

Ekstraktmorfoloogia meetodiga tuletatud keele tehnoloogia vadjä sõnavara näitel

Kristian Kankainen

2019

Sisukord

1	Sissejuhatus	4
2	Teoreetilised lähtekohad	6
2.1	Arvutimorfoloogia eesmärk ja lingvistiline motiveeritus	6
2.2	Paradigmaatiline morfoloogia	11
2.3	Vadja kirjakeel ja normatiiv	11
2.4	Ortograafia	12
2.5	Morfofonoloogia	12
2.6	Klassikaline paradigmaatiline morfoloogia	13
2.7	Morfeemi staatus ja definitsioon	13
2.8	Muuttüüp, tüüpsõna ja muutkond	13
3	Ekstraktmorfoloogia meetod	14
3.1	Sissejuhatus	14
4	Vadja morfoloogiliste tüüpsõnade analüüs	16
4.1	Ekstraktmorfoloogiaga leitud tüüpsõnad	16
4.2	Põhivormid ja analoogiavormid	16
4.2.1	Käändsõnad	17
4.2.2	Tegusõnad	17
4.3	Üks võimalik muuttüüpide süsteem	17
4.4	Muuttüüpide produktiivsus	17
5	Programmkoodi tuletamine	18
5.1	Keskne kirjeldus Lexical Markup Framework vormingus	18
5.1.1	Sõnaartiklite esitamine LMFis	19
5.1.2	Tüüpsõnamallide esitamine LMFis	19
5.2	Grammatical Framework morfoloogiakomponent	19
5.2.1	Leksikon	19
5.2.2	Muuttüübid	19
5.2.3	Arutelu	19
5.3	Integreerimine Giella-taristuga	20
5.3.1	Leksikon	20
5.3.2	Muuttüübid	20
5.3.3	Õigekirjakontrollija	21
5.3.4	Arutelu	21
6	Kokkuvõte	22
7	Põhimõisted ja lühendid	23
8	Kirjandus	24

9	The use of Extract Morphology for Automatic Derivation of Language Technology for Votic	26
10	Lisad	27

1 Sissejuhatus

Magistritöö esimene eesmärk on luua H. Heinsoo Sõnakopitaja esitatud sõnavarast morfoloogiline sõnastik, mis sisaldab sõnavara kõiki muutvorme. Selleks vajalik arvutimorfoloogiline kirjeldus ehitatakse sellisel moel, et see taandub tüüpsõnade muutvormitabelite esitamisele, mitte grammatiliste reeglite esitamisele. Niiviisi ehitatud teooriavaba(m) arvutimorfoloogiline kirjeldus võimaldab luua erinevaid keeletehnoloogiaid automaatselt programmkoodi tuletamise teel. Esitatakse kolme tehnoloogia automaatset tuletamist: 1) ühe keeletehnoloogilise taristusse integreerimise kaudu õigekirjakontrollija, 2) vadja keele arvutimorfoloogia moodul ühe loomulike keelte grammatikate koostamiseks mõeldud programmeerimiskeelele ja 3) morfoloogia tehnoloogiaülene kirjeldus ühe rahvusvahelise standardi abil.

Kuna kõik tuletatud keeletehnoloogia edaspidine täiendamine ja täpsustamine käib ainult lekseemide muutvormitabelite täiendamise ja täpsustamise kaudu, peab esimese eesmärgi juurde lisama seda, et magistritöös loodud leksikograafiline süsteem võimaldab keeleaktivistide rühmal töötada oma sõnavara ja keeletehnoloogia kallal edaspidi ka ilma spetsialistist keeleteadlase ja keeletehnoloogi abil. Kas seda vadja keele puhul ka juhtub, jääb tuleviku näidata.

Magistritöö teine eesmärk on analüüsida leitud tüüpsõnad mitmel viisil: 1) kirjeldada nende morfofonoloogiat keeleajalooliste arengute taustal, 2) leida tüüpsõnade põhivormid ja analoogiavormid, 3) esitada üks võimalik muuttüüpide süsteem ja võrrelda seda seni esitatutega ja viimalt 4) analüüsida muuttüüpide produktiivsust.

—

Magistritöö loob viisi ehitada arvutimorfoloogia puhtalt lekseemide sõnavormide esitamise teel ning teisendada ehitatud arvutimorfoloogilise mudeli automaatselt kahte keeletehnoloogilisse raamistikku.

Magistritöö kasutab loodud süsteemi selleks, et kirjeldada vadja keele normatiivsed morfoloogilised tüüpsõnad.

Tööd ajendab mõtteviis minimeerida tööd: loodud normatiivne morfoloogiline tüübistik on aluseks automaatselt tuletatud keeletehnoloogiale, kui normatiiv muu-

tub, muutub ka keeletehnoloogia. Töö paneb leksikaalse ressursi esikohale ja kõik leitud sisulised vead õiendatakse otse ressursis, mitte keeletehnoloogilistes tarkvarades eraldi.

2 Teoreetilised lähtekohad

Kuna töö opereerib arvutilingvistika, deskriptiivse ja dokumentaalse lingvistika ääremail, peame selgitama töö teoreetilised lähtekohad. Siinsele kompendiumiks on ka põhimõisted seletatud pt 7 Põhimõisted ja lühendid.

Töö püüdleb olla võimalikult teooriavaba, lastes vadja lekseemide sõnavormide tähtkoostised ise määrata nende paradigmade koostamisreeglid. (See on olla deduktiivne esialgse morfoloogia postuleerimises, vastandudes induktiivsele, s.o mingist grammatilisest kirjeldusest lähtudes.)

Tööl on siiski teoreetilised lähtekohad, mis tulenevad ühelt poolt arvutimorfoloogia nõuetest ja teisalt klassikalisest paradigmaatilisest morfoloogiakäsitlusest. Järgmiselt püüan argumenteerida, et arvutimorfoloogia ei pea olema lingvistisest teoriast ajendatud. Seejärel tutvustan tööle kõige lähedamini asetsevat morfoloogilist käsitlust.

2.1 Arvutimorfoloogia eesmärk ja lingvistiline motiveeritus

Arvutimorfoloogia eesmärgiks on siin töös valitud moodustada elektroonse vormisõnastiku:

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

Sellist vormisõnastikku võib kujutada erinevatel viisidel. Näiteks leksikaalse andmebaasina, kus iga lekseemi puhul on nenditud kõik selle muutevormid koos vastava grammatilise iseloomustusega. Matemaatilises mõttes kujutab vormisõnastik vaid seost muutevormide ja nende vastavate grammatiliste iseloomustuste vahel.

Arvutimorfoloogiad võivad seda seost, või vormisõnastiku funktsionaalsust, realiseerida arvutuslikult erinevatel viisidel ja ei pea olema lingvistiliselt motiveeritud.

Aga ühe täieliku vormisõnastiku koostamise maht on niivõrd suur (kui mitte lõpmatult suur) ja see on peamine motivatsioon organiseerida vormisõnastiku arvutusliku külje ühel või teisel moel lingvistiliselt.

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

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

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

Inkrementaalsete teooriate järgi lisandub iga (olgu IA järgi leksikaalselt loetletud või IP mõttes inferentsiaalse reeglga tuletatud) morfosüntaktilise tunnuse puhul sõnale ka selle vormiline eksponent (Stump 2001, lk 2). Vormilised eksponendid on üks-üheses seoses grammatiliste tunnustega ja need väljenduvad ükshaaval ehk ink-

rementaalselt.

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

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

Stumpi jagab oma klassifikatsiooni järgi Lieberi morfoloogilise teooria leksikaalseks ja inkrementaalseks. Halle ja Marantzi distributsioonilise morfoloogia teooria leksikaalseks ja realiseerivaks. Steele'i artikuleeritud morfoloogia teooria esindab inferentsiaalset ja inkrementaalset suunda.

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

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

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

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

Käesolevas magistritöös rakendatud ekstraktmorfoloogia on oma organisatsiooni suhtes WP mudel, aga selle käsitus morfeemist on lähedasem Beardi teooriale.

Arvutilingvistikas on arvutimorfoloogiat üldiselt organiseeritud klassikalise morfeemikäsituse järgi. Seda ilmestab hästi Lauri Karttunen, kes nendib inimkeele mudelda-

mise puhul arvutimorfoloogias kaks väljakutset: 1) morfotaktika ehk sõnast väiksemate üksuste kombineerumine ja 2) morfoloogilised vaheldused ehk sõnast väiksemate üksuste kuju olenemine nende ümbritsevast kontekstist (Karttunen 2003).

Kusjuures (1) viitab otseselt klassikalisele morfeemikäsitusele.

Karttuneni artikkel on aga ajendatud Stumpi teooriast ja ta näitlikustab kuidas Stumpi teooria on võimalik rakendada kasutades lõplike automaatide formalismi.

Karttunen toob siiski välja olukorra, et arvutilingvistide morfoloogiad põhinevad arvutuslikel meetoditel, millega nad implementeerivad morfoloogiaid ja mitte ei põhine otse mingil teoorial. Seega võib öelda, et arvutimorfoloogia on laiem kui lingvistiline morfoloogia, kuna esimest piirab mitte teooria, vaid arvutusliku meetodi võimsus. Karttunen näitabki, et Stumpi inferentsiaalne-realiseeriv teooria on taandatav lõplike automaatide formalismi arvutusvõimsusele. (Karttunen 2003)

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

“Morphophonemes are represented just as the combinations of the corresponding letters (or phonemes) which we can observe in the surface forms.

On the one hand, such an interpretation of morphophonemes is crude, but on the other hand, it is a fact that anybody can observe.” (Koskenniemi 2013, lk 157)

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

ndail).

