

Tartu Ülikool

Maailma keelte ja kultuuride kolledž

Helina Tera

**NEUROMASINTÕLKE JÄRELTOIMETAMINE AUTOMAATSE  
KVALITEEDISKOORI TOEL**

Magistritöö

Juhendaja: Reelika Saar

Tartu

2022

# Sisukord

Sissejuhatus.....	3
1. Masintõlge .....	5
1.1. Masintõlkest üldiselt .....	5
1.2. Masintõlke liigid .....	7
2. Masintõlke järeltoimetamine .....	9
2.1. Masintõlke järeltoimetamisest üldiselt .....	9
2.2. Eelnevad uurimused masintõlke järeltoimetamisest.....	12
3. Masintõlke kvaliteedi hindamine.....	15
3.1. Masintõlke hindamisest üldiselt.....	15
3.3. Eelnevad uurimused masintõlke automaatsest kvaliteedi hindamisest .....	19
4. Uurimus neoromasintõlke järeltoimetamisest automaatse kvaliteediskoori toel.....	22
4.1 Valim.....	23
4.2 Küsimustiku kirjeldus .....	24
4.3 Küsimustiku tulemused.....	25
4.4 Katse kirjeldus.....	29
4.5 Katse tulemused .....	32
5. Järeldused.....	40
Kokkuvõte.....	44
Kasutatud kirjandus .....	46
Summary.....	50
Lisa.....	52

## Sissejuhatus

Kuna digilahendused on maailmas üha tähtsamal kohal, arenevad koos digilahendustega pea kõik valdkonnad. Erandiks pole ka tõlkimine – väga palju kasutatakse igapäevaelus tõlkimiseks masintõlkesüsteemide abi. Ka professionaalne tõlkimine liigub vaikselt selles suunas, et üha enam saaks rakendada töösse masintõlget, sest masintõlke väljundid võib hinnata juba üsna heaks (olenevalt süsteemist, mida kasutada). Praegusele üldisele masintõlke tasemele aitab suuresti kaasa üleminek neuromasintõlkesüsteemidele, mis põhinevad inimese ajuga sarnastel närvivõrkudel, seega annavad sellised süsteemid väljundeid, mis peale selle, et on tõenäoliselt korrektsemad, mõjuvad inimesele lugedes ka loomulikumalt, mitte enam nii masinlikult kui varasemad masintõlked.

Masintõlke üheks põhiliseks rakendamise viisiks professionaalse tõlkimise alal on masintõlke järeltoimetamine. See kujutab endast mitte lähteteksti nullist tõlkimist (inimtõlkija poolt), vaid masintõlkesüsteemi poolt sihtkeelde tõlgitud teksti järeltoimetamist. Seega asub tõlkija justkui toimetaja rolli, kuid erinevus on selles, et tuleb arvestada masintõlke iseärasustega – masintõlge ei pruugi anda korrektseid tulemusi, isegi kui näiliselt võib tõlge tunduda korralik. Järeltoimetamisel on väga oluline, et seda tööd tegev tõlkija on järeltoimetamise alal pädev, sest masintõlgitud tekstis vigade tuvastamine ei pruugi olla sugugi lihtsam ülesanne, kui teksti algusest lõpuni ise tõlkimine. Samuti on olulised teadmised järeltoimetamisest, näiteks see, kui palju millise teksti puhul vaja tööd teha on, sest järeltoimetamise käigus, peale selle, et on oht jätta vigu märkamata, on ka väga suur oht tekste üle toimetada. Samuti on oluline tehnoloogiline pädevus, sest töö toimub arvutipõhistes programmides.

Selleks, et masintõlke väljundi järeltoimetamine kulgeks protsessina ladusamalt, on välja töötatud süsteeme, mis aitavad masintõlget hinnata. Minevikus on need süsteemid olnud üsna subjektiivsed, kuna masintõlke väljundeid on hinnanud inimtõlkijad. Nüüd on asutud välja töötama funktsioone, mis hindavad masintõlget automaatselt ja annavad sellele kvaliteediskoori. Neid nimetatakse MTQE (*Machine Translation Quality Estimation*) funktsioonideks. Teoreetiliselt võiks sellest funktsioonist kasu olla näiteks tõlkijatele, kes

saaksid skoori põhjal lihtsamalt otsustada, kas tegemist on korrektse tõlkega, hea masintõlkega, mis vajab veidi toimetamist või halva masintõlkega, mis oleks lihtsam kustutada ning seejärel tekst ise tõlkida. Sellisest funktsioonist tõlkijad ise ka ise puudust tundnud, nagu näitas Moorkensi ja O'Brieni 2017. aastal läbiviidud uurimus, millest tuleb hiljem lähemalt juttu.

Käesoleva töö eesmärk on välja selgitada, kas automaatsest kvaliteediskoorist võib olla tõlkijatele järeltoimetamisel kasu või mitte. Töö teoreetiline osa hõlmab endas ülevaadet masintõlkest, masintõlke järeltoimetamisest ning masintõlke hindamisest, sealhulgas on kirjeldatud automaatseid kvaliteediskoore pakkuvaid funktsioone. Töö analüüs koosneb kahest osast. Esimeseks neist on küsimustik, kus küsiti tõlkijate arvamust nii masintõlke, masintõlke järeltoimetamise kui automaatse kvaliteediskoori kohta (kõik kolm teemat on olulised, sest selleks, et hinnata automaatset kvaliteediskoori, on ka oluline, et oleks teatud arusaamised masintõlkest ja selle järeltoimetamisest). Küsimustikule lisaks sooritasid tõlkijad katse, kus neil paluti järeltoimetada kaks erinevat lühikest teksti inglise-eesti keelesuunal. Ühele tekstidest oli kuvatud kvaliteediskoorid ja teisele mitte. Töös on esitatud põhjalikud tulemused nii küsimustiku kui katse kohta. Küsimustik on analüüsitud iga küsimuse kaupa eraldi, et anda võimalikult põhjalik ülevaade erinevatest vastustest ja küsimustiku vastused koosnevad nii kvalitatiivsetest kui kvantitatiivsetest andmetest. Katse analüüsiks on koostatud erinevaid tabeleid, milles võrreldakse muuhulgas tehtud muutuste arvu ja järeltoimetamiseks kulunud aega. Katse tulemusi võrreldi ka küsimustiku tulemustega, et välja selgitada, kas tõlkijate hoiakud ja tegelik töö käivad kokku. Katse lõpus küsiti tõlkijatelt ka arvamust äsjasooritatud katse kohta, ka need vastused on töö lõpus põhjalikult välja toodud.

Magistritöö lõpus on esitatud järeldused, mis annavad ülevaate tehtud tööst ning sisaldavad mõtisklusi selle üle, millised võiksid olla edasised uurimused sel teemal.

# 1. Masintõlge

## 1.1. Masintõlkest üldiselt

Kuna maailm on pidevas arengus, siis on pidevas muutuses ka erinevad tööviisid, suures osas (nii palju kui võimalik) liigutakse tehnoloogiliste lahenduste poole, st üha tähtsam on inimeste ja masinate koostöö. Ideaalseks näiteks on see, kuidas on läbi aastate arenenud ja muutunud töö tõlkevaldkonnas. Tõlkemahud on maailmas aastast aastasse üha enam kasvamas (vt Nitzke ja Hansen-Schirra 2021: 1) ja keelesuundi, millel tõlgitakse, on sadu. Selleks, et tõlkijad suudaksid tõlkemahtude kasvuga sammu pidada, on vaja leida ja arendada uusi võimalusi ning selle tarbeks on kasutusele võetud tehnoloogilised abivahendid, eesotsas masintõlkesüsteemidega, mida on arendatud juba aastakümneid, kuid mis on just 21. sajandi jooksul teinud hüppelisi edusamme. Masintõlkesüsteemide arendamine on just tänapäeval äärmiselt vajalik, sest vastasel juhul ei ole varsti enam piisavalt tööjõudu ning muuhulgas muudab edukas masintõlge tõlketuru soodsamaks ka tõlketööde ostjatele ning pakkujatele.

Masintõlge on tõlge, mida pakuvad erinevad masintõlkesüsteemid. Sellisesse süsteemi sisestatakse lähtetekst ning seejärel tõlgib süsteem sisestatud teksti sihtkeelde. Tavakasutajate jaoks on kõige populaarsem neist ilmselt Google Translate, millega ilmselt kõik on mingil ajahetkel oma elus kokku puutunud, kuna see pakub laialt hulgal keelesuundi. Professionaalsed tõlkijad saavad integreerida erinevaid masintõlkesüsteeme ka otse tõlkeabiprogrammidesse, sest nii saab nendest programmidest vajadusel ilma vaevata automaatselt tõlkeid näiteks järeltoimetamise tarbeks, millest tuleb hiljem lähemalt juttu.

Kuna masintõlge on väga kiirelt arenev ja nüüdseks laialt kasutuses nii professionaalsetes kui tavakasutajatele mõeldud keskkondades, siis nähakse selles edaspidi vaid üha rohkem potentsiaali hõlbustamiseks ka tõlkijate tööd. Samuti, kui tutvuda lähiaastatel kirjutatud masintõlketemaliste uurimistööde ja artiklitega, siis pigem skeptiline tõlkijate suhtumine on järjest asendumas positiivsema suhtumisega. Näiteks Fernández-Torné ja Matamala leidsid oma 2021. aasta töös „Human evaluation of three machine translation systems: from quality to attitudes by professional translators“, et nende katses osalejad

pidasid masintõlke kvaliteeti pigem madalaks, kuid samaaegselt pidasid nad siiski masintõlget küllaltki kasulikuks ning tunnevad selle vastu huvi ja võtavad seda üha enam oma töös kasutusse. Mingit rolli masintõlkesse suhtumise pöördumises võib mängida ka see, et uued tõlkijad, kes tõlkijana tööle asuvad on noored ja tehnoloogiaga koos üles kasvanud, seega on usaldus selliste süsteemide vastu tõenäoliselt suurem ja suhtumine entusiastlikum, kui vanema generatsiooni tõlkijatel, kes on harjunud tekste pigem algusest lõpuni ise tõlkima ja tehnoloogiat nii väga ei usalda.

Masintõlke idee tekkis juba 20. sajandi esimesel poolel, seega on see juba tükk aega arenduses olnud. Põhjaliku ülevaate masintõlke ajaloost annavad Nitzke ja Hansen-Schirra oma 2021. aastal ilmunud raamatus „A short guide to post-editing“. Tänapäeval on masintõlke arenguga jõutud kohta, kus selle kvaliteeti võib mõnel juhul pidada väga heaks. Masintõlke võrdväärst inimtõlkega on juba aastaid püütud tõestada või ümber lükata. Näiteks Läubli *et al.* 2018. aasta uuring näitab, et lausetasandil ei olnud katsealustel eelistust masin- ja inimtõlke vahel, kuid kogu teksti tasandil eelistatakse siiski pigem inimtõlget (Läubli *et al.*: 4794).

Masintõlke kvaliteet sõltub erinevatest teguritest, seal hulgas teksti tüübist, treeningandmete hulgast (ehk andmete hulgast, mille põhjal süsteemi on treenitud tõlkima) ja kvaliteedist, aga ka keelte sarnasusest (Nitzke, Hansen-Schirra 2021: 4). Seega mida paremini on süsteem treenitud teatud tekste tõlkima, seda paremat tõlget see pakub. Nitzke ja Hansen-Schirra (2021: 39) kirjutavad lisaks, et masintõlke jaoks peetakse harva sobivaks väga loomingulisi tekste, sest need vajavad paindlikke tõkelahendusi, paindlikkust ja loovust, näiteks luuletused, mis tuginevad rütmile ja riimile, mida tavalised masintõlkesüsteemid arvesse ei võta. Katre Sepp, kes on 2017. aastal uurinud statistilise masintõlke järeltoimetamist, on oma töös kirjutanud järgmist:

Tundub loomulik, et üks põhjuseid, miks masintõlkesse negatiivselt suhtutakse, ongi see, et masintõlget püütakse kasutada kõikvõimalike tekstide tõlkimisel, võtmata arvesse seda, et tegelikult on väga suur vahe selles, millist tüüpi tekst masintõlkele ette anda või mille jaoks süsteem on treenitud (Sepp 2019: 7).

## 1.2. Masintõlke liigid

Masintõlge sai alguse reeglipõhisest masintõlkest. Reeglipõhise masintõlke tööpõhimõte tugines sellele, et keelt püüti süsteemile n-õ selgeks õpetada süsteemi sisestatavate keelereeglite põhjal. See oli mahukas töö, kuna neid keelereegleid pidid koostama ja sisestama inimesed ise. Reeglipõhise masintõlkesüsteemi tõlkeprotsess koosneb sisestatud teksti analüüsimisest (morfoloogiliselt, süntaktiliselt ja semantilist) ja seejärel teksti genereerimisest. Reeglipõhiste süsteemide suureks miinuseks on, et uute reeglite arendamine ja korras hoidmine on keeruline ning samuti ei ole garanteeritud, et uute reeglite süsteemi lisamisel see sama hästi töötab, kui enne seda. (Chérargui 2012: 164) Seega hakati reeglipõhisele masintõlkele otsima alternatiive, kuna selle arendamine nõuab keeleteadlastelt väga suurt tööd (grammatikareeglite, sõnaraamatute lisamine jne).

Järgmiseks hakati arendama statistilist masintõlget (kasutatakse ka terminit *andmepõhine masintõlge*), mis lähtub tõlkides vastete tõenäosusest, kasutades selleks üks- või kakskeelseid korpuseid. Sellest sai alternatiivne lähenemine masintõlkele, et ületada süsteemi n-õ uute teadmiste omandamise probleem, mis iseloomustab reeglipõhist masintõlget (Chérargui 2012: 165). Sellised süsteemid kasutavad suuri kakskeelseid tõlkekorpuseid ehk paralleelkorpuseid, millele rakendatakse õppealgoritmide ning seejärel eraldatakse sealt korpustest uue materjali tõlkimiseks tõlkeparameetrid ja mudelid, võttes valiku aluseks peamiselt sõnade sageduse ja kombinatsioonid (Guerberof 2012). Tänu sellele, et statistiline masintõlge keskendub kombinatsioonidele, tekib tõlgetes vähem vigu, mis varem tekkisid näiteks sõnade mitmetähenduslikkusest, sest sõnade sagedus ja kombinatsioonid aitavad konteksti luua.

