

Ekstraktmorfoloogia meetodiga tuletatud  
keele tehnoloogia vadjä noomeni vormisõnastiku  
näitel

Kristian Kankainen

2019

# Sisukord

<b>1</b>	<b>Sissejuhatus</b>	<b>4</b>
<b>2</b>	<b>Vormisõnastiku koostamise teoreetilised lähtekohad</b>	<b>6</b>
2.1	Vadja ühiskeel ja korpuse planeerimine . . . . .	6
2.2	Sõnavara valik, paradigma moodustamine ja ühtlustamine . . . . .	7
2.2.1	Noomeni käänded ja ühtlustamine . . . . .	8
2.3	Morfeemi staatus ja definitsioon . . . . .	11
2.4	Klassikaline paradigmaatiline morfoloogia . . . . .	11
2.5	Arvutimorfoloogia eesmärk ja lingvistiline motiveeritus . . . . .	12
2.5.1	Morfoloogia formaalsete teooriate lingvistiline motiveeritus .	13
2.5.2	Arvutimorfoloogiate lingvistiline motiveeritus . . . . .	15
2.5.3	Sügavam epistomoloogiline põhjus: formaalse lingvistika ling- vistiline motiveeritus . . . . .	16
2.6	Muuttüüp, tüüpsõna ja sõnatüüp . . . . .	17
<b>3</b>	<b>Vormisõnastiku sisestamine Morfoloogialabori rakenduse abil</b>	<b>18</b>
<b>4</b>	<b>Ekstraktmorfoloogia meetod</b>	<b>19</b>
4.1	Sõnatüübi eraldamine . . . . .	19
	Tehnilise tüve eraldamine . . . . .	21
	Üldistamine sõnatüübi funktsiooniks . . . . .	21
4.2	Sõnatüübi ennustamine . . . . .	22
4.3	Võrdlus teiste meetoditega ja Eesti leksikograafilise traditsiooniga . .	23
<b>5</b>	<b>Vadja morfoloogiliste tüüpsõnade analüüs</b>	<b>25</b>
5.1	I käändkond . . . . .	25
5.2	II käändkond . . . . .	27
5.3	III käändkond . . . . .	32
5.4	IV käändkond . . . . .	35
5.5	V käändkond . . . . .	35
5.6	VI käändkond . . . . .	37
5.7	VII käändkond . . . . .	38
5.8	VIII käändkond . . . . .	39
5.9	IX käändkond . . . . .	42
5.10	X käändkond . . . . .	42
5.11	XI käändkond . . . . .	43
5.12	XII käändkond . . . . .	44
5.13	XIII käändkond . . . . .	45
5.14	XIV käändkond . . . . .	46
5.15	XV käändkond . . . . .	49
5.16	Ekstraktmorfoloogia üldistatud muuttüüpide algoritm . . . . .	50
5.16.1	Muuttüüp I . . . . .	50
5.17	Põhivormid ja analoogiavormid . . . . .	50

5.17.1	Käändsõnad . . . . .	51
5.17.2	Tegusõnad . . . . .	51
<b>6</b>	<b>Programmkoodi tuletamine</b>	<b>52</b>
6.1	Keskne kirjeldus Lexical Markup Framework vormingus . . . . .	52
6.1.1	Sõnaartikli esitamine LMFis . . . . .	54
6.1.2	Sõnatüübi malli esitamine LMFis . . . . .	55
6.2	Integreerimine Grammatical Framework'iga . . . . .	57
6.2.1	Morfoloogia moodul . . . . .	58
6.2.2	Leksikoni moodul . . . . .	60
6.3	Integreerimine Giella-taristuga . . . . .	60
6.3.1	Tüüpsõnad . . . . .	61
6.3.2	Leksikon . . . . .	62
6.3.3	Õigekirjakontrollija . . . . .	62
6.3.4	Arutelu . . . . .	62
<b>7</b>	<b>Kokkuvõte</b>	<b>63</b>
<b>8</b>	<b>Põhimõisted ja lühendid</b>	<b>64</b>
<b>9</b>	<b>Kirjandus</b>	<b>65</b>
<b>10</b>	<b>The use of Extract Morphology for Automatic Derivation of Language Technology for Votic</b>	<b>68</b>
<b>11</b>	<b>Lisad</b>	<b>69</b>

# 1 Sissejuhatus

Magistritöö eesmärk on koostada vадja keele noomeni jaoks morfoloogilise andmestiku ehk vormisõnastiku. Vormisõnastik täidab kaht suuremat funktsiooni: 1) on produktiivsele, tänapäeva keelekasutust peegeldav sõnaraamat ja 2) selle sisu on teisendatav programmkoodi, mis integreerib vадja keele morfoloogiamooduli kahte keeletehnoogilisse raamistikku.

Tööl on neli suuremat eesmärki ehk teesi:

1. näidata, et ekstraktmorfoloogia on intuitiivne viis luua arvutimorfoloogia (võimaldab mitte-(arvutus-)lingvistil luua ja muuta andmestikku)
2. ekstraktmorfoloogia meetodiga loodud arvutimorfoloogia on võimalik integreerida automaatse teisendusega erinevatesse keeletehnoogiatesse
3. ekstraktmorfoloogia meetodiga luuakse arhiveeritava ja püsivama väärtusega andmestik
4. kõrvutada ja ühitada vадja noomeni ekstraktmorfoloogia leiud traditsioonilise morfoloogiakäsitlusega
5. on laiendatav teise sõnaliigiga sõnade sisestamiseks
6. on laiendatav tõlkevastustega

Magistritöö eesmärk on luua vадja keelele morfoloogiline sõnastik sellisel moel, et keeletehnoogilised komponendid, mh õigekirjakontrollija, on sellest automaatselt tuletatavad. Sõnastik oleks loodava vадja kirjakeele õigekeelsuse baasiks ja aitaks keeleõppes ja revitaliseerimises kaasa.

Morfoloogiline sõnastik on süsteem, mis sisaldab kõigi sõnaartiklite kõiki muutvorme. Töö andmestik piirdub vадja noomeniga. Töö andmestik hõlmab u 900 noomenit, mis on valdavas osas pärit Vадda Sõnakopittõja sõnastikust (Heinsoo 2015) koos täiendustega teistest sõnaraamatutest.

Sõnade paradigmad, ehk käändvormide tüvemuutused, on saadud Tsvetkovi sõnaraamatust (Laakso 1989) milledele on lisatud muutelõpud levinud vадja keele õpikust

(Конькова ja Дьячков 2014). Sõnade kirja­pilti on ühtlustatud vadja loodava kirjakeele järgi (Heinsoo, isiklik kommunikatsioon).

Loodud sõnaraamatu kirjed on kõrvuti morfoloogiliste muutvormidega varustatud ka sõnatüübitähisega. Sõnatüübid leitakse ekstraktmorfoloogia meetodi abil automaatselt ja sõnatüü­bistik uuendatakse iga kord sõnaraamatu kirjeid muudetakse.

Ekstrakt-morfoloogiline klassifikatsioon põhineb ainult sõnade mu­tu­vormide tähtkoostisel ja sõnatüü­bistik on maksimaalne selles suhtes, et igal tüvevokaalil ja astmevahelduslikul tähel on omaette sõnatüüp.

Kõrvuti eksisteeriv sõnatüü­bistik võimaldab arvutimorfoloogia automaatse loomise ja niiviisi on sõnaraamatusse uute kirjete lisamise juures süsteemil võimalik genereerida ja kasutajale kuvada sõna potentsiaalseid muutvorme. Samuti on võimalik sõnatüü­bistiku kirjelduse põhjal automaatselt luua programmkoodi, mis täidab morfoloogia komponendi ülesannet ühes või mitmes programmeerimiskeeles.

Ekstraktmorfoloogia meetod sobitub magistritöö ülesandele hästi mitmel põhjusel:

- meetod põhineb otse sõnavormidel ja ei vaja reegleid (vadja keelel puudub morfoloogiat piisavalt täpselt kirjeldav grammatika)
- tekitatud sõnastik koos sõnatüü­bistikuga on võimalik automaatselt ümber kodeerida keeletehnoloogiasse (mh õigekirjakontrollija)
- loodud sõnastik on iseenesest kasutajaliides, mis võimaldab vadja keele huvilistel endil oma sõnavara rikastada (ei vaja programmeerimis- ega keeleteaduslike oskusi)
- sõnastikus esinevad muutvormide artikleid on võimalik varustada korpusesinemuste informatsiooniga ja õigekeelsuslike kommentaaridega

Magistritöö on struktureeritud järgmiselt:

## 2 Vormisõnastiku koostamise teoreetilised lähtekohad

Siin peatükis kirjeldatakse vormisõnastiku koostamise ja esialgseid vadjä kirjakeelsete muutvormide rekonstrueerimise põhimõtteid. Andmestiku sisestamist ja vormistamist vormisõnastikuks Morfoloogialabori rakenduse abil kirjeldatakse peatükis 3.

Vormisõnastiku definitsiooniks magistritöös on võetud Ülle Viksi järgi:

“Üks täielik vormisõnastik peaks esitama kõigi sõnade kõik muutevormid koos vastava grammatilise iseloomustusega. Ainult siis saab kasutaja sõnastikust ilma mingi vaevata ja täiesti kindlalt teada, milline on vajalik vorm antud sõnast või millise sõna millise vormiga on tegemist tundmatu sõnavormi puhul.” (Viks 1992, lk. 7).

### 2.1 Vadjä ühiskeel ja korpuse planeerimine

Selles peatükis seletatakse vadjä ühiskeeleks valitud aluseid ning korpuse planeerimise mõistet. Vadjä keelele ei loodud 1930-ndateil Nõukogude Liidus ühiskeelt, nagu seda tehti näiteks karjala, vepsa ja isuri keele jaoks. Vadjä kirjakeele loomise pürgimistest pärast Nõukogude Liidu lagunemist on kirjutanud lähemalt Ernits 2006 ja vadjälaste ilma kirjakeelela keele normi tunnetusest on rääkinud Elena Markus ja Fedor Rozhanskiy 2013.

Korpusplaneerimine on keeleplaneerimise üks osa ja koosneb (Kloss, Université Laval ja Centre international de recherche sur le bilinguisme 1970) järgi kolmest osast:

1. kirjamine (ingl. *graphization*) ehk kirjaviisi määramine
2. morfoloogia ühtlustamine ja standardiseerimine
3. sõnavara moderniseerimine ja rikastamine

Neid osi ei pea vaatama järjestiku etappidena keele moderniseerimise poole, vaid (Coulmas 1989) on nimetanud seda pidevaks adapteerimise protsessiks.

Käesolev magistritöö jääb morfoloogia ühtlustamise ja standardiseerimise alla, ent tahest-tahtmata on kasutatud kirjaviisiga tehtud valik, kus on järgitud Heinsoo õppematerjalide kirjaviisi. Valitud kirjaviis on siiski automaatselt teisendatav teise levinud, Konkova õppematerjalidele vastavaks, mistõttu peaks magistritöös loodud

materjal olema mõlema õppematerjali jaoks kasulik. Ainsad erinevused kahe kirjaviiisi vahel on  $\ddot{u} \rightarrow y$  ja  $\check{c} \rightarrow c$ .

Soodustamaks Vadja keeles kirjutamist ja sedakaudu keele elavdamist, on magistritöö käigus ilmunud mõlemaid ortograafiaid toetav klaviatuuripaigutised arvutile ja nutiseadmeile (**kankainen\_annõmmõ\_2019**). Vadja keele jaoks kasutatud erinevatest kirjaviisidest on kirjutanud pikemalt Ernits 2010.

Morfoloogia ühtlustamist iseloomustab siinse töö puhul sõnavormide rekonstrueerimine Jõgõperä murdeliste muutvormide põhjal Vaipooli murdealale iseäralike joontele vastavaks. Standardiseerimise suhtes ei võtta magistritöö süsteemset seisukohta, vaid on oma valikutes pigem kirjeldava iseloomuga. Jõgõperä ja Luuditsa murrete erinevustest on kirjutanud mh Fedor Rozhanskiy ja Elena Markus 2015. Ühtlustamist kirjeldatakse lähemalt järgmises alajaotuses.

Antud magistritöö ei pürgi vadja keele sõnavara moderniseerida ega rikastada – hõlmatud sõnavara on varem juba avaldatud. Võimalikest vadja kirjakeele sõnavara rikastamise eri viisidest on kirjutanud Ernits 2010. Vene laensõnade adapteerimisest Jõgõperä murdes ja selle problemaatikast on kirjutanud Rozhanskiy 2009.

Kuna magistritöö loob ühtse töökeskkonna, mille abil määratakse morfoloogia ainult sõnade muutvormide esitamise kaudu ja ei vaja programmeerimisoskusi, siis loodab siinne autor sellele, et edaspidine korpuse planeerimine võiks toimuda vadjalaste endi ja vadja keele huviliste eestvõttel – muuda sõna muutvormi vormisõnastikus ja see muutub ka keeletehnoloogias.

## 2.2 Sõnavara valik, paradigmade moodustamine ja ühtlustamine

Sõnavara valik hõlmab Heinsoo õppematerjali sõnastikus esitatud nimisõnu ja adjektive (u 420 sõna). Sõnavara on veel laiendatud autori enda silma järgi Tsvetkovi sõnastikust ettesattunuga (u 460 sõna, mh *asfaltti*, *bibli* ja *biblioteekkõ*). Kokku on vormisõnastikusse koondatud 882 sõnaartiklit.

Sõnaartiklitele on lisatud Tsvetkovi sõnaraamatus esitatud põhikäändevormid (SG NOM, SG GEN, SG PAR, SG ILL ja PL NOM, PL GEN, PL PAR, harva ka PL ILL). Juhul kui

Tsvetkovi sõnaraamatus on esitatud mitut paralleelset põhikäändevormi, on eelistatud pikemaid vorme, nt SG ILL *asfalttisõ* lühikese illatiivi asemel *asfaltti*. Niiviisi on morfoloogiliste sõnatüüpide kaardistamise protsessis välditud paralleelvorme, kuigi loodud lõpplahendus ehk arvutimorfoloogiline süsteem võimaldab sõnaartiklitel esitada ka paralleelseid muuttüüpe.

Alljärgnevalt seletatakse põhikäändevormide ühtlustamise ja analoogiavormide koostamise põhimõtteid.

### 2.2.1 Noomeni käänded ja ühtlustamine

Käänete valiku ja muutelõppude puhul on järgitud Konkova õpikus esitatut (Konkova ja Djačkov 2014, lk. 10), mis langeb kokku Heinsoo õppematerjalide põhimõtetega (Heinsoo 2015, lk. 88). Käänded on näitlikustatud tabelis 1.

Hõlmatud käändeid on seega 24: ainsuse ja mitmuse nominatiiv, genitiiv, partitiiv, illatiiv, inessiiv, elatiiv, allatiiv, adessiiv, ablatiiv, translatiiv, terminatiiv ja komitatiiv.

Välja on jäetud essiivi, abessiivi, ekstsessiivi ja instruktiivi käänded, mida Ariste on pidanud produktiivsete käänetenähtetena Ariste 1968, lk. 17. Markus ja Rožanski grammatika Маркус ja Рожанский 2011 ei käsitle terminatiivi ja komitatiivi käänetenähtetena vaid käände ja järelasendi vahepealsetena, mille analüüsi nad on põhjendanud pikemalt Markus ja Rozhanskiy 2014.

Alljärgnevalt seletatakse kuidas Tsvetkovi sõnastikust saadud põhivormid on ühtlustatud ja mille põhjal analoogiakäänded on moodustatud.

**Nominatiiv** Tsvetkovi sõnaraamatus antud vormile on tüvelõpuvokaali puudumise korral tavaliselt lisatud lühike vokaal vastavalt vokalismile (Heinsoo ja Konkova eeskuju järgi).

**Genitiiv** Tsvetkovi sõnaraamatus antud vormi lõpuvokaal on ühtlustatud vastavalt Heinsoo ja Konkova sõnastikes esitatuga või rekonstrueeritud vastavalt sõna vokalismile.



	kääne	eespoolne	eespoolne	tagapoolne	tagapoolne	tagapoolne
SG NOM	pää	ärče	lafka	lammõz	ivuz	
SG GEN	pää	ärjä	lavga	lampa	ivusõ	
SG PAR	pääte	ärčä	lafka	lammassõ	ivussõ	
SG ILL	pähhe	ärčäse	lafkasõ	lampasõ	ivussõsõ	
SG INE	pääz	ärjez	lavgõz	lampaz	ivusõz	
SG ELA	pääss	ärjess	lavgõss	lampass	ivusõss	
SG ALL	päälle	ärjelle	lavgõllõ	lampallõ	ivusõllõ	
SG ADE	pääll	ärjell	lavgõll	lampall	ivusõll	
SG ABL	päält	ärjelt	lavgõlt	lampalt	ivusõlt	
SG TRA	päässi	ärjessi	lavgõssi	lampassi	ivusõssi	
SG TER	päässaa	ärjessaa	lavgõssaa	lampassaa	ivusõssaa	
SG COM	pääka	ärjeka	lavgõka	lampaka	ivusõka	
PL NOM	pääd	ärjed	lavgõd	lampad	ivusõd	
PL GEN	päije	ärčije	lafkojõ	lampajõ	ivussijõ	
PL PAR	päit	ärčiit	lafkoit	lampait	ivussiit	
PL ILL	päise	ärčiise	lafkoisõ	lampaisõ	ivussiisõ	
PL INE	päiz	ärčiiz	lafkoiz	lampaiz	ivussiiz	
PL ELA	päiss	ärčiiss	lafkoiss	lampaiss	ivussiiss	
PL ALL	päille	ärčiille	lafkoillõ	lampaillõ	ivussiillõ	
PL ADE	päill	ärčiill	lafkoill	lampaill	ivussiill	
PL ABL	päilt	ärčiilt	lafkoilt	lampailt	ivussiilt	
PL TRA	päissi	ärčiissi	lafkoissi	lampaissi	ivussiissi	
PL TER	päissaa	ärčiissaa	lafkoissaa	lampaissaa	ivussiissaa	
PL COM	päika	ärčijka	lafkoika	lampaika	ivussijka	

Tabel 1: Noomeni käänded koos käändelõppudega ees- ja tagapoolse vokalismi kujul.

