

KVANTITATIIVNE PÖÖRE – VANA TÜLI UUES KUUES

MARI AIGRO

Senisest suurema koguse andmete kaasamine teadustöösse tekitab uue positivismilainena suurt entusiasmi nii teadlastes kui ka ühiskonnas. Mõlemad lähtuvad üldist tunnustust leidvast teadusfilosoofilisest põhimõttest, et kvantifikatsioon teeb teadusharu süstemaatilisemaks ja seega „teaduslikumaks”. Siit loomulikult ei järeldu, et kvalitatiivsed meetodid ei küündi teaduseni. Pigem püütakse selle teesiga väljendada kvantifikatsiooni suhtelist tähtsust: kui võimalik, võiks teadlane vaadata kvantitatiivsete meetodite poole, kuid kui teadusharu on mingil määral „loomult” kvalitatiivne või kui puhttehniliselt pole (veel) kvantitatiivsete meetodite kasutamine võimalik, ei tasu teadustööle ka uusi meetodeid kunstlikult külge pookida.

Lisaks paratamatule kvalitatiivse ja kvantitatiivse metodoloogia vahelisele vastasseisule toob suurte andmestike ja andmetöötlusmeetodite levik paljudes teadusharudes kaasa olulisi küsimusi teoreetilise ja kvantitatiivse lähenemise suhte kohta. Kvantitatiivse pöörde käigus võib teoreetiliste arutelude pädevus teadustöös sattuda kergesti kahtluse alla, sest juhtrollis nähakse nüüd andmeanalüüsi. Veelgi enam – uues valguses võib lihtsasti paista, et teadustöö esmase tööriista roll ongi alati andmeanalüüsile kuulunud ning et kaasaegne tehnoloogia võimaldab sellel lihtsalt kehtestada oma õiguspärast staatust teadmise loomise mudelis. Selle mustriga järgi on teooriat näinud andmetega vastandumas nii loodusteadlased (Darwin, Seward 1903; Burton 1963) kui ka humanitaarteadlased, nende hulgas keeleteadlased (Yngve 1996).

Nii nagu teisteski teadusharudes, on teooria ja andmete vastandamine küsimusi tekitanud ka keeleteaduses. Vastandust lähemalt analüüsides selgub, et mitu fundamentaalset aspekti on ebaselged ja mitte kõik vaidluspooled ei räägi päris samast asjast. Esiteks teeb arutelu segasemaks see, et uus, metodoloogiline konflikt tavatseb seguneda vanema, formaalse ja funktsionaalse lingvistika vahelise ideoloogilise lõhega, mis on keeleteaduses kestnud juba 60 aastat. Kvantitatiivsete meetodite poolt kõnelejad ei vaidle alati vastu mitte lihtsalt kvalitatiivsetele töödele, vaid keeleteaduse formaalsele traditsioonile, mis ehitab väga detailseid ja paljude arvates liiga spekulatiivseid mudeleid. Kahe eraldiseisvalt olulise diskussiooni segunemine hägustab argumente ja teeb lahenduste leidmise keeruliseks. Seda, et konflikt on aktuaalne just kvantitatiivse liikumise ja teooria vahel, illustreerib hästi 2017. aasta novembris toimunud keeleteaduse andmepöördele keskendunud konverentsi pealkiri, mis küsis: „Mis saab teooriast?”, mitte „Mis saab kvalitatiivsetest meetoditest?”. Sellest kasvab välja ka teine probleem, mis puudutab kasutatavate terminite mitmetimõistetavust. Teoreetilise ja empiirilise lähenemise vastandamisel on suhteliselt selge, mida täpselt mõeldakse andmestike ja

analüüsimetodite all, kuid pole just kuigi selge, mida peetakse silmas mõiste all *teoreetiline lingvistika* ja kuidas nähakse seda andmetega konfliktis olevat.

Käesolevas artiklis arutan neid kahte aspekti. Kõigepealt demonstreerin, kuidas vana lõhe kahe suure suuna, formaalse ja funktsionaalse lingvistika vahel on üle kandunud uude metodoloogilisse vaidlusse, saanud kummalisel kombel andmerefunktsiooni osaks ning hakanud vaidluspooli kriitilistelt arutelupunktidelt kõrvalistele teemadele juhtima. Seejärel näitan, kuidas meetodi ja teooria suhet erinevates keeleteaduse harudes uurides selgub, et meetodi valik ega teooria roll ei ole üheski neist piiratud. Seega pole need kaks kontseptsiooni teaduses mitte komplementaarsed, vaid paralleelsed nähtused ja sisuline konflikt nende vahel puudub. Artikli peamine sõnum on, et sõltumata sellest, millist meetodit uurija kasutab, ei pea ega tohi teooria roll seetõttu marginaliseeruda. Need kaks teaduse aspekti ei saa esineda teineteisest lahus, sest üht ei saa mõtestada ilma teiseta.

1. Andmepõhine lähenemine teaduses

Enne keeleteadusele keskendumist on oluline mõista kvantitatiivse pöörde rolli üldiselt teadustes, sest keeleteadlased pole kaugeltki esimesed, kes oma teadusharu uude meetodi valguses taasavastavad.

Vaatlus- ja teooriapõhised teadlased vastandusid üksteisele juba enne andmeanalüüsi laialdasemat levikut. Näiteks Charles Darwin kaebas kirjas Henry Fawcettile 1903. aastal, et geoloogias on liiga palju juttu sellest, et teadlane peaks teoretiseerimise asemel vaatlema, kommenteerides seda nii: „[---] iga vaatlus, millest on kasu, toetub või vastandub mõnele [teoreetilisele] vaatele” (Darwin, Seward 1903). Üks eelmise sajandi selgemaid näiteid andmeanalüüsi laiema leviku mõjudest kuulub aga hoopis geograafia valdkonda, sest 1950. aastate kvantitatiivne (matemaatiline) revolutsioon oli niivõrd äkiline ja tugev, et see sai erialastes ajakirjades sageli ka ise diskussiooni fookuseks. Geograafid vaidlesid, miks ja kuidas võiks uus, mõõtmistele orienteeritud suund kasulik olla, sealjuures tundusid uued meetodid eriti vastuolulised inimgeograafias, mille andmeid peeti olemuselt kvalitatiivseks. Pärast umbes kümme aastat kestnud tuliseid vaidlusi said revolutsioonilisena tundunud ideed tavateadmise osaks ja torm vaibus. (Burton 1963)

Käesoleva artikli perspektiivist on eelkõige oluline märkida, et kuigi esimesed matemaatilise geograafia esindajad pidasid andmeid ainsaks oluliseks teaduse komponendiks, siis Jan Burtoni (1963) sõnul tõstis üleüldine kvantifitseerimistendents teoreetilised raamistikud lõpuks varasemast pigem rohkem au sisse. Revolutsiooni eesmärk polnud lihtsalt uus meetod, vaid geograafia klassikalise teadusmeetodi mõistes teaduslikumaks muutmine, mistõttu andmetel põhinevate vaatluste levides kinnistus ka teooria staatus funktsionaalse „faktisöelana”, mille abil tuli andmestikust esile tõusvaid fakte sorteerida ja mõtestada. „Kõik geograafia harud, mis väidavad end olevat teaduslikud, vajavad arenevat teooriat ning kõik harud, mis vajavad arenevat teooriat, vajavad kvantitatiivset metoodikat” (Burton 1963: 158).

