

EESTI TEADUSTE AKADEEMIA JA EESTI KIRJANIKE LIIDU AJAKIRI

## TEOREETILINE KEELETEADUS JA KVANTITATIIVSED MEETODID

ANN VEISMAN, JANE KLAVAN, HALDUR ÕIM

**N**ovembris 2017 toimus Tartu Ülikoolis V konverents sarjast „Teoreetiline keeleteadus Eestis” (23.–24. XI 2017).<sup>1</sup> Konverentsi pealkirjaks sai seekord „Kvantitatiivne pööre keeleteaduses – mis saab teooriast?”. Selline pealkiri oli ajendatud viimasel ajal keeleteaduses toimunud ulatuslikust suundumusest kvantitatiivsete meetodite poole. Tavaliseks on muutumas suuremahulised korpusuuringud koos keeruliste statistiliste mudelitega, üha enam püütakse teha eksperimente, et selgitada välja tegelik keekekasutus vastandina keeueurija enesevaatlusele. Seejuures on võimalik uurimisküsimusi esitada ja saadud andmeid tõlgendada mitmetest teoreetilistest raamistikest lähtudes. Mõned küsimused, millele soovisime konverentsil vastuseid otsida: kas keeleteaduslik mõte areneb tänu kvantitatiivsele pöördle või sellest hoolimata? Kas radikaalne empiiriline lähenemine võib piirata uurimisküsimuste esitamist? Kas mõni osa keelest jääb pimedale alale, sest puudub võimalus empiirilise materjali kogumiseks, või toob rangelt empiiriline lähenemine välja hoopis uusi tahke keeueuurimises ja -teooriates? Ning viimaks: kas on võimalik keeleteaduslik uurimus kui deskriptsioon ilma teoreetilise raamita? Soovisime arutleda nende küsimuste üle võimalikult erinevatest vaatenurkadest, oodatud olid nii kitsamad kui ka üldisemad lähenemised teemale. Konverentsi kutsutud esinejad olid professor Martin Ehala (Tartu Ülikool ja Helsingi ülikool), Tartu Ülikooli lektor, vanemteadur

<sup>1</sup> Alates 2001. aastast on konverentsid toimunud iga nelja aasta tagant, vt ka kodulehte: <http://teoreetilinekeeleteadus.ut.ee/> (22. VIII 2018).

Virve-Anneli Vihman ning Tartu Ülikooli humanitaarvaldkonna prodekaan, dotsent Liina Lindström.

Keele ja Kirjanduse teemanumbrisse oleme koondanud konverentsi-ettekannetest valminud artiklid. Artiklite autorid esindavad keeleteaduse eri valdkondi ja suundumusi ning seetõttu on ka vastused üleskutses esitatud küsimustele äärmiselt eriilmelised. Selline mitmekesine pilt on iseloomulik tänapäeva keeleteadusele üldiselt, nii Eestis kui ka mujal maailmas.

## 1. Milles küsimus?

Tuleb tunnistada, et konverentsi pealkiri on teadlikult mõnevõrra provokaatiivne, kuid see on siiski õigustatud. On fakt, et varasemaga võrreldes on viimase kümnekonna aasta jooksul silmanähtavalt kasvanud selliste publikatsioonide, sealhulgas väitekirjade arv, mille uurimismaterjal on hangitud korpustest või katseliste meetoditega ning tulemuste töötlemiseks on kasutatud statistilisi või muid kvantitatiivseid meetodeid. See tähendab meetodeid, mille tulemused on arvuliselt väljendatavad ja mille põhjal on võimalik teha järeldusi uuritavate keelenähtuste kohta, aga ka luua keeletöötlusvahendeid, mis ei pruugi huvi pakkuda ainult keeleteadlastele – nt andmekaeve (ingl *data mining*) programme, mida saab kasutada tekstidest sisuliste andmete hankimiseks mitmesugusel eesmärgil. Iseenesest on see loomulik areng, sest ühe korraliku teaduse tulemused peavad olema kui just mitte mõõdetavad, siis igatahes kontrollitavad. Inglise keeles tehakse seni vahet, mis on *science* ja mis *philosophy*.

Keeleteadus on läbi aegade ammutanud metodoloogiat mõlemast allikast. Seda nii valitsevate suundumuste veendumuste tõttu (milliseid keele aspekte ja kuidas tuleks uurida) kui ka keele kui uurimismaterjali mitmekesise olemuse tõttu (foneetika uurimisala on palju selgemalt mõõdetav ja kontrollitav ja seega loodusteadustele lähemal kui näiteks semantika). Keeleteaduse kahe-tisele olemusele (keel kui loodusnähtus ja keel kui inimlik, sotsiaalne nähtus) sekundeerib laiemas plaanis empirismi ja ratsionalismi vastasseis. Empirismi *vs.* ratsionalismi rollile keeleteaduse vooludes on juhtinud tähelepanu mitmed kognitiivsed keeleteadlased (vt nt Geeraerts 1993), kes toovad välja kognitiivse keeleteaduse toetumise empirismile, vastandina generatiivsele paradigmale, mille aluspõhimõtted sobivad pigem ratsionalismiga. Geoffrey Sampson<sup>2</sup> (2013) rõhutab empirismi kahte külge seoses keeleteadusega: nn deskriptiivset ja normatiivset.

Deskriptiivse empirismi järgi on teadmiste omandamine seotud kogemustega. Sellele vastandub keeleteaduses nativistlik vaade, mille järgi kaasasündivuse osa on keele omandamisel olulisem kui omandatav kogemus (nt generatiivse keeleteaduse vaimus on sel teemal üks tuntumaid töid Steven Pinkeri raamat „Keeleinstinkt” („The Language Instinct”, 1994)). Nativismi rollist lapsekeeleuuringute arengus kirjutab siinses teemanumbris põhjalikumalt Virve-Anneli Vihman. Normatiivne empirism on seotud keele uurimisega, empiristlikke vaateid jagav keeleteadlane ammutab ka ise teadmise kogemusest ehk keelematerjalist, mis on jagatav teiste keeleteadlaste ja

<sup>2</sup> Oluline on siin märkida, et Sampson ei esinda kognitiivseid ega generatiivseid keeleteadlasi.

uurijatega (ning seega taasesitatav ja kontrollitav). Ratsionalismile toetuva (normatiivset empirismi mitteoluliseks pidava) keeleteadlase jaoks on keeleteadlane ise sama hea andmeallikas kui kes tahes teine emakeelne kõneleja. Nagu Sampson möönab, ei välista deskriptiivse empirismi eitamine (nativism) normatiivse empirismi järgimist (Sampson 2013: 282). Ka Mari Aigro toob oma siinses artiklis välja, et generatiivne lähenemine ei välista kvantitatiivseid meetodeid. Kuid Sampson näitab ka veenvalt (analüüsidest ajakirja *Linguistics* kaastöid), et keeleteaduses on just deskriptiivse empirismi laienemine kaasa toonud järjepideva ja olulise kasvu normatiivse empirismi järgijate hulgas (Sampson 2013).

