TARTON YLIOPISTO
HUMANISTISTEN TIETEIDEN JA TAIDEAINEIDEN TIEDEKUNTA
VIRON JA YLEISEN KIELITIETEEN LAITOS
Suomalais-ugrilaisten kielten osasto

Krista Kotkamäe

Viro-suomi ihmis- ja konekäännöksen vertailu
Viron kansallismuseon Kohtaamisia-näyttelyn tekstien perusteella

Maisterintutkielma

Ohjaaja Juha-Matti Aronen

Tartto 2023

Sisällys

Johdanto 5
1. Yleiskatsaus konekäännösteknologiaan 7
1.1. Keskeiset käsitteet 7
1.2. Konekääntimien kehitys ja eri lähestymistavat 9
1.2.1. Suorakäännösjärjestelmät 10
1.2.2. Sääntöpohjainen konekääntäminen 10
1.2.3. Tilastollinen konekääntäminen 11
1.2.4. Esimerkkipohjainen konekääntäminen 12
1.2.5. Neuroverkkokonekääntäminen 12
1.3. Konekääntämisen ongelmakohdat 13
2. Konekäännöksen arviointi 16
2.1. Ihmiskäännös *vs* konekäännös 16
2.2. Konekäännöksen tutkimisesta 18
2.3. Kvantitatiiviset arviointimenetelmät 19
2.4. Kvalitatiiviset arviointimenetelmät 21
2.4.1. TAUSin DQF & QT21-virheluokittelu 23
2.5. Arviointimenetelmien vahvuudet ja heikkoudet 26
3. Konekäännöksen virheanalyysi 28
3.1. Analyysiaineiston kuvaus 28
3.2. Tutkimusmenetelmä 29
3.3. Virhetyyppien yhteenveto 32
3.3.1. Virheiden jakauma teksteittäin 34
3.4. Tarkkuus 35
3.4.1. Väärä käännös 35
3.4.2. Kääntämätön 38
3.4.3. Poisto 39
3.4.4. Alikääntäminen 41
3.4.5. Liikakääntäminen 42
3.5. Sujuvuus 43
3.5.1. Kielioppi 44

3.5.2. Oikeinkirjoitus 48
3.5.3. Välimerkkien käyttö 50
3.6. Tyyli 51
3.6.1. Kömpelö tyyli 51
3.6.2. Epäidiomaattinen tyyli 53
3.7. Muu 54
3.8. Konekäännöksen ja ihmiskäännöksen samankaltaisuus 55
Tiivistelmä 60
Lyhenteet 64
Lähteet 65
Eesti-soome inim- ja masintõlke võrdlus ERMi näituse ”Kohtumised” tekstide põhjal. Resümee 70
Liite 1: analyysimateriaalin virheluokittelu 72

Autorsuse kinnitus

Kinnitan, et olen käesoleva lõputöö ise kirjutanud ning toonud korrektselt välja teiste autorite panuse. Töö on kirjutatud lähtudes Tartu Ülikooli eesti ja üldkeeleteaduse instituudi lõputöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Johdanto

Tämä maisterintutkielma käsittelee konekääntämisen arviointia vertailemalla konekäännöstuotosta ihmisen tekemään käännökseen. Tutkimuksen tavoitteena on selvittää, minkä tyyppisiä käännösvirheitä tekevät yleisessä käytössä olevat neuroverkkokonekääntimet juuri viro–suomi-kieliparissa. Konekääntimillä uskotaan olevan morfologisen runsauden takia hankaluuksia suomalais-ugrilaisten kielten kääntämisen kanssa.

Tommi Nieminen antoi vuonna 2020 Euroopan Komission #TranslatingEurope työpajassa pidetyssä esityksessään Machine translation and the human translator yleiskatsauksen konekääntämisen tämänhetkiseen tilanteeseen Suomessa. Hänen mukaansa Suomessa tällä hetkellä konekääntämisessä taloudellisesti merkittävät kieliparit ovat suomi–englanti, suomi–ruotsi ja englanti–suomi, joihin löytyy riittävästi koulutusaineistoa. (Nieminen 2020: 4, 6.) Hän nosti erikseen esiin englanti–suomi-kieliparin, jossa neuroverkkokonekäännöksen kehitys on ollut nopeaa vuodesta 2017 alkaen (Nieminen 2020: 13). ”Suomea pidetiin pitkään vaikeasti konekäännettävänä kielenä monimutkaisen taivutuksen ja joustavan sanajärjestyksen vuoksi” (Nieminen 2020: 17).

Myös Tarton yliopiston kieliteknologian tutkimusryhmän kehittämän konekääntimen Tartu NLP:n vetäjä Mark Fišel kirjoitti vuonna 2016, että vironkielisen syötön ja/tai tuotoksen kanssa herää ongelmia sijapäätteiden ym. morfologisen tiedon runsauden sekä joustavan ja muuttuvan sanajärjestyksen takia. Tartu NLP neuroverkkokonekäännin olikin silloin vasta kehityksensä alkuvaiheessa ja yhteys neuroverkkokonekääntimen edeltäjään tilastolliseen konekääntimeen vielä vahva. (Fišel 2016.)

Tämän tutkimuksen avulla pyritään selvittämään sitä, sillä suomen ja viron kielen konekääntämistä on juuri kvalitatiivisilla menetelmillä tutkittu aika vähän. Tutkielma koostuu kolmesta sisältöluvusta, joista ensimmäisessä annan yleiskatsauksen konekäännösteknologiaan. Aluksi definioin konekääntämiseen liittyvät keskeiset

käsitteet. Sen jälkeen selvennän konekääntimien kehityksen taustaa ja eri lähestymistapoja sekä keskityn tarkemmin neuroverkkokonekääntämiseen. Ensimmäisen luvun lopussa esitän konekääntämisen ongelmakohtia. Tutkielman teoreettinen tausta perustuu pääasiassa kolmeen lähteeseen: The Routledge Encyclopedia of Translation Technology (toim. Chan Sin-wai), Machine Translation (kirj. Thierry Poibeau) ja Translation and technology (kirj. Chiew Kin Quah).

Toisessa sisältöluvussa käsittelen konekäännöksen arviointia. Pohdin ihmis- ja konekäännöksen lähtökohtia ja vertailen niitä keskenään. Esittelen muutamalla esimerkillä konekääntämisen eri tutkimussuuntia ja -menetelmiä. Sitten teen yleiskatsauksen konekääntämisen arviointimenetelmiin ja kerron hieman niiden vahvuuksista ja heikkouksista. Tämä sisältöluku perustuu aikaisempiin konekäännöstutkimuksiin ja menetelmäkuvauksiin.

Kolmas sisältöluku keskittyy tutkielmani virheanalyysiin. Aluksi annan taustatietoa analyysiaineistosta ja käytetystä tutkimusmenetelmästä. Aineisto on peräisin Viron kansallismuseon Kohtaamisia-näyttelyn teksteistä. Sitä seuraa konekäännösvirheiden luokittelu virhetyyppeihin ja siinä tehdyt havainnot. Tutkielmassa on sovellettu pääasiassa kvalitatiivista tutkimusmenetelmää, konekäännöksen subjektiivista virheanalyysiä. Sen rinnalla tutkin konekäännöstuotosta kvantitatiivisella menetelmällä. Lasken vertailevasti kolmen eri neuroverkkokonekääntimen (Tartu NLP:n, Tilde NMT:n ja Google NMT:n) Levenshteinin editointietäisyyden eli samankaltaisuuden ihmiskäännöksen kanssa. Analyysiluvun lopussa esitän sillä saatuja tuloksia.

Maisterintutkielmani tutkimuskysymykset ovat:

1. Mitkä ovat eri konekääntimien ongelmakohdat?
2. Millä tavoin voi arvioida konekäännöstuotosta?
3. Minkä tyyppisiä virheitä konekäännin tekee verrattuna ihmisen tekemään käännökseen?
4. Minkä verran konekäännös muistuttaa ihmiskäännöstä?

1. Yleiskatsaus konekäännösteknologiaan

Konekäännös on nopeasti kehittyvä kieliteknologian ala, johon on suuri käytännöllinen tarve. Sen historia ulottuu 1950-luvulle asti, jolloin kehitettiin niin sanottu ensimmäisen sukupolven konekäännösjärjestelmä, joka pystyi tuottamaan sanasanaisia käännöksiä. Kyseessä oli pieni järjestelmä, mutta se innosti investoimaan siihen alaan ja aiheutti myönteisen suhtautumisen täysautomaattisiin järjestelmiin, joiden avulla olisi mahdollista tuottaa korkealaatuisia käännöksiä. (Quah 2006: 60.) Suunnilleen samaan aikaa alettiin myös Virossa testaamaan konekääntämistä ja yritettiin kääntää matemaattisia tekstejä venäjästä viroon (Kaalep, Koit 2010: 727). Nykyajan konekäännöksestä puhuttaessa tarkoitetaan varsinkin neuroverkkokonekäännintä, joka valloitti maailman tämän vuosisadan toisella kymmenellä tekoälyn kehittymisen myötä (Forcada 2017: 291). Annan seuraavaksi yleiskatsauksen konekääntämisen kehitykseen ja sen eri lähestymistapoihin ja keskityn tutkimuskohteenani olevaan neuroverkkokonekäännökseen. Viimeisessä kappaleessa käsittelen konekääntämisen ongelmakohtia.

1.1. Keskeiset käsitteet

Käytän tutkielmassani konekääntämiseen liittyviä termejä. Olen kerännyt tähän konekääntämisen keskeisiä käsitteitä, joiden määritelmät ovat peräisin termikannoista IATE, TEPA-termipankki ja Tieteen termipankki.

Dekooderi (engl. *decoder*) – ”ohjelma tai laite, joka muuntaa koodatun datan takaisin alkuperäiseen muotoonsa” (TEPA-termipankki 2023).

Ekvivalenssi (engl. *equivalence*) – ”käännöksen ja lähtötekstin yksiköiden vastaavuus ja kommunikatiivinen samanarvoisuus” (Tieteen termipankki 2023).

Etukäteiseditointi (engl. *pre-editing*) – ”konekääntämisen työvaihe, jossa lähtötekstiä editoidaan sen saattamiseksi usealle tulokielelle helpommin käännettävään muotoon” (Tieteen termipankki 2023).

Interlingua (engl. *interlingua*) – ”apriorinen apukieli” (Tieteen termipankki 2023).

Jälkieditointi (engl. *post-editing*) – ”käännösprosessin osa, jossa ihminen tarkastaa ja korjaa konekäännöksen” (Tieteen termipankki 2023).

Kohdeteksti, tuloteksti (engl. *target text*) – ”käännösprosessin tuloksena syntyvä teksti” (Tieteen termipankki 2023).

Kohdistus (engl. *alignment*) – ”monikielisten tekstien tai ilmausten kohdistaminen” (Tieteen termipankki 2023).

Konekäännin, konekäännösohjelma (engl. *machine translation software*, *machine translation engine*) – ”ohjelma, joka kääntää automaattisesti luonnollisesta kielestä toiseen” (TEPA-termipankki 2023).

Konekäännös, konekääntäminen (engl. *machine translation*) – ”tekstin tai puheen kääntäminen konekääntimen avulla” (TEPA-termipankki 2023).

Konekäännösjärjestelmä (engl. *machine translation system*) – ”applications or online services that use machine-learning technologies to translate large amounts of text from and to any of their supported languages” (suom. sovellukset tai verkkopalvelut, jotka käyttävät koneoppimisteknologiaa kääntääkseen isoa määrää tekstiä mistä tahansa kielestä ja mihin tahansa kieleen, joita ne tukevat) (IATE 2023).

