiid dáhpáhuv-vat-vearbbaid hámiide mat eai leat geavahusas ja elliid-láhtttenvearbbaid hámiide mat illá leat geavahusas?

Lean válljen ahte NDS ii berreše ekspernet geavaheddjiiid dáhpáhuvvat- ja elliid-láh- ttenvearbbaid vuosttaš ja nuppi persovdnahámiide. Ėovvedus dasa lea ráddjet genererema sátnegirji gáldofiillas, nu ahite dákkár vearbbain leat dušše goalmmát persovnna hámit sojahanparadigmas mii addojuvvo geavaveaddjái. Ráddjen dahkko sátnegirji gáldofiillas attribuhttain context="dat". Sojahanparadigmas atribuhtta addá dat-pronomena konteakstan, dan sajis go son: dat goarjú. Maiddái giellaahppanprográmmia genereremis, geahča kapihttala 3.4.1, lea gáldofiillain ráddjenvejolašvuolta, nugo čuovvovaš attribuhttaárvu oaiivvilduvvon vearbbaidaide main galget genererejuvvo dušše goalmmát-
persovdnahámít ja daidda lassin moadde infihtiitahámí:

*gen_only*="+Inf,+Sg3,+Pl3,+ConNeg,+PrfPrc".\(^{50}\)

### 3.3.3.7 Finihtta vearbasojahanparadigmat eará geahččanguovllus

Giella-sme leksikontransduseris dálkevearbar mat mannet joatkkaleksikonaide maid sojahann-bálgát leat räädjejyuvon. Muđui buot vearbalemmat ožžot buot sojahanhámiid mat leat grammatihkkagirjjiid tabeallain, main leat +Sg1-ja +Sg2-hámít bajimuras ja buot ovcci persovnna-logu-hámít orrot leamen seamm dehálaččat. SIKOR davvisámi teavsttat ad-det nuppelágan gova finihtta vearbb sojahahanparadigmas. Juohke šánneris +Sg3-hámít dominerejit, 48,8 % rájes (oskuteavsttain) gitta 69,6 % rádjai (juridiikkalaš teavsttain).

<table>
<thead>
<tr>
<th>1. ja 2. persovnnaid hámit</th>
<th>3. persovnna hámit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sihkkaris hámit</td>
<td>Maid dáit hámit (buot hámit)</td>
</tr>
<tr>
<td>+Prs+Sg2, +Prt+Pl1</td>
<td>+Prt+Sg2</td>
</tr>
<tr>
<td>+Prt+Pl2, +Prs+Pl2</td>
<td>+Prs+Pl1</td>
</tr>
<tr>
<td>+Prt+Du1, +Prt+Du2</td>
<td>+Prs+Du1</td>
</tr>
</tbody>
</table>

| 3321 = 47 % | 4192 = 59 % | 5992 = 84 % | 6694 = 94 % |

Tabealla 3.3: Vearbalemmat main leat finihtta vearbahámít SIKORis, N = 7120. 94 % dain vearbalemmmain leat SIKORis 3. persovdnahámít.

Teavsttaid sisdoallu diedusge váikkuha vearbbaid sojahheapmái. Čáppagirjíaalašvuoda teavsttain lea 67,8 % finihtta vearbbain +Sg3-hámít, go teavsttaid čálit leat válljen čál-lit goalmmát persovnnas. Buoret gova oażžut go geahččat vearbalemmaid juohkáseami (tabealla 3.3). SIKOR teavsttain leat 7120 iešgudetlágan vearbalemma finihtta hámís, ja dain dušše 47 % leat 1. ja 2. persovnnaid sihkkaris hámit. Eahpesihkkaris hámit leat dákkárat mat leat homonymat juogo 3. persovnna hámiin dahje infinitiiivvain, geahča tabealla 3.5. Go válddán daid hámiid mieldel, de 59 % vearbalemmmain leat 1. ja 2. persovnnaid hámit.

