

Ekstraktmorfoloogia meetodiga tuletatud  
keeletehnoloogia 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, paradigmade 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	Muuttüüp, tüüpsõna ja sõnatüüp . . . . .	12
<b>3</b>	<b>Vormisõnastiku sisestamine Morfoloogialabori rakenduse abil</b>	<b>13</b>
<b>4</b>	<b>Ekstraktmorfoloogia meetod</b>	<b>15</b>
4.1	Sõnatüübi eraldamine . . . . .	15
4.1.1	Tehnilise tüve eraldamine . . . . .	17
4.1.2	Üldistamine sõnatüübi funktsiooniks . . . . .	17
4.2	Sõnatüübi ennustamine . . . . .	18
<b>5</b>	<b>Vadja morfoloogiliste tüüpsõnade analüüs</b>	<b>20</b>
5.1	I käändkond . . . . .	20
5.2	II käändkond . . . . .	22
5.3	III käändkond . . . . .	27
5.4	IV käändkond . . . . .	31
5.5	V käändkond . . . . .	31
5.6	VI käändkond . . . . .	33
5.7	VII käändkond . . . . .	34
5.8	VIII käändkond . . . . .	35
5.9	IX käändkond . . . . .	38
5.10	X käändkond . . . . .	38
5.11	XI käändkond . . . . .	40
5.12	XII käändkond . . . . .	40
5.13	XIII käändkond . . . . .	41
5.14	XIV käändkond . . . . .	43
5.15	XV käändkond . . . . .	45
<b>6</b>	<b>Programmkoodi tuletamine</b>	<b>47</b>
6.1	Keskne kirjeldus Lexical Markup Framework vormingus . . . . .	47
6.1.1	Sõnaartikli esitamine LMFis . . . . .	49
6.1.2	Tüüpsõna malli esitamine LMFis . . . . .	50
6.2	Integreerimine Grammatical Framework'iga . . . . .	52
6.2.1	Morfoloogia moodul . . . . .	53
6.2.2	Leksikoni moodul . . . . .	55
6.3	Integreerimine Giella-taristuga . . . . .	55

6.3.1	Tüüpsõnad . . . . .	56
6.3.2	Leksikon . . . . .	57
6.3.3	Õigekirjakontrollija . . . . .	57
7	Kokkuvõte	58
8	Põhimõisted ja lühendid	59
9	Kirjandus	60
10	The use of Extract Morphology for Automatic Derivation of Language Technology for Votic	63
11	Lisad	64

# 1 Sissejuhatus

Magistritöö eesmärk on luua süsteem, millega saab koostada vadja keele noomeni jaoks morfoloogiline andmestik ehk vormisõnastik. Vormisõnastik täidab magistritöös kolm suuremat eesmärki: 1) on keeleõppes kasutatav teksti moodustamist toetava sõnaraamatuna, 2) selle sisu on automaatselt teisendatav programmikoodi, mis integreerib vadja keele morfoloogia kahte keeletehnoloogilisse raamistikku ja 3) on lihtsasti täiendatav ja muudetav ilma programmeerimisoskusteta.

Esimese eesmärgi täitmiseks on vormisõnastikusse lisatud Heinike Heinsoo (2015) vadja õppematerjali sõnastikust kõik noomenid (u 420 sõna), millele on juurde lisatud teistest sõnaraamatutest (u 460 sõna). Igale sõnale on lisatud selle põhivormid Tsvetkovi sõnaraamatust (Laakso 1989) ja nende põhjal kõik käändevormid Konkova (Konkova ja Djačkov 2014). Magistritöö vältel koostatud andmestik on kirjeldava iseloomuga ja ei pretendeeri lõplikule morfoloogia esitusele (vrd eesmärk 3).

Teiseks eesmärgiks on olnud vormisõnastiku automaatne teisendatavus keeletehnoloogiaks. Vadja keele kasutajale on ehk kõige praktilisema väärtusega teisendatud õigekirjakontrollija. Programmeerija jaoks teisendab magistritöös koostatud süsteem vormisõnastiku programmikoodiks, mis täidab vadja morfoloogilise toe kahes keeletehnoloogilises taristus (Grammatical Framework ja Giella taristu). Täidetud eesmärgi tõttu on tulevikus võimalik luua vadja keele jaoks keeleõperakendusi nendes taristutes. Juba on katsetatud vene–vadja vestmiku loomist.

Kolmandaks eesmärgiks on olnud süsteemi kasutajasõbralikkus. Vormisõnastik on täiendatav ja muudetav lihtsal viisil, mis ei eelda erilisi oskusi peale muutvormide tundmise. Vormisõnastiku täiendamise puhul peab kasutaja valima uue sõna jaoks tüüpsõna, mille järgi see morfoloogiliselt käitub. Muutmise korral peab ainult sõnavormid ümber kirjutama. Sellega on magistritöös loodud süsteemiga tahetud anda vadja keelele kala asemel õnge, et edaspidine vadja keele arendamine võiks toimuda teiste eestvedamisel, mitte jääda arvutuslingvisti või keeletehnoloogi taha toppama.

Püstitatud eesmärkide täitmiseks on kasutatud ekstraktmorfoloogia meetodit (Ahlberg, Forsberg ja Huldén 2014). Meetod sobitub ülesandele (3) hästi, kuna ei põhine

morfo(fono)loogiliste reeglite kirjeldustele, vaid otse sõnavormidele, millest meetod eraldab (ekstraheerib) muuttüübikirjelduse. Sõnavormide muutmise muudab seega muuttüübi kirjeldust.

Valitud meetodi kompromissiks on muuttüüpide paljususe: *čiutto* (ee särk) käändub muuttüübi *katto* (ee katus) järgi, mitte *tüttö* (ee tüdruk). Ainuse erinevus on vokaalharmoonias. Ekstraktmorfoloogia meetodiga leitud muuttüübigistik sarnaneb Eesti leksikograafiatraditsioonis Elmar Muugi Väiksele õigekeelsus-sõnaraamatule (1933), mille muuttüübigistik hõlmab 895 eesti keele noomeni tüüpi. Ülle Viks ei ole pidanud neid muuttüüpideks, vaid struktureeritüüpideks (Viks 1976, lk. 282). Magistritöö vajava vormisõnastik hõlmab 232 vajava keele noomeni tüüpi.

Magistritöö on struktureeritud järgmiselt. Peatükis 2 kirjeldatakse vormisõnastiku koostamise teoreetilisi lähtekohti, selle asetust korpuse planeerimises (ptk 2.1); vormisõnastiku sõnavara valiku, paradigmade moodustamise ja ühtlustamise põhimõtted (ptk 2.2); meetodist tulenev erinev morfeemi definitsioon (ptk 2.3) ja selle sarnasus klassikalise paradigmaatilise morfoloogia käsitlusega (ptk 2.4).

Peatükis 3 tutvustatakse vormisõnastiku kasutajaliidest Morfoloogialabor. Peatükis 4 kirjeldatakse ekstraktmorfoloogia meetodit. Peatükis 5 analüüsitakse meetodiga saadud muuttüübigistik käändkonnniti. Peatükis 6 kirjeldatakse vormisõnastiku teisendamist keeletehnoloogiaks. Selleks kasutatud keskset kirjeldust (ptk 6.1) ja programmkoodi generaatoreid kahe taristuga integreerimiseks (ptk 6.2 Grammatical Framework ja ptk 6.3 Giella taristu ja õigekirjakontrollija). Peatükis 7 võetakse kokku.

## 2 Vormisõnastiku koostamise teoreetilised lähtekohad

Siin peatükis kirjeldatakse vormisõnastiku koostamise ja vadja kirjakeelsete muutvormide rekonstrueerimise põhimõtteid. Andmete sisestamist nn Morfoloogialabori rakendusega kirjeldatakse peatükis 3 ja talletatud andmete kuju kirjeldatakse peatükis 6.1.

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 Vadja ühiskeel ja korpuse planeerimine

Selles peatükis seletatakse vadja ühiskeeleks valitud aluseid ning korpuse planeerimise mõistet. Vadja keelele ei loodud 1930-ndatel Nõukogude Liidus ühiskeelt, nagu seda tehti näiteks karjala, vepsa ja isuri keele jaoks. Vadja kirjakeele loomise pürgimistest pärast Nõukogude Liidu lagunemist on kirjutanud lähemalt Ernits 2006 ja vadjalaste 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 magistrیتöös loodud materjal olema mõlema õppematerjali jaoks kasulik. Ainsad erinevused kahe kirjaväisi vahel on  $\ddot{u} \rightarrow y$  ja  $\check{c} \rightarrow c$ .