Konekäännöstuotos (engl. *machine translation output*) – ”käännetty teksti, joka tuotetaan käyttämällä tekoälyä sisällön automaattiseen kääntämiseen lähdekielestä kohdekieleen” (IATE 2023).

Koneoppiminen (engl. *machine learning*) – ”toiminta, jossa ohjelma oppii toistuvista tapahtumista ja muuttaa itse toimintaansa annetun tavoitteen suuntaan saamansa palautteen perusteella” (TEPA-termipankki 2023).

Kooderi (engl. *encoder*) – ”ohjelma tai laite, joka muuntaa datan toiseen muotoon määritettyjen sääntöjen mukaan” (TEPA-termipankki 2023).

Lähtöteksti, lähdeteksti (engl. *source text*) – ”käännösprosessin lähtökohtana oleva teksti” (Tieteen termipankki 2023).

Neuroverkko (engl. *artificial neural network, neural network*) – ”mukautuva tietojenkäsittelyn järjestelmä, joka koostuu suuresta joukosta yksinkertaisia, toisiinsa yhteydessä olevia laskenta- ja tiedonkäsittely-yksikköjä” (TEPA-termipankki 2023).

Rinnakkaiskorpus, paralleelikorpus (engl. *parallel corpus, bitext*) – ”kaksikielinen korpus” (Tieteen termipankki 2023).

Syväoppiminen (engl. *deep learning*) – ”tekoälyn oppimismenetelmä, jossa hyödynnetään usean kerroksen neuroverkkoja ja jonka tavoitteena on luoda algoritmin avulla neuroverkko, joka pystyy ratkaisemaan sille annetut ongelmat” (IATE 2023).

Tietokoneavusteinen kääntäminen, CAT (engl. *computer-assisted translation*) – ”kääntämisen menetelmä, jossa kääntäjä hyödyntää erityisesti kääntämisen tehostamiseen suunniteltuja tietoteknisiä työkaluja” (esimerkiksi käännösmuisteja ja terminhallintavälineitä) (Tieteen termipankki 2023).

Välikieli (engl. *pivot language*) – ”kieli, jonka kautta kääntäminen tai tulkkaus tehdään, jos sitä ei tehdä suoraan lähtötekstistä” (Tieteen termipankki 2023).

1.2. Konekääntimien kehitys ja eri lähestymistavat

Konekääntämisen kehityksessä voi laajasti erottaa neljä vaihetta. 1950-luvun ensimmäisten testausten jälkeen edettiin vuosina 1960–1980 konekääntämisen kehityksessä suorakäännösjärjestelmin (engl. *direct approach*). 1980-luvulla sai alkunsa sääntöpohjainen konekääntäminen (RBMT, engl. *rule-based machine translation*), joka kehittyi suorakääntämisen lähestymistavasta. 1990-luvulla päästiin korpuspohjaisiin menetelmiin: tilastolliseen (SMT, engl. *statistical machine translation*) ja esimerkkipohjaiseen konekääntämiseen (EBMT, engl. *example-based machine translation*). (Quah 2006: 58–64.) 2000-luvun alussa kehitettiin muun muassa ns.

hybridijärjestelmiä, joissa yhdistettiin tilastollisten mallien vahvuuksia ja lingvistisiä sääntöjä. (Yu Shiwen, Bai 2015: 187; Poibeau 2017: 114.) Nykyiset huipputeknologiset syväoppimiseen perustuvat eli neuroverkkokonekääntimet (NMT, engl. *neural machine translation*) pääsivät vauhtiin 2010-luvulla. (Poibeau 2017: 121.) Kuvaan seuraavaksi lyhyesti konekääntämisen eri lähestymistapojen toimintaperiaatteita.

1.2.1. Suorakäännösjärjestelmät

Konekääntämisen ensimmäisenä sukupolvena pidetään suorakäännösjärjestelmiä, jotka osoittautuivat melko alkeellisiksi ja pitkää käsittelyaikaa vaativiksi. (Quah 2006: 69.) Tämäntyyppiset järjestelmät luotiin struktuuriltaan ja lauserakenteeltaan samanlaisia kieliä varten (Kaalep, Koit 2010: 727). Suorakäännösjärjestelmät toimivat laajojen kaksikielisten sanastojen, morfologisen analyysin ja tekstinkäsittelyohjelmien varassa (Quah 2006: 70). Nämä sanastopohjaiset käännösjärjestelmät käänsivät sanasanaisesti lähtökielistä tekstiä kohdekieleen ja käyttivät tiettyjä sääntöjä sanajärjestyksen mukauttamiseen kohdekieleen sopivaksi. Syntaktista analyysia ei tässä lähestymistavassa toteutettu. (Poibeau 2017: 27.)

1.2.2. Sääntöpohjainen konekääntäminen

Toista sukupolvea konekääntämisen arkkitehtuurissa edusti sääntöpohjainen konekäännin, jossa oli kaksi lähestymistapaa: siirto (engl. *transfer*) ja interlingua (Quah 2006: 68). Näissä käännösmalleissa tapahtui eri määrin lähtökielen analyysi ja kohdekielen synteesi (Yu Shiwen, Bai 2015: 187). RBMT:n siirtomallit olivat monimutkaisempia kuin suorakäännösjärjestelmät, koska ne sovelsivat jonkin verran syntaktista analyysia. Mikäli syntaktisen komponentin avulla selvisi tieto lähtö- ja kohdekielestä, siirtomallisen konekääntimen tuotoksen olisi pitänyt olla idiomaattisempi kuin suorakäännöksen. RBMT:n interlingua -lähestymistavassa käytettiin välikieltä, jonka kautta toteutettiin käännös lähtökielestä kohdekieleen. (Poibeau 2017: 27.) Interlingua-malli mahdollisti monikielisten käännösjärjestelmien kehittämisen (Yu Shiwen, Bai 2015: 188). RBMT tarvitsi sekä perusteellisia sanakirjoja morfologisiin,

syntaktisiin ja semanttisiin tietoihin että käsin laadittuja kielioppeja (Liin ym. 2012: 21). Sääntöpohjainen konekäännin jäi vallitsevaksi vuosikymmeniksi (Poibeau 2017: 40).

1.2.3. Tilastollinen konekääntäminen

Internetin laaja levikki ja tietokoneiden tehon kehitys 1990-luvun alussa mahdollistivat konekääntämisessä tilastollisten menetelmien soveltamisen, sillä ne vaativat isoja tekstimääriä. (Poibeau 2017: 30.) Kaksikielisiin korpuksiin perustuvassa konekääntämisessä oli käytössä kaksi menetelmää: tilastollinen ja esimerkkipohjainen. Toisin kuin RBMT:ssä näissä ei sovellettu konekääntimissä tekstien analyysissa ja käännösekvivalenssien valinnassa lingvistisiä sääntöjä. (Quah 2006: 76–77.) Rinnakkaiskorpukset sisälsivät kohdistettuja lähtökielisiä tekstejä ja näiden kohdekielisiä käännöksiä, jotka olivat struktuuriltaan yhdistettyjä virketasolla. Laajojen digitaalisten tekstikorpusten käyttämisen avulla oli mahdollista tunnistaa ekvivalenssien malleja. Tilastollisilla laskelmilla selvitettiin kohdistettujen kaksikielisten tekstien eri käännösekvivalenssien todennäköisuudet. (Quah 2006: 64.) SMT:ssä segmentoitiin lähtöteksti ensin sanoiksi ja lausekkeiksi ja verrattiin näitä sen jälkeen isoon kohdistettuun kaksikieliseen rinnakkaiskorpukseen (Quah 2006: 77). Tämä kouluttamiseksi tai koodaukseksi kutsuttu vaihe sisälsi kielitietojen koodausta. (Poibeau 2017: 98–99.) Lauseke ei tässä kontekstissa välttämättä tarkoittanut sisällöltään yhtenäistä yksikköä kuten kielitieteessä, vaan satunnaista vierekkäisten sanojen riviä virkkeessä. Käännettäessä lauseke kerrallaan voi vähentää yksittäisten sanojen monitulkintaisuudestaan johtuvia väärinkäsityksiä sekä kieliopillista kömpelyyttä. (Kaalep, Koit 2010: 731.) Algoritmi valitsi käännöksen, jonka se käsitti todennäköisimmäksi yhdistämällä SMT:n kahdesta komponentista, käännös- ja kielimallista saatuja tietoja (Quah 2006: 78). SMT:n kolmas tärkeä vaihe oli dekooderi, jossa järjestelmä yritti dekoodata syötevirkettä ja löytää parhaan mahdollisen käännöksen. Se tarkoitti mahdollisten valintojen lajittelemista, jossa jokaiselle sanalle löytyi monta käännösekvivalenssia. (Poibeau 2017: 99.)

1.2.4. Esimerkkipohjainen konekääntäminen

Esimerkkipohjainen konekäännin muistutti tietokoneavusteista käännöstyökalua (engl. *computer-aided translation tool*), käännösmuistia (engl. *translation memory*) (Quah 2006: 83) ja kohdekielisen käännöksen luominen tapahtui siinä kolmessa vaiheessa. Järjestelmä yritti ensin löytää lähtökielen korpuksesta käännettävän lauseen osia. Sen jälkeen etsittiin kaksikielisistä käännösteksteistä kohdekielisiä käännösekvivalensseja. Lopuksi järjestelmä yritti yhdistää käännösfragmentit virheettömäksi virkkeeksi. (Poibeau 2017: 80.) SMT konekääntimet olivat kielistä riippumattomia, koska niitä ei ollut sovellettu tiettyjä kieliä varten (Liu, Zhang 2015: 201). Tilastollisten menetelmien ansiosta konekäännös oli siitä lähtien saatavissa isommalle määrälle käyttäjiä. Sitä käytettiin esimerkiksi Googlen konekääntimessä (Poibeau 2017: 30).

1.2.5. Neuroverkkokonekääntäminen

Neuroverkkokonekäännös kuten tilastollinenkin on rinnakkaiskorpukseen perustuva konekäännin, jota on koulutettava erittäin isoilla datamäärillä. Käännösmuistien tapaiset korpukset koostuvat käännösyksiköistä, joissa on yhdistetty lähtökielen segmentit (useimmiten lauseet) ja näiden käännökset. Siinä mielessä NMT muistuttaa edellisen sukupolven konekäännintä SMT:tä, mutta se käyttää aivan toisenlaista lähestymistapaa: neuroverkkoja. (Forcada 2017: 292.) NMT:ssä hyödynnetään hierarkkista syväoppimisen menetelmää, joka aloittaa analysoitavan objektin eli virkkeen analysoinnin perustasolta, merkeistä tai sanoista, jotta se pystyisi tunnistamaan monimutkaisempia rakenteita kuten sanarivejä tai lausekkeita, kunnes lopuksi onnistuu analysoida koko virke. Tarkoituksena on antaa järjestelmän tehdä itsenäisesti päätelmiä tietojen perusteella. (Poibeau 2017: 122.) Neuroverkkojen kouluttaminen on aikaa vievä prosessi, joka voi kestää päivistä kuukausiin (Forcada 2017: 295).

