

Keeletoimetamise tulevikust

Kas tehnika ja tehisintellekti võidukäigus jääb kohta ka toimetamisele?



MAREK DREVING, Euroopa Parlamendi
eesti kirjaliku tõlke üksuse juhataja

Iga ennustus põhineb paratamatult senistel teadmistel ja juba aasta või paari pärast võib tulevik paista hoopis teistsugune kui praegu. Eksimisevõimalus on aga alati suur. Töötanud viimased 16 aastat Euroopa Liidu (EL) institutsioonides, täpsemalt Euroopa Parlamendis eesti keelde tõlkimise vallas, ei söanda ma oletada, kuidas võiks kulgeda keeletoimetamise areng Eestis näiteks raamatute kirjastamisel, meedias või seaduste õiguskeele toimetamisel.

Suure osa toimetajate töömahust moodustab aga tõlgete toimetamine ning ELi institutsioonide eesti tõlketeenistuste tegevus ei ole midagi eraldiseisvat muust eesti keelde tõlkimisest ja tõlketekstide toimetamisest või laiemalt keele ja keele- tehnoloogia alal toimuvast, vaid on selle märkimisväärse suurusega osa. Mõnes asjas – näiteks tehniliselt, aga ka terminoloogiatöös – võime end ka avangardi hulka arvata. Nii saab meie kogemuse põhjal ehk ka mõne üldistuse teha.

KEELETOIMETAMINE ELis

ELi institutsioonides töötab praegu ligikaudu poolteistsada eesti keelde tõlkijat. Oma tõlketeenistus on kõikides ELi institutsioonides: kõige suurem Euroopa Komisjonis (ligikaudu 50 tõlkijat), mõnevõrra väiksemad Euroopa Parlamendis, Euroopa Liidu Nõukogus ja Euroopa Kohtus (igapähe 25 tõlkija ringis). Oma tõlketeenistus on ka nõuandvatel komiteedel Brüsselis, samuti Euroopa Kontrollikojal.

2018. aastal tõlkisime kõikide institutsioonide peale kokku eesti keelde ligikaudu 250 000 lehekülge tekste. Kui arvestada, et keskmine raamat on 250–300 lehekülge, siis on see võrreldav tuhatkonna raamatuga.

Ent kui palju on toimetajaid? Sellele võiks vastata kas 0 või 150. Eraldi keeletoiimetaja ametikohti ei ole üheski ELi institutsioonis. Mõnes on olemas küll korrektorid, teistes pole ka neid.

Kas see tähendab, et me tõlkeid üldse ei toimetagi? Nii see siiski pole. Toimetamistööd teevad tõlkijad ise. Kolleegi tõlke toimetamine kuulub peaaegu kõikide tõlkijate igapäevatöö hulka. Mõni teeb seda muidugi rohkem, teine vähem – sõltuvalt kogemusest, huvist ja oskustest.

Tõlketöö maht kõigub päris palju ja harilikult ei võeta ELi institutsioonidesse tööle väga kitsalt spetsialiseerunud ametnikke. Ühes isikus tõlkija-toimetaja profiil on laiem kui ainult keeletoiimetajal.

Kindlasti on sellel negatiivne külg: oskusliku keeletoiimetaja panus oleks lõpptulemust silmas pidades muidugi hea. Kuid on ka positiivset: niisugune olukord sunnib kõiki tõlkijaid keeleasjadega kursis olema ja pidevalt õppima. Kuidas teisi saaks kolleegi tööd toimetada ja oma pakutud parandusi talle selgitada? Nii mõjutab see loodetavasti ka iga tõlkija enda töö tulemust.