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

Morfoloogia ühtlustamist iseloomustab siinses töös sõnavormide rekonstrueerimine Jõgõperä murdeliste muutvormide põhjal Vaipooli murdealale iseäralikele joontele vastavaks. Standardiseerimise suhtes ei võta magistrیتöö 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 magistrیتöö ei pürgi vadja keele sõnavara moderniseerimisele ega rikastamisele – 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 magistrیتöö loob ühtse töökeskkonna, mille abil määratakse morfoloogia ainult sõnade muutvormide esitamise kaudu ja ei vaja programmeerimisioskusi, siis loodab töö 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 adjektiive (u 420 sõna). Sõnavara on veel laiendatud autori enda silma järgi Tsvetkovi sõnastikust ettesattunuga (u 460 sõna, mh *asfaltti*, *bibli* ja *biblioteekkõ*). Kokku on vormisõnastikku 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 mitu 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 produktiivseteks kääneteks Ariste 1968, lk. 17. Markus ja Rožanski grammatika Маркус ja Рожанский 2011 ei käsitle terminatiivi ja komitatiivi käänetenäidatena vaid käände ja järelasendi vahepealsetena, mille analüüsi nad on põhjendanud pike-malt 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 (Heinsoo 2015, lk. 88, Konkova ja Djačkov 2014, lk. 10).

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



kääne	eespoolne	eespoolne	tagapoolne	tagapoolne	tagapoolne
SG NOM	pää	ärče	lafkõ	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 sõ-nastikele vastavalt või rekonstrueeritud vastavalt sõna vokalismile. Mitmuse partitiivi puhul on eelistatud pikemat, muutelõpuga varianti puhta *-i*-mitmusliku tüve asemel. Mõlemad vormid esinevad paralleelsete variantidena Heinsoo ja Konkova õppema-terjalides (Heinsoo 2015, lk. 88, Konkova ja Djačkov 2014, lk. 10).

**Illatiiv** Loodava kirjakeele ühtlustamise huvides on vormisõnastiku koostamisel eelis-tatud läbinähtava käändelõpuga vormi *-se/-sõ* lühikese illatiivi asemel. Selline eelistus on siiski vastuolus olukorraga tänapäeva Luuditsa murdes, kus tavaliselt esineb lühi-ke vorm ja pikk vorm esineb väga harva Markus ja Rozhanskiy 2014, lk. 247. Heinsoo ja Konkova õppematerjalides on mõlemad vormid esitatud rööpselt (Heinsoo 2015,

lk. 88, Konkova ja Djačkov 2014, lk. 10).

**Inessiiv** 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-* ja *-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** 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** 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** 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** 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** 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** 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 tüvel (Ariste 1968, lk. 34). Markuse ja Rožanski järgi põhineb see tänapäeva keeles genitiivi tüvel ja harva allatiivi tüvel (Markus ja Rozhanskiy 2014, lk. 247). Seetõttu võib magistritöös valitud SG INE tüve põhjal konstrueeritud muutevorm olla vale, kuigi valik ühtib Konkova ja Heinsoo õppematerjalides esitatuga, so tugevaastmelise geminaadiga.

**Komitatiiv** 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) (Konkova ja Djačkov 2014, lk. 10).

## 2.3 Morfeemi staatus ja definitsioon

Magistritöös ei käsitleta morfeemi levinud lingvistilise seisukoha järgi kui *väikseimat tähenduslikku üksust*, vaid töö järgib definitsiooni, et morfeem on lekseemi *mistahes fonoloogilise kuju muutust, millega kaasneb tähenduslik muutus* (Beard 1987, lk. 31; Beard 1995, lk. 49), mis on sarnane klassikalise paradigmaatilise morfoloogia lähene-misele.

## 2.4 Klassikaline paradigmaatiline morfoloogia

Matthews (1991, lk. 186) tõstab esile kaks paradigmaatilise morfoloogia käsitlust, klassikalise ja uue ehk strukturalistliku. Mõlemad käsitlused mõistavad sõna kui selle sõnavormide komplekti ehk paradigmata (Matthews 1991, lk. 186). Kahe käsitluse peamiseks erinevuseks on see, et strukturalistlikus käsitluses koosneb sõna morfeemidest, aga klassikalises käsitluses pole sõnast väiksemat tähenduslikku üksust (*ibid.*, lk. 196). Sõnavormide grammatilised tähendused on klassikalise teooria järgi vaid sat-

tumused (ingl. *accidents*) (Matthews 1991, lk. 189).

Vanades ladina keele õpikutes ja grammatikates võidi esitada reegleid, mille abil ühe lekseemi kõik muutvormid moodustada. Reeglid opereerisid ainult sõnavormide tähtkoostisel. Näiteks võis reegel ühe sõnavormi lõputähti asendada muude tähtedega, et saada teine sõnavorm. Et asendatavatele tähtkoostistele ei omandatud mingit tähendust, näitlikustab ka see, et mõne reegli algvormiks võidi valida suvaline sõnavorm, mis oma tähtkoostise poolest kõige paremini sobis. (*ibid.*, lk. 195–196).

Magistritöös rakendatud ekstraktmorfoloogia meetod koosneb sarnastest sõnavormi koostamisreeglitest, nn muutvormimallidest. Meetodit tutvustatakse lähemalt pt 4.

## 2.5 Muuttüüp, tüüpsõna ja sõnatüüp

Selles töös kasutatakse termineid *muuttüüp*, *tüüpsõna* ja *sõnatüüp* sama tähendusega ja osutab lekseemi kõiki muutvorme. *Tüüpsõna* on valitud lekseem, millega nimetatakse üht *sõnatüüpi*. *Sõnatüüp* on seega tüüpsõnast üldisem mõiste ja tähistab muutvormide koostamismalle, millega saab moodustada sellesse sõnatüüpi kuuluvate sõnade kõiki muutvorme.

### 3 Vormisõnastiku sisestamine Morfoloogialabori rakenduse abil

Morfoloogialabor<sup>1</sup> on eraldiseisev veebirakendus morfoloogiliste sõnaraamatute koostamiseks. See on loodud Språkbanken'is<sup>2</sup> prototüübina, et integreerida ekstraktmorfoloogia nende leksikaalsesse taristusse Karp.

Labor põhineb ekstraktmorfoloogial ja lihtsustab uute sõnade morfoloogilise informatsiooni lisamist sõnaraamatu koostamisel. Rakendus ennustab sisestatud sõnale sobivaid muuttüüpe ja moodustab muuttüübile vastavalt kõik muutvormid. Kasutaja peab kontrollima, kas muutvormid on õiged või valima teise muuttüübi. Rakendus on ühendatud korpuspäringute süsteemiga KORP ja näitab muutvormi juures kasutajale selle esinemisarvu korpustes.

Magistritöö jooksul on koostatud ja kasutatud kaht vadja korpust: kirjakeelset korpust, mis koosneb Heinsoo õpematerjalide tekstidest ja murdeid kajastavat korpust, mis koosneb vadja sõnaraamatu Grünberg *et al.* 2013 näitelausetest. Korpused on päritavad Eesti Keeleressursside Keskuse korpuspäringute süsteemist KORP<sup>3</sup> ja avaldatud Kankainen 2018.

Morfoloogialaboriga salvestatud andmed ja nende kuju vastab suuresti pt 6.1 kirjeldatule. Siinkirjutajal oli au viibida Språkbanken'is prototüübi ehitamise ajal (2018. a kevadel) ja sai panustada andmete kuju standardiseerimisesse.

Magistritöö vadja vormisõnastiku sisestamise töövoog on olnud järgmine:

1. sisestada lekseemi põhivorm
2. valida olemasolev tüüpsõna olemasolevate sõnatüüpide põhjal
3. sõnatüübi (ja tüüpsõna) puudumisel sisestada lekseemi kõik muutvormid

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

<sup>1</sup><https://spraakbanken.gu.se/morfologilabbet/>

<sup>2</sup>Rootsi keelepank ehk keeleressursside keskus

<sup>3</sup><https://korp.keeleressursid.ee/?mode=vadja>

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 sõnatüübi, mis on tööprotsessi kiirendanud. Tihti on ka olnud abiks võimalus valida nt vokalisti suhtes vale sõnatüüp ja genereeritud muutvormides ainult korrigeerida vokalisti kajastavad osad.