NMT koostuu kooderista, järjestelmän osasta, joka analysoi koulutustietoja, ja dekooderista, joka tuottaa automaattisesti virkkeen käännöksen perustuen tietoihin, joita on analysoitu kooderilla. Järjestelmä yrittää selvittää ja ryhmittää sanoja, jotka esiintyvät samanlaisessa käännöskontekstissa. (Poibeau 2017: 123.) Sekä SMT:ssä että NMT:ssä kohdelause on lähtölauseen käännös tietyllä todennäköisyydellä. Kaikki kohdelauseet

voivat periaatteessa olla lähtölauseen käännöksiä. Molemmassa konekääntimessä valitsee dekooderi mahdollisimman todennäköisimmän kohdelauseen. NMT laskee kohdelauseen todennäköisyyden ottamalla huomioon kunkin lähtölauseen kohdesanan todennäköisyyden ja kohdelauseessa esiintyvät edelliset sanat. (Forcada 2017: 300–301.) Konekääntämisen neuroverkon arkkitehtuurissa on tärkeä sija myös huomiomekanismilla (engl. *attention mechanism*), joka auttaa luomaan yhteyksiä kooderin ja dekooderin välillä. Se perustuu siihen, että kaikki lähtölauseen sanat eivät ole käännösprosessissa jatkuvasti yhdenvertaisia, vaan jotkut niistä ovat erittäin tärkeitä. Huomiomekanismi auttaa järjestelmää aina keskittymään dynaamisesti lauseen olennaisiin käännettäviin osiin. Se on erittäin hyödyllistä pitemmissä lauseissa, mutta myös typologisesti etäisten kielten kääntämisessä. NMT:n kehitys on ollut nopeaa ja vallankumouksellista kuten 1990-luvulla SMT:n, mutta se on tapahtunut suuremmallakin vauhdilla. Syväoppimisen menetelmä on konekääntämisessä todistanut tehokkuutensa ensin lyhyiden virkkeiden kääntämisessä läheisten sukulaiskielten välissä ja tullut myöhemmin toimeen myös pitkien lauseiden ja etäisten kielten kanssa. (Poibeau 2017: 126–128.)

1.3. Konekääntämisen ongelmakohdat

Konekääntimen tuloksekkuuden päättelyyn vaikuttaa mm. se, minkä verran ihmisen tulee puuttua järjestelmän työprosessiin. Käsitteellä ihmisavusteinen konekääntäminen (engl. *human-aided MT*) tarkoitetaan koneen tuottamaa käännöstä, joka voi tarvita ihmisen avustusta joko lähtötekstin etukäteiseditoinnissa tai konekäännöstuotoksen jälkieditoinnissa. (Quah 2006: 11.) Etukäteiseditoinnin tarkoitus on yksinkertaistaa käännettävää tekstiä, jotta konekäännöstuotos olisi parempilaatuinen ja sen ansiosta pienenisi myös jälkieditoinnin tarve. (Quah 2006: 45.) Käännöslaadun takaamiseksi voi konekääntimeen syöttää myös kontrolloitua kieltä (engl. *controlled language*), josta ihminen on poistanut polysemian ja monimutkaiset lauserakenteet. (Quah 2006: 12.) Sillä karsitaan mahdollisten kielivirheiden määrää (Quah 2006: 48). Ajan mittaan konekääntimien kehityksessä on pyritty siihen, että ihmisen tulisi mahdollisimman vähän tai ei ollenkaan puuttua käännösprosessiin, jotta järjestelmä olisi täysautomaattinen.

Ihmisen konekääntimeen työprosessiin puuttumisen tarvetta on pidetty ongelmakohtana. Seuraavaksi käsittelen kokoavasti eri konekääntimien ongelmakohtia.

Konekääntämisen alkuvaiheen isoimmaksi ongelmaksi osoittautui sanojen kaikkien käyttökontekstien huomioon ottaminen eli kielen monimerkityksisyys. (Poibeau 2017: 29, 44.) Tilastollinen analyysi voi auttaa määrittelemään sanojen eri merkityksiä ja konteksteja, mutta esteenä voi olla lauseiden merkitykset ja näiden väliset suhteet. NMT yrittää ratkaista kyseisen ongelman. (Poibeau 2017: 119.) Esitän NMT:n haasteet tämän luvun loppupuolella ja tuon ensin esiin edellisten sukupolvien konekääntimien ongelmakohdat.

Suorakäännösjärjestelmät eivät tulleet toimeen polysemian kanssa. Samoin oli ongelmia kääntämisessä etäisten kielten välillä. (Quah 2006: 69.) RBMT:n puutteet liittyivät pääasiassa sääntöjen ja sanastojen riittävyyteen, näiden rakentamisen menetelmiin ja kustannuksiin sekä monimerkityksisyyteen ja idiomaattisuuteen toimentuloon (Yu Shiwen, Bai 2015: 192–193). Sääntöpohjainen konekäännin vaati älykkäänä järjestelmänä kielisääntöjen manuaalista muokkaamista, joka on kallista, aikaa vievää ja monimutkaista. (Liu, Zhang 2015: 201; Poibeau 2017: 40.) RBMT:n heikkous oli myös se, että konekäännin vaati lauseen kokonaista analyysia, ja jos vaikka vain yhtä osaa ei onnistunut analysoimaan, niin käännösvastinetta ei tarjottu koko lauseelle (Poibeau 2017: 79).

Tilastollisen konekääntimen luominen vaati sääntöpohjaiseen verrattuna vähemmän ihmiskättä (Liin ym. 2012: 21) ja se huomioi myös kielelle ominaisia idiomeja ja fraseologisia ilmaisuja (Fišel 2016). Vaikka SMT:n kuten NMT:nkin toimintaperiaate oli antaa konekääntimen päättää itsenäisesti tietojen perusteella parhaasta representaatiosta, se ei toiminut yhtä sujuvasti kuin syväoppimista soveltavissa järjestelmissä, koska liian monia parametreja täytyi sopeuttaa käsin. (Poibeau 2017: 122.) Myös konekääntimessä käytettävät rinnakkaiskorpukset eivät olleet isosta koosta huolimatta riittäviä kattamaan harvinaisia tai monimutkaisia kielivivahteita (Poibeau 2017: 106). Jos kaksikielinen korpus oli liian pieni, niin järjestelmä ei pystynyt tehokkaasti luomaan hyviä käännöksiä (Quah 2006: 80).

Korpusten datamäärä voi kielittäin vaihdella huomattavasti. Rinnakkaiskorpuksia luotiin automaattisilla menetelmillä webissä saatavissa olevien tekstien ja näiden kohdekielisten vastineiden perusteella. Prosessin konemaisuuden takia rinnakkaistekstien laatu ei ollut varmistettu, mutta tekstien iso määrä takasi sen, että näiden joukosta löytyi myös hyvälaatuisia. (Poibeau 2017: 72.) Kielten laajuus voi säätää korpuksille ja siten myös konekääntimelle rajoituksia. Jos tietystä kielestä ei onnistunut kerätä riittävästi dataa, käytettiin esimerkiksi englantia välikielenä ja hankittiin aineistoa sen kautta. Se ei kuitenkaan ratkaissut ongelmaa täysin, koska englanninkieliset tekstit voivat joskus olla keskinkertaisia ja ns. tuplakääntäminen voi aiheuttaa virheiden todennäköisuutta. (Poibeau 2017: 112.)

Neuroverkkojen kouluttaminen on monimutkaisempaa kuin tilastollisessa lähestymistavassa ja se vie enemmän aikaa (Forcada 2017: 302). Syväoppimisen menetelmillä on saatu huomattavia tuloksia luonnollisen kielen käsittelyssä, mutta siitä huolimatta järjestelmä epäonnistuu usein virkkeen kokonaisrakenteen tunnistamisessa. NMT:n merkittävimmät ongelmakohdat ovat neuroverkkojen kouluttamisen vaikeuksien lisäksi koulutusaineistosta puuttuvien eli tuntemattomien sanojen sivuuttaminen ja tiettyjen sanaryhmien kääntämättömyydet, jotka voivat aiheuttaa outoja ja epätäsmällisiä käännöksiä. (Poibeau 2017: 125–126.) NMT:n konekäännöstuotos on usein SMT:n tuotoksen kaltainen monissa ikävissä virhetyypeissä kuten epäjohdonmukaisuus numeroissa, linkeissä ja termeissä, erisnimien väärä kääntäminen jne. Se voi olla myös erilaista sisältäen yllättäviä virheitä semantiikassa. (Forcada 2017: 303.) Neuroverkkojen toimimista on yhä hankala ymmärtää ja analysoida, koska datan sisäinen kuvaus on vain numeerista ja monimutkaista eikä ihmiselle suoraan luettavaa (Poibeau 2017: 128).

2. Konekäännöksen arviointi

Konekäännöksestä on tullut osa arkielämäämme, minkä takia konekäännösten laatu on yhä tärkeämpi. Käännöstuotoksen laadun arviointi on tärkeä osa konekääntimien kehitystyöstä (Quah 2006: 130; Kit, Wong 2015: 214). Konekääntämistä on tutkittu paljon, eri näkökulmista ja kieliparien kohden, enemmän siitä lähtien, kun NMT on tullut laajemmin käyttöön. Tapoja, miten voi tutkia konekäännöstuotosta, on monia. Voi arvioida esimerkiksi sen tarkkuutta, sujuvuutta tai sopivuutta tiettyihin tekstityyppeihin.

Konekäännöksen ensisijaisessa laadun arvioinnissa käytetään useimmiten automaattisia arviointimenetelmiä, koska ne vaativat mahdollisimman vähän henkilöresurssia ja tulokset selviävät nopeasti. Matemaattisten laskelmien ohella on ajoittain tarve tutkia konekäännöstä myös kvalitatiivisilla menetelmillä, jolloin konekäännöstuotosta arvioi ihminen. Molemmassa lähestymistavassa on vertauskohteena referenssikäännös eli verrataan konekäännöstä ihmisen tekemään käännökseen. (Quah 2006: 135–136.)

Seuraavaksi esittelen konekääntämisen eri arviointimenetelmiä ja annan esimerkkejä tutkimuksista, joissa käsitellään konekääntämistä eri näkökulmista. Ensin pohdin hyvän käännöksen käsitettä ja ihmis- ja konekäännöksen lähtökohtia.

2.1. Ihmiskäännös *vs* konekäännös

Käännösten arviointi edellyttää sopimista käsityksestä, mitä ylipäänsä tarkoittaa hyvä käännös ihmiskääntäjän ja konekäännöksen näkökulmasta. Varsin yleisesti ajatellaan, että hyvä käännös muistuttaa sävyltään, tyyliltään ja yleiseltä rakenteeltaan lähtötekstiä, se on kielellisesti virheetön ja idiomaattinen. Lukijan ei tulisi tajuta, että kyseessä on käännös. Vaikka käännöksen tavoitteena on lähtötekstin samankaltaisuus, niin sanasanainen käännös taas ei ole hyvää käytäntöä, koska lopputulos voi olla vaikeaselkoinen ja epäidiomaattinen. Kääntäjä tutustuu yleensä käännettävään kokonaistekstiin, ennen kuin ryhtyy työhön. Hyvän kohdekielisen tekstin onnistumisessa auttaa myös taustatietojen kuten tekstin käyttötarkoituksen ja kontekstin tuntemus.

Kääntäminen on näin ollen monimutkainen prosessi, johon tarvitaan korkeatasoisia kognitiivisia ja lingvistisiä taitoja. Nämä taidot eivätkä tarvittavat taustatiedot ole konekääntimen saatavissa. (Poibeau 2017: 16–17.)