Peale selle kätkeb ELi tõlkija töö väga sageli toimetamise aspekti, sest palju tuleb töötada tõlkemäludest ja andmebaasidest saadud vastetega: hinnata nende kvaliteeti ja kõlblikkust, kontrollida nende korrektsust ning teha muudatusi ja parandusi, et vasteid oma tõlkes kasutada. Näiteks Euroopa Parlamendis on tõlgete taaskasutuse määr ligi 50%. Peaaegu poole sellest, mida tõlgime, leiame juba kusagilt mujalt varem tõlgituna – vahel täiesti täpse, aga sagedamini osalise vastena kas meie enda, Euroopa Komisjoni või nõukogu tõlgetest, juba vastuvõetud ja avaldatud õigusaktidest või mujalt.

Varasemate tõlkesegmentide taaskasutus hõlbustab tõlkija tööd ja tagab järjepidevuse, sest nii ei tõlgita sama lähtekeelset lauset mitu korda mitmel moel. Teisalt aga seob see tõlkija teise inimese tehtud tõlkega, millest on tihti üsna raske end lahti rebida. Seepärast kuuleme mõnigi kord tõlkijat toimetajale väitmas: „Aga [tõlke]mälust tuli nii!“

Toimetamistööd teevad tõlkijad ise. Kolleegi tõlke toimetamine kuulub tõlkijate igapäevatöö hulka.

Euroopa Parlamendis on tõlgete taaskasutuse määr ligi 50%. Peaaegu poole sellest, mida tõlgime, leiame juba kusagilt mujalt varem tõlgituna.

Automatiseeritus ja varasemate tõlgete taaskasutus kiirendab tööd, kuid tekitab ka olukorra, kus tõlkija – tegutsedes korraka nii tõlkijana kui ka tõlkeprogrammi pakutud fraaside ja lõikude toimetajana – ei loo tõlgitava teksti mõtte edastamiseks oma teadmiste ja keeletunnetuse abil uut teksti, vaid korrastab ja sõnastab ümber juba olemasolevat. Võib ka juhtuda, et ta jätab selle tegemata, kui kiirustades tundub, et olemasolev käib ju küll, või suruvad tähtajad ja uued tööd peale või on juba valmis lausest lihtsalt raske lahti lasta ja mõelda, kas tegelikult ütleksin seda kuidagi teistmoodi, kui oleksin selle ise algusest peale kirja pannud.

TERVIKPILT KAOB

Mida tähendab see tulevikku silmas pidades? On kindel, et tagasi ei tule aeg, kui toimetajal oli iga teksti juures mahti end teema ja taustaga kurssi viia, tegutseda tervikut arvestades, pöörduda tagasi juba läbitöötatud osade juurde jne. Selline ideaalne maailm on praeguse tekstimahu ning töö- ja avaldamiskiiruse juures võimatu.

Praegu töötame veel klassikaliste tõlkeprogrammidega, milles tõlkimine ja üha enam ka toimetamine käib segmenthaaval (kuigi ka traditsiooniline paberil toimetamine elab veel). Tõlketellimused saabuavad enamasti ikka harjumuspäraste Wordi või Exceli dokumentidena, mille originaali saab lahti võtta, et näha, milline peab lõpptulemus jääma. Lähituleviku suurim muutus, mis tegelikult juba käib – seni küll vaevaliselt ja tagasilöökidega – on aga niinimetatud XMListumine.

See tähendab, et edaspidi puudub meil tõlkides ja toimetades üldse võimalus vaadata, millise teksti, trükise või tootega on tegu. Selle asemel näeme üksnes

Näeme üksnes pusletükikesi ning sõltume konteksti hoomamisel sellest, mida tellija on heaks arvanud või taibanud saatesõnaks kaasa panna.

pusletükikesi ning sõltume konteksti hoomamisel sellest, mida tellija on heaks arvanud või taibanud saatesõnaks kaasa panna. Mis aga kõige küsitavam: kõik see, mille süsteem või arvutiprogramm on leidnud andmebaasidest või mis on viitetekstidena kaasa pandud, on sellisel juhul juba tõlkija ja toimetaja jaoks n-ö lukus ning seda enam muuta või parandada ei saa.

