

# ERAND REEGEL JA SÕNASTIK AVATUD MORFOLOOGIAMUDELIS

Ülle Viks

## 1. VAATEPUNKT

Alljärgnevas artiklis püüan probleeme käsitleda eelkõige eesti keele morfoloogilise süsteemi uurijana ja modelleerijana, kelle käsituses on arvutilingvistika võimalused. Keelesüsteemi formaliseerimine ja realiseerimine arvutis pole eesmärk omaette, vaid küllaltki tõhus vahend püstitatud hüpoteeside kontrollimiseks ning mudeli kõigi allsüsteemide testimiseks.

**Avatud** morfoloogiamudel tähendab seda, et mudeli rakendusulatus arvutis ei ole piiratud mingi kindla hulga sõnadega, vaid et süsteem saab hakkama ka talle tundmatute sõnadega – nii nagu inimene oskab käänata-pöörata ka neid sõnu, mida ta kunagi varem kasutanud pole. Inimese morfoloogiapädevusse kuulub teatud hulk sõnavorme, mida ta tunneb ja oskab lauses kasutada, kuid lisaks neile kindlasti ka teatud hulk reegleid, mille abil ta on võimeline vajaduse korral sõnavorme moodustama. Sama põhimõte on aluseks ka morfoloogia modelleerimisel arvutis.

## 2. TAUST: MORFOLOOGIASÜSTEEM JA SELLE KIRJELDUS

Morfoloogiasüsteemi kesksed protsessid on sõnavormide süntees ja analüüs. Morfoloogilise **sünteesi** sisend on algvorm ehk lemma ja sõnavormi grammatilised tähendused (nt JALG + 'sg el') ning väljund on vastav sõnavorm (JALAST). Morfoloogilise **analüüsi** sisend ja väljund on vastupidised: sisendiks on sõnavorm (JALAST) ning väljundiks on lemma koos vastavate grammatiliste tähendustega (JALG + sg el').

Sünteesi käigus valitakse sisendile vastavad üksused (lemma – tüvevariant, grammatilised tähendused – morfoloogiline formatiiv) ja ühendatakse need sõnavormiks. Analüüsi käigus liigendatakse sõnavorm üksusteks ja leitakse tüvevariandile vastav lemma ning formatiivivariandile vastavad grammatilised tähendused.

Tegelikult muidugi asi nii lihtne ei ole: analüüsi- ja sünteesi- protsesse võib vaadelda pööratavatena ainult teatava lihtsustuse

korral. Protsessid, mis toimuvad sisendi ja väljundi vahel, on kummagi suuna puhul paljuski erinevad. Kuid ometi on tegemist ühe ja sama allsüsteemiga. Analüüs ja süntees kasutavad suures osas samu andmeid ja neid protsesse on võimalik selliselt arvutis realiseerida, et nad kasutavad mitmeid ühiseid programmimooduleid. Seetõttu saab neist ka koos rääkida.

Põhilised komponendid, mis on olemas kõigis morfoloogia-süsteemide kirjeldustes – nii traditsioonilistes kui automaatsetes – on grammatika ja sõnastik. **Grammatika** sisaldab reegleid, mis kirjeldavad üksustega süsteemis toimuvaid protsesse: morfoloogilist sünteesi ja analüüsi. **Sõnastik** kujutab endast üksuste loendit, mis sisaldab süsteemi funktsioneerimiseks vajalikke andmeid iga üksuse kohta.

Sõnastik ja grammatika kirjeldavad üht ja sama nähtust – antud juhul morfoloogilist süsteemi –, kuid teevad seda erineval viisil. Grammatika üldistab ja annab süsteemist ülevaate, mis ei pruugi olla ammendav; sõnastik võimaldab ammendavat kirjeldust, mis pole aga ülevaatlik. Avatud morfoloogiamudel üritab integreerida grammatika ja sõnastiku erinevad võimalused morfoloogiasüsteemi kirjeldamisel. See, mis on morfoloogias üldine ja reeglipärane, esitatakse formaalsete reeglite kujul, s.t grammatikana, ja see, mida ei saa (või pole mõtet) reeglistada, esitatakse sõnastikus. Sõnastik sisaldab sel juhul ainult erandeid.