Empirismi ja ratsionalismi rolli arvutilingvistika arengus on suurepäraselt kokku võtnud Mare Koit (2006). Ta näitab, et reeglipärane (ratsionalistlik) keeletöötlus on mõttekas n-õ väikeste mängusüsteemide loomisel ja siis, kui napib ressursi andmetöötluks (andmete hulk, arvutustehnika võimsus). Väikeste mudelite eelis keeleteadlasele on nende hoomatavus ja selgus. Andmepõhiste keelemudelite eelis on Koidu järgi efektiivsus, need suudavad modelleerida keelelisi nähtusi, mille kohta selge arusaam puudub. Kokkuvõtteks näitab Koit, kuidas ratsionalistlik ja empiristlik lähenemine arvutilingvistikas koos töötavad (Koit 2006: 49–51). Tugevalt empiristlikku lähenemist esindab siinses kogumikus Jane Klavani artikkel, mis näitab muuhulgas, kuidas andmemahud ja andmeanalüüs on keeleteaduses üha olulisemaks muutunud ning arvutimudelite kõrval vajab keeleteadlane kinnitust oma leidudele eksperimentaalseid meetodeid kasutades.

Lisaks aluspõhimõtetele on probleem ilmselt ka selles, et keeleteooriad (mille järgi keelt kirjeldada, mille mõistestikku kasutada) tegelevad justnagu oma probleemide ja kontseptsioonidega, oma vaidlustega, ega pööra piisavat tähelepanu sellele, mis toimub empiirilises uurimistöös ja keeleteaduse teistes paradigmatel. Näiteks võib siin tuua, et generativistide-funktsionalistide debatiid on haruldased ja kipuvad enamasti liiva jooksmale. Kummagi koolkonna aluspõhimõtted (kaasasündivus-õpitavus, ratsionalism-empirism, modulaarsus-holistilisus; ehk kokkuvõttes vastused küsimusele „mis on keel?“) on radikaalselt erinevad, kuid n-õ „oma poolust“ peetakse vaikimisi üldkehtivaks. Ja kui funktsionaalse koolkonna kui ajalisel hilisema keeleteadusliku voolu esindajad on harjunud ennast määratlema vastanduses eelneva, generatiivse koolkonnaga, siis generatiivsel koolkonnal pole olnud kunagi vajadust ennast funktsionalismiga kõrvutades määratleda, samas kuulevad nad teiselt poolt pidevalt teravat kriitikat ja vastandumist.

Siin peame peatuma terminil *teoreetiline keeleteadus*. Seda võib mõista mitmeti. Esmalt teoreetilise ja praktilise vastanduse kaudu. Sel juhul viitab termin mis tahes teoreetilistele arutlustele keele olemuse ja keel(t)e kirjeldamiseks vajalike mõistete üle. Teooriaid luuakse selleks, et püstitada nende abil hüpoteese, mida (praktiliselt) kontrollida. Just nii on mõistetud teoreetilist keeleteadust konverentsisarja kavandades. Kuid terminil on ka kitsam tähendus. Selle all peetakse tihti silmas vaid generatiivsesse suunda kuuluvat keeleteadust, mis vähemalt oma algusaegadel vastandus eelkõige kaas-aegsele rakenduslikule keeleteadusele (milleks loeti nt sotsiolingvistikat, teise keele omandamise uurimist, aga ka foneetikat) ja kindlaid keeli kirjeldavale keeleteadusele (nn deskriptiivne lingvistika). Generatiivne lähenemine võttis

ülesandeks uurida keelt teoreetilisel tasandil, kirjeldada keelevõimet, leida keeleuniversaale (nn universaalne grammatika, mis on iga keele alus). Noam Chomsky järgi peab lahus hoidma nn sisemise keele (ingl *I-language*) uurimise, mis tegeleb keele kui mentaalse objektiga, ja välise keele (ingl *E-language*) uurimise, mis tegeleb keele kui konkreetse keelekogukonnas tavaks saanud väljendusharjumuste (keelekasutuse) uurimisega.

Teoreetilise keeleteaduse kõrval käibib mõiste *üldine keeleteadus*, mis pärineb Ferdinand de Saussure'i tuntud loengusarjast, hiljem välja antud raamatuna („Cours de linguistique générale”), mis on nüüd ka eesti keeles lugejatele kättesaadav (Saussure 2017). Kuigi Saussure tutvustas ideed uurida keelt kui sellist, arenes see uurimisalana välja hiljem, just generatiivse keeleteaduse tuules (Sampson 2017).

Ehkki leidub keeleteadlasi, kes on kriitilised keele kui sellise uurimise mõttekuse suhtes üleüldse (näiteks Sampson (2017): „[k]eelte kohta ei ole lihtsalt mitte midagi väga öelda teoreetilisel tasandil”), ei saa tänapäeva keeleteaduse seisukohast siiski ammu enam väita, et keele üle teoretiseerimine on vaid kitsalt generatiivse suuna pärusmaa (kuigi selle jaoks on see selgelt keskne küsimus). Praktiliselt vajame arutelu keele üle kas või selleks, et selgitada välja (või pigem: pidevalt uuendada ja parandada oma arusaama sellest), milliste terminite (mõistete, kategooriate) abil on võimalik üksikkeeli kirjeldada (sh vaielda, kui suures osas saavad või ei saa eri keelte kirjeldused toetuda ühtlustatud terminitele, vt radikaalse käsitlusena Haspelmath 2010). Näiteks pakub keel(t)e kirjeldamiseks täiesti uusi mõisteid Ronald Langackeri kognitiivne grammatika (vt nt Langacker 2008).