\(^{50}\)Maiddái substantiivvain lea sătnegirjji gáldoviillas ráddjenvejolásvuoluhta. Substantiivvat mat dábaalččat eai geavahuuvvo mánggaáillogu hámis, eai genererejuvvo mánggaáillogus, vaikko hámpi gávdno analyssáhtiit. Dát stüvrevjuvvo sătnegirjji gáldoviillas attribuhttain seamma vuogi mieldle go vearbbain: *gen_only*="Sg".

68
Tabeallas 3.3 84 % veearablemain leat +Sg3 finihtta hámit. Olles 94 % veearablemain leat 3. persovdnahámit go váiddán mierle +Pl3-hámiiid, mat eai leat sikhkaris hámiid gaskkas. Vaikko maiddái dáiid loguid báidná teavsttaid sisdoallu, de dat doarju gova das ahle buot veerbbat geavahuvvojité 3. persovnnas, muhto dušše oassi veerbbain geavahuvvojité 1. dahje 2. persovnnas. Dán juohkáseami vuodul sáhttá veerbbaiide jurddašit earálagan paradigmastruktuurvrra, mií vuorúhá 3. persovnná veerbaarhámiiid, ja dušše oassi veerbbain ožžot 1. ja 2. persovdnahámiid.

Dát boadus heive bures Janda ja Nesset (2010) evttolhussíi goovditét sojahanparadigmáid radiálá kategoriijan mas lea guovddáš prototiípa mas lea gullevásvohtta lahtuide mat leat seamma kategoriijas muhto ravddabut. Ruoššagiela vuodul soai rakhadeaba dákkár vurdujuvven huksehusa, mas leat prototiíppalaččamus lahttu gurut bealde:

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


<table>
<thead>
<tr>
<th>Vearbaleksikonat</th>
<th>Finihtta hámit</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEXICON WEATHER-VERB</td>
<td>+Sg3</td>
</tr>
<tr>
<td>LEXICON HUMAN-SUBJECT-VERB</td>
<td>+Sg1, +Sg2, +Sg3</td>
</tr>
<tr>
<td></td>
<td>+Du1, +Du2, +Du3</td>
</tr>
<tr>
<td></td>
<td>+Pl1, +Pl2, +Pl3</td>
</tr>
<tr>
<td>LEXICON DEFAULT-VERB</td>
<td>+Sg3, +Pl3</td>
</tr>
</tbody>
</table>

Tabealla 3.4: Leksikontransduseris: Dánina lágidemiin dušše dihto veerbbat oččošedje buot finihtta sojahanhámiiid.

<table>
<thead>
<tr>
<th>Ovda-nemearka</th>
<th>1. ja 2. persovnnaid hámit</th>
<th>Eará finihtta hámit</th>
<th>Infinihtta hámit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bst</strong></td>
<td>dáhpáhuva +Imprt+Sg2</td>
<td>+Prs+Pl3</td>
<td>Prs+ConNeg</td>
</tr>
<tr>
<td></td>
<td>dáhpáhuuvet +Prt+Sg2, +Imprt+Pl2</td>
<td>+Prt+Pl3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>dáhpáhuuve +Prs+Du1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bhst</strong></td>
<td>gánnáht +Imprt+Sg2</td>
<td>+Prs+Prt+Sg1</td>
<td>PrfPrc +Prt+ConNeg</td>
</tr>
<tr>
<td></td>
<td>?gánnáhat +Prs+Pl1, +Prs+Sg2</td>
<td>+Prs+Pl3</td>
<td>Inf</td>
</tr>
<tr>
<td></td>
<td>gánnáhan +Prs+Sg1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gánnáhit +Prs+Pl1, +Prt+Sg2</td>
<td>+Prs+Pl3</td>
<td>Inf</td>
</tr>
<tr>
<td><strong>Ktr</strong></td>
<td>savvo +Imprt+Sg2</td>
<td>+Prs+Sg3</td>
<td>VGen, +Prs+ConNeg</td>
</tr>
<tr>
<td></td>
<td>savvot +Prs+Sg2, +Prs+Pl1</td>
<td></td>
<td>Inf</td>
</tr>
</tbody>
</table>