Kuivõrd magistritöö jooksul on vajja sõnatüüpe ühtlustatud, on seda tehtud tagantjärele valides juba salvestatud sõnadele uue sõnatüübi. Kuna seda on pidanud tegema kõikide sõnade puhul, mis kuuluvad uue, ühtlustatud sõnatüübi alla, on see olnud aeganõudev protsess. Ühtlustamise ajal tekkinud mõtted, kuidas seda protsessi lihtsustada ja kiirendada, on edastatud Språkbanken'i arendajatele.

## 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 (ekstraheerimise) korral eraldab meetod lekseemi muutvormide tabelist selle 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	vormiinfo
<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>katoisõ</i>	SG ELA
<i>kattoisõ</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 morfoloogilise informatsiooniga.

tehniline tüvi	muutvormimall	vormiinfo
<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 muutujad  $x_1 = kat$  ja  $x_2 = o$  vastab sisendist leitud, allajoonitud tehnilisele tüvele).

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.

#### 4.1.1 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*.

#### 4.1.2 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

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

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

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. Sõnatüüp 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 programmkoodi 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 vormisõnastiku koostamisel. Seda on tehtud kuna magistritöö eesmärgiks on olnud esmane integreerimine keeletehnoloogilistesse raamis-

tikesse 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 uute sõnade sisestamisel neile pakutud sõnatüüpide järjestamiseks kasutajale.

Vormisõnastikku salvestatud lekseemid kannavad informatsiooni tehnilise tüve ja sõnatüübimalli kohta. Mida rohkem lekseeme jagavad 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 tavaid. Seda informatsiooni on võimalik kasutada selleks, et seada sõnatüübimallide muutujatele kitsendusi ja piirata selle ühtivate tehniliste tüvede hulka, ning tundmatu sõnavormi puhul ennustada selle kuulumist ühe või teise sõnatüüpi alla. (*ibid.*, 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 tundmatul sõnavormil ühtida sõnatüübiga. (*ibid.*, lk. 2580).

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

## 5 Vadja morfoloogiliste tüüpsõnade analüüs

See osa kirjeldab ekstraktmorfoloogiaga leitud vadja keele morfoloogilised tüüpsõnad ja jaotab need käändkondadesse. Käändkondade süsteemiks on kasutatud Ariste (1968) grammatikas esitatud.

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).

Käändkondadesse klassifitseerimine ei ole olnud range, vaid selle eesmärgiks on olnud koondada kokku sarnaselt käituvaid sõnu gruppideks. Seejärel on sõnu gruppide siseselt on seejärel püütud analoogia põhjal ühtlustada Tsvetkovi sõnaraamatus esinevat variatsiooni kirjakeele lihtsama õppimise eesmärgil. Need sõnad, mida ei ole suudetud klassifitseerida Ariste käändkondadesse, esitatakse eraldi.

Peatüki viimases osas analüüsitakse kui hästi Silfverberg, Liu ja Hulden 2018 esitatud ekstraktmorfoloogia üldiste muuttüüpide algoritm kajastab käändkondi.

Kõigis käändkondades on rekonstrueeritud lõpukaolised vokaalid ja ühtlustatud peamiselt lõpuvokaali õ:a vaheldumine. Ühtlustatud komponente on detailsemalt kirjeldatud iga käändkonna juures.

Alljärgnevatel kirjeldustel on lisatud tüüpsõnade muutvormimallid kujul, milles nende tehniline tüvi on allajoonitud. Selline esitus võimaldab lugejal paremini näha läbi ekstraktmorfoloogia muutvormimalli funktsioneerimist ja moodustada teiste sama tüüpsõna alla kuuluvate sõnade käändvorme.

### 5.1 I käändkond

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

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

Koostatud vormisõnastiku sõnade SG ILL vorme on esialgselt üsna kunstlikult üht-

lustatud 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 tüüpsõnade paljus on ajendatud ühelt poolt vokalismist (muutelõpu erinevus *või:võisõ* ja *täi:täise*). Teisalt aga pika türevokaali lühene-mise tõttu i-mitmuse vormides (*pääd:päije* ja *tüüd:tüije*), mistõttu jääb ühe tüüpsõna alla -ä-lised muutelõpud ja teise -ü-lised.

Üks avatud küsimus ja tähelepanek on SG ja PL vormide kokkulangemine difton-giga sõnade puhul.

### Ekstraktmorfoloogia tüüpsõnad

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-poolseid -aa-tüvelisi sõnu.

**puu** *puu*, *puutõ*, *puhhu*, *puuss*, *puud*, *puiõ*, *puut*, *puisõ*, *puiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *puu*, *kuu*, *luu*, *suu* ja *pihlpuu*. Tüüpsõna-mall kirjeldab tagapoolseid -uu-tüvelisi sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-seid -ii-tüvelisi sõnu.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *pää* ja *bulipää*. Tüüpsõnamall kirjeldab eespoolseid -ää-tüvelisi sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-poolseid -oo-tüvelisi sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-seid -ee-tüvelisi sõnu.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *tüü, vüü, üü* ja *püü*. Tüüpsõnamall kirjeldab eespoolseid -üü-tüvelisi sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid diftongiga sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall 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 probleemaa-  
tikast vadja Jõgõperä murdes on kirjutanud Rozhanskiy 2009.

### Ekstraktmorfoloogia tüüpsõnad

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega č:dž.

**süüčči** *süüči, süüččiä, süüččise, süüččiss, süüčid, süüččije, süüččiit, süüččise, süüččiiss*

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega čč:č.

**järčü** *järjü, järčüä, järčüse, järjüss, järjüd, järčüje, järčüit, järčüise, järčüiss*

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega rč:rj.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega ff:f, mille lõpuvokaal on i.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *suuto, vaahto* ja *lehto*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega t:Ø.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega t:Ø, mille lõpuvokaal on i.

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

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

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *lako*, *luku*, *mako*, *maku*, *suku*, *vako* ja *čako*. Tüüpsõnamall 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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega kk:k, mille lõpuvokaal on i.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega kk:k.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega kk:k, mille lõpuvokaal on i.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *flakku*, *herkku*, *jõkilikko*, *kakko*, *kakku*, *kiičku*, *kolčku*, *kukko*, *kurčku*, *kuuzikko*, *lepikko*, *liivikko*, *luikko*, *lukku*, *lõõčku*, *majakko*, *musikko*, *mäčizikko*, *naizikko*, *oomnikko*, *pettelikko*, *rehtelkakku*, *seukko*, *võrkko*, *õzrikko* ja *čerikko*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega kk:k.

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

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

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *põlto* ja *mõlto*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega lt:ll.

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

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

**löölü** löölü, löölüä, löölüse, löölüss, löölüd, löölüje, löölüit, löölüise, löölüiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *löölü*, *süüzü* ja *jürü*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *airo*, *aju*, *anõ*, *čaaju*, *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* ja *ahjo*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

**bagazi** bagazi, bagazia, bagazisõ, bagaziss, bagazid, bagazijõ, bagaziit, bagaziisõ, bagaziiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *bagazi*, *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* ja *almõzi*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *i*.

**postaljon** postaljoni, postaljonia, postaljonisõ, postaljoniss, postaljonid, postaljonijõ, postaljoniit, postaljoniisõ, postaljoniiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *postaljon* ja *parad*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega nt:nn, mille lõpuvokaal on *i*.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *lento*, *lintu*, *rokkalintu* ja *kanto*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega nt:nn.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega p:v.



**hapo** hapo, happoa, happosõ, haposs, hapod, happojõ, happoit, happoisõ, happoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega p:pp.

**vilppi** vilpi, vilppiä, vilppise, vilpiss, vilpid, vilppije, vilppiit, vilppiise, vilppiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme vilppi ja šlääppi. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega pp:p, mille lõpuvokaal on i.

**hüppü** hüpü, hüppüä, hüppüse, hüpüss, hüpüd, hüppüje, hüppüit, hüppüise, hüppüiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega pp:p.

**lippu** lipu, lippua, lippusõ, lipuss, lipud, lippujõ, lippuit, lippuisõ, lippuiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme lippu, lõppu, puippu ja kippu. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega pp:p.

