Väitekiri algupärase ja tõlgitud keele eristatavusest

> Marju Taukar. Algupäraste ja tõlgitud eestikeelsete tekstide sarnasused ja erinevused. (Humanitaarteaduste dissertatsioonid 56.) Tallinn: Tallinna Ülikool, 2020. 125 lk. [Võrguväljaanne.]

Marju Taukari doktoritöös uuritakse algupäraste ja tõlgitud eestikeelsete tekstide sarnasusi ja erinevusi. Töö ajendiks on laialt levinud subjektiivne arvamus, et tõlkeprotsessi tulemusel sündinud keel erineb mingite tunnuste, nt vormi või stiili poolest algupärasest keelest. Enamasti põhinevad sellised arvamused lugejate isiklikel kogemustel ja eelarvamustel, ent empiirilisi uurimusi, mis vaatleksid tõlgitud tekste erinevatest vaatenurkadest ja kasutaksid ka kvantitatiivseid meetodeid, Eestis varem tehtud ei ole. Marju Taukari väitekiri täidabki eesti tõlkelise keele uurimisel valitsenud lünga, seades eesmärgiks leida tekstides keelelisi tunnuseid, mis viitavad tõlkelisusele ja võimaldavad määrata, mille poolest erineb tõlgitud tekst algupärasest eestikeelsest tekstist.

Töö põhiosa koosneb neljast peatükist, millest kõige originaalsem ja mahukam on neljandas peatükis esitatud kvantitatiivne korpuspõhine uuring tõlkelise ja algupärase keele eristatavusest masinõppe abil. Töö alguses antakse ülevaade tõlkelise keele olemusest ja varasematest kvantitatiivsetest tõlkelise keele uurimustest maailmas. Viimastel aastakümnetel on tõlketeaduses tehtud mitmeid korpuspõhiseid uurimusi, kus proovitakse teada saada, mis iseloomustab tõlkeprotsessis sündinud keelt. Varasemad tõlkeuurimused keskendusid sageli väiksemale hulgale konkreetsetele keeleaspektidele, nt universaalsetele keelenditele või kollokatsioonide erinevustele. Hilisemad, kvantitatiivsed korpuseuurimused püüavad läheneda tõlkelisele keelele terviklikumalt ning põhinevad sageli masinõppelistel meetoditel. Selliste uuringute eesmärk on tuvastada ja klassifitseerida tunnused, mille abil saab konstrueerida statistilisi mudeleid tõlkelise keele selgitamiseks. Doktoritöös kirjeldatud uurimuste näitel saab ülevaate sellest, kuidas kvantitatiivsed uuringud on maailmas aja jooksul arenenud ja milliseid tunnuseid on kõige sagedamini tekstide klassifitseerimisel kasutatud.

Teine peatükk keskendub tõlkeprotsessile ja jälgib ühelt poolt tõlkija tegevust tõlkeprotsessis ja teiselt poolt toimetaja rolli tõlkelise keele kujunemisel. Tõlkeprotsessi uurimisel on alates 1980. aastatest kasutatud erinevaid meetodeid. Kõige esimesed olid valjusti mõtlemise protokollid, mille puhul tõlkija sõnastab tõlkimisprotsessi käigus tekkinud probleemid ja nende lahendused valjusti välja hääldades. Tänapäeval kasutatakse ekraanipildi salvestamist, mis näitab ära tõlkija konkreetsed tegevused tõlkimise ajal, aga