Arutelu keele üle ei ole üldisemalt ahvatlev mitte ainult nende jaoks, kes generatiivse keeleteaduse vaimus tahavad kirjeldada keeleuniversaale (universaalset grammatikat) ja inimese keelemeele olemust. Ka generativistidele vastanduvad funktsionalistliku ja kognitiivse keeleteaduse esindajad otsivad keelele omaseid üldisi printsiipe. Erinevus seisneb selles, et ühtede jaoks on tegemist kindlapiiriliste reeglitega (mille abil keelt genereerida), mis ennustavad ja seletavad kõikide keelte toimimismehhanismi. Teiste jaoks on tegemist inimese olemusest tulenevate üldiste seaduspärasustega, mis aitavad mõista, miks keel – nii üldiselt kui ka üksikult – on selline, nagu ta on (sellest *funktsioon* ja *kognitiivne*), kuid mis ei pruugi sugugi kõikides keeltes ja kultuurides täielikult kattuda. Esimete jaoks saab reeglid tuletada mis tahes keelest (seega on inglise keele uurimine sel eesmärgil täpselt sama hea kui eesti või hiina keele uurimine), teiste jaoks on olemuslikult oluline keeleandmete mitmekesisus, teooriate kontrollimine paljude keelte andmetel. See tähendab, et nii funktsionalistide kui ka kognitiivsete keeleteadlaste hulgas leidub neid, kes tegelevad teoretiseerimisega keele üle (kuid nad ei lahuta seda keelekasutusest).

## 2. Kvantitatiivne või kvalitatiivne?

Kvantitatiivne-kvalitatiivne on keeleteaduse kontekstis mitmeti mõistetavad sõnad/hinnangud/meetodid: kvantitatiivne võib tähendada nii lihtsalt reaalsel andmetel põhinevat (st empiirilist, ingl *evidence-based* 'tõendusmaterjalipõhine') kui ka massiivsetel andmetel põhinevat (statistiliselt analüüsitud)

keeleteadust. Esimesel juhul vastandub see kvalitatiivsele kui mitteempiirilisele, uurija teoreetilisele (või ka introspektiivsele) arutlusele sellest, kuidas asjad võiks olla. Sellist vastandust kasutab siinses numbris nt Mari Aigro oma artiklis. Teisel juhul on oluline andmete saamise viis ja hulk: kvantitatiivne keeleteadus on selline, mis püstitab hüpoteesi, kogub selle kohta andmeid ja analüüsib neid statistiliste meetodite abil. Sellist lähenemist peavad oma artiklis silmas Jane Klavan ja Heiki-Jaan Kaalep. Kvalitatiivne on kasutusel keeleteaduses ka analüüsimeetodile viitamiseks, andmed on sel juhul empiirilised, kuid mitte massilised, analüüs on detailne, mitte statistiline. Kvalitatiivse keeleteaduse võimaluste üle arutlevad oma ühisartiklis Marge Käsper ja Raili Marling.

Kui küsida, kas keeleteaduslik mõte areneb tänu kvantitatiivsele pöördele või sellest hoolimata, siis otsene vastus saab siin olla ühene: kvantitatiivsed (korpusepõhised, statistilised ja katselised) uuringud on mõjutanud teoreetilise mõtte arengut positiivselt, ja mida aeg edasi, seda olulisemaks nende roll muutub, seda nii uute mõistete kui ka tervete probleemivaldkondade toomises keeleteoreetikute huvifääri. Samuti võib öelda, et lõhe teoreetilise ja empiirilise vahel ei ole lootusetult suur ning teoreetilise mõistestiku arengu ja empiirilistes (kvantitatiivsetes, eksperimentaalsetes) uuringutes toimunu vahel võib leida piisavalt seoseid.

### 3. Keele uurimine-kirjeldamine ja teooriad

Järgnevalt arutleme täpsemalt selle üle, kuidas keele uurimine ja kirjeldamine on seotud teooriatega. Üldlevinud definitsiooni järgi on teooria ülesanne seletada nähtus(t)e olemust, mida vastav teadus uurib. Ent teaduse kui tegevusala seisukohalt on seletamine tuumaks, mis annab teadusele nime, piiritleb uurimisobjekti ja ka eesmärgid. Tegelik „igapäevane” tegevus igas teaduses – kui tegevusvaldkonnas – on uurimistöö ja sealt lähtuvate probleemide lahendamine, andmete hankimine, uurimismeetodite arendamine, täiustamine. Teooriaid, mis mingis teadusvaldkonnas üritavad sama nähtust seletada, võib olla ja tüüpiliselt ongi mitu (vrd „Teoreetiline keeleteadus Eestis IV. Mitut teooriat me Eestis vajame?”). Iga teooria sõnastab eesmärgi – taotletava tulemuse – ja teatud üldised nõuded, mida see tulemus peab täitma. Teooria loojad-autoriteedid võivad küll tahta määrata ka seda, mida ja kuidas uurida sellise tulemuse saavutamiseks, aga konkreetse teaduse arengu aluseks on „igapäevauurimistöö” omaette tegevusena, teaduse objekti kohta selliste andmete hankimine, mis võimaldavad teooriaid ja neis sisalduvaid konkreetseid väiteid arendada (või ümber lükata). Selle vältimatuks osaks on meetodite ja esitusviiside-formalismide väljatöötamine, mida kasutatakse andmete töötlemiseks ning tulemuste kirjeldamiseks teooria nõuetele vastavas vormis. Selles arengus toimivad omad reeglipärasused, mida teooria kui abstraktne konstruktsioon täielikult ei kontrolli.

Teisalt: mitte igasugune uurimine, st mingite andmete kogumine ja töötlemine omaette, ilma eelnevalt sõnastatud eesmärgita – mida nende andmetega soovitakse uuritava nähtuse kohta välja selgitada, tõestada või ümber lükata (st teoreetilise seisukoha või hüpoteesita) – ei ole eraldi võetuna teoreetiliselt

relevantne teaduslik uurimine. Kvantitatiivsed korpusepõhised ja katselised meetodid keekekasutuse uurimises on küllap ahvatlenud tegema ka selliseid uurimusi, kus ei ole eelnevalt selgelt tajutud, rääkimata formuleerimisest, mis probleemidele missuguse teoreetilise kontseptsiooni raames peaks see uuring vastuse andma.