**lamppi** lampi, lamppia, lamppisõ, lampiss, lampid, lamppijõ, lamppiit, lamppiisõ, lamppiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme lamppi, pappi, suppi, ukroppi ja kaappi. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega pp:p, mille lõpuvokaal on i.

**sese** seze, seseä, sessese, sezess, sezed, seseje, seseit, seseise, seseiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme sese ja läsü. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega s:z.

**siso** sizo, sissoa, sissoisõ, sizoss, sizod, sisojõ, sisoit, sisoisõ, sisoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme siso ja nisu. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega s:z.

**mahsu** mahzu, mahsua, mahsusõ, mahzuss, mahzud, mahsujõ, mahsuit, mahsuisõ, mahsuiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme mahsu ja haisu. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega s:z, mille tüvi konsonantklustri tõttu ei gemineeru.

**kursi** kurzi, kursia, kursisõ, kurziss, kurzid, kursijõ, kursiit, kursiisõ, kursiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega s:z, mille tüvi konsonantklustri tõttu ei gemineeru, mille lõpuvokaal on i.

**rusko** ruzgo, ruskoa, ruskosõ, ruzgoss, ruzgod, ruskojõ, ruskoit, ruskoisõ, ruskoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme rusko, tuisku, usko ja pääsko. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega sk:zg.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega ss:s, mille lõpuvokaal on *i*.

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

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

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

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

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega t:d, mille lõpuvokaal on *i*.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *kotko*, *laatko* ja *itku*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tk:dg.

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

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

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *rätte*, *tüttõ* ja *nenärätte*. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega tt:t.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *hattu*, *juttu*, *katto*, *kuttsu*, *laatto*, *lanttu*, *palto*, *porttu*, *Tarttu* ja *čiutto*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tt:t.

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

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

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tt:t.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *famili* ja *bašni*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutusega sõnu, mis eeldatavasti palatalisatsiooni tõttu käituvad eriliselt.

### 5.3 III käändkond

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

Vormisõnastikus esineb III käändkonnas siiski palju sõnu, mille esimeses silbis on *-u-*, mis Ariste järgi peaks kuuluma hoopis V käändkonda (Ariste 1968, lk. 44). Need sõnad on üldjuhul jäetud muutmata kujule ja on järgitud Tsvetkovi tüvekujusid. Näiteks *muuzikkõ*, *vunukkõ* (tüüpsõna *borovikkõ*); *suma* (tüüpsõna *baldõhina*); *bruudõ*, *pudrõ*, *surmõ* (tüüpsõna *azzõ*).

Tsvetkovi sõnaraamatus esineb rohkesti sõnu, mille mitmusetüvi on *-õi-*. Paljud nendest on vormisõnastikus ühtlustatud *-oi-* kujule.

Sellesse käändkonda kuuluvad paljud vene laensõnad, mille lõpuvokaalide ühtlustamine on olnud keeruline. Üldiselt on nende kuju jäetud muutmata ja need esinevad samal kujul, nagu Tsvetkovi sõnaraamatus.

#### Ekstraktmorfoloogia tüüpsõnad

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutusega sõnu

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega fk:vg.

**mahsõ** mahza, mahsa, mahsasõ, mahzõss, mahzõd, mahsojõ, mahsoit, mahsoisõ, mahsoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega hs:hz.

**vihtõ** viha, vihta, vihtasõ, vihõss, vihõd, vihtojõ, vihtoit, vihtoisõ, vihtoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega ht:h.

**aitõ** ajja, aita, aitasõ, aijõss, aijõd, aitojõ, aitoit, aitoisõ, aitoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega it:jj.

**aikõ** aiga, aika, aikasõ, aigõss, aigõd, aikojõ, aikoit, aikoisõ, aikoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme aikõ, jalkõ, liikõ, lõnkõ, nahkõ, rahkõ,  
vilkõ, vinkõ, võlkõ ja aastaikõ. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuu-  
tusega k:g.

**sika** siga, sikka, sikkasõ, sigass, sigad, sikojõ, sikoit, sikoisõ, sikoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega k:g, mis geminaadi tõttu esinevad SG ILL vormis tugevas  
astmes.

**borovikkõ** borovika, borovikka, borovikkasõ, borovikõss, borovikõd, borovikkojõ, borovikkoit,  
borovikkoisõ, borovikkoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme borovikkõ, domovikkõ, durakkõ, fartuk-  
kõ, fiizikkõ, fookusnikkõ, frištikkõ, gribanikkõ, harakkõ, itikkõ, joožikkõ, kaamenšik-  
kõ, kabakkõ, kamalikkõ, katolikkõ, kelkkõ, loomikkõ, kopekkõ, latikkõ, luzikkõ, luukkõ,  
markkõ, muuzikkõ, mõiznikkõ, noorikkõ, nuužnikkõ, obakkõ, paikkõ, palkkõ, pinžakkõ,  
podarkkõ, poštmarkkõ, rankkõ, rohosirkkõ, tarkkõ, tikkõ, tubakkõ, urokkõ, vakkõ, vu-  
nukkõ ja bašmukkõ. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega kk:k.

**siltõ** silla, silta, siltasõ, sillõss, sillõd, siltojõ, siltoit, siltoisõ, siltoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega lt:ll.

**paha** paha, pahha, pahasõ, pahõss, pahõd, pahhojõ, pahhoit, pahhoisõ, pahhoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid tüvemuutuseta sõnu, mille lõpuvokaal on a ning mis on SG PAR ja SG ILL  
vormis tugevas astmes.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *baldõhina, barabana, fotokartočka, grana, griba, kala, kana, liha, lina, litra, maja, raha, suma, sõna, tara, telefona, televizora, tila, vana* ja *astia*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *a*.

**azzõ** *azža, azža, azzasõ, azzõss, azzõd, azzõjõ, azzõit, azzõisõ, azzõiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *azzõ, baddõ, bahvõlõ, blaahõ, bobrõ, borkkanõ, bruudõ, čirjavõ, čirjõ, deelõ, dobrõ, filmõ, glaizõ, grammõ, gribavihmõ, iivõ, jumalõ, jurmõ, kabjõ, kaglõ, kagrõ, kajagõ, kambõlõ, kanavõ, karjõ, kassõ, katagõ, kavalõ, kvartirõ, laddõ, ladvõ, lahjõ, lahnõ, lainõ, laivõ, liivõ, linnõ, listõ, maailmõ, maamõ, mahlõ, mannõ, marjõ, matalõ, metlõ, muragõ, mussõmarjõ, naglõ, negrõ, niglõ, ostanofkõ, paglõ, progonõ, pudrõ, pupuškõ, rauhõ, saappõgõ, sarjõ, saunõ, siglõ, sisavõ, slifkõ, summõ, surmõ, suukkurliivõ, sõbrõ, šuubõ, ženihõ, taičinõ, trubõ, vihmõ, vikahtõ, vilõ, õravõ, õzrõ* ja *akkunõ*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *õ*.

**bank** *banka, banka, bankasõ, bankõss, bankõd, bankojõ, bankoit, bankoisõ, bankoiss*  
Tüüpsõna 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, penisioner, sammõl, zanaresk, žurnalist, tarelk, vaahtõr, viks* ja *ahvõn*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, millel puudub lõpuvokaal.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *pintõ, rantõ, rintõ* ja *kantõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega nt:nn.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *rapa, sõpa* ja *napa*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega p:v, mille lõpuvokaal on *a*.

**aapõ** *aava, aapa, aapasõ, aavõss, aavõd, aapojõ, aapoit, aapoisõ, aapoiss*

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega p:v, mille lõpuvokaal on *õ*.

**liippõ** *liipa, liippa, liippasõ, liipõss, liipõd, liippojõ, liippoit, liippoisõ, liippoiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *liippõ* ja *kauppõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega pp:p.

**partõ** *parra, parta, partasõ, parrõss, parrõd, partojõ, partoit, partoisõ, partoiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *partõ* ja *kõrtõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega rt:rr.

**kraaskõ** kraazga, kraaska, kraaskasõ, kraazgõss, kraazgõd, kraaskojõ, kraaskoit, kraaskoisõ, kraaskoiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *kraaskõ*, *laiskõ*, *nagriskaaskõ*, *niskõ*, *pas-kõ* ja *kaaskõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega sk:zg.