Suurte andmestike ja arvutisimulatsioonide levikuga võib teadusmaastikul märgata veel teist ja hilisemat kvantitatiivset pööret, mille intensiivsus

varieerub teadusharuti. Ka selle puhul on küsitav, kuivõrd see klassikalise teadusmeetodi alustalasid raputab. Näiteks räägiti meditsiinis kolmteist aastat tagasi vajadusest lisada teadusmeetodisse uus, tehnoloogial põhinev vaheamm: pärast hüpoteesi püstitamist peaks iga uuring viima läbi planeeritud eksperimendi kõigi võimalike versioonide simulatsioonid, mis näitaks, millisel kujul oleks tegelikku eksperimenti kõige optimaalsem teha (Satava 2005). Kvantlaine harjal on tehnoloogiaajakirjanduses saanud kuulsaks geeniuurija ja -järjestaja Craig Venter, keda tuuakse esile kui teadlast, kes klassikalise teooria-andmed-kinnitus mudeli asemel n-ö õngitseb teooria välja hoopis andmetest. Nii kuulutas tehnoloogiaajakiri Wired tema näitel 2008. aastal, et teooriaajastu on teaduses lõppenud (Anderson 2008). Tundub aga, et Wired ei pööranud väga palju tähelepanu Venteri enda jutule, sest viimane räägib ühes intervjuus andmete juhitud teaduse asemel hoopis ümberpööratud teadusmeetodist, milles teooria järgneb andmete kogumisele ning mis pole teaduspraktikas sugugi ebatavaline. „Darwin ei mõelnud evolutsiooniteooriat välja Galapagosel ringi kõndides. [...] Alles siis, kui ta koju tagasi jõudis ja kogutud materjali uuris, märkas ta nendevahelisi erinevusi ja asetask need konteksti.” (Shreeve 2004)

Kaardistades uusi teadusmetoodilisi mõttesuundi, tasuks mainida ka vastupidist liikumist füüsikas, mida illustreerib Richard Dawidi 2015. aastal Münchenis korraldatud konverents „Miks usaldada teooriat?”. Konverentsil püüdis üks füüsikute koolkond teist koolkonda veenda, et ka need teooriad (nagu näiteks stringiteooria), mida sisuliselt pole võimalik ei empiirilisel tõestada ega ümber lükata ja mis seega on Karl Popperi (1959) järgi halb teaduspraktika, saavad siiski olla valiidsed teadussuunad, mille kallal tuleb edasi töötada (Dawid 2015).

Nõnda on andmeanalüüsi levik paljudes teadusharudes toonud lisaks muudatustele metodoloogias kaasa ka lihtsamaid uuendusi meetodis. Fundamentaalseid muutusi pole see aga teooria ja andmete vahelises suhtes siiski põhjustanud. Kuigi 1950-ndate alguses tundus ehk geograafias, et teooria tuleb vana ja tõlmuse tööriistana uuest ja võimekast andmepõhisest teadusmudelist välja jätta, siis lõpuks selle roll, detailsus ja integreeritus uurimustega hoopis paranes, sest andmestike laienedes tekib suurem kogus uusi fakte ja seega ka suurem vajadus süstemaatilise ja detailsema „faktisõela” järele. Geenitehnoloogias hakkas sama idee uuesti sensatsiooniline tunduma, kuid sisulisel vaatlusel selgub, et kuigi kronoloogiliselt võivad teooria ja andmed mudelis ümber pöörduda, pole sisulist muutust nendevahelises hierarhias toimunud. Ühtlasi, nagu illustreerib stringiteooria näide, võib probleeme tekkida ka vastassuunal: kui väga abstraktsed teooriad pole tehniliselt ümberlükatavad, kas see tähendab, et nendega töötamine on automaatselt aja raiskamine?

2. Uued ja vanad lõhed lingvistikas

Andmeanalüüsi levides on paljud keeleteadlased hakanud end nägema osana liikumisest, mis populariseerib andmeanalüüsi ja kvantitatiivseid meetodeid. See on paljudele humanitaariaharudele juba tuttav nähtus, aga keeleteaduse teeb eriliseks asjaolu, et see oli juba enne andmeanalüüsi laiemat levikut

lõhestunud. Konflikt, mille üheks pooleks on formaalne keeleteadus ja teiseks pooleks nimetatakse vahel funktsionaalset, vahel deskriptiivset ja vahel sotsiolingvistilist suunda, ulatub juba umbes 60 aasta tagusesse aega, mille alguseks loen siinkohal Noam Chomsky esimese transformatsiooniteemalise raamatu ilmumist (1957). Lõhe on püsinud äärmiselt aktuaalsena tänapäevani, sest mõlemad harud on aktiivsed ja produktiivsed, vaatamata sellele, et nende vahel on sügavad teadusfilosoofilised eriarvamused, mis algavad uurimisobjekti olemusest. Järgnev alapeatükk näitab, et nimetatud konflikt pole mitte ainult vaibunud, vaid seda on hakatud ekslikult segama uute kvantitatiivsete meetodite kasutuselevõtu diskussiooniga.

2.1. Teadusfilosoofiline lõhe keset keeleteadust

Nagu öeldud, on varasemate eriarvamuste keskmeks vastandus formaalsete lingvistide, eelkõige generatiivsete koolkondade ja funktsionaalsete lähenemiste pooldajate vahel. William Croft (1991: 5) võrdleb funktsionaalset ja formaalset suunda, öeldes, et kuigi mõlemad püüavad kategoriseerida keelesüsteemi ja kaardistada üldiselt kehtivaid piire, mis defineerivad võimaliku inimkeele, siis nende metodoloogilised erinevused tingivad selle, et ka nende arusaam otsitavast süsteemist erineb radikaalselt. Mõlemad süüdistavad teineteist „valedele andmetele keskendumises”: formalistid arvavad, et funktsionaalne suund keskendub ekslikult baasstruktuuri asemel obskuursetele keelenäidetele (Croft 1991: 6), samas näiteks sotsiolingvistilis-deskriptiivne suund süüdistab formaalseid lähenemisi sageli kultuurikonteksti ja muude sotsiaalsete tegurite eiramises (Romaine 1994).

Teisalt aga on aspekte, mille poolest kõik formalistide ja ülejäänud keeleteadlaste vahelised vastasseisud üksteisega sarnanevad. Paljud lingvistid ühinevad formalistide vastu, sest nad ei nõustu generatiivse liikumise ideega sellest, et keelevõime (ingl *competence* või *I-language*) ja keelekasutus (*E-language*) on keele erinevad aspektid, mida peaks uurima eraldi (Chomsky 1986). Generatiivse keeleteooria järgi on keelekasutus üks kahest võimalikust loomuliku keele kontseptsioonist, mis näeb keelt inimestevahelise, tavade poolt kujundatud sotsiaalse objektina (Ludlow 2011: 45). Keelevõime on aga aju keelekeskus (ingl *faculty of language*, Hauser jt 2002), loodusteaduslik psühholoogiline objekt, mis sisaldab kõigile maailma keeltele ühiseid alusreegleid ja mille olemasolu tõestuseks peetakse muuhulgas meie võimet kasutada ja mõista lauseid, mida me iial kuulnud pole. Keele varieerumist ja muutumist uurivad teadlasi ei ärrita ainult asjaolu, et Chomsky eristab keelevõime uurimist keelekasutusest, vaid ka see, et ta kuulutab keelevõime uurimise lingvistika peamiseks eesmärgiks, sest loodusteaduste eeskujul otsib see suund „optimaalset vormi”, lihtsaid aluspõhimõtteid (Chomsky 2002: 56–57). Lõhe keelevõime ja keelekasutuse uurijate vahel on niisiis tõsine, sest suur osa neist, kes Chomsky definitsiooni järgi on keelekasutuse uurijad, ei pea keele kasutamise uurimist sotsiaalses kontekstis vähem teaduslikuks kui aju keelekeskusele keskendumist ega ka n-ö pärislingvistikast eraldiseisvaks tegevuseks. Paljud on veendunud, et inimkeele uurimine ilma sotsiaalse kontekstita pole üldse võimalik ning selle praktiseerimine ei saa meile anda mõtestatud teadmist

keele toimimise printsiipidest (Mesthrie 2012; Van Herk 2012). Õli valab tulle asjaolu, et keelevõime uurijatel on komme vastandada oma tööd keelekasutuse uurijate tööle, öeldes, et nende lähenemine on eriline, sest see küsib keele kohta miks-küsimusi ning on keskendunud keelte sarnasuste, mitte erinevuste otsimisele – seaduspärade otsimine ja maailma süstematiseerimine on ometigi teaduse olemus (Chomsky 2002; Ludlow 2011). Keele kasutusele orienteeritud teadlased viitavad aga asjaolule, et nõnda jäiga fookusega ning eriti keeleandmetele keskendumata jääb suur osa loodud teadmisest (eelkõige laboritingimustes detailideni välja töötatud formaalsed mudelid) hoopis spekulatsioonide valdkonda ega küündi teaduseni (Levshina 2015).

2.2. Looduslik või kultuuriline objekt?

Kuigi konflikt on vana ja keeruline, mängivad selles väga olulist rolli kaks vastanduskohta: ühelt poolt vaatlevad need kaks leeri keelt kui sisuliselt erinevat uurimisobjekti ja teiselt poolt on neil väga erinevad meetodid, s.o andmekasutustraditsioonid.