**Partitiiv** Tsvetkovi antud vormi lõpuvokaal on ühtlustatud Heinsoo ja Konkova sõnastikele vastavalt või rekonstrueeritud vastavalt sõna vokalismile. Mitmuse partitiivi puhul on eelistatud pikemat *-it*-muutelõppu ainult *-i*-mitmusliku tüve asemel, mis mõlemad esinevad paralleelsete variantidena Konkova õppematerjalis.

**Illatiiv** Loodava kirjakeele ühtlustamise huvides on eelistatud läbinähtava käändelõpuga vormi *-se/-sõ* lühikese illatiivi asemel. Selline eelistus on vastuolus Markus ja Rozhanskiy 2014, lk. 247 on väljatoodule, et tänapäeva Luuditsa murdes esineb tavalist lühike vorm ja pikem vorm väga harva. Konkova esitab mõlemad vormid rööpselt.

**Inessiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-z*. Mitmuse vormid

on rekonstrueeritud PL ILL põhjal.

Ariste järgi on vadja keele inessiivile omaseks tunnuseks see, et geminaatklusiilid *-kk-*, *-pp-* ja *-tt-*, geminaatafrikaadid *-tts-*, *-ttš-*, geminaat *-ss-* ning konsonantkluster *-hs-* esinevad alati tugevas astmes (Ariste 1968, lk. 23). Seda printsiipi järgides on SG INE käändevormi vastavad tüved rekonstrueeritud tugevaastmelisteks.

**Elatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-ss*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal.

**Allatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-lle* või *-llõ* vastavalt sõna vokalismile. Mitmuse vormid on sarnaselt rekonstrueeritud PL ILL vormi põhjal.

**Adessiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-ll*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal.

**Ablatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-lt*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal.

**Translatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-ssi*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal.

**Terminatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud SG INE vormi alusel, kusjuures muutelõpp *-z* on asendatud lõpuga *-ssaa*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal.

Ariste järgi võib terminatiiv põhineda illatiivi või allatiivi tüvele (Ariste 1968, lk. 34). Markus ja Rožanski järgi põhineb see tänapäeva keeles genitiivi tüvele ja harva

allatiivi tüvele (Markus ja Rozhanskiy 2014, lk. 247).

Ülaltoodud märkuste järgi võib magistritöös valitud SG INE tüve põhjal konstrueeritud muutevorm olla vale, kuigi valik ühtib Konkova ja Heinsoo õppematerjalides esitatuga, so tugevaastmelise geminaadiga.

**Komitatiiv** On analoogiavorm ja selle muutevormid ainsuses on rekonstrueeritud PL NOM vormi alusel, kusjuures muutelõpp *-d* on asendatud lõpuga *-ka*. Mitmuse vormid on rekonstrueeritud PL ILL põhjal, kusjuures *-i*-tüveliste sõnade puhul on moodustunud pikk *i* muudetud *-ij-* vastavalt Konkova esitatud vormile (*--ijka*).

### 2.3 Morfeemi staatus ja definitsioon

Morfeemi ei käsitleta siin töös levinud lingvistilisest seisukohast kui *väikseimat tähenduslikku üksust*, vaid klassikalistele paradigmaatilistele lähenemistele omaselt kui *mistahes tähtkoostise muutust, millega kaasneb tähenduslik muutus* (Beard 1987; Beard 1995).

### 2.4 Klassikaline paradigmaatiline morfoloogia

Matthews tõstab esile kaks WP-mudelit: klassikaline ja uus ehk strukturalistlik. Klassikaline kuulub X ajastusse ja selle kõrghetk oli õpikutes X sajandil, ajal kui keelt õpetati kooloniates (Matthews 1991 lk X). Uue rajas Matthews ja seda on Karttuneni sõnul edasi arendanud Zwicky, Anderson ja Stump (Karttunen 2003, lk. 2).

Klassikaline on lähedasem siinses magistritöös rakendatud ekstraktmorfoloogia-le. Klassikalise ja strukturalistliku lähenemise suurim vahe seisneb kahes asjaolus. Eeskätt ei näinud klassikalised grammatikud sõnast väiksemat ja tähenduslikku üksust, elik ei arvestatud morfeemiga. Arvestati ainult vormikülje üksustega, tähtedefoneemide ja silpidega. Sõna oli väikseim tähenduslik element keeles ja seda nähti tervikuna, hõlmates terve oma paradigma kuuluvaid vorme. (Matthews 1991 lk X–Y)

Matthews toob välja (klassikalise) paradigmaatilise suuna kolm head omadust. ....

Õpikutes ja grammatikates välja toodud reeglid, mille abil sai ühe lekseemi para-

digma moodustada, opereerisid ainult sõnavormide tähtkoostisel. Näiteks võidi ühe sõnavormi lõputähti asendada teiste tähtedega, et saada teine sõnavorm. Et asendavatele tähtkoostistele ei omandatud mingit tähendust, näitlikustab see, et mõne reegli algvormiks võidi valida selline sõnavorm, mis oma tähtkoostise poolest kõige paremini sobis. (Matthews 1991 lk X–Y).

Sõna kui selle vormide tervik; pedagoogiline praktika ja paradigma üldistuse ülekantavus uutele sõnadele Matthews 1991. Matthewski jätab mudeli vormipõhiseks ja mitte morfeempõhiseks, selle kohta edasi järgmises allosas.

## 2.5 Arvutimorfologia eesmärk ja lingvistiline motiveeritus

Arvutimorfologia eesmärgiks on siin töös valitud moodustada elektroonse vormisõnastiku:

“Üks täielik vormisõnastik peaks esitama kõigi sõnade kõik muutevormid koos vastava grammatilise iseloomustusega. Ainult siis saab kasutaja sõnastikust ilma mingi vaevata ja täiesti kindlalt teada, milline on vajalik vorm antud sõnast või millise sõna millise vormiga on tegemist tundmatu sõnavormi puhul.” (Viks 1992, lk. 7).

Sellist vormisõnastikku võib moodustada erinevatel viisidel. Näiteks leksikaalse andmebaasina, kus iga lekseemi puhul on nenditud kõik selle muutevormid koos vastava grammatilise iseloomustusega, või näiteks reeglite komplektina, mida rakendades saab koostada lekseemi muutvorme vastavalt nende grammatilistele iseloomustustele.

Matemaatilises mõttes kujutab vormisõnastik vaid *seost* muutevormide ja nende vastavate grammatiliste iseloomustuste vahel.

Arvutimorfoloogiad võivad seda seost (või vormisõnastiku funktsionaalsust) realiseerida arvutuslikult erinevatel viisidel ja ei pea olema lingvistilis-grammatiliselt motiveeritud. Kuna üks täielik vormisõnastik on mahult niivõrd suur (kui mitte lõpmatult suur), on selle mahu kompaktsem ja ülevaatlikum esitus peamiseks motivat-

siooniks organiseerida selle koostamise reeglite abil, mis on ühel või teisel moel põhjendatud lingvistiliste-grammatiliste reeglipärasustega.

### 2.5.1 Morfoloogia formaalsete teooriate lingvistiline motiveeritus

Eelmise sajandi keskpaiku jagas Charles Francis Hockett kõik seni Ameerikas sajandi algusest saadik ilmunud grammatikad kahe üldise mudeli järgi, IA (ingl. *Item-and-Arrangement*, üksus ja distributsioon v järjestus v korraldus) ja IP (ingl. *Item-and-Process*, üksus ja protsess ehk protsessimorfoloogia). Kõrvalmärkusena tõi ta välja ka kolmanda, “vanema ja väarikama” mudeli, WP (ingl. *Word-and-Paradigm*, sõna ja paradigma), aga jättis selle oma käsitlusest välja (Hockett 1954, lk. 210). Hockett võrdleb IA ja IP mudelite eelseid ja argumenteerib, et IA toonane populaarsus seisneb eeskätt selles, et ajastu eelistab formaalseid mudeleid. Kuna IA-mudel oli juba formaliseeritud tahtis Hockett nüüd formaliseerida sellest vanema IP-mudeli (Hockett 1954, lk. 214) ning sellest sai hiljem, Fred Karlssoni sõnade järgi, generatiivse lingvistika peamiseks mudeliks (Karlsson *et al.* 2002, lk. 126).

IP-mudel põhineb (morfoloogilise) protsessi mõistel, millega ühest algvormiks valitud kujust (ingl. *base*) luuakse teine vorm (Hockett 1954, lk. 210). IA tekkis vastureaktsioonina IP protsessimõiste suunalisusele – enam ei tahetud tõsta esile üht vormi algsemaks teistest vormidest (Hockett 1954, lk. 211). IA põhineb morfeemi mõistel, mida Hockett iseloomustab kui keele väikseimat grammatiliselt olulist üksust, ja selle distributsiooni määramisel (Hockett 1954, lk. 212). Hockett nendib, et ka IA mudeli puhul tuleb siiski teha kohati suvalisi valikuid selle üle, mis kuulub morfeemi tasandile ja mis kuulub distributsiooni tasandile (Hockett 1954, lk. 212).

Gregory Stump on arendanud Hocketti IP ja IA kaheksjagamise klassifikatsiooni edasi tänapäevaste morfoloogiliste teooriate põhjal. Nimetades IAd ümber leksikaalseks (ingl. *lexical*) ja IPd inferentsiaalseks (ingl. *inferential*) lisab ta klassifikatsioonile veel sisemise telje: inkrementaalsed (ingl. *incremental*) ja realiseerivad (ingl. *realizational*) teooriad. (Stump 2001, lk 1-2)

Inkrementaalsete teooriate järgi lisandub iga (olgu IA puhul leksikaalselt loetletud või IP puhul inferentsiaalse reeglga tuletatud) morfosüntaktilise tunnuse puhul

sõnale ka selle vormiline eksponent (Stump 2001, lk. 2). Vormilised eksponendid on üks-üheses seoses grammatiliste tunnustega ja need väljenduvad ükskhaaval elik inkrementaalselt.

Realiseerivate teooriate juures ei pea vormiline eksponent iga morfosüntaktilise tunnuse puhul eraldi ja koheselt väljenduma, vaid vormiline väljendus võib realiseeruda tunnuste suuremate komplektide puhul või üldse kui sõna kõik tunnused on teada (Stump 2001, lk. 2).

Realiseerivad teooriad võimaldavad niiviisi suurema paindlikkuse vormiliste väljendujate *realiseerimisel*, loobudes vormiliste väljendujate üks-ühesest seosest morfosüntaktiliste tunnustega.

Stumpi jagab oma klassifikatsiooni järgi Lieberi morfoloogilise teooria leksikaalseks ja inkrementaalseks. Halle ja Marantzi distributsioonilise morfoloogia teooria leksikaalseks ja realiseerivaks. Steele'i artikuleeritud morfoloogia teooria esindab inferentsiaalset ja inkrementaalset suunda. (Stump 2001, lk. 2–3).

Stumpi enda ja Matthews, Zwicky ning Andersoni teooriaid nimetab ta WP teooriateks, mis on inferentsiaalsed ja realiseerivad (Stump 2001, lk. 3).

Robert Beard on nimetanud ülaltoodud viimaste autorite arendatud realiseerivaid teooriaid eru-morfoloogiaks (ingl. *'split' morphology*) (Beard 1987, lk. 20) ja pakkunud välja morfoloogia veel võimsama eraldamise, mis põhineb tema morfoloogia lahususe hüpoteesil (ingl. *Separation Hypothesis*) (Beard 1995).

Morfoloogia lahususe hüpoteesil põhinevate teooriate ja realiseerivate (eru-)morfoloogia-teooriate vahe on fundamentaalne ja lähtub nende käsitlest süntaksi ja semantika vahekorra. Kõige ilmekalt paistab nende vahe morfeemi definitsioonis, küsimuses kas morfeem on keele väikseim vormiline tähenduslik üksus või mitte.

Beardi teoorias ei ole morfeem grammatiliselt tähenduslik, vaid defineeritud kui mistahes muutusena lekseemi fonoloogilises kujus (Beard 1987, lk. 31). Seega on tema teoorias ainult lekseemid tähenduslikud märgid ning grammatilised afiks (morfeemid) on seda vaid sattumuslikult (Beard 1987, lk. 17).

Käesolevas magistritöös rakendatud ekstraktmorfoloogia on oma organisatsiooni suhtes sõna ja paradigma mudel, aga selle käsitus morfeemist on lähedasem Beardi

teooriale.

### 2.5.2 Arvutimorfoloogiate lingvistiline motiveeritus

Arvutilingvistikas on arvutimorfoloogiat üldiselt organiseeritud klassikalise morfeemikäsituse järgi. Seda ilmestab hästi Lauri Karttunen, kes nendib inimkeele mudeldamise puhul arvutimorfoloogias kaks väljakutset: 1) morfotaktika ehk sõnast väiksemate üksuste kombineerumine ja 2) morfoloogilised vaheldused ehk sõnast väiksemate üksuste kuju olenemine nende ümbritsevast kontekstist (Karttunen 2003).

Mille mõlemad väljakutsed viitavad otseselt klassikalisele morfeemikäsitusele.

Karttuneni artikkel on vastus Stumpi teooriale ja ta näitlikustab selles kuidas Stumpi teooria on võimalik rakendada kasutades lõplike automaatide formalismi.

Karttunen toob välja olukorra, et arvutimorfoloogiad põhinevad arvutuslikel formalismidel, millega nad implementeerivad morfoloogiaid ja mitte ei põhine otse mingil lingvistilisel teoorial. Ta ütleb et morfoloogiauurija üllitiste peamine eesmärk on olla veenev, et tema teooria annab läbinägelikuma (ingl. *insightful*) ja elegantsema kirjelduse kui teised teooriad ja formalismid (Karttunen 2003, lk. 2). Praktilised küsimused nagu sõnavaraline katvus, arvutuskiirus ja mälumaht ei ole relevantssed akadeemilisele morfoloogiauurijale (Karttunen 2003, lk. 2).

Seega võib öelda, et arvutimorfoloogia on laiem kui lingvistiline morfoloogia, kuna esimest ei piira mitte teooria, vaid arvutusliku meetodi võimsus. Karttunen tõestab artiklis, et Stumpi inferentsiaalne-realiseeriv teooria on taandatav lõplike automaatide formalismi arvutusvõimsusele.

Sellest võib järeldada, et arvutilingvistikas on lingvistilise teooria roll pigem olla ajendiks kui tõetruuks postulaadiks, kuigi kindlasti on teooria ja selle implementatsioonilise praktika vahekord raskesti eraldatavad ja ajas muutuvad. Kuigi tendentsi tõetruuduse vähenemisele võib siiski täheldada tänapäeval ka Kimmo Koskenniemi töös, kus ta on hiljuti oma taandatud kahetasemelises morfoloogiamudelil püüdnud morfofoneemi mõiste juures loobuda selle tähendusliku külje lingvistilisest realismist, omastades seda puhtalt vormile:

“*Morphophonemes* are represented just as the *combinations of the corresponding letters* (or phonemes) which we can observe in the surface forms. On the one hand, such an interpretation of morphophonemes is crude, but on the other hand, it is a fact that anybody can observe.” (Koskeniemi 2013, lk. 157)

### 2.5.3 Sügavam epistemoloogiline põhjus: formaalse lingvistika lingvistiline motiveeritus

Sügavama epistemoloogilise põhjuse, miks arvutimorfoloogiaid on ajendanud pigem lingvistiline motivatsioon ja mitte arvutusteoreetilised võimalused, arwab siinkirjutaja leiduvat strukturealse lingvistika formaliseerimisperioodi alguses, mis algas enne arvutusmasinate leiutamist (1940.–1960.-ndate aastatel) ja ammu enne arvutite arvutus- ja mäluvõime võimsuse plahvatuslikku suurenemist (1980.–2000.-ndail).

Formaalseid teooriaid ja seega teooriate formaliseerimist peetakse teaduse lipulaevaks (Auroux *et al.* 2006, lk. 2026). Teooriate formaliseerimisprotsessi jagab Pieter Seuren neljaks etapiks, kus esimene koosneb uuritava ainese tüüpide (ehk kategooriate) leidmisest ning nendele esituskuju määramisest (Auroux *et al.* 2006, lk. 2027). (Teisisõnu tegeleb see *type-token distinction*’i probleemiga). Teine etapp käib sageli käsi-käes esimese etapiga ja hõlmab tüüpide taksonoomia määramist, ehk selle määramist, mis andmed kuuluvad mis tüübi alla millal ja mis tingimustes (Auroux *et al.* 2006, lk. 2027). Kolmas etapp koosneb struktuuri määramisest tüüpide esinemisele, ehk kuidas kategooriaid on võimalik omavahel kombineerida (Auroux *et al.* 2006, 2027 jj) näiteks puu- või sõltuvusstruktuuride abil. Neljas ja viimane etapp koosneb ühe ennustava ja kirjeldava väärtusega formaalse teooria ülesseadmisest algoritmina ehk sammsammulise tegevusjuhisenä (Auroux *et al.* 2006, lk. 2031).

Arvutimorfoloogia on arvutiprogramm (või mitme programmi komplekt), mis tahest-tahtmata hõlmab seelaadset formaalset sammsammulist tegevusjuhust.

Probleem, miks arvutimorfoloogiad juhenduvad lingvistilistest teooriatest ja mitte puht-arvutuslikest võimalustest asub formaliseerimisprotsessi 3. ja 4. etapi vahel.



Millisel viisil tuleb põhjendada struktuuri määravaid reegleid?

Zellig Harris (kes oli Noam Chomsky juhendaja) kirjeldab oma *magnum opus* teoses grammatika formaliseerimise lähenemist, mis põhjendab struktuuriliste reeglite määramise ühe formaalse avastamismenetluse abil keeleainese korpusesinemustest. See on, formaalse teooria sammsammulised reeglid tuletatakse puhtalt struktuuri- de esinemistest korpusanalüüsi teel. Selline väga töömahukas grammatika loomise menetlusviis sai tema kaasaegsetelt kõva kriitikat olles nii ilmselgelt ebarealistlik ja ebapraktiline. Harris oli tundlik kriitikale ja mainib oma raamatu lõpus viisi, kuidas korpusesinemustest eraldi püstitatud reegleid saab hoopis vastupidises suunas *testida* korpustekstide peal. See pani aluse generatiivsele grammatikale, mida arendas edasi tema kasvandik Noam Chomsky teoses *Syntactic Structures* (1957). (Auroux *et al.* 2006, lk. 2031).