Süsteem eeldab, et kõik, mis on kord tehtud ja käiku lastud, on puhas kuld ning sobib samasugusena uuesti kasutada. Aga kui see nii ei ole, kui on tehtud viga või kui kontekst on siiski teistsugune?

Näen selles lähituleviku suurimat konflikti IT-arendajate ja keeleinimeste vahel, ka konflikti tehnoloogia või tehisintellekti ning pärisintellekti, st elusa inimese aju ja taju vahel. Põhiküsimuseks näib kujunevat see, kas tulevikus õnnestub saavutada mõistlik tasakaal tehnika ja tehnoloogia arengu ning teisalt selge, õige ja ilusa keelega seotud vajaduste vahel, mida ei ole kõike võimalik algoritmidesse kirjutada.

Milline roll jääb selle juures tõlkijale ja toimetajale? Isegi kui amet ja töökoht säilivad, siis kui palju on kasu õppinud, kogenud ja intelligentse inimese tarkusest, kui tehnika ei lase tal enam neid oskusi ja teadmisi kasutada?

MIDA JA KUI PALJU TOIMETADA?

On veel üks küsimus, mis hakkab päevakorra kerkima juba praegu, kuid arvata-vasti muutub tulevikus veel teravamaks: mida kui hästi tõlkida ning mida kui palju ja kui põhjalikult toimetada?

Selline küsimus võib esmapilgul tunduda kohatu ja isegi rumal. Miks peab üldse valima? Teeme ikka alati kõike hästi ja korralikult!

Teose, dokumendi, infosisu autor peab oma teksti kirja pannes eeldatavasti silmas potentsiaalset lugejat, teksti kasutajat. Seda arvestades ta kas viimistleb oma teksti põhjalikult või hoopis visandab kiiruga, kirjutab selle lihtsalt ja ladusalt või täpsustab ennekõike hoolikalt kõiki juriidilisi nüansse, erialast sõnavara jne.

Niisamuti peab tõlkija ja toimetaja teadma või ära tundma, kellele ja milleks on autori tekst mõeldud, ning sellest lähtudes tekstisse suhtuma. Kui harilikult tähendab see ennekõike sobiliku registri ja sõnavara valikut, siis üha enam on sellele lisandumas teisi, raskemini määratletavaid asjaolusid. Sellest, kas tegu on kiir- või masintõlkega ning milline on teksti eesmärk (kas juriidiline analüüs või noorteürituse flaiier?) ja kui püsiv see on (kas raamat või sotsiaalmeedia säuts?), hakkavad sõltuma teksti kvaliteedinõuded. Seda tingib asjaolu, et info maht on tänapäeval piiramatult, samal ajal kui tõlke- ja toimetamisressursid – eriti väikese-arvulise kõnelejaskonnaga keelte puhul – on vägagi piiratud.

Info üleküllus tähendab, et me ei suuda ammu enam kõike meile pakutavat pealiskaudselt sirvidagi, rääkimata süvenemisest. See sunnib nii infosisu loojaid, selle edastajaid ja vahendajaid (sh tõlkijad, toimetajad) kui ka kasutajaid üha enam võtma arvesse selle sisu eesmärki, püsivust, olulisust, kasutajaskonda ning seega ka kvaliteeti ja hinda. Praegu veel me nii kaugel ei ole – veebi ja trükki jõuab valimatult kõik. Suur osa sellest toimetaja käe alt läbi ei käi, kuid samal ajal

kulutatakse väärtuslikku ressursi võib-olla ka asjatult kusagil mujal, kus selle töö tulemus püsib vaid mõne tunni või seda kasutavad vaid üksikud lugejad.

Ilukirjanduse tõlkimisel oleks mõeldamatu jätta kvaliteet teisejärguliseks: kiiruga kokku visatud ja pealispindsel toimetatud tõlge sunnib romaani ärritunult käest panema. Niisamuti on lubamatu seaduste ja muude õigusliku toimega haldusdokumentide tõlkimisel või toimetamisel juhinduda üksnes kiirusest – sellised tekstid ja dokumendid eeldavad täpsust ning kontrollitud ja järjepidevalt kasutatavaid termineid.