Tabealla 3.5: Giella-sme: Vearbba dábáleamos homonymiijat. Tabeallas leat giellamodealla analysat, ja 1. ja 2. persovnnaid hámit eai gávdno SIKORis. Sátnehápmi gánnáhat gávdno oktii SIKORis, čállinmettåhussan gánnáha sajis cealkagis: ... das leat nu mángga bàsttuwuoða ahte gánnáhat [gánnáha] yeehččat... (Åvvir). Bst = bárrastávvalvearba, Bhst = bárahisstávvalvearba, Ktr = kontrákta vearb.
hámít livčče morfologalaš analysan. Vearbba biehttalanhámí analyxs lea álki válljet dahje sikhkut danne go dat gáibida ahle biehttalanvearba lea hámi gurut beadle. Dušše pre-seannsa +Pl3 ja infinitiivahámíid lea vättis disambigueret nugo bajilčállagiin, čuoggdáidlist-tuin ja cealkagiin main leat eanet oalgecealkagat.

3.3.4 Giellamodealla gáržžideami čoahkkáigeassu

Áidna realisttalaš vuohki sámegielaide råhkadit analysáhtora, lea luks FST, mii lea hui fámolaš reaidu sihke analyseret ja genereret sátnehámíid. Teakstakorpusis eat gávtnaše buot sániiid, muhto FST:s lea fas badjelmearálšaš genereren čuolbman: Mii genereret sániiid mat eai gávdno.

Badjelmearálšaš genereren lea guovttelágan. FST generere morfologalaččat dohkálšaš sáníid mat eai leat dán rádjai geavahuuvvon korpusis. Hárve báikenamaid, vierissáníid dahje odda dohkkehuvvon tearpmaid sátnehámít leat gergosat geavahuvvot, go olmmoš čállá teavsttaid odda domeainain.


Grammatihkkagirjjuid sojahanparadigmat addet sátneluohká maksimála paradigma, ja luksedettiin FST dát sojaheamit addojuvvojtit dábálaččat buot sátneluohká lemmaide, go mihttomearrin lea gokčat teakstakorpusa sáníid nu bures go vejolaš.

3.4 Giellamodealla viiddideapmi

Árbevirolaš CALL (engl. ‘Computer-Assisted Language Learning’ = giellaahppan dihtora vehkiin) vuogádagat fáillet hárjehusaid nugo mánggaválljema (= eng. ‘multiple choice’), sániiid dahje cealkagiid oktiibidjama, sániiid dahje govaid coahkkaleami ja eaŋkilsániiid deavdima. Máhcahat muitala dábálaččat dušše leago vástádus seammá go čoavddavástá dus. ICALL (engl. ‘Intelligent CALL’ = jierbmás CALL) geavaha giellateknologalaš analysáhtora, ja daïmna lágni máňtá sihke genereret bargobihtáid ja analyseret geavahaaddji vástádusa ja nu dakkidade addit máhcahaga das maid geavahaaddji lea cállán.

Dán kapihttalis lea sáhka heivehit FST nu ahte dat sáhttá addit metalingvisttalaš máhcahaga giellaohpahallái, sihke morfologalaš ja syntávssalaš máhcahaga. Kapihttalis 3.4.1 válddahalan vuogádaga mii generere morfologalaš hárjehusaid geavahaaddjái, FST:in oktan gáldofiillaiguin mat leat XML-formáhtas. Dát vuollekapihtal oktan artiikkaliin VI lea mielde nákkosgirjjis maiddái danne go dat čilge vuogádaga mii lea vuoddun čuovvovaš guovtti vuollekapihttalii oktan artiikkaliiguin VII ja VIII. Vuogádagas geavahaaddjí oazžu metalingvisttalaš meattáhusmáhcahaga, mat leat vuodduduvvon sátnemámiid morfologalaš proseassade. Kapihttalis 3.4.2 cájehan mo sáhttá viiddidit giellamodealla nu ahte lea vejolaš addit metalingvisttalaš máhcahaga maiddái dasa maid geavahaaddji lea cállán, ja kapihttalis 3.4.3 viiddidan vuogádaga maiddái addit metalingvisttalaš máhcahaga giellaohpahalli syntávssas.