Luonnollisen kielen sovitus koneella on kielen monimerkityksisyyden takia hankala. Sanastoista löytyvät sanojen eri merkitykset, mutta konekääntimelle aiheuttaa vaikeuksia käännöskontekstiin sopivien vastineiden poimiminen. Ihmiskääntäjälle polysemia ei tuota ongelmia, koska hän ymmärtää kontekstin. (Poibeau 2017: 19–21.) Herää kysymys, odotammeko ihmisen ja koneen tekemästä käännöksestä samanlaista lopputulosta. Konekäännöksen tarkoitus on luoda vakiokäännöksiä eikä tavoitella ainutkertaisuutta (Poibeau 2017: 72). Siitä huolimatta odotamme konekäännöstuotoksen olevan kieliopillisesti virheetön ja sisällöltään tarkka. Viimeistä edellytämme myös ihmisen tekemästä käännöksestä, mutta sen lisäksi jotakin selittämätöntä, mikä tekee käännöksestä ainutlaatuisen.

Konekäännöksen käsitteen määritelmä ”tekstin tai puheen kääntäminen konekääntimen avulla” on jo annettu luvussa 1.1., mutta ihmiskäännöksen käsitekin ansaitsee tarkemman selvennyksen. Nykyään on melko yleistä, että ammattikääntäjät käyttävät työssään erilaisia käännöstyökaluja kuten käännösmuistiohjelmia, digitaalisia sanastoja, termikantoja tai muita vastaavia resursseja. Sitä lähestymistapaa kutsutaan yhteisesti tietokoneavusteiseksi kielen kääntämiseksi (engl. *machine-aided translation*). (Quah 2006: 13–14.) Samalla ihmiskääntäjät ovat konekääntämisen paineen alla, sillä heidän odotetaan käyttävän konekäännösjärjestelmiä, koska se voi tietyissä oloissa nopeuttaa käännösprosessia ja säästää kustannuksiakin. (Quah 2006: 18)

Ihmiskäännös eroaa konekäännöksestä lähinnä siinä, että ihminen tekee tietoisia käännösvalintoja, joita hän pystyy perustelemaan. Toisin sanoen ihmiskäännöksessäkin voi käyttää konetta apuvälineenä, mutta käännösstrategian valitsee kuitenkin kääntäjä itse. Konekäännöstuotoksen vaikuttajana riippuen konekäännintyypistä voi olla konekäännösohjelmaan syötettyjen kielisääntöjen, koulutusaineiston tai yleisesti ottaen tarvittavan datamäärän vajavaisuus kuten luvusta 1.3. kävi ilmi. Ihmiskääntäjään vaikuttavat etenkin käännettävän kohdekielen taito, muun muassa muiden kielten osaaminen, ja yleiset tiedot ja käännösalan pätevyys. Käännöslaatu voi riippua myös

alkutekstistä ja sen kielellisestä laadusta. Tämä koskee sekä kone- että ihmiskäännöstä. Huonolaatuisesta lähtötekstistä on hankalaa, joskus jopa mahdotonta tuottaa hyvää kohdekielistä tekstiä.

2.2. Konekäännöksen tutkimisesta

Konekääntäminen tarjoaa tutkimuskohteena runsaasti aineistoa. Toisaalta tutkitaan eri järjestelmien toimivuutta ja kehitystä, toisaalta esimerkiksi käännöstuotoksen laatua kielellisestä näkökulmasta tai suhtautumista konekääntämiseen. Esittelen muutamalla esimerkillä tämän alan eri tutkimussuuntia. Tavoitteena on tuoda esiin ei niinkään tutkimustuloksia kuin eri tutkimuskohteita ja -menetelmiä.

Anna Fernández-Torné ja Anna Matamala vertailivat espanja–saksa-kieliparin kolmen eri konekääntimen tuotosta ja keskittyivät ihmiseen vaikuttaviin tekijöihin arviointiprosessissa. He tutkivat mm. konekäännöstuotoksen sujuvuutta, tarkkuutta sekä jälkieditoinnin tarvetta ja sen työläyttä (engl. *post-editing effort*) ja ihmisarvioijien suhtautumista siihen. Konekäännöstuotoksen laatua arvioitiin sekä segmentin että kokonaisen tekstin tasolla. Vertailukohteina olivat aihekohtaisten SMT ja NMT-konekääntimien ja yleisen konekääntimen tuotos. (Torné, Matamala 2021: 97.) Tutkimuksen tavoitteena oli kerätä vertailevasti sekä kvalitatiivista että kvantitatiivista dataa (Torné, Matamala 2021: 100). Kvalitatiivisessa tutkimuksessa sovellettiin pääasiassa TAUS:in (engl. *Translation Automation User Society*) ohjeita ja kvantitatiivisessa osuudessa automaattisia arviointimenetelmiä BLEU (engl. *Bilingual Evaluation Understudy*), METEOR (engl. *Metric for Evaluation of Translation with Explicit Ordering*) ja TER (engl. *Translation Error Rate*). (Torné, Matamala 2021: 102–103.) Tutkimuksessa todettiin, että automaattisilla arviointimenetelmillä ja ihmisarvioinnilla saatujen tulosten välillä on hyvä korrelaatio (Torné, Matamala 2021: 117).

Suomea käsittelevistä konekäännöstutkimuksista erottui Ari Gröhnin pro gradu, joka keskittyy pääasiassa siihen, miten NMT soveltuu erityyppisten tekstien englanti–suomi-kääntämiseen. Gröhn vertaili kahden eri konekäännösohjelman tuotosta ihmisen tekemään referenssikäännökseen. Tavoitteena oli tutkia, löytyykö teksteistä kielellisiä

piirteitä, joiden avulla voi ennustaa konekääntimen sopivuutta tiettyjen tekstien kääntämiseen. Tutkimuksessa yhdistettiin automaattisia ja manuaalisia menetelmiä. Tekstityyppiluokittelussa sovellettiin Biberin viittä dimensiota, automaattisessa arvioinnissa LeBLEU (engl. *Letter-edit-BLEU* tai *Levenshtein-BLEU*) -menetelmää ja manuaalisessa virhekategorisoinnissa TAUSin DQF-MQM-luokittelua (engl. *Dynamic Quality Framework-Multidimensional Quality Metrics*). Gröhn toteutti myös lausepituuden analyysin, jonka kautta hän halusi tunnistaa korrelaatiota käännösten keskimääräisen LeBLEU-tulosten ja erityyppisten tekstien lausepituuksien välillä. (Gröhn 2019: 1–2.) Tutkimuksesta ilmeni, että automaattisiin menetelmiin vertailuaineistona tarvittavia, vapaasti käytettävissä olevia ihmiskäännöksiä oli hankala löytää myös sen takia, että tutkimuskohteena oli suomen kieli. (Gröhn 2019: 39.)

Viron kielen sopivaisuutta konekäännökseen on tutkinut esimerkiksi Euroopan Komission käännöstoimen viron kielen pääosaston kääntäjä Ingrid Sibul. Hän kuvaili vuonna 2018 aikakauslehdessä Õiguskeel ilmestyneessä artikkelissa, miten hän arvioi Euroopan Unionin virastoja varten luodun NMT:n eTranslation (eTranslation 2022) tuottamia käännöksiä. Sibul vertaili keskenään kuuden eri aihepiirin ja erityylisen tekstin tilastollista ja neuroverkkokonekäännöstuotosta ihmisen tekemään käännökseen. Ensin hän toteutti subjektiivisen analyysin eli analysoi eri konekääntimillä tehtyjä käännöksiä manuaalisesti. Sen jälkeen hän mittasi automaattisen arviointimenetelmän Levenshteinin editointietäisyyden (LE, engl. *Levenshtein distance*) algoritmin avulla konekäännettyjen tekstien etäisyyden ihmiskäännöksestä. Samalla hän arvioi konekäännöstuotoksia lisäksi neljällä eri automaattisella arviointimenetelmällä (BLEU, NIST, TER, CharCut). (Sibul 2018: 5.)

2.3. Kvantitatiiviset arviointimenetelmät

Konekäännöksen arvioimiseen on monia mahdollisuuksia ja yhtä varmaa vakiomenetelmää siihen ei ole, koska arviointituloksia tarvitaan eri syistä. Yleinen käsitys kuitenkin on, että arviointi on järjestelmien kehityksen kannalta oleellista. (Quah 2006: 129.) Kääntäjillä ja kehittäjillä on konekäännöksen arvioinnissa eri tavoitteet. Kehittäjät tarvitsevat esimerkiksi tietoa järjestelmän mahdollisista rajoituksista kuten sen

tekstityyppien ja aihepiirien kääntämisen kattavuudestaan. Kääntäjiä taas kiinnostaa konekääntimen tuottavuus käännöstyössä, missä määrin voi luottaa sen tuotokseen. (Kit, Wong 2015: 217.)

Jo ensimmäisten konekääntimien kehittämisen aikana heräsi kysymys, miten eri järjestelmiä voi verrata keskenään ja mitata niiden laatua ja paranemista. (Poibeau 2017: 131.) 1990-luvun loppuun saakka konekääntämisen arviointi perustui pääasiassa ihmisarvioijien työhön, joka on aikaa vievää, kallista ja osittain epäjohdonmukaista. Sen takia yritettiin luoda automaattisia arviointimenetelmiä, jotka eivät vaatisi ihmisen puuttumista arviointiprosessiin. (Poibeau 2017: 134.) Aikaisempi konekäännöksen arviointi keskittyi vain kahteen aspektiin: ymmärrettävyyteen ja tarkkuuteen (Quah 2006: 131). Konekäännöksen kehittäjät tarvitsivat arviointimenetelmiä, jotka olisivat luotettavia, nopeita, toistettavia, mahdollisimman edullisia (Poibeau 2017: 131) sekä kielestä riippumattomia ja ihmisarviointeihin verrattavia (Quah 2006: 136).

Konekäännöksen automaattisilla arviointimenetelmillä verrataan useimmiten konekäännöksen ja ihmisen luoman referenssikäännöksen tilastollisia arvoja (Koponen 2010: 1). Niiden avulla ei arvioida käytännössä niinkään konekäännöksen laatua kuin sen samanlaisuutta ihmiskäännökseen. (Kit, Wong 2015: 232.) Eri menetelmät eroavat siinä, minkä perusteella niissä päätetään käännösten läheisyydestä, ja ne on tarkoitettu varsinkin tietyn konekääntimen seuraavien versioiden tai samanlaisten konekääntimien vertailevaan arviointiin. Menetelmien yksityiskohdissa voi olla eroja, mutta niiden perusta on yleensä konekäännöstuotoksen vertailu ihmiskäännökseen: verrataan testilauseita, jotka on käännetty sekä koneen että ihmisen toimesta. (Kaalep, Koit 2010: 736.) Kuten seuraavasta katsauksesta käy ilmi, esiintyy myös muita lähestymistapoja.

Esittelen muutamaa laajalti käytössä olevaa automaattista arviointimenetelmää, joita voi yleisesti kutsua tekstin samankaltaisuuden mittareiksi (engl. *text similarity metrics*) (Kit, Wong 2015: 225). Käytetyin näistä on sen rajoituksista huolimatta sanaperustainen BLEU-menetelmä (Poibeau 2017: 136). Sillä saadaan arviointitulokset pikaisesti, se on kielistä riippumaton ja vaatii vähemmän resurssia. BLEU-menetelmällä saatujen ja ihmisen tekemän arviointitulosten välillä on korrelaatio. (Papineni ym. 2002: 311) Sitä pidetään BLEU:n yhtenä vahvuutena (Papineni ym. 2002: 318). BLEU:n tärkein

lähtökohta: ”*the closer a machine translation is to a professional human translation, the better it is*” (suom. mitä lähempänä on konekäännös ammattimaista ihmiskäännöstä, sitä parempi se on) (Papineni ym. 2002: 311).