Kas aga sama rangeid nõudmisi on tarvis rakendada kõigele? Kas pelgalt kiire ülevaate saamiseks selle kohta, mida mingi delegatsioon oma visiidil tegi, nägi ja kuulis ning mida õhtul sõi, on tingimata tarvis, et visiidi protokoll oleks viimistletud niisama latusaks nagu hea romaan või terminites niisama täpselt nagu määrus, millest vastuvõtmise järel saab automaatselt kõikide ELi liikmesriikide õiguse osa? Kena ju oleks, aga kui aega ega vaba ressursi pole nii palju? Eriti veel, kui teame, et seda sirvib pärast mõni üksik huviline, et teada saada üksnes põhiline. Kas piisaks sel juhul ka toorest ja toimetamata tõlkest või kiiresti kirja pandud protokollist, kus mõte on küll edasi antud, kuid sõnastus ei pruugi olla kõige latusam?

Niisugused valikud toovad kaasa kolm suurt probleemi.

Kuidas ja mille põhjal eristada tekste, mis nõuavad parimat kvaliteeti, mille puhul võib piirduda keskpärasega ja mille jaoks piisab toor-masintõlkest.

Esiteks, väga raske on määratleda ja veel raskem reeglitenä kirja panna, kuidas ja mille põhjal eristada tekste, mis nõuavad parimat kvaliteeti, mille puhul võib piirduda keskpärasega ja mille jaoks piisab toor-masintõlkest (lisaks kõik muud varjundid nende vahel). Kes peaks seda üldse otsustama: kas autor, tellija, toimetaja, tõlkija või hoopis sihtrühm ehk lugeja?

Teiseks, tõlkijad ja toimetajad on professionaalid, kelle ametiaue ei luba kehva teksti kergesti käest anda. Kuidas toimetajatööd tehes otsustada, et selle teksti teen nii hästi, kui oskan, tollest teisest aga jooksen kolm korda kiiremini üle?

Kolmandaks, kuidas saavutada see, et tellija teadlikult rahulduks pealispindse või konarliku tekstiga või seda sellisena koguni telliks? Erasektoris saab seda reguleerida teenuse hinnaga, ent avalikus sektoris?

Siit jõuame järgmise tähtsa küsimuseni: kas oleme nõus sellega, et meie keele-ruumi jõuavad ja jäävad ringlema tekstid, mis on teadlikult koostatud, tõlgitud või toimetatud mingite teistsuguste, kehvemate kriteeriumite kohaselt? Kas ja

kui palju oleks lubatud niisugustes tekstides keelenormist ja õigekirjareeglitest kõrvale kalduda? Kas aktsepteeriksime kohmakat sõnastust, kuid õigekiri olgu korras? Kas suudame sellisel juhul hoida sellist odavkeelt kusagil eraldi, piiratud kasutusruumis kinni ja vältida selle levikut ametikeelde, argikeelde ja lõpuks võib-olla koguni ka keelenormi? Kardan, et ei suuda. Näeme seda toimumas juba praegu. Lugegem kas või toimetamata intervjuusid ajalehtede veebiväljaannetes või kuulakem ametnike ja ministrite keelekasutust.

UUS MUUTUJA VÖRRANDIS – MASINTÖLGE

Viimasel paaril aastal on ELi institutsioonide eesti keele tõlketeenistustes hakanud toimetamis- ja tõlkimistööd üha enam mõjutama masintõlge. Esimesed katsetused sel alal tehti maailmas üsna ammu ja siis reeglipõhiselt. Kiiresti ilmnes aga, et kõiki keelereegleid, erandeid ja nüansse pole võimalik algoritmideks teisendada, ja masintõlge jäi kauaks soiku, kuni arvutite võimsus lubas proovida statistilist meetodit. Seda kasutas näiteks Google ja sellel põhines algul ka Euroopa Komisjoni masintõlge, mida saavad kasutada teisedki, sh Euroopa Parlament. Lihtsustatult otsiti andmebaasidest sarnaseid mõne sõna pikkusi juppe nende esinemissageduse alusel. Eesti keelt (koos ungari, soome ja saksa keelega) peeti üsna lootusetuks, sest lause ülesehitust, käändelõppe jms tuleb tõlkides palju muuta.