**klaassõ** klaasa, klaassa, klaassasõ, klaasõss, klaasõd, klaassojõ, klaassoit, klaassoisõ, klaassoiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *klaassõ* ja *bruussõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega ss:s.

**poduškõ** podužga, poduška, poduškasõ, podužgõss, podužgõd, poduškojõ, poduškoit, poduškoisõ, poduškoiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega šk:žg.

**dovariššõ** dovariša, dovarišša, dovariššasõ, dovariššass, dovarišad, dovariššojõ, dovariššoit, dovariššoisõ, dovariššois

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega šš:š.

**sata** saa, satta, sattasõ, saass, saad, satojõ, satoit, satoisõ, satoiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *sata*, *sõta* ja *muta*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega t:Ø.

**liitkõ** liidga, liitka, liitkasõ, liidgõss, liidgõd, liitkojõ, liitkoit, liitkoisõ, liitkoiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *liitkõ* ja *butkõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tk:dg.

**kandidaattõ** kandidaata, kandidaatta, kandidaattasõ, kandidaatõss, kandidaatõd, kandidaattojõ, kandidaattoit, kandidaattoisõ, kandidaattoiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *kandidaattõ*, *lauttõ*, *pliittõ*, *riittõ*, *žiivõttõ*, *taattõ* ja *gaazapliittõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tt:t.

**inostranttsõ** inostrantsa, inostranttsa, inostranttsasõ, inostrantsõss, inostrantsõd, inostranttsojõ, inostranttsoit, inostranttsoisõ, inostranttsois

Tüüpsõna hõlmab vormisõnastiku lekseeme *inostranttsõ*, *liittsõ*, *tablittsõ*, *vattsõ*, *õtsõ* ja *bolnittsõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega tts:ts.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega ut:vv, mille lõpuvokaal on *u*.

**poutõ** povva, pouta, poutasõ, povvõss, povvõd, poutojõ, poutoit, poutoisõ, poutoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *poutõ* ja *lautõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega ut:vv, mille lõpuvokaal on õ.

**štanad** štanojõ, štanoit, štanoisõ, štanass, štanad, štanojõ, štanoit, štanoisõ, štanoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolset tüvemuutusega *plurale tantum* sõna.

## 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 Ariste (1968, lk. 44) järgi kahesilbilised sõnad, mille türevokaal on -a ja millel esineb esimeses silbis -o-, -u- või -õ-.

Ariste järgi kuuluvad siia käändkonda enamik nendest sõnadest, mille esimene silp sisaldab -õ-d, kuigi need võiksid kuuluda ka III käändkonda (1968, lk. 44).

### Ekstraktmorfoloogia tüüpsõnad

**kuha** kuha, kuha, kuhasõ, kuhõss, kuhõd, kuhhijõ, kuhhiit, kuhhiisõ, kuhhiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolset tüvemuutusega sõna.

**poikõ** poiga, poika, poikasõ, poigõss, poigõd, poikijõ, poikiit, poikiisõ, poikiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *poikõ*, *rookõ* ja *lõukõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega k:g.

**biblioteekkõ** biblioteeka, biblioteekka, biblioteekkasõ, biblioteekõss, biblioteekõd, biblioteekkijõ, biblioteekkiit, biblioteekkiisõ, biblioteekkiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *biblioteekkõ*, *hoikkõ*, *ikolookkõ*, *jaanikukkõ*, *kolkkõ*, *konjõkkõ*, *kukkõ*, *rokkõ*, *sukkõ* ja *bambukkõ*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega kk:k.

**kõik** kõika, kõikka, kõikkasõ, kõikõss, kõikõd, kõikkijõ, kõikkiit, kõikkiisõ, kõikkiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega kk:k, millel puudub lõpuvokaal.

**kultõ** kulla, kulta, kultasõ, kullõss, kullõd, kultõjõ, kultõit, kultõisõ, kultõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega lt:ll.

**vimpõ** vimma, vimpa, vimpasõ, vimmõss, vimmõd, vimpijõ, vimpiit, vimpiisõ, vimpiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *vimpõ* ja *kumpõ*. Tüüpsõnamall kirjeldab  
tagapoolseid sõnu tüvemuutusega mp:mm.

**kompjutera** kompjutera, kompjutera, kompjuterasõ, kompjuterass, kompjuterad, kompjuterijõ,  
kompjuteriit, kompjuteriisõ, kompjuteriiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *kompjutera*, *kuja*, *loba*, *õmpõja*, *õpõttõja*  
ja *izõra*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille lõpuvokaal  
on *a*.

**maamuna** maamuna, maamuna, maamunasõ, maamunass, maamunad, maamunõjõ,  
maamunõit, maamunõisõ, maamunõiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *maamuna*, *muna* ja *kalindõra*. Tüüpsõna-  
mall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille mitmuse tüvi on *-õi-*.

**mokom** mokoma, mokoma, mokomasõ, mokomõss, mokomõd, mokomijõ, mokomiit,  
mokomiisõ, mokomiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid tüvemuutuseta sõnu, millel puudub lõpuvokaal.

**bukvõ** bukva, bukva, bukvasõ, bukvõss, bukvõd, bukvijõ, bukviit, bukviisõ, bukviiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *bukvõ*, *duumõ*, *formõ*, *ilmõ*, *koirõ*, *konnõ*,  
*kuhjõ*, *kuivõ*, *kummõ*, *kuuluvõ*, *kõrvõ*, *moodõ*, *muudrõ*, *mõizõ*, *obrazõ*, *postõ*, *program-*  
*mõ*, *roojõ*, *soolõ*, *sveezõ*, *trubačistõ*, *tuimõ*, *tuttavõ*, *velosipedõ*, *vohmõ*, *voolõ*, *võimõ* ja  
*bombõ*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu, mille lõpuvokaal  
on *õ*.

**propkõ** probga, propka, propkasõ, probgõss, probgõd, propkijõ, propkiit, propkiisõ, propkiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *propkõ*, *skopkõ* ja *jupkõ*. Tüüpsõnamall  
kirjeldab tagapoolseid sõnu tüvemuutusega pk:bg.

**rooppõ** roopa, rooppa, rooppasõ, roopõss, roopõd, rooppõjõ, roopõit, roopõisõ, roopõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega pp:p.



**kurp** *kurva, kurpa, kurpasõ, kurvõss, kurvõd, kurpijõ, kurpiit, kurpiisõ, kurpiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega rp:rv.

**usa** *uza, ussa, ussasõ, uzass, uzad, usijõ, usiit, usiisõ, usiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega s:z, mille lõpuvokaal on a.

**õhsõ** *õhza, õhsa, õhsasõ, õhzõss, õhzõd, õhsijõ, õhsiit, õhsiisõ, õhsiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *õhsõ* ja *sorsõ*. Tüüpsõnamall kirjeldab  
tagapoolseid sõnu tüvemuutusega s:z, mille lõpuvokaal on õ.

**luiskõ** *luizga, luiska, luiskasõ, luizgõss, luizgõd, luiskijõ, luiskiit, luiskiisõ, luiskiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega sk:zg.

**mussõ** *musa, mussa, mussasõ, musass, musad, mussõjõ, mussõit, mussõisõ, mussõiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *mussõ* ja *kapussõ*. Tüüpsõnamall kirjeldab  
tagapoolseid sõnu tüvemuutusega ss:s.

**moškõ** *možga, moška, moškasõ, možgõss, možgõd, moškijõ, moškiit, moškiisõ, moškiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *moškõ* ja *krõškõ*. Tüüpsõnamall kirjeldab  
tagapoolseid sõnu tüvemuutusega šk:žg.

**lootõ** *loovva, loota, lootasõ, loovvõss, loovvõd, lootijõ, lootiit, lootiisõ, lootiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega t:vv.

**biskvittõ** *biskvita, biskvitta, biskvittasõ, biskvitõss, biskvitõd, biskvittajõ, biskvittiit, biskvittiisõ, biskvittiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid sõnu tüvemuutusega tt:t.

## 5.6 VI käändkond

Kuuendasse käändkonda kuuluvad Ariste sõnul need sõnad, mis lõppevad *-õa/-eä/-iä*.  
Jõgõperä murret iseloomustab ta diftongi puudumisega sõna lõpus. Ariste 1968, lk. 44.