Statistilisel masintõlkel oli kõigele vaatamata siiski üsna palju puudusi. Näiteks, kuna statistilised masintõlkesüsteemid põhinevad sõnade sagedaste koosinemise mustrite leidmisel, tõlgitakse vähem kasutatavad sõnad suurema tõenäosusega valesti (Silva 2012: 1). Seetõttu arendati masintõlget kiirelt edasi. Uusim suundumus masintõlkesüsteemide vallas on neuro- ehk närvivõrkudel põhinev masintõlge (Sepp 2017: 8). Neuromasintõlge kasutab tehnilikku närvivõrku selleks, et ennustada sõnade järjestuse tõenäosust, sageli tervete lausete

kujul (DeepAI 2019). 2016. aastal tuli esimesena neuromasintõlkega välja Google Translate (Nitzke, Hansen-Schirra 2021: 21). Mitmed hiljutised uuringud ja hindamised näitavad, et neuromasintõlke kvaliteet ületab paljude keelepaaride puhul eelnevaid masintõlke tüüpe (Koponen *et al.* 2019: 62). Kuid ka neuromasintõlke puhul tuleb olla ettevaatlik – neuromasintõlke väljund võib tunduda korrektne, aga ei pruugi seda olla, sest mida paremad on tõlked (näiliselt), seda keerulisem on nendes vigade tuvastamine (Nitzke, Hansen-Schirra 2021: 27). Sama on leidnud ka Ingrid Sibul, kes tõdes oma töös, et neuromasintõlke tekstid on soravamad kuid ei pruugi olla usaldusväärsed (Sibul 2018: 5).

Osad süsteemid kasutavad siiani paralleelselt statistilist masintõlget ja neuromasintõlget, kuid üldine suundumus on siiski liikuda ainult neuromasintõlke peale.

Kuigi masintõlge ei ole veel arenenud (ega ei prognoosita veel lähiajal ka arenema) punkti, kus seda saaks saajaprotsendiliselt usaldada ja ilma toimetamata professionaalseks otstarbeks kasutusse võtta, siis sellegipoolest on tehtud suuri edusamme. Osasid tõlketekste kasutatakse juba sellisel kujul, nagu need masintõlkesüsteemist tulevad, olenevalt sellest, kes on sihtrühm ja kui korrektset tõlget vajatakse. Enamustel juhtudel on siiski vaja inimtõlkija-toimetaja sekkumist, et toimetada tekst üle kas tervelt või täielikult, enne, kui see kasutusse võetakse. Järgmisena tulebki juttu masintõlke järeltoimetamisest, kui uuest põhilisest suundumusest tõlkevaldkonnas.

## 2. Masintõlke järeltoimetamine

### 2.1. Masintõlke järeltoimetamisest üldiselt

Kui varem kasutati masintõlkesüsteeme suuresti tõlkimisel lihtsalt abivahenditena (või mitte professionaalsetes olukordades, vaid näiteks kodukasutuses), sest tervete lausete puhul ei andnud need väga täpseid tulemusi, siis tänu masintõlkesüsteemide kiirele arengule kasutatakse nüüd juba tekste, mis on üleni nende süsteemide poolt tõlgitud. See säästab tõlkija nullist tõlkimise vaevast ja sealjuures asub tõlkja hoopis toimetaja rolli, sest tema ülesanne on masintõlke poolt koostatud tekst järeltoimetada. Üks järeltoimetamise definitsioon, mille on oma töös pakkunud O'Brien (2011: 197-198) on järgmine: „järeltoimetamine on masintõlke toortõlke parandamine inimtõlkija poolt vastavalt konkreetsetele suunistele ja kvaliteedikriteeriumitele. Čulo *et al.* (2014: 35) on öeldud, et järeltoimetamine on valdkond, kus kohtuvad inimtõlkija ja masin, samuti masintõlke ja tõlketeaduse valdkonnad.

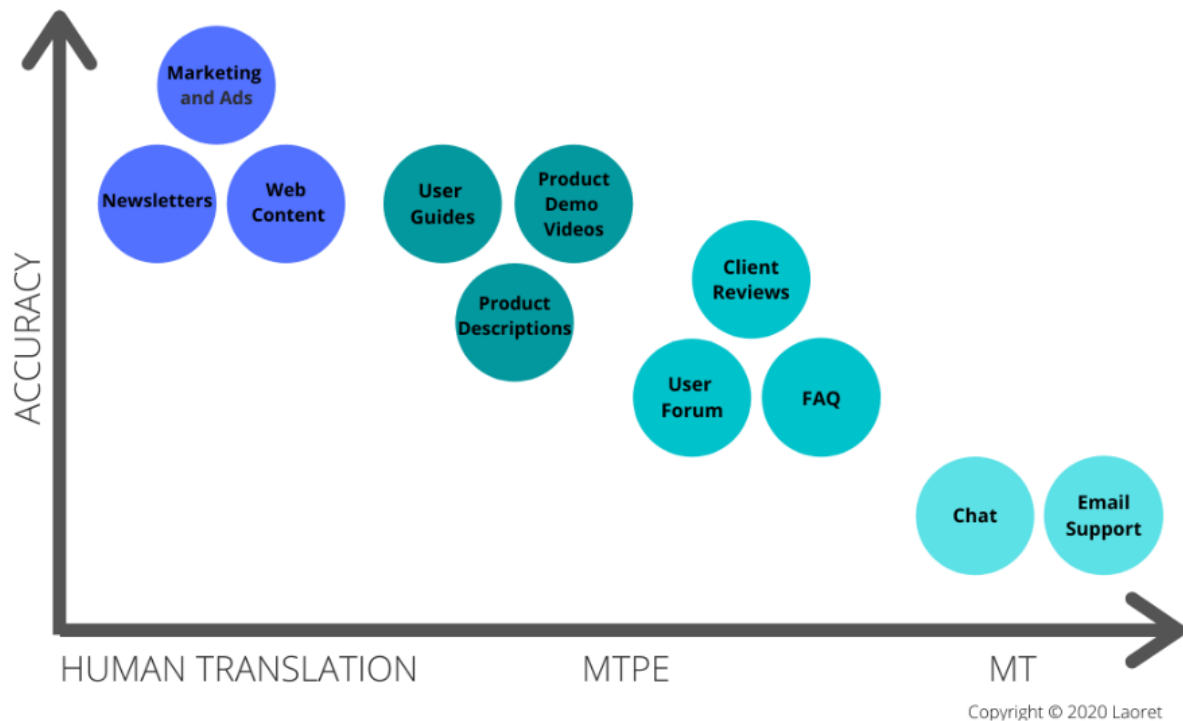
Järeltoimetamine on tänapäeval vastupandamatu, kuna see kiirendab potentsiaalselt tõlkijate tööd. Sellegipoolest võib töö kiirus olla järeltoimetamisel võrreldes inimtõlkega küll suurem, kuid kvaliteet halvem. Yamada (2019: 87) on kirjutanud, et tõlke järeltoimetamise kvaliteet peab vastama samadele standarditele, mis inimtõlke kvaliteet. Masintõlke järeltoimetamist käsitleb tõlkestandard ISO 18587:2021. Seal on esitatud näiteks pädevusvaldkonnad, mis on järeltoimetajale selles töös vajalikud. Nende hulka kuuluvad tõlkekompetentsus, keeleoskus ja teksti mõistmine, informatsiooni otsimise pädevus, kultuuriline pädevus, tehniline pädevus ja valdkondlik pädevus. Seega ei ole järeltoimetamine sugugi kergem töö, kui nullist tõlkimine. On olukordi, kus järeltoimetajal on korraga vaja rakendada isegi rohkem oskuseid kui tõlkijal.

Masintõlgitud tekst ei pruugi alati olla kasutatav, näiteks loogikavigade tõttu, mistõttu on selles töös vajalik olla tähelepanelik ja märgata igasuguseid vigu, sest neuromasintõlge võib tunduda petlikult korrektne ja mõned asjad võivad seetõttu kahe silma vahele jääda. Näiteks tõlgitud lause võib lähteteksti teades tunduda igati loogiline, kuid tegelikult võib masintõlkes olla kasutatud mõnda vale terminit, mis on küll sarnase tähendusega, kuid siiski

mittesobilik. Siinkohal peab mainima, et järeltoimetamisel on väga oluline järeltoimetajate kakskeelsus, et mõista nii lähte- kui sihtteksti, sest vastasel juhul võivad kontekstuaalsed vead jääda märkamata. Sealhulgas ei ole oluline mitte ainult kakskeelsus üksi, vaid ka tõlgitava teksti valdkonna tundmine näiteks sõnavaralises mõttes.

Koponen *et al.* (2019: 66-67) on välja toonud, et järeltoimetamisel ei piirduta alati ainult vigade parandamisega, vaid toimetajad võivad ka isiklike eelistuste põhjal tekste „üle toimetada“, st teevad masintõlgitud tekstis muudatusi, mis tegelikult olid masintõlgitud tekstis grammatiliselt ja terminoloogiliselt juba korrektsed. Seega ei ole järeltoimetamine ainult vigade parandamine, vaid selle käigus tuleb leida kognitiivne tasakaal lähteteksti mõistmise, soovitatud tõlke hindamise ja sihtteksti loomise vahel (Green *et al.* 2013: 440).

Järeltoimetamisel eristatakse kahte erinevat meetodit – osaline järeltoimetamine (*light post-editing*) ja täielik järeltoimetamine (*full post-editing*). Osaline järeltoimetamine tähendab seda, et toimetatud ehk lõplik tekst võib tunduda ebaharilik ja -loomulik, kuna see on genereeritud arvuti poolt. Grammatika ja süntaks võivad olla väärad, niikaua kui tähendus on arusaadav (Nitzke, Hansen-Schirra 2021: 30). Kuigi see ülesanne võib tunduda kergena, on sellise järeltoimetamisviisi puhul just kõige suurem oht teksti liialt üle toimetada. Kuna osalise järeltoimetamise käigus tehakse ainult võimalikult minimaalselt muudatusi, et tekst arusaadavaks muuta, siis kasutatakse seda peamiselt tekstide puhul, mis ei ole mõeldud avalikkusele (näiteks ettevõttesiseselt e-kirjade või vestluslogide tõlkimine), vaid lihtsalt selleks, et informatsiooni kiirelt edasi anda. Täielik järeltoimetamine tähendab, et peale kõige, mis kehtib osalise järeltoimetamise puhul, on oluline ka stiililine pool. See tähendab, et süntaks, grammatika ja kirjavahemärgid peavad olema korrektsed. Siiski on oluline kasutada võimalikult palju toortõlget ja mitte liiga peensusteni toimetada (Nitzke, Hansen-Schirra 2021: 32). Kuna täielikult järeltoimetatud tekst on korrektne, siis seda saab kasutada samadel tingimustel, kui nullist inimtõlkijate poolt tõlgitud tekste (kuid seetõttu, et aluseks on masintõlkesüsteemi väljastatud toortõlge, sobivad pigem siiski tehnilisemad tekstid, mitte loovtekstid). Joonisel 1 on näidatud, milliste tekstide puhul on mõistlikum rohkem masintõlget kasutada ja millised tekstid nõuavad rohkem täpsust.



Joonis 1. Tekstide täpsuse skaala. Allikas: <https://laoret.com/blog/light-post-editing-or-full-post-editing/>

Peale eelmainitu tuleb tõlkijal otsustada, kas järeltoimetamine on mingi teksti puhul üldse mõistlik lahendus või võtab selle parandamine rohkem aega, kui võtaks traditsiooniline tõlkimine. Pidev ebakvaliteetne masintõlge võib ajendada professionaalseid tõlkijaid järeltoimetamisest loobuma ja pöörduma tagasi nullist tõlkimise juurde. (Béchara *et al.* 2021: 1). Tõlkija peab hindama ka masintõlke riske. Kui töö tellija ei ole täpsustanud, et töö tuleks teha järeltoimetamise meetodil, siis peab tõlkija kaaluma läbi riskid, mis võivad tekkida. Näiteks tasuta veebipõhised süsteemid salvestavad kõik sisestatud tekstid teenusepakkuja serverisse ning sealt võivad need olla kättesaadavad kolmandatele isikutele. Seega, kui tegemist on tekstiga, mis sisaldab tundlikku infot, tuleks selliseid süsteeme vältida, või kui siiski otsustada masintõlget kasutada, siis kindlasti tuleks enne küsida luba tõlketöö tellijalt. (Nitzke, Hansen-Schirra 2021: 54-55).

Nende põhjuste tõttu on masintõlke uurimine väga oluline masintõlkesüsteemide arendajatele, et selgitada välja masintõlke tugevusi ja nõrkusi ning nende nõrkustega edasi

töötada ja neid parandada. Lisaks sellele saavad uurimustest kasu ka tõlketööde pakkujad (nt tõlkebürood), sest nad saavad selliste uurimustööde põhjal paremini otsustada, kui kasulik on neile järeltoimetamise integreerimine töösse ja kuidas seda tööd sobivamalt tasustada, sest kui varem on kasutusel olnud tähemärgipõhine tasustamine, siis nüüd, järeltoimetamise arenedes, liigutakse üha enam tunnitasu suunas. Korralikust järeltoimetatavast masintõlkest võib olla abi ka tõlkijatele, kelle töökiirust see potentsiaalselt märgatavalt tõsta võiks ja tööd hõlbustaks.

Masintõlke järeltoimetamist saab rakendada ka n-ö ringkasutusse. See tähendab, et masintõlget võib kasutada ka tehnilisel tasandil, osana protsessist, mille eesmärk on täiendavalt treenida ja täiustada masintõlkesüsteeme. Tekste saab masintõlkesüsteemidega automaatselt tõlkida ja seejärel inimtõlkija poolt järeltoimetada. Edaspidi saab kasutada seda toimetatud väljundit uute masintõlkesüsteemide treenimiseks. (Sayers *et al.* 2021: 50). Siinkohal tuleb mainida, et halva inimtõlkija poolt järeltoimetatud teksti puhul võib tekkida olukord, et masintõlkesüsteeme treenitakse ebakvaliteetsete tekstidega, kuid võib eeldada, et see süsteem on nii läbi mõeldud, et kui mitmed tõlkijad need väljundid üle vaatavad, siis ei tohiks enam nendesse vigu sisse jääda.

## **2.2. Eelnevad uurimused masintõlke järeltoimetamisest**

Masintõlke järeltoimetamist on käsitletud paljudes erinevates uurimistöodes, sealhulgas ka Eestis kirjutatud magistritöodes/projektides.

Katre Sepp kirjutas 2017. aastal oma magistritöö teemal „Masintõlke järeltoimetamine“, kus ta analüüsis ühe tõlkebüroo raames tõlkijate suhtumist masintõlkesse ja järeltoimetamisse, kas masintõlge kiirendab ja tõhustab tõlkimist (inglise-eesti keelesuunal) ning kas masintõlget kasutades esineb rohkem või vähem vigu kui lihtsalt tõlkemälu abil tõlkides. Töös rakendati statistilist masintõlget. Uurimistöös viis Sepp läbi

katse, kus ta uuris katsealuste hoiakuid järeltoimetamise suhtes ning viis läbi järeltoimetamise katse, kus ta mõõtis katsealuste töökiirust ning analüüsis vigu. Töö tulemustes tõi Sepp välja, et suurema tõlkimiskogemusega tõlkijad suhtuvad masintõlkesse negatiivsemalt kui väiksema tõlkimiskogemusega tõlkijad. Samuti on professionaalsetel tõlkijatel oma meetodid juba välja kujunenud ning masintõlke lisamine aeglustab nende tööd. Katsest aga selgus, et masintõlke integreerimine tõstis töö tempot keskmiselt 25% (valitud tekstide puhul). Tähtsat rolli mängis ka teksti pikkus – mida lühemad laused, seda parem on masintõlge. Järeltoimetamise katses selgus ka, et masintõlge võib kaasa tuua erinevaid vigu, näiteks sõnajärjes või kohmakas sõnastuses.