Ent tehisintellekti ajastu saabudes läks komisjon paar aastat tagasi katseliselt üle närvivõrkudel põhinevale masintõlkele ehk neurotõlkele, tehes seda esimesena just „rasketes“ keeltes, sh eesti keeles. Sellega muutus meie igapäevane töö lausa üleöö, sest neurotõlge on teatud tekstide puhul vägagi kasutuskõlblik.

Väga lihtsustatult öeldes ei otsi arvuti enam andmebaasist sarnaseid tekstijuppe ega paku neid varasema esinemissageduse järgi, vaid tuletab endale tekstikorpuse põhjal ise reeglid, võttes arvesse sõnajärge, süntaksit, sõltuvusvahetordi ja muid keerulisi keeleaspekte. Meile tuttavaid grammatikareegleid ei ole ette antud, tegu on iseõppiva tehisintellektiga.

Tulemused on hämmastavad, sest neurotõlge suudab anda ka eesti keeles grammatiliselt korrektseid, õigete käände- ja pöördelõppudega ladusaid lauseid, mis sageli on ka sisult korrektseid.

Neurotõlge suudab anda ka eesti keeles grammatiliselt korrektseid, õigete käände- ja pöördelõppudega ladusaid lauseid, mis sageli on ka sisult korrektseid.

Muidugi ei tähenda see, et nüüd on tõlkija ja toimetaja ametil lõpp ning masin teeb kõik ise ära. Kaugel sellest – ka kasutuskõlblikku masintõlget tuleb tähelepanelikult toimetada ja ega masin kõike ka veel hästi oska, ta teeb ikka nalja ka ning vahel pöörab mõtte vastupidisekski.

Toimetaja jaoks on aga selle juures üks tähtis muutus. Varem oli kohe näha, kui masintõlke lause oli kasutu – näiteks siis, kui isegi sisu ja tähenduse mõttes korrektsete osade õigesse käändessegi panek võtnuks sama kaua või rohkem aega, kui nende ise õigesti kirjapanek. Nüüd aga, isegi kui sisu on veidi mõttest mööda või tähenduselt koguni vastupidine, on masintõlke pakutavas kasutuskõlblikku palju, sest grammatiliselt on lause korrektne ning tuleb üles leida pigem sisuvead.

See kõlab lihtsamalt kui tegelikkus. Esiteks on võimalus kergesti sisuliselt eksida. Lause näib ilus, ladus ja korrektne – see uinutab. Mõni oluline asi võib olla välja jäänud, mõnikord on midagi hoopis ülearu, kuid seda ei pruugi märgata. Kui aga pikalt lähtetekstiga võrdlema jääda, kaotab masintõlge mõtte, sest aega kulub rohkem kui vilunud tõlkijal enda peaga tõlkimisele.

Siit tekibki alaline dilemma, mis tuleb masintõlke toimetajal enam-vähem sekundi või paariga lahendada iga lause juures: kas võtan masintõlke ja hakkan seda korda tegema või tõlgin parem ise otsast peale ise?

Samuti järgib masintõlge paratamatult ikkagi rohkem lähtekeele sõnajärge, kui seda teeks hea tõlkija, kes ei tõlgi sõnu, vaid mõtet, või asjatundlik toimetaja. Ent kui lause on saadud korrektseks, ilmsed vead on leitud ja parandatud, kas siis enam jätkub aega ja jaksu või oskustki veel ka lause stiil, sõnajärg ja ladusus üle käia ja korda teha? Need on kaks asja, mida me masintõlget kasutades peame veel õppima ja millega harjuma. On selge, et masintõlge ei kao kuhugi, vaid järjest täiustub ja selle kasutusala laieneb. Seepärast on lähiajal vaja tegeleda põhjalikumalt sellest tulenevate veatüüpidega ja järeldoimetamise koolitusega.