Eelnevaga olen ma tahtnud öelda seda, et arvutimorfoloogia koostamispõhimõtted põhineda morfeemil ja morfotaktilistel reeglitel ja mitte puhtalt muutvormide nentimisel korpuse põhjal, on eeskätt ajalooliste traditsioonide järjepidevus. Käesolev töö ei järgi neid traditsioone.

See traditsioon on kristalliseerunud ka pealkirjas “This volume grows out of a special session that we organized at the January 2009 Annual Meeting of the Linguistic Society of America entitled “Computational Linguistics: Implementation of Analyses against Data”.” (Bender ja Langendoen 2010).

## 2.6 Muuttüüp, tüüpsõna ja sõnatüüp

Eesti traditsiooni järgi on muuttüüp tüüpsõnast üldisem. Kuidas siin töös terminoloogiliselt ümber käia, kas *muuttüüp* või *tüüpsõnamall*?

Muuttüübistik sõltub selle aluseks võetud klassifikatsioonist, ekstraktmorfoloogiat võiks vaadata kui lihtsalt üht väga formaalselt defineeritud muuttüübistikku.

Huldenil on omakorda üks väga formaalne viis, kuidas vähendada ekstraktmorfoloogiaga leitud muuttüüpide arvu. Kas see on hoopis muuttüübistik?

### **3 Vormisõnastiku sisestamine Morfoloogialabori rakenduse abil**

Sisestamise töövoog on olnud:

1. sisestada lekseemi põhivorm
2. valida olemasolev tüüpsõna ennustatud sõnatüüpide põhjal
3. sõnatüübi (ja tüüpsõna) puudumise juhul sisestada lekseemi kõik muutvormid

Punkti (3) juhul loob rakendus ekstraktmorfoloogia abil uue sõnatüübi, mis on koheselt järgmise sisestatava lekseemi puhul ennustatav ja valitav.

Töö algusfaasis ja tühja andmestikuga alustamise tõttu ei ole loomulikult olnud võimalik valida paljudele sisestatud lekseemidele õiget sõnatüüpi ning on pidanud sisestama kõik muutvormid. Töö jooksul on sõnatüüpide arv kasvanud ja rakenduse abil on saanud valida õige sõnatüüp, mis on tööprotsessi kiirendanud. Tihti on ka olnud abiks valida nt vokalisti suhtes vale sõnatüüp ja genereeritud muutvormides ainult korrigeerida vokalisti kajastav tähtkoostis.

## 4 Ekstraktmorfoloogia meetod

Ekstraktmorfoloogia on juhendatud masinõppe meetod, mis koosneb kahest komponendist: *sõnatüübi eraldamise meetodist* ja tundmatule sõnale õige *sõnatüübi ennustamise meetodist*. Meetod on *juhendatud*, sest sisestatud andmed, ehk muutvormitabelid, peavad olema korrektsed.

Meetod on välja pakutud kui lihtsama ja loomulikuma viisina lingvistil määratleda arvutimorfoloogia, kuna see põhineb lekseemide sõnavormide esitamisel ja ei juhindu morfoloogiliste-grammatiliste reeglite kirjutamisest (Forsberg 2016, lk. 14).

Sõnatüübi eraldamise korral üldistab meetod sisestatud muutvormitabelit ja eraldab sellest lekseemi tehnilise tüve ja sõnatüübimalli. Meetodi seda osa kirjeldatakse lähemalt pt-s 4.1.

Eraldatud sõnatüübimallide ja tehniliste tüvede põhjal on võimalik üldistada nende iseärasusi ja luua statistiline ennustusmudel. Ennustusmudeliga on võimalik määrata tundmatu sõnavormi kuuluvust ühe või teise sõnatüübi alla. Meetodi seda osa kirjeldatakse lähemalt pt-s 4.2.

Magistritöös kasutatakse ekstraktmorfoloogia meetodit eesmärgiga, et luua vajja keelele keeletehnoloogiat. Selleks otstarbeks tuletatakse eraldatud sõnatüübimallide põhjal programmikoodi, mis teostab mallidest tuleneva sõnavormide analüüsimise ja sünteesimise võime. Programmikoodi automaatset tuletamist kirjeldatakse pt-s 6.

### 4.1 Sõnatüübi eraldamine

Sõnatüübi eraldamise mehhanism põhineb tehnilise tüve tuvastamisel. Tehniline tüvi moodustub nendest tähtedest, mis ilmnevad lekseemi igas muutvormis.

Veel ilma detailidesse takerdumata näitlikustatakse siinkohal lugejale meetodi sisendit ja väljundit. Sisendiks on ühe lekseemi muutvormitabel tervikuna (tabel 2). Väljundiks on meetodi poolt leitud tehniline tüvi ja sõnatüübimall, mille põhjal on võimalik rekonstrueerida sisendi muutvormitabel (tabel 3). Tabelitele viidatakse alljärgnevas tekstis mitmel korral.

muutvorm	tunnused
<i>katto</i>	SG NOM
<i>katod</i>	PL NOM
<i>kato</i>	SG GEN
<i>kattojõ</i>	PL GEN
<i>kattoa</i>	SG PART
<i>kattoit</i>	PL PART
<i>kattose</i>	SG ILL
<i>kattoisõ</i>	PL ILL
<i>kattoz</i>	SG INE
<i>kattoiz</i>	PL INE
<i>katossõ</i>	SG ELA
<i>kattoissõ</i>	PL ELA
<i>katollõ</i>	SG ALL
<i>kattoillõ</i>	PL ALL
<i>katol</i>	SG ADE
<i>kattoil</i>	PL ADE
<i>katoltõ</i>	SG ABL
<i>kattoiltõ</i>	PL ABL
<i>katossi</i>	SG TRAN
<i>kattoissi</i>	PL TRAN
<i>kattossaa</i>	SG TERM
<i>kattoissaa</i>	PL TERM
<i>katoka</i>	SG COM
<i>kattoika</i>	PL COM

Tabel 2: Sisendi muutvormide tabel koos morfoloogiliste tunnustega.

tehniline tüvi	muutvormimall	tunnused
<u>kat</u> t o	$x_1 + t + x_2$	SG NOM
<u>kat</u> o d	$x_1 + x_2 + d$	PL NOM
<u>kat</u> o	$x_1 + x_2$	SG GEN
<u>kat</u> t o jõ	$x_1 + t + x_2 + jõ$	PL GEN
<u>kat</u> t o a	$x_1 + t + x_2 + a$	SG PART
<u>kat</u> t o ite	$x_1 + t + x_2 + it$	PL PART
<u>kat</u> t o sõ	$x_1 + t + x_2 + sõ$	SG ILL
<u>kat</u> t o isõ	$x_1 + t + x_2 + isõ$	PL ILL
<u>kat</u> t o z	$x_1 + t + x_2 + z$	SG INE
<u>kat</u> t o iz	$x_1 + t + x_2 + iz$	PL INE
<u>kat</u> o sse	$x_1 + x_2 + ssõ$	SG ELA
<u>kat</u> t o issõ	$x_1 + t + x_2 + issõ$	PL ELA
<u>kat</u> o llõ	$x_1 + x_2 + llõ$	SG ALL
<u>kat</u> t o illõ	$x_1 + t + x_2 + illõ$	PL ALL
<u>kat</u> o l	$x_1 + x_2 + l$	SG ADE
<u>kat</u> t o il	$x_1 + t + x_2 + il$	PL ADE
<u>kat</u> o ltõ	$x_1 + x_2 + ltõ$	SG ABL
<u>kat</u> t o iltõ	$x_1 + t + x_2 + iltõ$	PL ABL
<u>kat</u> o ssi	$x_1 + x_2 + ssi$	SG TRAN
<u>kat</u> t o issi	$x_1 + t + x_2 + issi$	PL TRAN
<u>kat</u> t o ssaa	$x_1 + t + x_2 + ssaa$	SG TERM
<u>kat</u> t o issaa	$x_1 + t + x_2 + issaa$	PL TERM
<u>kat</u> o ka	$x_1 + x_2 + ka$	SG COM
<u>kat</u> t o ika	$x_1 + t + x_2 + ika$	PL COM

Tabel 3: Väljundi tüüpsõnamall (kusjuures  $x_1 = kat$  ja  $x_2 = o$  vastab sisendist leitud ühisosajadale).

Ekstraktmorfoloogia meetod eraldab sisendtabelist sõnatüübimalli ja tehnilise tüve lekseemi muutvormide tähtkoostise põhjal, s.o sõnavormide tähtede põhjal. Meetod arvestab palatalisatsiooniga kuivõird vadjä ortograafias seda märgitakse, rõhuga meetod ei arvesta, kuna on sisendis märkimata. Silpide arvuga meetod ei arvesta.

Tehniliseks tüveks loetakse sõna need tähtede jadad, mis esinevad (korduvad) üle kõigi selle muutvormide (allajoonitud tähed tabelis 3, veerg 1). Tehniline tüvi koosneb ühest või enamast osast (tähtede jadast), antud näite puhul kahest. Samasse sõnatüüpi kuuluvad lekseemid erinevad vaid oma tehnilise tüve poolest. Sõnatüübi muutvorme eristab teine-teisest see tähtkoostis, mis ei kordu üle kõigi muutvormide (vrd morfeemi definitsioon pt 2.3).

Tehnilise tüve leidmine koosneb kahest etapist: 1) pikima ühisosajada eraldamine ning 2) ühisosajada ühene jaotamine muutvormide vahel tehniliseks tüveks. Sõnatüüp koosneb muutvormimallidest ja koostatakse tehnilise tüve jaotuse põhjal, kus-

juures üldistatakse tehnilise tüve osad muutujateks ja sõnatüüp funktsiooniks.

**Tehnilise tüve eraldamine** Tehniline tüvi on ekstraktmorfoloogias defineeritud kordse pikima ühisosajadana (*Multiple Longest Common Subsequence*). Näites (tabel 2) esinevad üle kõigi muutvormide järgmised tähed: *k*, *a*, *t* ja *o*. Aga kuna *t*-sid esineb eri muutvormides rohkem kui üks, ei ole selge millele neist vastab pikima ühisosajada täht *t*. Näite puhul on võimalik kaks pikima ühisosajada jaotust: *kat* ja *o* või *ka* ja *to*.

Võimalike mitmesuste lahendamiseks kasutatakse ekstraktmorfoloogias ühestamise heuristikat. Ühestamine eelistab mh pikemat esiosa (Hulden 2014, lk. 33), mistõttu valitakse näite puhul tehniliseks tüveks osad *kat* ja *o*.

**Üldistamine sõnatüübi funktsiooniks** Tehnilise tüve fikseerimise järel moodustatakse muutvormimallid. Muutvormides asendatakse tehnilise tüve osad muutujatega (tabelis 3, veerg 2). Näiteks

$$katto \rightarrow \underline{kat} + t + \underline{o} \rightarrow x_1 + t + x_2$$

$$katod \rightarrow \underline{kat} + \underline{o} + d \rightarrow x_1 + x_2 + d$$

vastavalt SG NOM ja SG GEN vormidele.

On ilmne et, kui asendatakse muutvormimallides muutujad vastavate tehnilise tüve osadega, siis taas-moodustatakse muutvormid *katto* ja *katod*. Muutujaid aga asendades teiste tähtedega, nt *čiut* ja *o*, moodustatakse teise lekseemi, *čiutto* käändetabel. Näiteks

$$x_1 + t + x_2 \rightarrow \underline{čiut} + t + \underline{o} \rightarrow čiutto$$

$$x_1 + x_2 + d \rightarrow \underline{čiut} + \underline{o} + d \rightarrow čiutod$$

vastavalt SG NOM ja SG GEN muutvormimallidele.

Järelikult kuuluvad mõlemad lekseemid sama sõnatüübi alla ja erinevad ainult oma tehnilise tüve poolest.

Lekseemi iseloomustab ekstraktmorfoloogias sõnatüüp ja tehniline tüvi. Sõna-

tüüp on funktsioon, mille muutuja(te)ks on tehniline tüvi. Sõnatüübi muutvorme eristab teine-teisest see tähtkoostis, mis ei kordu üle kõigi muutvormide (vrd morfeemi definitsioon pt 2.3).

Kuna see vajab spetsiifilist teadmist tehnilise tüve moodustamise kohta, ei ole tehnilisel tüvel opereeriv funktsioon sobilik praktiliseks inim-kasutamiseks. On võimalik moodustada kasutamist hõlbustav funktsioon, mille sisendiks on terve sõnavorm.

Tehniline tüvi ei sõltu definitsiooni järgi ühest konkreetsest muutvormist, vaid lekseemi tervest muutvormitabelist (paradigmast). Seetõttu on ekstraktmorfoloogia lemma suhtes neutraalne. Lemma valik on teisejärguline ja seda on võimalik hiljem vahetada.

Kui lemmaks on valitud kindel muutvorm (nt SG NOM), on võimalik koostada funktsioon, mille sisendiks on muutvorm tervikuna ja väljundiks on tehniline tüvi. Funktsiooni koostamise aluseks on lemmaks valitud vormi muutvormimall, mille järgi sisend-sõnavormi tähtkoostis jaotatakse ühestamisheuristika põhjal (Ahlberg, Forsberg ja Hulden 2014, lk. 572). Niiviisi saab aheldada lemmavormi sisendiks võtva funktsioon kokku tehnilise tüve sisendiks võtva funktsiooniga ja moodustada kasutust hõlbustava funktsiooni, mille sisendiks on sõna lemmavorm ja väljundiks muutvormide tabel. Seda lähenemisviisi kasutatakse magistritöös Grammatical Framework programmikoodi tuletamise juures (pt 6.2).

## 4.2 Sõnatüübi ennustamine

Sõnatüübi ennustamine on eraldiseisev protsess ja põhineb ainult eraldatud tehniliste tüvede muutujate tähtkoostise analüüsimisel. Ennustamine koosneb piirangute seadmisest tehnilise tüve muutujate kujule.

Sõnatüübi ennustamismeetodit on magistritöös kasutatud vaid kaudselt, Morfoloogialabori veebileidese kaudu vormisõnastik koostamisel. Seda on tehtud kuna magistritöö eesmärgiks on olnud esmane integreerimine keeletehnoloogilistesse raamistikesse ja õigekirjakontrollija loomine, kus ennustamisvõimet ei ole otseselt tarvis. Ennustamismeetod sarnaneb *Grammatical Framework*'i *Smart Paradigm*-funktsiooni

mõistega (Détrez ja Ranta 2012) ja on magistritöö üks võimalik edasiarenduse suund. Forsberg ja Hulden (Forsberg ja Hulden 2016a) on rakendanud meetodit tundmatu sõnavormi sõnatüübi ennustaja (ingl. *morphological guesser*) ehitamiseks, mida Morfoloogialabori veebileides kasutab sõnatüüpide järjestamiseks kasutajale.

Vormisõnastikku salvestatud lekseemid kannavad informatsiooni tehnilise tüve ja sõnatüübimalli kohta. Mida rohkem lekseeme jagab üht ja sama sõnatüübi malli, seda rohkem on informatsiooni selle sõnatüübi tehniliste tüvede kujude kohta.

Tehnilised tüved ei jaotu sõnatüüpide vahel arbitraarselt, vaid nende kuju peegeldab tihti morfofonoloogilisi nähtuseid ja ortograafilisi tavasid. Seda informatsiooni on võimalik kasutada selleks, et seada sõnatüübimallide muutujatele kitsendusi ja piirata sellega ühtivate tehniliste tüvede hulka, ning tundmatu sõnavormi puhul ennustada selle kuulumist ühe või teise sõnatüüpi alla. (Forsberg ja Hulden 2016a, lk. 2579).

Forsberg ja Hulden kirjeldavad statistilist viisi, millega nad loovad tehnilise tüve kuju tähtkoostist kitsendava regulaarse grammatika. Statistik arvutab salvestatud andmete põhjal piirmäära, mida mitte ületades loetakse tehnilise tüve tähtkoostise algus- või lõpposa (või terviklikult) suletud klassiks. Suletud klassi puhul peab leitud tähtkoostis antud positsioonis esinema, et lubada tundmatu sõnavorm ühtida sõnatüübiga. (Forsberg ja Hulden 2016a, lk. 2580).

Meetodi näitlikustamiseks võib magistritöö vältel koostatud vormisõnastiku põhjal märkida seda, et kõik *katto*-sõnatüübi alla kuuluvad lekseemid moodustaksid suletud klassi: tehnilise tüve esimene muutuja peab lõppema *t*-tähega: *juttu*, *hattu*, *katto*, *kuttsu*, *laatto*, *lanttu*, *paltto*, *porttu*, *Tarttu*, *čiutto*.

#### 4.3 Võrdlus teiste meetoditega ja Eesti leksikograafilise traditsiooniga

Tavapärane on arvutimorfoloogia koostada reeglite kirjutamise kaudu, seejuures teha kõigepealt klassifikatsioon, mille järgi on leida kõik eri klassid elik üksused milledele rakenduvad samad reeglid. Seda on teinud nt Viks eesti arvutimorfoloogia koostamiseks (mh Viks 1976) ja sarnase viisi metodoloogiat on püüdnud formaliseerida

Koskenniemi (Koskenniemi 2013).



## 5 Vadja morfoloogiliste tüüpsõnade analüüs

See osa kirjeldab ekstraktmorfoloogiaga leitud vadja keele morfoloogilised sõnatüübid ja jaotab need käändkondadesse. Käändkondade süsteemiks on kasutatud Ariste 1968. Tsvetkovi sõnaraamatus esinevat variatsiooni on analoogia põhjal ühtlustatud kirjakeele lihtsama õppimise eesmärgil. Peatüki viimases osas analüüsitakse milmoel Silfverberg, Liu ja Hulden 2018 esitatud ekstraktmorfoloogia üldiste muut tüüpide algoritm kajastab käändkondi.

Ariste käändkonnad põhinevad muutustel, mis kajastuvad järgmistes käändeis: SG NOM ja PL NOM, SG GEN ja PL GEN, SG PAR ja PL PAR, SG ILL ja PL ILL ning SG ELA ja PL ELA Ariste 1968, lk. 42.

Üle käändkonnniti on rekonstrueeritud lõpukaolised vokaalid ja ühtlustatud peamiselt lõpuvokaali õ:a vaheldumine. Detailsemalt on ühtlustatud komponente kirjeldatud iga käändkonna juures.