Tulemegi esmalt teooria juurde, sest siit lähtuvad ka hilisemad küsimused teooria ja meetodite seoste kohta. Määratluse järgi uurib keeleteadus inimkeelt, niisiis on keeleteooria ülesanne seletada, mis on inimkeel. Aga mida see seletus peab endast kujutama? Mida peab seletama, kui seletatakse keele olemust? Asi pole muidugi selles, et reaalseid keeli on tuhandeid, vaid selles, mida inimkeele kui sellise all niisuguses määratluses mõeldakse. Praegusi arusaamu sellest (ja nende erinevusi) on kajastatud ka Eesti keeleteaduslikus kirjanduses, näiteks annab Helle Metslang Keele ja Kirjanduse kaksiknumbris „Humanitaarteadused ja rahvus” sellest hea ülevaate, ehkki teises, üldisemas humanitaarteaduste kontekstis (Metslang 2018). Kognitiivse keeleteaduse arengut ja probleeme on hiljuti põhjalikult käsitletud nt Divjak jt (2016); varasematest vt nt Õim 2002; Õim, Tragel 2007. Olgu siin lisatud vaid mõned kommentaarid.

Mingis mõttes on iga keel kahtlemata autonoomne struktuur, muidu ei saaks grammatikaid kirjutada ega õpetada. Sama kehtib ka inimkeele kohta – taas: muidu ei saaks ühes keeles kirjutatud tekstide sisu arusaadaval viisil teise keelde tõlkida, olgu keelte sõnavara ja grammatiline struktuur nii erinev kui tahes. Seega lähtudes otstarbest – funktsioonidest, mida keeled (neid kasutavates inimkogukondades) täidavad – on võimalik kõiki keeli pidada samahästi kui üheks keeleks, mis eksisteerib paljude variantidena. Iga keel on ühes ja samas mõttes suhtlusvahend, millel on spetsiifiline otstarve ja struktuur (suhelda võidakse ka žestide, miimika, pooside jms abil). Seejuures on keel inimestele ainuomane: selle aluseks on miski, mida võime nimetada kas inimese keelevõimeks või keele kasutuspädevuseks. Esiteks, on hästi teada, et väljaspool inimkogukonda sündinud inimlaps inimkeelt kasutama ei õpi, kuid inimkogukonnas kasvades õpib ta ära – seejuures ilma mingi sihipärase õpetamiseta – selle keele, mida kogukond kõneleb, sõltumata sellest, mis keelt rääkisid tema bioloogilised vanemad. Ja teiseks tasub meenutada, et ehkki algselt toimus keel artikuleeritud kõne, st hääle vahendusel, seega vahetus kontaktis olevate suhtlejate vahel, ei ole see teatavasti ammu (vähemalt kirja leiutamise peale) enam obligatoorne ega ka mitte valdav. Teiste suhtlusvahendite ja -viiside kohta see samavõrd ei kehti.

Seega on keel (kui keelevõime realisatsioon) midagi muud kui lihtsalt hetkel vajaliku teabe edastamise vahend, see on nii faktiinfo kui ka mõtete, arvamuste ja tunnete, taotluste, keeldude-käskude, väljamõeldiste jne vahetamise vahend. Ja mis eriti oluline: nende säilitamise, „salvestamise” vahend mitte ainult suhtluses vahetult osalenute jaoks, vaid – tekstidena – ka teiste jaoks. Keele abil vahendatud teave salvestub mällu samamoodi kui muul viisil (nt vahetu meelelise kogemusena) hangitud teave ja on vajaduse korral taas kasutatav, muuhulgas ka keele (tekstide) kui suhtlusmeediumi abil teistele edastatav. Erinevates kultuurides võivad need vajadused olla vägagi erinevad, nii sõnavara (semantika), grammatiliste konstruktsioonide (süntaksi) kui ka suhtlussündmuste ülesehituse (pragmaatika) osas – nende varieeruvustega

tegelebki nt viimasel ajal aktiveerunud kultuurilingvistika (ingl *cultural linguistics*). Aga varieerumine ei saa inimkeelte puhul väljuda teatud piiridest, ja nende piiridega tegelebki teoreetiline keeleteadus.

Keele käsitlemisest primaarselt mitte fikseeritud tunnustega formaalse struktuurina, vaid suhtlemist, vastastikust arusaamist võimaldava vahendina algas hargnemine nii teoreetilises keeleteaduses üldse kui ka nt kognitiivses keeleteaduses. Kuidas keeled niisugust suhtlusfunktsiooni täidavad, mis faktorid tulevad mängu lisaks keele sõnavara ja grammatika valdamisele? Seda, et keeled on väga erinevad nii sõnavara kui ka grammatika osas, aga et see ei takista ühe keele tekste teise tõlkimast (st teksti mõtet ümber kodeerimast), teati juba ammu enne tänapäevases mõttes keeleteooriate sündi. Tänapäeva keeleteooriad ei anna siin ühest vastust, pigem pakutakse erinevaid kontseptsioone. Nende võrdlemiseks ja väidete kontrollimiseks on katselised ja kvantitatiivsed meetodid väga teretulnud.

Kokkuvõtvalt võib öelda, et keeleuurimise teeb teaduslikuks keeleteooriast lähtumine, ja keeleteoorial on siin kahesugune roll.

1. Keeli uuritakse, et luua teooriat. Ent teooria ei ole eesmärk omaette. Teooria esmane ülesanne on seletada, mis on inimkeel, kuid see seletus peab aitama ka välja töötada rakendusi, alates nt emakeele- ja võõrkeeleõpetusest kuni erinevate keelte arvutitöötluseni.

2. Teooria on see alus, millest lähtudes keelenähtusi uuritakse ja kirjeldatakse, nende erinevusi seletatakse. Teooriat arendatakse selle uurimistöö käigus. Keeruliseks teeb ettevõtmise esiteks see, et erinevaid inimkeeli on ju tuhandeid, ja teiseks see, et eri situatsioonides kasutatakse ka üht ja sama keelt väga erinevalt.

