

DISSERTATIONES PHILOLOGIAE ESTONICAE
UNIVERSITATIS TARTUENSIS

1



**EESTI KEELE
KLASSIFIKATOORNE
MORFOLOOGIA**

Ülle Viks

TARTU 1994

DISSERTATIONES PHILOLOGIAE ESTONICAE
UNIVERSITATIS TARTUENSIS

1



**EESTI KEELE
KLASSIFIKATOORNE
MORFOLOOGIA**

Ülle Viks

TARTU 1994

Väitekiri on valminud Eesti Teaduste Akadeemia Eesti
Keele instituudis

Dissertant: Ülle Viks
Eesti Keele Instituut
Roosikrantsi 6
EE0100 Tallinn
email: ylle @ kki.ee

Oponendid: Paul Alvre, DrPhil, Tartu
Tiit-Rein Viitso, DrPhil, Tartu

Kaitsmine toimub 14. juunil 1994 kell 11 Tartu Ülikooli
nõukogu saalis, Ülikooli 18, EE2400 Tartu

Nõukogu sekretär

/Mare Sepp/

Tartu Ülikooli Kirjastuse trükikoda
Tiigi 78, EE2400 Tartu
Tellimus nr. 191. Trükiarv 120.

© Ülle Viks, 1994

Sisukord

Sissejuhatus	5
Morfoloogiasüsteem	7
Morfoloogilised üksused	8
Üksuste variantsus	10
Morfoloogiline klassifikatsioon	13
Tüübituvastus	16
Sõnavormide süntees ja analüüs	17
Grammatika ja sõnastiku integratsioon	19
Arvutirakendused	24
Publikatsioonid	26
Classificatory Morphology of Estonian. Abstract	29

Sissejuhatus

Käesolev väitekiri koosneb väikemonograafiate sarjast, milles on esitatud klassifikatoorse morfoloogiamudeli piirjooned ja eesti sõnamuutmise kirjeldus neist põhimõtetest lähtudes. Väitekirja lisana on esitatud morfoloogiamudeli realisatsioon sõnastikuvormis. Väitekirja põhiosad on:

- I. Klassifikatoorse morfoloogia üldiseloostus:
Klassifikatoorse morfoloogia põhimõtted. [5]
- II. Verbimorfoloogia:
Klassifikatoorse morfoloogia. Verb. [8]
- III. Noomenimorfoloogia:
Klassifikatoorse morfoloogia. Noomen. [9]
- IV. Grammatika ja sõnastiku integratsioon:
Väike vormisõnastik I. Sissejuhatus & grammatika. [16]
- Lisa. Morfoloogiasõnastik:
Väike vormisõnastik II. Sõnastik & lisad. [17]

Eesti keel kuulub morfoloogiakeelte hulka, kus grammatilisi tähendusi väljendatakse eelkõige morfoloogiliste vahenditega, sõna erinevate muutevormide kaudu. Eesti keelt iseloomustab muutevormide suur hulk: noomenil 14-15 käänat ainsuses ja mitmuses koos rohkete paralleelvormidega, verbil ligi 60 lihtvormi, pluss analüütilised aja- ja eitusvormid.

Teine eesti keele oluline eripära on morfoloogiliste üksuste suur variantsus. Varieeruda võivad nii morfoloogilised formaatiivid kui ka tüved, kusjuures tüves võib põimuda mitu variantsuse liiki. Tüvi võib vahel jääda muutumatuks, kuid enamasti leiavad seal aset kas astmemuutused või lõpumuutused või mõlemad korraga. Tüvevariandid ei ole seejuures seotud alati samade muutevormidega, vaid võivad paradigmas paikneda erinevate vaheldusmallide järgi. Variantide paljusest tingitult tõuseb keelekirjelduse üheks olulisemaks osaks üksusevariantide kombinatoorika kirjeldamine.

Käesolev töö esitab eesti morfoloogia eripära arvestava klassifikatoorse morfoloogia mudeli ning selle alusel eesti keele morfoloogilise süsteemi kirjelduse nii grammatika kui ka sõnastiku kujul. Sõnastik kuulub lahutamatult morfoloogiakirjelduse juurde, sest grammatika ja sõnastiku meetodid ja võimalused on erinevad. Grammatika kirjeldab morfoloogiasüsteemi nähtuste ja seoste kaupa, mida väljendatakse reeglite või tabelite abil. Sõnastik kirjeldab sama süsteemi üksiksõnade kaupa, esitades sõnaartiklis kõik andmed ühe sõna kohta. Grammatika taotleb ülevaatlikkust, kusjuures üksiksõnad on grammatikas esindatud ainult näidete tasandil. Sõnastik taotleb ammendavust, kusjuures reeglitel pole sõnaartiklis kohta. Ainult sõnastiku ja grammatika integreerimine üheks tervikuks annab võimaluse kirjeldada morfoloogiasüsteemi niihästi ülevaatlikult kui ka ammendavalt.

Klassifikatoorne morfoloogiamudel kirjeldab nii morfoloogiasüsteemi struktuuri (s.t üksusi ja nendevahelisi seoseid) kui ka funktsioneerimist (s.t sõnavormide analüüsi ja sünteesi). Mudel keskendub eelkõige erinevuste väljatoomisele sõnade morfoloogilises käitumises, kasutades meetodina sõnade klassifitseerimist morfoloogiliste tunnuste alusel.

Meetodi valik on põhjendatud ühelt poolt eesti grammatikate ja sõnastike traditsiooniga, kus on ikka tüüpsõnu kirjelduse abivahendina kasutatud. Teiselt poolt (ja see asjaolu ongi traditsiooni toetanud) on eesti morfoloogias põimunud nii mitmesugused keelelised protsessid, et ilma klassifitseerimiseta on selles raske orienteeruda.

Klassifikatsiooni liigitusaluste valiku põhikriteeriumiks on morfoloogiliste tunnuste seosed lingvistiliste reeglitega. See tagab klassifikatsiooni sobivuse morfoloogiliste protsesside (analüüsi ja sünteesi) kirjeldamise abivahendina.

Siinne kokkuvõte on struktureeritud klassifikatoorse morfoloogiamudeli sisemist loogikat järgides, mitte aga väitekirja liigenduse alusel, mille on tinginud materjali käsitlemise loogika (eraldi osadena verb ja noomen, eraldi sõnastik). Seosed kahe loogika vahel luuakse viidete abil.

Morfoloogiasüsteem

[5:3-22, 30-43; 8:18-22; 9:4-15, 16:22-35; 4; 10]

Morfoloogia on keelesüsteemi osa, mille objektiks on sõnavorm — see on väikseim fonoloogiline järjend, mis ühendab endas leksikaalse ja grammatilise tähenduse. Morfoloogilised protsessid on kahesuunalised. Morfoloogiline analüüs hõlmab protsesse, mis fonoloogilisest järjendist lähtudes tuvastavad sõnavormi grammatilised tähendused ja algvormi (kui leksikaalse tähenduse kandja). Morfoloogiline süntees hõlmab protsesse, mis algvorme ja grammatilisi tähendusi sidudes moodustavad foneemijärjendena realiseeruvaid sõnavorme.

Klassifikatoorne morfoloogia on paradigmakeskne, mis tähendab, et tema põhiobjektiks pole mitte üksik sõnavorm, vaid morfoloogiline sõna. See on abstraktne üksus, millesse on koondunud kõik ühe algvormiga seotud paradigmaatilised sõnavormid (muutetvormid). Algvorm ehk lemma on sõnavorm, mis esindab morfoloogilist sõna (edaspidi *sõna*) nii morfoloogilise analüüsi ja sünteesi reeglites kui ka sõnastikus.