Noomeni sõnavarast on välja jäetud komparatiivsed vormid (mõlõpi, vanepi).

Alljärgnevat kirjeldustes on lisatud sõnatüübi muutvormimallid kujul, kus tehniline tüvi on allajoonitud. Selline esitus võimaldab lugejal paremini näha ekstraktmorfoloogilise muutvormimalli funktsioneerimist läbi.

### 5.1 I käändkond

Esimesse käändkonda kuuluvad (Ariste 1968, lk. 40) järgi ühetüvelised, ühesilbilised sõnad. Ekstraktmorfoloogia eraldab kõik sellesse käändkonda kuuluvad sõnatüübid, aga koondab ka mitmesilbilised liitsõnad, mille järelkomponent kuulub siia käändkonda (nt *pihlpuu* nagu *puu*, *bulipää* nagu *pää*).

SG PAR vormidele on lisatud lõpuhäälilik -õ või -e vastavalt sõna vokalismile.

Koostatud vormisõnastiku sõnade SG ILL vorme on esialgselt üsna kunstlikult ühtlustatud kirjakeele jaoks: *pää:pähhe*, *puu:puhhu* ja *maa:mahha* ning diftongiga sõnade *puhul* või:*võisõ*, *täi:täise*.

Näide Tsvetkovi sõnaraamatus esinevast SG ILL vormi variatsioonist: *kuu:kuusõ*; *üü:ühese*; *vüü:vühe~vühese~vüüse*; *püü:pühe~pühese*.

Ekstraktmorfoloogia eraldatud sõnatüüpide paljus on ajendatud ühelt poolt vokalismist (muutelõpu erinevus *või:võisõ* ja *täi:täise*). Teisalt aga pika türevokaali lühendamise tõttu i-mitmuse vormides (*pääd:päije* ja *tüüd:tüije*), mistõttu jääb ühte sõnatüüpi -ä-lised muutelõpud ja teise -ü-lised.

Üks avatud küsimus ja tähelepanek on SG ja PL vormide kokkulangemine diftongiga sõnade puhul.

### Ekstraktmorfoloogia sõnatüübid

**maa** *maa, maatõ, mahha, maass, maad, maijõ, mait, maisõ, maiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid -aa-tüvelisi sõnu.

**puu** *puu, puutõ, puhhu, puuss, puud, puijõ, puit, puisõ, puiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *puu, kuu, luu, suu, pihlpuu*. Sõnatüübimall kirjeldab tagapoolseid -uu-tüvelisi sõnu.

**pii** *pii, piite, pihhe, piiss, piid, piije, piit, piise, piiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid -ii-tüvelisi sõnu.

**pää** *pää, pääte, pähhe, pääss, pääd, päije, päit, päise, päiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *pää, bulipää*. Sõnatüübimall kirjeldab eespoolseid -ää-tüvelisi sõnu.

**soo** *soo, sootõ, sohho, sooss, sood, soijõ, soit, soisõ, soiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid -oo-tüvelisi sõnu.

**tee** *tee, teete, tehhe, teess, teed, teije, teit, teise, teiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid -ee-tüvelisi sõnu.

**tüü** *tüü, tüüte, tühhe, tüüss, tüüd, tüije, tüit, tüise, tüiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *tüü, vüü, üü, püü*. Sõnatüübimall kirjeldab eespoolseid -üü-tüvelisi sõnu.

**täi** *täi, täite, täise, täiss, täid, täije, täit, täise, täiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid diftongiga sõnu.

või *või, võitõ, võisõ, võiss, võid, võijõ, võit, võisõ, võiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid diftongiga sõnu.

## 5.2 II käändkond

Teise käändkonda kuuluvad kahesilbilised sõnad, mille türevokaal on -o, -u, -ü, -i või -õ ning rohkem silpidega sõnad, mille türevokaal on -o Ariste 1968, lk. 42.

Sellesse käändkonda kuuluvad paljud vene laensõnad. Kirjakeele puhul puudub ülevaade vene laensõnade mugandamisest. Mugandamisstrateegiatest ja problemaatikast vajda Jõgõperä murdes on kirjutanud Rozhanskiy 2009.

### Ekstraktmorfoloogia sõnatüübid

auči *audži, aučia, aučisõ, audžiss, audžid, aučijõ, aučiit, aučisõ, aučiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega č:dž.

süüčči *süüči, süüččiä, süüččiise, süüččiss, süüčid, süüččiye, süüččiit, süüččiise, süüččiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega čč:č.

koffi *kofi, koffia, koffisõ, kofiss, kofid, koffijõ, koffiit, koffisõ, koffiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega ff:f.

suuto *suuo, suutoa, suutosõ, suuoss, suuod, suutojõ, suutoit, suutoisõ, suutoiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *suuto, vaahto, lehto*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega t:Ø.

vahti *vahi, vahtia, vahtisõ, vahiss, vahid, vahtijõ, vahtiit, vahtisõ, vahtiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega t:Ø ja mille lõpuvokaal on i.

alku *algu, alkua, alkusõ, alguss, algud, alkujõ, alkuit, alkuisõ, alkuiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *alku, lohko, pehko, plehku, touko, vihko, vinku, alko*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega k:g, kusjuures tüvemuutus esineb konsontklustris, mistõttu gemineerumist ei toimu SG PAR ja SG ILL tüvedes.

**lako** lago, lakkoa, lakkosõ, lagoss, lagod, lakojõ, lakoit, lakoisõ, lakoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *lako*, *luku*, *mako*, *maku*, *suku*, *vako*, *čako*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega k:g, kusjuures tüvi gemineerub SG PAR ja SG ILL vormides.

**läikki** läiki, läikkiä, läikkise, läikiss, läikid, läikkije, läikkeit, läikkeise, läikkeiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega kk:k ja mille lõpuvokaal on *i*.

**tükkü** tükü, tükküä, tükküse, tüküss, tüküd, tükküje, tükküit, tükküise, tükküiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega kk:k.

**viki** viki, vikkiä, vikkise, vikiss, vikid, vikkije, vikkiit, vikkiise, vikkiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega kk:k ja mille lõpuvokaal on *i*.

**flakku** flaku, flakkua, flakkusõ, flakuss, flakud, flakkujõ, flakkuit, flakkuisõ, flakkuiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *flakku*, *herkku*, *jõkilikko*, *kakko*, *kakku*, *kiikku*, *kolkku*, *kukko*, *kurkku*, *kuuzikko*, *lepikko*, *liivikko*, *luikko*, *lukku*, *lõõkku*, *majakko*, *musikko*, *mäčizikko*, *naizikko*, *oomnikko*, *pettelikko*, *rehtelkakku*, *seukko*, *võrkko*, *õzrikko*, *čerikko*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega kk:k.

**galstukki** galstuki, galstukkia, galstukkisõ, galstukiss, galstukid, galstukkijõ, galstukkiiit, galstukkiiisõ, galstukkiiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *galstukki*, *kokki*, *kolkki*, *luukki*, *pukki*, *vokki*, *fraakki*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega kk:k ja mille lõpuvokaal on *i*.

**põlto** põllo, põltoa, põltosõ, põlloss, põllod, põltojõ, põltoit, põltoisõ, põltoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *põlto*, *mõlto*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega lt:ll.

**greebeni** greebeni, greebeniä, greebenise, greebeniss, greebenid, greebenije, greebeniit, greebeniise, greebeniiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *greebeni*, *Helsengi*, *jevi*, *kiikeri*, *kiisseli*, *meebeli*, *näteli*, *Reeveli*, *retsepti*, *rööveli*, *špeili*, *väli*, *vääri*, *ängeli*, *bibli*. Sõnatüübimall kirjeldab eespoolseid tüvemuutuseta sõnu ja mille lõpuvokaal on *i*.

**lõülü** lõülü, lõülüä, lõülüse, lõülüss, lõülüd, lõülüje, lõülüit, lõülüise, lõülüiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *lõülü*, *süčüzü*, *jürü*. Sõnatüübimall kirjeldab eespoolseid tüvemuutuseta sõnu.

**airo** airo, airoa, airosõ, aiross, airod, airojõ, airoit, airoisõ, airoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *airo*, *aju*, *anõ*, *čaažu*, *čello*, *elo*, *haadu*, *heeno*, *hlaamu*, *ilo*, *javo*, *jõulu*, *kahu*, *kalmo*, *karu*, *kehno*, *kirstu*, *koivu*, *konno*, *laulu*, *lello*, *morško*, *murru*, *nagru*, *ohtõgo*, *paju*, *paksu*, *pallo*, *passibo*, *pojo*, *saadu*, *savvu*, *siivo*, *škoulu*, *talo*, *varjo*, *vello*, *vilu*, *viro*, *vooro*, *õhtõgo*, *ahjo*. Sõnatüübimall kirjeldab tagapoolseid tüvemuutusega sõnu.

**bagaži** bagaži, bagažia, bagažisõ, bagažiss, bagažid, bagažijõ, bagažiit, bagažiisõ, bagažiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *bagaži*, *balhoni*, *baroni*, *biljardi*, *bobuli*, *buljoni*, *divani*, *dohtõri*, *farfori*, *flaneli*, *gimnazi*, *gitari*, *glazi*, *haili*, *inspektori*, *itkuri*, *jaani*, *kammõri*, *kongressi*, *kuhni*, *lusti*, *makarooni*, *malari*, *mandõri*, *naapuri*, *nojaabri*, *nuumõri*, *paperi*, *plaastõri*, *pošti*, *stooli*, *suukkuri*, *taari*, *tormi*, *tunni*, *vagzõli*, *vari*, *vinkuri*, *almõzi*. Sõnatüübimall kirjeldab tagapoolseid tüvemuutusega sõnu ja mille lõpuvokaal on *i*.

**poštaljon** poštaljoni, poštaljonია, poštaljonisõ, poštaljoniss, poštaljonid, poštaljonijõ, poštaljoniit, poštaljoniisõ, poštaljoniiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *poštaljon*, *parad*. Sõnatüübimall kirjeldab tagapoolseid tüvemuutusega sõnu.

**sünti** sünni, süntiä, süntise, sünniss, sünnid, süntije, süntiit, süntiise, süntiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega nt:nn ja mille lõpuvokaal on *i*.

**lento** lenno, lentoa, lentosõ, lennoss, lennod, lentojõ, lentoit, lentoisõ, lentoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *lento*, *lintu*, *rokkalintu*, *kanto*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega nt:nn.

**vipu** vivu, vippua, vippusõ, vivuss, vivud, vipujõ, vipuit, vipuisõ, vipuiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega p:v.

**hapo** hapo, happoa, happosõ, haposs, hapod, happojõ, happoit, happoisõ, happoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega p:pp.

**vilppi** vilpi, vilppiä, vilppise, vilpiss, vilpid, vilppije, vilppiit, vilppiise, vilppiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *vilppi*, *šlääppi*. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega pp:p ja mille lõpuvokaal on *i*.

**hüppü** hüpü, hüppüä, hüppüse, hüppüss, hüppüd, hüppüje, hüppüit, hüppüise, hüppüiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega pp:p.

**lippu** lipu, lippua, lippusõ, lipuss, lipud, lippujõ, lippuit, lippuisõ, lippuiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *lippu*, *lõppu*, *puippu*, *kippu*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega pp:p.

**lamppi** lampi, lamppia, lamppisõ, lampiss, lampid, lamppijõ, lamppiit, lamppiisõ, lamppiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *lamppi*, *pappi*, *suppi*, *ukroppi*, *kaappi*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega pp:p ja mille lõpuvokaal on *i*.

**sese** seze, sesseä, sessese, sezess, sezed, seseje, seseit, seseise, seseiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *sese*, *läsü*. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega s:z.

**siso** sizo, sissoa, sissoisõ, sizoss, sizod, sisojõ, sisoit, sisoisõ, sisoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *siso*, *nisu*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega s:z.

**mahsu** mahzu, mahsua, mahsusõ, mahzuss, mahzud, mahsujõ, mahsuit, mahsuisõ, mahsuiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *mahsu*, *haisu*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega s:z, mille tüvi konsonantklustri tõttu ei gemineeru.

**kursi** kurzi, kursia, kursisõ, kurziss, kurzid, kursijõ, kursiit, kursiisõ, kursiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega s:z, mille tüvi konsonantklustri tõttu ei gemineeru ja mille lõpuvokaal on *i*.

**rusko** ruzgo, ruskoa, ruskosõ, ruzgoss, ruzgod, ruskojõ, ruskoit, ruskoisõ, ruskoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *rusko*, *tuisku*, *usko*, *pääsko*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega sk:zg.

**rissi** risi, rissiä, rissise, risiss, risid, rissije, rissiit, rissiise, rissiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega ss:s ja mille lõpuvokaal on *i*.

**passi** pasi, passia, passisõ, passiss, pasid, passijõ, passiit, passiisõ, passiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *passi*, *komissi*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega ss:s ja mille lõpuvokaal on *i*.

**karjušši** karjuši, karjuššia, karjuššisõ, karjušsiss, karjušid, karjuššijõ, karjuššiit, karjuššiisõ, karjuššiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *karjušši*, *latõšši*, *potašši*, *falšši*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega šš:š ja mille lõpuvokaal on *i*.

**täti** tädi, tättiä, tättise, tädiss, tädid, tätije, tätiit, tätiise, tätiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega *t:d* ja mille lõpuvokaal on *i*.

**kotko** kodgo, kotkoa, kotkosõ, kodgoss, kodgod, kotkojõ, kotkoit, kotkoisõ, kotkoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *kotko*, *laatko*, *itku*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega *tk:dg*.

**kittsi** kitsi, kittsiä, kittsise, kitsiss, kitsid, kittsije, kittsiit, kittsiise, kittsiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *kittsi*, *komferenttsi*, *pletti*, *biletti*. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega *tt:t* ja mille lõpuvokaal on *i*.

**rätte** räte, rätteä, rättese, rätess, räted, rätteje, rätteit, rätteise, rätteiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *rätte*, *tüttõ*, *nenärätte*. Sõnatüübimall kirjeldab eespoolseid sõnu tüvemuutusega *tt:t*.

**hattu** hatu, hattua, hattusõ, hatuss, hatud, hattujõ, hattuit, hattuisõ, hattuiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *hattu*, *juttu*, *katto*, *kuttsu*, *laatto*, *lanttu*, *paltto*, *porttu*, *Tarttu*, *čiutto*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega *tt:t*.

**dokumentti** dokumenti, dokumenttia, dokumenttisõ, dokumenttiss, dokumentid, dokumenttijõ, dokumenttiit, dokumenttiisõ, dokumenttiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *dokumentti*, *fabrikantti*, *Franttsi*, *fundamentti*, *kajutti*, *kametti*, *kanfetti*, *katti*, *kometti*, *komfetti*, *komnõtti*, *lautti*, *magnetti*, *minutti*, *muzõkantti*, *protestantti*, *protsentti*, *Rootsi*, *bankrutti*. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega *tt:t* ja mille lõpuvokaal on *i*.

**komifet** komifeti, komifettiä, komifetti, komifetiss, komifetid, komifettije, komifettiit, komifettiise, komifettiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus. Sõnatüübimall kirjeldab tagapoolseid sõnu tüvemuutusega *tt:t*.

**famili** famila, famila, familasõ, familõss, familõd, familojõ, familoit, familoisõ, familoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *famili*, *bašni*. Sõnatüübimall kirjeldab tagapoolseid tüvemuutuseta sõnu, mis eeldatavasti palatalisatsiooni tõttu käituvad eriliselt.

### 5.3 III käändkond

Kolmandasse käändkonda kuuluvad kahesilbilised sõnad, mille tüvevokaal on *-a* ning rohkem silpidega sõnad, millel esineb esimeses silbis *-a-*, *-õ-* või *-i-* Ariste 1968, lk. 42.

Avatuid küsimusi-tähelepanekuid:

- siin on sõnu millel on *-u-* 1. silbis, need peaksid käima hoopis V alla
- Tsvetkovil palju *-õi-*mitmusetüvega, need on ühtlustatud *-oi-*
- paljud laensõnad kuuluvad siia alla, nende lõpuvokaalidega on häda

#### Ekstraktmorfoloogia sõnatüübid

**tuhattõ** tuhatta, tuhatta, tuhattasõ, tuhatõss, tuhatõd, tuhattojõ, tuhattoit, tuhattoisõ, tuhattoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**lafkõ** lavga, lafka, lafkasõ, lavgõss, lavgõd, lafkojõ, lafkoit, lafkoisõ, lafkoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**mahsõ** mahza, mahsa, mahsasõ, mahzõss, mahzõd, mahsojõ, mahsoit, mahsoisõ, mahsoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**vihtõ** viha, vihta, vihtasõ, vihõss, vihõd, vihtojõ, vihtoit, vihtoisõ, vihtoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**aitõ** ajja, aita, aitasõ, aijõss, aijõd, aitojõ, aitoit, aitoisõ, aitoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**aikõ** aiga, aika, aikasõ, aigõss, aigõd, aikojõ, aikoit, aikoisõ, aikoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *aikõ*, *jalkõ*, *liikõ*, *lõnkõ*, *nahkõ*, *rahkõ*, *vilkõ*, *vinkõ*, *võlkõ*, *aastaikõ*.

**sika** siga, sikka, sikkasõ, sigass, sigad, sikojõ, sikoit, sikoisõ, sikoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.