### 3. ERAND

Erandi mõistet kasutatakse keelesüsteemi kirjeldamisel küllalt tihti, kuid selle sisu tavaliselt ei defineerita, vaid kasutatakse kui väljastpoolt antut. Eesti kirjakeele seletussõnaraamat (1991) annab erandile sellise definitsiooni: “milleski üldisest v. normaalsest kõrvalekalduv v. erinev nähtus, omadus, asi, isik vms.”

Erand saab oma konkreetse sisu seoses taustsüsteemiga: ta on erinev millestki, ta on miski, mis pole see. Ja see see on enamasti määratletud kui üldine, tavaline, normaalne, reeglipärane, ühtlane – see on midagi umbmäärast, millel on pigem statistiline iseloom.

Sõnu ‘erandlik’ ja ‘irregulaarne’ võib lingvistilises kontekstis lugeda sünonüümideks, mis aga tavaliselt seostuvad erinevate antonüümidega: *erandlik* vastandub rohkem sõnadele ‘üldine’, ‘tüüpiline’ ja ‘tüübipärane’, *irregulaarne* vastandub sõnadele ‘regulaarne’ ja ‘reegipärane’ ‘Tüüp’ ja ‘reegel’ aga omavahel enam süno-

nüümid ei ole. Neid saab täpsemalt määratleda ning nende kaudu on võimalik määratleda ka erandeid.

**Reegel** kirjeldab morfoloogias mingit üksikut seaduspärasust, mis kehtib teatud omadusi kandvate sõnade puhul. Need omadused esindavad reegli rakendumistingimusi. Nt üht eesti keele tüvemuutust tugeva astme nõrgenemisel kirjeldab reegel  $k \rightarrow g$ , mis rakendub juhul, kui  $k$ -le eelneb lühike vokaal ja  $n$  ning  $k$ -le järgneb  $i$  (rakendumistingimused on kaldkriipsu / taga), nt  $l'anki \rightarrow langi$ ,  $t'ankima \rightarrow tangib$ . Koos häälikumuutusega nõrgeneb ka välde ( $3 \rightarrow 2$ ).

$$k \rightarrow g / Vn\_i$$

$$3 \rightarrow 2 / \text{nõrk aste}$$

**Erandit** saab defineerida reegli kaudu: erand on objekt, mille puhul reegli rakendumistingimused on küll täidetud, kuid mis ometi reeglile ei allu. Nt sõna  $t'ank$  võib nimetada esitatud häälikumuutusreegli suhtes erandiks, sest sellises kontekstis peaks toimuma tüvemuutus (nagu sõnades  $l'ank$  või  $t'ankima$ ), aga sõnas  $t'ank$  jääb  $k$  nõrgenemata ja nõrgeneb ainult välde. Seega sõna  $t'ank$  on selle reegli erand.

Erandite hulk tuleneb otseselt sellest, kui palju ja kui keerulisi reegleid fikseeritakse. Kui taotleda väga täpseid reegleid ja ammentavat kirjeldust, võib eranditest täiesti loobuda – reegel võib ju kehtida ka üheainsa sõna puhul (sel juhul tuleb tema rakendumistingimusena esitada terve sõna). Kui aga taotleda lihtsaid reegleid ja ülevaatlikku kirjeldust, on erandeid rohkem.

Näiteks on täiesti võimalik reeglistada ja ka arvutis realiseerida tüvemuutused sõnades  $pr'aad$ ,  $r'aad$  ja  $l'aad$ . Et arvutis tekitada algvormist  $pr'aad$  nõrga astme tüvi  $pr'ae$ , tuleb rakendada terve komplekt järjestikuseid reegleid, kusjuures osa vahevorme keeles ei eksisteeri (trell # tähistab sõnapiiri, reaalsed vormid on alla joonitud):

$0 \rightarrow i / VVd\_ \#$	$\underline{pr'aad} \rightarrow \underline{pr'aadi}$
$3 \rightarrow 2 / \text{nõrk aste}$	$\underline{pr'aadi} \rightarrow \text{praadi}$
$d \rightarrow 0 / VV\_i$	$\text{praadi} \rightarrow \text{prai}$
$V1V1 \rightarrow V1 / VV\_V$	$\text{prai} \rightarrow \text{prai}$
$i \rightarrow e / d \rightarrow 0$	$\text{prai} \rightarrow \text{prae}$
$0 \rightarrow ' / 1 \text{ silp}$	$\text{prae} \rightarrow \underline{pr'ae}$