See tähendab, et teooria määrab nii selle, kuidas keelenähtusi uuritakse (metodoloogia ja metoodika), kui ka selle, mida ja milleks uuritakse. Eri aegadel on arusaamad nii ühest kui ka teisest olnud üpris erinevad. Keeleuurimise areng läbi sajandite on kulgenud lainetena, kord on huvi laineharjal keele sisu, kord vorm. Tegelikult võiks seda lainetuse asemel nimetada spiraaliks, sest iga järgmise etapi aluseks on olnud see, mis on tehtud enne, ja peamise eeskujuna on enamasti valitud üle-eelmine. Mõjutajaks on tihti olnud ka naaberteadused: antropoloogia, filosoofia, psühholoogia, arvutiteadus ja arvutilingvistika (piisab, kui vaadata, kuidas on sündinud XX sajandi eri keeleteooriad, kust on tulnud vastav lähenemisviis ja mõistestik; ühe näitena vt järgmist peatükki).

Kuid veelgi olulisem kui niisuguse lainetuse fikseerimine on ehk see, mida see lainetus ütleb keele olemuse kohta. Keel ei ole lihtsalt vorm, mis võimaldab väljendada soovivat sisu, ja keele vormi ei määra väljendamist vajav sisu üksi. Keel – iga üksik keel – on sisu ja vormi ühtsus, mistõttu on loomulik, et keeleuurimine on läbi sajandite kõikunud nende kahe poole vahel. Vastanduse teravus on ometi aja jooksul üha vähenenud. Seda on oluline silmas pidada, kui püüda oletada, mis suunas võib oodata keeleteooria arenemist. Isegi kui järgmiseks laineiks kujuneb taas vormi uurimise rõhutamine, mis on väga tõenäoline, peetakse silmas sisuga palju tihedamini seotud vormi kui eelmises vormi rõhutanud laines. Nii on see olnud kõigi eelmiste lainetega.

#### 4. Kuidas ja milleks mudeldada keelt?

Järgnevalt toome näite ühest mõjukast keelega tegelevast uurimisvaldkonnast, mis eeldab „päriskeeleteaduselt” ühevõrra nii kvantitatiivset lähenemist kui ka sobivaid teoreetilisi kontseptsioone.

Praktiline „katseline” valdkond, milles tehtav töö on pakkunud vägagi olulist materjali keeleteooriate jaoks, on arvutimudelite arendamine, mis võimaldavad kasutajal (inimesel) suhelda arvutisüsteemiga loomulikus keeles – nn keelt mõistvate süsteemide arendamine. See on valdkond, mida siin ei ole ilmselt põhjust pikalt tutvustada, selle kohta on ka Keeles ja Kirjanduses ilmunud piisavalt kirjutisi (nt Öim, Koit 2014, 2017). Oluline on see valdkond siinses kontekstis aga seetõttu, et loodavad süsteemid on otseses mõttes vaadeldavad kui katsepolügoonid, mis haaravad keelt tervikuna, mitte selle üksikuid tasandeid või aspekte, nii struktuurses kui ka funktsionaalses mõttes. Lähteidee on, et kui soovime arvutile üle kanda tekstide mõistmise võime, siis peame esmalt üle kandma oma teooria selle kohta, kuidas mõistmissüsteem toimib inimeses, selleks et panna arvuti samamoodi toimima – sest nii kirjalikud kui ka suulised tekstid, mille mõistmist üritame arvutile õpetada, on produtseerinud inimesed ja mõeldud inimestele. Seejärel võime kontrollida, kuidas see teooria arvutimudelina töötab. Tasub meenutada, et mitte ainult keele mõistmise, vaid ka inimliku tunnetus- ja arutlusvõime modelleerimist arvutil on üritatud juba enne tänapäevaste keeleteooriate sündi. Näiteks kognitiivteaduse loojate ja edendajate hulgas mängisid 1950.–1960. aastatel olulist rolli ka tehisintellekti arendajad (vt Öim 2008).

Toome kaks näidet, kuidas üldisema arvutiteadusliku taustaga teoreetilised mõisted on võetud kasutusele keele mudeldamisel ja sealt jõudnud teoreetilisse keeleteadusesse. Esimesena vaatleme freimi mõiste käibele tulekut. Keeleteaduses on see praegu tuntud eelkõige sõnade, lausete ja terviktekstide tähenduse esitusvahendina (*freimisemantika*). Selle mõiste (mitte ainult termini) tõi kasutusse Marvin Minsky (nt 1975; esimesed raportid, preprintid ilmusid juba palju varem) algselt tehisintellektisüsteemide jaoks, mis pidid visuaalse pildi – foto, filmi – analüüsi põhjal ära tundma, mis objekt on pildil (maja, auto, inimene, puu jne). Tollal oli vastava uurimisvaldkonna üldnimetuseks *kujundite äratundmine* ja see oli küberneetika üks olulisemaid rakendusvaldkondi. Minsky pakkus välja idee, kuidas formaalselt – st arvutis töödeldavalt – esitada andmeid nt tüüpilise maja, auto vms objekti kohta. Seejuures tüüpilisuse idee võttis ta tollal (kognitiivses) psühholoogias juba kasutatava prototüübi mõiste käsitlusest. Seda esitusviisi nimetaski Minsky kirjeldatava objekti freimiks. Objekti kirje kujutas endast tüüpiliste tunnuste (nt mõõtmed, kuju, liigendatus) ja nende võimalike väärtuste loendit.

Keeleteaduses – keeleteoorias – võttis freimi mõiste kasutusele Charles Fillmore (1968, 1977) lause tähenduse (süvastruktuuri) kirjeldamiseks, sest see oli tollal semantikas põletav probleem. Idee oli aga sama. Lausefreimi algmõiste keskmeks süvastruktuuris oli predikaat (kui puhtalt semantiline üksus), millel võis olla kindel hulk kindlas rollis argumente. Freim kujutas endast seega teatud situatsioonide, tegevuste jms tüüpstruktuuri, mida antud predikaadi-argumentide kombinatsiooni abil sai väljendada. Predikaadi argumentide vormistus pindstruktuuri lausetes, mis neid situatsioone või tegevusi



esitasid, võis varieeruda keeliti. Siit kulges kiire levik arvutilingvistikasse ja ka teoreetilise keeleteaduse erinevatesse kontseptsioonidesse. Lähemalt vt nt Winograd 1975.