**borovikkõ** borovika, borovikka, borovikkasõ, borovikõss, borovikõd, borovikkojõ, borovikkoit, borovikkoisõ, borovikkoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *borovikkõ, domovikkõ, durakkõ, fartuk-kõ, fizikkõ, fookusnikkõ, frištikkõ, gribanikkõ, harakkõ, itikkõ, joožikkõ, kaamenšik-kõ, kabakkõ, kamalikkõ, katolikkõ, kelkkõ, koomikkõ, kopekkõ, latikkõ, luzikkõ, luukkõ, markkõ, muuzikkõ, mõiznikkõ, noorikkõ, nuužnikkõ, obakkõ, paikkõ, palkkõ, pinžak-kõ, podarkkõ, poštimarkkõ, rankkõ, rohosirkkõ, tarkkõ, tikkõ, tubakkõ, urokkõ, vakkõ, vunukkõ, bašmukkõ.*

**siltõ** silla, silta, siltasõ, sillõss, sillõd, siltojõ, siltoit, siltoisõ, siltoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**paha** paha, pahha, pahasõ, pahõss, pahõd, pahhojõ, pahhoit, pahhoisõ, pahhoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**azzõ** azža, azža, azzasõ, azzõss, azzõd, azžojõ, azžoit, azžoisõ, azžoiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *azzõ, baddõ, bahvõlõ, blaahõ, bobrõ, bork-kanõ, bruudõ, čirjavõ, čirjõ, deelõ, dobrõ, filmõ, glaizõ, grammõ, gribavihmõ, iivõ, ju-malõ, jurmõ, kabjõ, kaglõ, kagrõ, kajagõ, kambõlõ, kanavõ, karjõ, kassõ, katagõ, kava-lõ, kvartirõ, laddõ, ladvõ, lahjõ, lahnõ, lainõ, laivõ, liivõ, linnõ, listõ, maailmõ, maamõ, mahlõ, mannõ, marjõ, matalõ, metlõ, muragõ, mussõmarjõ, naglõ, nigrõ, niglõ, ostanof-kõ, paglõ, progonõ, pudrõ, pupuškõ, rauhõ, saappõgõ, sarjõ, saunõ, siglõ, sisavõ, slifkõ, summõ, surmõ, suukkurliivõ, sõbrõ, šuubõ, ženihõ, taičinõ, trubõ, vihmõ, vikahtõ, villõ, õravõ, õzrõ, akkunõ.*

**bank** banka, banka, bankasõ, bankõss, bankõd, bankojõ, bankoit, bankoisõ, bankoiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *bank, bluud, bluudõčk, boran, fartõl, fialk, figur, fortočk, frikadelk, golod, greettsin, gupk, invaliid, kaban, kamal, kamin, kanal, kipun, kluub, kohin, lihoratk, mašin, mašinist, muudõr, omõn, pagan, peñsioner, sammõl, zanavesk, žurnalist, tarelk, vaahtõr, viks, ahvõn.*

**baldõhina** baldõhina, baldõhina, baldõhinasõ, baldõhinass, baldõhinad, baldõhinojõ, baldõhinoit, baldõhinoisõ, baldõhinoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *baldõhina, barabana, fotokartočka, grana, griba, kala, kana, liha, lina, litra, maja, raha, suma, sõna, tara, telefona, televizora, tila, vana, astia.*

**pintõ** pinna, pinta, pintasõ, pinnõss, pinnõd, pintojõ, pintoit, pintoisõ, pintoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *pintõ, rantõ, rintõ, kantõ.*

**rapa** rava, rappa, rappasõ, ravass, ravad, rapojõ, rapoit, rapoisõ, rapoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *rapa, sõpa, napa.*

**aapõ** aava, aapa, aapasõ, aavõss, aavõd, aapojõ, aapoit, aapoisõ, aapoiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**liippõ** liipa, liippa, liippasõ, liipõss, liipõd, liippojõ, liippoit, liippoisõ, liippoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: liippõ, kauppõ.

**järčü** järjü, järčüä, järčüse, järjüss, järjüd, järčüje, järčüit, järčüise, järčüiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**partõ** parra, parta, partasõ, parrõss, parrõd, partojõ, partoit, partoisõ, partoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: partõ, kõrtõ.

**kraaskõ** kraazga, kraaska, kraaskasõ, kraazgõss, kraazgõd, kraaskojõ, kraaskoit, kraaskoisõ, kraaskoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: kraaskõ, laiskõ, nagriskaaskõ, niskõ, pas-kõ, kaaskõ.

**klaassõ** klaasa, klaassa, klaassasõ, klaasõss, klaasõd, klaassojõ, klaassoit, klaassoisõ, klaassoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: klaassõ, bruussõ.

**poduškõ** podužga, poduška, poduškasõ, podužgõss, podužgõd, poduškojõ, poduškoit, poduškoisõ, poduškoiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**dovariššõ** dovariša, dovarišša, dovariššasõ, dovariššass, dovarišad, dovariššojõ, dovariššoit, dovariššoisõ, dovariššoiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**sata** saa, satta, sattasõ, saass, saad, satojõ, satoit, satoisõ, satoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: sata, sõta, muta.

**liitkõ** liidga, liitka, liitkasõ, liidgõss, liidgõd, liitkojõ, liitkoit, liitkoisõ, liitkoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: liitkõ, butkõ.

**kandidaattõ** kandidaata, kandidaatta, kandidaattasõ, kandidaatõss, kandidaatõd, kandidaattojõ, kandidaattoit, kandidaattoisõ, kandidaattoiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: kandidaattõ, lauttõ, pliittõ, riittõ, žiivõttõ, taattõ, gaazapliittõ.

**inostranttsõ** inostrantsa, inostranttsa, inostranttsasõ, inostrantsõss, inostrantsõd, inostranttsõjõ, inostranttsõit, inostranttsõisõ, inostranttsõiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *inostranttsõ*, *liittsõ*, *tablittsõ*, *vattsõ*, *õttsõ*, *bolnittsõ*.

**jõutu** jõvvu, jõutua, jõutusõ, jõvvuss, jõvvud, jõutujõ, jõutuit, jõutuisõ, jõutuiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**poutõ** povva, pouta, poutasõ, povvõss, povvõd, poutojõ, poutoit, poutoisõ, poutoiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *poutõ*, *lautõ*.

**štanad** štanojõ, štanoit, štanoisõ, štanass, štanad, štanojõ, štanoit, štanoisõ, štanoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.4 IV käändkond

Neljandasse käändkonda kuuluvad mitmed sõnad, mis on ainsuses eespoolse vokalismiga, ent mitmuses on tagapoolsed Ariste 1968, lk. 43. Selliseid sõnu Heinsoo loodavas kirjakeeles ei esine (isiklik kommunikatsioon).

## 5.5 V käändkond

Viiendasse käändkonda kuuluvad kahesilbilised sõnad, mille tüvevokaal on *-a* ja millel esineb esimeses silbis *-o-*, *-u-* või *-õ-*. Kattumise kohta III käändkonna sõnadega, mille esimene silp sisaldab *-õ-*, mainib Ariste, et enamik neist kuulub siia. Ariste 1968, lk. 44.

Avatuid küsimusi-tähelepanekuid:

- 5 käändkonna liikmed Aristel *-õi-* on suuresti muudetud *-ii-*
- 'mussõ' leiti mitu pl 'mussii' VKSi näitelauseste hulgast

## Ekstraktmorfoloogia sõnatüübid

**kuha** kuha, kuha, kuhasõ, kuhõss, kuhõd, kuhhijõ, kuhhiit, kuhhiisõ, kuhhiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**poikõ** poiga, poika, poikasõ, poigõss, poigõd, poikijõ, poikiit, poikiisõ, poikiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *poikõ*, *rookõ*, *lõukõ*.

**biblioteekkõ** biblioteeka, biblioteekka, biblioteekkasõ, biblioteekõss, biblioteekõd, biblioteekkijõ, biblioteekkiit, biblioteekkiisõ, biblioteekkiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *biblioteekkõ*, *hoikkõ*, *ikolookkõ*, *jaanikuk-kõ*, *kolkkõ*, *konjõkkõ*, *kukkõ*, *rokkõ*, *sukkõ*, *bambukkõ*.

**kõik** kõika, kõikka, kõikkasõ, kõikõss, kõikõd, kõikkijõ, kõikkiit, kõikkiisõ, kõikkiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kultõ** kulla, kulta, kultasõ, kullõss, kullõd, kultõjõ, kultõit, kultõisõ, kultõiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**vimpõ** vimma, vimpa, vimpasõ, vimmõss, vimmõd, vimpijõ, vimpiit, vimpiisõ, vimpiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *vimpõ*, *kumpõ*.

**kompjutera** kompjutera, kompjutera, kompjuterasõ, kompjuterass, kompjuterad, kompjuterijõ, kompjuteriit, kompjuteriisõ, kompjuteriiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *kompjutera*, *kuja*, *loba*, *õmpõja*, *õpõttõja*, *izõra*.

**maamuna** maamuna, maamuna, maamunasõ, maamunass, maamunad, maamunõjõ, maamunõit, maamunõisõ, maamunõiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *maamuna*, *muna*, *kalindõra*.

**mokom** mokoma, mokoma, mokomasõ, mokomõss, mokomõd, mokomijõ, mokomiit, mokomiisõ, mokomiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**bukvõ** bukva, bukva, bukvasõ, bukvõss, bukvõd, bukvijõ, bukviit, bukviisõ, bukviiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *bukvõ*, *duumõ*, *formõ*, *ilmõ*, *koirõ*, *kon-nõ*, *kuhjõ*, *kuivõ*, *kummõ*, *kuuluvõ*, *kõrvõ*, *moodõ*, *muudrõ*, *mõizõ*, *obraazõ*, *postõ*, *prog-rammõ*, *roojõ*, *soolõ*, *sveezõ*, *trubačistõ*, *tuimõ*, *tuttavõ*, *velosipedõ*, *vohmõ*, *voolõ*, *võimõ*, *bombõ*.

**propkõ** probga, propka, propkasõ, probgõss, probgõd, propkijõ, propkiit, propkiisõ, propkiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *propkõ*, *skopkõ*, *jupkõ*.

**rooppõ** roopa, rooppa, rooppasõ, roopõss, roopõd, rooppõjõ, roopõit, roopõisõ, roopõiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kurp** kurva, kurpa, kurpasõ, kurvõss, kurvõd, kurpijõ, kurpiit, kurpiisõ, kurpiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**usa** *uza, ussa, ussasõ, uzass, uzad, usijõ, usiit, usiisõ, usiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õhsõ** *õhza, õhsa, õhsasõ, õhzõss, õhzõd, õhsijõ, õhsiit, õhsiisõ, õhsiiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *õhsõ, sorsõ*.

**luiskõ** *luizga, luiska, luiskasõ, luizgõss, luizgõd, luiskijõ, luiskiit, luiskiisõ, luiskiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**mussõ** *musa, mussa, mussasõ, musass, musad, mussõjõ, mussõit, mussõisõ, mussõiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *mussõ, kapussõ*.

**moškõ** *možga, moška, moškasõ, možgõss, možgõd, moškijõ, moškiit, moškiisõ, moškiiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *moškõ, krõškõ*.

**lootõ** *loovva, loota, lootasõ, loovvõss, loovvõd, lootijõ, lootiit, lootiisõ, lootiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**biskvittõ** *biskvita, biskvitta, biskvittasõ, biskvitõss, biskvitõd, biskvittajõ, biskvittiit, biskvittiisõ, biskvittiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.6 VI käändkond

Kuuendasse käändkonda kuuluvad Ariste sõnul need sõnad, mis lõppevad *-õa/-eä/-iä*.

Ta toob eraldi välja Jõgõperä murde erinevused üheainsa näitesõnaga. Ariste 1968, lk. 44. Vadja kirjakeeles püütakse järgida ... TODO kirjutada.

Avatuid küsimusi-tähelepanekuid:

- käändkonna liikmete pluurali tüved on ühtlustatud – kas jätta nii või taastada Tsvetkovi variatiivsus?

### Ekstraktmorfoloogia sõnatüübid

**kerkiä** *kerkiä, kerkiätä, kerkiäse, kerkiäss, kerkiäd, kerkije, kerkiit, kerkiise, kerkiiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *kerkiä, pehmiä, pimmiä, siitiä, jämiä*.

**terve** *terve, tervettä, tervese, tervess, terved, tervije, terveit, terveise, terveiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kanka** *kanka, kankatõ, kankasõ, kankass, kankad, kankõjõ, kankõit, kankõisõ, kankõiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *kanka, kõrka, maikka, makka, ruska, valka, õika, harma*.

## 5.7 VII käändkond

Seitsmendasse käändkonda kuuluvad kahesilbilised sõnad, mille SG NOM lõpp on *-i*, ent mille türevokaal on *-e/-õ* Ariste 1968, lk. 45.

Avatuid küsimusi-tähelepanekuid:

- 7 käändkonna kohta TODO kirjuta et isuri mõju tõttu on *-i:-e:-iä* levinud, aga normeerime nagu Aristel ja Tsvetkovil ka paralleelina tihti
- eespoolsed on *i:e:eä* ja tagapoolsed on *i:õ:õa*
- väci:väe aga mida teha lahti:lahe? – VKSis esineb Lu lahõ

### Ekstraktmorfoloogia sõnatüübid

**väci** *väe, väččeä, väččese, väess, väed, väččije, väččiit, väččiise, väččiiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *väci, mäci*.

**lahti** *lahõ, lahtõa, lahtõsõ, lahõss, lahõd, lahtijõ, lahtiit, lahtiisõ, lahtiiss*  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**laki** *lagõ, lakkõa, lakkõsõ, lagõss, lagõd, lakijõ, lakiit, lakiisõ, lakiiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *laki, nõki, jõki*.

**kurki** *kurgõ, kurkõa, kurkõsõ, kurgõss, kurgõd, kurkijõ, kurkiit, kurkiisõ, kurkiiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *kurki, õnki, kanki*.

**põski** *põzgõ, põskõa, põskõsõ, põzgõss, põzgõd, põskijõ, põskiit, põskiisõ, põskiiss*  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**irvi** *irve, irveä, irvese, irvess, irved, irvije, irviit, irviise, irviiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *irvi, järvi, leemi, nimi, pilvi, čivi*.

**sarvi** *sarvõ, sarvõa, sarvõsõ, sarvõss, sarvõd, sarvijõ, sarviit, sarviisõ, sarviiss*  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *sarvi, savi, Soomi, sõrmi, taimi, talvi, tammi, õnni, õvvi, põlvi*.

**enči** *enñe, enčëä, enčese, enñess, enñed, enčije, enčiit, enčiise, enčiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**siipi** *siive, siipeä, siipese, siivess, siived, siipije, siipiit, siipiise, siipiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kuusi** *kuuzõ, kuussõ, kuussõsõ, kuuzõss, kuuzõd, kuusijõ, kuusiit, kuusiisõ, kuusiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**tauti** *tavvõ, tautõa, tautõsõ, tavvõss, tavvõd, tautijõ, tautiit, tautiisõ, tautiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.8 VIII käändkond

Kaheksandasse käändkonda kuuluvad -ä-tüvelised sõnad Ariste 1968, lk. 46.

Avatuid küsimusi-tähelepanekuid:

- 8 käändkond on väga variatiivne türevokaali suhtes (eined, leived, čenned aga sepäd,
- eine (Heinsoo, Konkova ning Rozhanskiy ja Markus) aga einä (VKS)
- läikkiv on ühtlustatud läikkive
- Tsvetkovil paljud geminatsioonid puudu (õjja)
- tegija-liides on eespoolsete sõnade puhul ühtlustatud -jä:-jä:-jä, mitte -je:-jä:-jä, VKSis esineb rohkem -jä Lu/Li/J märgenditega (Konkoval eespoolseid sõnu ei esine)
- kuigi Tsvetkovil on häälduspäraselt ülesmärgitud 'õmpõlia' ja 'müüjä', on need läbivalt ühtlustatud (lisatud -j- nii SG NOM kui ka PL vormidele)
- tegija-liides on tagapoolsete sõnade puhul ühtlustatud -ja:-ja:-ja (kuigi Konkoval esineb -jõ:-ja:-ja)

### Ekstraktmorfoloogia sõnatüübid

**ičä** *iä, iččä, iččäse, iäss, iäd, ičije, ičiit, ičiise, ičiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**selče** sellä, selčä, selčäse, selless, selled, selčije, selčiit, selčiise, selčiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**eine** einä, einä, einäse, einess, eined, einije, eniit, eniise, eniiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *eine*, *esimespäive*, *fökle*, *irvieine*, *lehme*, *läikkive*, *läsive*, *nätilpäive*, *petäje*, *piime*, *pominpäive*, *pädre*, *päive*, *rehte*, *seine*, *silme*, *tühje*, *äije*, *ämme*, *čülme*.

**läkine** läkinä, läkinä, läkinäse, läkiness, läkined, läkineje, läkineit, läkineise, läkineiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *läkine*, *däade*.

**rissimä** rissimä, rissimä, rissimäse, rissimäss, rissimäd, rissimije, rissimit, rissimise, rissimiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *rissimä*, *emä*.

**pähčen** pähčenä, pähčenä, pähčenäse, pähčeness, pähčened, pähčenije, pähčeniit, pähčeniise, pähčeniiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *pähčen*, *ičäv*.

**räpäle** räpälä, räpälä, räpäläse, räpäless, räpäled, räpälöje, räpälöit, räpälöise, räpälöiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**läsijä** läsijä, läsijä, läsijäse, läsijess, läsijed, läsijije, läsijiit, läsijiise, läsijiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *läsijä*, *müüjä*, *tečejä*, *köühä*.

**kitai** kitaja, kitaja, kitajasõ, kitajass, kitajad, kitajojõ, kitajoit, kitajoisõ, kitajoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**slona** slona, slona, slonasõ, slonass, slonad, slonijõ, slonõit, slonõisõ, slonõiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õmpõlija** õmpõlija, õmpõlija, õmpõlijasõ, õmpõlijass, õmpõlijad, õmpõlijijõ, õmpõlijait, õmpõlijaisõ, õmpõlijaiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kompjutera** kompjutera, kompjutera, kompjuterasõ, kompjuterass, kompjuterad, kompjuterijõ, kompjuteriit, kompjuteriisõ, kompjuteriiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *kompjutera*, *kuja*, *loba*, *õmpõja*, *õpõttõja*, *ižora*.

**čenče** čennä, čenčä, čenčäse, čenness, čenned, čenčije, čenčiit, čenčiise, čenčiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.



**tünke** tüngä, tüнкä, tüнкäse, tüנגess, tünged, tüнкije, tüнкiit, tüнкiise, tüнкiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**änte** ännä, äntä, äntäse, änness, änned, äntije, äntiit, äntiise, äntiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**leipe** leivä, leipä, leipäse, leivess, leived, leipije, leipiit, leipiise, leipiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**seppe** sepä, seppä, seppäse, sepäss, sepäd, seppiije, seppiit, seppiise, seppiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *seppe*, *leppe*.

**ärče** ärjä, ärčä, ärčäse, ärjess, ärjed, ärčije, ärčiit, ärčiise, ärčiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *ärče*, *märče*.

**pesä** pezä, pessä, pessäse, pezäss, pezäd, pesije, pesiit, pesiise, pesiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *pesä*, *rissisä*, *česä*.