Keelt võib uurimisprotsessis näha nii kultuurilise kui ka loodusliku objektina. Arusaam keelest kui kultuurilisest objektist on süstemaatilisemalt kristalliseerunud William Labovi töödes, kuid idee on palju vanem. Edward Sapir on öelnud: „[---] keele omandamine on iseenesest väga erinev protsess kõndima õppimisest. Viimase puhul ei mängi kultuur, s.o traditsiooniline kogum sotsiaalseid kasutusviise, erilist rolli.” (Sapir 1921: 1) Kultuurilisena näeb keeleteadust ka Victor Yngve (1996), kelle järgi on see loodusteadustest (ja nende teadusmeetodist) lahknenu nõnda kaugemale ning selle terminoloogia (nt „sõna”, „nimisõna” ja „lause”) on nii põhjalikult kultuuriline, et arvukate teooriate ja suundade vahel on võimatu objektiivselt ühe või teise kasuks otsustada. Ainsa lahendusena näeb Yngve seda, et keeleteadus loobuks tervenisti loogikapõhisest ja abstraktsest lähenemisest ning keskenduks ainult keele füüsilisele väljendusele, s.o muutuks foneetikakeskseks.

Ligi 60 aastat tagasi alguse saanud generatiivne koolkond (Chomsky 1957, 1965) näeb aga keelt kui ajus mingil määral fikseeritud (ingl *hard-wired*) objekti ning oma uurimistraditsiooni kui lingvistika loodusteaduslikku haru. Ka see pole uus idee – Chomsky ise nimetab generatiivse mõtte inspiratsiooniallikana Wilhelm von Humboldti (1836, Chomsky 1965 kaudu), kes seadis keeleuurijate lähteülesandeks kirjeldada, kuidas on keele piiratud grammatikaga võimalik luua lõputu hulk lausungeid, paigutades juba seeläbi keele kaudselt pigem ajju kui ühiskonda. Kolmanda ja vähem esindatud arusaama järgi paigutub keeleteadus sarnaselt matemaatikaga kultuuri- ja loodusteaduste vahele – loomulik keel on küll täielikult mõistetav vaid keelekasutusena, st sotsiaalse üksusena, ega saa seega olla loodusteaduse uurimisobjekt, kuid erinevalt kultuuriteadustest on keeleteaduses kasutusel välised, abstraktsed mõisted ja universaalid (Koster 2006).

Kahe keelekäsitluse konflikt väljendub viimasel ajal näiteks Daniel Everetti kriitikas Chomsky suunal. Everett, olles avastanud, et pirahani keeles puudub aju keelekeskuse aluskomponendiks peetav rekursioonifunktsioon, väidab, et keel on läbi ja lõhki kultuuriline tööriist, mida kasutatakse

kogukonna liikmete mõtete ja käitumise kujundamiseks (2012). Eriti kõnekas näide sellest, kuidas kaks liikumist on 60 aastaga üksteisest aina enam kaugenud, on generatiivse keeleteaduse pooldaja Steven Pinkeri kommentaar, mis kasutab sajanditagust Sapiri võrdlust vastupidise eesmärgiga: „Keel pole rohkem kultuuriline leiutus, kui seda on inimeste püstises asendis kõndimise võime” (Pinker 1994: 18).

Kirjeldatud uurimisobjekti erinevusest välja kasvaval lõhel on aga väga oluline ja mõnes mõttes ka ootuspärane tagajärg, nimelt lõhe kahe suuna metoodikas. Formaalsele koolkondadele omistatakse ehk liiga eksklusiivselt, kuid kohati siiski õigustatult kvantitatiivsete meetodite eiramise traditsiooni. Eraldi rünnakuobjektiks on saanud formaalses lingvistikas sageli (kvalitatiivselt ja informaaalselt) kasutatavad grammatilisus- ja vastuvõetavushinnangud üksikutele lausungitele. Carson Schütze (2016) esitab rea argumente, mis näitavad, et kui hinnangumeetodit korrektselt ja kriitiliselt kasutada – see tähendab muuhulgas, et hinnangu andjaks on keelerääkijate kogukond, mitte keeleteadlane ise või tema lähim kolleeg (nagu see formaalsetes-generatiivsetes suundades sageli tavaks on olnud) –, siis on tegu kasuliku, kvantitatiivse ning mitte sugugi ainult formaalse keeleteadusega piiritletud meetodiga. Peter Ludlow (2011) selgitab, miks ideoloogilisest lõhest on saanud muuhulgas ka metodoloogiline lõhe. Kuigi ta kinnitab, et generatiivne, keelevõimele fokuseeritud keeleteadus võib tehniliselt kasutada nii kvantitatiivseid kui ka kvalitatiivseid meetodeid, siis keelekorpused võivad „liiga madalal rippuva puuviljana” olla teooriate arendamisel probleemsed, sest sisaldavad mõne konkreetse nähtuse puhul vigu ja lünki.

2.3. Vana konflikt uue nurga alt

Nõnda on konflikt „formaalne *versus* deskriptiivne/funktsionaalne/sotsioloogiline lingvistika” sujuvalt üle kandunud metoodika valda ning saanud konfliktiks „formaalne *versus* empiiriline lingvistika”. Formaalsete lähenemiste suunas tuleb aina rohkem kriitikanooli n-ö andmemahukatest lingvistika-harudest, mida defineeritakse aina sagedamini nende vastandina. Hea näide on statistiline loomuliku keele töötlus (ingl *Natural Language Processing* ehk NLP) – näiteks Christopher Manning ja Hinrich Schütze seovad kohe NLP õpiku alguses meetodivaliku uurimisobjekti keelevõime ja keelekasutuse temaatikaga, öeldes, et „[---] generatiivsed keeleteadlased väidavad, et keelevõimet saab kirjeldada isolatsioonis, kuid empiirilised lähenemised sellega ei nõustu, soovides kirjeldada keele tegelikku kasutust” (Manning, Schütze 1999: 6). See tsitaat illustreerib hästi kahe vaidluse segunemist, sest autorid viitavad selgelt sellele, et keelevõimet uuritakse mitteempiiriliselt ja keele tegelikku kasutust empiiriliselt (see, kas nende „keele tegelik kasutus” kattub keelekasutuse kontseptsiooniga, pole hetkel oluline). Kuigi säärane üldistus on laialt levinud (Croft 2002), ei ole siiski võimalik öelda, et empiirilised meetodid on kuidagimoodi eksklusiivselt just kasutuspõhiste (sh funktsionaalsed ja kognitiivsed raamistikud) või deskriptiivsete lähenemistega seotud. Seda tundub väljendavat ka korpuslingvistika ja andmeliikumise populariseerija Stefan Gries (2015, 2017), kes küll vastandab empiirilist suunda generatiiv-

sele, kuid viitab sealjuures alati ettevaatlikult ainult formaalsete suundade mitteempiirilisele traditsioonile, mitte nende uurimisobjekti mitteempiirilisele olemusele.

Ludlow küll proovib formalistide suunas tulnud süüdistustele vastata, öeldes, et formaalne keeleteadus ei näe keelekompetentsi sugugi isoleerituna ülejäänud keele aspektidest ning on vägagi huvitatud sellest, kuidas keelt päriselt kasutatakse. Formalistide uurimisobjektiks on lihtsalt eraldi süsteem (universaalne grammatika ehk UG), mis on küll keele tegelikus kasutuses vaid üks faktor, kuid võimaldaks seda siiski osaliselt kirjeldada ja selgitada (Ludlow 2011: 87). Formalistide uurimisobjekt on niisiis UG, kuid uurimisküsimus ja -ala puudutavad keele tegelikku kasutust. Ludlow' seisukoht illustreerib formalistide tavapärasest vastust kõnealusel konfliktis: tegelikult pole lõhet olemas ja formaalne lähenemine ei peagi empiiriliseks muutuma. Kaks lähenemist lihtsalt uurivad erinevaid keeleaspekte ega vastandu üksteisele (vt ka Chomsky 2000). Probleem on loomulikult selles, et see vastus ei tegele ei sisuliste vastuoludega ega sellega, kuidas formaalsete teooriatega lisaks iseendale ka ülejäänud keeleteaduse rolli ja funktsiooni defineeritakse. Seega pole lõhe eitamine kõlanud vastaspoolele veenvalt.