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

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

Probleem, miks arvutimorfoloogiad juhinduvad lingvistilistest teooriatest ja mitte puht-arvutuslikest võimalustest asub formaliseerimisprotsessi 3. ja 4. etapi vahel. Millisel viisil tuleb põhjendada struktuuri määravaid reegleid?

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

Eelnevaga olen ma tahtnud öelda seda, et arvutimorfoloogiatega koostamispõhimõtted

põhineda (morfoloogilistel) reeglitel ja mitte puhtalt muutvormide nentimisel korpusel põhjal, on eeskätt ajalooliste traditsioonide järjepidevus. Käesolev töö ei järgi neid traditsioone.

2.2 Paradigmaatiline morfoloogia

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

Klassikaline on lähedasem siinses magistratöös rakendatud ekstraktmorfoloogiale. Klassikalise ja strukturalistlike lähenemiste suurim vahe seisneb kahes asjaolus. Eeskätt ei näinud klassikalised grammatikud sõnast väiksemat ja tähenduslikku üksust. Arvestati ainult vormikülje üksustega, tähtede-foneemide ja silpidega. Sõna oli väikseim tähenduslik element keeles ja seda nähti tervikuna, hõlmates terve oma paradigma kuuluvaid vorme. (Matthews 1991 lk X–Y)

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

Õpikutes ja grammatikates välja toodud reeglid, mille abil sai ühe lekseemi paradigma moodustada, opereerisid ainult sõnavormide tähtkoostisel. Näiteks võidi ühe sõnavormi lõputähti asendada teiste tähtedega, et saada teine sõnavorm. Et asendatavatele tähtkoostistele ei omandatud mingit tähendust, näitlikustab see, et mõne reegli algvormiks võidi valida selline sõnavorm, mis oma tähtkoostise poolest kõige paremini sobis. (Matthews 1991 lk X–Y).

2.3 Vadja kirjakeel ja normatiiv

Vadja keelele ei loodud kirjandust 1930-ndatel, nagu seda tehti Nõukogude Liidus näiteks karjala, vepsa ja isuri keele jaoks.

Siiski on vadja keelel hulganisti lingvistilisi kirjeldusi, nagu grammatikaid (mh Ahlqvist 1856; Airila 1934; Tsvetkov 2008; Ariste 1968; Маркус ja Рожанский 2011), sõnaraamatuid (mh Tsvetkov 1995; Ariste 1943; Laakso 1989; Raag 1982; Pomberg ja

H. Heinsoo 1991; Grünberg *et al.* 2013; Heinike Heinsoo 2015) ja ka etnograafilisi töid (mh Kass 1961; Mälk 1977).

Kirjeldused ei aita siiski kaasa tänapäeva keeleõppija küsimustele *kuidas kirjutatakse sõna TÜTTÖ mitmuse omastavas?*. Selleks on vaja tänapäevase vadja keele morfoloogia standardiseerimist ehk normatiivset kirjeldust.

Käesolev töö ei pürgi looma lõplikku normatiivi, kuivõrd ta loob süsteemi, mis oskab vastata morfoloogilistele küsimustele. Aga loodud süsteemi peamine eesmärk on siiski võimaldada muuta ja jätkata tööd normatiivi arendamiseks ja mille ümber saaks keeleaktivistid ise koonduda, ilma et selleks oleks niivõrd vaja ei lingvistilist ega keeletehnoloogilist spetsialisti.

Püüd luua vadja morfoloogiale normatiivne alus lihtsustab paljudele küsimustele vastusi leida, nt mis käändeid arvestada. Siiski on tööga loodud *keeletehnoloogia tuletamise süsteem* avatud ka teistsugustele lähenemistele keeleainesele.

2.4 Ortograafia

On järgitud Heinsoo loodud ortograafiat mille jaoks on Kankainen teinud vadja klaviatuuripaigutise (Kankainen, ilummas). Vadja erinevatest kirjaviisidest on kirjutatud Ernits 2010 ja erinevatest kirjakeele loomise pürgimistest Ernits 2006.

2.5 Morfofonoloogia

Tavaliselt jagatakse arvutilingvistikas morfoloogia ja fonoloogia eraldi nii, et morfoloogia tasandil on abstraktne esitus, nn morfofoneemid, mille pindesitused tulenevad eraldi fonoloogilistest reeglitest.

Niiviisi saaks esitada mõlemad fonoloogilised vormid *tšiuttoa* ja *tüttöä* ühe ja sama morfoloogilise kujuga TŠIUTTO+A ja TÜTTÖ+A. Kusjuures käändelõpu +A pindesitus *a*-na või *ä*-na sõltuks vastavalt sellest, kas lemmas esineb tagapoolsed või eespoolsed vokaalid.

See töö ei arvesta morfofonoloogilise tasandiga. Peatükis 4.3 näidatakse üht võimalikku viisi koondada tüüpsõnu kokku abstraktsemal tasandil, mis mingil määral

arvestab ka morfofonoloogilisi reeglipärasusi.

2.6 Klassikaline paradigmaatiline morfoloogia

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

2.7 Morfeemi staatus ja definitsioon

Morfeemi ei käsitleta siin töös levinud lingvistilisest seisukohast kui *väikseimat tähenduslikku üksust*, vaid klassikalistele paradigmaatilistele lähenemistele omaselt kui *mistahes tähtkoostise muutust, millega kaasneb tähenduslik muutus* (Beard 1987; Beard 1995). Morfeemipõhist suunda ajab nt Stump 2001.

2.8 Muuttüüp, tüüpsõna ja muutkond

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

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

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

3 Ekstraktmorfoloogia meetod

See osa kirjeldab töös rakendatud ekstraktmorfoloogia meetodit. Töö kasutab ekstraktmorfoloogiat kaheks otstarbeks, esiteks vadja keele morfoloogiliste tüüpsõnade väljaselgitamiseks ja kirjeldamiseks ja teisalt programmkoodi automaatseks tuletamiseks saadud kirjelduse põhjal. Neid kahte rakendust kirjeldatakse lähemalt vastavates peatükkides *Vadja morfoloogiliste tüüpsõnade analüüs* ja *Programmkoodi tuleamine*.

3.1 Sissejuhatus

Ekstraktmorfoloogia on juhendatud masinõppe meetod, mis üldistab lekseemide muutvormitabeleid ja eraldab neist tüüpsõnamallid. See on *juhendatud*, sest sisendiks olevad muutvormitabelid peavad olema korrektselt koostatud.

Selles töös käsitletakse meetodi abil saadud mudelit siiski pigem lihtsa kirjeldusena. See on tüüpsõnakirjeldus, mis on osa sõnastikust – lekseemi paradigma kirjeldusena. Ja sellest kirjeldusest

Tüüpsõnamall koosneb muutvormide mallidest ja vastab seega morfoloogilise paradigma mõistele. Tüüpsõnamalli abil on võimalik moodustada ka tundmatu sõna kõik muutvormid. Kuna kaks või enam lekseemi võivad jagada üht ja sama tüüpsõnamalli (s.o kuuluda sama paradigmasse), on võimalik ekstraktmorfoloogia meetodiga üldistada lekseemide iseärasusi ja luua nendest tüüpsõnade produktiivsuse mudeli. Produktiivsusemudeliga on võimalik ennustada uue ja tundmatu sõnavormi kuuluvust ühe või teise tüüpsõna alla.

Veel ilma detailidesse takerdumata näitlikustatakse siinkohal lugejale meetodi sisendit ja väljundit. Sisendiks on ühe lekseemi muutvormitabel tervikuna (vt tabel 1). Väljundiks on meetodi poolt leitud tüüpsõnamall (vt tabel 2). Tabelitele viidatakse alljärgnevas tekstis mitmel korral.

muutvorm	tunnused
<i>katto</i>	SG NOM
<i>katod</i>	PL NOM
<i>kato</i>	SG GEN
<i>kattoi</i>	PL GEN
<i>kattoje</i>	PL GEN
<i>kattoa</i>	SG PART
<i>kattoi</i>	PL PART
<i>kattoite</i>	PL PART
<i>kattoise</i>	SG ILL
<i>kattoise</i>	PL ILL
<i>kattoz</i>	SG INE
<i>kattoiz</i>	PL INE
<i>katosse</i>	SG ELA
<i>kattoisse</i>	PL ELA
<i>katolle</i>	SG ALL
<i>kattoille</i>	PL ALL
<i>katol</i>	SG ADE
<i>kattoil</i>	PL ADE
<i>katolte</i>	SG ABL
<i>kattoilte</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 1: Sisendi muutvormide tabel koos morfoloogiliste tunnustega.

ühisosajada	muutvormimall	tunnused
<u>kat</u> t o	$x_1 + t + x_2$	SG NOM
<u>kat</u> o d	$x_1 + x_2 + d$	PL NOM
<u>kat</u> o	$x_1 + x_2$	SG GEN
<u>kat</u> t o i	$x_1 + t + x_2 + i$	PL GEN
<u>kat</u> t o je	$x_1 + t + x_2 + je$	PL GEN
<u>kat</u> t o a	$x_1 + t + x_2 + a$	SG PART
<u>kat</u> t o i	$x_1 + t + x_2 + i$	PL PART
<u>kat</u> t o ite	$x_1 + t + x_2 + ite$	PL PART
<u>kat</u> t o se	$x_1 + t + x_2 + se$	SG ILL
<u>kat</u> t o ise	$x_1 + t + x_2 + ise$	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	$x_1 + x_2 + sse$	SG ELA
<u>kat</u> t o isse	$x_1 + t + x_2 + isse$	PL ELA
<u>kat</u> o lle	$x_1 + x_2 + lle$	SG ALL
<u>kat</u> t o ille	$x_1 + t + x_2 + ille$	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 lte	$x_1 + x_2 + lte$	SG ABL
<u>kat</u> t o ilte	$x_1 + t + x_2 + ilte$	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 2: Väljundi tüüpsõnamall (kusjuures $x_1 = kat$ ja $x_2 = o$ vastab mallist leitud ühisosajadale).