Sellele reeglikomplektile alluvad 1-silbilised noomenid, mille algvormi lõpus 'aad' Kuid lisaks neile kolmele sõnale, mis käituvad esitatud reeglite kohaselt, leidub veel samasuguse struktuuriga sõnu, mille puhul reeglite rakendustingimused on samuti täidetud, kuid mis käituvad teisiti: üks sõna saab teise tüvevokaali (*s'aad* → *s'aadu* → ... *s'ao*) ja kolmes sõnas (neist kahel on olemas reeglipärased homonüümid) nõrgeneb ainult välde (*kr'aad* → *kraadi*, *r'aad* → *raadi*, *l'aad* → *laadi*). Seega on neil reeglitel erandeid kokku rohkemgi kui reeglipäraseid sõnu.

Kui aga sellest reeglikomplektist tüvemuutuste moodulis loobuda, osutuvad eranditeks hoopis varem reeglitele allunud sõnad (*pr'aad* jne). Endised erandid (*kr'aad* jne) alluvad kahele üldisemale reeglile, millest üks lisab tüvevokaali *i* ja teine muudab ainult välde (sõna *s'aad* : *s'ao* jääb erandiks ka nüüd):

0 → i / VVd_#	<u>kr'aad</u> → <u>kr'aadi</u>
3 → 2 / nõrk aste	<u>kr'aadi</u> → <u>kraadi</u>

See lihtsam ja üldisem reeglikomplekt kehtib suure hulga sõnade puhul silpide arvust ja konkreetsetest vokaalidest sõltumata (*dek'aad*, *promen'aad*, *kat'ood*, *antip'ood*, *pl'eed*, *mop'eed*, *g'iid*, *insektits'iid*, *sl'aid*, *sinus'oid* jne).

Selle reegli kõrval on aga veel reeglid, mis kehtivad sõnade puhul, mille algvormi lõpus on *b* või *g* (vältereegel jääb kirjutamata):

0 → i / VVb_#	<i>amf'iib</i> → <i>amfiibi</i> , <i>sn'oob</i> → <i>snoobi</i>
0 → i / VVg_#	<i>privil'eeg</i> → <i>privileegi</i> , <i>meteorol'oog</i> → <i>meteoroloogi</i>

Kummalgi neist on oma erandid, nt *t'iib* : *tiiva*, *h'oob* : *hoova*; *r'oog* *r'oa*, *h'uug* : *huua*. Kui üldistada kõik kolm reeglit kujule 0 → i / VVG\_# (kus G = nõrk klusiil), siis ühinevad ka nende erandid.

Eelnevatest näidetest võib teha järelduse, et erandit kui niisugust polegi olemas: erandid tekivad süsteemi kirjeldamisel ja moodustavad süsteemi selle osa, mis jääb reeglite mõjupiirist välja poole.

Kui reegli kaudu saab erandit küllalt konkreetsetl määratlada, siis tüübi kaudu defineerimisel on erandi mõiste segasem. Esiteks

on klassifikatsioon ise suhteline mõiste, sest tema liigitusaluseid saab valida erinevalt: võib valida ainult formatiivivariantide komplekti ja tüvemuutuste liigi nagu "Väikeses vormisõnastikus" (=VVS, Viks 1992), võib aga valida klasse eristama ka konkreetse tüvemuutusreegli nagu Muugi süsteemis (Muuk 1933). Liigitusaluste hulgast ja üldistusastmest sõltub ka klasside hulk

Klassifikatsiooni vahendusel on seega põhimõtteliselt võimalik eristada kõiki iseärasusi sõnade muutmisel – nii regulaarseid kui ka irregulaarseid. Seda juhul, kui igale teistmoodi muutuvale sõnale omistada oma tüübinumber. Praktilise lahendusena arvutis on see täiesti sobiv, aga lingvistilisest aspektist mitte kõige parem, sest nii lähevad kaduma sarnasusseosed sõnade muutmise vahel.