Krete Saak kirjutas 2019. aastal oma magistriprojekti teemal „Tõlkemälu ja masintõlke vastete mõju kognitiivsele koormusele tõlkimisprotsessis“. Selle tööga uuris ta seda, millised on mõjud just Eesti tõlkijatele, kuna seda teemat on Eestis vähe uuritud (ning kognitiivne mõju põhineb tugevalt tõlke siht- ja lähtekeeltel). Katses oli 20 osalejat, kes tõlkisid teksti inglise-eesti suunal, nende kognitiivset koormust hinnati nii järeltoimetamisel kui osalise vastega segmentide tõlkimisel. Töös kasutati silmaseire meetodit. Katse tulemused näitasid, et kasutades tõlkides tõlkemälu ja masintõlget, siis täpse vastega segmente tõlkides on kognitiivne koormus tõlkijale väiksem (vähemalt poole väiksem võrreldes segmentidega, mis tuli algusest peale ise tõlkida). Järeltoimetamise uuringust aga selgus, et kognitiivne koormus tõlkijale on sarnane nullist tõlkimisega.

Katrin Shuyler kirjutas 2021. aastal magistr töö teemal „A Case Study on Post-editing Machine Translation: Tasks, Challenges, and Attitudes“. Oma töö raames viis Shuyler läbi kvalitatiivse juhtumiuuringu, milles uuris tõlkijate seisukohta seoses masintõlke järeltoimetamisega. Katsealused läbisid järeltoimetamiskatse ja seejärel tehti nendega intervjuud. Tulemuste analüüs näitas, et järeltoimetamisel peetakse raskeks otsustamist ja ka tõlketarkvara kasutamist (kuna kõik katsealused ei olnud tõlketarkvara kasutamiseга tuttavad). Katsealused suhtusid masintõlke kasutusse tavatingimustes (tööväliselt, oma tarbeks) pigem positiivselt, kuid professionaalsetes tingimustes siiski pigem skeptiliselt. Töös esitati tulemuste põhjal ka ettepanekuid, kuidas masintõlget hõlpsamalt töösse integreerida,

sh lisada täpsed juhised tõlkijatele töö kohta ja koolitada tõlkijaid masintõlkesüsteeme kasutama.

### 3. Masintõlke kvaliteedi hindamine

#### 3.1. Masintõlke hindamisest üldiselt

Masintõlke kvaliteedi hindamine tähendab masintõlke väljundile hinnangu andmist selle põhjal, kui hea või usaldusväärne see on. Halvasti koostatud masintõlke hindamiseks ja ümberkirjutamiseks võib järeltoimetajal kuluda rohkem aega kui nullist tõlkimiseks (Béchara *et al.*: 2021: 1). Béchara *et al.* (2021: 1) leiavad, et masintõlke kvaliteeti hindav süsteem võib aidata vältida frustratsiooni ja vähendada aega, mis kuluks järeltoimetajal otsustamiseks, kas masintõlge on järeltoimetamist väärt. Seega saaks kvaliteediskoori järgi potentsiaalselt otsustada, millised tõlked vajavad ainult järeltoimetamist ja mis oleks vaja nullist tõlkida.

Masintõlke väljundi hindamine ja veaanalüüs on väga olulised, samas keerulised ülesanded. Masintõlkesüsteemide hindamise olulisus seisneb selles, et selle tulemused näitavad, kui usaldusväärne on väljund ja neid tulemusi kasutatakse süsteemi täiustamiseks (Dorr *et al.* 2011). Käsitsi masintõlke hindamine on kallid ja aeganõudev, seetõttu on aastate jooksul uuritud palju erinevaid automaatse hindamise võimalusi (Toompalu 2011: 4). Kui varem on seda tehtud pigem käsitsi või kasutades erinevaid hindamismeetrikaid, nagu sõna veamäär WER (Word Error Rate), positsioonist sõltumatu sõna veamäär PER (Position independent word Error Rate), BLEU (Bilingual Evaluation Understudy), NIST (National Institute of Standards and Technology) ning METEOR (Metric for Evaluation of Translation with Explicit Ordering) (Toompalu 2011: 4), siis nüüd on hakatud masintõlkesüsteemidesse integreerima kvaliteedihinnangu võimalust. See tähendab, et masintõlkesüsteem annab ise automaatse skoori selles olemasoleva informatsiooni põhjal tõlke usaldusväärseuse ja kvaliteedi kohta.

Selleks, et masintõlkesüsteem annaks väljundile ehk koostatud tõlkele automaatse kvaliteediskoori, on vaja, et masintõlke hindamine toimuks samuti automaatselt. Automaatse hindamise alla liigitatakse süsteeme, mis hindavad masintõlget ilma inimeste osaluseta, mis tähendab, et inimese osalus automatiseeritud mõõtemetodites seisneb vaid ülesande

seadistamisel, nt andmete kogumises, annotatsioonides ja võrdlustõlgete kogumises (Chatzikoumi 2019: 140).

Ka automaatseid hindamismeetodeid on vaja hinnata, et teada, kui hästi hindamissüsteemid toimivad ja kuidas neid edasi arendada. Selleks on välja töötatud erinevad kriteeriumid, millel hindamine põhineb ja millele süsteemid vastama peaksid. Chatzikoumi (2019: 140) mainib, et kuigi sellised kriteeriumid on välja töötatud võrdlustõlkepõhiste meetrikate jaoks, siis kehtivad need samamoodi ka kvaliteedi hindamise meetrikate puhul. Chatzikoumi (2019:140) on oma töös toonud välja ka Banjeree ja Lavie (2005) poolt koostatud järgmised kriteeriumid, millele automaatne hindamissüsteem peaks vastama: kõrge vastavus inimeste poolt tehtud hindamistega, mis on kvantifitseeritud seoses tõlkekvaliteediga; tundlikkus kvaliteedinüansside suhtes süsteemide vahel või sama süsteemi väljundite suhtes selle arengu eri etappides; tulemuste järjepidevus (saada sarnased tulemused sarnaste tekstide puhul, mida sama süsteem on tõlkinud); usaldusväärsus (tagatud on kooskõla hindamiskoori ja tulemuste vahel); suur hulk erinevaid esindatud valdkondi ning kiirus ja kasutatavus. Koehn (2010) on seda nimekirja täiendanud järgnevate soovitustega: madalad kulud, võimalus optimeerida süsteemi otsest toimivust mõõdiku suhtes, skooride intuiitse tõlgendamise võimalus, järjepidavus ja korrektsus (Chatzikoumi 2019: 140).

Automaatsel kvaliteedi hindamisel on seega inimhindamisega võrreldes mitmeid eeliseid – see on kiirem, madalamate kuludega, vajab vähem inimeste sekkumist, on taaskasutatav ning ei vaja kakskeelseid hindajaid (Banjeree, Lavie 2005: 65). Samuti peetakse automaatset kvaliteedi hindamist inimhindamisest objektiivsemaks, sest ei ole võimalik tagada, et inimene annab samale tekstile kaks korda järjest täpselt sama hinnangu, või kindel olla, et mitu inimest nõustuksid teksti hindamise osas. Samas automaatne meetrika (juhul kui hindamisparameetreid pole muudetud) annab samale tekstile alati sama hinnangu (Chatzikoumi 2019: 144). Kuid kuna ka automatiseeritud meetrika põhineb inimtõlgetel, ehk kasutab tõlkele hinnangu andmiseks süsteemis olevaid võrdlustõlkeid, mis on tehtud inimeste poolt, siis ei saa seda siiski täiesti objektiivseks nimetada (seda tehakse ainult selleks, et rõhutada erinevust inimhindamise ja automatiseeritud hindamise vahel).

Automatiseeritud hindamisel on ka omad negatiivsed küljed. Üheks neist on see, et süsteemis peab olema piisavalt sarnaseid võrdlustõlkeid, sest muidu pole millegi baasil hinnangut anda ning sel juhul ei pruugi hinnangul olla mingit väärtust. Varasemad uuringud näitavad, et masintõlke arendajatel on keeruline ka saadud hinnangut tõlgendada ja võimatu anda teavet masintõlkesüsteemi täpsete tugevuste ja puuduste kohta (Zhou *et al.* 2008: 1121). Kuid need puudused mängivad suuremat rolli pigem masintõlke arendamisel, kui tõlkijate jaoks masintõlke järeltoimetamisel. Samas oleks ka tõlkijale kasulik, kui süsteem oskaks hinnangut ise tõlgendada ja pakkuda, kus võiks madalama hinnangu puhul viga peituda (kas viga on grammatiline või näiteks on mõni termin selline, mis tuleks kindlasti üle kontrollida).

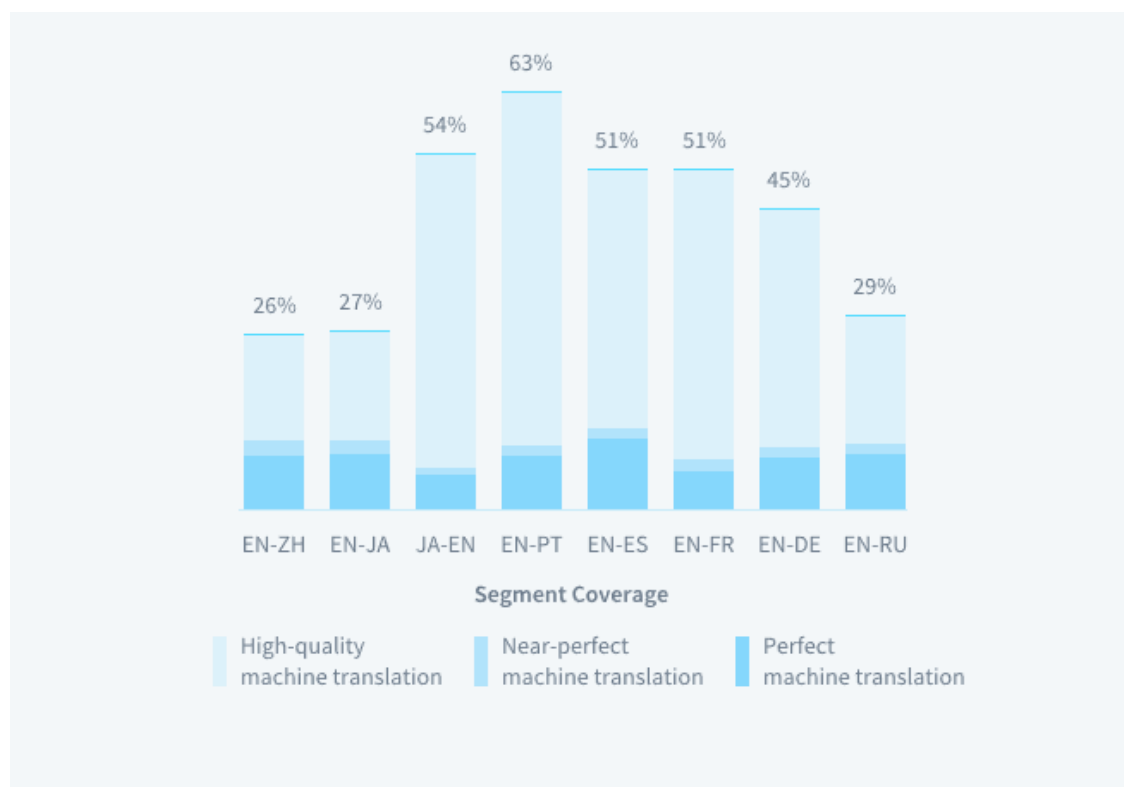
### **3.2. Masintõlke automaatsest kvaliteedi hindamisest (MTQE)**

Kui eelnevalt oli juttu erinevatest hindamissüsteemidest ja automaatsetest meetrikatest, mida varasemalt kasutatud on, siis see töö keskendub automaatsele masintõlke väljundi kvaliteedi hindamisele (MTQE ehk *Machine Translation Quality Estimation*).

2017. aastal viisid Moorkens ja O'Brien läbi küsimustikuuringu „Assessing User Interface Needs of Post-Editors of Machine Translation“, mille eesmärgiks oli välja selgitada, milliseid järeltoimetamisega seotud funktsioone tõlkijad masintõlkesüsteemidelt ootavad. Uuringus oli osalejaid 223 ning tulemused näitasid, et 81% (ehk 189) vastanutest sooviks, et oleks mingisugune usaldusskoor, mis näitaks masintõlke väljundi eeldatavat kvaliteeti. Nitzke ja Hansen-Schirra (2021) tõid oma raamatus välja, et mõned masintõlkesüsteemid juba pakuvad sellist võimalust. Nende hinnangute eesmärk on kiirendada otsust masintõlke väljundi kasulikkuse kohta, st kas usaldada ja/või järeltoimetada masintõlget või masintõlge kustutada ja selle asemel tekst ise tõlkida (Nitzke-Hansen Schirra 2021: 47).

Erinevalt eelnevalt kirjeldatud tavapäraest masintõlke hindamiseks kasutatavatest meetrikatest ei nõua MTQE inimese sekkumist, et tagada hinnang kvaliteedile (Bechara *et al.*

2021: 61) ning väljundi hindamine ei tugine võrdlustõlgetele. Masintõlke automaatset kvaliteedi hindamise teenust saab kasutada näitaks tõlkeabiprogrammis Memsources. Memsources'i kodulehel on välja toodud, et nende hindamissüsteemi saab kasutada rohkem kui 90 keelepaariga ning kõigi masintõlkesüsteemidega, mida Memsources toetab. Memsources'i esmane testimine on andnud tulemused, et kuni 60% segmentidest saavad kvaliteediskooriks 75%–100%. Tänu sellele peaks teoreetiliselt järeltoimetamine olema kiirem, sest sellise skoori saanud tõlked peaksid nõudma vaid toimetamist, mitte algusest lõpuni ise tõlkimist. Alltoodud tulpdiagrammil on välja toodud mõned keelesuunad, mille peal Memsources on masintõlke kvaliteedi hindamist testinud ja milliseid tulemusi need andnud on:



Joonis 2. Erinevate keelesuundade masintõlke kvaliteedi hindamise tulemused. Allikas: Memsources'i koduleht

Memsources'i kodulehel on veel välja toodud, et masintõlkesüsteemi poolt koostatud tõlgetele on võimalik saada nelja erinevat kvaliteediskoori, need on järgmised:

100% - ideaalne masintõlge, tõenäoliselt toimetamist ei vaja

99% - ideaalilähedane masintõlge, võimalik, et vajab vähest toimetamist, enamasti on probleem kas vormingu või kirjavahemärkidega

75% - hea masintõlge, sobib järeltoimetamiseks

MT – tõenäoliselt madala kvaliteediga masintõlge, ei ole soovitatav järeltoimetada, vaid kasutada näiteks abimaterjalina. (Memsources'i koduleht, viimati külastatud 25.04.2022).