4 Vadja morfoloogiliste tüüpsõnade analüüs

See osa kirjeldab ekstraktmorfoloogiaga leitud vadja keele morfoloogilisi tüüpsõnu ja analüüsib nende vastavust vadja keele grammatikatega (?) ja ajaloolise morfoloogiaga (?).

4.1 Ekstraktmorfoloogiaga leitud tüüpsõnad

See alaosa loendab leitud tüüpsõnad sõnaliigiti. Analüüsitakse tüüpsõnade alla kuuluvaid sõnu struktuurselt (kui mitu silpi, silpide struktuur).

Analüüsides on võimalik luua arvutikirjeldus hüpoteetilise vadjakeelse sõna üle õigekirjakontrollija jaoks.

4.2 Põhivormid ja analoogiavormid

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

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

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

4.2.1 Käändsõnad

4.2.2 Tegusõnad

4.3 Üks võimalik muuttüüpide süsteem

Silfverberg ja Hulden (2018) on kirjeldanud üht võimalikku formaalset viisi, kuidas ekstraktmorfoloogia tüüpsõnade arvu vähendada. Siin alaosas esitatakse selle põhjal loodud vadja muuttüübistik ja võrreldatakse seda Eesti muuttüüpide traditsiooniga.

Eesti muuttüüpide traditsioonist on kirjutanud mh Viks 2018.

4.4 Muuttüüpide produktiivsus

Kristiina Kross (Ross) nimetab produktiivsuseks “mingi morfoloogilise nähtuse võimet allutada endale uusi sõnu” (Kross 1984). Siin allosas seatakse eelmises osas leitud muuttüübid pingeritta selle järgi, kui mitu tüüpsõna nendele allub.

Kas selleks on vaja defineerida, mis on *uus sõna*? Näiteks kõik uuemad vene keele laenud.

Kas produktiivsuse pingerida on võimalik jagada mingi kriteeriumi järgi avatuteks ja suletuteks muuttüüpideks?

5 Programmkoodi tuletamine

Programmikoodi tuletamise 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 faktuaalselt *mida* ning tuletatud programmkood kirjeldab konkreetset *kuidas* seda teadmist rakendada.

Töös kasutatakse keskseks kirjelduseks leksikaalset ressursi, mille peamine osa koosneb ekstraktmorfoloogiaga leitud tüüpsõnade mallidest.

Keskse kirjelduse leksikaalset ressursi hoitakse rahvusvahelise standardi vormingus *Lexical Markup Framework* (ISO/TC 37/SC 4 2007).

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

5.1 Keskne kirjeldus Lexical Markup Framework vormingus

Sissejuhatav tekst, mis on e-sõnastike ja leksikaalsete andmebaaside rahvusvaheline standard Lexical Markup Framework (ISO/TC 37/SC 4 2007) ja milleks seda kasutatakse. (märksõnu: semantika eeldefineeritud märgenduskeel; koostöövõime)

Standardi märgenduskeel koosneb mitmest eriotstarbelisest laiendimoodulist (vt nt Francopoulo 2013). Siinne töö kasutab kahte: morfoloogia moodul (*LMF Morphology Extension*) ja morfoloogiliste paradigmade moodul (*LMF Morphological Pattern Extension*).

Morfoloogiamooduli eesmärgiks on kirjeldada morfoloogiat mahu kaudu, s.o kirjeldada lekseemi loendades kõiki selle muutvorme.

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

Sama nähtuse kirjeldamine nii mahus kui ka sisus võib tunduda liigsena, ent nii viisi võimaldatakse rohkem informatsiooni hoidmist.

Näiteks võib iga lekseemi muutvormi kohta hoida informatsiooni nende reaalsest korpusesinemustest. Niiviisi on võimalik klassifitseerida tüüpsõnade teoretiseeringutaset, kui ühe ja sama tüüpsõna alla kuuluvate lekseemide korpusleiud kinnitavad igat selle muutvormi, ei ole see teoretiseeritud.

Peale sõnaartiklite ja morfoloogilise informatsiooni hoitakse leksikaalses ressursis ka globaalset informatsiooni, nagu keele nimetus ja kood.

5.1.1 Sõnaartiklite esitamine LMFis

Iga sõnaartikkel ehk leksikaalne kirje kannab informatsiooni lekseemi sõnaliigi kohta, selle valitud lemma vorm ning morfoloogiamooduliga esitatud muutvormitabeli.

5.1.2 Tüüpsõnamallide esitamine LMFis

5.2 Grammatical Framework morfoloogiakomponent

Mis on see, mida mina teen. Seejärel, mis on programmeerimiskeel Grammatical Framework ja milleks seda kasutatakse.

Morfoloogiakomponendi programmkood on jaotatav kaheks tükiks, leksikoniks ja muuttüüpide funktsioonid. Järgnevalt neist detailsemalt. Viimases alaosas on arutelu GFide võimalustest ja edasiarendusvõimalustest.

5.2.1 Leksikon

5.2.2 Muuttüübid

5.2.3 Arutelu

Loodud morfoloogiakomponenti on kasutatud interaktiivses vадja-vene-вадja vestmikus.

5.3 Integreerimine Giella-taristuga

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.

5.3.1 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*.

5.3.2 Muuttüübid

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

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

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

5.3.3 Õigekirjakontrollija

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

5.3.4 Arutelu

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.

6 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 keelekümbuskoolis 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 kesksele 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.

7 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 $aabe$. 23

Lemma on suvaliselt valitud grammatiliste tunnuste komplekt, mida kasutatakse lekseemi viitamiseks. 23

Mikrostruktuur on sõnastiku sõnaartikli sisemine struktuur. 23

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. 23

Tehniline tüvi on tähtkoostiste järjend, millega saab tüüpsõnamalli muutvormide muutujad asendada elik väärtustada ja niiviisi koostada ühe konkreetse sõna kõik vormid. 23

Tüüpsõnamall 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. 14, 23

8 Kirjandus

- Ahlqvist, August (1856). *Wotisk grammatik jemte språkprof och ordförteckning: (Föredr. d. 15 Oktober 1855)*. [Helsingfors: s.n. 162 lk. kokku.
- Airila, Martti (1934). *Vatjan kielen taivutusoppi. 1: Nominien taivutus*. Vähäisiä kirjelmiä 87. Helsinki: Suomalaisen kirjallisuuden seura. 55 lk. kokku.
- Ariste, Paul (1943). *Vadja lemmüz: mõningaid vadja sõnaseletusi: avec résumés français*. Helsinki: s.n. 1 lk. kokku.
- (1968). *A grammar of the Votic language*. Indiana University publications. Uralic and Altaic series vol. 68. Bloomington : The Hague: Indiana University ; Mouton. 121 lk. kokku.
- Auroux, Sylvain *et al.*, toim. (2006). *History of the Language Sciences: An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present*. Köide 3. ISBN: 3-11-016736-0.
- Beard, Robert (1987). „Morpheme order in a lexeme/morpheme-based morphology“. *Lingua* 72.1, lk. 1–44.
- (1995). *Lexeme-morpheme Base Morphology: A General Theory of Inflection and Word Formation*. SUNY Series in Linguistics. OCLC: 940540414. State University of New York Press.
- Beesley, Kenneth R ja Lauri Karttunen (2003). *Finite state morphology*. Stanford, Calif.: CSLI Publications. ISBN: 1-57586-433-9 978-1-57586-433-4 1-57586-434-7 978-1-57586-434-1.
- Chomsky, Noam (1957). *Syntactic Structures*. OCLC: 979583577. ISBN: 978-3-11-021832-9. URL: <https://doi.org/10.1515/9783110218329> (vaadatud 19.02.2019).
- Erelt, Mati, Tiiu Erelt ja Kristiina Ross (2007). *Eesti Keele Käsiraamat*. 3., täiend. tr. Tallinn: Eesti Keele Sihtasutus. 726 lk. kokku. ISBN: 978-9985-79-210-0.
- Ernits, Enn (2006). „Vadja liikumisest ja kirjakeelest“. *Keel ja Kirjandus* 49.1, lk. 85–87. URL: <https://www.digar.ee/viewer/et/nlib-digar:81648/143905/page/85>.
- (2010). „Vadja kirjaviisist ja sõnaloomest“, lk. 17.
- Forsberg, Markus ja Mans Hulden (2016). „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 tšeelee 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.
- Hockett, Charles F. (august 1954). „Two Models of Grammatical Description“. *WORD* 10.2, lk. 210–234. ISSN: 0043-7956, 2373-5112. DOI: 10.1080/00437956.1954.11659524. URL: <http://www.tandfonline.com/doi/full/10.1080/00437956.1954.11659524> (vaadatud 03.09.2018).
- 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 (vaadatud 13.06.2017).
- Karlsson, Fred *et al.* (2002). *Üldkeeleteadus*. Tallinn: Eesti Keele Sihtasutus. 362 lk. kokku. ISBN: 978-9985-811-96-2.