Čuovvovaš golmma vuollekapihttal reaidduin ii leat ulbmil ahte buot sánit gielas galget leat mielde, nugo ovddit kapihtiin, muhto reáddude lea válljejuvvon sátnevuoerká ja morfologijja mat leat relevánttat giellaohpahallái. Kapihttalis 3.4.1 leat heiveheamit vuosttažettiin vuogádaga gáldofiillain mat leat XML-formáhtas, kapihttalis 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 geavalhuvvon huksét guokte ICALL vuogádaga morfologiijaohphahallamii. Go olmmoš oahphahallá sámegiela, de lea dárbbašlaš hárjehallat ollu grammatihka, ja prográmmmat ledje jurddášuvvon luohkkálanja oahphahussii lassin.


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

Govus 3.10: Rivttes vástádus lea mánggaidlogu akkusatiivahápmi guliiid. Máhcahat addá dieduid sáni morfologalaš proseassaid birra mat gusket máddagii, dássemolsumii, diftoŋŋanjuolgamii ja gehčosii: "guolli lea bárrastávval sátni mii galgá leat geahnhohis dásis. Muitte ahte diftoŋŋa njuolgá id-gehčosa dihte." (Artihkkala VI govus 3.)

Vuogádagas leat golbma dehálaš oasi: leksikon-, hárjehus- ja máhcahatfiillat, mat buot leat XML-formáhtas. Leksikonfiillain leat vállyjeuvvon lemmat (1200 substantiivva, 750 vearbba, 300 adjektiivva ja dasa lassin muhtun pronomenat, lohkosánit ja advearb-
Govus 3.11: QA-hárjehusa templáhta. (Govus bohta artihkkalis VI.)

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


Govvosa 3.11 hárjehusminsttar generere QA-hárjehusaid nugo "Maid doai bargga-śeidde luomus? Luomus moai ... [yearbb biñifiivahápmi]." ACTIVITY-seahtas leat 87 yearbbba maid lean meroštallan heivvolážžan, ja ovttas iešgudetlágan pronomeniiin, hárjehusminsttar generereśii oktiiboot 783 iešgudetlágan hárjehusa. Hárjehusaid evalu-remis (340 soaittáhagas välljejuvvon QA-bára) adde evaluerejaddjit juohke hárjehussii
3.4. GIELLAMODEALLA VIIDDIDEAPMI

gaskamearálačcat 2,9 čuoggá rivttes grammaathi, 2,8 dohkálaš semantihhkak ja 2,9 peda-gogalaš heivvolašvuodas (3 lea buoremus čuokkesloholku).

Vuogádat buohtastahlättä geavaheaddji vástádusa sátnehámiin maid FST lea genereren grammaathi, sánis mii lea definerejuvvon rivttes vástádussan. Lea eanet go okta rivttes vástádus, jus FST lea genereren variánttaitaid. Vuogádat sáhttá dohkkehit čállinhámiid main eai leat sámegiela sierra bustávat jus geavaha stávenložžema (govus 3.6).

Vuogádaga meattáhusmáhcahagas ii leat alladásat vuorróváikkuhkus, go vuogádat registere dušše ahnte geavaheaddji lea čállán dohkálaš vástádusa dahje boasstuvástádusa. Dattetge geavaheaddji oazžu relevánta meattáhusmáhcahaga jus vástádus ii dohkkehuvo. Meattáhusmáhcahat lea huksejuvven dinain lehto buot morfologalaš proseeassat leat listejuvvon, ja dat fas kombinerejuvvoj guvttiin paramehteriiin, mat leat hárjehustiipa ja lemma iešvuođat mat leat relevánttát máhcahahkkii. Lemma iešvuođat leat mättatiipa, difonoŋŋa, konsonánta, vokálarádji ja soggevokála, ja diedut leat vurkejuvvon leksikonfiillas XML-elemeanti attribuhttan, nugo guolli-substantiiviivas:

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

Gouvosa 3.10 máhcahat sisttisdoallá 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 difonoŋŋa galgá njuolgat (eng. ‘diphthong simplification’), jus lemma XML-elemeanttas lea merkejuvvon difonoŋŋa ja soggevokála i (3). Mánggaidlogu akkusatiivva geazžus (eng. ‘suffix’) lea id (4).