Sõltuvalt sellest, millised grammatilised kategooriad milliste sõnadega seostuvad, jaguneb eesti sõnavara kolme morfoloogilise klassi: 1) noomenid, mis võivad muutuda käändes ja arvus; 2) verbid, mis võivad muutuda kõneviisis, ajas, tegumoes, isikus ja kõnelaadis; 3) muutumatud sõnad.

Grammatiliste kategooriate elementide võimalikud kombinatsioonid määravad morfoloogilise paradigma koosseisu ja sisemise struktuuri. Igale tähenduste kombinatsioonile (nt 'mitmus' + 'omastav') vastab paradigma liige ehk muutetvorm. Muutetvormid paiknevad paradigmas teatud kindla korra järgi. Igal muutetvormil on oma paradigmaatiline positsioon ja sellele vastav grammatiline tähendus, mis moodustub kategooria elementide grammatilistest tähendustest.

Sõnavormide poolt vaadates kujutab sõna morfoloogiline paradigma endast ühe sõna kõigi muutetvormide korrastatud hulka. Semantiliselt püsiv element paradigmas on sõnatüvi, mis kannab

sõna leksikaalset tähendust, semantiliste muutujate rollis esinevad morfoloogilised formatiivid, mis kannavad sõnavormide grammatilisi tähendusi.

Paradigma maht ja koosseis ei ole jäigalt fikseeritud. Võib eristada täisparadigmat, millesse põhimõtteliselt kuuluvad sõna kõik muutevormid, ja alaparadigmat, mis hõlmab mingit kindlat osa täisparadigmast, nt tingiva kõneviisi paradigma, minevikuparadigma, finiiitne paradigma, mitmuse paradigma jne.

Kui sõna vormistikust puudub mõnele paradigmaatilisele positsioonile vastav muutevorm, on tegemist defektse paradigmaga, nt noomenil *mõlema* puudub ainsuse nimetav, verbil *pidama* 'kohustatud olema' puuduvad umbisikulise tegumoe vormid jne.

Kui ühele paradigmaatilisele positsioonile vastab mitu sõnavormi, on tegemist rööp- ehk paralleelvormidega. Need on erikujulised, kuid ühe ja sama tähendusega vormid, nt *s'appa/sid* & *s'appi* 'pl p' S'EPP. Paralleelvorme esineb rohkesti noomeni paradigmat, kus mitmuses on enamasti paralleelselt kasutusel *te*-mitmuse ja vokaalmitmuse vormisarjad, nt *hammas/tesse* & *h'amba/isse*, *hammate/s* & *h'amba/is* jne. Kui regulaarsete sarjadena esinevad paralleelvormid vastavad mõne teise muuttüübi tingimustele, võib rääkida paralleelparadigmadest, s.t sõna muutub kahe erineva muuttüübi järgi, nt *s'ulge/ma* : *sule/b* (nagu *'õppi/ma* : *õpi/b*) ja *s'ulge/ma* : *s'ulge/b* (nagu *ela/ma* : *ela/b*).

Morfoloogilised üksused

[5:30-33; 16:35-36; 18:153-157]

Paradigmakeskse lähenemise seisukohalt koosneb iga muutevorm kahest osast: sõnatüvest ja morfoloogilisest formatiivist. See on minimaalne morfoloogiline liigendus (vormiliigendus), mille puhul tüvi kannab sõnavormi leksikaalset tähendust ja morfoloogiline formatiiv kannab kogu grammatiliste tähenduste kompleksi: *hammas/tega* → HAMMAS + 'pl kom', *h'amba/ga* → HAMMAS + 'sg kom'. Kui sõnavormi grammatilisel tähendusel puudub tüvele liituv fonoloogiline realisatsioon, räägitakse nullformatiivist, nt *h'amba(/0)* → HAMMAS + 'sg g'.

Detailsem morfoloogiline liigendus jagab formatiivi üksikuteks morfeemideks, mis vastavad üksikutele grammatilistele tähendustele: *hammas/te/ga* → HAMMAS + 'pl' + 'kom', *h'amba/ga* → HAMMAS + 'kom'.

Üksuste kombinatoorikat (ehk sõnavormi sisestruktuuri) kirjeldavad morfoloogilised reeglid. Need määravad, millised üksused, millises järjestuses ja mis tingimustel võivad ühes sõnavormis koos esineda.

Eesti keeles on näiteks noomenivormide morfoloogiline üldstruktuur järgmine: TÜVI + ARV + KÄÄNE. Et kõik muud üksused peale tüve võivad teatud juhtudel ka ilma fonoloogilise väljenduseta olla, siis on reaalseid morfoloogilisi kombinatsioone neli: TÜVI + ARV + KÄÄNE (*'aasta/te/ga*), TÜVI + KÄÄNE (*'aasta/ga*), TÜVI + ARV (*'aasta/te*), TÜVI (*'aasta*). Verbi morfoloogiline struktuur on mõnevõrra keerulisem ja hõlmab rohkem grammatilisi kategooriaid. Kombinatsioonide arv kahekordistub, kui arvestada ka partiklit *gi/ki*, mis võib liituda kõigile sõnavormidele, v.a side- ja hüüdsõnad ning eitussõna.

Sõnavormi liigendus võib minna ka sügavamale. Tuletusliigenduse puhul kirjeldavad morfoloogilised reeglid lisaks sõnavormi puhtmorfoloogilisele struktuurile ka tüve tuietusstruktuuri: millised derivatsioonisufiksud ja millises järjestuses võivad juurele liituda.

Liitsõnade morfoloogiline struktuur on eesti keeles üsna keeruline ja suhteliselt vähe uuritud ala. Liitsõna jaguneb leksikaalseid tähendusi kandvateks komponentideks: üheks või mitmeks täiendsõnaks ja üheks põhisõnaks, vrd nt *k'aitse+vägi*, *õhu+k'aitse+vägi* ja *õhu+k'aitse+s'uur+tüki+vägi*.

Morfoloogiliselt muutub üldjuhul ainult põhisõna, kuid üksikutel sõnadel võib muutuda ka täiendosa, nt *k'uus+sada* : *kuue+saja/te*, *'emb+k'umb* : *emma/st+kumma/st*. Täiendsõna on enamasti kujul, mis ei sisalda formatiive (*r'aud+t'õe*, *raua+m'aak*, *lüh+rom'aan*), aga ta võib ka sisaldada nii tuletusliiteid kui morfoloogilisi formatiive (*l'aul |mis+t'und*, *vali |ja+m'ees*, *ve|tel+pääste*).

Siinne käsitlus keskendub morfoloogiale ja piirdub sõnavormi kahendliigendusega tüveks ja formatiiviks.

Kuigi eesti keelt loetakse valdavalt aglutinatiivseks keeleks, on siin küllalt palju fusiivseid momente ning tüve ja formatiivi piir pole alati selgelt eristatav, näiteks võivad tüve lõpuvokaal ja formatiivi algusvokaal kokku sulada. Tüüpiline on see vokaalmitmuse puhul (*j'alga + u* → *j'algu* või *j'algu*∅, *jala + ul* → *jal/ul* või *jalul*), aga esineb ka näiteks verbi minevikuvormides (*tule + in* → *tu/in* või *tuli/n*).

Mõnikord ei võimalda ka ortograafia tegelikke morfoloogilisi piire kajastada — seda tingib geminaadi kirjutusviis ühe tähega, kuigi morfeemi piir võib minna kahe kokkusulanud foneemi vahelt (*kät/tega* ja *l'eid/tud*).

Kõigil sellistel juhtudel on piiri märkimine tinglik. Siinses käsitluses on eelistatud formatiivi terviklikkuse säilitamist, seega: *j'algu*, *jal/ul*, *tu/in*, *kät/tega*, *l'eit/ud*.