**Tüübierandid** tekivad klassifikatsiooni optimeerimisel. Et tüübigraaf oleks ülevaatlik ja informatiivne, võetakse ühe tüübi alla kokku sõnarühmad, mille muutmine on suures osas ühesugune ja erineb teiste sama tüübi sõnade omast ainult üksikute vormide osas (Viks 1978). Need on põhitüübi suhtes erandid.

Erandlik võib olla kasvõi ainult üks vorm või teatud vormide tüved või teatud vormide formatiivid. Suurem osa erandliku sõna muutevorme moodustatakse aga täiesti regulaarselt. Nt sõna *olema* võib ühendada *tulema*-tüübiga, sest tal on ainult kahes muutevormis teistsugune kaju 'on 'Ind Pr Sg 3' ja 'Ind Pr Pl 3' (vrd 'on – tule/b, tule/vad). Kõik muud vormid on tal sarnased *tulema*-tüübi vastavate vormidega (*ole/n – tule/n, 'olla – t'ull/a, ol'nud – tul'nud* jne). Sõna *rubla* võib viia *kõne-* (*tubli*)-tüübi erandiks, sest tal on ainsuse osastavas täiendav paralleelvorm *rubla e rubla/t*, mida teistel sõnadel pole.

Kuna sõnade jaotamisel morfoloogilistesse tüüpidesse on aluseks mitu erinevat liigitusalust, siis võivad ühe tüübi erandid olla erandlikud mitmest eri aspektist. Nt VVS-i *elama*-tüübi (27) erandite hulka kuuluvad *ajama*, millel on impersonaalvormides erandlik tüvevariant 'ae ('ae/takse jne, vrd *ela/takse* jne), ja *m'uum*a, millel on impersonaalis erandlikud *d*-algulised formatiivivariandid (*m'uu/dud* jne, vrd *ela/tud* jne). Tüübierand on seega veel hägusam mõiste kui reeglierand, sest tüübierandite hulk ja valik sõltub sellest, millised tüübid omavahel optimeerimise käigus ühendatakse (optimeerimata klassifikatsioonil erandeid pole). Avatud morfoloogiamudelis tüübierandeid sellisel kujul ei kasutata. Kõik klassifikatsiooniga seotud erandid kuuluvad vastavate reeglite juurde.

#### 4. REEGLID JA SÖNASTIKUD

Reeglid, mis kirjeldavad morfoloogiasüsteemi seaduspärasusi ja seal toimuvaid protsesse, jagunevad oma liigi ja rakendussfääri alusel mitmeks allsüsteemiks, millest igaüks vajab ka oma sõnastikku. Põhilised allsüsteemid (reeglimoodulid), mis on olulised eesti morfoloogia jaoks, on järgmised: morfotaktika, allotaktika, tüvemuutused ja tüübituvastus (Viks 1994). Neist kaks viimast võimaldavad süsteemi teha avatuks.

Kõige universaalsem morfoloogia allsüsteem on **morfotaktika**. Morfotaktika reeglid määravad ühelt poolt ära kogu keele morfoloogilise süsteemi struktuuri: grammatiliste kategooriate valiku ja realiseerumisvõimalused ning sõnaklassidele vastavad morfoloogilised paradigmad (nt verbi- ja noomeniparadigma). Teiselt poolt määravad morfotaktika reeglid ära sõnavormi sisemise struktuuri: millised üksused millises järjestuses ja mis tingimustel võivad ühes sõnavormis koos esineda (nt noomenil TÕVI + ARV + KÄÄNE).

Morfotaktilist infot esindab morfoloogiasõnastikus sõnaliigi tähis, millele vastavate paradigmade kirjeldused ise on alati väljaspool sõnastikku – grammatikas. Ainuüksi morfotaktika reeglite abil on täiesti võimalik kirjeldada morfoloogiasüsteemi funktsioneerimist aglutinatiivsete keelte puhul, kus on küll rohkesti muutevorme, kuid kus üksused ei varieeru.

Et eesti keeles võivad varieeruda kõik morfoloogilised üksused, nii tüved kui ka formatiivid, siis on vaja kontrollida variantide omavahelist sobivust. Reegleid, mis kirjeldavad üksusevariantide (allomorfide) valiku ja koosinemise tingimusi sõnavormis, võiks nimetada morfotaktika eeskujul **allotaktika** reegliteks (morfotaktika – morfeemide kombinatoorika, allotaktika – allomorfide kombinatoorika).