Dát vuogádat lea beaktel ráhkadit morfologalaš hárjehusaid gielaide main lea rikkes morfologiija, muhto unnán čálalaš resurssat. Sáhttá geavahit FST mii lea juo huksejuvven sátnedárkkistanprográmma várás, dahje luksit sierra FST mas leat lemmat ja morfologiija mat leat relevánttai giellaoahlpuhussii, ja dinain genereret ollu hárjehusaid automáhtalačcat. Sihke addaga ja buktaga lea vejolaš prosesseret mánggain paramehteriin. Dákkár paramehter sáhttá ovدامéarkka dihte leat dihto suopmanai hámit ja sojaheamit. Hámiide dahje leemaaide mat gullet dihto suopmanii lasihuvvo +Dial/-gilkor (eng. ‘Dialect’ = suopman), ja sáhttá ráhkadit sierra suopman-FST:id go kompilerenproseeassas hilgu bálgaíd mat leat merkejuvvon eará suopmanii +Dial/-gilkoriin. Seammá ládje lea vejolaš genereret grammaathi, hárjehusaid dihto čállinváriánttain, jus gielas
le eanet go okta čállinuvohki geavahusas, nugo eamiálbmotgielain mángji lea, geahča 
kapihttala 4 ovdamearkkaíd.

3.4.2 Oahpahalli morfologalaš gaskagiela giedahallan (Art. VII)

Divvun-prográmma lea heivehuvvon eatnigielčállái dainna lágiin ahnte vurdo ahnte čálli 
diehtá mo sátni jietnaduvvo, muhto ii vealtameahttumit dieđe sání rivvtes čállinhámi. 
Giellaoahpahalli gielas, dáš gohčoduvvon L2 (eng. ‘Language 2’ = giella maid olmmoš 
aahppá maŋŋel go lea oahppan vuosttaš gielas (= L1)), leat dávja eará meattáhusat go 
eatnigielaga gielas. Giellaoahpahalli sáhttá lasiht boaesttugehčosa sántní dahje ii muitte 
váldit vuhtii buut morfofonologalaš beliidi. Boadus lea sátni man jietnadeapmi ii su-
lastahte jovssushámi. Govvosis 3.12 oaidnit ahnte giellaoahpahalli lea lasihan i-gehčosa 
eadni-substantiivi, muhto ii leat muitán soggevokála rievvdaemi i:á, ja čállá eadnii illa-
tiivahápmiin, dan sajis go eadnái. Divvun-prográmma evttoanalgoritma ii bastte addit 
rivvtes sántnehámi čállái, vaikko divvungaska ii leat go okta, danne go prográmma ii vuorut 
i:á meattáhusa.

 Govvus 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 dábbalaš baosttuhamiidi oktan iešgudetlágan 
meattáhusgikorriiguin vai lea vejolaš addit metalingvisttalaš máhcaha. Govvosa 3.12 
ovdamearkkas Giella-sme analysáhtor ii atte makkárge analysa eadnii-sántní, muhto L2-
FST:in sátni oččošii dákkár analysa: eadnii+N+Sg+Ill+IllVErr (eng. ‘Illative Vowel 
Error’ = illatiivva vokálameattáhus). +IllVErr-gikor sáhttá báhcahit metalingvisttalaš 
máhcaha: ‘Muitte soggevokála rievvdaemi illatiivvas.’.

Dábálaš giellaoahpahalli baosttučállin lea maiddái čállit a gokko galggašii leat á. 
Kapihttala 3.2.2 čoavddus lea stávenložžen, mainna sáňtášii sihke gávdnat gude lemma

Govus 3.13: L2-FST sisttisdoallá gaskagiela gehčosiid ja morfonologoiija: Transduser addá gehčosiod s ja i ja báhcahagaid ḧEG dásemolsuma ja ḧVOWCH vokálareivvadeami várás. Lassebálgát dääd báhcahagaid haga leat gilkorastoffuvvon +IllVErr ja +CGErr (eng. ‘Consonant Gradation Error’ = dásemolsumanteattáhus). (Artihkkala VII govus 3.)