Siinkohal tuleb mainida, et kvaliteediskoore ei kuvata alati protsentides ja selleks ei kasuta universaalset ühtset süsteemi, vaid erinevad süsteemid kasutavad erinevaid lahendusi. Näiteks on Euroopa Komisjonil oma masintõlkesüsteem eTranslation, mis on uue funktsioonina veel küll katsetamisjärgus, kuid see süsteem kuvab kvaliteediskooreid vahemikus 0–1, täpsusega kaks kohta peale koma, mis annab kõrgema täpsusega skoori kui Memsources. Samas puuduvad hetkel piisavad uuringud näitamaks, kas tõlkijal on abi sellisest täpsusest, kui ei teata täpselt, millel skoor põhineb ja kas skoori, mis on väiksem kui Memsources'i 75%, peaksid tõlkijad üldse vajalikuks näha, sest sellisel juhul on tõenäoline, et tekstis tuleks nagunii mõndagi muuta.

### **3.3. Eelnevad uurimused masintõlke automaatselt kvaliteedi hindamisest**

Joss Moorkens *et al.* uurisid oma artiklis „Correlations of perceived post-editing effort with measurements of actual effort“ muuhulgas seda, kuidas mõjutab n-õ usaldusskoori nägemine järeltoimetamist. Uuring koosnes kolmest erinevast etapist, mille esimeses etapis hindasid kuus kogenud tõlkijat töös esitatud tekste selle järgi, kui palju tööd nende arvates peaks järeltoimetamisele kuluma (tekstid olid inglise – portugali keelesuunal). Töö teises etapis, mõned nädalad hiljem, paluti neil tõlkijatel samad tekstid järeltoimetada, et tõlkijate tööd mõõta (kasutati silmaseire meetodit, mõõdeti aega ja kasutati TER meetrikat). Nende andmete põhjal koostati tekstidele usaldusskoored, seega erineb uuringus kasutatud hindamissüsteem automaatselt masintõlke hinnangust selle poolest, et põhineb

inimhinnangutel ning skoor ei näita mitte otseselt kvaliteeti vaid seda, kui palju tööd hinnanguliselt jältoimetamisele kuluma peaks. Uuringu kolmandas etapis paluti sama jältoimetamisülesanne läbida minimaalse jältoimetamiskogemusega üliõpilastel, kes lisaks ülesandele nägid ka eelnevalt määratud skooore (skoorid olid liigitatud valgusfoorivärvides). Tulemustes selgus, et usaldusskoori nägemine ei tekitanud tõlkijate töös märkimisväärseid erinevusi (autorid toovad välja, et üks põhjuseid võib olla see, et skoorid olid antud inimhindajate poolt, seega subjektiivsed ja võib-olla mitte nii usaldusväärsed).

Carlos Teixeira ja Sharon O'Brien uurisid oma töös „The Impact of MT Quality Estimation on Post-Editing Effort“ kuidas masintõlke automaatne hinnang tõlkijaid jältoimetamisprotsessis mõjutab. Katsealusteks olid 20 professionaalset tõlkijat ja uuringu keelesuunaks inglise-hispaania (katses oli kokku 4 jältoimetatavat teksti). Töö potentsiaalseks eesmärgiks oli tõlkijate tööd masintõlke abil hõlbustada (sh ka kognitiivset koormust vähendada) ning tekitada tõlkijates masintõlke vastu suuremat usaldust. Uuringus kasutati nelja erinevat kvaliteedihinnangut: ilma hinnanguta, täpne (automaatne) kvaliteedihinnang, ebatäpne (automaatne) kvaliteedihinnang ning inimhinnanguline kvaliteedihinnang. Hinnangud olid esitatud protsendisüsteemis. Katsealuste tööd mõõdeti ajaliselt, klahvivajutuste järgi ja silmaseire meetodil. Uuringu tulemused näitasid, et sellel, mis tüüpi hinnangut kuvati, ei olnud katsealuste jaoks erinevust, kuid mida paremat hinnangut tekstile kuvati, seda vähem aega katsealused selle teksti peale kulutasid. Samas leidsid töö autorid, et ei saa kindlalt väita, kas hinnangu nägemisel oli katsealustele mingi mõju.

Hannah Béchara *et al.* artiklis „The Role of Machine Translation Quality Estimation in the Post-Editing Workflow“ uuriti, kuidas mõjutab automaatne masintõlke kvaliteedi hindamine professionaalses keskkonnas masintõlke jältoimetamist. Uurimuse raames korraldati kasutajauuring, milles osalesid professionaalsed tõlkijad, kes jältoimetasid erinevaid lauseid (kokku 260 lauset). Katses osalenute produktiivsust hinnati kolmel viisil - nullist tõlkimisel, masintõlke jältoimetamisel ning jältoimetamisel koos kvaliteediskoori nägemisega. Mõõdeti nii lausete tõlkimiseks/jältoimetamiseks kulunud aega kui ka klahvivajutusi. Laused olid neile esitatud erinevate värvide järgi, et välja selgitada, kuidas

selline meetod tõlkijate tööd kiirendada võib. Värvide, mida kasutati, olid kollane (lausel ei ole masintõlget, tuleb nullist tõlkida), sinine (masintõlge on esitatud ilma kvaliteediskoorita, st tõlkija peab ise otsustama, kui palju ja mida muuta), roheline (on olemas masintõlge ja kvaliteediskoor on 75 (100-st) või rohkem) ja punane (on olemas masintõlge ja kvaliteediskoor, mis on alla 75). Töös esitatud lähtelaused olid inglise keeles ning sihtkeeleks oli hispaania keel. Tulemusest selgus, et masintõlke nägemisest võib tõlkijale töövoos olla kasu eelkõige siis, kui pakutud masintõlge on madalama skooriga, kuid tulemused näitasid ka seda, et hea kvaliteediskooriga masintõlge võib tööd kiirendada.

## **4. Uurimus neoromasintõlke järeltoimetamisest automaatse kvaliteediskoori toel**

Käesoleva uurimuse inspiratsiooniks on eelnevalt mainitud tööd, mis käsitlevad sarnaseid teemasid, nagu ka käesolev (nii järeltoimetamise kui MTQE funktsiooniga seotud tööd). Nende tööde põhjal on koostatud uurimustöö katses kasutatud meetodid, mis sellesse töösse kõige paremini sobivad. Sealhulgas on katse koostatud nii, et oleks kõigile katsealustele jõukohane ja ei tekiks raskusi, mis katsealuste fookuse potentsiaalselt kõrvale suunaksid, nagu näiteks võõraste tehnoloogiliste süsteemidega töötamine, ja meetodid on kohandatud vastavalt ressurssidele, kuna katses osalemine on vabatahtlik, katsealuste enda vabast ajast ja mitte tasustatud. Seega on katse koostatud katsealuste jaoks võimalikult mugavas formaadis ning piisava pikkusega, et see ei võtaks liiga kaua aega.

Uurimusel on kaks erinevat uurimisobjekti – esimeseks neist on küsimustik, milles uuritakse katsealuste suhtumist ja hoiakuid masintõlke järeltoimetamise ja MTQE funktsiooni kohta – selle eesmärk on uurida, kas masintõlke järeltoimetamine on üldse töö, mida tõlkijad on nõus tegema, või pigem soovitakse sellest hoiduda. Teiseks uurimisobjektiks on järeltoimetamise katse, mis paluti katsealustel sooritada küsimustiku ühe osana.

Katse kohta on püstitatud järgmised küsimused:

- Kas kvaliteediskoori nägemine mõjutab tõlkijat tegema vähem/rohkem muudatusi ja aitab otsuste tegemisele kaasa võrreldes hinnangu mitte nägemisega?
- Kas järeltoimetamise kiirus on erinev võrreldes tekste, millel on kuvatud kvaliteediskoor ja tekste, millel kvaliteediskoori ei kuvata?
- Kuidas suhtuvad tõlkijad MTQE integreerimisse järeltoimetamise protsessi?

Töös on püstitanud ka hüpotees, mis võiks katse tulemuste põhjal alust leida ning see on järgmine: Automaatse kvaliteediskoori kuvamine hõlbustab tõlkija tööd järeltoimetamisel, kuna aitab tõlkijal otsustada, millised laused on järeltoimetamist väärt ja kui palju ning milliseid tuleks tõlkida nullist.

## 4.1 Valim

Nii küsimustiku täitjad kui ka katses osalejad said valitud selle põhjal, et neil oleks mingisugune masintõlke järeltoimetamise kogemus juba olemas. Seega said katsealusteks nii tudengid kui täiendusõppijad, kes on osalenud masintõlke järeltoimetamise kursusel ning ka tõlkijad, kes on tõlkealal juba mõnda aega töötanud ja seeläbi järeltoimetamisega kokku puutunud. Katses osales 11 katsealust, kellel on kõigil erinev tõlkestaaž vahemikus alla aasta kuni 10 aastat. Katsealuste vanused koos tõlkestaažiga on välja toodud järgmises tabelis:

<b>Katsealune</b>	<b>Vanus</b>	<b>Tõlkestaaž aastates</b>
<b>Katsealune 1</b>	23	1–2
<b>Katsealune 2</b>	28	1–2
<b>Katsealune 3</b>	26	1–2
<b>Katsealune 4</b>	23	1–2
<b>Katsealune 5</b>	33	6–10
<b>Katsealune 6</b>	25	1–2
<b>Katsealune 7</b>	27	<1
<b>Katsealune 8</b>	50	<1
<b>Katsealune 9</b>	41	<1
<b>Katsealune 10</b>	26	3–5
<b>Katsealune 11</b>	23	<1

Tabel 1. Tõlkijate vanused ja tõlkestaaž aastates.

Tabelist on näha, et katsealuste hulgas on üsna erinevates vanustes ja erineva tõlkestaažiga inimesi. Väiksema tõlkestaažiga katsealuseid on rohkem seetõttu, et tegemist on tudengitega, kes on alles oma karjääri alguses.

## 4.2 Küsimustiku kirjeldus

Küsimustik, mis katsealustele saadeti, koostati LimeSurvey keskkonnas, mis on eelkõige teadusasutustele (haridusasutused, erinevad institutsioonid) loodud veebirakendus, kus saab mugavalt koostada küsimustikke. Kuna keskkonnas saavad katsealused jääda anonüümseks ja tagatud on andmete turvalisus, siis on see hea valik selliste uurimuste koostamiseks. Samuti on LimeSurvey'st võimalik faile alla laadida ning hiljem ka üles laadida, seega saab kogu küsimustiku ja katse läbi viia samas keskkonnas, mis on mugav nii katse koostajatele kui katsealustele. Katsealused pääsesid küsimustikule ligi lingi kaudu, mis neile edastati masintõlke järeltoimetamise kursusel või e-kirja kaudu.

Küsimustik hõlmas endas küsimusi järeltoimetamise kohta ja ka järeltoimetamise katset. Küsimustiku esimese poole küsimuste valimise põhimõtteks oli see, et need juhataksid võimalikult hästi sisse katse, mida katsealused peale neile küsimustele vastamist sooritama asusid. Seega hõlmasid need küsimused nii masintõlget, masintõlke järeltoimetamist kui ka küsimusi MTQE funktsiooni tundmise kohta. Küsimustikku sidus kahe küsimuste osa vahel asuv katse, mis tuli katsealustel sooritada enne, kui nad said edasi liikuda viimaste küsimuste juurde, mis käsitlesid äsja sooritatud katset ning hõlmasid küsimusi näiteks arvamusest MTQE funktsiooni kasulikkuse kohta ning nende enda üldise arvamuse kohta. Küsimustiku tulemusi saab mõneti võrrelda eelnevate töödega, nagu näiteks Katre Sepa (2017) magistritöö, millest oli juttu alapeatükis 2.2 – ka tema koostas oma töö raames küsimustiku ning järeltoimetamisega seotud katse, kuid statistilise masintõlke kohta. Käesolevas töös on see-eest kasutatud neuromasintõlget. Küsimustiku tulemustest võib kasu olla näiteks masintõlkesüsteemide arendajatele, kes arendavad neuromasintõlke süsteeme ja soovivad teada, kas tõlkijate suhtumine on masintõlkesüsteemide arenedes nendesse muutunud kuid näiteks ka õppejõududele, kes tudengitele selliseid süsteeme tutvustada saavad – mida positiivsem tagasiside, seda rohkem oleksid tudengid sellistest tööviisidest huvitatud.

Järgnevas alapeatükis on küsimuste kaupa välja toodud katsealuste vastused. Kuna vastustest sai nii kvantitatiivseid andmeid (nt skaalal 1 – 10 oma pädevuse hindamine) kui ka kvalitatiivseid andmeid (nt katsealuste isiklik arvamus MTQE funktsioonist), siis on andmed

esitatud kõige sobivamatel viisidel lähtudes küsimuse sisust. Osad vastused on esitatud tabelitena ning kirjalike vastuste puhul on võimalusel sarnased vastused kokku koondatud ja ülejäänud ükshaaval välja toodud.

### 4.3 Küsimustiku tulemused

**1. Kas kasutate tõlkides masintõlkesüsteemide abi (nt lülitate tõlkeabiprogrammis masintõlke funktsiooni sisse)? Muu vastuse korral kirjutage oma vastus palun kommentaarina.**

10 katsealust vastasid, et kasutavad masintõlkesüsteemide abi.

1 katsealune vastas, et ei kasuta masintõlkesüsteeme.

**2. Kui palju kasutate oma töös masintõlget? (1 - ei kasuta üldse; 10 - kasutan väga tihti).**

Kuna küsimusele sai vastata valides enda jaoks sobiva variandi skaalal 1–10, siis on tulemused esitatud tabeli kujul ning järgmiselt:

Katsealune	1	2	3	4	5	6	7	8	9	10	11
Vastus	5	2	5	8	10	6	8	3	3	5	7

Tabel 2. Töös masintõlke kasutamine skaalal 1–10.

Tabeli tulemustest on näha, et on katsealuseid, kes kasutavad masintõlget väga tihti ja ka neid, kes kasutavad seda pigem vähe. Kui esimesele küsimusele vastas üks katsealune, et ei kasuta masintõlkesüsteemide abi, siis skaalal 1–10 siiski keegi 1-te ei valinud, seega võib järeldada, et mingil määral kasutavad kõik vastanud siiski masintõlkesüsteeme. 5 katsealust vastasid, et kasutavad masintõlget pigem rohkem kui vähem (>5) ning 3 katsealust valisid oma vastuseks täpselt keskmise 5-e. Keskmiseks vastuseks nende andmete põhjal tuleb 5,6.