Samuti ei toimu just sageli sisuliste vastuolude analüüsi funktsionaalsel suunal. Selle asemel võib aina rohkem märgata tendentsi, kus vastuseis, mis tegelikult seisneb mitmetes sisulistes punktides, koondatakse hoopiski ühteainsasse punkti, millega dominantsele ja prestiižsele, seni end ärritaval kombel „teaduslikuks lingvistikaks” kuulutanud formaalsele suunale on lihtne vastulööki anda. Selle koondumispunktiks on andmed ja nende kasutamine. Nõnda saab uueks populaarseks vastuseisuks „empiirilised lingvistid formaalsete vastu”. See, et tegu on diskursiivselt lihtsustatud, tegelikke vastuolusid eirava suunaga, väljendub seisukohtades, mis sarnanevad ülalpool toodud Manningi ja Schütze (1999) omadega, kus pannakse kokku meetodi kriitika ja uurimisobjekti definitsiooni kahtluse alla seadmine. Kui neil oleks probleem eelkõige formalistide metoodikaga, siis puudutaks arutelu seda, kuidas ja kas saab grammatilisushinnanguid kvantitatiivselt koguda ja statistiliselt analüüsida, millised eksperimendid oleks keelekeskuse uurimiseks kõige sobivamad ja miks on probleemne ehitada keelevõime teooria ainult mõnda keelt emakeelena kõnelevatelt „konsultantidelt” pärinevale infole. Ühtlasi moodustaks kvalitatiivsetele meetoditele rajatud ja filoloogias välja kasvav deskriptiivse keeleteaduse osa kvantitatiivsele (korpus)liikumisele igas mõttes loomulikuma vastandi, sest üldiselt nõustuvad need osapooled uurimisobjekti olemuses ja kummalegi ei tunduks, et võrreldakse liiga erinevaid objekte. Metoodika diskussioon on piisavalt oluline, et seda peaks saama pidada, ilma et peaks vaidlema selle üle, kas keele olemus seisneb selle ühiskondlikus või neuroloogilises aspektis, ja seda saaks teha ehk efektiivsemalt, kui uurimisobjektis nõustuvad koolkonnad arutaks omavahelisi metodoloogilisi traditsioone. Loomulikult ei võta kõik funktsionaalset suunda esindavad, kvantitatiivselt meelestatud lingvistid formaalse suuna uurimisobjektiga vastandumist nii lihtsal kujul omaks (Stefan Gries) ning paljud siiski mõistavad kvantitatiivset pööret eelkõige metoodikaliikumisenä, mitte terveid keeleteaduse harusid tasalülitava uurimisobjekti reformina.

Kahe diskussiooni segamisel on tõsised tagajärjed. Esiteks identifitseerivad paljud lingvistid end siiski formaalse ja empiirilise keeleteadlasena ega võta tõsiselt liikumist, mis süüdistab mitteempiirilises kõiki aju keelekeskusele orienteeritud uurimusi, jäädes seega meelega eemale olulistest metodoloogilistest aruteludest, mida see foorum peab. Teiseks, kui segunevad kaks diskussiooni, millest esimene küsib, mida peaks uurima, ja teine küsib, kuidas peaks uurima, ei jõua kumbki veenva lahenduseni, sest nagu järgnev peatükk selgitab, lingvistika valdkondadel ja lähenemistel puuduvad metoodilised piirangud. Kolmandaks kannatab aga keerise keskel oleva noore ja lootust andva kvantitatiivse liikumise võime tutvustada vajalikes kohtades uusi meetodeid ja näidata, kuidas neid sujuvalt vanasse teoreetilisse uurimistraditsiooni sobitada.

3. Metodoloogia ja keeleteaduse harud

Kahe konflikti puutepunkt seisneb andmekasutustraditsioonides, mis keeleteaduse harudes erinevad. Seega näitamaks, et arutelu uurimisobjektist ja arutelu metodoloogiast on kaks väga erinevat diskussiooni, mida ei tohiks omavahel segi ajada, tuleb lühidalt analüüsida, kas peale pelga traditsiooni on ka mõni reaalne põhjus, miks teatud tüüpi metoodika kasutamine on mõnes keeleteadusharus piiratud või raskendatud.

Kõige jämedakoelisem jaotus, mis jagab lingvistika valdkondadeks mitte inimkeele alusel (süntaks, morfoloogia, fonoloogia jms), vaid selle uurimisviiside lõikes, lahutab Nicholas Evansi ja Alan Denchi (2006) arvates keeleteaduse kolmeks: formaalne, deskriptiivne ja tüpoloogiline lingvistika. „Formaalne” ei tähista siinkohal üksnes generatiivseid lähenemisi (nagu eelnevas artikli osas on formaalse all kitsamalt silmas peetud), vaid kõiki raamistikke, mis otsivad keelest mustreid ning loovad neist süsteeme, millele saab omistada „mudeli” staatust. Selle jaotuse alusel kuuluvad sildi „formaalne” alla ka erinevad seaduspärasusi ja printsiipe otsivad kasutus põhised teooriad. Deskriptiivse keeleteaduse alla kuuluvad vaid need lähenemised, mille eesmärk on üldlingvistiliste mustrite asemel keskenduda mõne konkreetse keele nüansside kirjeldamisele. Tüpoloogide eesmärk on aga välja töötada ühtne keelekontseptsioonide raamistik, mille abil kirjeldada kõiki maailma keeli ja võrrelda erinevaid grammatikaid. Tüpoloogia allharuks peab see käsitlus ka universaale otsivat lingvistikat, mis on William Crofti (2002) sõnul tüpoloogia ja formaalse lingvistika puutepunkt.

Kui me aktsepteerime kirjeldatud kolme alamkategoriat (uurimisregistrit), siis näeme, et metodoloogilised piirangud neil kategooriatel puuduvad. Deskriptiivne keeleteadus on ehk kõige ilmselgemal viisil nii kvantitatiivne kui ka kvalitatiivne – Evans ja Dench (2006) toovad selle põhitööriistadena välja loomuliku keele korpused ja eksperimendid, sh vastuvõetavushinnangud. Tüpoloogiast mõeldakse ehk kõige sagedamini kui kvantitatiivsest alast, sest selle andmestiku moodustavad keskmisest suurem arv keeli ja see on orienteeritud keeletunnuste vaheliste seoste otsimisele. Balthasar Bickel (2007) eristab selle kõrval kvalitatiivset tüpoloogiat, mis otsib muutujaid keelestruk-

tuuride sarnasuste ja erinevuste kirjeldamiseks, ning teoreetilist tüpoloogiat, mis töötab välja tunnustevahelisi seoseid põhjendavaid teooriaid.

Formaalset keeleteadust seostatakse vastupidiselt tüpoloogiale eelkõige kvalitatiivse lähenemisega. Põhjuseks on arvukad varasemad tööd, mis võisid andmeid täielikult (ja teadlikult) eirates teha inimese keele võime kohta järeldusi vaid mõne üksiku ingliskeelse lause põhjal, ning grammatilisus- ja vastu võetavushinnangud, mille kogumisprotsess oli sageli informaalne ja juhuslik.

Olukord aga muutus. Kui varem oli mudeli ehitamisel tavaks kasutada vaid selliseid (tehislikke) näidislauseid, mille grammatilisus oli igale keele rääkijale ilmselge (ja mille valiidsust seega ei peagi kvantitatiivselt mõõtma) (Chomsky 1975), siis juba 1970-ndatel hakkas ilmema, et ilmselgete ja seega liiga lihtsate lausenäidete abil on raske ehitada süsteemi, mis suudaks ennustada kõiki võimalikke lauseid. Moodsatelt teoreetiliselt mudelitelt oodatakse aina enam võimet selgitada keele varieeruvust (Schütze 2016). Isegi Chomsky ei pea enam lausete grammatilisust süntaktilises analüüsis kuigi kõnekaks faktoriks, öeldes, et ka mittegrammatilised laused moodustavad osa keelest ja muuhulgas näiteks keele poeetilise traditsiooni (Chomsky 2017; vt ka Ferreira, Patson 2007 ja „piisavalt hea” teooria).