Üksuste variantsus

[5:33-38; 8:55-58; 9:25-39, 50-60; 18:157-159; 7]

Eesti keelele on iseloomulik, et morfoloogilised üksused võivad sõnavormides realiseeruda erinevate variantidena, millel on üks ja sama (leksikaalne või grammatiline) tähendus, kuid erinev fonoloogiline kuju. Varieeruda võivad sõnavormi mõlemad osad, nii tüved kui ka formatiivid. Tüve variantsus avaldub paradigma sees, s.t sama sõna erinevate muutevormide võrdlemisel (*kalle* : *k'alde*). Formatiivi variantsus avaldub paradigmade vahel, s.t ühe ja sama muutevormi võrdlemisel erinevate sõnade puhul (*kalla/de* - *kal/le/te*).

Tüvede varieerumisel on kaks aspekti. Üks iseloomustab variantide omavahelist fonoloogilist seost — tüvem muutust, mis on kirjeldatav teisendusreeglite abil. Tüvem muutused pole otseselt seotud konkreetsete muutevormidega. Samad reeglid toimivad nii noomeni- kui verbimorfoloogias ja ka derivatsiooni puhul. Teine aspekt iseloomustab variantide seost paradigmaatiliste positsioonidega, mis on kirjeldatav vaheldusmallide abil. Vaheldusmallid näitavad, millistes muutevormides ühe sõna erinevaid tüvevariante kasutatakse.

Teisendusreeglid. Eesti keeie tüvemuutuste süsteem on küllalt keeruline ja selle adekvaatseks kirjeldamiseks on vaja mitut liiki teisendusreegleid. Põhilised tüvemuutused on:

a) tüve **astmemuutus**, mille puhul tüvi esineb kas tugevas või nõrgas astmes; astmeid eristab kõigepealt vältemuutus, millega võivad kaasneda mitmesugused häälikuteisendused, nagu klusiilide geminatsioon, assimilatsioon, kadu jne, nt *t'elli/ma - telli/b* (3-2), *p'aika/ma - paiga/ta* (3-2, k-g), *kr'unti/ma - krundi/b* (3-2, t-d), *vanne - v'ande* (2-3, nn-nd), *puue - p'uude* (2-3, 0-d).

b) tüve **lõpumuutused**, mille puhul tüvi esineb kas lemmatüvena või muutetüvena ja mis hõlmab selliseid häälikuteisendusi, nagu lõpukadu või lisandus, häälikuvahetus, *ne/se*-vaheldus jne, nt *v'eok - v'eoki* (0-i), *m'under - m'undri* (er-ri), *soolane - soolase* (ne-se), *sipelgas - sipelga* (s-0), *paigas - paigase* (0-e).

Tüve astme- ja lõpumuutused võivad esineda eraldi, nagu eeltoodud näidetes, kuid sagedamini avalduvad nad koos ühes sõnas, nt *p'aik - paiga - p'aika* (3-2-3, k-g-k, 0-a), *kaigas - k'aika* (2-3, g-k, s-0), *pundar - p'untra* (2-3, d-t, ar-ra), *kannel - k'andle* (2-3, nn-nd, el-le), *suue - s'uudme* (2-3, 0-d, e-me).

Astme- ja lõpumuutusreeglitele võivad lisanduda veel (c) morfonoloogilise **distributsiooni** reeglid, mis kirjeldavad selliseid muutusi tüves, mille esilekutsujaks on konkreetne tüvele liituv formaativ oma fonoloogiliste omadustega. Näiteid:

- tüve pikk vokaal lüheneb vokaali ees: *id'ee + id* → *id'e/id*,
- pikk madal vokaal kõrgeneb a ja e ees: *joo + akse* → *juu/akse*,
- tüvevokaal sulab kokku mitmusevokaaliga: *j'alga + V* → *j'algu*.

Analoogiliselt on kirjeldatavad ka mõned formaatiivide varieerumised, mis sõltuvad tüve lõpphääliku(te)st, nt

- tugev klusiil nõrgeneb helilise konsonandi järel: *hüpel + ta* → *hüpel/da*,
- konsonantide vahele tekib siirdevokaal: *n'aer + v* → *n'aer/ev*.

Vaheldusmallid. Tüvede vaheldusmallid, mis iseloomustavad tüvevariantide kasutamist erinevates muutevormides, on vastavalt tüve muutuste liigile samuti kaheksaüks:

a) **astmevaheldusmall** iseloomustab tugeva (T) ja nõrga (N) astme positsioone paradigmas, vrd

h'inda/ma - hinna/ta - h'inda/b - hinna/tud (TNTN)

s'undi/ma - s'undi/da - sunni/b - sunni/tud (TTNN)

huige - h'uike - huige/t - h'uike/d - h'uike/id (NTNTT)

l'uik - luige - l'uike - luige/d - l'uike/sid (TNTNT) (ABBBB)

b) **lõpuvaheldusmall** iseloomustab lemmatüve (A) ja muute-tüvede (B,C) positsioone paradigmas, vrd

r'audne - r'audse - r'audse/t - r'audse/te (ABBB)

ase - ase/me - ase/t - ase/me/te (ABAB)

soolane - soolase - soolas/t - soolas/te (ABCC)

hammas - h'amba - hammas/t - hammas/te (ABAA) (NTNN)

l'uik - luige - l'uike - l'uike/de (ABBB) (TNTT)

Eesti keele tüvevaheldusmallid on keeleajalooliselt olnud suu-rel määral morfonoloogiliselt tingitud. Astmevaheldust mõjutas sõnavormi silbistruktuur: kahe silbilistes tüvedes põhjustas kinnine teine silp nõrga astme tekke, lahtise silbi korral kujunes tugev aste. Lõpuvaheldus oli rohkem seotud kindlate vormidega: kahetüvelistel sõnadel oli teatud vormides kasutusel vokaaltüvi, teatud vormides konsonanttüvi. Tänapäevaks on keeles toimunud palju ajaloolisi häälikumuutusi (lõpu- ja sisekaod jne), mille tagajärjel tüvevaheldusi tinginud tegurid võivad olla kadunud, kuid vaheldusmallid ise on säilinud. Seetõttu on tüvevaheldusmallid saanud suurel määral sõna individuaalseks iseärasuseks.

Samal kombel on ka formatiivide variandid algselt olnud morfonoloogiliselt tingitud, sõltudes formatiivi positsioonist sõnavormi silbistruktuuris, silbi rõhustaatusest ja eelnevatest häälikutest. Keeleajaloolised arengud on sellegi sõltuvuse paljudel juhtudel kaotanud ja jätnud formatiivvariantide valiku sõna individuaalseks iseärasuseks, vrd *norse/te, morse/de*.

Morfoloogiline klassifikatsioon

[5:23-38; 8:5-17, 23-37; 9:16-46; 16:37-54; 17:478-797;
18:159-161; 2; 3; 6; 10]

Allotaktika. Et varieeruda võivad mõlemad morfoloogilised põhiüksused, nii tüved kui ka formatiivid, siis ei piisa sõnavormide analüüsi ja sünteesi kirjeldamisel morfotaktika reeglitest, mis töötavad üksuste tasemel. Üksused võivad küll anda korrektse morfotaktilise struktuuri, kuid üksuste variandid võivad olla omavahel sobimatud. Näiteks struktuur TŪVI + 'pl g' on igati normaalne. Sõnal LIIGE on kaks tüvevarianti: *liige* ja *l'ikme*, mitmuse omastava formatiivil on variandid: *te* ja *de*. Seega on kokku neli erinevat võimalust variante kombineerida. Korrektne vorm selle sõna jaoks on aga ainult üks — *l'ikme/te*, ja mitte **l'ikme/de* (vrd *k'like/de*), **liige/te* (vrd *hõige/te*), **liige/de* (vrd *kolge/de*).