Seega jääb masintõlkesüsteemide kasutamine tulemuste põhjal keskmiselt poole peale, mis näitab süsteemide mõõdukat kasutamist.

### **3. Kuidas suhtud masintõlke järeltoimetamisse? Kirjelda lühidalt oma mõtteid.**

4 katsealust on arvamusel, et masintõlge on omal kohal ja heaks lahenduseks siis, kui tõlke kvaliteet ei ole väga oluline või on vaja mingit infot väga kiiresti edastada. Samuti on välja toodud, et oma rolli mängib ka masintõlkesüsteem, mida kasutatakse, sest ka nende väljundite kvaliteeditase võib olla väga erinev.

1 katsealune kirjutas, et lihtsamate tekstide tõlkimisega saab masintõlkesüsteem mõnikord päris hästi hakkama ning järeltoimetada tuleb väga vähe.

1 katsealune kirjutas, et masintõlge on hea, kuna teeb esmase töö tõlkija eest ära.

1 katsealune kirjutas, et tal ei ole järeltoimetamise osas väga tugevat seisukohta, sest kasutab seda pigem harva, kuid kui kasutab, siis reeglina tuleb toimetada ja muuta praktiliselt iga lauset.

4 katsealust on arvamusel, et masintõlke järeltoimetamine on sama keeruline ja ajakulukas või isegi keerulisem ja ajakulukam, kui nullist tõlkimine. Seega võib järeltoimetamine olla kurnav ja ebameeldiv, sest tööle lisandub veel üks element tähelepanelikkuse näol ning on pidev dilemma, mida parandada ja mida mitte.

3 katsealust kirjutasid, et nende jaoks on lihtsam tekstid otsast lõpuni ise tõlkida, kasutades masintõlget üksnes sõnaraamatu või sünonüümide pakkuja rollis.

1 katsealune toob välja, et järeltoimetamisel tekitab lisapingeid klientide väärarusaam, et masin on suurema töö juba ära teinud.

1 katsealune lisab siiski, et kuna masintõlge on pidevas arengus, siis ilmselt tulevikus muutub järeltoimetamine oluliselt lihtsamaks ja seda võetakse rohkem kasutusse.

**4. Kui pädevaks pead end masintõlke järeltoimetamisel? Vali oma sobivaim variant (1- ei pea end pädevaks; 10 - väga pädev).**

Katsealune	1	2	3	4	5	6	7	8	9	10	11
Vastus	6	6	7	5	6	4	3	1	4	5	7

Tabel 3. Pädevus järeltoimetamisel skaalal 1 – 10.

Selle küsimuse tulemused näitavad, et väga pädevaks ei pea end järeltoimetamisel keegi, sest kõrgeimaks hinnanguks on vaid 7, madalaimaks aga 1, mis on vähim, mida valida sai. Keskmiseks hinnanguks tuleb nende vastuste põhjal 4,9, mis jääb natukene alla keskmise. Sellegipoolest on lausa 5 katsealust paigutanud ennast skaalal üle keskmise pädevuse alasse ning 2 täpselt keskele.

**5. Kas oled kokku puutunud korrektse masintõlkega, mis ei vajagi järeltoimetamist? Muu vastuse korral kirjuta oma vastus kommentaarina.**

4 katsealust vastasid, et on kokku puutunud korrektse masintõlkega.

5 katsealust vastasid sellele küsimusele eitavalt.

2 katsealust kirjutasid oma vastuse kommentaarina ning tõdesid, et on harva korrektse masintõlkega kokku puutunud küll aga seda pigem lihtsamate tekstide (lihtlausete) puhul. 1 katsealune tõdeb samas, et suhtub ka nendesse alati ettevaatlikkusega.

**6. Kas oled masintõlke järeltoimetamisel kokku puutunud raskesti tuvastatavate vigadega?**

Sellele küsimusele vastasid kõik 11 katsealust jaatavalt.

1 katsealune lisas kommentaarina juurde, et need võivad olla sellised vead, mida ei pruugi leida, kui lähteteksti väga hoolega ei vaata. Näiteks jääb märkamata mõni väike tähendusvarjund, mis on teksti jaoks tegelikult väga oluline.

1 katsealune tõdes, et vead jäävad märkamata vähese kogemuse tõttu järeltoimetamisel.

## **7. Millised on masintõlke järeltoimetamise suurimad eelised ja puudused?**

10 katsealust 11-st tõid välja, et masintõlke järeltoimetamise suurimateks eelisteks on kiirus ja aja kokkuvõid. Teised eelised, mida välja toodi, on järgmised:

- hea masintõlkesüsteemi puhul ei pea tavaliselt muretsema grammatika pärast;
- masintõlge teeb suure töö ära ja jääb ainult järeltoimetamise vaev;
- masintõlge aitab nn valge lehe painest üle saada, st annab inspiratsiooni isegi siis, kui tekst on vigane;
- masintõlge võib anda ideid tõlkeprobleemide lahendamiseks või keerulisematest kohtadest aru saamiseks.

Puuduste osas oli läbivaks vastuseks see, et halva masintõlke kvaliteedi puhul võib järeltoimetamine rohkem aega võtta kui nullist tõlkimine. Teised puudused, mida välja toodi, on järgmised:

- peab olema väga tähelepanelik, et vigu sisse ei jääks, sest tekstis võib olla selliseid sõnu, millest pilk üle libiseb, sest need tunduvad esmapilgul õiged;
- raskesti leitavate vigade võimaluse tõttu tekib võlts turvatunne;
- masintõlkest endast otseselt mittesõltuv puudus on selle kasutamine liiga lõdva randmega, sest see ei sobi igale poole;
- terminoloogilised probleemid, kuna programmid ei suuda mõnikord tuvastada tähenduslikke erinevusi ning terminile võib tulla vaste, mis ei pruugi konkreetse konteksti sobida;
- klient peab täpselt määratlema, millist kvaliteeti oodatakse, sest ei ole veel ühte kindlat järeltoimetamise reeglistikku.

Vastuste põhjal on näha, et järeltoimetamisele osatakse nimetada pigem rohkem erinevaid puuduseid, kui eeliseid. Kõik väljatoodud puudused on väga asjakohased ning

olnud läbivaks teemaks ka eelnevates uurimustes, mis samal teemal läbi viidud (nt uurimused, mida tutvustas alapeatükk 3.3). Samas oli enamustes vastustes läbivalt mainitud, et järeltoimetamise eeliseks on kiirus, seega oli kõigis vastustes esindatud siiski nii eelised kui puudused.

#### **8. Kas oled kuulnud funktsioonist MTQE (Machine Translation Quality Estimation)?**

8 katsealust vastasid sellele küsimusele, et on sellest funktsioonist kuulnud

3 katsealust vastasid sellele küsimusele, et ei ole sellest funktsioonist varem kuulnud.

#### **9. Kas oled oma töös MTQE funktsiooni kunagi kasutanud? Kui jah, siis kirjelda lühidalt kus ja kuidas sellega kokku oled puutunud.**

8 katsealust vastasid, et ei ole oma töös MTQE funktsiooni kasutanud.

3 katsealust vastasid, et on MTQE funktsiooni kasutanud. Nendest 2 katsealust kirjutasid, et on kasutanud MTQE funktsiooni, kuna näiteks Memsources'i tasulises versioonis on see funktsioon automaatselt sisse lülitatud. 1 katsealune kirjutas, et on kasutanud seda ülikoolis, kus ta kirjutas MTQE teemal seminaritöö, milles võrdles automaatseid kvaliteedimäärajaid järeltoimetajate saadud tulemustega.

### **4.4 Katse kirjeldus**

Katse eesmärk oli välja selgitada, kas kvaliteediskooride lisamine järeltoimetamiseks mõeldud tekstile vähendab tööks kulunud aega ja kas rohkem muudatusi tehakse tekstis, kus on segmentidele lisatud kvaliteediskoorid, või tekstis, kus skoori juures pole. Katse viidi läbi

2022. aasta maikuu ning katses osales 11 katsealust ehk katsel oli sama valim, mis küsimustikul.

Katsesse valitud tekstide masintõlkimise tarbeks kasutati tõlkeabiprogrammi Memsources. Memsources'i tasuta versioonis saab tekste tõlkides valida erinevate masintõlkesüsteemide vahel või rakendada ka mitut korruga. Samuti on võimalik manuaalselt masintõlkesüsteeme juurde lisada, kui seal pakutud valikus ei leidu sobivat. Selleks tuleb Memsources'is menüüs klõpsata Memsources Translate'i peale, kus saab hõlpsalt erinevaid masintõlkesüsteeme sisse-välja lülitada. Selle katse raames testiti erinevate masintõlkesüsteemide ja ka nende kombinatsioonide väljundeid, et kvaliteediskooride hulgas oleks esindatud võimalikult palju erinevaid skooride. Masintõlkesüsteemidest kasutati Google Translate'i, DeepLi ja Microsoft Translatorit, põhiliselt seetõttu, et nendes on olemas inglise-eesti keelesuund ja need pakuvad üsna korralikke tõlkeid, sest neil süsteemidel on suured andmebaasid, mille põhjal need süsteemide algoritmid õppida saavad. Peale masintõlkesüsteemide paika panekut valib Memsources'i süsteem oma algoritmide järgi ise, milliste süsteemide tõlkeid see tekstile annab. Lisaks ei kasutatud selle uurimuse raames tekstide tõlkimisel ühtegi vastetega tõlkemälu, sest töö eesmärk on uurida just masintõlke enda väljundeid ja nende automaatseid kvaliteediskooride.

Katse tarbeks valiti tekstid, mille segmentidele annaks MTQE funktsioon võimalikult erinevad skoorid ja seejärel tõlgiti need tekstid Memsources'i abil ära. Tekste valides lähtuti ka leheküljel 11 asuvast Joonisest 1, millel on välja toodud, et kasutusjuhendite tõlkimine jääb täpselt masintõlke järeltoimetamise ja inimtõlke vahele, st sellised tekstid nõuavad piisavalt täpsust, et mitte ainult masintõlkida, kuid samas võib masintõlge anda piisavalt hea väljundi, mis nõuab inimtõlkija poolt vaid järeltoimetamist. Katse aluseks sai seega kaks lühemat tekstilõiku, mis on mõlemad pärit ühe kodumasina kasutusjuhendist. Tekstid on valitud hoolikalt, sest tähtis on see, et need oleksid sõnade arvult sama pikad ning tekstides käsitletav teema sama raskustasemega, sama tekstitüüpi ning võimalikult sarnase sisuga. See on oluline, sest sel juhul saab eeldada, et tekstide järeltoimetamiseks kulub sama kaua aega, seega saab hiljem kahte teksti võrrelda nii aja kui tehtud muudatuste alusel. Esimese teksti

pikkuseks on 183 sõna ning teise teksti pikkuseks 189 sõna. Mõlema teksti lähtekeeleks on inglise keel ja sihtkeeleks eesti keel.

Kuigi Memsources'i on limiteeritud kasutusõigustega võimalik ka kasutada tasuta, siis Memsources Translate'i funktsioon on tasuline ning ei ole kõigile kättesaadav. Samuti, nagu tõi Shuyler (2021) oma töös välja, tekib tõlkijatel siiani tehnoloogilisi raskusi erinevate süsteemide kasutamisel. Seega, kuna käesoleva töö eesmärk ei ole tõlkeabiprogrammide pädevuse välja selgitamine, vaid automaatse kvaliteediskoori uurimine, peaks katse olema katsealustele koostatud võimalikult tuttavas keskkonnas ja sooritamiseks mugav. Katse sai koostatud DOCX formaadis, et seda saaks avada tekstitöötlusprogrammides (millest leidub ka mitmeid vabavaralisi variante) ning katsealused saaksid valida omale kõige sobivama keskkonna, kus katset sooritada. Memsources'ist on võimalik ka alla laadida kakskeelne DOCX formaadis fail, mis tegi katse koostamise väga lihtsaks, sest tabelid olid allalaaditud formaadis juba vormistatud nii, nagu on need kuvatud ka tõlkeabiprogrammides.

Peale tekstide allalaadimist koostati fail, mille algusesse lisati põhjalikud juhised katses osalemise kohta. Sellele järgnes kaks teksti, mille vasakul poolel on lähtekeelne tekst ning paremal poolel sihtkeelne tekst, mis vajab järeldoimetamist. Esimesel tekstil on kõige parempoolsemas tulbas näha ka MTQE funktsiooni poolt pakutud automaatne kvaliteediskoor.

Remember that this User Manual is also applicable for several other models.	Pidage meeles, et see kasutusjuhend on rakendatav ka mitmete teiste mudelite puhul.	75
Differences between models will be identified in the manual.	Erinevused mudelite vahel tuvastatakse käsiraamatus.	MT
CONTENT	SISU	MT
<b>Safety Warnings</b>	<b>Ohutushoiatused</b>	100
<b>Technical Aspects Of Your Machine</b>	<b>Teie masina tehnilised aspektid</b>	75
<b>Using Your Machine</b>	<b>Masina kasutamine</b>	100

Joonis 3. Vaade katse tekstist, mis kuvati koos kvaliteediskooridega.

Katse on täies mahus nähtaval töö lõpus lisana. Katse sooritamisel paluti katsealustel võimalusel kasutada muutuste jälitaja (Track Changes) funktsiooni, et hiljem selle järgi

tehtud muudatuste hulka vaadata (kuid see polnud nii oluline, sest MS Wordis on võimalik ka originaali ja muudetud teksti võrrelda, mis annab muudatustest samasuguse pildi). Muudatuste hulka arvestati hiljem analüüsi tarbeks nii sõnade kui ka kirjavahemärkide baasil – sinna hulka loeti nii kustutatud sõnad, lisatud sõnad kui muudetud sõnad, samuti ka kustutatud ja/või lisatud kirjavahemärgid.

Katset sooritades paluti katsealustel mõõta ka iseseisvalt võimalikult täpselt (vähemalt minuti täpsusega) aega, mis neil kummagi teksti järeltoimetamiseks kulus ning seejärel kirjutada need ajad küsimustikus viidatud kohta. Et katsealustel ei läheks meelest kahe teksti vahel esimesele tekstile kulunud aega kirja panna, oli nende kahe teksti vahel selleks meeldetuletus. Vastasel juhul võib tekkida situatsioon, et tööhoos asub katsealune teise teksti juurde ning alles selle lõpus meenub, et esimese teksti peale kulunud aeg ei saanudki kirja panna.

Kuna katse oli osa küsimustikust, siis said katsealused katse faili alla laadida otse küsimustikust ning seejärel said nad laadida sooritatud katsega faili sinna samale küsimustiku lehele üles. Peale katse sooritamist ja üles laadimist pidid katsealused vastama veel mõnele küsimusele, mis käsitlesid äsja sooritatud katset ja MTQE funktsiooni. Vastused neile küsimustele on välja toodud küsimustiku vastustest eraldi, alapeatükis 4.5 peale katse tulemusi.