Tsvetkovi sõnaraamatus esineb suur variatiivsus mitmuse tüves. Variatiivsust on  
vormisõnastikus vähendatud ja ühtlustatud kolmele tüvele: tüüpsõna *kerkiä* tüvega  
*-ii-*, tüüpsõna *terve* tüvega *-ei-* ja tüüpsõna *kanka* tüvega *-õi-*.

### Ekstraktmorfoloogia tüüpsõnad

**kerkiä** *kerkiä, kerkiätä, kerkiäse, kerkiäss, kerkiäd, kerkije, kerkiit, kerkiise, kerkiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *kerkiä, pehmiä, pimmiä, siitiä ja jämiä*.  
Tüüpsõnamall kirjeldab eespoolseid tüvemuutusetä sõnu.

**terve** *terve, tervettä, tervese, tervess, terved, tervije, terveit, terveise, terveiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid tüvemuutusetä sõnu.

**kanka** *kanka, kankatō, kankasō, kankass, kankad, kankōjō, kankōit, kankōisō, kankōiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *kanka, kõrka, maikka, makka, ruska, valka, oika ja harma*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutusetä sõnu.

## 5.7 VII käändkond

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

### Ekstraktmorfoloogia tüüpsõnad

**väci** *väe, väččēä, väččese, väess, väed, väččije, väččiiit, väččiiise, väččiiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *väci* ja *mäci*. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega č:Ø.

**lahti** *lahō, lahtōa, lahtōsō, lahōss, lahōd, lahtijō, lahtiit, lahtiisō, lahtiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega ht:h.

**laki** *lagō, lakkōa, lakkōsō, lagōss, lagōd, lakijō, lakiit, lakiisō, lakiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *laki, nōki* ja *jōki*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega k:g.

**kurki** *kurgō, kurkōa, kurkōsō, kurgōss, kurgōd, kurkijō, kurkiit, kurkiisō, kurkiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *kurki, ōnki* ja *kanki*. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega k:g, kusjuures tüvemuutus esineb konsontklustris, mistõttu gemineerumist ei toimu SG PAR ja SG ILL tüvedes.

**pōski** *pōzgō, pōskōa, pōskōsō, pōzgōss, pōzgōd, pōskijō, pōskiit, pōskiisō, pōskiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega sk:zg.

**irvi** *irve, irveä, irvese, irvess, irved, irvije, irviit, irviise, irviiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *irvi, järvi, leemi, nimi, pilvi* ja *čivi*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu.

**sarvi** *sarvõ, sarvõa, sarvõsõ, sarvõss, sarvõd, sarvijõ, sarviit, sarviisõ, sarviiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *sarvi, savi, Soomi, sõrmi, taimi, talvi, tammi, õnni, õvvi* ja *põlvi*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega *nč:nñi*.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega *p:v*.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega *s:z*.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid sõnu tüvemuutusega *t:d*.

## 5.8 VIII käändkond

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

### Ekstraktmorfoloogia tüüpsõnad

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega *č:Ø*.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega *lč:ll*.

**eine** *einä, einä, einäse, einess, eined, einije, einiit, einiise, einiiss*

Tüüpsõna hõlmab vormisõnastiku lekseeme *eine, esimespäive, fõkle, irvieine, lehme, läikkive, läsive, nätilpäive, petäje, piime, pominpäive, pädre, päive, rehte, seine, silme, tühje, äjje, ämme* ja *čülme*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *e* ja mitmuse tüvi on *-ii-*.

**läkine** läkinä, läkinä, läkinäse, läkiness, läkined, läkineje, läkineit, läkineise, läkineiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *läkine* ja *dääde*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *e* ja mitmuse tüvi on *-ei-*.

**räpäle** räpälä, räpälä, räpäläse, räpäless, räpäled, räpälöje, räpälöit, räpälöise, räpälöiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *e* ja mitmuse tüvi on *-öi-*.

**rissimä** rissimä, rissimä, rissimäse, rissimäss, rissimäd, rissimije, rissimit, rissimise, rissimiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *rissimä* ja *emä*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu, mille lõpuvokaal on *ä*.

**pähčen** pähčenä, pähčenä, pähčenäse, pähčeness, pähčened, pähčenije, pähčeniit, pähčeniise, pähčeniiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *pähčen* ja *ičäv*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu, millel puudub lõpuvokaal.

**läsijä** läsijä, läsijä, läsijäse, läsijess, läsijed, läsijije, läsijiit, läsijiise, läsijiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *läsijä*, *müüjä*, *tečejä* ja *köühä*. Tüüpsõnamall kirjeldab eespoolseid tüvemuutuseta sõnu.

**kitai** kitaja, kitaja, kitajasõ, kitajass, kitajad, kitajojõ, kitajoit, kitajoisõ, kitajoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

**slona** slona, slona, slonasõ, slonass, slonad, slonijõ, slonõit, slonõisõ, slonõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

**õmpõlija** õmpõlija, õmpõlija, õmpõlijasõ, õmpõlijass, õmpõlijad, õmpõlijijõ, õmpõlijait, õmpõlijaisõ, õmpõlijaiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

**kompjutera** kompjutera, kompjutera, kompjuterasõ, kompjuterass, kompjuterad, kompjuterijõ, kompjuteriit, kompjuteriisõ, kompjuteriiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *kompjutera*, *kuja*, *loba*, *õmpõja*, *õpõttõja* ja *izõra*. Tüüpsõnamall kirjeldab tagapoolseid tüvemuutuseta sõnu.

**čenče** čennä, čenčä, čenčäse, čenness, čenned, čenčije, čenčiit, čenčiise, čenčiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespoolseid sõnu tüvemuutusega nč:nn.

**tünke** tünga, tünkä, tünkäse, tüngess, tünged, tünki, tünkiit, tünkiise, tünkiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega nk:ng.

**änte** ännä, äntä, äntäse, änness, änned, äntije, äntiit, äntiise, äntiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega nt:nn.

**leipe** leivä, leipä, leipäse, leivess, leived, leipije, leipiit, leipiise, leipiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega p:v.

**seppe** sepä, seppä, seppäse, sepäss, sepäd, seppi, seppiit, seppiise, seppiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *seppe* ja *leppe*. Tüüpsõnamall kirjeldab  
eespoolseid sõnu tüvemuutusega pp:p.

**ärče** ärjä, ärčä, ärčäse, ärjess, ärjed, ärčije, ärčiit, ärčiise, ärčiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *ärče* ja *märče*. Tüüpsõnamall kirjeldab  
eespoolseid sõnu tüvemuutusega rč:rj.

**pesä** pezä, pessä, pessäse, pezäss, pezäd, pesije, pesiit, pesiise, pesiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *pesä*, *rissisä* ja *česä*. Tüüpsõnamall kir-  
jeldab eespoolseid sõnu tüvemuutusega s:z, mille mitmuse tüvi on -ii-.

**lisä** lizä, lissä, lissäse, lizäss, lizäd, lissäije, lissäit, lissäise, lissäiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega s:z, mille mitmuse tüvi on -äi-.

**mätä** mää, mättä, mättäse, määss, määäd, mättije, mättiit, mättiise, mättiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega t:Ø.

**mettse** metsä, mettsä, mettsäse, metsäss, metsäd, mettsije, mettsiit, mettsiise, mettsiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid sõnu tüvemuutusega tts:ts.

**nenä** nenä, nennä, nenäse, nenäss, nenäd, nenije, neniit, neniise, neniiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid tüvemuutuseta sõnu, mille tüvi gemineerub -nn-.

**čülä** čülä, čüllä, čüläse, čüläss, čüläd, čülije, čüliit, čüliise, čüliiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid tüvemuutuseta sõnu, mille tüvi gemineerub -ll- SG PAR vormis.

**pühä** *pühä, pühhä, pühhäse, pühhäss, pühäd, pühhiije, pühhiit, pühhiise, pühhiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid tüvemuutuseta sõnu, mille tüvi gemineerub -hh- SG PAR vormis.

**üvä** *üvä, üvvä, üvväse, üväss, üväd, üvvije, üvviit, üvviise, üvviiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab eespool-  
seid tüvemuutuseta sõnu, mille tüvi gemineerub -vv- SG PAR vormis.