Õige valiku tegemiseks on vaja kontrollida variantide omavahelist sobivust. Reegleid, mis kirjeldavad üksusevariantide (allomorfade) valiku ja koosinemise tingimusi sõnavormis, nimetatakse **allotaktika reegliteks** (morfotaktika — morfeemide kombinatoorika, allotaktika — allomorfade kombinatoorika).

Klassifikatsioon. Kõige ökonoomsem viis allotaktika reeglite esitamiseks on morfoloogiline klassifikatsioon. Klassifikatsioon võib olla detailsem või jämedam, sõltuvalt sellest, kui paljusid tunnuseid liigitusalustena arvestada. Nii on nt Muugi sõnastikus 845 tüüpi, Õigekeelsussõnaraamatus 115, ja Väikeses vormisõnastikus 38 tüüpi. ÕS ja eriti Muuk arvestavad sõnade liigitamisel ka tüvemuu- tusi — sellest tuleneb ka tüüpide suur arv.

Klassifikatoorne morfoloogia tüvemuu- tusi liigitusalusena ei arvesta, sest need moodustavad eraldi allsüsteemi: grammatika tase- mel kirjeldavad tüvevariantide omavahelisi seoseid teisendusreeg- lid (astme- ja lõpumuutusreeglid), sõnastiku tasemel on kõik tüve- variandid iga märksõna juures antud.

Klassifikatsioon toob esile ainult erinevad allotaktilised kombinatsioonid ja liigitab kogu sõnade hulga kolme tunnuse järgi:

- 1) tüve astmevaheldusmall,
- 2) tüve lõpuvaheldusmall,
- 3) formatiivvariantide komplekt paradigmas.

Nende tunnuste alusel tehakse kõigepealt kolm iseseisvat lähteklassifikatsiooni, mis iseloomustavad sõnamuutmist igaüks ise aspektist. Lähteklassifikatsioonide ühendamisel tekib koondklassifikatsioon - tüübistik, mis ühendab kõik kolm tunnust.

Koondklassifikatsiooni klasse nimetatakse **muuttüüpideks**. Ühte muuttüüpi kuuluvad seega sõnad, mis käituvad ühtmoodi kõigi morfoloogiliste tunnuste järgi: neil on ühesugused formatiivvariandid kõigis vormides ning ühesugune tüve lõpuvaheldusmall ja ühesugune astmevaheldusmall paradigmas.

Muuttüüpide morfoloogilised iseloomustused esitatakse **tüübikirjeldustena**. Et muuttüübi kirjeldus oleks võimalikult kompaktna ja ülevaatlik, ei hõlma see kõiki muutevorme, vaid ainult **põhivormid**. See on minimaalne hulk muutevorme, milles tulevad esile kõik vajalikud morfoloogilised tunnused. Ülejäänud vormide moodustamiseks fikseeritakse analoogiareeglid.

Tüübikirjeldus kujutab endast paradigma põhivormide esitust sellisel kujul, mis fikseerib kõik tüübile omased variandikombinatsioonid (ehk allotaktilised konfiguratsioonid). Põhivorm antakse tüübikirjelduses liitähisena, miile iga element esindab üht varieeruvat nähtust:

- a) tüvekuju lõpumuutuse seisukohalt: lemmatüvi (A), muutetüved (B, C, D), lõpumuutuseta tüvi (X);
- b) tüve aste: tugev (T), nõrk (N), astmevahelduseta tüvi (X);
- c) formatiivvariant.

Nt	sg n:	AN	liige	AN	hõige	AT	k'iik
	sg g:	BT	l'iikme	AT	h'õike	BN	kiige
	sg p:	ANt	liige/t	ANt	hõige/t	BT	k'iike
	pl g:	BTte	l'iikme/te	ANte	hõige/te	BTde	k'iike/de
	pl p:	BTid	l'iikme/id	ATid	h'õike/id	BTsid	k'iike/sid

Paradigma ülejäänud liikmed jagunevad põhivormide juurde kuuluvateks analoogiarühmadeks. Ühe analoogiarühma piires on kõigil muutevormidel ühesugune allotaktiline konfiguratsioon: täpselt sama tüvevariant mis põhivormil ja analoogiareegli abil leitavad formatiivvariandid. Nt mitmuse omastava analoogiarühma kuuluvad kõik teised mitmusevormid alates illatiivist:

pl g → pl ill, pl in, pl el, pl all, pl ad, pl abl, pl tr, pl ter, pl es, pl ab, pl kom

l'iikme/te → *l'iikme/tesse, l'iikme/tes, ...*

hõige/te → *hõige/tesse, hõige/tes, ...*

k'iike/de → *k'iike/desse, k'iike/des, ...*

Klassifikatsiooni optimeerimine. Klassifikatsiooni vahendusel saab eristada kõiki võimalikke iseärasusi sõnade muutmisel, kui omistada igale teistmoodi muutuvale sõnale oma tüübinumber. Praktilise lahendusena on see täiesti sobiv, aga lingvistilisest aspektist mitte kõige parem, sest nii lähevad kaduma samasugused sõnade muutmise vahel.

Et tüübistik oleks ülevaatlik ja informatiivne, on ühe tüübi alla kokku viidud sõnarühmi, mille muutmise üksikute vormide osas erineb teiste sama tüübi sõnade omast. Need on põhitüübi suhtes erandid. Sageli on erandlik ainult üks vorm või teatud vormide tüvi, või teatud vormide formatiivid. Suurem osa vorme moodustatakse aga täiesti regulaarselt. Nt sõna *olema* on *tulema*-tüübi erand, sest tal on kahes muutevormis erandlik kuju 'on' 'ind pr sg3' ja 'ind pr pl3' (vrd 'on - *tule/b, tule/vad*). Kõik muud vormid on tal samased *tulema*-tüübi vastavate vormidega (*ole/n - tuie/n, oll/akse - tull/akse* jne).

Tüübierand on suhteline mõiste, sest erandite hulk ja valik sõltub sellest, millised tüübid omavahel ühendatakse. Siinjuures on püütud saavutada optimaalne suhe tüüpide ja erandite vahel: ühelt poolt, et tüüpide hulk oleks võimalikult väike ja ülevaatlik, ning teiselt poolt, et ka erandite hulk ei paisuks liiga suureks. Eesti keele optimaalne morfoloogiline klassifikatsioon sisaldab 38 muuttüüpi, neist 26 noomenitüüpi ja 12 verbitüüpi, lisaks veel 14 tüübinumbrita sõna (pronoomenid). Erandlike vormidega sõnu on fikseeritud 84, kusjuures erandeid sisaldavaid tüüpe on 20 ja eranditeta tüüpe 18.

Erandid ise on väga kirju materjal, põhjustatud erinevatest asjaoludest. Erand võib näiteks olla morfotaktiline (kui mõni muutevorm puudub, nt *mõlema*) või allotaktiline (kui variantide kombinatsioon on ebatüüpiline, nt *m'aa/de* & *maa/de* - nõrk aste) või morfoloogiline (kui tüvemuutus on ebareeglipärane, nt *k'üt/ma* : *k'õe/fakse*).

Tüübituvastus

[8:38-44; 9:47-49; 18:161-162; 13]

Eesti keeles on suur hulk sõnu, mille muutmisvõimalused on määratud tüve enda fonoloogilise struktuuri poolt. Nt tüved, milles on vähemalt 3 siipi ja lõpus *-lik*, saavad tüvevokaaliks *u* ning on järgsilbis astmevahelduslikud (*õnnel'ik* - *õnneliku* - *õnnel'ikku*); tüved mis lõpevad pika vokaaliga, saavad ainsuse osastavas formaatiiviks *d* (*m'aa* - *m'aa/d*, *rokok'oo* - *rokok'oo/d*). Tüve teatud fonoloogilised omadused on eeltingimuseks, et sõna saab nii või teisiti muutuda.