## **4.5 Katse tulemused**

Selles alapeatükis tutvustatakse katse tulemusi. Kuna katsealused on väga erineva tõlkestaažiga, siis on ka tulemused iga katsealuse kohta väga erinevad, sealhulgas ajalised erinevused ja ka tehtud muudatuste hulgad. Sellegipoolest, kuna katses osalejaid on piisav hulk, võib katse tulemuste põhjal teha paikapidavaid järeldusi.

Katse tulemused on esitatud tabelitena, kus on esitatud iga katsealuse kohta andmed eraldi ning nende põhjal ka arvutatud keskmised väärtused. Tabelites on 1. teksti all peetud silmas teksti, millele kuvati kvaliteediskoorid ning 2. teksti all teksti, millel kvaliteediskoore ei kuvatud. Kõik katsealused on tabelites kuvatud alati samas järjestuses, seega kui on soov kõrvutada näiteks ühe katsealuse katseks kulunud aeg tehtud muudatuste arvuga, siis need andmed leiab alati sama katsealuse numbri alt.

Järgnevas tabelis on näha, kui palju aega (minutites) kulus katsealustel ühe või teise teksti järeltoimetamiseks ning milline oli kõigi katsealuste tulemusi arvestades keskmine ajakulu. Kuna katsealustel paluti ise aega võtta ja enamus vastuseid esitati vaid minutites, siis on tulemustes kõik esitatud ajad, millele olid peale minutite veel juurde lisatud ka sekundid, ümardatud lähima poole- või täisminutini.

<b>Katsealuse nr</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>Keskmine aeg (min)</b>
<b>1. tekstile kulunud aeg</b>	27	17,5	17	10	9	9	20	16,5	15	5,5	31	16,1
<b>2. tekstile kulunud aeg</b>	45	17	30	5	15	9,5	20	11,5	13	6,5	53	20,5

Tabel 4. Mõlema teksti järeltoimetamiseks kulunud ajad.

Nagu tabelist näha, siis katses kogutud andmete põhjal kulus 1. teksti järeltoimetamisele keskmiselt veidi üle 16 minuti, kuid 2. teksti järeltoimetamiseks kulus katsealustel 20,5 minutit. Sealhulgas toimetas 6 katsealust kvaliteediskooridega teksti vähema ajaga, kui ilma skoorideta teksti ning 4-1 katsealusel kulus kvaliteediskooridega teksti järeltoimetamiseks rohkem aega. 1 katsealustest järeltoimetas mõlemad tekstid sama ajaga ehk 20 minutiga. Keskmiselt kulus kvaliteediskooridega teksti järeltoimetamiseks 21,5% võrra vähem aega kui ilma kvaliteediskoorideta teksti järeltoimetamiseks. Veel on huvitav see, kui erineva kiirusega katse sooritati – 1. tekst järeltoimetati kõige kiiremini 9 minutiga

ja kõige aeglasemini 31 minutiga, 2. tekst järeltoimetati kõige kiiremini lausa 5 minutiga ja kõige aeglasemini 53 minutiga.

Järgmises tabelis on välja toodud, kui paljudes segmentides tegid katsealused muudatusi tekstis, millel olid nähtaval kvaliteediskoorid. Andmed on esitatud nii muudetud segmentide koguarvuna kui ka kvaliteediskooride järgi jaotatuna. Tekstis, millel olid juures kvaliteediskoorid, oli kokku 27 segmenti. Kvaliteediskoorid, mis selles tekstis olid esindatud, olid 100%, 75% ja MT. Samasugused kvaliteediskoorid olid esindatud ka tekstis, mille kvaliteediskoorid olid katsealustele varjatud. Kuna väga keeruline oleks leida kahte (järjestikust, mitte suvaliste lausetega) teksti, kus kahe teksti puhul oleks esindatud kõik skoorid võrdselt, siis on mõlemas tekstis erinev hulk erinevate kvaliteediskooridega segmente. Sellegipoolest on andmed võrreldavad, sest mõlema teksti puhul domineerib kvaliteediskoor 75% ning skooridega 100% ja MT on esindatud vaid mõned üksikud segmendid. Kvaliteediskooridega teksti segmendid jaotuvad vastavalt skooridele järgmiselt:

- kvaliteediskooriga 100% on 6 segmenti,
- kvaliteediskooriga 75% on 17 segmenti,
- kvaliteediskooriga MT on 4 segmenti.

Katsealuse nr	1	2	3	4	5	6	7	8	9	10	11	Keskmiselt muudetud segmente
<b>100%</b>	0	2	2	2	0	0	2	2	0	0	2	1
<b>75%</b>	12	13	14	13	11	7	13	13	9	8	16	11,7
<b>MT</b>	3	1	3	3	3	1	2	2	1	3	3	2,3
<b>Kokku</b>	15	16	19	18	14	8	17	17	10	11	21	15

Tabel 5. Kuvatud kvaliteediskooridega teksti jaotus muudetud segmentide arvude järgi.

Keskmiselt tehti esimeses tekstis muudatusi 15-s segmendis. Sealhulgas kõige väiksem muudetud segmentide arv oli 8 ja kõige suurem muudetud segmentide arv oli 21.

Protsentuaalselt tehti keskmiselt muudatusi 55% segmentides. See tähendab, et keskmiselt muudeti veidi üle poolte segmentidest. Nagu tabelist näha, siis kvaliteediskooriks 100% saanud segmentide puhul jaotusid katsealused pooleks – pooled ei muutnud mitte ühtegi 6-st segmentist ning pooled muutsid neist segmentidest 2-te segmenti. Seega keskmiselt muudeti protsentuaalselt vaid 16,6% nendest segmentidest. Kvaliteediskooriks 75% saanud segmentidest muudeti keskmiselt 11,7 segmenti, ehk ümmarguselt 12. Kuna kokku oli 17 segmenti, mis said skooriks 75%, siis protsentuaalselt muudeti keskmiselt 69% nendest segmentidest. Kvaliteediskooriks MT saanud segmente oli selles tekstis vaid 4, neist keskmiselt muudeti 2,3 segmenti ja seega protsentuaalselt muudeti 57,5% nendest segmentidest.

Järgmises tabelis on esitatud andmed selle kohta, kui palju segmente muudeti varjatud kvaliteediskooridega tekstis. Selle teksti segmentide koguarv oli 21. Olgugi et katsealused neid skooore selle teksti puhul ise ei näinud, on siiski põnev võrrelda neid andmeid teise teksti andmetega, mille skooore katsealuseid nägid (kas varjatud kvaliteediskooridega teksti puhul muudetakse rohkem või vähem segmente kui teksti puhul, kus kvaliteediskoorid on kuvatud). Teise teksti puhul on segmentide jaotus järgmine:

- kvaliteediskooriga 100% on 4 segmenti,
- kvaliteediskooriga 75% on 15 segmenti,
- kvaliteediskooriga MT on 2 segmenti.

<b>Katsealuse nr</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>Keskmiselt muudetud segmente</b>
<b>100%</b>	0	1	1	1	0	0	0	0	0	0	0	0,3
<b>75%</b>	7	9	13	11	12	5	7	7	7	9	11	8,9
<b>MT</b>	1	0	1	1	2	0	2	2	1	0	1	1
<b>Kokku</b>	8	10	15	13	14	5	9	9	8	9	12	10,2

Tabel 6. Varjatud kvaliteediskooridega teksti jaotus muudetud segmentide arvu järgi.

Varjatud kvaliteediskooridega teksti puhul oli kõige väiksem muudetud segmentide arv 5 ja kõige suurem muudetud segmentide arv 15. Keskmiselt tehti selles tekstis muudatusi 10-s segmendis. Protsentuaalselt tehti keskmiselt muudatusi 48% segmentides, seega peaaegu pooltes segmentides. Kvaliteediskoori 100% saanud segmentides muutsid 4-st segmendist 1-te ainult 3 katsealust, mis teeb keskmiselt muudetud segmentide protsendiks vaid 7,5%. Kvaliteediskooriga 75% segmentidest muudeti keskmiselt 8,9 segmenti ehk 59% segmentidest. Kvaliteediskoori MT saanud 2-st segmendist muudeti keskmiselt 1-te segmenti, seega 50% segmentidest.

Järgmises tabelis on näidatud muudatuste hulgad, mis kummaski tekstis tehti. Kuna tekstides oli peaaegu sama suur sõnade hulk, siis on hea võrrelda, kas kvaliteediskooridega tekstis tehti vähem või rohkem muudatusi võrreldes varjatud kvaliteediskooridega tekstis. Välja on toodud ka keskmised muudatuste hulgad.

Katsealune	1	2	3	4	5	6	7	8	9	10	11	Keskmiselt muudatusi
1. tekst	57	55	78	73	81	27	73	46	43	35	124	62,9
2. tekst	34	40	112	43	125	33	42	44	35	33	73	55,8

Tabel 7. Tekstides tehtud muudatuste hulk.

Nagu tabelist näha, siis 8 katsealust tegid tekstis, millel olid kuvatud kvaliteediskoorid, rohkem parandusi, kui kvaliteediskoorideta tekstis. Keskmise muudatuste arv on samuti just kvaliteediskooridega teksti puhul suurem (kuigi seda ainult mõne muudatuse võrra). Samas on katsealuseid, kes just vastupidi tegid palju rohkem muudatusi varjatud kvaliteediskooridega tekstis. Suur muudatuste erinevus erinevate katsealuste tekstides tuleneb sellest, et osad katsealused muutsid ainult üksikuid sõnu, kuid osad muutsid mõnes kohas terveid segmente.

Katse lõpus küsiti katsealustelt veel äsja sooritatud katse kohta mõned küsimused ning paluti kirjeldada ka nende üldiseid mõtteid MTQE funktsiooni kohta. Need küsimused koos vastustega on töös järgmisena esitatud.

### **1. Kas pidasid kvaliteediskoori kasulikuks? Põhjenda oma vastust.**

11 katsealust vastasid, et kvaliteediskoori nägemine oli mõne lause puhul kasulik.

2 katsealust vastasid, et nende jaoks skoori nägemine kasulik ei olnud.

1 katsealune kirjutas, et kvaliteediskoor toetas järeltoimetamist.

2 katsealust kirjutasid, et maksimaalse skooriga (100%) segmentide puhul oli tihti tegemist üksikute sõnade või fraasidega (nt pealkirjad), mis on üsna levinud ning üldiselt ei vaja ka tõlkimisel pikka mõtlemist.

2 katsealune kirjutasid, et kvaliteediskoorist võib kasu olla pigem madalama skoori puhul (MT), sest see muudab tähelepanelikumaks.

1 katsealune kirjutas, et madal skoor võib tõsta tähelepanu küll selle segmendi osas, kuid seetõttu võib jälle tekkida oht, et libised üle mõnest 100%-st segmendist, millel selles kontekstis tegelikult ei ole õige tõlge.

3 katsealust vastasid, et selle teksti puhul nad kvaliteediskoorist ei lähtunud ja selle teksti puhul ei tekkinud ka selliseid kõhklusi, mis oleksid pannud skoori vaatama (mõne teise teksti puhul oleks seda rohkem kasutanud).

1 katsealustest kirjutas, et talle tundus, et ka mõned 100%-sed vasted vajasid parandamist.

1 katsealune kirjutas, et üldiselt ei ole tema jaoks vahet, milline on kvaliteedihinnang, sest iga lause peab mõttega siiski üle kontrollima.

**2. Kui palju lisab kvaliteediskoor sinu jaoks väärtust järeltoimetamisele (nt lihtsustab tööd, aitab kaasa otsustamisel, kui palju toimetada jne)? Vali sobiv hinnang (1 - ei lisa üldse; 10 - kvaliteediskoor on väga kasulik)**

Katsealune	1	2	3	4	5	6	7	8	9	10	11
Vastus	3	3	5	10	6	6	6	3	3	8	7

Tabel 8. Kvaliteediskoori väärtus katsealustele skaalal 1–10.

Selle küsimuse tulemused näitavad, et kvaliteediskoori kohta on arvamusi seinast seinast, kuid suurem osa vastanutest, ehk 6 vastanut, peab seda siiski pigem positiivseks lisaväärtuseks järeltoimetamise protsessile. Keskmiseks hinnanguks tuli nende andmete põhjal 5,5.

### **3. Kas juba kasutad/hakkaksid kasutama seda funktsiooni ka tulevikus?**

1 katsealune vastas, et ta juba kasutab seda funktsiooni oma töös.

6 katsealust vastas sellele küsimusele, et nad kasutaksid seda funktsiooni tulevikus kindlasti uuesti.

3 katsealust vastasid, et nad ei oska öelda, kas hakkaksid seda funktsiooni tulevikus kasutama.

1 katsealune vastas, et ta ei kasuta ega hakka seda funktsiooni tulevikus kasutama.

### **4. Sinu mõtted automaatselt kvaliteediskoorist. Oodatud on igasugused kommentaarid ja tagasiside!**

5 katsealust leidsid, et see on kasulik ja hea abivahend ning kindlasti võib järeltoimetamisele kaasa aidata, sealhulgas 3 katsealust kirjutasid, et need skoorid aitasid neil hinnata, kui valvas peaks järeltoimetamisel olema ja millistele segmentidele peaks rohkem

tähelepanu pöörama. 1 katsealune lisas, et kui neid skoores ei olnud, siis kontrollis ta igat segmenti palju põhjalikumalt ja tõenäoliselt mõtles ka üle.

2 katsealust leidsid samuti, et see on kasulik, eeldusel, et kontrollitakse üle ka segmentid, mis on saanud 100% hinnangu.

2 katsealust leidsid, et see võib olla kasulik, kuid selle liigne usaldamine võib põhjustada vigadest nn üle libisemist.

1 katsealune kirjutas, et see võib anda tähendusele ja lauseehitusele hea baasi, aga mõnede lausete puhul pigem segab, sest endale võib lause tunduda segane ja tekib tunne, et tahaks kõik ümber teha.

1 katsealune kirjutas, et ta ei vaadanud skoores üldse, sest kogemus on näidanud, et need ei ole usaldusväärsed, kuna tekstides, mida ta on pidanud tõlkima, on tihti ka 100% skooridega segmentide sees olnud vigu.

## 5. Järeldused

Selle uurimistöö raames koostati nii katse kui ka küsimustik, mille sisud on omavahel seotud. Selle eesmärk oli kõrvutada katsealuste suhtumised masintõlke järeltoimetamisse (automaatse kvaliteediskoori toel) sellega, kuidas reaalne töö välja näeb. Nii saab võrrelda, kas arvamused ja tööviisid omavahel kattuvad või mitte.