Teiseks võib vaadelda vähem tuntud, aga mitte vähem teoreetilise ja katsetelise keeleteaduse kulgu mõjutanud sündmuse (ingl *event*) mõiste tulekut teoreetilisse keeleteadusesse, selle arengut ja praegust sisu keeleteoreetilises mõistestikus. See on näide ühest konkreetsest (justnagu) tavakeelesõnast, millest välja kasvanud mõiste on kasutusel nii keeleteaduses, psühholoogias kui ka arvutilingvistikas (tekstitöötluses, andmekaeves).

Keeleteaduslikus semantikas on sündmuse kategooria iseenesest olnud kasutusel juba vähemalt John Lyonsi töödest alates (nt Lyons 1977) ja seda on kasutanud ka kognitiivse suunitlusega autorid. Ent neis käsitlustes oli sündmus leksikaalse semantika kategooria, millega iseloomustati teatud verbide kui predikaatide (ja nende kaudu lausete) tähendusi: sündmus on verbiga tähistatav entiteet (tegevus, protsess, ...), millel on ajaline mõõde, st selle käigus midagi toimub, muutub, sel on üldreeglina fikseeritav algus ja lõpp, põhjus ja tagajärg jne (vt ülevaadet sündmuse mõiste sellisest käsitlusest Pajusalu 2009: 58–80).

Ent kui semantikas jõuti lausete juurest tekstide tähenduse käsitlemiseni, omandas sündmuse mõiste hoopis teise sisu: sündmused on üks teadmusstruktuuri liik, mille abil inimesed liigendavad tunnetatavas – või tekstides kirjeldatud – maailmas toimuvat „mõttekateks” osisteks ja salvestavad sellistena mällu. Siinses kontekstis on eriti oluline, et iga sündmust saab liigendada ja kirjeldada eri aspektidest ja eri detailsuse tasanditel, sündmust esitav teadmusstruktuur mälus ei säilita sellega edastatud teksti sõnastust ega vormilist liigendust, mille abil see teadmine sisestati. Lühidalt: sündmuse mõiste osutus vägagi oluliseks nii keeleteoreetilises semantikas kui ka pragmaatikas (selle problemaatika keeleteadusepoolseid käsitlusi vt nt Bohnemeyer, Pedersen 2011). Formaalseteks esitusvahenditeks pakuti mitmeid struktuure, mis kajastasid sündmuse ajalist dünaamikat: stsenaariumid, plaanid jt (Schank, Abelson 1977); aga sobivad teiste hulgas ka vastavalt modifitseeritud freimistruktuurid (Õim, Koit 2014). Inimmälu probleemid olid samal ajal vägagi aktuaalsed ka kognitiivses psühholoogias ja kahtlemata mõjutasid arenguid teoreetilises keeleteaduses. Eriti oluline oli siinses kontekstis semantilise ja episoodilise mälu eristus (vt nt Tulving 1994: 85–97). Sündmus on tüüpiline episoodilise mälu struktuuriüksus.

Tekstide arvutitöötluses muutus sündmuse mõiste aktuaalseks pärast seda, kui ajakirjandus muutus elektrooniliselt kättesaadavaks. Ajalehe üks põhiülesandeid on avaldada uudiseid selle kohta, mis on kusagil toimunud. Need uudised on tüüpilised sündmuste kirjeldused. Oluline on see, et sedasama objektiivses maailmas reaalset toimunud sündmust võidakse kirjeldada väga erinevatest aspektidest, fokuseerides tähelepanu nt osalistele (mida tegi persoon X), kohale (mis toimus kohas Y) või põhjustele ja tagajärgedele. Eri autorid/väljaanded võivad ka näiteks persooni X tegevusi kirjeldades fokuseerida tähelepanu eri detailidele, lähtudes seejuures hinnangust sellele tegevusele.

Niisugune sündmuste kajastus meedias on alati huvitanud poliitika- või majandusanalüütikuid, kuid samavõrra võib see huvitada muid mingi sünd-

muse (selle koha, osaliste, toimumisdetailide) uurijaid, alustades ajaloolastest ja lõpetades nt koduloouurijatega. Et uurimismaterjal kujutab endast tüüpiliselt tekste, kujuneski aktuaalseks probleem, kuidas automaatselt tuvastada, kas erinevates tekstides on kirjutatud samast sündmusest (*sündmusanalüüs, andmekaeve*). See omakorda nõudis vastavaid vahendeid sündmuste esitamiseks ja töötlemiseks (vt nt Orasmaa 2016).

Need kaks näidet illustreerivad tekste töötlevate arvutimudelite loomisel tekkivate vajaduste tagasimõju keeleteaduslikule semantikale ja pragmaatikale: uurimisteemadesse ilmuvad mõisted, nagu *suhtlussündmus, suhtlusruum, suhtlusstrateegia*, mille kirjeldamiseks on vaja oma formaalseid vahendeid (Koit 2017).

## 5. Ülevaade artiklitest

Ajakirja teemanumbrisse oleme kogunud kaheksa artiklit. Kogumikku alustab **Virve-Anneli Vihmani** artikkel laste keeleomandamisest („Pilk laste keeleomandamisele. Teoreetilistest alustest ja kvantitatiivsetest lähenemistest”). Lastekeele uuringud on viimaste aastakümnete jooksul läbi teinud suure muutuse nii teoreetilistes tõekspidamistes kui ka tehnika arengust tingitud andmekogumis- ja analüüsimeetodites. Vihman näitab nii varasema nativistliku lähenemise ahvatlevust kui ka seda, kuidas reaalsed andmed võivad elegantse teoreetilise arusaama kummutada. Samas toob ta välja, et kaasasündivuse-õpitanuse debatt pole triviaalne, see vajab muuhulgas pidevat arutlust ka selle üle, kuidas andmeid analüüsida ja tõlgendada.