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

1. Leksikontransduseris adden lassebálgáid gehčosiidda mat leat norpma ologgbealde, ovdamearkka dihte bárahisstávvals substantivevva illatiivagehčosa i, mii addá hivs-set-substantiiivva illatiivahámi hivssegii dan sajis go čállingiela hivssegii.


4. Konkatinerejin nuppiin transduseriiin mas leat buot báikenamat unna álgobustávain ja gilkor +LowercaseErr mii karakterisere čállinmeattáhusa mas báikenamma
3.4.3 Oahpahalli syntávssalaš gaskagiela giedahallan (Art. VIII)


Vuogádaga čuolbma lei ahte giellaohpahalli garvá váttis giellaohksehusaid iežas vástá-dusas. Čoavddus lei addit eavttuid vástádussii: geavaheaddji ferte vástádusastis geavahit

dihto lemmaid mat leat vižžuovvon stuorra lemmaseahtain (govus 3.14). QA-hárjehusaid evaluaremis, maid dahken autenttalaš geavahanloggain, boahá ovdan ahite QA-hárjehusat main leat eavttut, geasuhit studeanttaid eanemusat danne go dat buorebut čájeha ahite lea morfologalaš hárjehallamis sáhka, iige rievtti miedle duolhtamáilmmi ságastallamis.


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

### 3.4.4 Giellamodealla viiddideami čoahkkáigeassu

Dán kapihttalá fáddá lei mo sáhttá viiddidit giellamodealla nu ahte gokčá maiddái giellaahpahalli gaskagiela. Morfologalaččat dát čovdojuvvo go FST:i lasiha bálgáid mat addet systemáhtalaš boasttuhámiid, main leat iešgudetláagan meattáhusgilkorat mat dahket vejolažžan sihhe addit rivttes analysa ja dasa lassin addit metalingvisttalaš máhcahaga.

Giella-sme syntávssalaš analysáhtorata disambigueremnjuolggadusat vuodduuvvet dasa ahte cealkagat čuvvot eatnigiela syntávssa, go ulbmi lea teakstakorpusa disambigueren (kapihtal 2.4). Dán kapihttalis lean čájehan mo syntávssalaš analysáhtorata njuolggadusaid sáhttá heivelit giellaahpahalli syntákssii. Analysáhtor máhttá teavstta disambigueret, vaikko cealkagis váilu kongrueansa dakko gokko galggašii leat, dahje objeakta lea čállojuvvon nominatiivahápmin. Analysáhtor maiddái lasiha dákkár dáhpáhusaide gilkora mii sáhttá báhcabiit metalingvisttalaš dieduid čállái. Áibbas friddja čállojuvvon teavsttain déit livččii lean hástaleaddji, muhto giellaahpahalli atta lea vástádus ICALL-vuogádaga jearaldahkki, ja jearaldat eaktuda vástádusas dihto vearbasoaheami ja kásusa, juoga mii dahká olu álkibun addit máhcahaga.

Morfologalaš viiddideapmi dahká ahte dárbbása unnit syntávssalaš viiddideami go
3.5 Giellamodealla, dan geavahus ja duolta

Giellamáilbmi


Ovtta leksikontransduseris sáhttá kompileret mánga iešgužetlágan FST. Gilkoriiguín sáhttá hehttet dihto lemmáid dahje bálgažd boalitimis miele analysesáhtoríi normatiiva giela várás, dahje dihto suopmana várás. Diakrihtalas leavggaiquin sáhttá addit gáržžidusa sáni posišuvdni ja hámpái go goallostuuvvo eará sáníide mearusossin.