**lisä** lizä, lissä, lissäse, lizäss, lizäd, lissäije, lissäit, lissäise, lissäiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**mätä** mää, mättä, mättäse, määss, määd, mättije, mättiit, mättiise, mättiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**mettse** metsä, mettsä, mettsäse, metsäss, metsäd, mettsije, mettsiit, mettsiise, mettsiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**nenä** nenä, nennä, nenäse, nenäss, nenäd, nenije, neniit, neniise, neniiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**čülä** čülä, čüllä, čüläse, čüläss, čüläd, čülije, čüliit, čüliise, čüliiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**pühä** pühä, pühhä, pühhäse, pühäss, pühäd, pühhiije, pühhiit, pühhiise, pühhiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**üvä** üvä, üvvä, üvväse, üväss, üväd, üvvije, üvviit, üvviise, üvviiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õja** õja, õjja, õjasõ, õjass, õjad, õjijõ, õjit, õjisõ, õjiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õma** õma, õmma, õmmasõ, õmõss, õmõd, õmmijõ, õmmiit, õmmiisõ, õmmiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kõva** kõva, kõvat, kõvasõ, kõvass, kõvad, kõvijõ, kõviit, kõviisõ, kõviiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.9 IX käändkond

See käändkond on spetsiifiline Kattila murdele ja seda ei käsitleta siin töös.

## 5.10 X käändkond

Kümnendasse käändkonda koondub suur osa kahetüvelisi sõnu, mille SG NOM lõpp on -i, ent mille türevokaal on -e/-õ. Ariste märgib, et SP PAR on mitu erinevat realiseerimist, kuigi nende moodustamisviis põhimõtteliselt järgib sama malli. Ariste 1968, lk. 47.

Avatuid küsimusi-tähelepanekuid:

- 10 käändkond Ariste sõnul on sg par väga variatiivne
- ühtlustatud on -i:-õ:-tõ lõpuvokaalid
- kuigi voosi:voovvõ hääldub vuuvvõ on see märgitud voovvõ

## Ekstraktmorfoloogia sõnatüübid

**lumi** lumõ, luntõ, lumõsõ, lumõss, lumõd, lumijõ, lumiit, lumiisõ, lumiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**iiri** iire, iirte, iirese, iiress, iired, iirije, iiriit, iiriise, iiriiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: iiri, meeli, meri, peeni, süli, veri, ääni, ääri, čeeli.

**juuri** juurõ, juurtõ, juurõsõ, juurõss, juurõd, juurijõ, juuriit, juuriisõ, juuriiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: juuri, kaani, koori, lõhi, noori, ooli, pooli, sooli, suuri, tuli, tuuli, uni, hooli.

**uhsi** uhzõ, ussõ, uhsõsõ, uhzõss, uhzõd, uhsijõ, uhsiit, uhsiisõ, uhsiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: uhsi, lahsi.

**kusi** kuzõ, kussõ, kussõ, kuzõss, kuzõd, kussijõ, kussiit, kussiisõ, kussiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**uusi** uuvvõ, uttõ, uuvvõsõ, uuvvõss, uuvvõd, uusijõ, uusiit, uusiisõ, uusiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *uusi*, *voosi*, *kuusi*.

**ühs** ühe, ühte, ühtese, ühess, ühed, ühsije, ühsiit, ühsiise, ühsiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**ülči** ülle, ülčiä, ülčese, ülless, ülled, ülčije, ülčiit, ülčiise, ülčiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *ülči*, *jälči*. Sõnatüübimall kirjeldab taga-  
poolseid sõnu tüvemuutusega .

**kahs** kahõ, kahtõ, kahtõ, kahõss, kahõd, kahsijõ, kahsiit, kahsiisõ, kahsiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**varsi** varrõ, varttõ, vartõsõ, varrõss, varrõd, varsijõ, varsiit, varsiisõ, varsiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**mesi** mee, mette, mettese, meess, meed, mesije, mesiit, mesiise, mesiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *mesi*, *süsi*, *vesi*, *čäsi*.

**susi** suõ, suttõ, suttõsõ, suõss, suõd, susijõ, susiit, susiisõ, susiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.11 XI käändkond

Üheteistkümnendasse käändkonda kuuluvad need sõnad, mille sg nom lõpp on -Z, ent mille vokaaltüvi on -s- Ariste 1968, lk. 48.

Avatuid küsimusi-tähelepanekuid:

- -Z-lõpu sandhi nähtus on kõigi liikmete puhul ühtlustatud -z lõpulisteks
- kas seda peab mainima, et Jõgõperä murdes on -s-, Kattila murdes on -hs- ja teistes murretes on -ss-

### Ekstraktmorfoloogia sõnatüübid

**rihenneüz** rihenneüse, rihenneüsse, rihenneüsesse, rihenneüsess, rihenneüsed, rihenneüsije, rihenneüsiit, rihenneüsiise, rihenneüsiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**makuz** makusõ, makussõ, makuzõsõ, makuzõss, makuzõd, makusijõ, makusiit, makusiisõ, makusiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *makuz*, *nagriz*, *paganuz*, *kolauz*.

## 5.12 XII käändkond

Kaheteistkümnes käändkond koondab need sõnad, mille SG NOM lõpp on *-n/-ne/-nõ*, ent mille vokaaltüves on *-se-/-ze-/-sõ-/-zõ-* sõltuvalt astmevaheldusele Ariste 1968, lk. 49.

Avatuid küsimusi-tähelepanekuid:

- pl tüvi ühtlustatud *-s-* igal pool TODO üle vaadata *s:z* vaheldus pluuralis, kas see on *s* kui 1. silp on pikk v kinnine? (Tsvetkovil pole reeglipäraselt vaid variatsiooniline)
- kas pl gen peaks vahelduma *-z-* (iloin)? või *-s-* (keskolin)?
- talviisijõ, talviiziit
- õpõin on väga erandlik sõna

### Ekstraktmorfoloogia sõnatüübid

**õpõin** õpõizõ, õvõissõ, õpõzõsõ, õpõizõss, õpõizõd, õpõizijõ, õpõiziit, õpõiziisõ, õpõiziiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**entin** entize, entisse, entizese, entizess, entized, entisije, entisiit, entisiise, entisiiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *entin*, *esimein*, *inimin*, *jäin*, *miltäin*, *rehelin*, *sinin*, *venäläin*, *eglin*.

**süsiin** süsize, süsiisse, süsize, süsize, süsize, süsize, süsize, süsize, süsize, süsize  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õnnõliin** õnnõlizõ, õnnõliissõ, õnnõlizõsõ, õnnõlizõss, õnnõlizõd, õnnõlisijõ, õnnõlisiit, õnnõlisiisõ, õnnõlisiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**čimolain** čimolaizõ, čimolaissõ, čimolaizõsõ, čimolaizõss, čimolaizõd, čimolaisijõ, čimolaisiit, čimolaisiisõ, čimolaisiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *čimolain*, *greekklain*, *hatukkõin*, *iirikkõin*, *illõkkõin*, *iloin*, *jõkain*, *kehnoikkõin*, *keskolin*, *kõikõllain*, *kõrvõlin*, *leivekkõin*, *luin*, *lättilain*, *magnettiin*, *main*, *mokomõin*, *mustõlain*, *nain*, *partõin*, *perennain*, *prikukkõin*, *puin*, *roottsilain*, *ruskolain*, *saunlain*, *soomõlain*, *sopuin*, *sukulain*, *talviin*, *tarttulain*, *tõin*, *ukrainalain*, *virolain*, *õhtõgoin*, *ühellain*, *audžikkõin*.

**koivuin** koivuizõ, koivuissõ, koivuizõsõ, koivuizõss, koivuizõd, koivuizijõ, koivuiziit, koivuiziisõ, koivuiziiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *koivuin*, *kultõin*, *kõltõin*, *pakkõin*, *rohoin*, *uutin*, *voosin*, *kalttõin*.

### 5.13 XIII käändkond

Kolmeteistkümnendasse käändkonda kuuluvad need sõnad, mis lõpevad pika vokaa-  
liga SG NOM. Lisaks kuuluvad siia mõned sõnad, mis lõppevad diftongiga SG NOM.  
Ariste 1968, lk. 49.

Avatuid küsimusi-tähelepanekuid:

- Aristel pole *seemen* vaid on seemee:seemenee:seemeetä
- Tsvetkovil pole süä:süäme vaid on süä:süä:süttä/süät
- Konkoval on võttim:võttimõ:võttima (Tsvetkovil on näitelauses võti)

Veel kuuluvad siia käändkonda ordinaalid kolmest edasi Ariste 1968, lk. 50. Nu-  
meraalide puhul on järgitud Rozhanskiy ja Markuse välja toodud:

- SG NOM lõpp on *-iz*

#### Ekstraktmorfoloogia sõnatüübid

**čümme** čümmene, čümnenä, čümnenese, čümmess, čümmed, čümnenije, čümneniit, čümneniise, čümneniiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**čümnenäz** čümnenettemä, čümmeneii, čümnenettemäse, čümnenettemäss, čümnenettemäd, čümnenettemije, čümnenettemiit, čümnenettemiise, čümnenettemiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**nelläz** nellettemä, nellettä, nellettemäse, nelletteemäss, nelletteemäd, nellettemije, nellettemiit, nellettemiise, nellettemiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**seemen** seemne, seement, seemnese, seemness, seemned, seemniije, seemniit, seemniise, seemniiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**süä** süä, süttä, süttäse, süäss, süäd, süttiije, süttiit, süttiise, süttiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**tütär** tüttäre, tütärte, tüttärese, tüttäreess, tüttäred, tüttäriije, tüttäriit, tüttäriise, tüttäriiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**õnnõto** õnnõttoma, õnnõtota, õnnõttomasõ, õnnõttomass, õnnõttomad, õnnõttomijõ, õnnõttomiit, õnnõttomiisõ, õnnõttomiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *õnnõto*, *hoolito*.

**kahõsa** kahõssõmõ, kahõssõma, kahõssõmasõ, kahõssõmass, kahõssõmad, kahõssõmijõ, kahõssõmiit, kahõssõmiisõ, kahõssõmiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kuuvvaiz** kuuvvõttõma, kuuvvõt, kuuvvõttõmasõ, kuuvvõttõmass, kuuvvõttõmad, kuuvvõttõmijõ, kuuvvõttõmiit, kuuvvõttõmiisõ, kuuvvõttõmiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kõlmaz** kõlmõttõma, kõlmaissõ, kõlmõttõmasõ, kõlmõttõmass, kõlmõttõmad, kõlmõttõmijõ, kõlmõttõmiit, kõlmõttõmiisõ, kõlmõttõmiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

## 5.14 XIV käändkond

Neljateistkümnenda käändkonna sõnad lõpevad *-aZ/-äZ*, *-iZ* või *-e/-õ* Ariste 1968, lk. 50.

Avatuid küsimusi-tähelepanekuid:

- *-Z*-lõpu sandhi nähtus on kõigi liikmete puhul ühtlustatud *-z* lõpulisteks
- plurale tantum 'ivusõd' kustutatud sest 'ivuz' olemas

- Tsvetkovi antud paralleelvariantidest on valitud vaid üks (korpuse, analoogsete sõnade ülekaalu ning Heinsoo ja Konkova põhjal)
- valitud 'lähe' tugevaastmeline sg tüvi, sest VKSis esineb ühes Li näitelauses
- -kõz-liides muudetud eespoolseks vastavate sõnade juures

### Ekstraktmorfoloogia sõnatüübid

**õgaz** õkka, õgassõ, õkkasõ, õkkass, õkkad, õkkajõ, õkkait, õkkaisõ, õkkaiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**puhaz** puhta, puhassõ, puhtasõ, puhtass, puhtad, puhtajõ, puhtait, puhtaisõ, puhtaiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *puhaz*, *ahaz*.

**lähe** lähte, lähett, lähtese, lähtess, lähted, lähteje, lähteit, lähteise, lähteiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**rüiz** rüčče, rüisse, rüččese, rüččess, rüččed, rüččije, rüččiit, rüččiise, rüččiiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**rakõ** rakkõ, rakõttõ, rakkõsõ, rakkõss, rakkõd, rakkõjõ, rakkõit, rakkõisõ, rakkõiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**rikaz** rikka, rikassõ, rikkasõ, rikkass, rikkad, rikkajõ, rikkait, rikkaisõ, rikkaiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**bulbukõz** bulbukka, bulbukassõ, bulbukkasõ, bulbukkass, bulbukkad, bulbukkajõ, bulbukkait, bulbukkaisõ, bulbukkaiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *bulbukõz*, *čirkõz*, *liivõkõz*, *mansikõz*, *musicõz*, *nenäkõz*, *õnnõkõz*, *baabukõz*.

**kalliz** kalli, kallissõ, kallisõ, kalliss, kallid, kallejõ, kalleit, kalleisõ, kalleiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**pallõz** palla, pallassõ, pallasõ, pallass, pallad, pallajõ, pallait, pallaisõ, pallaiss  
Sõnatüüp hõlmab vormisõnastiku lekseeme: *pallõz*, *rahvõz*, *taivõz*, *võõrõz*, *ahnõz*.

**poolõz** poola, poolassõ, poolasõ, poolass, poolad, poolajõ, pooloit, pooloisõ, pooloiss  
Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**hammõz** hampa, hammassõ, hampasõ, hampass, hampad, hampajõ, hampait, hampaisõ, hampaiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *hammõz*, *lammõz*, *ammõz*.

**lõunõ** lõuna, lõunattõ, lõunasõ, lõunass, lõunad, lõunajõ, lõunait, lõunaisõ, lõunaiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kangõz** kanka, kangõssõ, kankasõ, kankass, kankad, kankajõ, kankait, kankaisõ, kankaiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kauniz** kauni, kaunissõ, kaunisõ, kauniss, kaunid, kaunijõ, kauniit, kauniisõ, kauniiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**angõriaz** angõria, angõriassõ, angõriasõ, angõriass, angõriad, angõrijõ, angõriit, angõriisõ, angõriiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**raskõz** raska, raskassõ, raskasõ, raskass, raskad, raskojõ, raskait, raskaisõ, rakaiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kasõ** kassõ, kasõttõ, kassõsõ, kassõss, kassõd, kassõjõ, kassõit, kassõisõ, kassõiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**vetelüz** vetelüse, vetelüsse, vetelüsse, vetelüsess, vetelüsed, vetelüssije, vetelüssiit, vetelüssiise, vetelüssiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *vetelüz*, *jänez*.

**põrzõz** põrsa, põrzassõ, põrsasõ, põrsass, põrsad, põrsojõ, põrsoit, põrsoisõ, põrsoiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**ratiz** ratti, ratissõ, rattisõ, rattiss, rattid, rattijõ, rattiit, rattiisõ, rattiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**katõ** kattõ, katõttõ, kattõsõ, kattõss, kattõd, kattõjõ, kattõit, kattõisõ, kattõiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kuõ** kutõ, kuõttõ, kutõsõ, kuttõss, kuttõd, kuttõjõ, kuttõit, kuttõisõ, kuttõiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**varvõz** varpa, varvassõ, varpasõ, varpass, varpad, varpajõ, varpait, varpaisõ, varpaiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *varvõz*, *turvõz*.



**pere** pere, perette, perese, peress, pered, pereje, pereit, pereise, pereiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *pere*, *vene*, *erne*.

## 5.15 XV käändkond

Viieteistkümnes käändkond koondab sõnu nagu *lühüd*, *õhud*, *koollu*, *ilozuZ*, *rikkauZ*

Ariste 1968, lk. 51.

Avatuid küsimusi-tähelepanekuid:

•

### Ekstraktmorfoloogia sõnatüübid

**čämmel** čämmele, čämmelte, čämmeläse, čämmeless, čämmeled, čämmelije, čämmeliit,  
čämmeliise, čämmeliiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**oonõ** oonõ, oonõt, oonõsõ, oonõss, oonõd, oonijõ, ooniit, ooniisõ, ooniiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**kannõl** kannõlõ, kannõltõ, kannõlasõ, kannõlõss, kannõlõd, kannõlijõ, kannõliit, kannõliisõ,  
kannõliiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**peremmeez** peremmehe, peremmeesse, peremmehese, peremmehess, peremmehed, peremmehije,  
peremmehiit, peremmehiise, peremmehiiss

Sõnatüüp hõlmab vormisõnastiku lekseeme: *peremmeez*, *meez*.

**märännü** märännü, märännüt, märännüse, märänness, märänned, märännije, märänneit,  
märänneise, märänneiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**čeväd** čevvä, čevätte, čevättese, čeväss, čeväd, čevvije, čevviit, čevviise, čevviiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**čäčüd** čäččü, čäčütt, čäččüse, čäččüss, čäččüd, čäččüje, čäččüit, čäččüise, čäččüiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

**ivusõd** ivussijõ, ivussiit, ivussisõ, ivusõss, ivusõd, ivussijõ, ivussiit, ivussiisõ, ivussiiss

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***kaatsõd*** *kaattsojõ, kaattsoit, kaattsoisõ, kaatsõss, kaatsõd, kaatsojõ, kaattsoit, kaattsoisõ, kaattsoiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***koollud*** *koollõ, koolluttõ, koollõsõ, koollõss, koollõd, koollujõ, koolluit, koolluisõ, koolluiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***koorrõ*** *koorrõ, koorrõt, koortõsõ, koorrõss, koorrõd, koortõjõ, koortõit, koortõisõ, koortõiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***kõlmõd*** *kõlmõ, kõlma, kõlmõsõ, kõlmõss, kõlmõd, kõlmijõ, kõlmiit, kõlmiisõ, kõlmiiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***olud*** *olu, oluttõ, olusõ, oluss, olud, olujõ, oluit, oluisõ, oluiss*

Sõnatüüp ei hõlma teisi lekseeme vormisõnastikus.

***toho*** *toho, tohtõ, tohtõsõ, tohoss, tohod, tohojõ, tohoit, tohoisõ, tohoiss*

Sõnatüüp hõlmab vormisõnastiku lekseeme: *toho, roho*.

## 5.16 Ekstraktmorfoloogia üldistatud muuttüüpide algoritm

Silfverberg ja Hulden (2018) on kirjeldanud üht formaalset viisi, kuidas ekstraktmorfoloogia tüüpsõnu kokku grupeerida ja seega nende arvu vähendada. Siin alaosas rakendatakse meetodit leitud tüüpsõnadele ja esitatakse selle põhjal loodud vadjaja muuttüübistik ja võrreldatakse leitud muuttüübistikku Ariste käändkondadega.