üha enam ka klaviatuurisalvestamist ning tõlkija silmade liikumise ja pupillide suuruse jälgimist. Viimased annavad ka teavet tõlkimisel toimuvate kognitiivsete protsesside kohta. Tõlkimisel toimub samaaegselt mitu erinevat protsessi ja tähelepanu jaotumine nende vahel tähendab tõlkijale suurt kognitiivset koormust. Tõlkeprotsessi uuringute abil on põhjendatud tõlkija keelevalikuid, nt miks kasutatakse tõlkides lähtetekstiga vormiliselt või tähenduslikult sarnaseid keelelisi elemente, isegi kui nad on valed, või kuidas toimub tõlkimise ajal monitoorimisprotsess, st tõlke korrigeerimine. Kuna lõpliku tõlketeksti valmimisel on peale tõlkija väga suur roll toimetaja tegevusel, siis tegeletakse doktoritöös ka küsimusega, kui oluliselt muutub tõlge pärast toimetamist ja milliseid parandusi teeb tõlkes toimetaja. Doktoritöö raames viidi läbi kvalitatiivne uurimus, kus võrreldi seitsme tõlke toimetamata ja toimetatud versiooni, millele järgnes kvantitatiivne korpuseuuring viie tekstipaariga. Kuigi kvalitatiivne analüüs näitas, et toimetamisel tehakse enim parandusi sõnavara osas, ei ole need siiski nii ulatuslikud, et kvantitatiivses analüüsis tugevalt esile kerkiksid. Uuringu tulemusel selgus, et toimetatud tekstides oli kokkuvõttes vähem sõnu, kuid pisut rohkem oli üks kord esinevaid sõnu. Tekstisõnade ja sõnatüüpide suhe seevastu palju ei erinenud. Seega võib uurimuse tulemusel öelda, et toimetamine ei mõjuta tõlgitud teksti kvantitatiivselt mõõdetavaid omadusi kuigi olulisel määral.

Väitekirja kolmas peatükk on pühendatud aastatel 2011–2013 läbi viidud tajuuuringule, mille käigus selgitati välja, kas lugejad suudavad eristada tõlgitud tekste algupärastest, ja kui suudavad, siis milliste tunnuste alusel. Tõlkekriitikas tuuakse sageli välja, et tõlke muudab halvaks mõni valesti tõlgitud sõna või ebaloomulik sõnastus, kuid tõlgitud tekste lugedes ei osata sageli konkreetselt selgitada, miks mõni tekst mõjub tõlgituna ja teine mitte. Seda teemat on mitmel pool maailmas käsitletud erineva vaatenurga alt, kuid Eestis sellised empiirilised uurimused puuduvad. Marju Taukari töö annab esimese sissevaate sellesse, kuidas lugejad tõlgitud teksti tajuvad ja milliseid omadusi nad tõlgitud tekstile omistavad. Uuringu tulemusena selgus, et tõlgitud ja algupärased tekstid ei olnud vastajate jaoks kergesti eristatavad ehk siis tõlkeline keel ei ole nii spetsiifiline, et lugeja tõlgitud teksti alati ära tunneks. Kui tekst liigitati tõlkeks, siis arvati tekstil midagi viga olevat, see tähendab, et tõlget eristab algupärasest tekstist kõrvalekaldumine lugeja poolt õigeks peetud normist. Tõlkest räägiti pigem negatiivses toonis, mis kinnitab levinud, peamiselt isiklikke eelistusi ja subjektiivseid hinnanguid peegeldavat eelhoiakut tõlgete suhtes. Seevastu kui tekst liigitati algupäraseks tekstiks, siis arvati selle sõnavara olevat rikkalikum, leidlikum ja eestipärasem kui tõlgitud teksti sõnavara. Kuna tõlgete ja algupäraste tekstide eristamine polnud katses osalenud inimeste abil kuigi tulemuslik, suunas see doktoritöö autorit arvutuslike meetodite poole, et teha kindlaks, milliseid tulemusi annab arvutipõhine tekstide klassifitseerimine.

Doktoritöö neljandas peatükis soovibki autor teada saada, kas masinõppe klassifitseerimisalgoritm suudab tõlgitud ja algupärast teksti eristada ning milliste tunnuste abil suudab see teksti klassifitseerida. Analüüs põhineb märkimisväärselt ulatuslikul algupäraseid ja tõlgitud eestikeelseid ilukirjandustekste sisaldaval korpusel (13 291 381 tekstisõna), milletaolist ei ole Eesti tõlketeaduses varem kasutatud. Klassifitseerimistunnustena on kasutatud korpuse kõige sagedasemaid sõnu, lemmasid, tähemärke ja nende järjendeid, mis esinesid vähemalt 95%-s korpuse tekstidest, samuti sõnaliike ja sõnaliigijärjendeid. Klassifitseerimiskatse

