



## Oma reeglid avastab igaüks ise

Heiki-Jaan Kaalep 60

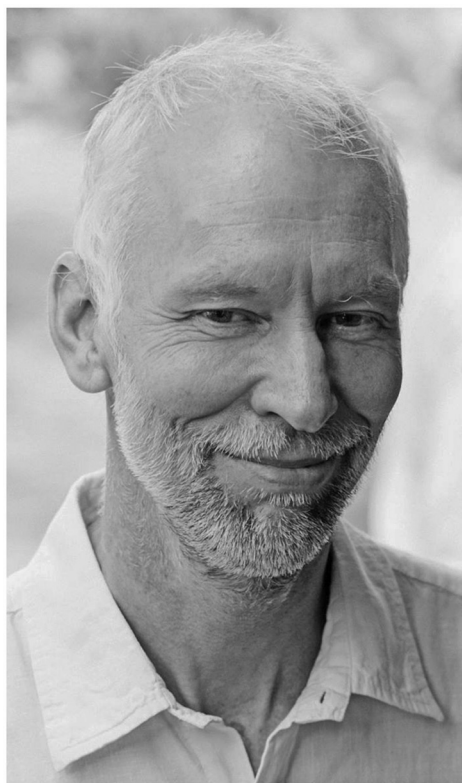


Foto: Cärol Liis Metsla.

*19. mail saab 60-aastaseks Heiki-Jaan Kaalep, kes keelerahva hulgas on eelkõige tuntud kui eesti morfoanalüüsi ja spelleri tegija ning ebamugavate küsimuste esitaja morfoloogia teooria teemal.*

### **Kuidas sa morfoloogia juurde sattusid?**

Minu tulemine keele juurde oli praktiline. Ülikooli diplomil on kirjas majandus-

teadlane-matemaatik, seejärel tegin informaatika magistri ja üldkeeleteaduse doktori.

Kui ma hakkasin programmeerimist õppima, siis selle võlu mu jaoks oli, et teed ühte asja, kirjutad mingit programmi-teksti, ja arvuti hakkab selle peale tegema midagi hoopis muud. Selline lapsik võlu, et vajutad lülitit ja hoopis teises kohas läheb lamp põlema. Sedasama lapsikut võlu üritasin siis kasutada keele töötlemisel, kus jõudsin mõtteni, et võiks ju proovida teha eesti keele spellerit. Ülle Viksi „Väike vormisõnastik“ oli olemas. Selle alusel tahtsin kontrollida, kas tekstis olev sõnavorm kuulub võimalike vormide loendisse või mitte, ehk panna puhtast lustist arvuti tegema midagi, mis lisaks võiks kasulik olla. Kohe aga selgus, et paljusid tekstides esinevaid sõnu, eriti tuletisi ja liitsõnu, sõnastikes ei ole. Kui tahtsin oma programmi ka neid ära tundma panna ja vaatasin selleks olemasolevaid keelekirjeldusi, siis avastasin, et needki ei andnud mulle vajalikku sõnavara katvust.

Nii tekkis mul huvi, et mis need päris reeglid on, mille järgi inimesed sõnu käänavad-pööravad, ja kuidas need erinevad lingvistikaraamatutes kirjas olevatest reeglitest. Olen olnud sunnitud oma reegleid ise avastama, sest keegi teine pole neid minu jaoks avastanud.

Ühel hetkel märkasin, et olen mõned oma reeglid valesti kirjutanud. Sattusin sellisesse tsükklisse, et lasen oma programmi päris tekstidest üle, ja tulemuse põhjal modifitseerin oma programmi. Nägin, kuidas ma ei suuda õigesti välja mõelda asju. Igale programmeerijale on selline olukord paras üllatus, sest programmeerimises üldiselt on ülesanne teada ja tuleb luua algoritm selle ülesande lahendamiseks.

damiseks. Mina ei teadnud oma ülesannet ise ja keegi ei öelnud mulle ka. Ainus võimalus oli oma programmi aeg-ajalt ümber kirjutada, vastavalt sellele, kuidas ma nägin teda valesti töötamas. Sellest tekkis vajadus ehitada oma programm niimoodi üles, et saaks reegleid kergesti muuta.

See teema on mind tõesti endasse imeanud. On kujunenud nii, et kui lahendan ühe asja ära, tekib kohe järgmine küsimus, millele ma ei oska vastata.