- Karttunen, Lauri (2003). „Computing with realizational morphology“. Teoses: *Computational linguistics and intelligent text processing*. Springer, lk. 203–214. (Vaadatud 18.08.2015).
- Kass, Asta (1961). „Käsitöö- ja rõivastusalane sõnavara vadjä keeles“. Tartu.
- Koskeniemi, Kimmo (22. juuli 2013). „An informal discovery procedure for two-level rules“. *Journal of Language Modelling* 1.1, lk. 155–188. ISSN: 2299-8470. DOI: 10.15398/jlm.v1i1.62. URL: <http://jlm.ipipan.waw.pl/index.php/JLM/article/view/62> (vaadatud 26.09.2016).
- Kross, Kristiina (1984). *Eesti noomeni muutmistüüpide produktiivsus: soome-eesti kontrastiivseminar, (Helsingi, 12.-20. okt. 1984): [ettekanne]*. Koostöös Keele ja Kirjanduse Instituut ja Eesti NSV Teaduste Akadeemia. Ars grammatica KKI-26. Tallinn: Eesti NSV Teaduste Akadeemia. 40 lk. kokku.
- Laakso, Johanna, toim. (1989). *Vatjan kääntheissanasto*. Lexica Societatis Fenno-Ugricae 22. Helsinki: Suomalais-ugrilainen seura. 103 lk. kokku. ISBN: 978-951-9403-21-2.
- 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.
- Mälk, Vaina (1977). *Vadja vanasõnad eesti, soome, karjala ja vene vastetega*. Koostöös Keele ja Kirjanduse Instituut. Tallinn: Eesti Raamat. 404 lk. kokku.
- Pomberg, Merle ja H. Heinsoo (1991). „Vadja tööriistade ja tarbeesemete nimetused Eesti Rahva Muuseumis: kursusetöö“. Tartu.
- Raag, Virve (1982). *A dictionary of Votic*. Uppsala: Fenno-ugrica suecana. 230 lk. kokku.
- Silfverberg, Miikka, Ling Liu ja Mans Hulden (20. august 2018). „A Computational Model for the Linguistic Notion of Morphological Paradigm“, lk. 12.
- Stump, Gregory T (2001). *Inflectional morphology a theory of paradigm structure*. Cambridge; New York: Cambridge University Press. ISBN: 978-0-511-01378-2 978-0-521-78047-6 978-0-511-48633-3. URL: <http://dx.doi.org/10.1017/CB09780511486333> (vaadatud 19.07.2015).
- Tsvetkov, Dmitri (1995). *Vatjan kielen Joenperän murteen sanasto*. Helsinki: Suomalais-ugrilainen seura ;Kotimaisten kielten tutkimuskeskus. ISBN: 978-951-9403-83-0.
- (2008). *Vadja Keele Grammatika*. Koostöös Jüri Viikberg, Ada Ambus ja Eesti Keele Instituut. Tallinn: Eesti Keele Sihtasutus. 169 lk. kokku. ISBN: 978-9985-79-216-2.
- 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.
- (2018). *Muuttüübid eesti ükskeelsetes sõnastikes*. URL: <https://www.eki.ee/teemad/tyybijutt.html> (vaadatud 04.03.2018).
- 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.

9 The use of Extract Morphology for Automatic Derivation of Language Technology for Votic

An English language summary of this work.

ühisosajada	muutvormimall	tunnused
<u>aa</u> põ	$x_1 + p\tilde{o}$	SG NOM
<u>aa</u> vōd	$x_1 + v\tilde{o}d$	PL NOM
<u>aa</u> va	$x_1 + va$	SG GEN
<u>aa</u> poi	$x_1 + poi$	PL GEN
<u>aa</u> pojō	$x_1 + poj\tilde{o}$	PL GEN
<u>aa</u> pa	$x_1 + pa$	SG PAR
<u>aa</u> poi	$x_1 + poi$	PL PAR
<u>aa</u> poitō	$x_1 + poit\tilde{o}$	PL PAR
<u>aa</u> paa	$x_1 + paa$	SG ILL
<u>aa</u> pasō	$x_1 + pas\tilde{o}$	SG ILL
<u>aa</u> poisō	$x_1 + pois\tilde{o}$	PL ILL
<u>aa</u> vaz	$x_1 + vaz$	SG INE
<u>aa</u> voiz	$x_1 + voiz$	PL INE
<u>aa</u> vōssō	$x_1 + v\tilde{o}ss\tilde{o}$	SG ELA
<u>aa</u> poissō	$x_1 + poiss\tilde{o}$	PL ELA
<u>aa</u> vōllō	$x_1 + v\tilde{o}ll\tilde{o}$	SG ALL
<u>aa</u> poillō	$x_1 + poill\tilde{o}$	PL ALL
<u>aa</u> vōl	$x_1 + v\tilde{o}l$	SG ADE
<u>aa</u> poil	$x_1 + poil$	PL ADE
<u>aa</u> vōssi	$x_1 + v\tilde{o}ssi$	SG ABL
<u>aa</u> poissi	$x_1 + poissi$	PL ABL
<u>aa</u> passi	$x_1 + passi$	SG TRA
<u>aa</u> poissi	$x_1 + poissi$	PL TRA
<u>aa</u> passaa	$x_1 + passaa$	SG TER
<u>aa</u> poissaa	$x_1 + poissaa$	PL TER
<u>aa</u> vaka	$x_1 + vaka$	SG COM
<u>aa</u> poika	$x_1 + poika$	PL COM

Tabel 3: Tüüpsõna *aapō* ekstraheeritud muutvormimall.

10 Lisad

Siin on esitatud kõik ekstraheeritud tüüpsõnamallide tabelid.

ühisosajada	muutvormimall	tunnused
<u>sini</u> n	$x_1 + n$	SG NOM
<u>sini</u> zed	$x_1 + zed$	PL NOM
<u>sini</u> ze	$x_1 + ze$	SG GEN
<u>sini</u> zije	$x_1 + zije$	PL GEN
<u>sini</u> sse	$x_1 + sse$	SG PAR
<u>sini</u> zii	$x_1 + zii$	PL PAR
<u>sini</u> ziit	$x_1 + ziit$	PL PAR
<u>sini</u> ze	$x_1 + ze$	SG ILL
<u>sini</u> zese	$x_1 + zese$	SG ILL
<u>sini</u> zije	$x_1 + zije$	PL ILL
<u>sini</u> zise	$x_1 + zise$	PL ILL
<u>sini</u> zež	$x_1 + zež$	SG INE
<u>sini</u> ziz	$x_1 + ziz$	PL INE
<u>sini</u> zess	$x_1 + zess$	SG ELA
<u>sini</u> ziss	$x_1 + ziss$	PL ELA
<u>sini</u> zelle	$x_1 + zelle$	SG ALL
<u>sini</u> zille	$x_1 + zille$	PL ALL
<u>sini</u> zell	$x_1 + zell$	SG ADE
<u>sini</u> zill	$x_1 + zill$	PL ADE
<u>sini</u> zelt	$x_1 + zelt$	SG ABL
<u>sini</u> zilt	$x_1 + zilt$	PL ABL
<u>sini</u> zessi	$x_1 + zessi$	SG TRA
<u>sini</u> zissi	$x_1 + zissi$	PL TRA
<u>sini</u> zeessaa	$x_1 + zeessaa$	SG TER
<u>sini</u> ziissaa	$x_1 + ziissaa$	PL TER
<u>sini</u> zeka	$x_1 + zeka$	SG COM
<u>sini</u> zika	$x_1 + zika$	PL COM

Tabel 4: Tüüpsõna *sinin* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>ai</u> kõ	$x_1 + \text{kõ}$	SG NOM
<u>ai</u> gõd	$x_1 + \text{gõd}$	PL NOM
<u>ai</u> ga	$x_1 + \text{ga}$	SG GEN
<u>ai</u> koi	$x_1 + \text{koi}$	PL GEN
<u>ai</u> kojõ	$x_1 + \text{kojõ}$	PL GEN
<u>ai</u> ka	$x_1 + \text{ka}$	SG PAR
<u>ai</u> kaa	$x_1 + \text{kaa}$	SG PAR
<u>ai</u> koi	$x_1 + \text{koi}$	PL PAR
<u>ai</u> koitõ	$x_1 + \text{koitõ}$	PL PAR
<u>ai</u> ka	$x_1 + \text{ka}$	SG ILL
<u>ai</u> kasõ	$x_1 + \text{kasõ}$	SG ILL
<u>ai</u> koisõ	$x_1 + \text{koisõ}$	PL ILL
<u>ai</u> gõz	$x_1 + \text{gõz}$	SG INE
<u>ai</u> koiz	$x_1 + \text{koiz}$	PL INE
<u>ai</u> gõssõ	$x_1 + \text{gõssõ}$	SG ELA
<u>ai</u> koissõ	$x_1 + \text{koissõ}$	PL ELA
<u>ai</u> gõllõ	$x_1 + \text{gõllõ}$	SG ALL
<u>ai</u> koillõ	$x_1 + \text{koillõ}$	PL ALL
<u>ai</u> gõl	$x_1 + \text{gõl}$	SG ADE
<u>ai</u> koil	$x_1 + \text{koil}$	PL ADE
<u>ai</u> gõltõ	$x_1 + \text{gõltõ}$	SG ABL
<u>ai</u> koiltõ	$x_1 + \text{koiltõ}$	PL ABL
<u>ai</u> gõssi	$x_1 + \text{gõssi}$	SG TRA
<u>ai</u> koissi	$x_1 + \text{koissi}$	PL TRA
<u>ai</u> kassaa	$x_1 + \text{kassaa}$	SG TER
<u>ai</u> koissaa	$x_1 + \text{koissaa}$	PL TER
<u>ai</u> gaka	$x_1 + \text{gaka}$	SG COM
<u>ai</u> koika	$x_1 + \text{koika}$	PL COM