Nii lapsekeeleuurijad kui ka murdeuurijad on sõltumata valitsevast teoreetilisest paradigmat paratamatult empiristid vastandina introspektiivse keeleteadusele. **Liina Lindströmi** ja **Maarja-Liisa Pilviku** artikkel „Korpuspõhine kvantitatiivne dialektoloogia” toob välja muutused murdeuurimises seoses murdekorpuste mahu suurenemisega ja statistiliste meetodite arenguga. Nad näitavad veenvalt, kuidas statistiline andmetöötlus annab murdeuurijale võimaluse varasemast oluliselt täpsemalt vastata mitmekesisitele küsimustele keele varieeruvuse kohta.

Järgneb artikkel taas alalt, mida introspektsiooni teel uurida ei saa – õppijakeelest ja selle uurimise eripärast keelekorpuste ajastul kirjutab **Mare Kitsnik**. Artikkel „Kuidas areneb eesti õppijakeel ja miks just nii – kas õppijakeelekorpus annab neile küsimustele vastused?” näitab, millist praktilist tulu toob õppijakeele andmete koondamine korpustesse. Korpusuuringud on aluseks tänapäevasele teaduslikult põhjendatud keeleõpetusele ja -õppele ning keeleoskuse hindamisele.

Küsimusega, milliseid katselisi meetodeid ja kuidas on kõige parem rakendada semantikas, tegeleb **Mari Uusküla** artikkel „Loetelukatse ja nimeandmiskatse psühholingvistika meetodina. Eelised ja kitsaskohad”. Uusküla toob oma pikaageelse kogemusele toetudes välja, kuidas kaks tuntud psühholingvistilist katset võivad anda olulist materjali keeleteaduslikeks uurimusteks. Samal ajal ei saa kunagi lootma jääda ühe või teise (katselise) meetodi ammendavusele või täiuslikkusele. Keeleteadlane peab olema tähelepanelik ja teadlikult kaaluma nii eeliseid kui ka kitsaskohti.

Katselistest meetoditest keeleteaduses ja meetodite mitmekesisuse vajadusest on juttu ka **Jane Klavani** artiklis „Kognitiivne keeleteadus arvude rägastikus”. Klavan on veendunud, et keeleteadus vajab hüpoteeside kontrollimiseks nii korpusandmete statistilist analüüsi kui ka keelekõnelejatega eksperimenteerimist. Need meetodid peaksid teineteist täiendama ja kindlustama keeleteaduse tulemuste valiidsuse.

Statistika rakendamise üle keeleandmete analüüsimisel arutleb ka **Heiki Jaan Kaalep** oma artiklis „Statistika koht keelemudelis”. Tema arutelu puudutab seda, kuidas keeleandmetele statistilist analüüsi rakendades ei tohi unustada eesmärki. Ta väidab, et nii keelekasutuse ülesehituslikke aspekte kirjeldavate kui ka keele produtseerimise mudelite tegemisel on tarvis üldist teoreetilist ettekujutust sellest, milline võiks olla keele produtseerimise mehhanism. Oma arutluse näitlikustamiseks on ta toonud kolm erinevat uurimust, mis kasutavad statistilist analüüsi.

Lisaks paradigmadevahelistele erinevustele mitmekesistavad keeleteaduse pilti ka kultuuriruumist tulenevad erinevused. **Marge Käsper** ja **Raili Marling** arutlevad oma artiklis „Diskursuseanalüüsi muutuv suhe kvantifitseerimisega inglisi- ja prantsuskeelses traditsioonis” diskursuseanalüüsi kui meetodi arengu üle keeleteaduses. Eriliseks teeb nende artikli kahe tugeva traditsiooni (inglisekeelse ja prantsuskeelse) võrdlemine. Diskursuseanalüüs on lisaks keeleteadusele kasutusel ka sotsiaalteadustes ning seegi on mõjutanud keeleteadusliku diskursuseanalüüsi kujunemist. Nii on diskursuseanalüüsi keskmes korraga nii tegeliku keelekasutuse analüüsi rõhutamine kui ka püüd leida kvalitatiivsele meetodile võimalikult „ühtne kuub”, st seda formaliseerida.

Kvantitatiivse pöörde ja sellega kaasnevate vaidluste üle keeleteaduses üldiselt arutleb **Mari Aigro** artiklis „Kvantitatiivne pööre – vana tüli uues kuues”. Ta näitab, et kvantitatiivsed meetodid iseenesest ei ole seotud ühegi keeleteadusliku paradigmana, nagu ka ei ole need mõne paradigma poolt välistatud. Vahelduseks tänapäeva Eesti keeleteaduses muidu levinud generatiivse keeleteaduse kriitilisusele on Aigro väited pigem funktsionalismikriitilised ja toetavad generatiivse keeleteaduse õigust samuti osaleda „kvantitatiivses pöördes”. Üheks keskseks väiteks on, et uurimismeetodi kriitikat (kvantitatiivsed empiirilised andmed *vs.* introspektsioon) ei tohi segamini ajada uurimisobjekti (keel üldiselt, ajuteadus *vs.* keel kui kommunikatsioonivahend) kriitikaga. Artiklit lugedes on hea meeles pidada, et funktsionalistide põhiväide ongi see, et me ei saa uurida keelt üldiselt lahus reaalsest keelekasutusest. See hoiak ühendab nii kognitiviste, funktsionaliste kui ka neid, keda artiklis nimetatakse deskriptivistideks. Nimetus *deskriptivist* ise viitab generatiivsele-formaalsele lähtekohale, et on võimalik lahutada keele kirjeldus üldiselt ja keelekasutuse kirjeldus.

Kokkuvõtteks võib öelda, et artiklid annavad mitmeid erinevaid sissevaateid tänapäeva keeleteaduses toimuvasse, kus ristuvad, põimuvad ja hargnevad paradigmad, teooriad, andmekogumis- ja analüüsimeetodid. Need ei esinda kaugeltki kogu pilti praeguse aja kirjust keeleteadusest ei maailmas ega ka mitte Eestis, kuid loodetavasti suudavad siiski panna lugeja kaasa mõtlema lõputult hargneval teemal, mis on keel ja kuidas seda uurida.