### **Ometi oli eesti morfoloogiat ju uuritud varemgi.**

Meelde tuleb Enn Tõugu kunagine analoogia füüsika ajaloost: termodünaamika seadused formuleeriti alles pärast aurumasina ehitamist. Aurumasina tegid insenerid, seda ei tehtud termodünaamika seaduste järgi. Ka lingvistidel polnud võimalik konstrueerida neid väiteid, mida mul vaja oli, kuni neil polnud massilisi tekstitöötlemise vahendeid.

Klassikalistel lingvistidel, kellel pole olnud suurte tekstimasside masinaga töötlemise võimalust, on olnud kas üks käsi selja taha seotud või lausa mõlemad, sest inimese vaimsed võimed, huvid ja tähelepanu on piiratud. Materjali on nii palju, et seda läbi lugeda ei jõua. Huvitavate asjade leidmiseks peab looma mingid filtrid. Neid filtreid ei saa aga keegi teine teha, sest ainult teadlane ise teab, mis teda huvitab, või alustades ei tea ta seda ka ise. Alles tulemusi vaadates saad vähehaaval aru, mida tegelikult otsid.

Nüüd oleme hoopis teistsuguses tehnoloogilises olukorras ja saame vaadata keelele teistmoodi.

### **Kuidas käib reeglite avastamine?**

Õnneks oli tekstikorpusti juba tehtud ja mul oli võimalik neid kasutada. Lasin ennast juhtida autentsest materjalist ja kirjutasin selle põhjal reegleid, mis võimaldasid korrektseid sõnu ära tunda.

Keele kõnelejal on tegelikult sama ülesanne. Ikka tuleb ette uusi sõnu, mille kasutust me pole varem näinud, aga mille käänamist keel nõuab. Samas on eesti keel selline, et erinevaid sõnu käänatakse erinevalt. Me peame kuskilt võtma informatsiooni, et kuidas seda sõna käänatakse, ja ei ole mitte mingit muud informatsiooni peale sõna kuju.

Wolfgang Wurzel ütles juba 1980-ndatel, et tuleb aru saada, et mõned sõnad ongi erandlikud. Neid muudetakse veidralt, aga nad on samal ajal sagedased, mistõttu neid mäletatakse ja nad on tõenäoliselt ka sõnastikus antud. Kõik ülejäänud sõnad käivad aga üldreegli järgi, kusjuures erinevate häälikulistite kujude jaoks on keeles erinevad reeglid.

Need reeglid loob enda jaoks iga inimene ise selle järgi, kuidas on mõnus sõnu grupeerida mingite ühiste tunnuste alusel. Praegu on olemas ka teooria selle kohta, kuidas nende sisemiste reeglite moodustamine käib. Charles Yang on esitanud küsimuse, kui palju peab sarnaselt käituvaid sõnu olema, et inimene mõtleks välja üldreegli nende kohta. Kui meile esitatakse kümme sõna, neist üheksa käänduvad ühtmoodi ja üks teistmoodi, siis tõenäoliselt õpime selle ühe ära, ja ülejäänud üheksa jaoks mõtleme nende sarnasuse alusel välja mingi reegli. Kui aga on kuus ühtmoodi ja neli teistmoodi, siis me suure tõenäosusega ei mõtle seda reeglit välja, vaid jätame kõik kümme sõna meelde, sest me ei ole enam kindlad, kas julgeksime oma reeglit üheteistkümnendale sõnale rakendada.

Üldreeglid ja erandite loendid on olemas kõigis keeltes. Järgmise sammuna saab uurida, millisel juhul juhtub nii, et inimene ei teagi, millist reeglit kasutada. Üksikutes keeltes on teada selliseid juhtumeid, kus inimesed on teatud tüüpi sõnade teatud vormide moodustamisel suurtes raskustes.

**Kuidas on võimalik, et keele kõnelejad mõtlevad välja lahenduse, millega neil endil raskusi tekib? Kas keel iseorganiseeruva süsteemina ei peaks seda ära hoidma?**

Just, keel organiseerub ja kasvab ise ja kaootiliselt. Nagu põõsas, kus ikka juhtub, et kaks oksa kasvavad risti. See on täiesti mõttetu, neil ei ole tarvis risti kasvada, aga ikkagi juhtub nii. Ja kui on juhtunud, siis põõsas kuidagi lahendab selle olukorra, näiteks nii, et üks okstest kärub.