Selleks et veenduda, et ajaloolisele traditsioonile vaatamata puudub igasuguse laias mõttes formaalse, st mudeleid loova lingvistika ja kvantitatiivsete meetodite vahel konflikt, tuleb vaid vaadata modernset formaalset lingvistikat. Näiteks Lisa Pearl ja Jon Sprouse (2013) kasutavad muuhulgas hoidjakeele korpusandmeid, et uurida teoreetilisi süntaktilisi „saari” ning nende rolli UG-s. David Willis (2011), kes on koostanud mitu ajaloolise ja modernse kõmri keele korpus, uurib nende põhjal formaalseid diakroonilisi eituse muutumise mudeleid minimalismi raamistikus. Omapärase segu generatiivsest teooriast ja tüpoloogiast moodustab eelmisel aastal lõppenud ReCoS projekt, mis uuris tüpoloogilise metoodikaga minimalismis olulisi tunnuseid (ingl *features*), et asetada nad omavahelisse hierarhiasse ja uurida nende rolli UG-s (Biberauer, Roberts 2017). Näiteid leidub ka kvantitatiivse metodoloogia sobivusest teiste formaalsete ja generatiivsete raamistikega (nt Bod, Kaplan 1998; Pullum, Scholz 2002; Miyao jt 2004). Nõnda on selge, et kõiki allharusid – deskriptiivset, tüpoloogilist ja formaalset keeleteadust – on võimalik viljelda nii kvantitatiivselt kui ka kvalitatiivselt ja kuigi levinud vaatenurk klassifitseerib kitsamas mõttes formaalse keeleteaduse loomu poolest kvalitatiivseks (Levshina 2015; Manning, Schütze 1999), pole sisuliselt formaalset keeleteadust korrektne näha vastandumas empiirilisele lingvistikale.

4. Teooria roll ja keeleteaduse allharud

Nagu algul mainitud, valitseb segadus selle ümber, mis täpselt on see „teooria”, mida nähakse andmetele vastandumas. Sageli on sildiga „teoreetiline keeleteadus” tähistatud kitsamas mõttes formaalseid suundi keeleteaduses (Evans, Dench 2006). See tekitab terminoloogilise segaduse ja loob vastanduse, mille paikapidamatust ülaltoodud peatükid juba analüüsisid. Kuidas mõista aga teooria ja andmete konflikti juhul, kui selle all peetakse silmas üldisemat, üldistavat ja hüpoteesi püstitavat klassikalise teadusliku meetodi

osa? Järgnev osa analüüsib eelnevalt toodud kolme keeleteadusliku suundumuse suhet teooriaga ning argumenteerib, et empiirilise ja teoreetilise vastuolu ei ole vettpidav ka teooria tavalises tähenduses.

Esiteks on oluline uuesti märkida, et teoreetilist komponenti tuleks selguse huvides eristada formaalsest – selgus on lõppude lõpuks keeleteaduse sisemise teoreetilise struktuuri abstraktsetes aruteludes tähtis faktor. Kui „formaalne” tähistab mudeleid ehitavat, aju keelekeskust uurivat lingvistikaharu, siis „teoreetiline” tähistab abstraktsustaset, mis on definitsiooni järgi iga klassikalist teadusmeetodit järgiva teadustöö osa. William Whewelli (1847) klassikaline teadusfilosoofiline esse defineerib teooria kui teadaolevate faktide teise pooluse; teadmise, mis erineb faktidest vaid selle poolest, kui kindlad me selles oleme. Ebakindel teadmine ehk töö teooriaosa on teaduse aluskomponent – see tõuseb teadlase iseseisvatest ideedest, mitte vaid puhtalt vaatlusest (kuigi vaatlus võib anda aluse ideedele, nagu artikkel Darwini näitel illustreeris), mõtestab vanu ja ennustab uusi andmeid ning saab ise nende kaudu kinnitust. Just sellest duaalsusest – kinnitatud teadmise (uued andmed, eelnevad faktid) ja kinnitamata teadmise (teooria) üksteisega sidumisest sünnibki uus teadus (Whewell 1847, 1860). Nõnda on isegi terminoloogia loomine ja kasutamine sügavalt teoreetiline tegevus (Castellvi 2003), sundides teadlast kodeerima maailma objektiivsete, kuid abstraktsete mõistete kaudu, ning mõtlema nii sellele, kuidas seostuvad need mõisted omavahel, kui ka nende seosele vaatlusandmetega.

Sellest teooria definitsioonist võib järeldada, et ükski teaduse, sh lingvistika allharu, sõltumata suunast ja meetodist, ei saa läbi ilma teooriata. Esiteks on kõigil vaja kasutada terminoloogiat, seda andmetega suhestada ning tõenäoliselt kas tegeleda mõne eksisteeriva ebakindla teadmise kogumiga või luua selline uus kogum, et teadustöö tulemused ja olulisus lõplikult perspektiivi seada. Nõnda pole olemas „tooreid andmeid”, sest andmed saavad mõtestatud ja töö saab maailmale relevantseks läbi teooria läätse (Kuhn 1970). Loomulikult ei ole teooria roll kõigis töodes võrdne ei koguselt ega kasutusviisilt, varieerudes taustsüsteemi ja uurimisobjekti vahel. Nii nimetab Bickel (2007) „teoreetiliseks tüpoloogiks” ala, mis võtab uurimisobjektiks tüpoloogilise teooria enda, kuigi üldise definitsiooni järgi on kõik kolm Bickeli (2007) tüpoloogia kategooriat – kvantitatiivne, kvalitatiivne ja teoreetiline tüpoloogia – mingil määral teoreetilised.

Ka Evans ja Dench (2006) nendivad, et eristatud kolm lingvistika haru on paratamatult teoreetilised (samal ajal on vaid üks neist formaalne). Deskriptiivne haru teoretiseerib tahes-tahtmata mõne üksiku keele sisemise struktuuri üle, sest kasutab mõisteid ja püüab mingil määral siiski üldistusi teha, vaatlusandmeid seletada. Teooria rolli varieerumine deskriptiivsetes teadustöodes on tekitanud koolkondlikke vastuolusid, eelkõige selles, mis puudutab andmetele teoreetiliste seletuste andmise määra. Victor Yngve (1996) sõnul soovib osa deskriptiivse suuna esindajatest lingvistikat teaduslikumaks teha radikaalse andmeliikumise ja korpuslingvistika abil. Ta kritiseerib neid, öeldes, et ka korpuslingvistid toetuvad samale hägusele ja kultuurilisele terminoloogiale nagu ülejäänud keeleteadlased ning et üleüldse on veider luua teaduse allharu, mis väidab oma spetsialiteediks olevat andmed kui sellised. Yngve kriitika võib aga mitu kümnendit hiljem osutada tarbetuks, sest

korpuslingvistid ei näe end sugugi alati nii teooriavabana teooria klassikalises tähenduses. Näiteks Dylan Glynn (2014) ja Stefan Gries (2010) arvavad, et korpuslingvistika sobib eriti hästi kognitiivse lingvistikaga, ja Gries nendib, et korpuslingvistid peaksid miks-küsimusele vastamiseks praegusest veelgi tihedamini pöörduma kognitiivse lingvistika ja konstruktsioonigrammatika raamistike poole (Gries 2010). Kuigi see on näide teooria rolli suurendamisest laiemas mõttes formaalse raamistiku poole pöördumise abil, siis, nagu mainitud, on teooria mõiste laiem ja selle töösse kaasamine on üldise teadusliku meetodi seisukohast uurimuse läbiviimise eeldus, sest nii saavad vastuse küsimused „miks see juhtub?” ja „miks see oluline on?”.

Ka kolmest allharust teise, s.o tüpoloogilise töö käigus sünnivad tahes- tahtmata teooriad, mis panevad paika keeli ühendavad parameetrid ja põhjendavad nende omavahelist korrelatsiooni. Laiemas mõttes formaalsete lähennemiste teooriad kirjeldavad aga aju keelekeskuse sisemist ehitust ja loomust. Formaalsete teooriate teadusfilosoofilise pädevuse võib küsimärgi alla seada nende liiga detailne terminoloogia või metakeel, mis teeb tekstid teistes raamistiketes töötavatele keeleteadlastele ligipääsmatuks (Evans, Dench 2006). Väidete ligipääsmatust võib näha ühe falsifitseerimise alamkategoriana, sest see takistab teeside ja neist moodustuva teooria empiirilist tõestamist või falsifitseerimist: „Kirjeldus, mis ei kasuta teistele koolkondadele ja ajastutele arusaadavat terminoloogiat, riskib sellega, et ta näeb küll hea välja, kuid sellega pole midagi peale hakata” (Evans, Dench 2006: 7).