### 5.16.1 Muuttüüp I

## 5.17 Põhivormid ja analoogiavormid

Selles osas selgitatakse välja vadjakeele tüüpsõnade põhi- ja analoogiavormid sõnaliigiti. Seda püütakse teha formaalselt põhinedes vaid ekstraktmorfoloogiaga leitud tüüpsõnamallidele.

M. Erelt, T. Erelt ja Ross 2007 järgi “[p]õhivormid on need vormid, mida pole võimalik teiste vormide alusel tuletada ning mille moodustamiseks tuleb iga sõnatüübi korral anda vastavad reeglid.” ja “[a]naloogiavormid on vormid, mida saab moodustada mingi põhivormi analoogial.”

Tegelikult on ekstraktmorfoloogia leitud LCS ainus põhivorm ja kõik muutvormid on sellest tuletatud analoogiavormid. Kuna aga läänemeresoome keelte keeleteaduses ei ole katkendlike põhivormide kasutamine traditsioonis (nagu seda on nt araabia keelte puhul), püütakse siin leida traditsioonilise käsitluse järgi põhi- ja analoogiavormid.

#### **5.17.1 Käändsõnad**

#### **5.17.2 Tegusõnad**

## 6 Programmkoodi tuletamine

Programmkoodi tuletamise (ingl. *source code generation*) all peetakse siin töös silmas mistahes protsessi, mille käigus tuletatakse mingi üldisema kirjelduse põhjal programmkoodi ühe või mitme konkreetse programmeerimiskeskona jaoks.

Üldine kirjeldus (või teisisõnu ontoloogia) kirjeldab faktuaalset *mis*-laadi teadmist ning tuletatud programmkood kirjeldab konkreetset *kuidas* neid teadmisi rakendada.

Töös kasutatakse keskseks kirjelduseks vormisõnastikku ja sellest ekstraktmorfoloogiaga eraldatud sõnatüübimalle. Kirjeldus on vormistatud rahvusvahelise standardi *Lexical Markup Framework* (ISO/TC 37/SC 4 2007) järgi leksikaalseks ressursiks, mis on XML vormingus.

Programmkoodi tuletavad nn generaatorid. Töös esitatakse kaht generaatorit, üks programmeerimiskeele Grammatical Framework jaoks ning teine Giella keeletehnoloogilise taristu integreerimise jaoks. Generaatorid on kirjutatud XQuery programmeerimiskeeles.

Mõlema programmkoodigeneraatori ühine arhitektuurne omadus on terminite tõlketabelite kasutamine. Tõlketabelite põhjal asendatakse leksikaalses ressursis kasutatud terminid sihtkeeles kasutatud terminitega. See võimaldab järgida ja austada eri taristute terminoloogilisi traditsioone. Näiteks nimetatakse sõnatüüpe LMFis ingliskeelse prefiksiga *as* (kui *asKatto*), aga Grammatical Framework'is eesliitega *mk* (ingliskeelsest tegusõnast *make*) kui *mkKatto* ja Giella taristus hoopis sõnaliigiga (*N\_KATTO*).

Samuti erinevad eri taristutes kasutatud grammatiliste tunnuste märgendid (nt ainsuse nominatiiv on LMFis ja GFis *singular nominative*, aga Giellas *+SG+NOM*).

### 6.1 Keskne kirjeldus Lexical Markup Framework vormingus

Lexical Markup Framework (LMF) on loomuliku keeletöötuse leksikonidele ja masinloetavatele sõnaraamatutele mõeldud rahvusvaheline ISO-standard. Standard koosneb märgenduskeelega defineeritud ühisest mudelist andmevahetuse juhtimiseks ja võimaldab erinevate ressursside ühendamist. (Francopoulo 2013, lk. 1).

Standard koosneb põhimoodulist ja mitmest eriotstarbelisest laiendimoodulist (Franco-poulo 2013, lk. 22). Magistritöös kasutatakse peale põhimoodulit veel kahte moodulit: morfoloogia moodul (*LMF Morphology Extension*) ja morfoloogiliste paradigmade moodul (*LMF Morphological Pattern Extension*). Valitud moodulitega on olnud võimalik kirjeldada ja salvestada nii vormisõnastiku andmed (lekseemide muutvormid) kui ka ekstraktmorfoloogiaga eraldatud andmed (sõnatüüpide protsessuaalsed koostamismallid ja tehniliste tüvede muutujad).

Morfoloogiamooduli eesmärgiks on kirjeldada morfoloogiat mahu kaudu, s.o kirjeldada lekseemi loendades kõik selle muutvormid.

Morfoloogiliste paradigmade mooduli eesmärgiks on seevastu kirjeldada morfoloogiat sisu kaudu, s.o kirjeldada neid kriteeriume ja reegleid, millega saab moodustada kõik ühe lekseemi muutvormid. Selles töös kirjeldatakse ekstraktmorfoloogia sõnatüübimallid antud mooduliga.

Sama nähtuse kirjeldamine nii mahus kui ka sisus võib tunduda liigsena, ent mõlemal kirjeldusviisil on omad head küljed. Lekseemide iga muutvormi loendamist peetakse magistritöös eeskätt dokumenteeriva ja arhiveeritava väärtusena, mis ei kahane pikaajalises perspektiivis – muutvormide loendamine on ka 50 või 150 aasta pärast informatiivne, keeletehnoloogia programmikood ei pruugi olla arusaadav ega jooksvatav. Veel võimaldab muutvormide loendamine edaspidises sõnastikutöös ka igale muutvormile lisada informatsiooni, nt selle reaalsete korpusesinemuste kohta.

Ekstraktmorfoloogiaga eraldatud sõnatüübimallide kirjeldamine samas vormingus võimaldab neid protsessuaalseid koostamiseeskirju teisendada programmkoodi generaatoritega. Generaatoreid võib hiljem lisada juurde uue keeletehnoloogia tuletamise jaoks ilma selleta, et lähteandmestikku peab muutma. See teebki LMF-i kirjelduse nõ tehnoloogianeutraalseks standardiks.

Peale sõnaartiklite ja morfoloogilise informatsiooni hoitakse leksikaalses ressursis ka globaalset informatsiooni, nagu keele nimetust ja keelekoodi, mida kasutatakse tuletatud programmkoodis peamiselt failide nimetamisel.

### 6.1.1 Sõnaartikli esitamine LMFis

Lekseemide sõnaartiklid LMFis esituses koosnevad muutvormide loendist, lemmaks valitud muutvormist, sõnaliigist ja sõnatüübist. Paralleelvormide puhul on sõnatüüpe rohkem kui üks.

Andmete struktuuri näitlikustav sõnaartikkel on LMFis XML märgendusformaadis esitatud joonises 1.

```
<LexicalEntry morphologicalPatterns="asKatto">
  <feat att="partOfSpeech" val="noun"/>
  <Lemma>
    <feat att="writtenForm" val="katto"/>
  </Lemma>
  <WordForm>
    <feat att="writtenForm" val="katto"/>
    <feat att="grammaticalNumber" val="singular"/>
    <feat att="grammaticalCase" val="nominative"/>
  </WordForm>
  <WordForm>
    <feat att="writtenForm" val="katod"/>
    <feat att="grammaticalNumber" val="plural"/>
    <feat att="grammaticalCase" val="nominative"/>
  </WordForm>
</LexicalEntry>
```

Joonis 1: Sõnaartikli *katto* esitamine LMFis (muutvormid on kajastatud vaid osaliselt).

Sõnatüübimall *asKatto* on märgitud *morphologicalPatterns* atribuudis. Mitme sõnatüübi puhul loendatakse need tühikutega eraldatuna. Sõnaliik on märgitud *partOfSpeech* elemendis (*noun*). Element *Lemma* sisaldab lemmaks valitud sõnavormi. *WordForm*-elemendid loendavad muutvorme koos nende grammatiliste tunnustega: arv (*grammaticalNumber* ja kääne (*grammaticalCase*).

Sõnavormide modaalsust on täpsustatud kirjakeelseks kujuks (*writtenForm*).

### 6.1.2 Sõnatüübi malli esitamine LMFis

Ekstraktmorfoloogiaga eraldatud sõnatüübid kirjeldatakse LMF-i morfoloogiliste paradigmade mooduli elementidega (*LMF Morphological Pattern Extension*, varem nimetatud ka *LMF Paradigm Pattern*). Iga sõnatüüp on märgitud identifitseeriva nimetusega, mille puhul on järgitud LMF-i tava lisada tüüpsõna ette ingliskeelne eeslide *as* (nt *asKatto*).

Peale nimetuse ja sõnaliigi kirjeldatakse LMF morfoloogiliste paradigmadega veel kõik ekstraktmorfoloogiaga eraldatud informatsioon: salvestatakse lekseemide tehniliste tüvede muutujad ja muutvormimallid.

Muutvormallid on LMF-is kirjeldatud nn transformatsioonielementidega (*TransformSet*), mis koosnevad muutvormi morfoloogistest tunnustest (*GrammaticalFeatures*) ja muutvormi koostamiseks vajalikest protsessidest (*Process*). Koostamisprotsessid osundavad lihtsaid, järjestatud konkatenatsioonitehteid (*addAfter*), mis liidavad lükkimise teel tehnilise tüve muutujad ja muutvormi ülejäänud tähtkoostised kokku.

Andmete struktuuri näitlikustav sõnatüüp on LMF-i XML märgendusformaadis esitatud joonises 2 (lk 56).

Morfoloogiliste paradigmade *Process*-elemendid on üks-üheses vastavuses ekstraktmorfoloogia meetodiga eraldatud sõnatüübimallidega.

```

<MorphologicalPattern>
  <feat att="id" val="asTšiutto"/>
  <feat att="partOfSpeech" val="nn"/>
  <TransformSet>
    <GrammaticalFeatures>
      <feat att="grammaticalNumber" val="singular"/>
      <feat att="grammaticalCase" val="nominative"/>
    </GrammaticalFeatures>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddVariable"/>
      <feat att="variableNum" val="1"/>
    </Process>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddConstant"/>
      <feat att="stringValue" val="t"/>
    </Process>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddVariable"/>
      <feat att="variableNum" val="2"/>
    </Process>
  </TransformSet>
  <TransformSet>
    <GrammaticalFeatures>
      <feat att="grammaticalNumber" val="plural"/>
      <feat att="grammaticalCase" val="nominative"/>
    </GrammaticalFeatures>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddVariable"/>
      <feat att="variableNum" val="1"/>
    </Process>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddVariable"/>
      <feat att="variableNum" val="2"/>
    </Process>
    <Process>
      <feat att="operator" val="addAfter"/>
      <feat att="processType" val="pextractAddConstant"/>
      <feat att="stringValue" val="d"/>
    </Process>
  </TransformSet>
</MorphologicalPattern>

```

Joonis 2: Tüüpsõnamalli tšiutto (mille alla kuuluvad mh *tšiutto* ja *katto*) esitus LMF-is. Esitus mudeldab muutvormimalle  $x_1 \oplus \mathbf{t} \oplus x_2$  ning  $x_1 \oplus x_2 \oplus \mathbf{d}$ .



## 6.2 Integreerimine Grammatical Framework'iga

Grammatical Framework (GF) on eriotstarbeline programmeerimiskeel, mis on loodud loomulike (ja tehislake) keelte grammatikate kirjeldamiseks. GFi iseloomustab muuhulgas see, et see kuulub funktsionaalsete programmeerimiskeelte paradigmasse ja on rajatud tüübiteooriale. (Ranta 2011, lk. VII)

Teoreetiliselt on GF formaalgrammatikana väljendusrikkuse poolest ekvivalentne PMCFG tüüpi grammatikaga (*Parallel Multiple Context-Free Grammar*), mis jääb mõõdukalt kontekstitundlike *mildly context-sensitive* ja täielkult kontekstitundlike *fully context-sensitive* grammatikate vahele (Ranta 2011, lk. 10).

Grammatical Framework toetab on keelespetsiifilise koodi eraldamist teekidesse, mida nimetatakse ressursiteekideks (Ranta 2008, Ranta 2011, lk. 97). Üle 32 keele on oma ressursiteegiga toetatud. Eesti keele ressursiteegi on loonud Listenmaa ja Kaljurand 2014.

Siin esitatud programmkoodi generaator loob vadja keele ressursiteegi morfoloogia ja leksikoni moodulid. Loodud vadja ressursiteek on internetis saadaval<sup>1</sup>.

Loodud generaator näitlikustab üht integreerimisviisi, mida iseloomustab tarkvara arendusmeetod pidev paigaldamine (*Continuous Integration*). Vormisõnastiku täiendamist (ja ka muutmist) saab kasutada eesmärgipäraselt selleks, et viia täiendused sisse Grammatical Framework'i vadja morfoloogiateeki ja sealtkaudu sellele põhinevatele rakendustele.

Generaator on jaotatav kaheks osaks, millest üks genereerib vadja morfoloogiameooduli ja teine vadja leksikonimeooduli.

Generaator koostatud võimalikult universaalseks ja ei sisalda midagi vadja keelele spetsiifilist. Lemmavormi valikut ei ole generaatorisse jäigalt sissekodeeritud, vaid on vahetatav generaatori parameetritega. Veel koostatakse moodulite failinimed vastavalt leksikaalses ressursist antud keelekoodi järgi.

---

<sup>1</sup><https://github.com/keeleleek/GF-Votic>

### 6.2.1 Morfoloogia moodul

Grammatical Framework'i morfoloogiamoodul koosneb nn paradigmafunktsioonidest, mille ülesandeks on luua sõnade paradigmad ehk muutvormitabelid (Ranta 2011, lk. 248, Détrez ja Ranta 2012, lk. 645).

Ekstraktmorfoloogiaga eraldatud sõnatüübimallide funktsioonid on olnud lihtne teostada paradigmafunktsioonidena. Igale sõnatübile on teostatud kaks eraldi funktsiooni: üks üldine, mis võtab sisendiks sõna muutvormi tervikuna ja väljastab sõna tehnilise tüve osad, teine on konkreetne ja selle sisendiks tehnilise tüve osad ja väljundiks on muutvormide tabel. Kuigi on võimalik kahest funktsioonist üks moodustada, on kaks funktsiooni eraldi hoitud võimalike tarkvaravigade otsimise tarbeks, et arendajal on võimalik testida enda moodustatud tehniliste tüvedega.

Sõnatüübist paradigmafunktsiooni genereerimiseks on vaja määrata funktsiooni nimi, liides (*interface*) ja keha (*body*). Genereeritud programmikoodi illustreeritakse joonises 3.

Morfoloogiamooduli nimi on koostatud leksikaalses ressursis oleva vadja keelekoodiga vastavalt GF moodulinimetustele *MorphoVot*.

```

resource MorphoVot = {

param
  Number = singular | plural ;
  Case = nominative ;
  NForm = NF Number Case ;

oper
  Noun : Type = {s : NForm => Str} ;

-----
-- Start of Noun section
-----

mkTšiuutto : Str -> Noun = \tšiuutto ->
  case tšiuutto of {
    tšiuut + "t" + o@(-(_+"t"+_)) => mkTšiuuttoConcrete tšiuut o ;
    _ => Predef.error "Unsuitable lemma for mkTšiuutto"
  } ;

mkTšiuuttoConcrete : Str -> Str -> Noun = \tšiuut,o ->
  { s =
    table {
      NF singular nominative => tšiuut + "t" + o ;
      NF plural nominative => tšiuut + o + "d"
    }
  } ;
}

```

Joonis 3: Vadja keele GF morfoloogiamooduli genereeritud programmikood (näites piiratud käänetega).

Funktsiooni nimetuseks on sõnatüübi identifikaator LMF*i* leksikaalsest ressursist, mille eesliide *as* vahetatakse GF*i* traditsiooni järgi *mk*.

Funktsiooni liides on deklareeritud *oper* osas ja parameetrite tüübid *param* osas. Parameetrite nimetused ja kategooriad peegeldavad grammatilisi tunnuseid LMF*is*.

Kuigi vadja vormisõnastik sisaldab magistritöö valmimise hetkel ainult nimisõnu, arvestab koodigeneraator teiste sõnaliikidega ja grupeerib sõnatüüpide paradigma-funktsioonid vastavateks sektsioonideks, mille ette lisatakse kommentaar sõnaliigi kohta. Koodi lugemise hõlbustamiseks järjestatakse veel sõnatüüpide funktsioonid

nende nimede järgi tähestikuliselt.

Üldisem, lemmavormil opereeriv funktsioon eraldab sisestatud muutvormi selle tehnilise tüve osadeks, mis seejärel edastatakse konkreetsele paradigmafunktsioonile. Võib märkida, et GFi regulaaravaldised töötavad mitte-aplalt (ingl. *non-greedy*), vastupidiselt paljude teiste programmeerimiskeeltele. Seetõttu lisatakse nt `@(-("_t"+_))` avaldisesse, mis muudab *t*-tähega ühtimise aplaks.

Tehnilise tüve muutujate nimetused saadakse sõnatüübi nimetusest.

Konkreetses funktsiooni nimetuse järel lisatakse *Concrete* ja see funktsioon tagastab muutvormide tabeli. Tabeli vasakul poolel on grammatilised tunnused ja paremal poolel on muutvormide mallid.

### 6.2.2 Leksikoni moodul

Grammatical Framework on oma olemuselt mitmekeelne ning iga rakendus peaks täitma nende oma semantika abstraktse grammatikaga. Kuna vormisõnastik ei sisalda hetkel tõlkeid, genereeritakse lihtne sõnaloend.

Iga sõnaartikli puhul koostatakse sõnaloendisse kirje, mille nimeks on lemma ja sõnaliigitähis ning väljakutse vastava sõnatüübi paradigmafunktsioonile. Juhul, kui sõnaartiklil on mitu sõnatüüpi, määratakse need kirje variantideks.

Genereeritud sõnaloendi programmikoodi illustreeritakse joonises ??.

```
fun
  lin katto_N = mkKatto "katto"
  lin čiutto_N = mkKatto "čiutto" ;
```

Joonis 4: Generated source code for the Votic GF lexicon.