Ühest küljest ei tohiks neid juhtumeid üldse olla. Teisest küljest, kui selline ristikasvamine on võimalik, siis võiks eesti keele taolises flekterivas ja muuttüüpi-dega keeles ju kõik kohad neid juhtumeid täis olla. Kui nii vaadata, on neid jälle üllatavalt vähe. See on lingvisti jaoks huvitav olukord: miks tekivad raskused just neis kohtades, ja kuidas inimesed neid lahendavad? Igal juhul võtab see lahendamine aega, ja me saame jälgida vahepealseid ebamugavaid olukordi.

**On lihtne näha, kuidas reeglite avastamine on lingvistile või keeletarkvara arendajale huvitav. Kas need avastatud reeglid on ka keele kõnelejale kuidagi kasulikud?**

Ka lingviste juhhib nende töös üldine ettekujutus, et keerukad nähtused on mingi väikese hulga lihtsate reeglite koosmõju tulemus. Küll oleks tore, kui leiaks need väikesed reeglid. Mõni on olemas ka, näiteks kuidas on eesti keeles tekkinud nii keeruline teisendus nagu *signa* : *sea*. See on tagasi viidav kahele lihtsale reeglile: *g* kaob, ja kui *g* kaob, siis kõrged vokaalid madalduvad. Iseasi on, et see keeleajalooline seletus ei ole eriti abiks keeleõppijale, sest selle taseme reegleid rääkides mees pidada on liiga keeruline.

Lingvist küll rõõmustab, et on avastanud tõelise seaduspära, millised on põhimõtteliselt keelte arenemise suunad, ja

milline on ka eesti keele tulevik. Kui seda liiki reegleid tunda, siis on võimalik natukenegi tulevikku ennustada. See uurimistöö ei anna keele õppimisele või kasutamisele suurt midagi, aga annab parema arusaamise keelest kui sellisest. Tänapäevalgi, kui *ühiksa* puhul jääb kõnes *h* hääldamata, siis *ü* madaldub ja saamegi kõnekeelse *öeksa*.

**Oled kirjutanud nii verbi kui ka noomeni algvormi valikust. Kas see valik mõjutab ainult keeleteooriat või ka kõnelejaid?**

Algvormi valik on näide sellest, kuidas eksitav lingvistiline väide võib tekitada keele kõnelejatele raskusi. Eriti meeldib mulle *ma*-tegevusnime pidamine verbi algvormiks. Need, kes pakkusid seda vormi sõnastikes märksõnana kasutada, teadsid ise, et see ei ole algvorm. Nad panid selle sinna pedagoogilistel kaalutlustel, et *ma*-tegevusnimest kui tugevastmelisest vormist oleks lihtsam moodustada teisi vorme. Kuna pärast seda on kõik sõnastikud sama traditsiooni järgi tehtud, on inimesed hakanud arvama, et see ongi algvorm. See tegelikkuses ei tööta, sest *ma*-tegevusnime vorm on nii haruldane, et seda pole võimalik õppimisel aluseks võtta, sest inimene pole teda tekstides näinud.

**Mida morfoloogia alal praegu teed?**

Morfoloogiatarkvara loomist alustasin üldotstarbelises programmeerimiskeeles, millega saaks kirjeldada mida iganes. Praegu püüan kasutada sellist matemaatilist abstraktsiooni nagu lõplikud muundurid. Neid on kõige parem endale ette kujutada nii, et muunduril on olekud, mille vahel sisendsümbolite mõjul toimuvad üleminekud, kusjuures emiteeritakse ka väljundsümbolid. Muundurit läbides teiseb üks sümboljada teiseks. Üleminek sihtolekusse sõltub ainult lähteolekust ja sisendsümbolist, mitte sellest, millist rada

pidi lähteolekusse satuti. Seega ei ole lõpliku muunduri olekul üldse mälu.

Tavalises programmis saan kirjutada ükskõik kui keerulisi reegleid, minna tagasi, alustada uuesti, mäletada eelnevat, mis jätab varju teatud keele omadused. Muundurid on jälle omaette viis keelt vaadata, mis sunnib nähtuste kirjeldamise peale rangemalt mõtlema.