Kuigi paljud nõustuvad, et iga keeleteaduslik töö on klassikalises mõttes teoreetiline, nõustub teooria täpse olemuse osas vähem inimesi. Nõnda väidavad mõned koolkonnad, et oluline on üksteisest eristada kirjeldavaid ja selgitavaid teoreetilisi raamistikke ning vaatamata levinud arvamusele ei pea üks ja sama raamistik või seda kasutatav uurimus samaaegselt vastama milline- ja miks-küsimustele (Dryer 2006). Küll aga on mõlemad küsimused paratamatult teoreetilised, sest ka terminoloogia on teooria.

Ühtlasi on problemaatiline väita, et teooria ja andmete puhul peaks üks teisest hierarhiliselt üle olema. Hierarhia aetakse aga sageli ekslikult segi kronoloogiaga. See, kas teadustöö protsessis peaks andmed kronoloogiliselt teooriale eelnema või järgnema, ei tundu olevat teadusmeetodi perspektiivist väga oluline, sest edukates avastustes võib kohata mõlemat mudelit. Teadus- harud erinevad üksteisest niivõrd palju, et üht mustrit on tulutu nõuda ja sageli on võimatu nende täpset kronoloogiat analüüsida. Näiteks rajatakse hüpotees mingile (informaalsele) andmestiku vaatlusele või teisest allikate andmetele ja hüpoteesi põhjal väljatöötatud teooriat hakatakse uuesti põhjalikult ja empiirilisel tõestama (evolutsiooniteooria, relatiivsusteooria) (vastuväiteks vt Hint 2016). Kui need kaks aspekti on teadusmeetodi seisukohast võrdselt olulised, siis on nende mõningane kronoloogiline segunemine oodatav: andmete põhjal võib tekkida uusi ideid, ideede põhjal saab koguda uusi andmeid.

Erinevalt kronoloogiast võib aga kahe elemendi hierarhilise suhte eeldamine olla teadusprotsessile ohtlik. Peter Ludlow (2011) nendib, et kui andmed asetada tähtsuse poolest juhtrolli, siis võib keeleteadlane liiga kiiresti teooria kõrvale lükata, kui üksik andmestik sellele pisutki vastu räägib, kuigi tegelikult saab vastuolu põhjalikumalt uurida ja teooriat modifitseerida. Ludlow'

sõnul võiks teooriate arengus oodata n-ö Kuhni kitsenemist – protsess, mis on meditsiinilistes uurimustes väga produktiivseks osutunud ning mis käivitub, kui andmed on loodud teoriaga vastuolus. Teooriat kitsendatakse andmetesse „mahtuma”, tekitades nõnda teooria uue versiooni, mis kirjeldab eelnevast versioonist väiksemat osa maailmast. Nõnda on kitsenenud mitmete haiguste definitsioonid, mis loodi reeglina liigselt laiama (Ludlow 2011: 86). Teine andmekeskse töös peituv oht seisneb selles, et tahes-tahtmata ei arutata läbi kasutatavat teoreetilist lähenemist ega terminikasutust, lugejaid ei kutsuta teoreetilisi valikuid kriitiliselt hindama ega paigutata terminoloogiat ja uurimistööd laiemasse teadmiste raami.

Vastupidine on situatsioon, kus juhtroll antakse teooriale – siia alla kuuluvad kõik need nii generatiivsed kui ka funktsionalistlikud tööd, mis ei defineeri end selgelt kvalitatiivse ega kvantitatiivse, kuid mis loovad kirjeldusi, esitavad väiteid, üldistusi ja järeldusi informaaalselt või juhuslikult kogutud, sageli liiga väikse andmestiku põhjal. Sarnaselt ülalkirjeldatud stsenaariumiga, tekitab ka see meetod töö teaduslikkuse seisukohalt rea riske ning just nende arutamine peakski olema andmepöörde põhidiskussiooni osa. Peaks arutlema, kas ja kuidas ning millistes keeleteadusharudes peaks n-ö tugitoolilingvistika asemel senisest rohkem kvantitatiivseid meetodeid kasutama, miks on hädavajalik koguda nii kvantitatiivseid kui ka kvalitatiivseid andmeid korrektselt ja kui võrd peab iga teooria olema empiirilisel falsifitseeritav (näiteks stringiteooria näite perspektiivis). Andmepöörde eesmärk ei peaks olema süvendada lõhet eksisteerivate valdkondade vahel ning populariseerida mõnda teoreetilist paradigmat teiste arvelt. Käesoleva artikli eesmärk on rõhutada, et see pole ei vajalik ega produktiivne, sest empiiriline keeleteadus ei vastandu üheski mõttes oma olemuselt keeleteaduse teoreetilisele poolele, ükskõik kas sildi „teoreetiline” all mõeldakse generatiivseid lähenemisi või klassikalise teadusliku meetodi üldistavat-mõtestavat osa.

5. Teooriavaba lingvistika

Vastuolu nägemine teadusfilosoofiliselt laia teooria kontseptsiooni ja keeleanemete vahel on ebatavaline, aga mitte ennekuulmatu. Martin Haspelmath (2010a, 2010b, 2016, 2017a) esindab keskmisest radikaalsemat keeleteadusuunda, mis usub, et kuna keeleteaduse peavool on ummikusse jooksnud, siis tuleb loobuda kõigist teoreetilistest raamistikest ning ümber defineerida viisid, kuidas kahe või enama keele vahele saab üldistusi tõmmata. See ei tähenda ainult kriitikat formalistide suunal, vaid puudutab ka lihtsaid terminoloogiakogumikke, nagu „Lingvistika baasteooria” (Dixon 2009; vt ka Dixon 1997; Dryer 2006), mis arendavad keelte kirjeldamise ja võrdlemise tarvis termineid.

Haspelmathi seisukohast on raamistikest ja teadusliku eeldusega terminitest loobumine vajalik, sest kõik teoreetilised raamistikud on definitsiooni järgi spekulatiivsed ja enamasti kinnitamatud, tehes võimatuks „keelte kirjeldamise omaenda raamis”. Iga raamistik, isegi kui selle vundament pole haruldasel kombel mõne suure keelkonna poole kaldu, võimaldab keelt vaadelda vaid kallutatud perspektiivis. Raamistikukeskne uurimistraditsioon on üht-

lasi põhjustanud olukorra, kus keeleteadlased kasutavad sama terminit keele kirjeldamiseks ja teiste keeltega võrdlemiseks. Need kaks – võrdlevad mõisted ja kirjeldavad kategooriad – tuleks aga Haspelmathi väitel lahus hoida ning meeles pidada, et kumbagi neist ei saa pidada loomulikeks mõisteteks (ingl *natural types*, nt tuberkuloos või argoon). Lingvistilised võrdlevad mõisted on Haspelmathi järgi hoopis jämedakoelised, mitteteaduslikud iseloomustused, nagu millegi nimetamine arengumaaks majandusteaduses. Niisiis ei ole neist võimalik järeldada ühegi nende poolt kirjeldatava elemendi tegelikke omadusi ning keelt ei ole võimalik uurida miks-küsimusele ehitatud teadusliku lähene-misega. (Haspelmath 2010b, 2017a)

Kuigi Haspelmath viitab pidevalt, et grammatilised süsteemid on ajaloo-liselt juhuslikud (2010a, 2017a), siis tema raamistikevaba keeleuurimist ei suuna siiski empiiriline veendumus, et keelte struktuuri ei mõjuta ajus fiksee-ritud printsiibid ning et keelte vahel seega mustrid puuduvad. Sotsiaalmeedias toimunud väitluses (Haspelmath 2017b), milles formaalne lingvist Richard Larson väitis, et teoriavaba teadus on *non sequitor*, vastas Haspelmath: „Ma nõustun, et mõned asjad [keele aspektid] on tõenäoliselt ajus fikseeritud, kuid ma ei näe, kuidas tavapärane generatiivse grammatika praktiseerimine aitaks meil neid mõista [---] Kuidas me saame teada, mis on UG-s?” Ilmneb, et Haspelmathi teoriavaba suund on empiirilise seoste puudumises veendumise asemel motiveeritud metodoloogilistest probleemidest, uskudes, et kuigi keel on sisemiselt süsteemne ja huvitav uurimisobjekt, siis praegused meetodid ja teooriad ei vii lingvistilise tõeni, vaid ainult suure hulga seosetute seisukohtadeni. Viidates, et ratsionaalsete mõttemehhanismide praktiseerimine ei ole piisav selleks, et paljastada maailmas peituvaid printsiipe, sarnaneb Haspelmath omapärasel kombel XVIII sajandi saksa filosoofilise liikumisega, mille esindajad Kant, Fichte ja Hegel arvasid, et inimese kogutud teadmised ei saa iial olla objektiivsed, sest nad on kõik loodud maailma tajumise kaudu, mis on oma olemuselt subjektiivne protsess. Tegelik reaalsus on seega tume ja tundmatu ala (Whewell 1847). Kuhni väitel kaotas mõttesuund populaarsuse, kui levis teaduslik meetod ja usk maailma saladuste paljastamiseks empiirili-selt kinnitatud meetodite ja fokuseeritud analüüsi abil (Kuhn 1970: 170–173).