Katse raames koostatud küsimustiku vastused andsid palju erinevaid tulemusi. Küsimustiku vastustest selgub, et masintõlke kasutamine on tõlkijate seas väga levinud, sest selgus, et kõik vastajad kasutavad seda mingil määral oma töös. Suhtumised masintõlke järeltoimetamisse jaotuvad umbes pooleks ning väga pädevaks ei pea end keegi, kuna keskmine hinnang oma isikliku pädevuse kohta skaalal 1–0 jäi 4,9 peale, ehk veidi alla poole. Mõned katsealused leidsid, et järeltoimetamisest võib abi olla, eelkõige kui on vaja teavet kiirelt edastada või kui on vaja tõlkida mingeid lihtsamaid tekste. Leitakse, et masintõlge võib mõne teksti puhul teha esmase töö tõlkija eest ära, seega muudab tõlkeprotsessi veidi lihtsamaks. Masintõlke järeltoimetamise eelistena toodi veel välja näiteks seda, et järeltoimetamine hoiab aega kokku ning seda, et masintõlge võib anda ideid tõlkeprobleemide lahendamiseks. Samas suhtutakse masintõlkesse ka skeptiliselt, enamasti seetõttu, et seda ei saa väga usaldada. Osad katsealused leidsid, et masintõlke järeltoimetamine võib olla isegi keerulisem kui nullist tõlkimine ning seetõttu oleks tihti lihtsam tekstid ise tõlkida, kasutades masintõlkesüsteeme üksnes abivahendina. Kokkupuude raskesti tuvastatavate vigadega on eranditult kõigil katsealustel, mis põhjendab paljude katsealuste skeptilist suhtumist. Masintõlke järeltoimetamisesse suhtutakse ka ettevaatlikkusega ja seda peetakse kognitiivselt väga kurnavaks, sest see nõuab suurt tähelepanelikkust. Seda seisukohta toetab Saaki (2019) kirjutatud magistriprojekt, mille tulemustest selgus, et järeltoimetamisel on tõlkijale sarnane kongnitiivne koormus, kui nullist tõlkimisel.

Suurimateks probleemideks järeltoimetamisel peetakse muuhulgas masintõlke kehva kvaliteeti ning vigade märkamise keerukust. Korrektse masintõlkega kokku puutumine on samuti pigem harv nähtus. Masintõlke järeltoimetamise puuduste sekka loetleti veel

järgmised: võib tekkida võlts turvatunne (ja vead jäävad märkamata), terminoloogilised probleemid ja järeltoimetamise liiga sage kasutamine (sest kõigil juhtudel see ei sobi).

Küsimustiku tulemuste põhjal võib järeldada, et tõlkijad oskavad masintõlke järeltoimetamist hinnata küll kui mõnikord kui kasulikku töömeetodit, kuid samas ei ole masintõlge veel jõudnud tasemeni, kus seda täielikult usaldada saab, seega tekib olukordi, kus järeltoimetamine töö hoopis keerulisemaks ja kurnavamaks muudab. Nagu Moorkensi ja O'Brieni (2017) küsimustikuuringus selgus, siis suur osa vastanutest sooviks, et masintõlke väljunditel oleks mingisugune usaldusskoor, mis aitaks paremini otsustada, kas masintõlge on järeltoimetamist väärt või mitte. Samuti selgus Shuyleri (2021) magistritöös, et järeltoimetamise protsessis on tõlkijate jaoks keeruline just otsustamine. Selleks lahenduseks võib potentsiaalselt olla MTQE funktsiooni integreerimine järeltoimetamise protsessi.

Peale küsimustiku täitmist sooritasid katsealused järeltoimetamise katse, millest ühele tekstile olid kuvatud kvaliteediskoorid, et näha, kas nende skooride olemasolu kiirendab katsealuste tööd ja kas skoores nähes tehakse tekstis rohkem või vähem muudatusi.

Esimese töös püstitatud küsimusega sooviti teada, kas kvaliteediskooride nägemine tõstab töökiirust võrreldes tekstiga, kus kvaliteediskoore kuvatud ei ole. Katse tulemustest selgus, et tekst, kus kvaliteediskoorid olid kuvatud, järeltoimetati 21,5% kiirema tempoga kui varjatud kvaliteediskooridega tekst. Sellest võib järeldada, et skooride kuvamisest oli katsealuste töö tempole kasu. Samuti lisasid katsealused töö lõpus, kus küsiti äsjasooritatud katse kohta mõned küsimused, et automaatne kvaliteediskoor on pigem hea abivahend ja toetab järeltoimetamist, sest aitab hinnata seda, kui tähelepanelik peaks järeltoimetamisel olema, nii kõrge skooriga segmentide kui madala skooridega segmentide puhul. Sama tulemuseni jõudsid ka Béchara *et al.* (2021), kelle töös selgus, hea kvaliteediskoor võib töö tempot tõsta ning kvaliteediskoorist võib eelkõige töövoos kasu olla madalate skooride puhul. Negatiivsetest külgedest toodi välja seda, et kõrge skooriga segmentidest on oht üle libiseda, kuigi tekst ei pruugi olla korrektne.

Teise töös püstitatud küsimusega sooviti teada, kas muudatusi tehakse rohkem kvaliteediskoore nähes või mitte nähes. Mõlema teksti puhul tehti muudatusi enam-vähem

pooltes segmentides (kuvatud skooridega tekstis muudeti 55% segmentidest ja varjatud skooridega tekstis 49% segmentidest), nii et olulist vahet sealt ei kujunenud. Sealhulgas muudeti mõlema teksti puhul protsentuaalselt kõige rohkem 75% skooriga segmente – kuvatud kvaliteediskooriga tekstis muudeti keskmiselt 69% segmentidest ja kuvamata skooriga tekstis keskmiselt 59% segmentidest. Kõige vähem muudeti 100% skooriga segmente, mida võiks sellise töö puhul ka eeldada, sest nagu Memsources'i enda kodulehel kirjas on, peaks 100%-sed vastad olema ideaalsed. Sellest kõigest võib järeldada, et katsealused teevad tööd võrdselt, olgu skoorid ees või mitte, aga kuna ajaline erinevus oli tekstide järeltoimetamisel päris suur, siis võib järeldada, et kvaliteediskooride nägemine aitab katsealustel kiiremini otsustada, millised segmendid vajavad rohkem muutmist ja millised mitte. Siinkohal saab kinnitust ka töös püstitatud hüpotees, et automaatse kvaliteedihinnangu kuvamine hõlbustab tõlkijate tööd järeltoimetamisel, kuna aitab neil otsustada, millised laused on järeltoimetamist väärt ja millised mitte.

Kolmas küsimus oli seotud sellega, kuidas suhtuvad katsealused MTQE integreerimisse järeltoimetamise protsessi (sh nende mõtted peale katse sooritamist). Mitmed katsealused kirjutasid, et kvaliteediskoor on hea abivahend ning toetab järeltoimetamist, sh aitab kergemini otsuseid langetada ja hinnata, kui valvas peaks erineva skooriga segmentide puhul olema. Samas lisati, et järeltoimetaja peab kindlasti üle kontrollima ka 100% vastad, sest vigadest on lihtne üle libiseda.

Kokkuvõtlikult võib järeldada, et kvaliteediskoori funktsioon on kasulik lisand, kuid kindlasti vajab veel palju arendamist ja uurimist. Kuna Sepp (2017) leidis oma uurimistöös, et masintõlke integreerimine töösse kiirendas tõlkijate tööd 25% võrra ja käesoleva töö tulemustes leiti, et masintõlkele kvaliteediskooride lisamine kiirendas järeltoimetamist veel omakorra 21,5% võrra, siis oleks põnev uurida, kui suured oleksid erinevused, kui kõrvutada inimtõlge, masintõlke järeltoimetamine (ilma skoorideta) ning kvaliteediskooridega masintõlke järeltoimetamine (siinkohal tuleb küll tõdeda, et kuna selliseid katseid, nagu selle töö raames tehti, saab väga erinevalt üles ehitada, siis võivad mõned teistsugused katsed anda täiesti erinevaid tulemusi, näiteks kui esitada katses skooridega ja skoorideta segmendid vaheldumisi). Samuti oleks põnev võrrelda erinevaid süsteeme, mis pakuvad

kvaliteediskoore, sest praegused süsteemid ei kasuta ühtset meetrikat. Seega oleks huvitav teada, kas ühed kvaliteediskoorid annavad paremaid tulemusi kui teised.

## Kokkuvõte

Selle töö eesmärk oli anda ülevaade masintõlkele automaatse kvaliteediskoori (MTQE) funktsioonist. MTQE funktsioon annab masintõlgitud tekstile automaatse kvaliteediskoori, mis võib aidata tõlkijatel otsustada, kas laused või segmendid on korrektsed, vajavad järeltoimetamist või tuleks tõlkijal algusest lõpuni ise tõlkida (madala kvaliteediskoori puhul). Otsustasin selle töö sellisel teemal kirjutada, sest MTQE süsteemid on alles hiljuti kasutusele võetud ja nende kohta ei ole veel palju uurimusi kirjutatud, seega on põnev uurida, kui kasulikud need tegelikult (reaalses töös) on.

Töö teoreetiline taust on jaotatud kolme peatükki. Esimene peatükk annab lühikese ülevaate masintõlkest üldiselt. Teine peatükk tutvustab masintõlke järeltoimetamist, sealhulgas tuuakse välja selle eelised ja puudused ja mõned näpunäited selle kohta, milliseid oskusi on vaja, et olla edukas järeltoimetaja. Viimane masintõlke järeltoimetamise alapeatükk tutvustab masintõlke järeltoimetamisest kirjutatud eelnevaid töid, mis on kõik Eestis kirjutatud magistritööd/projektid. Kolmandas peatükis antakse ülevaade masintõlke hindamisest, sealhulgas varem kasutatud meetoditest ja süsteemidest, mida on kasutatud. Seejärel tuleb juttu masintõlke automaatsest kvaliteedi hindamisest ning veidi põhjalikumalt tutvustatakse Memsources'i pakutatavat MTQE funktsiooni.

Töö empiiriline osa koosneb kahest osast – küsimustiku kirjeldusest ja selle tulemustest ja katse kirjeldusest ja selle tulemustest. Küsimustiku esimene osa koosnes küsimustest masintõlke, masintõlke järeltoimetamise ja eelneva MTQE funktsioonide kasutamise kogemuse kohta. Seejärel paluti katsealustel osa võtta järeltoimetamise katsest, kus nad pidid kahte (võimalikult sarnast) teksti järeltoimetama – ühele tekstile olid segmentide kõrvale kuvatud kvaliteediskoorid ning teisele tekstile mitte. Lähtetekst oli ingliskeelne ning sihttekst eestikeelne. Peale katse sooritamist asusid katsealused viimase küsimustiku osa juurde, mis hõlmas küsimusi äsjasooritatud katsest kohta ning nende arvamusest MTQE funktsiooni ja kvaliteediskooride kohta – täpsemalt uuriti, kas nende jaoks on see kasulik abivahend või mitte.

Küsimustiku tulemused näitavad, et katsealused küll kasutavad masintõlget, kuid vaid mõõdukalt, sest nad leiavad, et seda ei saa tihti usaldada. Järeltoimetamist nähakse kui pigem kognitiivselt kurnavat ülesannet, seega osad katsealused eelistavad tekste järeltoimetamise asemel ise otsast lõpuni tõlkida, kasutades masintõlkesüsteeme vajadusel abivahenditena. Järeltoimetamise katse tulemused näitasid aga seda, et kui kuvada tekstile lisaks kvaliteediskoorid, siis muutub töö tempo 21,5% võrra kiiremaks, võrreldes sellega, kui skoori ei kuvata. Samuti toetavad seda tulemust vastused küsimustiku lõpus asuvatele sooritatud katset käsitlevatele küsimustele – suur osa katsealuseid pidas kvaliteediskoori nägemist kasulikuks. Kahe erineva teksti muudatuste hulgad olid põhimõtteliselt samad, veidi enam tehti muudatusi tekstis, kus olid kvaliteediskoorid kuvatud.

Selle magistritöö lõpetuseks võib öelda, et MTQE funktsioonidest võib järeltoimetamisel kasu olla, nagu näitavad selle töö katse tulemused. Samuti on katsealused selle funktsiooni suhtes pigem positiivselt meelestatud. Kindlasti tuleks selliseid funktsioone veel uurida. Loodetavasti on tulevikus MTQE funktsioonide pakkujaid veel rohkem, et järgmistes uurimustes erinevate pakkujate funktsioone võrrelda.

## Kasutatud kirjandus

- Banerjee, S. ja Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. *In Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization* (lk 65–72). Ann Arbor, Michigan: Association for Computational Linguistics.
- Béchara, H. *et al.* (2021). The Role of Machine Translation Quality Estimation in the Post-Editing Workflow. *Informatics*, 8(3), 61. <https://doi.org/10.3390/informatics8030061>
- Chatzikoumi, E. (2019). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering* 2020 (26), 137–161. <https://doi.org/10.1017/S1351324919000469>
- Chéragui, M. A. (2012). Theoretical Overview of Machine translation. *Proceedings of the 4th International conference on Web and Information Technologies (ICWIT) 2012* (lk 160–169). Sidi Bel Abbès, Alžeeria: CEUR Workshop Proceedings.
- Čulo, O., Gutermuth, S., Hansen-Schirra, S. & Nitzke, J. (2014). The influence of post-editing on translation strategies. Balling, L. W. & Carl, M. (toim), *Post-editing of Machine translation: Processes and applications* (lk 200–218). Newcastle upon Tyne: Cambridge Scholars Publishing.
- DeepAI koduleht. *Neural Machine Translation*. Loetud aadressil <https://deepai.org/machine-learning-glossary-and-terms/neural-machine-translation>.
- Dorr, B., Snover, M. & Madnani, N. (2011). Chapter 5.1 introduction. Olive, J., McCary, J. & Christianson, C. (toim), *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation* (lk 801–803). New York: Springer.

- Fernández-Torné, A. & Matamala, A. (2021). Human evaluation of three machine translation systems: from quality to attitudes by professional translators. *Vigo International Journal of Applied Linguistics*, 18 (2021), 97–121. <https://doi.org/10.35869/vial.v0i18.3366>
- Green, S., Heer, J. & Manning, C. D. (2013). The Efficacy of Human Post-Editing for Language Translation. *CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (lk 439–448). New York: Association for Computing Machinery. <https://doi.org/10.1145/2470654.2470718>
- Guerberof Arenas, A. (2012). *Productivity and quality in the post-editing of outputs from translation memories and machine translation* (doktoritöö). Loetud aadressil <http://hdl.handle.net/10803/90247>
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press.
- Koponen, M., Salmi, L. & Nikulin, M. (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, 33, 61–90. <https://doi.org/10.1007/s10590-019-09228-7>
- Lavie, A. (2011). Evaluating the output of machine translation systems. *Proceedings of Machine Translation Summit XIII: Tutorial Abstracts*. Xiamen, China.
- Läubli, S., Sennrich, R. & Volk, M. (2018). Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (lk 4791–4796). Brüssel, Belgia: Association for Computational Linguistics.
- Memsources'i koduleht. *Machine Translation Quality Estimation*. Loetud aadressil <https://www.memsource.com/features/translation-quality-estimation/>.
- Mieke, L. (2020). *Light Post-Editing Or Full Post-Editing: Which Way To Go?* Loetud aadressil <https://laoret.com/blog/light-post-editing-or-full-post-editing/>.