Lean erenoamážit giežahallan FST giellamodealla stuorra hástalusa genereret dušše sátnehámiid mat geavahuvvojít gielas. Dát mearkkaša hámiid mat gávdnojít teakstadahje jietnakorpusis, dahje maid eatnigielat dohkheivččií hápmín maid son sáhtásti geavahit. Geavaheddjiidprográmmmaid gáldofillain lea vejolaš raddjet genererema prográmmmi mas lea sáhka, ja nu sáhttá ovdatearkka dihte raddjet giellaoahppama pedagogalaš sivaid geažil, nugo lean dahkan sátnegijái ja ICALL-prográmmiin, mii lea čilgejuvgon kapihtta-liin 3.2.2 ja 3.4.1.

Muhto mii hálíiidit giellamodealla leat nu lahka duolta sámegiela go vejolaš. Dan dihto lea dárbu gáržžidit giellamodealla sihke analysa ja genererema dáfuš, ja dan leat Giella-smX-analysáhtoríiin dahkan guvtti láhkai: Mii leat čujuhan lemmáid joatkakaleksikonaide main leat iešgudetlágan morfologalaš bálgažd, ja diakrihtalas leavggaiqid bokte stivret gudiin lemmain bálgažd mannet dihto morfologalaš sojahemiide, nugo oamastangehčosiidda. Gáržžidusat mat leat dahkon ovdatearkka dihte Giella-sme analysáhtoríi,
eai leat ollislacčat, ja sáhtääsi árvvoštallat eanet gáržžidusaid, nugo varbbaid finihtta ja
finihtta hámiin. Badjelmearálaš genereren lea čuolbma, ja morfologalaš analysáhtora
gáržžideapmi boahtá buot prôgrammaide buorrin.

Kapihttalis 3.3.3 lean erenoamážit geahččan guokte čuolmma mat gusket finiht-
ta varbahámiid generereremii. Imperatiivva biehttalanhápmi livččii hui ávkkašaš sikhkut
normatiiva giellamideallas, muhto hámit leat dattetge geavahusas, vaikko hui marginála,
ja dat geavalus gábidivččii dalle sierra Divvun-prôgramma. Nubbi ovdamearka lea
vearbagenitiiva, mii ii leat nu produktiiva, ja varbagenitiivva badjelmearálaš hámit cihket
dábálalš cállinmeattálhaid.

Lean maiiddái geahččan finihtta varbahámiid. Juohke modusis ja áiggis varrbaide
genererejuvvojít ovcci finihtta hámi. Buot varrbbat earret dálkevarrbbat ožžot dálá
analysáhtoris buot ovcci persovnna-logu-hámi juohke modusis. Mu evttóhus lea hukset
vearbaleksíkona eará láhkaí go dál, vai varrbbat maiidda ii leat duodaštuvvon olmmošlaš
subjeakta, ožžot dušše +Sg3 ja +Pl3 hámiid.

Giełłamodealla ulbmil lea máhttit analyseret buot sáníid teavsttain. Jus teavsttaid
čállit leat giellaohpahallit, de lea dárbu viuddidit giellamodealla fátmmastit maiiddái sin
gaskagiela hámiid mat spiehkastit jovssushámiin. Artihkkala VII fáddá lea mo sáhttá
lasihit boassttubálgáid oktan gilkoriin, mii maiiddái dahká vejolažžan addit metalingvist-
talaš máhcahaga cãllái. Máiiddái syntávässalaš analysáhtora njuolggadusaid sáhttá heivehit
giellaohpahalli syntáksii, ja artihkal VIII cãjeha ahnte ICALL gáržžiduvvon vuogádaga
siskkobealde, lea vejolaš addit máhcahaga cãllái maiiddái syntávässalaš meattálhusaide.
Kapihtal 4

FST addá vejolašvuodaid
eamiálbmotgielaide

4.1 Anárašgiella ovdamearkan


Anarâškielâ servi lea iskamin ruhtadit prošeavttta Čyeti čälled anarâškielân man ulbmil lea oazžut čuodi odda anárašgiela cállí. Ovdaprošeavttta jearahallamis boahtá ovdan ahte odda cállit dárdbašit sátnedárkkistanprográmma ja neahttasátnegirjjii, muhto maiddái sojahanparadigmaid veahkkin go eatnasat leat nuppegielagat (Olthuis ja Gerstenberger, boahtimín). Giellateknologijia lea vástådus sin dárbbuide.