Eri kieliparien konekäännösten automaattisesta arvioinnista käy ilmi, että tulokset saattavat poiketa kielittäin suuresti. Arviointituloksiin voi mahdollisesti vaikuttaa esimerkiksi koulutusaineiston rajoitettu määrä, morfologisesti rikkaat kielet, joita on tiettävästi hankalampaa käsitellä automaattisesti tai käännökset etäisten kielten välillä. (Poibeau 2017: 138.) Suomalaistutkijat ovat sen takia kehittäneet morfologisesti monimutkaisimpia kieliä (esim. suomea ja viroa) varten LeBLEU-menetelmän (Virpioja, Grönroos 2015: 412).

Referenssikäännöksiin perustuvista menetelmistä nostan lisäksi esiin Levenshteinin editointietäisyyden, jossa mitataan kahden segmentin eroa eli montako lisäystä, poistoa tai korvausta tulisi minimissään tehdä, jotta segmentit olisivat yhdenvertaisia (Garcia 2015: 72). Samankaltainen, editointietäisyyteen perustuva menetelmä on TER, jota käytetään jälkieditoinnin työläyden mittaamisessa (Kit, Wong 2015: 229).

Samankaltaisuuden mittareiden ohella käytetään konekäännöksen arvioinnissa myös laadun estimoinnin (engl. *quality estimation*) menetelmää, jossa sovelletaan koneoppimistekniikkaa. Se eroaa edellisessä luvussa esitetyistä metodeista siinä, että konekäännöstuotoksen laatua ”ennustetaan” vertailematta sitä ihmiskäännökseen (Kit, Wong 2015: 230). Viimeisemmät konekäännöksen kvantitatiiviset arviointimenetelmät ovat siirtymässä sanatasolta virketasoon. Konekäännöksen laatua voi arvioida myös menetelmillä, joilla konekäännöstuotosta ei verrata ihmisen tekemään käännökseen vain käytetään pelkkää konekäännösohjelmaa. (Fomicheva ym. 2020: 539, 551)

2.4. Kvalitatiiviset arviointimenetelmät

Konekäännöksen kvalitatiivisessa analyysissa sovelletaan suurin piirtein kahta lähestymistapaa: sisäistä (engl. *intrinsic*) ja ulkoista (engl. *extrinsic*) arviointia. Sisäinen arviointi käsittää konekäännöstuotoksen kielellistä laatua ja jakautuu laadun arvioinniksi (engl. *quality assessment*), käännösten järjestämiseksi (engl. *translation ranking*) ja

virheanalyysiksi (engl. *error analyses*). Ulkoisen arvioinnin tavoitteena on testata konekäännöstuotoksen käytettävyyttä tietyssä tehtävässä, jota konekäännös on tarkoitettu helpottamaan. Näissä voi eritellä kolmea menetelmää: tiedon poimiminen (engl. *information extraction*), ymmärrettävyyden testaaminen (engl. *comprehension test*) ja jälkieditointi. (Kit, Wong 2015: 221–225.) Käsittelen seuraavaksi näitä arviointimahdollisuuksia ja keskityn tarkemmin virheanalyysiin.

Sisäisessä lähestymistavassa, laadun arvioinnissa, eritellään kahta yleisintä kriteeriä: tarkkuutta ja ymmärrettävyyttä. Ihmisarvioijia pyydetään arvioimaan 5 pisteen asteikolla käännösten ”hyvyyttä”. Arviointi perustuu heidän vaistoonsa. Toisenlaisen metodin käännösten järjestäminen tapahtuu arvioijien mieltymyksien mukaan. Heidän on järjestettävä käännösvariantit parhaasta huonoimpaan. (Kit, Wong 2015: 222.)

Virheanalyysi taas lähestyy konekäännöstuotosta päinvastaisesta näkökulmasta, mittaamalla sen ”huonoutta”. Ensin tunnistetaan käännösvirheet ja ne luokitellaan sitten virhetyyppeihin. (Kit, Wong 2015: 223.) Virheiden kategorisointi on ihmisarvioinnissa laajalti käytössä oleva menetelmä, jota seuraa usein myös virheanalyysi (Chatzikoumi 2020: 150). Virheanalyysiä pidetään luotettavampana menetelmänä kuin laadun arviointia, koska virheiden tunnistaminen on yleensä objektiivisempaa ja johdonmukaisempaa (Schwarzl 2001, viitattu Kit, Wong 2015: 223 mukaan). Toisaalta tämän metodin hankaluus koskee juuri virheiden tunnistamista ja luokittelua. Ilmeiset virheet syntaksissa tai sananvalinnassa eivät aiheuta ongelmia, mutta arvioinnin subjektiivisuus nouse esiin, kun kyseessä on arvioijan kielelliset preferenssit, esimerkiksi ilmaisuissa jne. Virheiden luokittelu etukäteen määriteltyihin kategorioihin voi olla usein ongelmallista virhetyyppien epätarkkojen rajojen takia. Se johtuu siitä, että virhetyypit ovat luonteeltaan läheisesti yhdistettyjä. (Arnold et al. 1994; Flanagan 1994: 65–72, viitattu Kit, Wong 2015: 223 mukaan.) Yksi virhe voi sopia moneen eri kategoriaan ja sitten on tehtävä valinta. Virhe- eli subjektiivisen analyysiin voi vaikuttaa myös kielten läheisyys (Poibeau 2017: 142).

Konekäännöstä käytetään tyypillisesti tiedon saamiseksi vieraskielisistä teksteistä. Ulkoisen arvioinnin alle kuuluvan tiedon poimimisen menetelmän avulla voi päätellä konekäännöstuotoksen käytettävyyttä. Ihmisarvioijia pyydetään vertailemaan oleellisia

tapahtumatietoja kuten *kuka*, *missä* jne. semanttisiin rooleihin (*agentti*, *paikka* jne.), koska nämä tapahtumatiedot liittyvät kone- ja ihmiskäännöksen ja lähtötekstin lauseissa läheisesti toisiinsa. Ymmärrettävyyden testaamisessa arvioijille annetaan luettavaksi saman lähtötekstin kappaleen konekäännöstuotos ja ihmisen tekemä käännös ja vastattavaksi lista kysymyksiä. Tarkoitus on selvittää, missä määrin konekäännin onnistuu lähtötekstin kääntämisessä ottamalla huomioon tarkkuuden ja ymmärrettävyyden. (Kit, Wong 2015: 223–224.)

Konekäännöksen ihmisarviointia voi toteuttaa myös jälkieditoinnilla, joka tarkoittaa konekäännöstuotoksen korjaamista (Quah 2006: 46). Se voi tekstin käyttötarkoituksesta riippuen olla kiireistä minimikorjausta tiedon saamiseksi tai täysin hiottu ammattimaisen käännöksen kaltainen. (Quah 2006: 47.) Jälkieditoinnin työläyden mittarina on konekäännöstuotoksen virheiden korjaamiseen kuluva aika. Lopputulos ilmaisee konekäännöstuotoksen käytettävyyttä. Herää kysymys, onko tämä mittari vertailukelpoinen, kun jälkieditointiin kuluva aika riippuu monista ulkopuolisista tekijöistä kuten toimittajan keskittymismahdollisuudesta, työympäristöstä jne. Jälkieditoinnin puolueettomana metodina on editointietäisyys, jolla lasketaan konekäännöstuotokseen tarvittavien muutosten (lisäysten, poistojen, korvaamisen ja sanajärjestyksen muuttamisen) määrä. Jälkieditoinnin tarkoitus on antaa ihmistoimittajien tehdä minimaalisesti muutoksia, jotta konekäännöstuotos olisi merkitykseltään tarkka. (Kit, Wong 2015: 225.)

Seuraavaksi annan yleiskatsauksen tämän maisterintutkielman virheanalyysissä käytettyyn TAUSin DQF & QT21-virheluokitteluun ja lopuksi käsittelen vielä konekäännöksen eri arviointimenetelmien vahvuuksia ja heikkouksia.

2.4.1. TAUSin DQF & QT21-virheluokittelu

TAUSin DQF & QT21-menetelmää käytetään konekääntämisen arvioinnissa laajalti sekä käännösteollisuudessa että akateemisissa tutkimuksissa. (Görög 2014: 453) DQF on vuonna 2011 TAUSin kehittämä alusta, joka on ollut vapaasti akateemisessa käytössä vuodesta 2014 alkaen. Alustalta löytyvät käännöslaadun arvioimiseen tarvittavat työkalut, mittarit, raportit ja tiedot. (Chatzikoumi 2020: 147.) Esittelen tarkemmin

TAUSin DQF & QT21-virheluokittelua, koska olen valinnut sen tutkielmani kvalitatiivisen analyysiosuuden perustaksi.

TAUSin DQF & QT21-virheluokittelun mukaan lähtö- ja kohdeteksti rinnastetaan ja ruvetaan sen jälkeen toteamaan mahdollisia käännösvirheitä sijoittamalla ne eri virhetyyppeihin. DQF & QT21 on yhdistetty versio aiemmista menetelmistä MQM ja DQF. (Lommel, Arle ym. 2015: 4–5.) Virhetyyppejä on näissä metodeissa aikaisemmin luokiteltu eri tavoin, koska tulkintamahdollisuuksia käännösvirheiden jaotteluun on monia. (Lommel, Arle ym. 2015: 15–17.)

TAUSin DQF & QT21-virheluokittelu:

1. **tarkkuus** (engl. *accuracy*) – kohdeteksti eli käännös ei vastaa sisällöltään tarkasti lähtötekstiä
   1.1. **lisäys** (engl. *addition*) – käännös sisältää tekstiä, joka ei esiinny lähtötekstissä
   1.2. **poisto** (engl. *omission*) – käännöksestä puuttuu lähtötekstissä esiintyvää sisältöä
   1.3. **väärä käännös** (engl. *mistranslation*) – käännös ei kuva lähtötekstin sisältöä oikein
   1.4. **liikakääntäminen** (engl. *over-translation*) – käännös on spesifisempi kuin lähtöteksti
   1.5. **alikääntäminen** (engl. *under-translation*) – käännös on vähemmän spesifinen kuin lähtöteksti
   1.6. **kääntämätön** (engl. *untranslated*) – sisältö, joka tulisi kääntää, on jätetty kääntämättä
   1.7. **sopimaton tarkka käännösmuistivastine** (engl. *improper exact TM match*) – käännösmuistia käytettäessä tarjotaan tarkkana käännösmuistivastineena sopimatonta käännöstä
2. **sujuvuus** (engl. *fluency*) – tekstin muotoon tai sisältöön liittyvät virheet
   2.1. **välimerkkien käyttö** (engl. *punctuation*) – välimerkkien väärä käyttö
   2.2. **oikeinkirjoitus** (engl. *spelling*) – oikeinkirjoitusvirheet
   2.3. **kielioppi** (engl. *grammar*) – kielioppivirheet (ts. ei välimerkkien käyttöön eikä oikeinkirjoitukseen liittyvät virheet)

2.4. **kieliopillinen rekisteri** (engl. *grammatical register*) – väärän kieliopillisen rekisterin käyttö (esim. virallisen kielimuodon käyttö epävirallisen sijaan)

2.5. **epäjohdonmukaisuus** (engl. *inconsistency*) – epäjohdonmukaisuudet tekstitasolla

2.6. **linkki/ristiviittaus** (engl. *link/cross-reference*) – linkkien/ristiviittausten epäjohdonmukainen käyttö