Tabel 5: Tüüpsõna *aikõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>poi</u> kõ	$x_1 + \text{kõ}$	SG NOM
<u>poi</u> gõd	$x_1 + \text{gõd}$	PL NOM
<u>poi</u> ga	$x_1 + \text{ga}$	SG GEN
<u>poi</u> ki	$x_1 + \text{ki}$	PL GEN
<u>poi</u> kije	$x_1 + \text{kije}$	PL GEN
<u>poi</u> ka	$x_1 + \text{ka}$	SG PAR
<u>poi</u> kaa	$x_1 + \text{kaa}$	SG PAR
<u>poi</u> ki	$x_1 + \text{ki}$	PL PAR
<u>poi</u> kitõ	$x_1 + \text{kitõ}$	PL PAR
<u>poi</u> kaa	$x_1 + \text{kaa}$	SG ILL
<u>poi</u> kasõ	$x_1 + \text{kasõ}$	SG ILL
<u>poi</u> ki	$x_1 + \text{ki}$	PL ILL
<u>poi</u> kisõ	$x_1 + \text{kisõ}$	PL ILL
<u>poi</u> gõz	$x_1 + \text{gõz}$	SG INE
<u>poi</u> kiz	$x_1 + \text{kiz}$	PL INE
<u>poi</u> gõssõ	$x_1 + \text{gõssõ}$	SG ELA
<u>poi</u> kissõ	$x_1 + \text{kissõ}$	PL ELA
<u>poi</u> gõllõ	$x_1 + \text{gõllõ}$	SG ALL
<u>poi</u> killõ	$x_1 + \text{killõ}$	PL ALL
<u>poi</u> gõl	$x_1 + \text{gõl}$	SG ADE
<u>poi</u> kil	$x_1 + \text{kil}$	PL ADE
<u>poi</u> gõltõ	$x_1 + \text{gõltõ}$	SG ABL
<u>poi</u> kiltõ	$x_1 + \text{kiltõ}$	PL ABL
<u>poi</u> gõssi	$x_1 + \text{gõssi}$	SG TRA
<u>poi</u> kissi	$x_1 + \text{kissi}$	PL TRA
<u>poi</u> kassaa	$x_1 + \text{kassaa}$	SG TER
<u>poi</u> kissaa	$x_1 + \text{kissaa}$	PL TER
<u>poi</u> gaka	$x_1 + \text{gaka}$	SG COM
<u>poi</u> kika	$x_1 + \text{kika}$	PL COM

Tabel 6: Tüüpsõna *poikõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>tüt</u> t <u>ö</u>	$x_1 + t + x_2$	SG NOM
<u>tüt</u> t <u>ö</u> d	$x_1 + x_2 + d$	PL NOM
<u>tüt</u> t <u>ö</u>	$x_1 + x_2$	SG GEN
<u>tüt</u> t <u>ö</u> i	$x_1 + t + x_2 + i$	PL GEN
<u>tüt</u> t <u>ö</u> je	$x_1 + t + x_2 + je$	PL GEN
<u>tüt</u> t <u>ö</u> ä	$x_1 + t + x_2 + ä$	SG PAR
<u>tüt</u> t <u>ö</u> i	$x_1 + t + x_2 + i$	PL PAR
<u>tüt</u> t <u>ö</u> ite	$x_1 + t + x_2 + ite$	PL PAR
<u>tüt</u> t <u>ö</u> se	$x_1 + t + x_2 + se$	SG ILL
<u>tüt</u> t <u>ö</u> ise	$x_1 + t + x_2 + ise$	PL ILL
<u>tüt</u> t <u>ö</u> z	$x_1 + t + x_2 + z$	SG INE
<u>tüt</u> t <u>ö</u> iz	$x_1 + t + x_2 + iz$	PL INE
<u>tüt</u> t <u>ö</u> sse	$x_1 + x_2 + sse$	SG ELA
<u>tüt</u> t <u>ö</u> isse	$x_1 + t + x_2 + isse$	PL ELA
<u>tüt</u> t <u>ö</u> lle	$x_1 + x_2 + lle$	SG ALL
<u>tüt</u> t <u>ö</u> ille	$x_1 + t + x_2 + ille$	PL ALL
<u>tüt</u> t <u>ö</u> l	$x_1 + x_2 + l$	SG ADE
<u>tüt</u> t <u>ö</u> il	$x_1 + t + x_2 + il$	PL ADE
<u>tüt</u> t <u>ö</u> lte	$x_1 + x_2 + lte$	SG ABL
<u>tüt</u> t <u>ö</u> ilte	$x_1 + t + x_2 + ilte$	PL ABL
<u>tüt</u> t <u>ö</u> ssi	$x_1 + x_2 + ssi$	SG TRA
<u>tüt</u> t <u>ö</u> issi	$x_1 + t + x_2 + issi$	PL TRA
<u>tüt</u> t <u>ö</u> ssaa	$x_1 + t + x_2 + ssaa$	SG TER
<u>tüt</u> t <u>ö</u> issaa	$x_1 + t + x_2 + issaa$	PL TER
<u>tüt</u> t <u>ö</u> ka	$x_1 + x_2 + ka$	SG COM
<u>tüt</u> t <u>ö</u> ika	$x_1 + t + x_2 + ika$	PL COM

Tabel 7: Tüüpsõna *tüttö* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>koir</u> õ	$x_1 + \tilde{o}$	SG NOM
<u>koir</u> õd	$x_1 + \tilde{o}d$	PL NOM
<u>koir</u> a	$x_1 + a$	SG GEN
<u>koir</u> i	$x_1 + i$	PL GEN
<u>koir</u> ije	$x_1 + ije$	PL GEN
<u>koir</u> a	$x_1 + a$	SG PAR
<u>koir</u> aa	$x_1 + aa$	SG PAR
<u>koir</u> i	$x_1 + i$	PL PAR
<u>koir</u> itõ	$x_1 + it\tilde{o}$	PL PAR
<u>koir</u> aa	$x_1 + aa$	SG ILL
<u>koir</u> asõ	$x_1 + as\tilde{o}$	SG ILL
<u>koir</u> i	$x_1 + i$	PL ILL
<u>koir</u> isõ	$x_1 + is\tilde{o}$	PL ILL
<u>koir</u> õz	$x_1 + \tilde{o}z$	SG INE
<u>koir</u> iz	$x_1 + iz$	PL INE
<u>koir</u> õssõ	$x_1 + \tilde{o}ss\tilde{o}$	SG ELA
<u>koir</u> issõ	$x_1 + iss\tilde{o}$	PL ELA
<u>koir</u> õllõ	$x_1 + \tilde{o}ll\tilde{o}$	SG ALL
<u>koir</u> illõ	$x_1 + ill\tilde{o}$	PL ALL
<u>koir</u> õl	$x_1 + \tilde{o}l$	SG ADE
<u>koir</u> il	$x_1 + il$	PL ADE
<u>koir</u> õltõ	$x_1 + \tilde{o}lt\tilde{o}$	SG ABL
<u>koir</u> iltõ	$x_1 + ilt\tilde{o}$	PL ABL
<u>koir</u> õssi	$x_1 + \tilde{o}ssi$	SG TRA
<u>koir</u> issi	$x_1 + issi$	PL TRA
<u>koir</u> assaa	$x_1 + assaa$	SG TER
<u>koir</u> issaa	$x_1 + issaa$	PL TER
<u>koir</u> aka	$x_1 + aka$	SG COM
<u>koir</u> ika	$x_1 + ika$	PL COM

Tabel 8: Tüüpsõna *koirõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>lentü</u> z	$x_1 + z$	SG NOM
<u>lentü</u> sed	$x_1 + \text{sed}$	PL NOM
<u>lentü</u> se	$x_1 + \text{se}$	SG GEN
<u>lentü</u> si	$x_1 + \text{si}$	PL GEN
<u>lentü</u> sse	$x_1 + \text{sse}$	SG PAR
<u>lentü</u> ssi	$x_1 + \text{ssi}$	PL PAR
<u>lentü</u> sesse	$x_1 + \text{sesse}$	SG ILL
<u>lentü</u> sisse	$x_1 + \text{sisse}$	PL ILL
<u>lentü</u> sez	$x_1 + \text{sez}$	SG INE
<u>lentü</u> siz	$x_1 + \text{siz}$	PL INE
<u>lentü</u> sse	$x_1 + \text{sse}$	SG ELA
<u>lentü</u> sissõ	$x_1 + \text{sissõ}$	PL ELA
<u>lentü</u> göllõ	$x_1 + \text{göllõ}$	SG ALL
<u>lentü</u> koillõ	$x_1 + \text{koillõ}$	PL ALL
<u>lentü</u> göl	$x_1 + \text{göl}$	SG ADE
<u>lentü</u> koil	$x_1 + \text{koil}$	PL ADE
<u>lentü</u> göltõ	$x_1 + \text{göltõ}$	SG ABL
<u>lentü</u> koiltõ	$x_1 + \text{koiltõ}$	PL ABL
<u>lentü</u> gössi	$x_1 + \text{gõssi}$	SG TRA
<u>lentü</u> koissi	$x_1 + \text{koissi}$	PL TRA
<u>lentü</u> kassaa	$x_1 + \text{kassaa}$	SG TER
<u>lentü</u> koissaa	$x_1 + \text{koissaa}$	PL TER
<u>lentü</u> gaka	$x_1 + \text{gaka}$	SG COM
<u>lentü</u> koika	$x_1 + \text{koika}$	PL COM