6. Järeldused

Kvantitatiivsest pöördest tekkivat lainetust jälgides tundub, et kõik pole kulgenud päris ootuspäraselt ning vastasseis, mida nähakse empiirilise ja teoreetilise suuna vahel, tekitab küsimusi. Kui mõista teooriat klassikalise teadusmeetodi osana, st ebakindla faktikogumina, mis mõtestab ja ennus-tab andmeid, siis on konflikt ootamatu ja arusaamatu. Nimelt on need kaks poolust – teooria ja andmed – haruldaset ühtse, poolelise sajandi vanuse teadusmeetodi mudeli alustaladeks, mille kasulikkuses kahtlejaid on vähe. Kuidas on võimalik, et selline pealtnäha ohutu tööriist nagu statistiline ana-lüüs raputab tervet lääne kultuuri teaduse alustala? Vastuseks on, et see ei raputagi. Sarnaselt geograafia ja geenitehnoloogiaga on kvantitatiivses keele-teaduses antud konflikti tunnetamine ebatavaline ning enamik näeb teooriat siiski teadustöö osana, isegi kui see seisneb vaid terminoloogia ja võrreldavate

mõistete loomises ning miks-küsimuse asemel milline-küsimusele keskendumises. Tuleb lihtsalt silmas pidada, et andmestike töösse toomine ei vähendaks teoreetilise arutelu osakaalu ega sügavust. Sarnaselt geograafia näitega, kus kvantitatiivsete meetodite laiem levik viis hoopis põhjalikumate teoreetiliste käsitlusteni, jõuavad ka Małgorzata Fabiszak, Martin Hilpert ja Karolina Krawczak järeldusele, et vaatamata andmete tüübile „peame alati püüdma uurimistsükli lõpule viia, näidates, kuidas meie tulemused suhestuvad relevantsete teoreetiliste küsimustega” (Fabiszak jt 2016: 350).

Kui aga teooriat mõista „formaalse lingvistika” kitsamas tähenduses, st reeglina generatiivsete lähenemiste sünonüümina, siis ilmneb, et teooria ja andmete vastandumine on intensiivne, kuid mitte kuigi selge. Artikkel argumenteeris, et ükski keeleteaduse haru, sh generatiivne, pole metodoloogiliselt piiratud. Nõnda pole vastandumine sellisel kujul sisuline diskussioon, vaid hoopis omapärane sulam uuest metoodilisest diskussioonist ja 60 aasta vanusest kasutuspõhiste (funktsionaalse, kognitiivse jms) ja formaalsete-generatiivsete teooriate vastasseisust. Eraldiseisvalt on aga tegu kahe olulise aruteluga. Esimene diskussioon seisneks ideaalis sisulises arutelus sellest, millised on tehtavate keeleteaduste metodoloogilised probleemid ning millistelt keeleteaduse harudelt võiks oodata metodoloogilist „uuenduskuuri”, millised probleemid sellega kaasnevad ning milliseid see lahendab. Teine, koolkondlik diskussioon arutab aga hoopis, kas keel kui selline on kultuuriline või looduslik uurimisobjekt. Selle vastasseisu kahepoolne arutamine on keeruline, sest kui kasutuspõhise-deskriptiivse suuna esindajad rõhutavad kultuuri rolli inimkeele aspektide, sh grammatika kujundamisel, siis formalistid tavatsevad väita, et konflikti polegi olemas, sest kaks haru lihtsalt uurivad erinevaid keele tahke. Ehk on see just põhjustatud kommunikatiivsest „seinast”, et koolkondlikku vastasseisu jätkatakse kohati andmepöörde sildi all, mille käigus on kriitikanool suunatud justkui mitteempiirilistele meetoditele, kuid lõpuks keskendutakse ikka formalistide uurimisobjektile ja keelekompetentsi keelekasutusest eristamise teemaikale.

Käesoleva artikli eesmärk ei olnud analüüsida süvitsi kumbagi kahest nimetatud vastasseisust, vaid märkida, et kohati võib märgata kahe segunemist ning et see tendents on keeleteaduse arengu seisukohast pigem kahjulik kui produktiivne. Artikkel näitas, et teooria (ei klassikalises teadusliku meetodi ega ka formaalse lingvistika tähenduses) ei ole metoodilistel uuendustel ühelgi sisulisel moel jalus. Kvantitatiivsel pöördel on veel pidada mitmeid olulisi diskussioone, millest ükski ei puuduta formaalset lingvistikat, kuid mis on keeleteaduse arengu seisukohast äärmiselt olulised. Vastamata on näiteks küsimused sellest, mis on uue meetodi eelised ja nõrkused ning kuidas see mõjutab inimteaduste tuumaks peetava tähenduse otsimist ning kas ja kuidas on see siiski konfliktis korrektselt kasutatud kvalitatiivsete meetoditega. See, et teaduse tegemiseks on tarvis esiteks meetodit ja teiseks teooriat, on pigem üldine tõde kui vastamist vajav küsimus, sest paljud nõustuvad (ja nagu Gries 2010 on märkinud), et tarvis on ka kuidas- ja miks-küsimusi, ükskõik kui süvitsi töö statistilises analüüsis läheb. Kaks diskussiooni on niisiis eraldi võttes olulised, kuid ei muuda fakti, et teooria ja andmed on kaks teadust kandvat sammast – kumbki ei saa ilma teiseta mõtestatud.

Kirjandus

- Anderson, Chris 2008. The end of theory: The data deluge makes the scientific method obsolete. – *Wired Online*. <https://www.wired.com/2008/06/pb-theory/> (10. II 2018).
- Biberauer, Theresa, Roberts, Ian 2017. Parameter setting. – *The Cambridge Handbook of Historical Syntax*. Toim Adan Ledgeway, I. Roberts. Cambridge: Cambridge University Press, lk 134–162.
- Bickel, Balthasar 2007. Typology in the 21st century: Major current developments. – *Linguistic Typology*, nr 11, lk 239–251.
- Bod, Rens, Kaplan, Ronald 1998. A probabilistic corpus-driven model for lexical-functional analysis. – *Proceedings of the 17th International Conference on Computational Linguistics*, kd 1. Montreal, Quebec, Canada, August 10–14, 1998. Stroudsburg: Association for Computational Linguistics, lk 145–151.
- Burton, Ian 1963. The quantitative revolution and theoretical geography. – *Canadian Geographer*, kd 7, nr 4, lk 151–162.
- Castellvi, M. Teresa Cabre 2003. Theories of terminology: Their description, prescription and explanation. – *Terminology*, kd 9, nr 2, lk 163–199.
- Chomsky, Noam 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
- Chomsky, Noam 1975. *The Logical Structure of Linguistic Theory*. New York: Plenum.
- Chomsky, Noam 1986. *Knowledge of Language*. New York: Praeger.
- Chomsky, Noam 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.
- Chomsky, Noam 2002. *On Nature and Language*. Cambridge: Cambridge University Press.
- Chomsky, Noam 2017. Some puzzling foundational issues. – *Ettekanne. Generative Linguistics in the 21st Century: The Evidence and the Rhetoric*. University of Reading, 11. V 2017.
- Croft, William 1991. *Syntactic Categories and Grammatical Relations*. Chicago–London: The University of Chicago Press.
- Croft, William 2002. *Typology and Universals*. Cambridge: Cambridge University Press.
- Darwin, Charles, Seward, Anna 1903. *More Letters of Charles Darwin*. London: John Murray.
- Dawid, Richard 2015. Non-empirical confirmation. – *Ettekanne. Why Trust a Theory?* Ludwig Maximilian University, 7.–9. XII 2015. <https://www.whyltrusttheory2015.philosophie.uni-muenchen.de/program/index.html#dawid> (10. II 2018).
- Dixon, Robert M. W. 1997. *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Dixon, Robert M. W. 2009. *Basic Linguistic Theory*. Oxford: Oxford University Press.
- Dryer, Matthew 2006. Descriptive theories, explanatory theories, and basic linguistics. – *Catching Language: The Standing Challenge of Grammar Writing*. Toim Felix Ameka, Alan Dench, Nicholas Evans. Berlin: Mouton De Gruyter, lk 207–235.