Kõige ökonoomsem viis allotaktika reeglite esitamiseks on **morfoloogiline klassifikatsioon** (Viks 1977). Puhtal kujul allotaktika reegleid esindab VVS-i (Viks 1992) klassifikatsioon, kus liigitusaluseid on kolm: 1) formatiivvariantide komplekt paradigmas, 2) tüve astmevaheldusmallid, 3) tüve lõpuvaheldusmallid.

Allotaktilist infot esindab sõnastikus tüübinumber, mis juhatab VVS-i grammatikaosas esitatud tüübikirjelduse juurde. **Tüübikirjeldus** kujutab endast paradigma põhivormide esitust sellisel kujul, mis fikseerib kõik tüübile omased variandikombinatsioonid: iga põhivormi jaoks tema tüvevariant üldistatud kujul koos konkreetse

formatiivivariandiga. Iga sõna konkreetset tüvevariandid on antud sõnastikus.

Morfotaktika ja allotaktika reegleid rakendades on juba võimalik eesti keele automaatne morfoloogia tööle panna. Nii on VVS-i baasil realiseeritud EKI varasemad analüüsi- (Hein 1994) ja sünteesiprogrammid (Kuusik 1994). VVS-i reeglistikku (lisatud on liitsõnareeglid) kasutab ka Filosoofi morfoloogiline analüsaator (Kaalep 1996), mis töötab Microsofti Wordi eesti keele spelleris. Kuid morfotaktika ja allotaktika reeglitega piirduv süsteem on veel suletud süsteem, mis suudab sünteesida ja analüüsida ainult neid sõnu, mis on sõnastikus olemas. Iga uus sõna tuleb sõnastikku lisada – ja see nõuab lisatööd inimeselt, kes peab määrama muuttüübi ja sõnaliigi ning moodustama kõik vajalikud tüvevariandid.

Esimene samm avatud süsteemi poole on **tüvemuuatusreeglite** rakendamine (Kuusik 1995, 1996). Tüvemuuatuste reeglid moodustavad mitu erinevat allgrammatikat – vastavalt tüvemuuatuse liigile. Eraldi grammatikad on tüve sise- e astmemuuatuste jaoks, s.t tugeva ja nõrga astme moodustamiseks (mõlemas suunas: tugev → nõrk ja nõrk → tugev), eraldi grammatikad on tüve lõpumuutuste jaoks, s.t lemmatüve ja teiste tüvevariantide moodustamiseks (sammu mõlemas suunas: lemmatüvi → muutetüvi ja muutetüvi → lemmatüvi). Neile põhilistele grammatikatele lisanduvad veel grammatikad mitmusetüvede ja aditiivi geminaattüve moodustamiseks (mõlemad kahes suunas) ja ortograafiateisenduste jaoks.

Tüvemuuatusreeglite lülitamine süsteemi annab võimaluse hoida morfoloogiasõnastikus iga reeglipärase sõna jaoks ühtainsat tüvekuju, sest teisendusreeglid võimaldavad variante omavahel siduda ja ühest tüvevariandist teist moodustada. Sõnastikku peavad jääma ainult erandid – need tüvevariandid, mida allgrammatikates antud reeglite abil ei saa moodustada või mis moodustuksid valesti.

Sel teel väheneb sõnastiku maht oluliselt, sest tüvemuuatustega on seotud  $\frac{3}{4}$  eesti keele sõnavarast ja ühel sõnal on vähemalt kaks, aga sageli kuni viis tüvevarianti. Tüvemuuatuste süsteemi realiseerimine arvutis näitas, et formaalsete reeglitega on võimalik kirjeldada tüvemuuatusi  $\frac{90}{100}$  ulatuses VVS-i sõnavarast. Ainult 10% sõnadest on mingi reegli suhtes erandlikud: kõige rohkem erandeid annavad tüvevokaali lisamise reeglid (allgrammatikas lemmatüvi → muutetüvi) ja nõrgaastmelisest tüvevariandist tugeva

astme moodustamise reeglid (allgrammatikas nõrk aste → tugev aste).