Tabel 9: Tüüpsõna *lentüz* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>luzik</u> kõ	$x_1 + \text{kõ}$	SG NOM
<u>luzik</u> õd	$x_1 + \text{õd}$	PL NOM
<u>luzik</u> a	$x_1 + \text{a}$	SG GEN
<u>luzik</u> koi	$x_1 + \text{koi}$	PL GEN
<u>luzik</u> kojõ	$x_1 + \text{kojõ}$	PL GEN
<u>luzik</u> ka	$x_1 + \text{ka}$	SG PAR
<u>luzik</u> kaa	$x_1 + \text{kaa}$	SG PAR
<u>luzik</u> koi	$x_1 + \text{koi}$	PL PAR
<u>luzik</u> koitõ	$x_1 + \text{koitõ}$	PL PAR
<u>luzik</u> ka	$x_1 + \text{ka}$	SG ILL
<u>luzik</u> kasõ	$x_1 + \text{kasõ}$	SG ILL
<u>luzik</u> koisõ	$x_1 + \text{koisõ}$	PL ILL
<u>luzik</u> õz	$x_1 + \text{õz}$	SG INE
<u>luzik</u> koiz	$x_1 + \text{koiz}$	PL INE
<u>luzik</u> õssõ	$x_1 + \text{õssõ}$	SG ELA
<u>luzik</u> koissõ	$x_1 + \text{koissõ}$	PL ELA
<u>luzik</u> göllõ	$x_1 + \text{göllõ}$	SG ALL
<u>luzik</u> koillõ	$x_1 + \text{koillõ}$	PL ALL
<u>luzik</u> õl	$x_1 + \text{õl}$	SG ADE
<u>luzik</u> coil	$x_1 + \text{coil}$	PL ADE
<u>luzik</u> õltõ	$x_1 + \text{õltõ}$	SG ABL
<u>luzik</u> koiltõ	$x_1 + \text{koiltõ}$	PL ABL
<u>luzik</u> õssi	$x_1 + \text{õssi}$	SG TRA
<u>luzik</u> koissi	$x_1 + \text{koissi}$	PL TRA
<u>luzik</u> kassaa	$x_1 + \text{kassaa}$	SG TER
<u>luzik</u> koissaa	$x_1 + \text{koissaa}$	PL TER
<u>luzik</u> aka	$x_1 + \text{aka}$	SG COM
<u>luzik</u> koika	$x_1 + \text{koika}$	PL COM

Tabel 10: Tüüpsõna *luzikkõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>luik</u> ko	$x_1 + ko$	SG NOM
<u>luik</u> od	$x_1 + od$	PL NOM
<u>luik</u> o	$x_1 + o$	SG GEN
<u>luik</u> koi	$x_1 + koi$	PL GEN
<u>luik</u> kojõ	$x_1 + kojõ$	PL GEN
<u>luik</u> koa	$x_1 + koa$	SG PAR
<u>luik</u> koitõ	$x_1 + koitõ$	PL PAR
<u>luik</u> kosõ	$x_1 + kosõ$	SG ILL
<u>luik</u> koisõ	$x_1 + koisõ$	PL ILL
<u>luik</u> oz	$x_1 + oz$	SG INE
<u>luik</u> koiz	$x_1 + koiz$	PL INE
<u>luik</u> ossõ	$x_1 + ossõ$	SG ELA
<u>luik</u> koissõ	$x_1 + koissõ$	PL ELA
<u>luik</u> ollõ	$x_1 + ollõ$	SG ALL
<u>luik</u> koillõ	$x_1 + koillõ$	PL ALL
<u>luik</u> ol	$x_1 + ol$	SG ADE
<u>luik</u> coil	$x_1 + coil$	PL ADE
<u>luik</u> oltõ	$x_1 + oltõ$	SG ABL
<u>luik</u> koiltõ	$x_1 + koiltõ$	PL ABL
<u>luik</u> ossi	$x_1 + ossi$	SG TRA
<u>luik</u> koissi	$x_1 + koissi$	PL TRA
<u>luik</u> kassaa	$x_1 + kassaa$	SG TER
<u>luik</u> koissaa	$x_1 + koissaa$	PL TER
<u>luik</u> oka	$x_1 + oka$	SG COM
<u>luik</u> koika	$x_1 + koika$	PL COM

Tabel 11: Tüüpsõna *luikko* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>ai</u> tõ	$x_1 + \text{tõ}$	SG NOM
<u>ai</u> jõd	$x_1 + \text{jõd}$	PL NOM
<u>ai</u> ja	$x_1 + \text{ja}$	SG GEN
<u>ai</u> toi	$x_1 + \text{toi}$	PL GEN
<u>ai</u> tojõ	$x_1 + \text{tojõ}$	PL GEN
<u>ai</u> ta	$x_1 + \text{ta}$	SG PAR
<u>ai</u> toi	$x_1 + \text{toi}$	PL PAR
<u>ai</u> toitõ	$x_1 + \text{toitõ}$	PL PAR
<u>ai</u> taa	$x_1 + \text{taa}$	SG ILL
<u>ai</u> tasõ	$x_1 + \text{tasõ}$	SG ILL
<u>ai</u> toisõ	$x_1 + \text{toisõ}$	PL ILL
<u>ai</u> jaz	$x_1 + \text{jaz}$	SG INE
<u>ai</u> joiz	$x_1 + \text{joiz}$	PL INE
<u>ai</u> jõssõ	$x_1 + \text{jõssõ}$	SG ELA
<u>ai</u> toissõ	$x_1 + \text{toissõ}$	PL ELA
<u>ai</u> jõllõ	$x_1 + \text{jõllõ}$	SG ALL
<u>ai</u> toillõ	$x_1 + \text{toillõ}$	PL ALL
<u>ai</u> jõl	$x_1 + \text{jõl}$	SG ADE
<u>ai</u> toil	$x_1 + \text{toil}$	PL ADE
<u>ai</u> jõssi	$x_1 + \text{jõssi}$	SG ABL
<u>ai</u> toissi	$x_1 + \text{toissi}$	PL ABL
<u>ai</u> tassi	$x_1 + \text{tassi}$	SG TRA
<u>ai</u> toissi	$x_1 + \text{toissi}$	PL TRA
<u>ai</u> tassaa	$x_1 + \text{tassaa}$	SG TER
<u>ai</u> toissaa	$x_1 + \text{toissaa}$	PL TER
<u>ai</u> jaka	$x_1 + \text{jaka}$	SG COM
<u>ai</u> toika	$x_1 + \text{toika}$	PL COM

Tabel 12: Tüüpsõna *aitõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>ahkõr</u> õ	$x_1 + \tilde{o}$	SG NOM
<u>ahkõr</u> õd	$x_1 + \tilde{o}d$	PL NOM
<u>ahkõr</u> a	$x_1 + a$	SG GEN
<u>ahkõr</u> oi	$x_1 + oi$	PL GEN
<u>ahkõr</u> a	$x_1 + a$	SG PAR
<u>ahkõr</u> oi	$x_1 + oi$	PL PAR
<u>ahkõr</u> oitõ	$x_1 + oitõ$	PL PAR
<u>ahkõr</u> aa	$x_1 + aa$	SG ILL
<u>ahkõr</u> asõ	$x_1 + asõ$	SG ILL
<u>ahkõr</u> oisõ	$x_1 + oisõ$	PL ILL
<u>ahkõr</u> az	$x_1 + az$	SG INE
<u>ahkõr</u> oiz	$x_1 + oiz$	PL INE
<u>ahkõr</u> õssõ	$x_1 + \tilde{o}ssõ$	SG ELA
<u>ahkõr</u> oissõ	$x_1 + oissõ$	PL ELA
<u>ahkõr</u> õllõ	$x_1 + \tilde{o}llõ$	SG ALL
<u>ahkõr</u> oillõ	$x_1 + oillõ$	PL ALL
<u>ahkõr</u> õl	$x_1 + \tilde{o}l$	SG ADE
<u>ahkõr</u> oil	$x_1 + oil$	PL ADE
<u>ahkõr</u> õssi	$x_1 + \tilde{o}ssi$	SG ABL
<u>ahkõr</u> oissi	$x_1 + oissi$	PL ABL
<u>ahkõr</u> assi	$x_1 + assi$	SG TRA
<u>ahkõr</u> oissi	$x_1 + oissi$	PL TRA
<u>ahkõr</u> assaa	$x_1 + assaa$	SG TER
<u>ahkõr</u> oissaa	$x_1 + oissaa$	PL TER
<u>ahkõr</u> aka	$x_1 + aka$	SG COM
<u>ahkõr</u> oika	$x_1 + oika$	PL COM