MIS SAAB TOIMETAJA AMETIST?

Kui avalikku ruumi paiskuvate tekstide ja info hulk üha kasvab, kuid nende eluiga ja väärtus lühenevad, kui osa tekste loob inimese asemel tehisintellekt, siis mida see võiks tähendada toimetamise kui tegevusala ja toimetaja kui ameti tuleviku jaoks?

Esialgu saab siin esitada veel pigem küsimusi, kui pakkuda vastuseid.

Kui masintõlge muudab tõlkimise märksa kiiremaks, kas võime siis loota, et tõlkija-toimetaja saab selle võrra rohkem aega pühendada enda või kellegi teise teksti toimetamisele? See oleks tore.

Kui masintõlke kasutamine tähendab täiesti uut tüüpi tegevust, mis ei ole klassikalises mõttes ei tõlkimine ega toimetamine, vaid nn järeltoimetamine (ingl *post-editing*), kas siis kujuneb sellest ka uut tüüpi amet – järeltoimetaja? Seegi pole võimatu.

Kui veebiuudised, sotsiaalmeediapostitused või -säutsud võivad sattuda lugeja silme ette üksnes mõne minuti või tunni jooksul pärast nende avaldamist ning siis asendavad neid juba üha uued, kas on siis üldse mõtet neid keeleliselt toimetada? Pean siin silmas eelkõige meediaväljaannete tegevust.

Kuidas võib mõjuda see, kui igaüks võib kergesti kirjastada ise raamatu, millesse ta võib teksti kirja panna täpselt niisugusel tasemel, mileni tema keeleteadmised ulatuvad, ja toimetaja abi kasutamine ei tule sealjuures võib-olla pähegi? Kas siis ajapikku hakkamegi leppima ja harjuma sellega, et ka kõik muu, isegi ilu- ja tõlkekirjandus, võib olla kiiresti kokku visatud keelerämps?

Konarliku, kuid sisult korrektse seaduse tekstiga saavad juristid ilmselt hakkama, võib-olla isegi hõõruvad käsi, sest tõlgendamis- ja vaidlusruumi on nii rohkemgi. Mõne toote toimetamata kasutusjuhend võib eriti halval juhul olla koguni ohtlik, kuid tavalisem on see, et annab pigem ainst netis levitatavateks naljadeks. Isegi ajalehe või veebiväljaande vigu täis uudisnupu võib välja kannatada – nahunii on see varsti ununenud.

Kuid olen endiselt arvamusel, et raamat, ajalehe- või ajakirjaartikkel või isegi intervjuu, mis peaks kiiresti saadava teabe kõrval andma lugejale midagi ka intellektuaalselt, pakkuma naudingut, ei kõlba toimetamata kujul küll lugeda.

Seepärast usun, et toimetaja töö olemus teatud valdkondades – näiteks tõlkimisel – küll muutub ja teiseneb, võib-olla mõnes sektoris isegi kaob, sest seda tingib muutuv informatsioon, ent klassikaline keeletoimetamine – see tuum, mis hoiab alles ja elusana korrektse ja ilusa emakeele – jääb. Peab jääma.

Ning ehk mõne aja pärast, kui oleme vähe kvaliteetsest infolaviinist ja -mürast juba täiesti küllastunud, tõuseb meisterlikult koostatud ja toimetatud väärttekst taas suuremasse ausse, nii et ka keeletoimetajale jääb nende koostamisel ja avaldamisel väärikas koht.

Kui veebiuudised, sotsiaalmeediapostitused või -säutsud võivad sattuda lugeja silme ette üksnes mõne minuti või tunni jooksul pärast nende avaldamist ning siis asendavad neid juba üha uued, kas on siis üldse mõtet neid keeleliselt toimetada?