Kuigi kokkuvõtte sõnastiku mahus on tuntav, pole süsteem ainult tüve muutusreeglite abil veel sõnastikust vabaks saanud – sealt on tarvis leida iga sõna jaoks muuttüübi number, mis määrab sobiva tüübikirjelduse ja selle kaudu tüvevariantide kasutamise paradigmas (tüvede vaheldusmallid) ja formatiivvariantide valiku.

Tüübikuuluvuse saab kindlaks teha **tüübituvastuse** reeglitega (Viks 1995b), mis sõna algvormi (lemma) fonoloogilise kuju järgi omistavad talle tüübinumbri. Muuttüübi tuvastamisel arvestatakse sõna selliseid fonoloogilisi omadusi nagu silpide arv sõna algusest või pearõhulisest silbist, välde, lõpuhäälid, sisehäälid, vahel ka silbitüübid. Muuttüübiga koos selgub ka morfoloogiline sõnaklass: verb, noomen või muutumatu sõna.

Tuvastusreeglite lülitamine morfoloogiasüsteemi annab võimaluse piirduda sõnastikus ainult tuvastuseranditega, s.t sõnadega, mille algvormi struktuur ei võimalda antud reeglite abil tüüpi õigesti määrata. Kõige raskem on ära tunda muutumatuid sõnu, millel pole spetsiifilist tuletussufiksit, nt *'all*, *nõnda*, *vähe*, *s'est* jne. Palju erandeid annavad ka 2-silbilised *e*-lõpulised sõnad, mis võivad jaguneda mitme tüübi vahel, nt 01: *t'üüne* : *t'üüne*, 02: *h'omne* : *h'omse*, *raseda* : *raseda*, 04: *tase* : *taseme*, 05: *ranne* : *r'andme*, 06: *vanne* : *v'ande*, 09: *v'aene* : *v'aese*, 16: *kõne* : *kõne*. Kui rakendada tuvastusreegleid ortograafilisele sõnakujule, tekitab suuri raskusi pearõhu leidmine, millest sageli sõltub sõna tüübikuuluvus kõige rohkem, vrd *valang* (rõhk 1. silbil – tüüp 02) ja *volang* (rõhk 2. silbil – tüüp 22). Sõnu, mille algvorm ei sisalda piisavalt infot tüübi määramiseks, on VVS-i sõnade hulgas u 8%.

Tüübituvastuse abil muutub morfoloogiasüsteem avatuks ega sõltu enam sõnastikust. Kogu vajalik info süsteemi tööks tehakse kindlaks sõna enda fonoloogilise kuju põhjal. Kokkuvõttes väheneb sõnastiku osa oluliselt ja grammatika osa suureneb. See aga tähendab, et morfoloogilise sünteesi ja analüüsi tulemuslikkus ei sõltu enam sõnastiku mahust, vaid morfoloogiasüsteemi kirjelduse adekvaatsusest.

Avatud morfoloogiasüsteem funktsioneerib üldjoontes järgmiselt (Viks 1995a). Sünteesi puhul rakenduvad sisendsõnale (lemmale) kõigepealt tuvastusreeglid, mis määravad muuttüübi ja sõnaklassi, seejärel moodustatakse tüübikirjelduses ettenähtud tüveva-

riandid (tüve muutusreeglite abil) ja lisatakse tüübikirjelduses ettenähtud formatiivvariandid. Analüüsi puhul kasutatakse neidsamu reeglimooduleid, kuid teises suunas ja järjekorras. Sõnavormi liigendamisel saadud tüvevariantist moodustatakse (tüve muutusreeglite abil) kõik võimalikud algvormid ja kontrollitakse tuvas-  
tusreeglites, kas nad sobivad vastava tüübi sõnaks.

Igal reeglimoodulil on oma erandite sõnastik, mis vaadatakse läbi enne vastavasse moodulisse sisenemist. Kui mooduli sisendüksus on erandite sõnastikus olemas, siis saadakse sealt ka vajalik väljund ja järgnev reeglimoodul jäetakse vahele. Selline lahendus annab igale reeglimoodulile iseseisvuse ja võimaldab neid kasutada ka iseseisvalt – mõnes teises rakenduses.