Tabel 13: Tüüpsõna *ahkõrõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>omõn</u>	x_1	SG NOM
<u>omõn</u> ad	$x_1 + \text{ad}$	PL NOM
<u>omõn</u> a	$x_1 + \text{a}$	SG GEN
<u>omõn</u> oi	$x_1 + \text{oi}$	PL GEN
<u>omõn</u> ojõ	$x_1 + \text{ojõ}$	PL GEN
<u>omõn</u> a	$x_1 + \text{a}$	SG PAR
<u>omõn</u> aa	$x_1 + \text{aa}$	SG PAR
<u>omõn</u> oi	$x_1 + \text{oi}$	PL PAR
<u>omõn</u> oitõ	$x_1 + \text{oitõ}$	PL PAR
<u>omõn</u> aa	$x_1 + \text{aa}$	SG ILL
<u>omõn</u> asõ	$x_1 + \text{asõ}$	SG ILL
<u>omõn</u> oisõ	$x_1 + \text{oisõ}$	PL ILL
<u>omõn</u> õz	$x_1 + \text{õz}$	SG INE
<u>omõn</u> oiz	$x_1 + \text{oiz}$	PL INE
<u>omõn</u> õssõ	$x_1 + \text{õssõ}$	SG ELA
<u>omõn</u> oissõ	$x_1 + \text{oissõ}$	PL ELA
<u>omõn</u> õllõ	$x_1 + \text{õllõ}$	SG ALL
<u>omõn</u> oillõ	$x_1 + \text{oillõ}$	PL ALL
<u>omõn</u> õl	$x_1 + \text{õl}$	SG ADE
<u>omõn</u> oil	$x_1 + \text{oil}$	PL ADE
<u>omõn</u> õltõ	$x_1 + \text{õltõ}$	SG ABL
<u>omõn</u> oiltõ	$x_1 + \text{oiltõ}$	PL ABL
<u>omõn</u> õssi	$x_1 + \text{õssi}$	SG TRA
<u>omõn</u> oissi	$x_1 + \text{oissi}$	PL TRA
<u>omõn</u> assaa	$x_1 + \text{assaa}$	SG TER
<u>omõn</u> oissaa	$x_1 + \text{oissaa}$	PL TER
<u>omõn</u> aka	$x_1 + \text{aka}$	SG COM
<u>omõn</u> oika	$x_1 + \text{oika}$	PL COM

Tabel 14: Tüüpsõna *omõn* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>pliit</u> tõ	$x_1 + \text{tõ}$	SG NOM
<u>pliit</u> õd	$x_1 + \text{õd}$	PL NOM
<u>pliit</u> a	$x_1 + \text{a}$	SG GEN
<u>pliit</u> toi	$x_1 + \text{toi}$	PL GEN
<u>pliit</u> tojõ	$x_1 + \text{tojõ}$	PL GEN
<u>pliit</u> ta	$x_1 + \text{ta}$	SG PAR
<u>pliit</u> toi	$x_1 + \text{toi}$	PL PAR
<u>pliit</u> toitõ	$x_1 + \text{toitõ}$	PL PAR
<u>pliit</u> tasõ	$x_1 + \text{tasõ}$	SG ILL
<u>pliit</u> toisõ	$x_1 + \text{toisõ}$	PL ILL
<u>pliit</u> tõz	$x_1 + \text{tõz}$	SG INE
<u>pliit</u> toiz	$x_1 + \text{toiz}$	PL INE
<u>pliit</u> õssõ	$x_1 + \text{õssõ}$	SG ELA
<u>pliit</u> toissõ	$x_1 + \text{toissõ}$	PL ELA
<u>pliit</u> õllõ	$x_1 + \text{õllõ}$	SG ALL
<u>pliit</u> toillõ	$x_1 + \text{toillõ}$	PL ALL
<u>pliit</u> õl	$x_1 + \text{õl}$	SG ADE
<u>pliit</u> toil	$x_1 + \text{toil}$	PL ADE
<u>pliit</u> õltõ	$x_1 + \text{õltõ}$	SG ABL
<u>pliit</u> toiltõ	$x_1 + \text{toiltõ}$	PL ABL
<u>pliit</u> õssi	$x_1 + \text{õssi}$	SG TRA
<u>pliit</u> toissi	$x_1 + \text{toissi}$	PL TRA
<u>pliit</u> tassaa	$x_1 + \text{tassaa}$	SG TER
<u>pliit</u> toissaa	$x_1 + \text{toissaa}$	PL TER
<u>pliit</u> aka	$x_1 + \text{aka}$	SG COM
<u>pliit</u> toika	$x_1 + \text{toika}$	PL COM
<u>pliit</u> taa	$x_1 + \text{taa}$	SG PAR
<u>pliit</u> ta	$x_1 + \text{ta}$	SG ILL

Tabel 15: Tüüpsõna *pliitõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>mansik</u> õz	$x_1 + \tilde{o}z$	SG NOM
<u>mansik</u> kad	$x_1 + kad$	PL NOM
<u>mansik</u> ka	$x_1 + ka$	SG GEN
<u>mansik</u> kaijõ	$x_1 + kaijõ$	PL GEN
<u>mansik</u> assõ	$x_1 + assõ$	SG PAR
<u>mansik</u> kaitõ	$x_1 + kaitõ$	PL PAR
<u>mansik</u> kasõ	$x_1 + kasõ$	SG ILL
<u>mansik</u> kaisõ	$x_1 + kaisõ$	PL ILL
<u>mansik</u> kaz	$x_1 + kaz$	SG INE
<u>mansik</u> kaiz	$x_1 + kaiz$	PL INE
<u>mansik</u> kassõ	$x_1 + kassõ$	SG ELA
<u>mansik</u> kaissõ	$x_1 + kaissõ$	PL ELA
<u>mansik</u> kallõ	$x_1 + kallõ$	SG ALL
<u>mansik</u> kaillõ	$x_1 + kaillõ$	PL ALL
<u>mansik</u> kal	$x_1 + kal$	SG ADE
<u>mansik</u> kail	$x_1 + kail$	PL ADE
<u>mansik</u> kaltõ	$x_1 + kaltõ$	SG ABL
<u>mansik</u> kailtõ	$x_1 + kailtõ$	PL ABL
<u>mansik</u> kassi	$x_1 + kassi$	SG TRA
<u>mansik</u> kaissi	$x_1 + kaissi$	PL TRA
<u>mansik</u> kassaa	$x_1 + kassaa$	SG TER
<u>mansik</u> kaissaa	$x_1 + kaissaa$	PL TER
<u>mansik</u> kaka	$x_1 + kaka$	SG COM
<u>mansik</u> kaika	$x_1 + kaika$	PL COM

Tabel 16: Tüüpsõna *mansikõz* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>kat</u> t i	$x_1 + t + x_2$	SG NOM
<u>kat</u> i d	$x_1 + x_2 + d$	PL NOM
<u>kat</u> i	$x_1 + x_2$	SG GEN
<u>kat</u> t i jō	$x_1 + t + x_2 + jō$	PL GEN
<u>kat</u> t i a	$x_1 + t + x_2 + a$	SG PAR
<u>kat</u> t i i	$x_1 + t + x_2 + i$	PL PAR
<u>kat</u> t i itō	$x_1 + t + x_2 + itō$	PL PAR
<u>kat</u> t i se	$x_1 + t + x_2 + se$	SG ILL
<u>kat</u> t i ise	$x_1 + t + x_2 + ise$	PL ILL
<u>kat</u> t i z	$x_1 + t + x_2 + z$	SG INE
<u>kat</u> t i iz	$x_1 + t + x_2 + iz$	PL INE
<u>kat</u> i sse	$x_1 + x_2 + sse$	SG ELA
<u>kat</u> t i isse	$x_1 + t + x_2 + isse$	PL ELA
<u>kat</u> i lle	$x_1 + x_2 + lle$	SG ALL
<u>kat</u> t i ille	$x_1 + t + x_2 + ille$	PL ALL
<u>kat</u> i l	$x_1 + x_2 + l$	SG ADE
<u>kat</u> t i il	$x_1 + t + x_2 + il$	PL ADE
<u>kat</u> i lte	$x_1 + x_2 + lte$	SG ABL
<u>kat</u> t i ilte	$x_1 + t + x_2 + ilte$	PL ABL
<u>kat</u> i ssi	$x_1 + x_2 + ssi$	SG TRA
<u>kat</u> t i issi	$x_1 + t + x_2 + issi$	PL TRA
<u>kat</u> t i ssaa	$x_1 + t + x_2 + ssaa$	SG TER
<u>kat</u> t i issaa	$x_1 + t + x_2 + issaa$	PL TER
<u>kat</u> i ka	$x_1 + x_2 + ka$	SG COM
<u>kat</u> t i ika	$x_1 + t + x_2 + ika$	PL COM

Tabel 17: Tüüpsõna *katti* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>am</u> mõz	$x_1 + mõz$	SG NOM
<u>am</u> pad	$x_1 + pad$	PL NOM
<u>am</u> pa	$x_1 + pa$	SG GEN
<u>am</u> paijõ	$x_1 + paijõ$	PL GEN
<u>am</u> massõ	$x_1 + massõ$	SG PAR
<u>am</u> paitõ	$x_1 + paitõ$	PL PAR
<u>am</u> pasõ	$x_1 + pasõ$	SG ILL
<u>am</u> paisõ	$x_1 + paisõ$	PL ILL
<u>am</u> paz	$x_1 + paz$	SG INE
<u>am</u> paiz	$x_1 + paiz$	PL INE
<u>am</u> passõ	$x_1 + passõ$	SG ELA
<u>am</u> paissõ	$x_1 + paissõ$	PL ELA
<u>am</u> pallõ	$x_1 + pallõ$	SG ALL
<u>am</u> paillõ	$x_1 + paillõ$	PL ALL
<u>am</u> pal	$x_1 + pal$	SG ADE
<u>am</u> pail	$x_1 + pail$	PL ADE
<u>am</u> paltõ	$x_1 + paltõ$	SG ABL
<u>am</u> pailtõ	$x_1 + pailtõ$	PL ABL
<u>am</u> passi	$x_1 + passi$	SG TRA
<u>am</u> paissi	$x_1 + paissi$	PL TRA
<u>am</u> passaa	$x_1 + passaa$	SG TER
<u>am</u> paissaa	$x_1 + paissaa$	PL TER
<u>am</u> paka	$x_1 + paka$	SG COM
<u>am</u> paika	$x_1 + paika$	PL COM