Eesti morfoloogias on olulised sellised tüve fonoloogilise struktuuri omadused nagu silpide arv, välde või silbitüüp, sisehäälikud, lõpuhäälikud, vahel ka esimese silbi vokaal. Nende tunnuste kombinatsioon moodustab tüve **struktuurimalli**, mille põhjal on sageli võimalik otsustada, kuidas sõna morfoloogiliselt käitub: millised formatiivivariandid on paradigmas kasutusel ja millised on tüvevariantide vaheldusmallid. Seega struktuurimalli järgi saab tuvastada sõna tüübikuuluvust morfoloogilises klassifikatsioonis.

Struktuurimallide ja muutmisvõimaluste vaheline seos võib olla tugevam või nõrgem. Nt kõik verbid algvormi tüvestruktuuriga "3 silpi, I välde, lõpus *-ele*" muutuvad nagu *kõnele/ma* (*kõnele/ma* : *kõnel/da* e *kõnele/da* : *kõnele/n*). Samas aga verbid struktuuriga "2 silpi, III välde, lõpus *-u*" muutuvad enamasti nagu *m'uutu/ma*, st astmevahelduseta (*j'ahtu/ma* : *j'ahtu/b*), kuid väiksem osa sama struktuuriga verbe on astmevahelduslikud ja muutuvad nagu *'õppi/ma* (*m'ahtu/ma* : *mahu/b*). Esimesel juhul on tuvastusreegel absoluutne: tüve struktuur määrab üheselt sõna muuttüübi. Teisel tuvastusreegil on aga rida erandeid.

Sõnavormide süntees ja analüüs

[5:38-43; 8:45-51; 9:61-73; 16:55-57; 11; 18]

Sünteesi- ja analüüsialgoritmid on noomenil ja verbil ühised ning tuginevad morfoloogilises klassifikatsioonis sisalduvale infole.

Süntees. Selleks et mingist sõnast moodustada kõik tema muutevormid, tuleb läbida järgmine käik:

- 1) määrata sõna tüübikuuluvus (tuvastusreeglite abil või sõnaartikli põhjal);
- 2) võtta aluseks tüübinumbrile vastav tüübikirjelduse skeem;
- 3) asendada tüvevariantide tähised skeemis tegelike tüvevariantidega (moodustatud teisendusreegide abil või leitud sõnastikust);
- 4) liita tüvevariantidele tüübikirjelduses näidatud formatiivvariandid;
- 5) saadud põhivormidest moodustada vormianaloogia reeglite alusel ülejäänud muutevormid: võtta põhivormi tüvi ja liita formatiivid;
- 6) kui formatiiv on varieeruv, valida variant formatiivianaloogia reegli järgi;
- 7) kui tüve ja formatiivi piiril on lubamatuid järjendeid, teha morfoloogilise distributsiooni reeglitega määratud teisendused;
- 8) kui sõnal on erandlikke vorme, paigutada need tüübikirjelduse abil moodustatud vormide asemele.

Analüüs. Selleks et teha kindlaks, millise sõna millise vormiga on tegemist mingi sõnavormi puhul, tuleb läbida järgmine käik:

- 1) liigendada sõnavorm vastavalt morfofaktika reeglitele oletatavaks tüveks ja formatiiviks (k.a *G*-formatiiv);
- 2) formatiivide tabelist leida vastavad grammatilised tähendused ning formatiivvariandi allotaktilised piirangud (tüübinumbrite ja tüvekoodide kombinatsioonid);
- 3) tüvede sõnastikust leida vastav algvorm koos lisainfoga (tüübinumber ja tüvevariandi kood);
- 4) kontrollida formatiivvariandi piirangute ja algvormi lisainfo sobivust;
- 5) kui analüüs ei õnnestunud, siis liigendada sõnavorm teisest võimalikust kohast ning korrata formatiivi- ja tüvevariandi otsimist.

Paljudel juhtudel võib morfoloogiline analüüs anda sõnavormile mitu tõlgendust. Eesti keele küllalt ulatuslikus morfoloogilises homonüümiat on eristatavad kolm liiki:

a) paradigmasisene, mille puhul kattuvad sama sõna eri vormid:

purjekas — 'sg n' & 'sg in' PURJEKAS
elaside — 'ind ipf sg2' & 'ind ipf pl3' ELA/MA

b) vormisisene, mille puhul kattuvad eri sõnade samad vormid:

kuuega 'sg kom' — K'UUB & K'UUS
k'üttis 'ind ipf sg3' — K'ÜTTI/MA & K'ÜT/MA

c) vormidevaheine, mille puhul kattuvad eri sõnade eri vormid:

p'oega — 'sg p' P'OEG & 'sg kom' P'OOD
v'eost — 'sg p' V'EOS & 'sg el' VEDU
v'õita — 'sg ab' V'OI & 'inf' V'ÕIT/MA & 'ips pr neg' V'ÕID/MA

Olulisemad põhjused, mis homonüümiat tekitavad, on:

a) morfoloogiliste formatiivide endi homonüümiat:

s — 'sg in' & 'ind ipf sg3'
sid — 'pl p' & 'ind ipf sg2' & 'ind ipf pl3'
d — 'sg p' & 'pl n' & 'ind pr sg2'

b) sõnatüves toimuvad muutused:

p'oē — *p'ood* : *p'oē* & *puge/ma* : *p'oē/b*
lihase — *lihas* : *lihase* & *lihane* : *lihase*

c) sõnavormi erinevad liigendusvõimalused:

vihast — *viha/st* 'sg el' V'IHT & *vihast* 'sg p' VIHANE
j'ooksin — *j'oo/ksin* 'knd pr sg1' J'OOMA
& *j'ooks/in* 'ind ipf sg1' J'OOKS/MA
m'eest — *m'eesst* 'sg p' M'EES & *m'өө/st* 'sg el' MESI

Morfoloogiliste homonüümide hulgast sobiva valik pole lahendatav üksnes morfoloogia tasandil, sest sõnavormi tõlgendus sõltub kontekstist.

Grammatika ja sõnastiku integratsioon

[16; 17; 18:153-163]

Grammatika ja sõnastiku meetodid ja võimalused sõnamuutmise kirjeldamisel on põhimõtteliselt erinevad. Grammatika taotleb üldistust ja ülevaatlikkust. Seetõttu esitatakse seal info valdavalt reeglite ja tabelite kujul. Reeglite juurde kuuluvad tavaliselt loendid, mis lõpevad lühendiga *jne*, sest ammendavate sõnaloendite jaoks poie grammatikas kohta. Sõnastik seevastu peab fikseerima info iga üksiksõna kohta ilma keerukate reeglite abita. Seetõttu ei saa sõnastik anda ülevaatlikku pilti sõnade muutmisest. Ainult grammatika ja sõnastik koos on võimelised keele morfoloogiasüsteemi kirjeldama niihästi ülevaatlikult kui ka ammendavalt.

Mida suurem on morfoloogiakirjelduses reeglite osa, seda väiksem roll on sõnastikul. Kui grammatika puudub, peaks sõnastik sisaldama igast sõnast kõik vormid koos vastava grammatilise iseloomustusega, nt *kirjutamine* 'sg n' KIRJUTAMINE, *kirjutamise* 'sg g' KIRJUTAMINE, *kirjutamist* 'sg p' KIRJUTAMINE jne. Praktikas on selline sõnastik võimatu, eriti eesti keele puhul, kus igal sõnal on keskmiselt 33 vormi.