Muundurite lihtsus on selles mõttes võluv, et ei ole ju üldse ilmne, et morfoloogia põhineb nii lihtsal mehhanismil. Kas tõesti õnnestub kogu morfoloogia ära kirjeldada ilma mälu kasutamata? See ju omakorda näitaks, kui lihtne meie morfoloogia tegelikult on. Muundurite teine võluv omadus on, et neid saab vaadelda matemaatiliste objektidena, millele on võimalik luua muundurite algebra ehk teha tehteid, näiteks kaks muundurit kokku panna või ühest muundurist teist lahutada. Seda algebrat kasutades ma püüangi eesti morfoloogiat kirjeldada.

On terve hulk keelenähtusi, mida on väga tülikas kirjeldada tervikuna, kuid muundurite algebra kaudu on lihtne näha, milliste operatsioonidega on võimalik moodustada keerukaid vorme. Sama mehhanismi abil saab kirjeldada nii leksikoni kui ka teisendusreegleid. Kui jätkame juba kasutatud näitega, siis leksikonis saab loetleda selliseid sõnu nagu *sig*, *kägu*, *här* jt, ja kirjeldada teisendusreeglit, et *g* neis kaob. Me ei ütle kuskil, et *sig*a omastav on *sea*. On ainult üks vorm *sig*a ja kõik ülejäänud tuleb reeglitest. Ma ei väida, et täpselt sellised reeglid on inimesel peas. Mis inimese peas toimub, seda me ju tegelikult ei tea. Saan ainult öelda, et minu valitud formalismis on morfoloogiat kõige parem nii kirjeldada.

### **Kui kaugel selle tööga oled? Mis on lõppesmärk, kas uus ja parem speller?**

Muunduritel põhinev tehnoloogia on selles mõttes hea, et uue keele jaoks on

vaja teha ainult muundurid. Kogu tarkvara nende rakendamiseks näiteks teksti kontrollimisel on keelest sõltumatu. Ka speller on seega juba olemas, ilma et oleksin pidanud ise sellega jändama. Speller on praegu sama hea kui senised, sest olen oma reeglid lihtsalt uude formalismi ümber kirjutanud. Eesmärk on olnud muundurite abil parandada just sõnastikust puudevate sõnade kirjeldust, mis teoreetilise morfoloogia seisukohast on kõige huvitavam: kuidas muudavad inimesed neid sõnu, mida nad pole varem näinud ega saa võtta mälust, vaid peavad mingeid reegleid rakendama.

See on osutunud arvatust raskemaks. Olen küll oma artiklites kirjeldanud, kuidas see võiks käia mu seniste programmide põhjal, aga kui püüan neid piiratumas formalismis taasesitada ja sealjuures suuremat täpsust saavutada, siis olen komistanud probleemide otsa. Üks probleem on sõna rõhu määramine kirja pildi põhjal, kui sarnase häälikulise koostisega sõnu, näiteks *hevea* ja *korea* käänatakse erineva rõhu tõttu erinevalt: *heveaid*, aga *koreasid*. Või kas sõna *solaris* rõhk on esimesel või teisel silbil?

### **Siin tuleb mängu võõrsõnade muganemine, nagu *sümpptom* ja *sümpptom*.**

See on väga hea näide, sest see seisukoht, et eesti keeles sõnade muganemisel nihkub rõhk esimesele silbile, vajaks empiirilist kontrollimist. Võib-olla see tõesti kehtis sada aastat tagasi, aga ei saa olla kindel, et enam kehtib. Keel ikkagi muutub päris kiiresti.

### **Kas vältega on sama teema, näiteks sõna *teemant* osastavas käändes?**

Ma pole kindel, et asi on vältes. See on üks koht, kus väited eesti keele kohta on imelikud, et mitte öelda vastuolulised või valed. Ühest küljest öeldakse, et valde on takti omadus, ehk tähtis on kahe silbi järjend. Kuidas saame siis rääkida sõna

*teemant* teise silbi vältest? Seal ainult kaks silpi ongi, millele ei järgne midagi. Mulle tundub, et „faktid ei sega meil arvamast”. Midagi on meie väidetes väga viltu. On näha, et siin on midagi imelikku, mida meie teooriad hästi ei seleta. Mingi parameeter keeletehnoloogilise intuitsioonis seda valikut ju mõjutab, kas osastav on *teemandit* või *teemanti*, aga ma ei tea, mis see on. Ja ma ei tea, et keegi teaks.