Tabel 18: Tüüpsõna *ammõz* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>kõik</u> k	$x_1 + k$	SG NOM
<u>kõik</u> õd	$x_1 + \tilde{o}d$	PL NOM
<u>kõik</u> õ	$x_1 + \tilde{o}$	SG GEN
<u>kõik</u> ki	$x_1 + ki$	PL GEN
<u>kõik</u> kijõ	$x_1 + kij\tilde{o}$	PL GEN
<u>kõik</u> ka	$x_1 + ka$	SG PAR
<u>kõik</u> kaa	$x_1 + kaa$	SG PAR
<u>kõik</u> kia	$x_1 + kia$	PL PAR
<u>kõik</u> kitõ	$x_1 + kit\tilde{o}$	PL PAR
<u>kõik</u> kaa	$x_1 + kaa$	SG ILL
<u>kõik</u> kasõ	$x_1 + kas\tilde{o}$	SG ILL
<u>kõik</u> ki	$x_1 + ki$	PL ILL
<u>kõik</u> kisõ	$x_1 + kis\tilde{o}$	PL ILL
<u>kõik</u> õz	$x_1 + \tilde{o}z$	SG INE
<u>kõik</u> kiz	$x_1 + kiz$	PL INE
<u>kõik</u> õssõ	$x_1 + \tilde{o}ss\tilde{o}$	SG ELA
<u>kõik</u> kissõ	$x_1 + kiss\tilde{o}$	PL ELA
<u>kõik</u> õllõ	$x_1 + \tilde{o}ll\tilde{o}$	SG ALL
<u>kõik</u> killõ	$x_1 + kill\tilde{o}$	PL ALL
<u>kõik</u> õl	$x_1 + \tilde{o}l$	SG ADE
<u>kõik</u> kil	$x_1 + kil$	PL ADE
<u>kõik</u> õltõ	$x_1 + \tilde{o}lt\tilde{o}$	SG ABL
<u>kõik</u> kiltõ	$x_1 + kilt\tilde{o}$	PL ABL
<u>kõik</u> õssi	$x_1 + \tilde{o}ssi$	SG TRA
<u>kõik</u> kissi	$x_1 + kissi$	PL TRA
<u>kõik</u> kassaa	$x_1 + kassaa$	SG TER
<u>kõik</u> kissaa	$x_1 + kissaa$	PL TER
<u>kõik</u> aka	$x_1 + aka$	SG COM
<u>kõik</u> kika	$x_1 + kika$	PL COM

Tabel 19: Tüüpsõna *kõikk* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>kuk</u> kõ	$x_1 + \text{kõ}$	SG NOM
<u>kuk</u> õd	$x_1 + \text{õd}$	PL NOM
<u>kuk</u> a	$x_1 + \text{a}$	SG GEN
<u>kuk</u> ki	$x_1 + \text{ki}$	PL GEN
<u>kuk</u> kije	$x_1 + \text{kije}$	PL GEN
<u>kuk</u> ka	$x_1 + \text{ka}$	SG PAR
<u>kuk</u> kaa	$x_1 + \text{kaa}$	SG PAR
<u>kuk</u> ki	$x_1 + \text{ki}$	PL PAR
<u>kuk</u> kitõ	$x_1 + \text{kitõ}$	PL PAR
<u>kuk</u> kaa	$x_1 + \text{kaa}$	SG ILL
<u>kuk</u> kasõ	$x_1 + \text{kasõ}$	SG ILL
<u>kuk</u> ki	$x_1 + \text{ki}$	PL ILL
<u>kuk</u> kisõ	$x_1 + \text{kisõ}$	PL ILL
<u>kuk</u> õz	$x_1 + \text{õz}$	SG INE
<u>kuk</u> kiz	$x_1 + \text{kiz}$	PL INE
<u>kuk</u> õssõ	$x_1 + \text{õssõ}$	SG ELA
<u>kuk</u> kissõ	$x_1 + \text{kissõ}$	PL ELA
<u>kuk</u> öllõ	$x_1 + \text{öllõ}$	SG ALL
<u>kuk</u> killõ	$x_1 + \text{killõ}$	PL ALL
<u>kuk</u> õl	$x_1 + \text{õl}$	SG ADE
<u>kuk</u> kil	$x_1 + \text{kil}$	PL ADE
<u>kuk</u> õltõ	$x_1 + \text{õltõ}$	SG ABL
<u>kuk</u> kiltõ	$x_1 + \text{kiltõ}$	PL ABL
<u>kuk</u> õssi	$x_1 + \text{õssi}$	SG TRA
<u>kuk</u> kissi	$x_1 + \text{kissi}$	PL TRA
<u>kuk</u> kassaa	$x_1 + \text{kassaa}$	SG TER
<u>kuk</u> kissaa	$x_1 + \text{kissaa}$	PL TER
<u>kuk</u> aka	$x_1 + \text{aka}$	SG COM
<u>kuk</u> kika	$x_1 + \text{kika}$	PL COM

Tabel 20: Tüüpsõna *kukkõ* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>päiv</u> e	$x_1 + e$	SG NOM
<u>päiv</u> äd	$x_1 + äd$	PL NOM
<u>päiv</u> ä	$x_1 + ä$	SG GEN
<u>päiv</u> i	$x_1 + i$	PL GEN
<u>päiv</u> ije	$x_1 + ije$	PL GEN
<u>päiv</u> ä	$x_1 + ä$	SG PAR
<u>päiv</u> ää	$x_1 + ää$	SG PAR
<u>päiv</u> i	$x_1 + i$	PL PAR
<u>päiv</u> ii	$x_1 + ii$	PL PAR
<u>päiv</u> ää	$x_1 + ää$	SG ILL
<u>päiv</u> äse	$x_1 + äse$	SG ILL
<u>päiv</u> i	$x_1 + i$	PL ILL
<u>päiv</u> ise	$x_1 + ise$	PL ILL
<u>päiv</u> äz	$x_1 + äz$	SG INE
<u>päiv</u> iz	$x_1 + iz$	PL INE
<u>päiv</u> ässä	$x_1 + ässä$	SG ELA
<u>päiv</u> issä	$x_1 + issä$	PL ELA
<u>päiv</u> ällä	$x_1 + ällä$	SG ALL
<u>päiv</u> ille	$x_1 + ille$	PL ALL
<u>päiv</u> äl	$x_1 + äl$	SG ADE
<u>päiv</u> il	$x_1 + il$	PL ADE
<u>päiv</u> älte	$x_1 + älte$	SG ABL
<u>päiv</u> ilte	$x_1 + ilte$	PL ABL
<u>päiv</u> ässi	$x_1 + ässi$	SG TRA
<u>päiv</u> issi	$x_1 + issi$	PL TRA
<u>päiv</u> ässaa	$x_1 + ässaa$	SG TER
<u>päiv</u> issaa	$x_1 + issaa$	PL TER
<u>päiv</u> äka	$x_1 + äka$	SG COM
<u>päiv</u> ika	$x_1 + ika$	PL COM

Tabel 21: Tüüpsõna *päive* ekstraheeritud muutvormimall.

ühisosajada	muutvormimall	tunnused
<u>par</u> tõ	$x_1 + t\ddot{o}$	SG NOM
<u>par</u> rōd	$x_1 + r\ddot{o}d$	PL NOM
<u>par</u> ra	$x_1 + ra$	SG GEN
<u>par</u> toi	$x_1 + toi$	PL GEN
<u>par</u> tojõ	$x_1 + toj\ddot{o}$	PL GEN
<u>par</u> ta	$x_1 + ta$	SG PAR
<u>par</u> toi	$x_1 + toi$	PL PAR
<u>par</u> toitõ	$x_1 + toit\ddot{o}$	PL PAR
<u>par</u> tasõ	$x_1 + tas\ddot{o}$	SG ILL
<u>par</u> toisõ	$x_1 + tois\ddot{o}$	PL ILL
<u>par</u> rōz	$x_1 + r\ddot{o}z$	SG INE
<u>par</u> toiz	$x_1 + toiz$	PL INE
<u>par</u> rōssõ	$x_1 + r\ddot{o}ss\ddot{o}$	SG ELA
<u>par</u> toissõ	$x_1 + toiss\ddot{o}$	PL ELA
<u>par</u> rōllõ	$x_1 + r\ddot{o}ll\ddot{o}$	SG ALL
<u>par</u> toillõ	$x_1 + toill\ddot{o}$	PL ALL
<u>par</u> rōl	$x_1 + r\ddot{o}l$	SG ADE
<u>par</u> toil	$x_1 + toil$	PL ADE
<u>par</u> rōltõ	$x_1 + r\ddot{o}lt\ddot{o}$	SG ABL
<u>par</u> toiltõ	$x_1 + toilt\ddot{o}$	PL ABL
<u>par</u> rōssi	$x_1 + r\ddot{o}ssi$	SG TRA
<u>par</u> toissi	$x_1 + toissi$	PL TRA
<u>par</u> tassaa	$x_1 + tassaa$	SG TER
<u>par</u> toissaa	$x_1 + toissaa$	PL TER
<u>par</u> raka	$x_1 + raka$	SG COM
<u>par</u> toika	$x_1 + toika$	PL COM
<u>par</u> taa	$x_1 + taa$	SG PAR
<u>par</u> ta	$x_1 + ta$	SG ILL

Tabel 22: Tüüpsõna *partõ* ekstraheeritud muutvormimall.

```
<LexicalEntry morphologicalPatterns="asKatto">
  <feat att="partOfSpeech" val="nn"/>
  <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 kajastatud vaid osaliselt).

```

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

```

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