**kõva** *kõva, kõvat, kõvasõ, kõvass, kõvad, kõvijõ, kõviit, kõviisõ, kõviiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid tüvemuutuseta sõnu, mille tüvi gemineerub -vv- SG PAR vormis.

**õja** *õja, õjja, õjasõ, õjass, õjad, õjijõ, õjit, õjisõ, õjiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid tüvemuutuseta sõnu, mille tüvi gemineerub -jj- SG PAR vormis.

**õma** *õma, õmma, õmmasõ, õmõss, õmõd, õmmijõ, õmmiit, õmmiisõ, õmmiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus. Tüüpsõnamall kirjeldab taga-  
poolseid tüvemuutuseta sõnu, mille tüvi gemineerub -mm- SG PAR vormis.

## 5.9 IX käändkond

See käändkond on spetsiifiline Kattila murdele ja ei esine loodava vadja kirjakeele  
aluseks olevas murdes.

## 5.10 X käändkond

Kümnendasse käändkonda koondub suur osa kahetüvelisi sõnu, mille SG NOM lõpp  
on -i, ent mille tüvevokaal on -e/-õ (Ariste 1968, lk. 47). Ariste märgib, et kuigi SP PAR  
vormil on mitu erinevat realisatsiooni, järgib nende moodustamisviis põhimõtteliselt  
sama malli (*ibid.*, lk. 47). Ekstraktmorfoloogiaga eraldatud muutsõnamallide vähene  
arv näitab, et meetod pädeb nende põhimõttete tuvastamises.

Käändkonna muutsõnamallides on kirjakeele järgi ühtlustatud kõikide sõnade lõ-  
puvokaalid -i:-õ:-tõ ja -i:-e:-te vastavalt sõna vokaalharmooniale.

## Ekstraktmorfoloogia tüüpsõnad

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

**iiri** iire, iirte, iirese, iiress, iired, iirije, iiriit, iiriise, iiriiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *iiri*, *meeli*, *meri*, *peeni*, *süli*, *veri*, *ääni*, *ääri* ja *čeeli*.

**juuri** juurõ, juurtõ, juurõsõ, juurõss, juurõd, juurijõ, juuriit, juuriisõ, juuriiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *juuri*, *kaani*, *koori*, *lõhi*, *noori*, *ooli*, *pooli*, *sooli*, *suuri*, *tuli*, *tuuli*, *uni* ja *hooli*.

**uhsi** uhzõ, ussõ, uhsõsõ, uhzõss, uhzõd, uhsijõ, uhsiit, uhsiisõ, uhsiiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *uhsi* ja *lahsi*.

**kusi** kuzõ, kussõ, kussõ, kuzõss, kuzõd, kussijõ, kussiit, kussiisõ, kussiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**uusi** uuvvõ, uuttõ, uuvvõsõ, uuvvõss, uuvvõd, uusijõ, uusiit, uusiisõ, uusiiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *uusi*, *voosi* ja *kuusi*.

Märkus: kuigi lekseem *voosi:voovvõ* hääldub 'vuuvvõ' on see analoogiliselt teiste sõnade muutustega kirjakeeles märgitud *voovvõ*.

**ühs** ühe, ühte, ühtese, ühess, ühed, ühsije, ühsiit, ühsiise, ühssiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**ülči** ülle, ülčiä, ülčese, ülless, ülled, ülčije, ülčiit, ülčiise, ülčiiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *ülči* ja *jälči*.

**kahs** kahõ, kahtõ, kahtõ, kahõss, kahõd, kahsijõ, kahsiit, kahsiisõ, kahsiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**varsi** varrõ, varttõ, vartõsõ, varrõss, varrõd, varsijõ, varsiit, varsiisõ, varsiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**mesi** mee, mette, mettese, meess, meed, mesije, mesiit, mesiise, mesiiss

Tüüpsõna hõlmab vormisõnastiku lekseeme *mesi*, *süsi*, *vesi* ja *čäsi*.

**susi** suõ, suttõ, suttõsõ, suõss, suõd, susijõ, susiit, susiisõ, susiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

## 5.11 XI käändkond

Üheteistkümnendasse käändkonda liigitab Ariste need sõnad, mille SG NOM lõpp on -Z, ent mille vokaaltüvi sõltub murdest: Jõgõperä murdes realiseerub see kui -s-, Kattila murdes kui -hs- ja teistes vadja murretes kui -ss- (Ariste 1968, lk. 48). Vormisõnastikus on järgitud Jõgõperä murdele omast realisatsiooni -s-.

Vadja keele sandhi nähtust on kirjakeeles ühtlustatud -z lõpulisteks.

### Ekstraktmorfoloogia tüüpsõnad

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *makuz*, *nagriz*, *paganuz* ja *kolauz*.

**rihenneüz** rihenneüse, rihenneüsse, rihenneüsesse, rihenneüsess, rihenneüsed, rihenneüsije, rihenneüsiit, rihenneüsiise, rihenneüsiiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

## 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 *ibid.*, lk. 49.

Mitmuse vormide vokaaltüved on kunstlikult ühtlustatud tugevaastmeliseks (-s-). Tsvetkovi sõnaraamatus esitatud sõnades esineb suur varieeruvus tüveastme puhul ja siinkirjutaja ei ole suutnud leida mingit reeglipära. Sellega seotud on küsimus, kas PL GEN tüvi peaks vahelduma teistest PL käänete astmest.

Pika vokaaliga lõpevate sõnade puhul esineb ka varieeruvust: *süsiin* ja *õnnõliin* puhul on see pikk ainult SG NOM ja SG PAR käänetes, ning *magnettiin* ja *talviin* puhul on see pikk igas käändes.

### Ekstraktmorfoloogia tüüpsõnad

**õpõin** õpõizõ, õvõissõ, õpõzõsõ, õpõizõss, õpõizõd, õpõizijõ, õpõiziit, õpõiziisõ, õpõiziiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.



**entin** *entize, entisse, entizese, entizess, entized, entisije, entisiit, entisiise, entisiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *entin, esimein, inimin, jäin, miltäin, rehe-*  
*lin, sinin, venäläin* ja *eglin*.

**süsiin** *süsize, süsiisse, süsizese, süsizess, süsized, süsisije, süsisiit, süsisiise, süsisiiss*  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

**čimolain** *čimolaizõ, čimolaissõ, čimolaizõsõ, čimolaizõss, čimolaizõd, čimolaisijõ, čimolaisiit,*  
*čimolaisiisõ, čimolaisiiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *čimolain, greeklain, hatukkõin, iirikkõin,*  
*illõkkõin, iloin, jõkain, kehnokkõin, keskolin, kõikõllain, kõrvõlin, leivekkõin, luin, lät-*  
*tilain, magnetiin, main, mokomõin, mustõlain, nain, partõin, perennain, prikukkõin,*  
*puin, roottsilain, ruskolain, saunlain, soomõlain, sopuin, sukulain, talviin, tarttulain,*  
*tõin, ukrainalain, virolain, õhtõgoin, ühellain* ja *audžikkõin*.

**koivuin** *koivuizõ, koivuissõ, koivuizõsõ, koivuizõss, koivuizõd, koivuizijõ, koivuiziit,*  
*koivuiziisõ, koivuiziiss*  
Tüüpsõna hõlmab vormisõnastiku lekseeme *koivuin, kultõin, kõltõin, pakkõin, rohoin,*  
*uutin, voosin* ja *kalttõin*.

### 5.13 XIII käändkond

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

Käesolevas töös on siia käändkonda koondatud need sõnad, mis SG GEN käändes  
lõpevad *-me/-mä/-ma/-mõ/-ne/-re*. Suure variatiivsuse tõttu on käändkonda kuuluvad  
sõnad jäetud ühtlustamata. Variatiivsust esineb nii Tsvetkovi sõnaraamatu sees, kui  
ka õppevahendite vahel. Konkoval (2014) on esitatud *võttim:võttimõ:võttima* ent Aris-  
tel, Tsvetkovil? ja Heinsool on *võti*.

Näiteks Tsvetkovil on *süä:süä:süttä/süät*.

## Ekstraktmorfologia tüüpsõnad

**čümme** čümmene, čümmenä, čümmenese, čümmess, čümmed, čümmenije, čümmeniit, čümmeniise, čümmeniiss

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna hõlmab vormisõnastiku lekseeme *õnnõto* ja *hoolito*.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

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

Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

## 5.14 XIV käändkond

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

Vadja keele -Z-lõpu sandhi nähtus on kõigi liikmete puhul ühtlustatud -z-lõpulisteks SG NOM käändes.