## 6.3 Integreerimine Giella-taristuga

The Giellatekno infrastructure has been characterized in Moshagen, Pirinen ja Trotterud 2013 to be a *development environment infrastructure* (as opposed to a resource

infrastructure), offering a framework for building language-specific analysers and directly turn them into a wide range of useful programs.

From the point of view of our work on Votic morphology, the programs of interest are proofing tools and morphological analyzers.

For the integration with Giellatekno's infrastructure, several components are needed: a lexicon, paradigm descriptions in FST, automatic test declarations, and a Makefile that binds all the components together.

At the work's current stage all but the Makefile is being generated. Because of this, each component has only been tested on its own but not integrated in the infrastructure.

We don't include any examples of generated code for the Giellatekno infrastructure because of space considerations. We will only present and discuss the main design choices made.

Keeletehnoloogilise taristuga Giella integreeritakse selles töös peamiselt selleks, et saada kätte õigekirjakontrollija. Giella-taristu koosneb veel võimalustest. Taristut kasutavad peamiselt Giellatekno ja Divvun.

Integreerimine on jagatav kaheks peamiseks osaks: leksikoni integreerimeerimine ja tüüpsõnamallide integreerimine. Seejärel kirjeldatakse taristu poolt loodud õigekirjakontrollija tööpõhimõtet ja lõpetuseks on arutelu.

### 6.3.1 Tüüpsõnad

Paradigmade ehk tüüpsõnamallide esitus FST formalismis põhineb suuresti Forsbergi ja Huldeni (2016) tööle.

Paradigmad esitatakse relatsioonidena sõnavormi ja lemma koos analüüsiga vahel. Sellised relatsioonid sisaldavad lõpmatut hulka sõnalemmasid, millest mõistagi pole suurem osa vajakeelsed. Mis on siiski tähtis, on see, et relatsioonid mudeldavad paradigmasid.

Sõnade lõpmatu hulk piiratakse leksikonis antuga ja niiviisi saadakse leksikonis sisalduvate sõnade kõik sõnavormid. Nendest ja ainult nendest sõnavormidest koosnebki esialgne vadja õigekirjakontrollija.

### 6.3.2 Leksikon

“Formally, the lexc language is a kind of right-recursive phrase-structure grammar.” ja “A lexc description compiles into a standard Xerox finite-state network, either a simple automaton or a transducer.” (Beesley ja Karttunen 2003, lk 203).

Kuigi lexc fraasistruktuurigrammatikatega on võimalik paradigmasid (tüüpsõnamalle) mudeldada, ja tavaliselt selleks seda kasutataksegi Giella taristus, võtab see töö teise lähenemisenurga ja lihtsustab võimalikult palju leksikoni struktuuri.

Leksikon koosneb selles töös ainult kahest andest: *lemma* ja *tüüpsõna*.

### 6.3.3 Õigekirjakontrollija

Eelnevalt kirjeldatud integreerimine Giella-taristusse võimaldab taristul luua õigekirjakontrollija. Mis on õigekirjakontrollija, kus seda kasutatakse ja mida see kontrollib?

### 6.3.4 Arutelu

Loodud õigekirjakontrollija on eesmärgipäraselt jäetud lihtsakoeliseks. See märgib kõik sõnad valeks, mis ei sisaldu sõnastikus. See on lühiajaliseks kasutamiseks ja mõeldud ärgitada kasutajaid ise pakkuma täiendusi ja sõnaloomet vadja sõnastikusse.

## 7 Kokkuvõte

Magistritöö on kirjeldanud süsteemi, millega on ühelt poolt defineeritud vadja keele normatiivne morfoloogia ja mille põhjal teisalt tuletatakse automaatselt morfoloogiline keeletehnoloogia.

Morfoloogilise normatiivi vajadust ajendab Heinike Heinsoo läbiviidud kursused keelekümluskoolis Ämmesse Vunukassaa ja normatiiv on hõlpsasti muudetav-parendatav ilma programmeerimisoskusteta.

Saadud morfoloogilist tüübistikku on analüüsitud vadja keele grammatikatega ja põhjendatud ajaloolise morfoloogiaga.

—

Töö keskseks osaks on ekstraktmorfoloogiameetodiga saadud tüüpsõnakirjeldused. Kirjeldused kodeeritakse koos sõnastikuga ümber standardsesse vormingusse ja saadud leksikaalse ressursi järgi tuletatakse automaatselt programmkoodi kahe keeletehnoloogilise platvormi jaoks, ja tagatakse seega vadja keele tugi nendes platvormides.

Niivisi kasutatakse ekstraktmorfoloogia meetodit kasutajaliidesena, mille abil koostatakse arvutimorfoloogia ainult tüüpsõnade muutvormitabeleid sedastades – mitte programmeerides.

Magistritöös esitatud töövoog paneb leksikaalse ressursi keskele kohale ja tuletatud tehnoloogia sellest teiseks. Uue sõnavara ja vigade parandused tehakse ressursis, mitte mitmes tehnoloogias eraldi.

Kuna nii tüüpsõnade kirjeldused, kui ka ülejäänud sõnastik kodeeritakse rahvusvahelise standardi Lexical Markup Framework vormingusse, tagatakse võimaluse ressursi pikaajaliseks arhiveerimiseks. Leksikaalne ressurss on loetav ja arusaadav palju kauem, kui seda on programmeerimiskood.

Viimase tõttu püüab magistritöö ühendada arvutuslingvistika ja dokumenteeriva lingvistika valdkondi.

## 8 Põhimõisted ja lühendid

Siin loetletakse töös kasutatud mõisted ja lühendid koos nende tähendustega.

**Konkatenatsioon** ehk  $\oplus$  on tähtede ja tähtjärjendite lükkimine teine-teise järele, et moodustada uus tähtjärjend. Näiteks  $aa \oplus be$  moodustab  $aabe$ . 64

**Lemma** on suvaliselt valitud grammatiliste tunnuste komplekt, mida kasutatakse lekseemi viitamiseks. 64

**Mikrostruktuur** on sõnastiku sõnaartikli sisemine struktuur. 64

**Muuttüüp** on morfoloogilise klassifikatsiooni elementaariusus. See on sõnaklass, mis erineb teistest sõnaklassidest mingite erijoonte poolest sõnade muutmisel. 64

**Muutvormimall** kirjeldab üksiku muutvormi koostamisskeemi ja kannab selle grammatilised tunnused. On integraalne osa tüüpsõnamallist. Koostamisskeem koosneb muutujatest ja konstantidest, mille tähtkoostised lükitakse üks-teise järele. Muutujate tähtkoostised võivad olla mingil moel piiratud. 64

**Sõnatüübimall** on ekstraktmorfoloogiaga leitud tüüpsõna paradigma kirjeldus, mis koosneb iga muutvormi koostamismallidest ehk muutvormimallidest. Tüüpsõnamall on relatsioon tehnilise tüve ja kõigi selle paradigmasse kuuluvate muutvormide vahel. 19, 64

**Tehniline tüvi** on tähtkoostiste järjend, millega saab tüüpsõnamalli muutvormide muutujad asendada elik väärtustada ja niiviisi koostada ühe konkreetse sõna kõik vormid. 19, 64



## 9 Kirjandus

- Ahlberg, Malin, Markus Forsberg ja Mans Hulden (2014). „Semi-supervised learning of morphological paradigms and lexicons“. Teoses: *Proc. of EACL*, lk. 569–578. URL: <http://www.aclweb.org/anthology/E/E14/E14-1.pdf#page=595> (vaadatud 13.08.2017).
- Ariste, Paul (1968). *A grammar of the Votic language*. Indiana University publications. Uralic and Altaic series vol. 68. Bloomington : The Hague: Indiana University ; Mouton. 121 lk. kokku.
- Auroux, Sylvain *et al.*, toim. (2006). *History of the Language Sciences: An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present*. Kõide 3. ISBN: 3-11-016736-0.
- Beard, Robert (1987). „Morpheme order in a lexeme/morpheme-based morphology“. *Lingua* 72.1, lk. 1–44.
- (1995). *Lexeme-morpheme Base Morphology: A General Theory of Inflection and Word Formation*. SUNY Series in Linguistics. OCLC: 940540414. State University of New York Press.
- Beesley, Kenneth R ja Lauri Karttunen (2003). *Finite state morphology*. Stanford, Calif.: CSLI Publications. ISBN: 1-57586-433-9 978-1-57586-433-4 1-57586-434-7 978-1-57586-434-1.
- Bender, Emily M. ja D. Terence Langendoen (20. märts 2010). „Computational Linguistics in Support of Linguistic Theory“. *Linguistic Issues in Language Technology* 3.1. ISSN: 1945-3604. URL: <http://elanguage.net/journals/lilt/article/view/661> (vaadatud 29.03.2015).
- Chomsky, Noam (1957). *Syntactic Structures*. OCLC: 979583577. ISBN: 978-3-11-021832-9. URL: <https://doi.org/10.1515/9783110218329> (vaadatud 19.02.2019).
- Coulmas, Florian (1989). *Language adaptation*. OCLC: 231037158. Cambridge u.a.: Cambridge Univ. Pr. ISBN: 978-0-521-36255-9.
- Détrez, Grégoire ja Aarne Ranta (2012). „Smart paradigms and the predictability and complexity of inflectional morphology“. Teoses: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, lk. 645–653. URL: <http://dl.acm.org/citation.cfm?id=2380895> (vaadatud 21.04.2017).
- Erelt, Mati, Tiiu Erelt ja Kristiina Ross (2007). *Eesti Keele Käsiraamat*. 3., täiend. tr. Tallinn: Eesti Keele Sihtasutus. 726 lk. kokku. ISBN: 978-9985-79-210-0.
- Ernits, Enn (2006). „Vadja liikumisest ja kirjakeelest“. *Keel ja Kirjandus* 49.1, lk. 85–87. URL: <https://www.digar.ee/viewer/et/nlib-digar:81648/143905/page/85>.
- (2010). „Vadja kirjaviisist ja sõnaloomest“, lk. 17.
- Forsberg, Markus (2. detsember 2016). „What can we learn from inflection tables?“ Teoses: *BAULT 2016*. BAULT 2016. Helsinki. URL: <http://blogs.helsinki.fi/language-technology/news/bault-2016/> (vaadatud 07.12.2016).
- Forsberg, Markus ja Mans Hulden (2016a). „Deriving Morphological Analyzers from Example Inflections“. Teoses: *LREC*. URL: [http://www.lrec-conf.org/proceedings/lrec2016/pdf/1134\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2016/pdf/1134_Paper.pdf) (vaadatud 13.08.2017).
- (2016b). „Learning Transducer Models for Morphological Analysis from Example Inflections“. *Proceedings of StatFSM*. Association for Computational Linguistics, lk. 42. URL: <http://anthology.aclweb.org/W16-2405>.
- Francopoulo, Gil (2013). *LMF lexical markup framework*. London; Hoboken, NJ: ISTE Ltd ; John Wiley & Sons. ISBN: 1-84821-430-8 978-1-84821-430-9.

- Heinsoo, Heinike (2015). *Vad'd'a sõnakopittõja*. Koostöös Helena Miettinen *et al.* Helsinki ; Tartu: Mooses Putron muistosäätio, Tallinna Raamatutrükikoda). 182 lk. kokku. ISBN: 978-952-93-5025-4.
- Hockett, Charles F. (august 1954). „Two Models of Grammatical Description“. *WORD* 10.2, lk. 210–234. ISSN: 0043-7956, 2373-5112. DOI: 10.1080/00437956.1954.11659524. URL: <http://www.tandfonline.com/doi/full/10.1080/00437956.1954.11659524> (vaadatud 03.09.2018).
- Hulden, Mans (2014). „Generalizing inflection tables into paradigms with finite state operations“. Teoses: *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*, lk. 29–36. URL: <http://aclweb.org/anthology/W/W14/W14-2804.pdf>.
- ISO/TC 37/SC 4 (30. juuni 2007). *Language resource management—Lexical markup framework (LMF)*. 24613:2007 Rev.14. ISO. URL: [http://lirics.loria.fr/doc\\_pub/LMF\\_revision\\_14.pdf](http://lirics.loria.fr/doc_pub/LMF_revision_14.pdf) (vaadatud 13.06.2017).
- Karlsson, Fred *et al.* (2002). *Üldkeeleteadus*. Tallinn: Eesti Keele Sihtasutus. 362 lk. kokku. ISBN: 978-9985-811-96-2.
- Karttunen, Lauri (2003). „Computing with realizational morphology“. Teoses: *Computational linguistics and intelligent text processing*. Springer, lk. 203–214. (Vaadatud 18.08.2015).
- Kloss, Heinz, Université Laval ja Centre international de recherche sur le bilinguisme (1970). *Research possibilities on group bilingualism: a report*. OCLC: 920280520. Québec: Centre international de recherche sur le bilinguisme.
- Konkova, O ja Nikita Djačkov (2014). *Vadda čeele*. ISBN: 978-5-903562-68-8.
- Koskenniemi, Kimmo (22. juuli 2013). „An informal discovery procedure for two-level rules“. *Journal of Language Modelling* 1.1, lk. 155–188. ISSN: 2299-8470. DOI: 10.15398/jlm.v1i1.62. URL: <http://jlm.ipipan.waw.pl/index.php/JLM/article/view/62> (vaadatud 26.09.2016).
- Laakso, Johanna, toim. (1989). *Vatjan käänteissanasto*. Lexica Societatis Fenno-Ugricae 22. Helsinki: Suomalais-ugrilainen seura. 103 lk. kokku. ISBN: 978-951-9403-21-2.
- Listenmaa, Inari ja Kaarel Kaljurand (2014). „Computational Estonian Grammar in Grammatical Framework“. Teoses: *9th SaLTMiL Workshop on “Free/open-Source Language Resources for the Machine Translation of Less-Resourced Languages”*, lk. 13.
- Markus, E ja F Rozhanskiy (2014). „Comitative and Terminative in Votic and Lower Luga Ingrian“. *Linguistica Uralica* 50.4, lk. 241. ISSN: 0868-4731. DOI: 10.3176/lu.2014.4.01. URL: [http://www.kirj.ee/?id=24861&tpl=1061&c\\_tpl=1064](http://www.kirj.ee/?id=24861&tpl=1061&c_tpl=1064) (vaadatud 15.05.2018).
- Markus, Elena ja Fedor Rozhanskiy (18. september 2013). „The concept of the norm in a nearly extinct language without literacy“. Teoses: *Books of Abstracts*. 46th Annual Meeting of the Societas Linguistica Europaea. Köide Part II. University of Split, Croatia, 231–232.
- Matthews, Peter Hugoe (1991). *Morphology*. 2nd ed. Cambridge textbooks in linguistics. Cambridge: Cambridge University Press. 251 lk. kokku. ISBN: 978-0-521-41043-4 978-0-521-42256-7.
- Moshagen, Sjur N., Tommi A. Pirinen ja Trond Trosterud (2013). „Building an open-source development infrastructure for language technology projects“. Teoses: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013); May 22-24, 2013; Oslo University; Norway. NEALT Proceedings Series 16*. Linköping University Electronic Press; Linköpings universitet, lk. 343–352.
- Ranta, Aarne (2008). „Grammars as software libraries“. DOI: 10.1017/CB09780511770524.014.

- Ranta, Aarne (2011). *Grammatical framework : programming with multilingual grammars*. Studies in computational linguistics. CSLI Publications. ISBN: 978-1-57586-627-7 1-57586-627-7 978-1-57586-626-0 1-57586-626-9.
- Rozhanskiy, F (2009). „Zaimstvovannyje russkije sushtshestvitel'nye i problema ih adaptatsii v krakol'skom govore vodskogo jazyka. Russian Substantive Loanwords and Their Adaptation in Jõgõperä Votic“. *Linguistica Uralica* XLV.3, lk. 192–204. ISSN: 0868-4731, 1736-7506. URL: <https://www.ceeol.com/search/article-detail?id=203971> (vaadatud 19.12.2018).
- Rozhanskiy, Fedor ja Elena Markus (9. juuni 2015). „Dialectal variation in Votic: Jõgõperä vs. Luuditsa“. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 6.1, lk. 23. ISSN: 2228-1339, 1736-8987. DOI: 10.12697/jeful.2015.6.1.02. URL: <http://jeful.ut.ee/index.php/JEFUL/article/view/jeful.2015.6.1.02> (vaadatud 11.11.2015).
- Silfverberg, Miikka, Ling Liu ja Mans Hulden (20. august 2018). „A Computational Model for the Linguistic Notion of Morphological Paradigm“, lk. 12.
- Stump, Gregory T (2001). *Inflectional morphology a theory of paradigm structure*. Cambridge; New York: Cambridge University Press. ISBN: 978-0-511-01378-2 978-0-521-78047-6 978-0-511-48633-3. URL: <http://dx.doi.org/10.1017/CB09780511486333> (vaadatud 19.07.2015).
- Viks, Ülle (1976). „Verbide muuttüübid ja morfoloogilised tüübid“. *Keel ja Kirjandus* 5, lk. 276–288. URL: [http://www.digar.ee/viewer/download/34744/full?c=download\\_1\\_142787\\_full&json=true&format=full](http://www.digar.ee/viewer/download/34744/full?c=download_1_142787_full&json=true&format=full) (vaadatud 06.06.2017).
- (1992). *Väike Vormisõnastik. A Concise Morphological Dictionary of Estonian*. Koostöös Henno Rajandi ja Keele ja Kirjanduse Instituut. Tallinn: Eesti Teaduste Akadeemia. 837 lk. kokku. ISBN: 5-7976-1215-1.
- Конькова, Ольга Игоревна ja Никита Дьячков (2014). *Vad'd'a ceeli: учебное пособие по водскому языку*. Санкт-Петербург: Музей антропологии и этнографии им. Петра Великого. 84 lk. kokku. ISBN: 978-5-903562-68-8.
- Маркус, Елена Борисовна ja Федор Иванович Рожанский (2011). *Современный водский язык: тексты и грамматический очерк. Том 2, Грамматический очерк и библиография: [в 2-х томах]*. Koostöös Институт языкознания (Moskva). Санкт Петербург: Нестор-История. 381 lk. kokku. ISBN: 978-5-98187-834-3.

## **10 The use of Extract Morphology for Automatic Derivation of Language Technology for Votic**

An English language summary of this work.

## 11 Lisad

Siin on esitatud kõik ekstraheeritud tüüpsõnamallide tabelid.