## 5. AKTIIVNE – PASSIIVNE MORFOLOOGIA

Avatud morfoloogiamudel on vahetult seotud regulaarse ja irregulaarse ehk aktiivse ja passiivse morfoloogia eristusega, mille tõi eesti morfoloogiakäsitluse Toomas Help (1985, 1990). Aktiivse ja passiivse morfoloogia vastandusele on üles ehitatud ka "Eesti keele grammatika" (1995) morfoloogiapeatükk (autor Kristiina Ross). Tsiteerin (lk 126–127):

"Aktiivne morfoloogia hõlmab seda osa vormimoodustusest, milles reeglid rakenduvad lähtudes sõna algvormi fonoloogilis-derivatiivsest struktuurist, ilma et kõneleja vajaks õige vormi moodustamiseks mingit lisainfot selle sõna kohta. Suurem osa sõnavarast kuulub aktiivsesse morfoloogiasse. Passiivne morfoloogia hõlmab seda osa vormimoodustusest, milles algvorm ei määra üheselt, milliste reeglite abil sõnavorme saab moodustada. Passiivsete reeglite väljakujunemisel on olnud põhjuseks keeleajaloolised fonoloogilised muutused, kuid tänapäeva keele seisukohalt võib passiivseid reegleid pidada leksikaliseerunuks. See tähendab, passiivse morfoloogia alla kuuluvate vormide kasutamisel peab kõnelejal iga konkreetse sõna kohta teada olema, et selle sõna vajaliku vormi moodustamiseks ei saa kasutada tema algvormi struktuurist automaatselt johtuvaid reegleid. Tõenäoselt on passiivse vormimoodustuse korral kõneleja mälus olemas valmis kujul teatud hulk selle sõna vorme, millest kõneleja analoogia abil tuletab ülejäänud vormid. Grammatikakirjelduses on võimalik suhestada passiivse morfoloogia alla kuuluvad vormid algvormiga põhimõtteliselt samasuguste reeglite abil nagu aktiivse morfoloogia korral. Reaalses vormimoodustuses enamik neist reeglitest tõenäoliselt ei toimi, nad lihtsalt kirjeldavad valmis vormide vahelisi seoseid.

Aktiivse ja passiivse morfoloogia eristamise otstarbekus on ilmne. See võimaldab selgelt välja tuua ühelt poolt morfoloogia tuuma: need aktiivsed ja produktiivsed reeglid, mille abil saab va-

balt vorme moodustada ka tundmatutest sõnadest, ja teiselt poolt passiivsed ja ebaproduktiivsed reeglid, mis kirjeldavad seoseid vormide vahel, mida ei moodustata, vaid peetakse meeles.

Piiri asukoht aktiivse ja passiivse morfoloogia vahel ja kumagi esinemisvorm pole aga enam nii ilmselged. Teaduslik grammatika tahab anda kogu morfoloogiasüsteemile ammendavat kirjeldust. Seetõttu esitatakse reeglid kahes osas: aktiivsed reeglid kirjeldavad regulaarseid moodustusseoseid sõnavormide vahel ja passiivsed reeglid kirjeldavad irregulaarseid seoseid.

Avatud morfoloogiamudel arvutis tahab realiseerida morfoloogiasüsteemi kõige ökonoomsemal viisil. Seetõttu esitatakse arvutis ainult aktiivsed reeglid ning kõik passiivse morfoloogiaga seotud sõnavormid esitatakse üksnes erandite sõnastikes. Passiivsete reeglite realiseerimine arvutis on võimalik, kuid see teeks tarbijarakenduste töö liiga aeglaseks. Küll aga on passiivseid reegleid vaja arvutis lingvistilise uurimistöö jaoks, et kontrollida, kas reeglid on õigesti formuleeritud ja kas nad töötavad adekvaatselt.

Ühtlasi selgub sel teel ka optimaalne vahekord reeglite ja erandite vahel. Et erandite hulk ja valik on otseses sõltuvuses reeglite hulgast ja üldistusastmest, siis võib tegelik piir aktiivse ja passiivse morfoloogia vahel erinevates keelekirjeldustes ja rakendustes joosta erinevalt, sõltudes selle otstarbest, kasutatud kriteeriumidest ja tunnustest ning morfoloogiakirjelduse koostaja taotlustest ja arusaamadest.