- Moorkens, J. & O'Brien, S. (2017). Assessing user interface needs of posteditors of machine translation. Dorothy Kenny (toim), *Assessing user interface of post-editors of machine translation* (lk 109–130). New York, NY: Routledge.
- Moorkens, J., O'Brien, S., da Silva, I.A.L., de Lima Fonseca, N.B. & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation*, 29, 267–284.
- Nitzke, J. & Hansen-Schirra, S. (2021). A short guide to post-editing. *Translation and Multilingual Natural Language Processing 16*. Berliin: Language Science Press.
- O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation* 25(3), 197–215.
- Saak, K. (2019). *Tõlkemälu ja masintõlke vastete mõju kognitiivsele koormusele tõlkimisprotsessis* (magistritöö). Tartu: Tartu Ülikool, Maailma keelte ja kultuuride kolledž.
- Sayers, D., Sousa-Silva R., Höhn, S. *et al.* (2021). Language In The Human-Machine Era. *The Dawn of the Human-Machine Era: A forecast of new and emerging language technologies* (lk 52–57). <https://doi.org/10.17011/jyx/reports/20210518/1>
- Sepp, K. (2017). *Masintõlke järeldoimetamine* (magistritöö). Tallinn: Tallinna Ülikool, Humanitaarteaduste instituut.
- Sibul, I. (2018). Masintõlge ja eestikeelsed õigustekstid – ühe tõlkija senised kogemused Euroopa Komisjonis. *Õiguskeel*, 2018/2, 1–7.
- Silva, J. (2012). Statistical Machine Translation - the problem of unknown words. *Communications in Computer and Information Science*, 333, 176–187.
- Shuyler, K. (2021). *A Case Study on Post-editing Machine Translation: Tasks, Challenges, and Attitudes* (magistritöö). Tartu: Tartu Ülikool.

- Teixeira, C., O'Brien, S. (2017). The Impact of MT Quality Estimation on Post-Editing Effort. *Proceedings of the MT Summit XVI, Volume 2: Users and Translators Track* (lk 211–233). Nagoya, Jaapan.
- Toompalu, K. (2011). *Tõlke kvaliteedi hindamine* (magistritöö). Tartu: Tartu Ülikool, Arvutiteaduse instituut.
- Yamada, M. (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation*, 31, 87–106.
- Zhou M., Wang B., Liu S., Li M., Zhang D. & Zhao T. (2008). Diagnostic evaluation of machine translation systems using automatically constructed linguistic check-points. *In Proceedings of the 22nd International Conference on Computational Linguistics* (lk 1121–1128). Manchester, Ühendkuningriigid.

## Summary

The aim of this thesis was to give an overview of the MTQE (*Machine Translation Quality Estimation*) systems, which give automatic scores to machine translated texts, in order to help translators decide, whether the sentences/segments are correct, worth post-editing, or in case of a low score, should probably be deleted and translated by a human translator. I decided to write my thesis on this topic since MTQE systems are fairly new in the machine translation field, so research is needed in order to see how useful these are.

The theoretical background is divided into 3 parts. The first chapter gives a short overview of machine translation. The second chapter introduces machine translation post-editing, its strengths and weaknesses, what skills are needed from a translator in order to be a successful post-editor, and the final subchapter highlights some of the previous work on the subject. The third chapter focuses on machine translation evaluation, more specifically on previous systems that have been used to evaluate machine translation, and then introduces the MTQE feature (mostly focusing on Memsources' MTQE feature).

In the empirical research part of the thesis, the results of a questionnaire and the results of a post-editing experiment conducted for this thesis are presented. The first part of the questionnaire consisted of questions about machine translation, post-editing machine translation and background knowledge of MTQE features. Then, the subjects were asked to take part in a post-editing experiment, where they had to post-edit two similar texts – one with MTQE scores and one without it. The source text was in English and target text in Estonian. Lastly, in the second part of the questionnaire, the subjects were asked some questions to describe their thoughts on the translation experiment and whether they found the MTQE scores valuable or not.

Results of the questionnaire show that the subjects use machine translation, but only moderately, since they don't find it to be very trustworthy. Post-editing is seen as a cognitively demanding task, therefore many of the subjects prefer translating texts themselves. The results of the translation experiment however show, that when presented a

post-editing task including MTQE scores, the subjects take 21.5% less time to complete the task compared to when they don't see the scores. The results are supported by the answers to the questions asked from the subjects at the end of the questionnaire about the MTQE feature, as the majority of subjects found this feature to be a rather useful tool in helping them in the post-editing process. There were no significant changes in comparing the number of changes made by subjects in the texts.

To conclude, based on the results of this thesis, it seems that MTQE features could become useful in the post-editing process to translators, as the results of this thesis show an increase in post-editing speed. Also, most of the subjects were open-minded and positive towards this feature. But it should be kept in mind, that even segments, that receive a score of a 100% should still be double checked. There is certainly a need for more research to be done on this subject and, as time goes by, there will certainly be more providers of such features, so in the future, it would be interesting to compare these features.

Lisa

## Järeltoimetamise katse

Selles failis esitatud tabelites leiad **kaks teksti, mille masintõlke väljundid vajavad järeltoimetamist**. Esimesele neist on juurde lisatud masintõlke **automaatne kvaliteediskoor** (parempoolses tulbas), see tähendab, et segmendid on tõlgitud masintõlkega ning tõlkeabiprogramm on andnud tõlkele automaatse skoori, mida võid järeltoimetamisel abivahendina kasutada. **Mida suurem on skoor, seda paremaks on süsteem tõlke hinnanud** (kuid see, kui palju ja mida järeltoimetada või ümber tõlkida, jääb siiski toimetaja enda otsustada).

Kvaliteediskoorid, mida kuvatakse, on **100, 99, 75 ja MT (ehk skoor puudub)**.

100 - ideaalne masintõlge, tõenäoliselt toimetamist ei vaja

99 - ideaalilähedane masintõlge, võimalik, et vajab vähest toimetamist; enamasti esineb probleeme vormingu või kirjavahemärkidega

75 - hea masintõlge, sobib järeltoimetamiseks

MT – tõenäoliselt madala kvaliteediga masintõlge, ei ole soovitatav järeltoimetada, vaid kasutada näiteks abimaterjalina.

Teisel tekstil kvaliteediskoore juures ei ole, ka see tekst tuleb samamoodi järeltoimetada.

Katse on üles ehitatud samas formaadis, nagu kuvaks seda tõlkeabiprogrammid, seega kõige **vasakpoolses tulbas on lähtetekst ja sellest paremal on toimetamist vajav sihttekst**.

PS! Pea meeles, et tegu on järeltoimetamisega, seega **püüa hoiduda n-ö ületoimetamisest** (näiteks mõne sõna muutmine eelistuse tõttu, kuigi mõlemad on korrektsed) ja tee **ainult vajalikke muudatusi**.

Võimalusel kasuta toimetamisel **muutuste jälitaja funktsiooni** (inglisekeelse MS Wordi puhul Review – Track Changes / eestikeelse Wordi puhul Läbivaatus – Jälita muutusi).

Palun jälgi iseseisvalt, kui palju aega kummagi teksti toimetamine võttis (vaata kella pealt algus- ja lõpuaega või kasuta näiteks telefoni taimerit). Küsimustikus on koht, kuhu saad hiljem eraldi kirja panna kummagi teksti toimetamiseks kulutatud aja.

Kui tekstid on järeltoimetatud, siis salvesta fail ja **lae üles küsimustikus viidatud kohta**.

<b>Food Chopper &amp; Slicer</b>	<b>Toidu chopper &amp; slicer</b>	MT
User Manual	Kasutusjuhend	100
<b>Read this manual prior to initial operation of the product!</b>	<b>Lugege seda juhendit enne toote esmakordset käitamist!</b>	75
Dear Customer,	Lugupeetud klient,	MT
We hope that your product, which has been manufactured in modern facilities and passed through a strict quality control procedure, will give you very good results.	Loodame, et teie toode, mis on valmistatud kaasaegsetes rajatistes ja läbinud range kvaliteedikontrolli protseduuri, annab teile väga häid tulemusi.	75
We advise you to read through this manual carefully before using your product and keep it for future reference.	Soovitame teil enne toote kasutamist selle juhendi hoolikalt läbi lugeda ja hoida seda edaspidiseks kasutamiseks.	75
The User Manual will help you use your machine in a fast and safe way.	Kasutusjuhend aitab teil masinat kiiresti ja turvaliselt kasutada.	75
This product has been produced in environmentally friendly, modern facilities	Seda toodet on toodetud keskkonnasõbralikes ja kaasaegsetes rajatistes.	75
Read the User Manual before installing and starting your product.	Lugege enne toote installimist ja käivitamist kasutusjuhendit.	75
In particular, follow the instructions related to safety.	Eelkõige järgige ohutusega seotud juhiseid.	75
Keep this User Manual within easy reach for future reference.	Hoidke seda kasutusjuhendit edaspidiseks viitamiseks käeulatuses.	75
Also read all additional documents supplied with this product.	Lugege ka kõiki selle tootega kaasas olevaid täiendavaid dokumente.	75
Remember that this User Manual is also applicable for several other models.	Pidage meeles, et see kasutusjuhend on rakendatav ka mitmete teiste mudelite puhul.	75
Differences between models will be identified in the manual.	Erinevused mudelite vahel tuvastatakse käsiraamatus.	MT
CONTENT	SISU	MT
<b>Safety Warnings</b>	<b>Ohutushoiatused</b>	100
<b>Technical Aspects Of Your Machine</b>	<b>Teie masina tehnilised aspektid</b>	75

<b>Using Your Machine</b>	<b>Masina kasutamine</b>	100
<b>Maintenance and Cleaning</b>	<b>Hooldus ja puhastus</b>	100
Safety Warnings	Ohutushoiatused	100
Before using the chopper, read carefully the following operating instructions. It is suggested that you keep these instructions in a safe place for future reference.	Enne kopteri kasutamist lugege hoolikalt läbi järgmised kasutusjuhendid. Soovitav on hoida neid juhiseid turvalises kohas tulevaseks kasutamiseks.	75
Attention	Tähelepanu	100
Check whether the voltage indicated on the chopper is the same as the voltage in your home.	Kontrollige, kas kopteril näidatud pinge on sama, mis teie kodus olev pinge.	75
Do not place the appliance near any heat source which could cause damage.	Ärge asetage seadet ühegi soojusallika lähedusse, mis võib kahjustada.	75
Use the appliance on a stable work surface away from water.	Kasutage seadet stabiilsel tööpinnal veest eemal.	75
Do not take off the safety cover before the blade has completely stopped turning.	Ärge eemaldage turvakatet enne, kui tera on täielikult peatunud.	75
Take care with the knife as the blades are very sharp.	Hoolitse nuga eest, sest terad on väga teravad.	75

**Enne teise teksti juurde liikumist pane palun kirja aeg, mis kulus esimese teksti toimetamiseks!**

All repairs of electrical appliances must only be performed by skilled personnel. Improper repairs can result in considerable hazards for the user.	Kõik elektriseadmete remonditööd tohib teostada ainult kvalifitseeritud personalil. Ebakorrektsed remontid võivad põhjustada kasutajale märkimisväärsed ohted.
Attention	Tähelepanu
Before using your product, read carefully the following operating instructions. It is suggested that you keep these instructions in safe place.	Enne toote kasutamist lugege hoolikalt läbi järgmised kasutusjuhised. Soovitatakse neid juhiseid hoida turvalises kohas.
Check whether the voltage indicated on the base of the product is the same as the voltage in your home.	Kontrollige, kas toote alusele märgitud pinge on sama, mis teie kodus.
This appliance is not intended for use by persons (including children) with reduced physical, sensory or mental capabilities, or lack of experience and knowledge, unless they have been given supervision or instruction concerning use of the appliance by a person responsible for their safety.	See seade ei ole mõeldud kasutamiseks isikutele (sealhulgas lastele), kellel on vähenenud füüsilised, sensoorsed või vaimsed võimed või kellel puuduvad kogemused ja teadmised, välja arvatud juhul, kui nende ohutuse eest vastutav isik on neid juhendanud või juhendanud seadme kasutamist.
Values stated on the markings of the product or in the printed documents accompanying the product are obtained in the laboratory on basis of the relevant standards.	Toote märgistusel või tootega kaasas olevates trükitud dokumentides esitatud väärtused on saadud laboris asjakohaste standardite alusel.
These values may change according to the usage and ambient conditions of the product.	Need väärtused võivad muutuda vastavalt toote kasutamisele ja keskkonnatingimustele.
Power values are tested 240 V.	Võimsuse väärtused on testitud 240 V.
Children should be supervised to ensure that they do not play with the appliance.	Laste üle tuleb teostada järelevalvet, et nad ei mängiks seadmega.
Using Your Machine	Masina kasutamine
Put the main blade into the mixing bowl, onto the fixed shaft and upper blade onto the main blade.	Pange põhitera segamiskaussi, fikseeritud võllile ja ülemine tera põhitera peale.
Add ingredients.	Lisage koostisosad.
Fit the plate in position on the mixing bowl.	Paigaldage taldrik segamisnõusse.
Put the power unit on the plate fitted mixing bowl. Press down to start.	Asetage toiteplokk taldrikule paigaldatud segamisnõusse. Vajutage käivitamiseks alla.
In order to obtain a better quality of chopping, operate the appliance by using the "pulse" function. If the ingredients	Parema hakkimiskvaliteedi saavutamiseks kasutage seadet "pulse"-funktsiooni. Kui koostisained jäävad kausi küljele

stick on the side of the bowl, dislodge the pieces with a spatula and operate 2 or 3 times using the "pulse" function.	kinni, eemaldage tükid spaatliga ja kasutage 2 või 3 korda "pulse"-funktsiooni.
Attention	Tähelepanu
Do not operate the appliance when empty.	Ärge kasutage seadet tühjalt.
Maintenance and Cleaning	Hooldus ja puhastamine
Do not immerse the power unit into the water.	Ärge kastke toiteplokki vette.
Clean it with damp sponge.	Puhastage see niiske käsnaga.
Mixing bowl, safety cover and blades can be cleaned into the upper basket of the dishwasher.	Segamiskaussi, kaitsekatet ja terasid saab puhastada nõudepesumasina ülemises korvis.

Allikas: <https://www.manualslib.com/manual/894319/Beko-Bkk-2158.html>

## **Lõputöö autori kinnitus**

Olen lõputöö kirjutanud iseseisvalt. Kõigile töös kasutatud teiste autorite töödele, põhimõtteliste seisukohtadele ning muudest allikaist pärinevatele andmetele on viidatud.

Helina Tera, 23.05.2022

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Helina Tera,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Neuromasintõlke järeltoimetamine automaatse kvaliteediskoori toel” , mille juhendaja on Reelika Saar, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Helina Tera*

**23.05.2022**