2.7. **merkkien koodaus** (engl. *character encoding*) – merkkien väärä koodaus

3. **termistö** (engl. *terminology*) – erikoisalan termin väärä käyttö

3.1. **ristiriitaisuus termipankin kanssa** (engl. *inconsistent with termbase*) – käännöksessä on käytetty termejä, jotka eivät vastaa termipankissa esiintyviä termejä

3.2. **epäjohdonmukainen terminologian käyttö** (engl. *inconsistent use of terminology*) – terminologian epäjohdonmukaisuus tekstitasolla

4. **tyyli** (engl. *style*) – tekstin tyyliin liittyvät virheet

4.1. **kömpelö** (engl. *awkward*) – vaikeasti ymmärrettävä teksti

4.2. **yrityksen tyyli** (engl. *company style*) – käännös on ristiriidassa yrityksen tyyliohjeen kanssa

4.3. **epäjohdonmukainen tyyli** (engl. *inconsistent style*) – tyylin epäjohdonmukaisuus tekstitasolla

4.4. **kolmannen osapuolen tyyli** (engl. *third-party style*) – käännös on ristiriidassa vaaditun tyyliohjeen kanssa

4.5. **epäidiomaattinen** (engl. *unidiomatic*) – käännös on kieliopillisesti oikea, mutta se ei ole idiomaattinen

5. **muotoilu** (engl. *design*) – tekstin muotoiluun liittyvät virheet

5.1. **pituus** (engl. *lenght*) – merkittävä ero lähtö- ja kohdetekstin pituudessa

5.2. **paikallinen muotoilu** (engl. *local formatting*) – paikallisiin muotoilunormeihin liittyvät virheet

5.3. **merkkaus** (engl. *markup*) – käännösteksti on väärin merkattu ja se on aiheuttanut muotoiluvirheen

5.4. **puuttuva teksti** (engl. *missing text*) – käännöksessä löytyvä teksti on loppuviimeistelyn aikana kadonnut

5.5. **tekstin rajaaminen/laajeneminen** (engl. *truncation/text expansion*) – laajennettua käännöstekstiä on rajattu muotoilun rajoitusten syistä

6. **kulttuuriadaptaatio** (engl. *locale convention*) – käännöstekstin muoto ei vasta kulttuurikohtaisia käytäntöjä

6.1. **osoitteen muoto** (engl. *address format*)

6.2. **päivämäärän muoto** (engl. *date format*)

6.3. **valuutan muoto** (engl. *currency format*)

6.4. **mittayksiköiden muoto** (engl. *measurement format*)

6.5. **pikanäppäimet** (engl. *shortcut key*) – esim. ohjelmistoteksteissä

6.6. **puhelinnumeroiden muoto** (engl. *telephone format*)

6.7. **kulttuurikohtainen viite** (engl. *culture-specific reference*) – käännöksessä on käytetty kulttuurikohtaista viitettä, joka on kohderyhmälle epäselvä

7. **muu** (engl. *other*) – muut virhetyypit, jotka eivät sovi edellä mainittujen tyyppien alle. (TAUS 2022)

TAUSin DQF & QT21 -menetelmän mukaan voi tarvittaessa määritellä myös virheiden vakavuusasteita: kriittinen (engl. *critical*), suuri (engl. *major*), pieni (engl. *minor*), neutraali (engl. *neutral*). Lisäksi voi korostaa merkittäviä käännöksiä määräämällä vakavuusasteeksi erinomainen (*kudos*). (TAUS 2022)

2.5. Arviointimenetelmien vahvuudet ja heikkoudet

Konekäännöksen arviointimenetelmien avulla saatujen tulosten tulkitsemisessa tulisi huomioida näiden metodien vahvuudet ja heikkoudet. Kuten aikaisemmasta kävi ilmi, on historiallisesti ollut ensisijaista konekäännöksen ihmisarviointi. Ihmisen tekemän arvioinnin tärkeänä vaikuttajana on arvioija itse, koska hänen on täytettävä tietyt kriteerit, jotta arviointitulosten luotettavuus olisi taattu. Arvioijat voivat arviointityypin mukaan olla joko yksi- tai kaksikielisiä, ts. vain kohdekielen tai molempien, sekä lähtö- että kohdekielen äidinkielisiä tai melkein äidinkielisiä osaajia. (Chatzikoumi 2020: 146.) Ihmisarviointi taas vaatii paljon resursseja. Tämän puutteen ohella nousi esiin myös

ihmisarvioinnin subjektiivisuus, joka johtuu arvioijien vaihtelevasta kielitaitotasosta. Siitä seuraa vaihteleva arviointitulos. Sen takia suositaan usein automaattista arviointia, koska sitä pidetään pysyvämpänä, luotettavampana ja kustannustehokkaana. (Quah 2006: 136.)

Automaattisten menetelmien puutteena on se, että arviointi perustuu melko vähäiseen tietoon. Siinä huomioidaan lyhyitä sanarivejä ja esimerkiksi lauseiden tyyli, sujuvuus ja kielioppi jää täysin sivuun. Näin ollen voi täysin käsittämätön teksti, joka koostuu satunnaisista ja merkityksettömistä lauseista, saada hyvän arviointituloksen. (Poibeau 2017: 136.) Myös sanasanainen käännös saisi samasta syystä todennäköisemmin korkeamman arvosanan kuin vapaamuotoisempi käännös. (Kit, Wong 2015: 231.)

Automaattinen arviointi on tärkeää ensinnäkin konekääntimien kehittäjille, koska se mahdollistaa mitata järjestelmien toimivuutta ja kehitystä ajan myötä. Yleensä näiden tulokset korreloivat myös ihmisen yleisen käsityksen kanssa järjestelmän laadusta. Järjestelmä, joka saa ajan mittaan parempia arvosanoja, tarjoaa käännöksiä, joiden laatu todellakin ihmisarvioijien mielestä paranee. (Poibeau 2017: 138.) Automaattisiin mittareihin perustuva arviointi ei kuitenkaan kerro varmaa tietoa käännöslaadusta ja on tarkoitettu ensinnäkin järjestelmien järjestykseen asettamiseksi. Jos tarvitaan perusteellista ja informatiivista analyysia, ihmisen tekemä arviointi on jatkuvasti korvaamatonta. (Kit, Wong 2015: 232)

3. Konekäännöksen virheanalyysi

3.1. Analyysiaineiston kuvaus

Maisterintutkielmani analyysiaineisto on peräisin Viron kansallismuseon Kohtaamisia-näyttelyn teksteistä. Näyttelytekstien valinta johtuu monesta syystä: ne ovat julkisessa käytössä, niissä käsitellään eri aihepiirejä ja niihin löytyy suomenkieliset ihmiskäännökset. Konekäännöstutkimuksissa verrataan useimmiten konekäännöstä ihmisen tekemään käännökseen. Siksi aineistoksi oli valittava materiaalia, jonka myös ihminen on kääntänyt. Neljän tekstin perusteella luotu korpus koostuu 100 segmentistä. Segmenteiksi kutsun analysoitavia yksikköjä eli lauseita. Käytän tutkielmassani viitteitä T3, T10, T11 ja T14, jotka merkitsevät analyysiin käytettyjä tekstejä. Taulukosta 1 löytyy näiden tekstien kokonaispituudet sekä analysoitujen segmenttien määrä.

Tutkielmani korpuksen näyttelytekstien otsikot ovat:

Modernit ajat (T3)
Ihminen ja ympäristö (T10)
Oma sija (T11)
Laitamme ruokaa (T14)

Taulukko 1. **Analyysiaineiston tiedot.**

Tekstin viite	Tekstin aihepiiri	Segmenttien kokonaismäärä	Analysoitujen segmenttien määrä
T3	yhteiskuntaelämä	680	22
T10	ympäristö	620	24
T11	kotisisustus ja elinympäristö	171	27
T14	ruoka	324	27

Analysoitujen tekstien yhteinen piirre on se, että ne käsittelevät näyttelyteksteinä historiaa, jokainen omasta näkökulmastaan. Yhteiskuntaelämää kuvaavan tekstin T3 muotoilu on tutkimukseen valituista teksteistä kaikkein yleiskielisin. Teksteistä T10 ja

T14 taas löytyy monia erikoisalaan (puutarhanhoitoon ja ruoanlaittoon) liittyviä termejä. Teksti T11 eroaa muista teksteistä siten, että se sisältää historialliseen aikakauteen liittyvää sanastoa. Kaikkien tekstien kielellinen tyyli on kuvaava.

Analyysiaineisto koostuu myös ihmisen tekemistä käännöksistä. Vertailin vironkielisten lähtösegmenttien konekäännöstuotosta ihmiskäännöksiin, jonka on editoinut suomen kieltä äidinkielenä puhuva kielentarkistaja. En ollut tietoinen maisterintutkielmaan valittujen tekstien ihmiskääntäjien tarkoista taustatiedoista.

3.2. Tutkimusmenetelmä

Yhdistin konekäännöksen arvioinnissa kvalitatiivista ja kvantitatiivista menetelmää. Konekäännöksen automaattisten arviointimenetelmien käyttäminen edellyttää hyviä kieliteknologisia taitoja. Sen takia päätin käyttää kvantitatiivisessa analyysissa Levenshteinin editointietäisyyden laskelmaa, joka on helpommin sovellettava (AWSM 2022). Laskin LE:llä samankaltaisuuden ihmisen tekemän käännöksen kanssa. Kvalitatiivinen analyysi perustuu TAUSin DQF & QT21-virheluokitteluun. Olen mukauttanut TAUSin virheluokittelua valitsemalla tutkielmani kohteena olevat virhetyypit ja jättänyt muut pois. Samanlaista lähestymistapaa on käytetty myös muissa samantyyppisissä tutkimuksissa (esim. Gröhn 2019: 21–22). Päätyin tähän virheluokitteluun, koska halusin keskittyä analyysissa lingvistisiin kysymyksiin.

Tutkielmani virheluokittelu:

1. tarkkuus (poisto, väärä käännös, liikakääntäminen, alikääntäminen, kääntämätön);
2. sujuvuus (välimerkkien käyttö, oikeinkirjoitus, kielioppi);
3. tyyli (kömpelö, epäidiomaattinen);
4. muu.

Virhetyypit, jotka ovat jätetty pois tästä tutkimuksesta ja poisjättämisen perusteet:

- *tarkkuuden* luokassa *sopimaton tarkka käännösmuistivastine* (käännösmuistia ei ollut käytössä);

- *sujuvuuden* luokassa *epäjohdonmukaisuus* (analysoin lause- eikä tekstitasolla); *linkki/ristiviittaus*, *merkkien koodaus* (eivät olleet analyysin kannalta tärkeitä);
- *termistön* luokassa *ristiriitaisuus termipankin kanssa* (termipankkeja ei käytetty), *epäjohdonmukainen terminologian käyttö* (analysoin lause- eikä tekstitasolla);
- *tyylin* luokassa *yrityksen ja kolmannen osapuolen tyyli* (tyyliohjeita ei käytetty), *epäjohdonmukainen tyyli* (analysoin lause- eikä tekstitasolla);
- *muotoilun* ja *kulttuuriadaptaation* virhetyypit jäivät kokonaan pois, koska ne eivät olleet analyysin kannalta olennaisia.

En myöskään määrännyt tutkielmassasi virheiden vakavuusasteita, koska tavoitteena oli vain tunnistaa ja luokitella käännösvirheitä. Konekäännöksen virheiden vakavuusasteiden määrääminen on ilmeisesti tärkeämpää kaupallisille käyttäjille (esim. käännöstoimistoille).