## Kirjandus

- Bohnmeyer, Jürgen, Pedersen, Eric (toim) 2011. *Event Representation in Language and Cognition*. Cambridge: Cambridge University Press.
- Divjak, Dagmar, Levshina, Natalia, Klavan, Jane 2016. Cognitive Linguistics: Looking back, looking forward. – *Cognitive Linguistics*, kd 27, nr 4, lk 447–463.
- Fillmore, Charles 1968. The case for case. – *Universals in Linguistic Theory*. Toim Emmon Bach, Robert T. Harms. New York: Holt, Rinehart and Winston, lk 1–88.
- Fillmore, Charles 1977. Scenes-and-frames semantics. – *Linguistic Structures Processing*. Toim Antonio Zampolli. Amsterdam: North-Holland, lk 55–8.
- Geeraerts, Dirk 1993. Cognitive semantics and the history of philosophical epistemology. – *Conceptualizations and Mental Processing in Language*. (Cognitive Linguistics Research 3.). Toim Richard A. Geiger, Brygida Rudzka-Ostyn. Berlin: De Gruyter Mouton, lk 53–79.
- Haspelmath, Martin 2010. Comparative concepts and descriptive categories in crosslinguistic studies. – *Language*, kd 86, nr 3, lk 663–687.
- Koit, Mare 2006. Ratsionalism ja empirism keeletöötlustes: vastasseis või koostöö? – *Teoreetiline keeleteadus Eestis II*. (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 7.) Toim Ilona Tragel, Haldur Õim. Tartu: Tartu Ülikooli Kirjastus, lk 41–54.
- Koit, Mare 2017. Dialoogsüsteem ja läbirääkimiste struktuur. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, nr 13, lk 37–51.
- Langacker, Ronald W. 2008. *Cognitive Grammar. A Basic Introduction*. Oxford: Oxford University Press.
- Lyons, John 1977. *Semantics I–II*. Cambridge: Cambridge University Press.
- Metslang, Helle 2018. Kas rahvusel on kohta lingvisti peas? – *Keel ja Kirjandus*, nr 1–2, lk 50–55.
- Minsky, Marvin 1975. Minsky's frame system theory. – *TINLAP'75: Proceedings of the 1975 workshop on theoretical issues in natural language processing*. Stroudsburg: Association for Computational Linguistics, lk 104–116.
- Orasmaa, Siim 2016. *Explorations of the Problem of Broad-coverage and General Domain Event Analysis: The Estonian Experience*. (Dissertationes Mathematicae Universitatis Tartuensis 108.) Tartu: University of Tartu Press.
- Pajusalu, Renate 2009. *Sõna ja tähendus*. Tallinn: Eesti Keele Sihtasutus.
- Sampson, Geoffrey 2013. The empirical trend: Ten years on. – *International Journal of Corpus Linguistics*, nr 13, lk 281–289.
- Sampson, Geoffrey 2017. *The Linguistics Delusion*. Sheffield–Connecticut: Equinox Publishing Limited.
- Saussure, Ferdinand de 2017. *Üldkeeleteaduse kursus*. Tlk Tiit Kuuskmäe. Tallinn: Varrak.
- Schank, Roger C., Abelson, Robert P. 1977. *Scripts, Plans, Goals and Understanding. An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tulving, Endel 1994. *Mälu*. Tlk Jüri Allik. Tallinn: Aktsiaselts „Kupar”.
- Winograd, Terry 1975. Frame representations and the declarative/procedural controversy. – *Representation and Understanding. Studies in Cognitive Science*.

Toim Daniel G. Bobrow, Allan Collins. New York–San Francisco: Academic Press, lk 185–210.

Õim, Haldur 2002. Teoreetiline keeleteadus Eestis. – Teoreetiline keeleteadus Eestis. (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 4.) Toim Renate Pajusalu, Ilona Tragel, Tiit Hennoste, H. Õim. Tartu: Tartu Ülikooli Kirjastus, lk 11–23.

Õim, Haldur 2008. Kognitiivne pööre. – Keel ja Kirjandus, nr 8–9, lk 617–627.

Õim, Haldur, Koit, Mare 2014. Keele mõistmise modelleerimine arvutil. Semantika, ontoloogia ja sündmuste esitus kahe valdkonna näitel. – Keel ja Kirjandus, nr 3, lk 169–194.

Õim, Haldur, Koit, Mare 2017. Suundumusi inimsuhtluse keelises analüüsis ja modelleerimises I–II. – Keel ja Kirjandus, nr 1, lk 71–80; nr 2, lk 143–150.

Õim, Haldur, Tragel, Ilona 2007. Teoreetilise keeleteaduse arengust mujal ja meil XXI sajandi alguses. – Keel ja Kirjandus, nr 2, lk 98–115.

*Ann Veismann (sünd 1972), PhD, Tartu Ülikool, eesti ja üldkeeleteaduse instituut, eesti ja üldkeeleteaduse lektor, ann.veismann@ut.ee*

*Jane Klavan (sünd 1983), PhD, Tartu Ülikool, maailma keelte ja kultuuride kolledž, inglise keele ja lingvistika lektor, jane.klavan@ut.ee*

*Haldur Õim (sünd 1942), PhD, Tartu Ülikool, eesti ja üldkeeleteaduse instituut, arvuti-lingvistika vanemteadur, haldur.oim@ut.ee*

## **Theoretical linguistics and quantitative methods**

**Keywords:** quantitative turn in linguistics, theoretical linguistics, different methods in linguistics

The article serves as a preface to the present collection of articles, which are based on conference papers on “Quantitative turn in linguistics – what about theory?”. It discusses why the question has arisen and how a multitude of linguistic theories may impact our understanding of methods. Some essential theoretical concepts and their backgrounds are explained. There follows a detailed description of two examples of how a concept born in the framework of a certain theory has spread, thus acquiring an important role in the choice of method. Finally, a brief overview is given of each article included in the special issue.

*Ann Veismann (b. 1972), PhD, University of Tartu, Institute of Estonian and General Linguistics, Lecturer of Estonian and General Linguistics, ann.veismann@ut.ee*

*Jane Klavan (b. 1983), PhD, University of Tartu, College of Foreign Languages and Cultures, Lecturer in English Language, jane.klavan@ut.ee*

*Haldur Õim (b. 1942), PhD, University of Tartu, Institute of Estonian and General Linguistics, Senior Research Fellow in Computer Linguistics, haldur.oim@ut.ee*