Kui grammatika sisaldab tüübikirjeldusi ja analoogiareegleid ning neist sõltuvaid morfoloogilise sünteesi ja analüüsi reegleid, siis piisab sõnastikus sellest, kui iga sõna jaoks on antud algvorm, kõik erinevad tüvevariandid (kooskõlas tüübikirjeldustega), tüübinumber ja erandlikud vormid, nt *kirjutamine kirjutamise kirjutamis* S 12.

Kui grammatika sisaldab tüvemuutuste reegleid, siis võivad tüvevariandid sõnastikust välja jääda. Nt tüübis 12 saadakse tüvevariandid reeglite "ne → se" ja "se → s" abii ja vajalik sõnastikuinfo on *kirjutamine* S12.

Algvormide hulka sõnastikus saab vähendada paradigmaatilise derivatsiooni reeglite lisamisega grammatikasse. Nt reegel "verbi algvormi tüvi + *mine* → noomen S 12" asendab kõiki verbi *mine*-tuletisi, mis muutuvad 12. tüübi järgi. Vajalik sõnastikuinfo sõna *kirjutamine* jaoks on se! juhul: *kirjuta/ma* V 27.

Kui lisada grammatikasse tüübituvastusreeglid, väheneb sõnastiku algvormide hulk veelgi. Nt verbi tüvestruktuur "3 jne silpi, lõ-

pus mitte-*ele*" määrab sõna muutuma tüvevahelduseta: nagu *elama* V 27. Seega reegli "verb + 3 jne silpi, lõpus mitte-*ele* → V 27" olemasolu vabastab sõnastiku kõigist 3- ja enamasilbilistest verbi-dest, mis muutuvad 27. tüübi järgi ning sõna *kirjuta/ma* õigeks muutmiseks piisab ainult reeglite teadmisest.

Reeglite lisamine grammatikasse vähendab sõnastiku mahtu, kuid ei kaota vajadust sõnastiku järele, sest keelereeglitel on peaaegu alati rohkem või vähem erandeid. Kõik need sõnad, mille morfoloogiline käitumine on mingis mõttes erandlik ja ebaproduktiivne, vajavad lisainfot sõnastikus.

Sõna võib olla erandlik mitmes mõttes ja vastavalt on erinev ka vajalik lisainfo. Tuvastusreeglite erandid vajavad tüübitähist, mis suunab teise tüübikirjelduse juurde. Tüübierandid (ehk allotaktika reeglite erandid) vajavad väljatoomist erandlike vormidena. Teisendusreeglite erandid vajavad suunamist teise reegli juurde või erandlike tüvevariantide esitamist valmiskujul. Morfotaktika reeglite erandid vajavad infot puuduvate vormide kohta.

Milline on informatsioonilise koormuse optimaalne jaotus sõnastiku ja grammatika vahel, sõltub morfoloogiakirjelduse adre-ssaadist ja otstarbest ning võib olla igal konkreetsel juhul erinev.

"Väike vormisõnastik". Sõnastiku ja grammatika integrat-siooni on taotletud "Väikese vormisõnastiku" (VVS) koostamisel. See on eesti keele esimene spetsiaalne morfoloogiasõnastik ning kujutab endast klassifikatoorse morfoloogia mudeli realisatsiooni nii trükisena kui ka elektroonilisel kujul.

VVS-i raamatuversioon koosneb kahest võrdväärsest ja oma-vahe! lahutamatult seotud osast: grammatika ülevaatest ja sõnasti-kust. Grammatika annab ettekujutuse muutmissüsteemist tervikuna ja sellesse kuulub üldine morfoloogiline info:

- 1) ülevaade grammatilistest kategooriatest ja paradigmade kirjeldused;
- 2) morfoloogiliste formatiivide tabelid;
- 3) muuttüüpide tabelid;
- 4) tüübikirjeidused koos näidetega;
- 5) sõnavormide sünteesi ja analüüsi algoritmid;
- 6) juhised regulaarsete tuletiste moodustamiseks.

Sõnaartikkel sõnastikus esitab andmed iga konkreetse sõna kohta:

- 1) märksõna;
- 2) muud tüvevariandid, mis sõna muutevormides võivad esile tulla;
- 3) morfoloogiline indeks, mis koosneb sõnaliigi tähisest, tüübinumb-rist ja tüvevitmuse korral lisanduvast mitmuse vokaalist;
- 4) tüübist hälbivad morfoloogilised omadused: sõna mitmuslikkus ja erandlikud vormid;
- 5) valik liitsõnu ja sõnaühendeid näite tasandil.

VVS on üles ehitatud nii, et tema abil saab lihtsate ja täiesti formaalsete reeglitega kirjeldada eesti keele morfoloogilise süsteemi funktsioneerimist — sõnavormide sünteesi ja analüüsi. Eduka sünteesi jaoks piisabki ainult sõnaartiklite ja grammatika infost. Analüüsi jaoks on seda siiski vähe. Tüve muutuste tõttu võivad sama sõna vormid olla väga erinevad, nt *uba* : 'oa/d, k'üt/ma : k'õe/akse.

Niisuguste vormide kokkuviimiseks on sõnastikku lisatud viiteartiklid. Viitemärksõnadeks on need tüvekujud, mis põhimärksõna omast erinevad nii palju, et ei satu alfabeedis tema kõrvale:

'oa < uba, k'õe/akse < k'üt/ma.

Peale grammatika ja sõnastiku sisaldab VVS ka lisasid, mis annavad morfoloogiasüsteemile täiendava vaate ning esitavad sõnastikus sisalduva info teisel kujul:

Lisa 1: tüübiloendid — märksõnad muuttüüpide kaupa (pöörd-järjestuses),

Lisa 2: liitsõnade ja sõnaühendite seotud komponendid,

Lisa 3: liitsõnakomponentide sagedusloendid,

Lisa 4: võõrkomponentide tabelid.

VVS-i lähtematerjalina on kasutatud ÕS-i ("Õigekeelsussõnaraamat" Tallinn 1976), mis on aga arvuti abil põhjalikult ümber töötatud. Kogu materjal, mida VVS otseselt või kaudselt püüab kajastada, on järgmine:

- 1) kõigi märksõnade kõik võimalikud muutevormid;
- 2) regulaarsed tuletised;
- 3) liitsõnades ja sõnaühendites avalduvad sõna- või tüvekujud;
- 4) regulaarsed võõrsõnakomponendid.

Muutevormid. Morfoloogilise info esitamisel võtab VVS arvesse kõiki võimalikke sõnavorme, sealhulgas ka teoreetilisi, mis tegelikus keelekasutuses ei pruugi üldse esineda, kuid mida formaalsete reeglite järgi saab moodustada. Sellised on näiteks paljude verbide umbisikulise tegumoe või käskiva kõneviisi või 1. ja 2. isiku vormid (*p'ungu/ti, j'uhtu, s'eisne/n, järøldu/d*), noomenitel sageli lühikese sisseütleva või tüvemitmuse vormid või mitmusevormid üldse (*tibu : t'ippu, p'õid : põll/el, mesi : me/teks*). Kogu see tegelike ja teoreetiliste muutevormide hulk pole sõnastikus fikseeritud mitte otseselt, vaid võimalusena, mille realiseerib kasutaja sõnartiklites ja grammatikas sisalduvate andmete ja sünteesiaalgoritmi järgi.