Aloitin konekäännösanalyysin valitsemalla aineistosta satunnaisia segmenttejä, joihin sisältyi vähintään yksi käännösvirhe. Sen jälkeen etsin segmenteistä käännösvirheet ja luokittelin ne virhetyyppeihin. Sitten laskin segmenttien pituudet ja vertailin näitä havaittujen virheiden määrään. Tieto segmenttien pituudesta oli tärkeä, jotta olisi mahdollista tehdä johtopäätöksiä segmenttipituuksien ja virheiden määrän korrelaatiosta. Analyysin kvantitatiivisessa osuudessa laskin kolmella eri konekääntimellä saatujen käännösvastineiden LE:n. Tutkielmassani analysoidut lähtösegmentit ja näiden suomenkieliset kone- ja ihmiskäännökset löytyvät liitteestä 1.

Valitsin viro–suomi-kieliparin konekäännöksen virheanalyysia käsittelevää tutkielmani varten Tarton yliopiston kieliteknologian tutkimusryhmän kehittämän konekääntimen Tartu NLP:n. Tämä on monikielinen, 29 kieltä sisältävä yleisessä käytössä oleva neuroverkkokonekäännin, jonka työkielistä 23 on suomalais-ugrilaisia kieliä. Sitä on koulutettu kaikkien näiden 29 kielen käännösesimerkeillä. Pienempien suomalais-ugrilaisten kielten koulutusaineiston määrä on pienempi ja se vaikuttaa myös näiden kielien käännöslaatuun. (Tartu NLP.) Viron ja suomen aineiston sisältö on monipuolisempi ja sen saatavuus parempi, mutta toki ei yhtä hyvä kuin esimerkiksi englannin. Tartu NLP on silti ainutlaatuinen konekäännin, koska suurin osa näistä

pienistä vaarassa olevista kielistä eivät olleet aiemmin saatavilla julkisissa konekääntimissä. (Kolk 2023.)

Käytin konekäännösten toteuttamiseen käännösmuistiohjelmaa Memsource, jota varten on luotu Tartu NLP -liittymä. Konekääntimiä kehitetään jatkuvasti kouluttamalla näitä yhä uusilla käännösesimerkeillä, joten konekäännöstuotos voi ajan mittaan muuttua. Analysoitavat segmentit on konekäännetty ensimmäistä kertaa huhtikuussa 2021 ja sen jälkeen taas helmikuussa 2022. Eri vaiheessa kerätyt käännökset mahdollistavat sen tutkimisen, miten konekäännöstuotos on 10 kuukauden aikana muuttunut.

Kvalitatiivinen eli subjektiivinen virheanalyysi on tehty huhtikuussa 2021 konekäännetyistä segmenteistä. Tämän lisäksi tutkin konekäännöstuotosta kvantitatiivisella menetelmällä: otin Tartu NLP:n ohella käsiteltäväksi kahden neuroverkkokonekääntimen, Tilde NMT:n ja Google NMT:n konekäännökset. Analyysissa käytetyt konekääntimet perustuvat kaikki neuroverkkoihin ja ne eivät ole aihekohtaisia, vain yleisessä käytössä olevia. Sen takia ei ole mahdollista kuvata tarkemmin näiden konekääntimien koulutusaineiston sisältöä eikä määrää. Kvalitatiivinen virheanalyysi on hyvin aikavievää, joten tämän tutkielman laajuus rajoittaa kaikkien kolmen konekääntimen perusteellista tutkintaa. Kvalitatiivisen ja kvantitatiivisen analyysin yhdistäminen antaa monipuolisemman kuvan viro–suomi-kieliparin konekääntämisen nykytilanteesta.

Olennaista on nostaa esiin kielitaustani konekäännöksen tutkija ja arvioijana. Äidinkieleni on viro ja olen opiskellut kandidaattiopinnoissa pääaineena sekä viroa että suomea. Maisterivaiheessa keskityn suomen kieleen. Minulla on noin neljän vuoden kokemus suomi–viro-kieliparin asiatekstien kääntäjänä. Maisterintutkielmassasi analysoin viro–suomi-kieliparin käännöksiä, mutta virheanalyysissa on ollut tukenani suomea äidinkielenä puhuva ohjaajani, joten äidinkieleni ei olennaisesti ole vaikuttanut tutkimustuloksiani.

3.3. Virhetyyppien yhteenveto

Vironkielisten lähtösegmenttien pituus on keskimäärin 84,3 merkkiä: lyhyin segmentti 6 ja pisin 208 merkkiä. Lähtö- ja kohdesegmenttien pituudet on laskettu ilman välilyöntejä. Havaitsin 100 analysoidusta segmentistä yhteensä 333 käännösvirhettä.

Eniten (28 kpl) löytyi segmenttejä, jotka sisälsivät kaksi käännösvirhettä. Kaksi virhettä sisältävien lähtösegmenttien pituus on keskimäärin 64,9 merkkiä. Kolme käännösvirhettä sisältäviä segmenttejä oli 20 kpl (lähtösegmenttien keskipituus 87,1 merkkiä). 16 segmenttiä sisälsi neljä virhettä ja 14 segmenttiä yhden virheen. Näiden lähtösegmenttien keskipituus on vastaavasti 94,6 ja 61,6 merkkiä. Yli neljä virhettä sisältävien segmenttien määrä oli yhteensä 22 kpl ja niiden lähtösegmentit olivat pitempiä kuin 1–4 virhettä sisältävät segmentit (97,5–161 merkkiä). Yksi segmentti sisälsi keskimäärin 3,3 käännösvirhettä. Näin ollen yli puolet, 78 % segmenteistä sisälsi 1–4 käännösvirhettä. Lyhyemmät segmentit sisälsivät yleensä vähemmän ja pidemmät enemmän käännösvirheitä. Tarkemmat tiedot lähtösegmenttien keskipituuksien ja virheiden määrän jakautumisesta segmenteittäin löytyy taulukosta 2.

Taulukko 2. **Virheiden määrä segmenteittäin verrattuna lähtösegmenttien keskipituuksiin.**

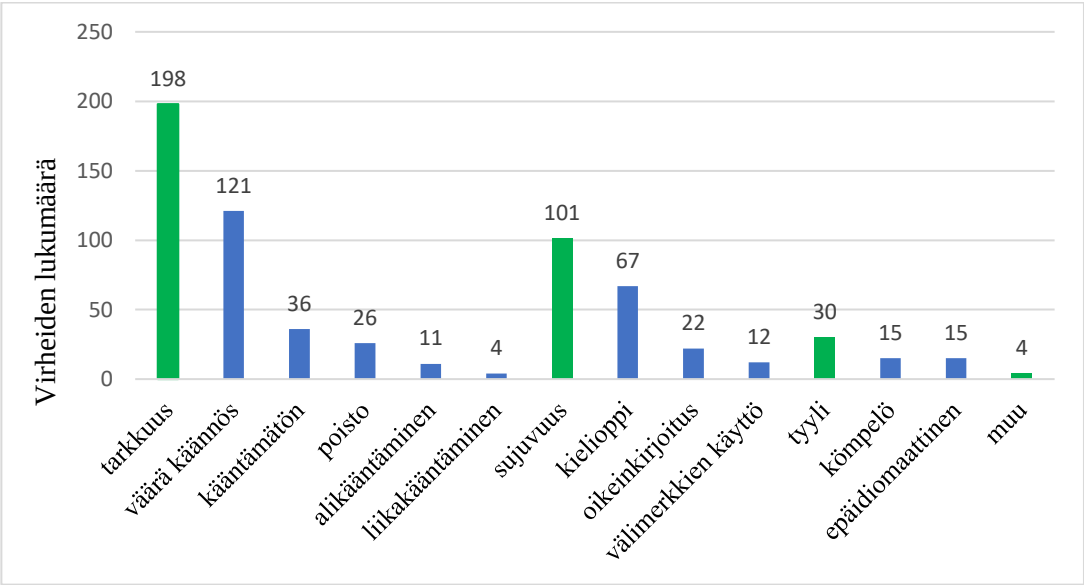
Segmenttien määrä (kpl)	Virheiden määrä (kpl)	Virheiden kokonaismäärä (kpl)	Lähtösegmenttien keskipituus (merkkiä)
14	1	14	61,6
28	2	56	64,9
20	3	60	87,1
16	4	64	94,6
8	5	40	97,5
6	6	36	106,5
4	7	28	142,3
3	8	24	125,3
1	11	11	161
yht. 100	kesk. 3,3	yht. 333	kesk. 84,3

Konekäännöksen kvalitatiivisessa analyysissa havaitut käännösvirheet sijoittuivat pääasiassa kolmeen päätyyppiin: tarkkuus, sujuvuus ja tyyli. Joissakin tapauksissa oli tarvetta käyttää luokittelua muu. Tarkkuuden, sujuvuuden ja tyylin virhetyypit jakautuivat taas alatyyppeihin: poisto, väärä käännös, liikakääntäminen, alikääntäminen, kääntämätön, välimerkkien käyttö, oikeinkirjoitus, kielioppi, kömpelö, epäidiomaattinen. Taulukossa 3 on esitetty virhetyypit pää- ja alatyypeittäin, näissä sisältyvien virheiden määrät sekä prosenttiosuudet. Suurin osa, 59,5 % kaikkien tekstien virheistä oli tarkkuusvirheitä. Toiseksi suurin virhetyyppi oli sujuvuus (30,3 %). Tyylivirheiden osuus oli 9 % ja kategorian muu alle sijoittui 1,2 % virheistä.

Taulukko 3. **Konekäännöksessä esiintyneiden virhetyyppien jako.**

Virhetyyppi	Virheiden määrä (kpl)				Virheiden kokonais-määrä (kpl)	Virheiden osuus (%)
	T3	T10	T11	T14		
tarkkuus	**46**	**51**	**52**	**49**	**198**	**59,5**
väärä käännös	29	23	45	24	121	36,3
kääntämätön	8	14	1	13	36	10,9
poisto	7	10	2	7	26	7,8
alikääntäminen	2	3	2	4	11	3,3
liikakääntäminen	-	1	2	1	4	1,2
sujuvuus	**17**	**20**	**31**	**33**	**101**	**30,3**
kielioppi	16	7	25	19	67	20,1
oikeinkirjoitus	1	9	2	10	22	6,6
välimerkkien käyttö	-	4	4	4	12	3,6
tyyli	**3**	**9**	**9**	**9**	**30**	**9**
kömpelö	2	5	6	2	15	4,5
epäidiomaattinen	1	4	3	7	15	4,5
muu	**1**	**1**	**1**	**1**	**4**	**1,2**
yhteensä	**67**	**81**	**93**	**92**	**333**	**100**

Virhetyyppien jakaumaa lukumäärittäin kuvaa piirros 1. Vihreällä on merkitty virhetyyppien pääluokat ja sinisellä näiden alaluokat.



Piirros 1. **Konekäännöksen virhetyyppien jakauma lukumääräisesti.**

3.3.1. Virheiden jakauma teksteittäin

Teksteittäin oli jossain määrin eroavaisuuksia virhejakaumassa, erityisesti virhetyyppien alaluokissa. Tarkkuusvirheitä oli kaikissa teksteissä suurin piirtein saman verran (46–52 kpl). Alatyyppiensä osalta muista teksteistä erottautui kotisisustusta ja elinympäristöä käsittelevä teksti T11, josta löysin huomattavasti enemmän vääriä käännöksiä (45 kpl) kuin muista (23–29 kpl) teksteistä. Käänteinen tilanne ilmeni kääntämättömyyksien tyypissä: näitä löytyi tekstistä T11 vain yksi, kun taas muista teksteistä 8–14 kpl. Myös poistoja löytyi tekstistä T11 vähemmän (2 kpl) kuin muista (7–10 kpl).