**Oskus eriti suulises kõnes käänata *teemanti* võiks ju olla parameeter, millega näidata enda kuulumist haritud seltskonda.**

Ka see on lahtine küsimus, millised koolis õpitud sõnadest on rahva hulgas olulised näited ja millised mitte. Võib-olla osastav *teemanti* siiski ei ole oluline, sest neid, kes pole seda ära õppinud, on nii palju, et selle vastu eksimist ei panda tähelegi.

**Kas kõige rohkem pannakse tähele sõnavalikut?**

Sõnavalikut pannakse isegi vähem tähele. Eraldiseisvad paronüümipaarid, nagu *järgi-järele* ja *õieti-õigesti*, on nii raskesti õpitavad, et pooled inimesed ei tea, kumb neist peaks haritust näitama. Reeglite rikkumisi – nagu *autosi* ja *tegematta* – märgatakse ikka rohkem. Need kuuluvad kooliõpetuse morfoloogiasse, sest on hulk murdeid, kus *d* jääbki mitmuse osastava lõpust ära, samuti ei ole kõnes sageli kuulda, kas ta seal on või mitte. Keele kirjeldajana ma ei saaks öelda, et *d* puudumine vale on. *Tegematta* ja *tegemata* vahel pole häälduses üldse mingit vahet. Kirjapildi erinevus on aga tõesti asi, mis koolis ära õpitakse ja mida kõik tähele panevad.

**Kuidas saaks ühtse konventsiooni kujunemisele kaasa aidata?**

Otseselt ei olegi vaja kaasa aidata. Keele kõneleajate hulgas on n-ö korrektne keele-

kasutus niigi väärtustatud. Kui interneti-foorumites kasutatakse norminguvastaseid vorme, siis on väga suur tõenäosus, et sind ei võeta tõsiselt. Su argumente ei usuta, sest ortograafiavigade tegemine jätab mulje, et sa pole piisavalt haridust saanud. Mitte et sa oleks emakeeletunnis reegleid õppinud, vaid piisavalt eestikeelset materjali lugedes oleks kõik need reeglid sulle ise külge jäänud. Sa ei saakski neid vigu teha, sest oled õigete vormidega nii ära harjunud.

Näen ka spontaanset survet õigekirja järgida. Paljudes foorumites on vabatahtlikud keeletehnikud, kes parandavad tagantjärele postitusi, ja ka inimesed ise parandavad oma postitusi aeg-ajalt. Osutub, et on oluline jätta endast mulje kui haritud inimesest.

**Oled lisaks ülikoolile tegutsenud ka erasektoris: su Filosoft oligi vast esimene keeletehnoloogia-ettevõtte Eestis. Kuidas neid kaht sektorit võrdled?**

Spellerit ja morfoanalüsaatorit tegin tõesti eraettevõttes. Aga nüüd on need antud vabavaraliseks, mina olen ülikoolis ja konkureerin oma varasema tööga. See on seotud sellega, et morfoloogiaalased lahendamata küsimused osutusid nii huvitavaks, et praktilise töövahendi tegemine ei motiveeri enam. Erasektor ei maksaks palka lihtsalt huvitava küsimuse uurimise eest, eriti kui uurid aasta aega ja siis selgub, et ei tulnud midagi välja.

**Kuidas on morfoloogiatarkvara andmine vabavaraliseks tagantjärele vaadates välja tulnud?**

Mitte nii hästi, kui oleks võinud loota. Eesmärk oli, et ilmuks inimesed, kes seda täiustavad, aga seda ei ole juhtunud. Võib-olla on morfoloogiatarkvara niisuguseks täiendamiseks sisemiselt liiga keeruline. Selle poolest on vabavara küll hea, et iga soovija ei pea selle kasutamiseks enam luba küsima.

**Tasuta asju oled loonud veel. Kas Keeleveeb kui tasuta alternatiiv tekkima hakavatele tasulistele sõnastikuportaale ole maailmaparandus?**

Oletasin, et internetis väljas olevate asjade eest raha küsimist pole mõtet proovida. Kuna ma ise ei taha maksta, küllap siis ei taha teised ka. Pigem oli selline idealistlik seisukoht, et kui teha hea asi ja seda tasuta jagada, küll siis tuleb kunagi hiljem sellele kuskilt ka raha. Ja see on tõeks läinud. Meil on olnud projekte, kus oleme lubanud tulemuse Keeleveebis tasuta avaldada, ja see on tõenäoliselt olnud üks põhjusi, miks need projektid on rahastuse saanud.