osutus väga edukaks: vaid ühe tunnuse klassifitseerimistäpsus jäi alla 75%. Parimateks tunnusteks (klassifitseerimistäpsus 94–98%) osutusid üksiksõnad ja -lemmad ning pikemad tähemärgijärjendid. See tulemus viitab sellele, et korpuses olnud algupäraseid ja tõlgitud tekste eristas kõige paremini sagedaste sõnade jaotus. Veidi vähem edukad tunnused olid lühemad tähemärgijärjendid, sõnaliigid ja sõnaliigijärjendid, mis võiksid osutada algupäraste ja tõlgitud tekstide grammatilistele erinevustele. Vähem edukad tunnused olid ka sõnade ja lemmade järjendid, mis võiksid viidata erinevate kollokatsioonide kasutamisele tõlgetes ja algupärandites.

Väitekirja keskmes olev klassifitseerimiskatse tõestab, et algupärased ja tõlgitud tekstid on masinõppemeetoditega üksteisest eristatavad, kuid see tulemus ei ole kuigi lihtsalt kvalitatiivselt tõlgendatav. Et mõista, mille poolest algupärased ja tõlgitud eestikeelsed tekstid täpsemalt erinevad, on töös vaadeldud ka seda, millised sõnad ja lemmad on eelistatud algupärastes tekstides ja millised tõlgetes. Silmatorkavaim tulemus on rõhumäärsõnade (nagu *vist*, *ju*, *küll*, *hoopis*, *ometi*) rohkus algupärastele tekstidele iseloomulike sõnade loendi tipus. Need on sõnad, mis paljude korpuses olevate tõlketekstide lähtekeeltes puuduvad. Seega võib nende vähesust tõlgetes pidada lähteteksti interferentsi ilminguks. Kergemini tõlgendatavat infot algupäraste ja tõlgitud tekstide erinevuste kohta annab ka kahe alamkorpuse sõnaliikide võrdlus. Kõige selgemini tuleb esile asesõnade suurem hulk tõlketekstides, mida taas võiks seostada interferentsiga: erinevalt paljude korpuses olevate tõlketekstide lähtekeeltest on eesti keelele grammatiliselt omane asesõnade väljajätt.

Peale algupärasusest või tõlkelisusest tulenevate joonte võivad tekstide grupeerumises sagedaste sõnade jaotuse põhjal mängida rolli ka stilistilised tegurid: korpusetekstide klasteranalüüs grupeeris lisaks algupärastele ja tõlgitud tekstidele ka samade autorite ja/või tõlkijate tekstid (kuid vähemal määral samast lähtekeelest tõlgitud tekstid).

Kokkuvõttes avab Marju Taukari väitekiri Eesti tõlketeaduses uudse eksperimentaalse ja kvantitatiivse uurimissuuna ning demonstreerib selle potentsiaali erinevat tüüpi uurimisküsimustele vastamisel. Nagu autor isegi töö kokkuvõttes ütleb, on uurimuse edasiarendamise võimalused avarad: näiteks pakuks huvi täpsemate algupäraseid ja tõlgitud tekste eristavate tunnuste väljaselgitamine, sealhulgas tõlketeoorial põhinevate hüpoteeside kontrollimise teel, ning tõlgitud teksti uurimise sidumine tõlkimise olemuse ja tõlkeprotsessi uurimisega. Kahjuks ahendavad korpuspõhise tõlkeuurimise võimalusi autoriõigustega kaitstud tekstide kasutamisega seotud juriidilised piirangud, mis teeb töö empiirilise materjali veelgi tähelepanuväärsemaks. Kuna tõlkelist eesti keelt sellises mahus varem uuritud ei ole, siis puudus seni teadmine, kas ja kuidas erineb tõlgitud tekst algupärasest tekstist. Uuringu tulemused aitavad paremini mõista tõlkimist ja selle käigus toimuvaid protsesse ning tõlkija teadlike või alateadlike valikute tagamaid. Väitekiri annab väärtuslikku mõtlemisainet nii tõlkijatele kui ka tõlkijate koolitajatele ning aitab loodetavasti kaasa nii tõlkeõpetuse kui ka tõlkija elukutse maine parandamisele.

TERJE LOOGUS, HEETE SAHKAI