- Evans, Nicholas, Dench, Alan 2006. Catching language. – *Catching Language: The Standing Challenge of Grammar Writing*. Toim Felix Ameka, A. Dench, N. Evans. Berlin: Mouton De Gruyter, lk 1–41.
- Everett, Daniel 2012. *Language: the Cultural Tool*. New York: Pantheon.
- Fabiszak, Małgorzata, Hilpert, Martin, Krawczak, Karolina 2016. Usage-based cognitive-functional linguistics: From theory to method and back again. – *Folia Linguistica*, kd 50, nr 2, lk 345–353.
- Ferreira, Fernanda, Patson, Nikole D. 2007. The „good enough” approach to language comprehension. – *Language and Linguistics Compass*, kd 1, nr 1–2, lk 78–83.
- Glynn, Dylan 2014. Polysemy and synonymy. – *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*. Toim D. Glynn, Justyna A. Robinson. Amsterdam: John Benjamins, lk 7–38.
- Gries, Stefan 2010. Corpus linguistics and theoretical linguistics. A love-hate relationship? Not necessarily. – *International Journal of Corpus Linguistics*, kd 15, nr 3, lk 327–343.
- Gries, Stefan 2015. Quantitative linguistics. – *International Encyclopedia of the Social and Behavioral Sciences*, kd 19. Toim James Wright. Amsterdam: Elsevier, lk 725–732.
- Gries, Stefan 2017. Syntactic alternation research. Taking stock and some suggestions for the future. – *Belgian Journal of Linguistics*, nr 31, lk 7–27.
- Haspelmath, Martin 2010a. Framework-free grammatical theory. – *The Oxford Handbook of Linguistic Analysis*. Toim Bernd Heine, Heiko Narrog. Oxford: Oxford University Press, lk 375–402.
- Haspelmath, Martin 2010b. Comparative concepts and descriptive categories in crosslinguistic studies. – *Language*, kd 86, nr 3, lk 663–687.
- Haspelmath, Martin 2016. The challenge of making language description and comparison mutually beneficial. – *Linguistic Typology*, kd 20, nr 2, lk 299–303.
- Haspelmath, Martin 2017a. How comparative concepts and descriptive linguistic categories are different (draft). – Zenodo. <http://doi.org/10.5281/zenodo.570000> (10. II 2018).
- Haspelmath, Martin 2017b. Do We Need a „Framework” for Syntax? A Conversation between Richard Larson and Martin Haspelmath. <https://dlc.hypotheses.org/966> (10. II 2018).
- Hauser, Marc D., Chomsky, Noam, Fitch, W. Tecumseh 2002. The faculty of language: What is it, who has it, and how did it evolve? – *Science*, kd 298, nr 5598, lk 1569–1579.
- Hint, Mati 2016. Mõõtmised ei loo teooriat. – *Keel ja Kirjandus*, nr 8–9, lk 627–637.
- Humboldt, Wilhelm von 1836. *Über die Verschiedenheit des Menschlichen Sprachenbaues*. Berlin. (Chomsky 1965 kaudu.)
- Koster, Jan 2006. Is linguistics a natural science? – *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdijk*. Toim Hans Broekhuis, Norbert Corver, Riny Huijbregts, Ursula Kleinhenz, J. Koster. Berlin: Walter de Gruyter, lk 350–358.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
- Levshina, Natalia 2015. *How to do Linguistics with R*. Amsterdam: John Benjamins Publishing Company.

- Ludlow, Peter 2011. *The Philosophy of Generative Grammar*. Oxford: Oxford University Press.
- Manning, Christopher D., Schütze, Hinrich 1999. *Foundations of Statistical Natural Language Processing*. Cambridge–London: MIT Press.
- Mesthrie, Rajend 2012. *The Cambridge Handbook of Sociolinguistics*. Cambridge: Cambridge University Press.
- Miyao, Yusuke, Ninomiya, Takashi, Tsujii, Jun'ichi 2004. Corpus-oriented grammar development for acquiring a head-driven phrase structure grammar from the Penn Treebank. – *Natural Language Processing – IJCNLP 2004. Lecture Notes in Computer Science 3248*. Toim K. Y. Su, J. Tsujii, J. H. Lee, O.Y. Kwong. Berlin–Heidelberg: Springer, lk 684–694.
- Pearl, Lisa, Sprouse, Jon 2013. Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. – *Language Acquisition*, kd 20, nr 1, lk 23–68.
- Pinker, Steven 1994. *The Language Instinct*. New York: William Morrow and Company.
- Popper, Karl 1959. *The Logic of Scientific Discovery*. London–New York: Routledge.
- Pullum, Geoffrey K., Scholz, Barbara C. 2002. Empirical assessment of stimulus poverty arguments. – *The Linguistic Review*, nr 19, lk 9–50.
- Romaine, Suzanne 1994. *Language in Society: An Introduction to Sociolinguistics*. Oxford–New York: Oxford University Press.
- Sapir, Edward 1921. *Language. An Introduction to the Study of Speech*. New York: Harcourt, Brace and company.
- Satava, Richard M. 2005. The scientific method is dead – long live the (new) scientific method. – *Surgical Innovation*, kd 12, nr 2, lk 173–176.
- Schütze, Carson T. 2016. *The Empirical Base of Linguistics*. Berlin: Language Science Press.
- Shreeve, James 2004. Craig Venter's epic voyage to redefine the origin of species. – *Wired Online*. <https://www.wired.com/2004/08/venter/> (10. II 2018).
- Van Herk, Gerard 2012. *What Is Sociolinguistics?* Cambridge–Oxford: Wiley.
- Whewell, William 1847. *The Philosophy of Inductive Sciences, Founded upon Their History*. London: J. W. Parker and Son.
- Whewell, William 1860. *On the Philosophy of Discovery, Chapters Historical and Critical*. London: J. W. Parker and Son.
- Willis, David 2011. A minimalist approach to Jespersen's cycle in Welsh. – *Grammatical Change: Origins, Nature, Outcomes*. Toim Dianne Jonas, John Whitman, Andrew Garrett. Oxford: Oxford University Press, lk 93–119.
- Yngve, Victor H. 1996. *From Grammar to Science. New Foundations for General Linguistics*. Amsterdam: John Benjamins.

Mari Aigro (snd 1990), doktorant, Tartu Ülikool, mari.aigro@ut.ee

Quantitative shift in linguistics – an old discussion revived

Keywords: quantitative methods, theoretical linguistics, formal linguistics, philosophy of science, scientific method

The new trend of incorporating a large amount of quantitative data into linguistic research has sparked a wave of scientific positivism, both among researchers as well as the public. In addition to the inevitable comparison and conflict between qualitative and quantitative methods it has raised questions about the relationship between data-based and what are commonly called theoretical approaches to language science, the former casting doubt on the validity of the latter. However, on a closer look it appears that the new forum, regarding itself as methodologically oriented, can sometimes shift its focus from the question of which methodology should be used and when, to the much older divide in linguistics, between the formal and usage-based approaches. The discussion then becomes focussed on the nature of language as a research object. This is why the question asked is: “What will become of theory?” rather than “What will become of non-empirical linguistics?”

The article proposes that there are two different definitions one can assign to the concept of “theory” – one can talk about the traditional part of the scientific model, i.e. the “why?” and “how?” question of every research project, or one can use it as a synonym for formal linguistics. The article claims that neither of them is in any conflict with empirical approaches. As a step in the scientific method it is in fact by definition unavoidable, because even without explicitly speculating about why a tendency occurs in data, a researcher is still forced to use theoretically based terminology. Conflict between formal approaches and data is equally misguided, because all branches of linguistics – descriptive, typological and formal – are capable of being, and often are, quantitatively empirical. Regardless of the amount of data, the role of theory cannot become marginalized, because one cannot exist without the other – theory provides data with meaning and data validates theory.

Mari Aigro (b. 1990), PhD Student, University of Tartu, mari.aigro@ut.ee