Keeleveebi ideoloogia üks oluline külg on, et sealsed sõnastikud on igäüks omaette. Ei ole proovitud teha üht supersõnastikku nagu EKI Sõnaveeb praegu, sest me vaatame kohe alguses, et see oleks väga raske. Iga uue sõnastiku lisamisel pead varasema ümber tegema, mida me ei jõua teha. Keeleveeb loodi suhteliselt kergesti ülalpeetavaks lahenduseks.

Asi algab veel kaugemalt: mis on need üksused, mida sa sõnastikku paned. Sageli ei ole need sõnad, vaid mingid suuremad üksused. Iga sõnastikutegija defineerib need ise oma vajadustest lähtudes, ja mingit õiget vastust me ei tea. Me ei näe ka läbi, kuidas kirjutatud keel mõjutab meie vaadet keelele. On ju sõnade, kollokatsioonide ja muude üksuste erinev käsitus tingitud paljuski sellest, et kirjutame häälikuliselt ja paneme tühikuid vahele. Kui meil oleksid hieroglüüfid, siis oleksid ka probleemid hoopis teised. Sõna mõiste tekib alles siis, kui hakatakse midagi üles kirjutama.

**Sõna definitsiooni mõjutab ju ka tähendus. Võtame ühe hiljutise näite. Kas sõna liiderlik tähenduses 'juhilik' on olemas-olevale sõnale antud uus tähendus või reeglipäraselt tuletatud uus sõna?**

Minu jaoks on kõige huvitavam, millise kriteeriumi järgi otsustada, kellel on õigus. Siin ma ei oska isegi kandidaate pakkuda. Võib-olla see polegi keeleteaduse küsimus, kas sõnale on antud uus tähendus või tuletatud uus sõna? Minu intuitsioon ütleb, et pigem tuletatud, sest sarnasus sõnaga *liider* on nii suur. Samas on teada piirang, et mõnda reeglipäraselt tuletist ei saa moodustada, kui tuletisega samakujuline ja samasse valdkonda kuuluva tähendusega sõna on juhuslikult juba olemas. Homonüümid iseenesest ei ole probleem, kui tähendused on piisavalt erinevad: näiteks sõna *laik* hiljutist laenamist ei seganud palju vanema samakujulise laenu olemasolu üldse, sest ta on piisavalt harv ja piisavalt teises valdkonnas. Aga näiteks *tus*-liiaga sõnad, nagu *huuletus* või *pilvitus* – me tegelikult ikka ei ütle *huuletus* luule puudumise kohta, see on ainult mängult.

Kriteeriumi otsimiseks tuleks vaadata sarnaste probleemidega sõnade gruppe, kust saaks teada sõnade üldise käitumise: millistest kriteeriumidest oleneb sõna omaksvõtmine või kinnistumine? Siin on kõvasti mittetriviaalset lingvistilist uurimistööd. Korpuses pole just palju olukordi, kus pakutava tuletisega samakujuline laen on olemas, aga harv ja võib-olla isegi ajalooline, samas kui tuletusalus on praegu populaarne. See oleks huvitav uurimisteema.

**Aga korpuses on ju igasugust teksti, kaasa arvatud vigast.**

Alati leidub igasugust materjali, ka vigast. Korpuse juurde ei saa minna, tehes nägu, et see on absoluutne tõde. Sa lähed sinna ikka oma teoreetiliste konstruktidega. Oma artiklis „Verbi algvorm ja teised vormid”<sup>1</sup> olen näidanud, kuidas verbivormide erinev sagedus eri korpustes

<sup>1</sup> H-J. Kaalep, Verbi algvorm ja teised vormid. – Keel ja Kirjandus 2021, nr 6, lk 510–531.

annab just uuritavas küsimuses erinevaid tulemusi. Minu väide on, et need korpused, kus lõputa vorm ei ole kõige sagedasem, ongi valed korpused. Andmeid mitte usaldada tundub väga ohtlik, aga mul on selleks olemas teoreetiline põhjendus: korpus on vale selles mõttes, et ta esindab keelt, mis just minu uurimisküsimuse mõttes on kallutatud.

Siin on jälle analoogia füüsika ajalooga: kuidas Galileo Galilei leidis, et kõik kehad kukuvad sama kiiresti, sõltumata nende massist. Kui meie andmed koosneks katsetest, kus viskame sulge ja kivi tornist alla, siis see korpus näitaks, et kivi kukub kiiremini. See tulemus oleks aga õhutakistuse võrra kallutatud. Galileo loomulikult vaakumit ei saavutanud, aga saavutas selle, et hõredamas õhus kiiruste erinevus vähenes, millest ta sai teha oma järeldused.