Tuletised. Tuletuse all on VVS-is mõistetud kõiki selliseid sõnamoodustusprotsesse, kus sufiksi lisamisel tekib uus morfoloogiline sõna, millel on olemas oma muutmispäradigma. Selle määratluse järgi osutuvad tuletisteks ka omadussõnade võrdlusastmed (*vana: vane|m, van|im*) ja kesksõnad (*luge/ma: luge|v, l'øe|tav, luge|nud, l'øe|tud*). Kesksõnad on kahepaikse iseloomuga, kuuludes ühtlasi nii tuletiste kui ka verbi muutevormide hulka.

Kõige lähemal morfoloogiale on nn **paradigmaatiline derivatsioon**. See hõlmab juhtumid, kus tuletusliide käitub analoogiliselt morfoloogilise formatiiviga: sufiksi lisamine ei muuda tuletusaluse konkreetset leksikaalset tähendust, vaid eelkõige tema kategoriaalset tähendust. Sufiks vahetab tüve sõnaliiki (nt verb + *mine* → substantiiv) või lisab mingi väga üldise semantilise komponendi (nt noomen + *ke* → deminutiivnoomen).

Kõige regulaarsemad paradigmaatilised tuletised esitatakse VVS-is analoogiliselt sõna muutevormidega: tuletised ise (nt verbi *luge/ma* puhul vormid *luge|mine, luge|ja, luge|mata, l'øe|tud* jts)

jäetakse märksõnadena esitamata, kuid antakse võimalus nende moodustamiseks grammatikas fikseeritud formaalsete reeglite abil. Paradigmaatilise derivatsiooni reeglid on koondatud tabelisse, kus iga sufiksi jaoks on näidatud:

- 1) tuletusalus: millisele vormile sufiks liitub;
- 2) morfoloogiline kirjeldus: kuidas tuletis muutub;
- 3) näited.

Nt

mine	V sup-tüvi	mi/ne -se -s/t	S 12 i	luge/ma — lugemine
ja	V sup-tüvi	ja	S(A) 1	kõnele/ma — kõneleja
m	A 'sg g'	m -a	A 2	suure — suurem

Sufiksi morfoloogilise kirjelduse kaudu saavad kõik vastavad tuletised endale automaatselt sellesama kirjelduse ja ülituvad üldisse morfoloogilisse süsteemi: igast niiviisi moodustatud tuletisest saab edasi moodustada kõik tema muutevormid. Need üksikud erandid, mis muutuvad teisiti kui reeglitega määratud, on sõnastikus eraldi märksõnadena sees.

Liitsõnad ja sõnaühendid. Liitsõnamoodustus on eesti keeles kohati üsna vaba ja liitsõnade arv praktiliselt lõpmatu. Seetõttu ei saa ükski sõnastik neid kõiki sisaldada. Et sõnastikku mitte koormata, kuid seejuures siiski säilitada liitsõnades peituv morfoloogiline info, on VVS-is märksõnadena esitatud mitte liitsõnad ise, vaid ainult liitsõnade seotud komponendid — need, mis mujal esile ei tule: a) esikomponendid, nt *viker(+k'aar)*, *sõemis(+n'ahk)*, *väri(+h'ein)*, *j'äätt(+m'aa)*, *üli(+p'üüdl'ik)*, *'umb(+s'õlm)*; b) järelkomponendid, nt *(v'iis+)aastak*, *(tähele+)panel'ik*, *(abi+) 'elluma* jne. Iga seotud järelkomponendi juurde kuulub morfoloogiline iseloomustus, nii nagu iseseisvatel sõnadelgi. Liitsõna seotud komponendid on kõik esitatud ka eraldi lisana (Lisa 2).

Liitsõnade varju on ÕS-is peidetud suur hulk sõnaühendeid (peamiselt ühend- ja väljendverbid), nt *järele+ahvimine* — *järele 'ahvima*, *j'õu+k'atsumine* — *j'õudu k'atsuma*, *võidule+p'ääs* — *võidule p'ääsema*. Ka sellistes ühendites sisalduv morfoloogiline info on VVS-is võimaluste piires säilitatud. Märksõnadena on antud ühendite seotud komponendid — tüve- või sõnakujud, mida muutevormide hulgas ei esine, aga mida ka iseseisvateks sõnadeks ei

saa pidada, nt *pl'ehku* (*_panema*), *p'eksa* (*_s'aama*) jne. Ka need vormid on lülitatud Lisasse 2.

Võõrsõnad. Suurem osa ÕS-i võõrsõnu on VVS-ist välja jäetud, põhiliselt need, mis morfoloogia seisukohalt huvi ei paku. Sellist kadu on püütud korvata samal kombel, nagu omakeelsete liitsõnade puhul. Võõrsõnaseeriade asemel antakse märksõnadena valik võõrliiteid ja liitsõna osiste taolisi elemente, nt *grafo-*, *mikro-*, *pre-*; *-gr'aaf*, *-iit*, *-s'æemia*. Järelkomponentide juurde kuulub samuti morfoloogiline iseloomustus.

Kuna VVS-is moodustab grammatika koos sõnastikuga funktsionaalse terviku ja tagab sellega kasutajale võimaluse ise sünteesida ja analüüsida sõnavorme, siis tänu sellele sisaldab sõnastik eesti keele sõnavormide kohta peidetud kujul ehk rohkemgi andmeid kui tema lähtesõnastik ÕS, kuigi on mahult tunduvalt väiksem ja kompaktsem. ÕS-is on märksõnu ca 114.000, VVS-is aga 36.200 (+ 9.300 viiteartiklit).

Arvutirakendused

[1; 11; 12; 14; 15; 18]

VVS-i elektrooniline versioon samaneb trükiversiooniga oma sisu poolest, kuid erineb vormilt. Teda saab kasutada mitmel erineval otstarbel:

1) VVS kujutab endast lingvistilist andmebaasi, kust uurijad võivad neile vajalikku materjali saada kätte kiiresti, ammendavalt ja sobival kujul (mitmesugused loendid erinevate tunnuste järgi ning eri viisil järjestatult), või mida saab töödelda mitmesuguste programmidega (nt statistiline, fonoloogiline, morfoloogiline analüüs);

2) VVS on kasutatav leksikograafilise andmebaasina, millest saab iga sõna muutmise kohta andmeid küsida, ning mille elemente saab kasutada ka uute sõnastike koostamisel (nt morfoloogilise info ülekanne teistesse sõnastikesse);

3) VVS talitleb lingvistilise protsessori komponendina suuremates infotötluse süsteemides.

VVS-i baasil on arvutis realiseeritud nii morfoloogiline süntees kui ka analüüs. Sünteesiprogrammi abil on koostatud eesti keele sõnavormide pöörd sõnastik (ca 1,2 miljonit sõnavormi), mis omakorda oli aluseks kahe uue sõnastiku tegemisel:

- 1) morfoloogiliste homonüümide sõnastik (ca 78.000 vormi),
- 2) riimisõnastik (ca 175.000 riimisarja 600.000 riimuvast sõnavormist).

Sünteesiprogrammil tugineb ka loodav eesti keele morfoloogilise nõustamise süsteem.

VVS-il baseeruvaid analüüsiprogramme on praegu teadaolevalt kaks. Üks on tehtud TÜ-s eesti keele automaatse õigekirjakontrollija (spelling checker) jaoks. Teine, keelereeglite tuginev analüüsiprogramm on tegemisel EKI-s.

Automaatse morfoloogia programmid on vajalikud kõigis eesti keelt kasutavates tekstitöötlussüsteemides, info automaatsel otsimisel, inimese ja arvuti dialoogi modelleerimisel, automaattõlke ja tehisintellekti süsteemides, eesti keele õppimise, õpetamise ja nõustamise süsteemides jne.

Kui automaatne morfoloogiasüsteem lähtub loomuliku keele enda reeglitest, siis on tal praktilise rakendusväärtuse kõrval ka omaette väärtus lingvistilise teooria kontrollimise seisukohalt.