Juhul kui Tsvetkovi sõnaraamatus on antud paralleelvorme, on neist püütud valida selline, mis esineb kas Heinsoo ja Konkova sõnastikes, või VKSi Vaipooli murdelistes näitelausestes. Näiteks on valitud *lähe* tugevaastmeline SG tüvi näitelausete põhjal.

### Ekstraktmorfoloogia tüüpsõnad

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

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

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

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

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

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

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

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

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

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

**hammõz** hampa, hammassõ, hampasõ, hampass, hampad, hampajõ, hampait, hampaisõ, hampaiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme hammõz, lammõz ja ammõz.

**lõunõ** lõuna, lõunattõ, lõunasõ, lõunass, lõunad, lõunajõ, lõunait, lõunaisõ, lõunaiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kangõz** kanka, kangõssõ, kankasõ, kankass, kankad, kankajõ, kankait, kankaisõ, kankaiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kauniz** kauni, kaunissõ, kaunisõ, kauniss, kaunid, kaunijõ, kauniit, kauniisõ, kauniiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**angõriaz** angõria, angõriassõ, angõriasõ, angõriass, angõriad, angõrijõ, angõriit, angõriisõ, angõriiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**raskõz** raska, raskassõ, raskasõ, raskass, raskad, raskojõ, raskait, raskaisõ, rakaiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kasõ** kassõ, kasõttõ, kassõsõ, kassõss, kassõd, kassõjõ, kassõit, kassõisõ, kassõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**vetelüz** vetelüse, vetelüsse, vetelüsse, vetelüsess, vetelüsed, vetelüssije, vetelüssiit, vetelüssiise, vetelüssiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme vetelüz ja jänez.

**põrzõz** põrsa, põrzassõ, põrsasõ, põrsass, põrsad, põrsojõ, põrsoit, põrsoisõ, põrsoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**ratiz** ratti, ratissõ, rattisõ, rattiss, rattid, rattijõ, rattiit, rattiisõ, rattiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**katõ** kattõ, katõttõ, kattõsõ, kattõss, kattõd, kattõjõ, kattõit, kattõisõ, kattõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kuõ** kuõtõ, kuõttõ, kuõtõsõ, kuttõss, kuttõd, kuttõjõ, kuttõit, kuttõisõ, kuttõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**varvõz** varpa, varvassõ, varpasõ, varpass, varpad, varpajõ, varpait, varpaisõ, varpaiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *varvõz* ja *turvõz*.

**pere** pere, perette, perese, peress, pered, pereje, pereit, pereise, pereiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *pere*, *vene* ja *erne*.

## 5.15 XV käändkond

Viieteistkümnendasse käändkonda koondab Ariste sõnu nagu *lühüd*, *õhud*, *koollu*, *ilozuZ*, *rikkauZ* Ariste 1968, lk. 51.

Ka selle käändkonna liikmete -Z-lõpu sandhi nähtus on ühtlustatud -z-lõpuliseks SG NOM käändes.

### Ekstraktmorfoloogia tüüpsõnad

**čämmel** čämmele, čämmelte, čämmeläse, čämmeless, čämmeled, čämmeliije, čämmeliit, čämmeliise, čämmeliiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**oonõ** oonõ, oonõt, oonõsõ, oonõss, oonõd, oonijõ, ooniit, ooniisõ, ooniiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kannõl** kannõlõ, kannõltõ, kannõlasõ, kannõlõss, kannõlõd, kannõlijõ, kannõliit, kannõliisõ, kannõliiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**peremmeez** peremmehe, peremmeesse, peremmehese, peremmehess, peremmehed, peremmehije, peremmehiit, peremmehiise, peremmehiiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *peremmeez* ja *meez*.

**märännü** märännü, märännüt, märännüse, märänness, märänned, märännije, märänneit, märänneise, märänneiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**čeväd** čevvä, čevätte, čevättese, čeväss, čeväd, čevviije, čevviit, čevviise, čevviiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**čäčüd** čäčcū, čäčütt, čäčcüse, čäčcüss, čäčcüd, čäčcūje, čäčcūit, čäčcūise, čäčcūiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**ivusõd** ivussijõ, ivussiit, ivussisõ, ivusõss, ivusõd, ivussijõ, ivussiit, ivussiisõ, ivussiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kaatsõd** kaattsojõ, kaattsoit, kaattsoisõ, kaatsõss, kaatsõd, kaatsojõ, kaattsoit, kaattsoisõ,  
kaattsoiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**koollud** koollõ, koolluttõ, koollõsõ, koollõss, koollõd, koollujõ, koolluit, koolluisõ, koolluiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**koorrõ** koorrõ, koorrõt, koortõsõ, koorrõss, koorrõd, koortõjõ, koortõit, koortõisõ, koortõiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**kõlmõd** kõlmõ, kõlma, kõlmõsõ, kõlmõss, kõlmõd, kõlmijõ, kõlmiit, kõlmiisõ, kõlmiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**olud** olu, oluttõ, olusõ, oluss, olud, olujõ, oluit, oluisõ, oluiiss  
Tüüpsõna ei hõlma teisi lekseeme vormisõnastikus.

**toho** toho, tohtõ, tohtosõ, tohoss, tohod, tohojõ, tohoit, tohoisõ, tohoiss  
Tüüpsõna hõlmab vormisõnastiku lekseeme *toho* ja *roho*.

## 6 Programmikoodi tuletamine

Programmikoodi 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 faktsiaalset *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.

Programmikoodi 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 programmkood 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).

Tüüpsõnamall *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 Tüüpsõna malli esitamine LMFis

Ekstraktmorfoloogiaga eraldatud tüüpsõnad 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 *as-Katto*).

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 51).

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

```

<MorphologicalPattern>
  <feat att="id" val="asTšiuutto"/>
  <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 tšiuutto (mille alla kuuluvad mh *tšiuutto* 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 täielikult kontekstitundlike *fully context-sensitive* grammatikate vahele (*ibid.*, 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>4</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 morfoloogiameetodi ja teine vadja leksikonimeetodi.

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>4</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 sõnatüübile 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.

Tüüpsõnamallist paradigmafunktsiooni genereerimiseks on vaja määrata funktsiooni nimi, liides (*interface*) ja keha (*body*). Genereeritud programmkoodi 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*i*s.

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 programmkoodi 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 vajja 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?

Loodud õigekirjakontrollija on eesmärgipäraselt jäetud lihtsakoeliseks. See märgib kõik sõnad valeks, mis ei sisaldu sõnastikus. See on lühiajaliseks kasutamiseks ja mõeldud ärgitama 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  $aabe$ . 60

**Lemma** on suvaliselt valitud grammatiliste tunnuste komplekt, mida kasutatakse lekseemi viitamiseks. 60

**Mikrostruktuur** on sõnastiku sõnaartikli sisemine struktuur. 60

**Muuttüüp** on morfoloogilise klassifikatsiooni elementaariusus. See on sõnaklass, mis erineb teistest sõnaklassidest mingite erijoonte poolest sõnade muutmisel. 60

**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. 60

**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. 15, 60

**Tehniline tüvi** on tähtkoostiste järjend, millega saab tüüpsõnamalli muutvormide muutujad asendada elik väärtustada ja niiviisi koostada ühe konkreetse sõna kõik vormid. 15, 60

## 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.
- 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.
- 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).
- 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.
- Grünberg, Silja et al., toim. (2013). *Vadja keele sõnaraamat =: Vaddaa tseelee sõna-tširja = Словарь водского языка*. 2., täiend. ja parand. tr. Tallinn: Eesti Keele Sihtasutus. 1823 lk. kokku. ISBN: 978-9985-79-553-8.
- 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.
- 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).

- Kankainen, Kristian (15. oktoober 2018). *keeleleek/votic-corpora: Version 1.0*. DOI: 10.5281/zenodo.1462648. URL: <https://zenodo.org/record/1462648> (vaadatud 30.04.2019).
- 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.
- 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. Kõide 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.
- Muuk, Elmar (1933). *Väike õigekeelsus-sõnaraamat*. 1. ja 2. trükk. Tartu.
- Ranta, Aarne (2008). „Grammars as software libraries“. DOI: 10.1017/CB09780511770524.014.
- (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.
- 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).

- Viks, Ülle (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.
- 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.