Bargu hukset anárašgiela FST álggahuvvui albma láhkai jagis 2014. Giella-smn lea dál dan muttus ahte dat dovdá 95,2 % teakstakorpusa 1,3 miljovnna sániin (tabeal-la 2.4). Giella-smX analysáhtoirigiuin lea vejolaš hukset iešgudetlagan geavahedddjijid-prográmmmaid (govus 4.1), ja Giella-smn lea juo laktojuvven dihtorjorgalanprográmmmaid mas lea davvisámegiella gáldogiellan (Antonsen ja earát, 2017). Dán prošeavttta ruhta-

52Complementary Aanaar Saami Language Education for adults
53http://gtweb.uit.no/mt-testing/ (14.02.2018)
4.2 Gielat Sámi olggobealde


FST huksen lea váštádas gielaide maiddái Sámi lagašguovllu olggobealde. Davvimáilmi eamiáblmotgielain lea hui rikkes morfologija ja teakstakorpus lea mendoon unni jus dihtor galggasíi oahppat sojahaminstariid. Ruonáeatnangielas (kalaallisut) leat 50 000 hálli ja ruonáeatnangillii leat almmuhuvvon ollu teavsttat, go giella geavahuvvo sïhke váldohálldahasas, beavvälaš aviissain ja vuoddoskuvllain. Muhto ruonáeatnangiella lea polysyntehtalaš giella, mas leat unnán sátnemáddagat ja ollu suorggideamit, ja dát dahká sátnetiippaid ja sániid gorálašvuoda erenoamáš unmin. Vuosttaš sátnedárkkistanprográmma ráhkaduvvui sátnelisttus mas ledje 350 000 sáni, ja dattetge dat govččai dušše 25 % sániin oktilaš aviissateavsttas. Lei čielggas ah teertii iskat hukset sátnedárkkistanpro-

54Norgga dutkaarádi juolludeapmi nr. 234299 ja Kone Foundation
55http://divvun.org/proofing/proofing.html

gramma eará lahkonanvugin, ja jagi 2006 válbmman ruonáeatnangia sátnedarkkistan-prográmmi Kukkuniiaat mii lea genererejuvven FST:in ja hukséjuvven Giellatekno- ja Divvun-joavkkuid dalá vuoodstrooktuvvras (geahča detáljaid Heilmann ja earát (2009)).

Juo jagi 2007 Kukkuniiaat govččai 90 % sániin oktilaš teavsttas (Oqaaserpassualeriffik, dovdmeeahltten jahki), ja da manjel leat prográmmas boahát odda buoriduvvon ver-shuvnmat.


Govus 4.2: Giellateknologalaš reaiddut eamiálbmogiidi gielaide. Gurut ravad-das lea njuolggadusvudot jorgalanprográmmma davvisámegielas dárogillii, Plains Cree sótnedárkkistanprográmmma stávvvalalfabehtain, lullisámegiela sátnegrigiri mii dovdá mor-fologiija (NDS) ja nuortalašgiela boallboveavdi mobiilateleovmnna várás.

Kapihtal 5

Konklušuvdna


Njealját kapihttalis lean čákahan mo FST huksen oktan Giella-vuoddostruktuvrrain vástida mángga davvimáïlnmi eamiálbmotgiela dárbbuide oazžut giellateknologalaš reaid-duid. Giella-vuoddostruktuvra maid sáhttá viežžat internealhtas, lea dahkan álkipun huksit FST, ja maiddái geavahit FST njuolga sátnegirjjis ja sátnedárkkistanprográmmas. Buot vuoddostruktuvrra buorádusat mat leat dahkkon sámegielaíd várás, bohtet maiddái eará eamiálbmot- ja vehádatgielaide buorrin. Dát vuoddostruktuvra ja FST lahkonan-
KAPIHTAL 5. KONKLUŠUVDNA

vuohki dahket alladásat geavaheddjiidprográmmmaid olámuddui maiddái eará vehádatgie-
laide main lea rikkes morfologijja, vátna teakstakorpus ja unnán hállit.