Milline on aktiivse ja passiivse morfoloogia vahekord inimese peas, on muidugi ainult oletus. Tõenäoliselt ei piirdu inimese keeleoskus reeglite ja erandite teadmise, oma osa on kindlasti ka korduva kasutuse kaudu kinnistunud prototüüpidel, assotsiatiivsetel seostel jne. Kuid sellegipoolest on usutav, et see eristus toimib.

Kahe morfoloogia piiri asukoht keelekasutuses on veelgi määramatum kui keelekirjelduste puhul, sõltudes suurel määral inimese keeleoskuse tasemest. Kui passiivse morfoloogia osa moodustavad selgeksõpitud (omandatud) sõnavormid ning aktiivse osa reeglid, siis algajal keeleõppijal on passiivne osa veel väike ja aktiivne reeglite osa suhteliselt suur – ta peab sageli mõtlema, kuidas üht või teist vormi moodustada. Mida rohkem sõnavorme on keeleõppija omandanud, seda suurem on passiivse morfoloogia osakaal. Reegleid on tal vaja kasutada ainult tundmatute sõnade puhul. Emakeelsel kõnelejal kuulub passiivsesse ossa lisaks erandlikele

sõnavormidele ilmselt ka teatud hulk suure kasutussagedusega sõnavorme, kusjuures samade sõnade vähemkasutatavad vormid moodustab ta tõenäoliselt reeglite abil. Kõigi vormide meelespidamine oleks ebaökoonoomne.

## KIRJANDUS

- Eesti keele grammatika I, 1995, Tallinn.
- Eesti kirjakeele seletussõnaraamat I, 2. vihik, 1991, Tallinn.
- Hein, I. 1994**, Practical Realisation of the Morphological Analysis. – Ü. Viks (ed.), Automatic Morphology of Estonian 1, Tallinn, lk 29–35.
- Help, T. 1985**, Linguistic Competence and Morphology. Preprint KKI-38, Tallinn.
- Help, T. 1990**, Eesti regulaarne ja irregulaarne verb. (Dissertation, MS in the Institute of the Estonian Language). Abstract, Tartu.
- Kaalep, H.-J. 1996**, ESTMORF: A Morphological Analyzer for Estonian – H. Öim (ed.), Estonian in the Changing World, Tartu, lk 43–98.
- Kuusik, E. 1994**, Morphological Synthesis of Estonian Based on the Agglutination Strategy. – Ü. Viks (ed.), Automatic Morphology of Estonian 1, Tallinn, lk 36–48.
- Kuusik, E. 1995**, Automatic Recognition of the Estonian Stem Changes. – Ü. Viks (ed.), Automatic Morphology of Estonian 2, Tallinn, lk 46–71.
- Kuusik, E. 1996**, Eesti tüvemuutuste süsteemi modelleerimine. Magistriväitekirj (käsitöö Eesti Keele Instituudis).
- Muuk, E. 1933**, Väike õigekeelsus-sõnaraamat, Tartu.
- Viks, Ü. 1977**, Klassifikatoorse morfoloogia põhimõtted. Preprint KKI-9, Tallinn.
- Viks, Ü. 1978**, Morfoloogilise klassifikatsiooni optimeerimisest. – Sõnast tekstini, Tallinn, lk 91–111.
- Viks, Ü. 1992**, Väike vormisõnastik I: Sissejuhatus & grammatika. A Concise Morphological Dictionary of Estonian I: Introduction & Grammar. Väike vormisõnastik II: Sõnastik & lisad. A Concise Morphological Dictionary of Estonian II: The Dictionary & Appendices, Tallinn.
- Viks, Ü. 1994**, Eesti keele morfoloogiline analüsaator. Automaatanalüüsi võimalused ja võimatused. – Keel ja Kirjandus nr 3, lk 150–163.
- Viks, Ü. 1995a**, About Rule-oriented Morphology of Estonian. Abstracts of Posters Presented at the 10th Nordic Conference of Computational Linguistics NODALIDA-95, Helsinki 29–30 May 1995, lk 28–30.
- Viks, Ü. 1995b**, Rules for Recognition of Inflection Types. – Ü. Viks (ed.), Automatic Morphology of Estonian 2, Tallinn, lk 23–45.