Korpus peame samamoodi mõistusega vaatama, ka seal on asjaolusid, mis meie jaoks on müra, nagu Galileo katses oli atmosfääri tihedus. Muidugi on oht nimetada müraks midagi, mis on tegelikult üksnes minu teooria jaoks ebamugavad andmed. Siin ongi argumenteerimise koht.

Samuti tuleb siit üks erinevus humanitaar- ja reaalaralade vahel. Reaalala teadlastel on küsimus teada, humanitaaridel mitte. Humanitaarid alustavad sellest, et leiavad ühe huvitava nähtuse, ja kõigepealt lihtsalt kirjeldavad seda, küsimused tekivad alles selle järel. Paljud lingvistid lähenevad ka korpusuuringutele nagu humanitaarid – kasutavad küll reaalarala vahendeid, aga lähenemisviis on jätkuvalt humanitaarne, nad ainult kirjeldavad.

Minu jaoks on korpusete lappamine ja asjade kokkulugemine triviaalne ülesanne, ilma milleta ei saa. Aga seda võib teha igaüks, ja kui kirjeldusest kaugemale ei jõuta, siis ei ole huvitav sellist artiklit lugeda.

Järgmine vajalik samm oleks katsuda seletada, kuidas leitu on seotud teo-

riaga. Kas need andmed on mõne senise teooria või mudeliga kooskõlas, vastuolus või mudeli suhtes invariantseid, mis oleks samuti huvitav tulemus.

### **Kuidas suhtud humanitaaridele statistika ja muude nn digiainetete õpetamisega?**

Arvutilingvistika eriala lõpetajad saavad juba praegu tööle arvutifirmadesse. Seal on vaja inimesi, kellel on teoreetiline ettekujutus keele kui süsteemi toimimisest ja kes samal ajal oskavad natukenegi programmeerida. Teisest küljest, kui tahta teha keeleteadust, siis tuleb natukenegi suuta korpusi läbi hekseldada ja tulemustele lihtsamat statistikat rakendada. Kui üliõpilane natukestki matemaatikat ega statistikat õppida ei taha, tekib mul mure, mida ta siis üldse saab õppida. Üks võimalus on tahta saada õpetajaks, kes räägib lastele eesti keele reeglitest. See on põhimõtteliselt aksepteeritav, aga siis peab need reeglid tema jaoks leidma keegi teine. Ka tal oleks kasu natukesest statistikast ja programmeerimisest, mis annaks võimaluse põhjendatult vaadata, mis reegleid ta õpetab või toimetamisel rakendab. Mis on nende taust, kust nad tulevad või kas nad päriselt vastavad sellele, kuidas inimesed tegelikult keelt kasutavad?

Keeleainetes oleks vaja rohkem rääkida, kuidas oleme jõudnud nende väidetele, mis meil praegu kehtivad, kaasa arvatud praeguse ortograafiani. Rääkida ka nendest ettepanekutest, mis ei ole läbi läinud, või kohtadest, kus oleks saanud otsustada teisiti.

Meie ortograafia on ühest küljest intuiitiivne, aga teisest küljest hääldusest erinev. Näiteks kaashäälikuühendis kirjutatakse iga häälik ühe tähega. See on nii kergesti omandatav, et me isegi ei märka, et see on ebahäälduspärane. On proovitud teha teistsuguseid reegleid, näiteks saami keelte skandinaaviapärase ortograafia, ja inimesed ei ole suutnud neid ära õppida. See

on väga õpetlik, miks nad pole suutnud. Lisaks on huvitav, kuidas mittehäälduspärane kirjaviis ei ole meid üldse seganud ega meie hääldust mõjutanud.

**Kas keele muutumine ohustab eesti keelt?**

Ainuke oht eesti keelele on rääkijate vähesus. Kui rääkijaid oleks rohkem, siis rääkigu nad millist eesti keelt tahes, see on ikkagi eesti keel. Isegi see, kui räägitakse

halba eesti keelt, ei ole eesti keelele kuidagi ohtlik.

**Aitäh intervjuu eest! Sinuga on alati tore lobiseda.**

Tore, kui saame lobiseda. Ukrainas oleks meievanused mehed praegu rindel või surnud.

ARVI TAVAST