Publikatsioonid

Töö tulemused ja arendused on avaldatud järgmistes publikatsioonides:

1. Õigekeelsussõnaraamatu arvutivariant. Verbid.
— Keel ja Kirjandus 1974, 10: 586-596 (1 ap)
2. Eesti verbid. Morfoloogiline klassifikatsioon. Loendid.
Estonian Verbs. Morphological Classification. Lists
(Preprint KKI-3). Tallinn 1974 (+ M. Remmel) (6,1 ap)
3. Verbide muuttüübid ja morfoloogilised tüübid.
— Keel ja Kirjandus 1976, 5: 276-288 (1,3 ap)
4. Sõnaliik kui niisugune.
— Keel ja Kirjandus 1977, 9: 521-525 (0,4 ap)
5. Klassifikatoorse morfoloogia põhimõtted.
"Ars grammatika" (Preprint KKI-9). Tallinn 1977 (2,78 ap)
6. Morfoloogilise klassifikatsiooni optimeerimisest.
— Sõnast tekstini. Tallinn 1978: 91-111 (1 ap)
7. Kuidas formaliseerida tüvemuutusi.
— Keel ja Kirjandus 1979, 11: 671-677 (0,7 ap)
8. Klassifikatoorne morfoloogia. Verb.
"Ars grammatika". Tallinn 1980 (3,35 ap)
9. Klassifikatoorne morfoloogia. Noomen.
"Ars grammatika". Tallinn 1982 (4,76 ap)
10. Ühest morfoloogilisest klassifikatsioonist.
— Keel ja Kirjandus 1982, 10-11: 517-525, 575-586 (2 ap)
11. Sõnavormide homonüümia eesti keeles.
— Keel ja Kirjandus 1984, 2: 97-105 (1 ap)
12. Morphological synthesis at automatic compilation of dictionaries.
— Symposium on automatic compilation of dictionaries
(Tallinn, November 25-27, 1985). Summaries.
Tallinn 1985: 84-87 (0,35 ap)
13. Fonoloogiliste ja morfoloogiliste mallide seosed.
— Arvutuslingvistika sektori aastaraamat 1988.
Tallinn 1990: 160-166 (0,4 ap)
14. Sõnastike andmebaas: milleks, mis ja kuidas.
— Arvutuslingvistika sektori aastaraamat 1988.
Tallinn 1990: 167-175 (0,5 ap)

15. Riimisõnastiku sünnist.
– Keel ja Kirjandus 1991, 10: 614-617 (+ J.Sang) (0,4 ap)
- 16.1. Väike vormisõnastik I. Sissejuhatus & grammatika.
Tallinn 1992 (4 ap)
- 16.2. A Concise Morphological Dictionary of Estonian I.
Introduction & Grammar. Tallinn 1992 (4 ap)
- 16.3. Краткий морфологический словарь эстонского языка I.
Введение & грамматика. Таллинн 1992 (4 ap)
17. Väike vormisõnastik II. Sõnastik & lisad.
Tallinn 1992 (54 ap)
(A Concise Morphological Dictionary of Estonian II.
The Dictionary & Appendices.
Краткий морфологический словарь эстонского языка II.
Словарь & приложения.)
18. Eesti keele morfoloogiline analüsaator. Automaatanalüüsi
võimalused ja võimatused.
– Keel ja Kirjandus 1994, 3: 150-163 (1,4 ap)

CLASSIFICATORY MORPHOLOGY OF ESTONIAN

Abstract

The Estonian language belongs to "morphological" languages in which the grammatical meanings are expressed by morphological means in the first place. Estonian is characterized by a great number of inflected forms. The Estonian noun has 14-15 cases in singular and plural, together with their numerous parallel forms, while the verb is represented by nearly 60 simple forms, apart from the analytical forms of tense and negation.

Another essential peculiarity of the Estonian language is the extensive variation of its morphological units. Variation may characterize both morphological formatives and stems, while a stem may be subject to different kinds of variation. Although a stem may also remain unchanged, in most cases it is patient of stem-grade alternations or stem-end alternations, or both. Moreover, the location of stem variants in paradigms can follow different alternation patterns. Owing to the multitude of variants the efficiency of the description of the language is largely dependent on its allotactic part dedicated to the comb inability of unit variants.

The present paper proposes a model of classificatory morphology to handle the morphological idiosyncrasies of the Estonian language, and a description of the Estonian morphology, based on this model.

The model of classificatory morphology describes a structure of morphological system (units and their relationships) as well as its functioning (analysis and synthesis of word forms). The model focuses, above all, on differences in the morphological behaviour of words employing the classification of words by morphological features as the main tool.

The choice of the criteria of classification is based mainly on the relations between morphological features and linguistic rules. This provides for the adequacy of the classification by describing morphological analysis and synthesis.

Classification reveals only the allotactic configurations of the unit variants in paradigms, while the interviant transformational relations are taken care of by special rules. The whole multitude of words are classified according to three features:

- 1) stem-grade alternation pattern,
- 2) stem-end alternation pattern,
- 3) paradigmatic set of formative variants.

On the basis of these features three separate classifications are drawn up each of which characterizes inflection from a different aspect. Next, these preliminary classifications are combined into a summary classification considering all three features. Classes in the summary classification are called inflection types.

The morphological character of an inflection type is presented in the form of a type description. To make the description as concise and comprehensive as possible only a minimum number of basic inflectional forms (5-7 for a noun, 4-9 for a verb) are included, in which all of the necessary morphological features are still manifest. The rest of forms can be obtained by means of fixed rules of analogy.

An optimum morphological classification of Estonian includes 38 inflection types: 26 for nouns, 12 for verbs (plus 14 pronouns without type numbers). There are 84 words having some irregular forms.

The classificatory description of the Estonian morphology is presented by a combined grammar and dictionary. The two are inseparable employing different methods and having different potentialities. As the grammar seeks clarity and generalization in the first place, it describes the morphological system as a set of phenomena and their relations, doing it by means of rules and tables. The dictionary, however, endeavours to be exhaustive, so it describes the system as a set of individual words providing each with full information in the corresponding entry. Only an integral presentation of both dictionary and grammar permits to describe the morphological system clearly as well as exhaustively.

This aim has been realized in A Concise Morphological Dictionary of Estonian (MDE) that consists of two equally essential and inseparably connected parts: a grammatical survey and a dictionary. The grammar part contains general morphological information:

- 1) a survey of grammatical categories and descriptions of paradigms;
- 2) tables of morphological formatives;
- 3) tables of inflection types;
- 4) type descriptions together with examples;
- 5) algorithms for the analysis and synthesis of word forms;
- 6) instructions for regular derivation.

Every dictionary entry supplies the following information:

- 1) entry word;
- 2) other stem variants that may appear in inflected forms of the same word;
- 3) a morphological index consisting of a part-of-speech marker, type number and plural vowel (required to form stem plural);
- 4) type-deviant traits: *pluralia tantum* and irregular forms;
- 5) an illustrative selection of compound words and word combinations.

Besides the grammar and the dictionary MDE also contains appendices to give an additional perspective to the morphological system by presenting the dictionary information in another way:

Appendix 1: type lists — reverse-ordered entry words classified by inflection types;

Appendix 2: the bound members of compound words and word combinations;

Appendix 3: frequency lists of compound word members;

Appendix 4: tables of foreign components of words.

There is also an electronic version of MDE that can be used as follows:

- 1) a linguistic database providing research material and being analyzed by means of various available programs;
- 2) a lexicographic database the elements of which may be used to compile new dictionaries;
- 3) part of linguistic processor in some other system of information processing.

On the basis of the computerized MDE both the synthesis and analysis of the Estonian word forms have been realized.