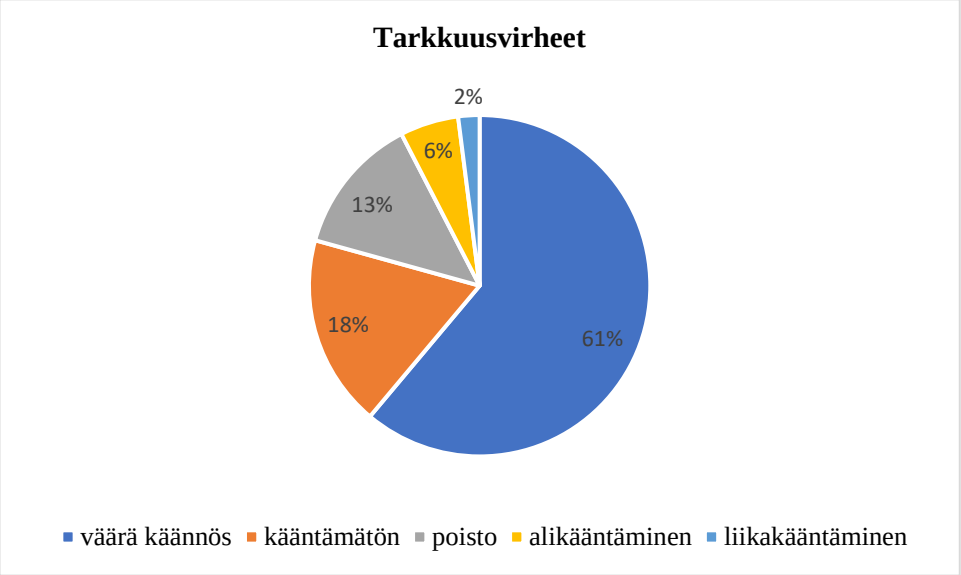
Sujuvuusvirheitä löytyi enemmän ruokaa käsittelevästä tekstistä T14 (33 kpl) ja tekstistä T11 (31 kpl). Kielioppivirheitä löytyi vähiten ympäristötekstistä T10 (7 kpl) kun taas muista 16–25 kpl. Sujuvuuden tyypissä erottui yhteiskuntaelämää käsittelevä teksti T3, josta löytyi vain yksi oikeinkirjoitusvirhe eikä välimerkkien virheellistä käyttöä. Muissa teksteissä esiintyi oikeinkirjoitusvirheitä 2–9 kpl ja välimerkkien virheellistä käyttöä 4 kpl. Tyylivirheitä löytyi vähiten tekstistä T3 (3 kpl), kun taas muista teksteistä 9 kpl. Väärää kohdekieltä eli virhetyyppiä muu esiintyi kaikissa teksteissä saman verran. Seuraavaksi käsittelen teksteissä havaittuja virhetyyppejä tarkemmin.

Esimerkkejä esitellessäni käytän seuraavia lyhenteitä:

LT = vironkielinen lähtöteksti
KK = konekäännös
IK = ihmisen tekemä käännös.

3.4. Tarkkuus

Kaikista virheistä (333 kpl) yli puolet (198 kpl) oli tarkkuusvirheitä. Ne jakautuivat viiteen alatyyppiin, joita kuvaa piirros 2. Isoimman alatyypin muodostivat väärät käännökset, joita oli 61 % tarkkuusvirheistä. 18 % tarkkuusvirheistä oli kääntämättömyyksiä, 13 % poistoja, 6 % alikääntäminen ja 2 % liikakääntäminen.



Piirros 2. **Tarkkuusvirheiden jakauma.**

3.4.1. Väärä käännös

Yli puolet (121 kpl) tarkkuuden tyyppiin kuuluvista virheistä oli vääriä käännöksiä, joista suurin osa (106 kpl) oli sananvalintavirheitä. Eniten löytyi sananvalintavirheitä tekstistä T11 (42 kpl), muista teksteistä suurin piirtein saman verran: 29 kpl (T3), 24 kpl (T14), 23 kpl (T10). Näitä lähemmin tutkiessani huomasin toistuvia virhemalleja. Havaitsin monesti, että tiettyjen sanojen lähtö- ja konekäännöstekstin kirjoitusasu on

samankaltainen. Tietyt sananvalintavirheet olivat toistuvia, tämä nousi esiin erityisesti vuosilukujen kääntämisessä. Seuraavaksi annan näistä tapauksista esimerkkejä taulukoissa 4 ja 5.

Taulukko 4. **Samankaltainen äänteellinen kirjoitusasu lähtötekstissä ja konekäännöksessä.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
sõjaväelennu*välju*	armeijan lento*väylistä*	sotilaslento*kentistä*	T10, 15
eestlaste	espanjalaisiin	virolaisten	T10, 47
argisemaid	arkistoista	arkisimpia	T14, 22
kulp	kulppu	kauha	T14, 63
kokale	kokaiinille	kokille	T14, 215

Taulukko 5. **Toistuvat sananvalintavirheet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Nii kaubeldi *17. sajandil* tulbisibulatega nagu väärtpaberitega, *18. sajandil* olid moes nelgid ning *19. sajandi* vaieldamatud lemmikud olid maikellukesed, mida kasvatati palju ka Raadil.	*1700-luvulla* oli myyty tulbisibulaa kuten arvopapereita, *1800-luvulla* moes nelgia ja *1900-luvun* kiistattomia suosikkeja oli maikellukesia, jotka kasvatettiin paljon myös Radiossa.	*1600-luvulla* käytiin kauppaa tulppaanisipuleilla aivan arvopaperien tapaan, *1700-luvulla* olivat muodissa neilikat ja *1800-luvun* kiistattomia suosikkeja olivat kielot, joita kasvatettiin paljon myös Raadilla.	T10, 34
mõisaaedades	mantereissa	kartanoiden puutarhoissa	T10, 61
hinnati	arvioitiin	arvostettiin	T11, 56

Lähes kaikki vuosisatoja sisältävät segmentit oli virheellisesti käännetty, ja näitä esiintyi yhteensä yhdeksän kertaa. Löytyi vain yksi tapaus, jossa vuosisata oli käännetty oikein (T10, 46). Sana *mõis* (suom. ’kartano’) oli kolmessa eri segmentissä konekäännetty vastineeksi *manner* ja sana *hindama* (suom. ’arvostaa’) merkityksessä ’pitää tärkeänä’ kahdessa segmentissä vastineeksi *arvioida*.

Huomasin samalla, että muutamassa segmentissä oli eri kontekstissa käytetty virheellisesti samaa käännösvastinetta. Konekäännöksessä oli esimerkiksi kahdessa eri segmentissä käytetty käännöstä *kääpiö*, toisessa sanan *voodilina* (suom. ’lakana’) (T11,

118) ja toisessa sanan *kärg* (suom. ’kenno’) (T10, 53) vastineena. Samanlainen tilanne oli sanan *keittiö* kanssa. Sitä oli käytetty sekä sanan *kukeseen* (suom. ’kantarelli’) (T14, 242) että *supp* (suom. ’keitto’) (T14, 68) vastineena.

Toistuvien sananvalintavirheiden ohella nostan esiin teksteissä monessa eri segmentissä esiintyviä sanoja, jotka oli äsken mainitusta virhemallista poiketen konekäännetty eri tavalla. Esimerkiksi sana *kulp* (suom. ’kauha’) oli konekäännetty vastineeksi *kulma* (T14, 21), *kulppu* (T14, 63) ja *kulp* (T14, 22).

Konekäännösteksteistä löytyi myös sananvalintavirheitä, joiden kohdalla kone oli tulkinnut lähtötekstiä väärin ja yhdistänyt alkusanan väärään käännösvastineen. Taulukossa 6 näkyy muutama esimerkki.

Taulukko 6. **Lähtötekstin sanantulkintavirheet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
õnnestus *kaitsepookimine* rõugete vastu	onnistuttiin *puolustamaan* rõugeita	onnistui isorokko*rokotus*	T10, 12
õlgedest (suom. *olki-oljen*) ja puidust valmistatud õlgtarud	*olkapäästä* (suom. *olka-olan*) ja puusta valmistetut õlat	*oljista* ja puusta valmistettuja olkipesiä	T10, 55
*part*epealne	*ankan* yläosa	*parsien* ylinen	T11, 21

Erityisen kiinnostava oli seurata, miten konekäännin toimii tapauksissa, joissa lähtötekstin sana on sille tuntematon. Näissä tilanteissa konekäännin käänsi sanan joko osittain, jätti jonkun osan sanasta kääntämättä tai tarjosi täysin tuntemattoman vastineen, jota ei kohdekielessä ole olemassakaan (taulukko 7). Huomasin, että tämä koski pääasiassa yhdyssanoja.

Taulukko 7. **Tuntemattomien sanojen kääntäminen.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
viigimarju	viikunmarjoja	viikunoita	T10, 24
kokakunst	kokataito	ruoanlaittotaide	T14, 15
peiupoistega	peukkupojien	sulhaspoikien	T14, 102
vana-aastaõhtul	vanha-iltana	uudenvuodenaattoiltana	T14, 138
serviise	serererejä	astiastoja	T11, 118
poisipea	poispää	poikatukka	T3, 184
puudritoos	puuritos	puuterirasia	T3, 212

3.4.2. Kääntämätön

Kääntämättömiä sanoja löytyi 36 kpl, eniten ympäristöä (T10, 14 kpl) ja ruokaa (T14, 13 kpl) käsittelevistä teksteistä (taulukko 8). Monissa tapauksissa oli kääntämättömään sanaan lisätty suomen kielen mallin mukainen sijapääte tai monikon tunnus.

Taulukko 8. **Kääntämättömät erikoisalan sanat.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
rõugete	rõugeita	isorokkorokotus	T10, 12
tulbisibulatega	tulbisibulaa	tulppaanisipuleilla	T10, 34
nelgid	nelgia	neilikat	T10, 34
pakktarudest	pakkitaruista	onkalopesistä	T10, 53
maikellukesed	maikellukesia	kielot	T10, 34

Ruokaa kuvaavasta tekstistä löytyi kulttuuriviittauksia, jotka oli jätetty lähtökieliseksi eli kääntämättä, vaikka osa ruoista on tuttuja Suomessakin (taulukko 9).

Taulukko 9. **Kääntämättömät kulttuuriviittaukset.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
räim	räim	silakat	T14, 9
verivorst	verivorst	verimakkaroista	T14, 9
rosoljekauss	rosoljekauss	rosollikulho	T14, 9
mulgipuder	mulgipuder	peruna-ohraryynipuuro	T14, 234
mulgi kapsad	mulgikapsat	sianliha-ohrasuurimohapankaali	T14, 234

Teksteistä löytyi myös yhdyssanoja, joiden jokin yhdysosa tai koko yhdyssana oli kääntämättä. Esimerkkejä näistä tapauksista on annettu taulukossa 10.

Taulukko 10. **Kääntämättömät yhdyssanat.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
köögiviljap*eenra*	vihannes*peenra*	vihannes*penkkiä*	T10, 28
ühepaja*toidud*	yksipaja*toidut*	patoja	T14, 68
*rukki*leib	*rukki*leipä	*ruis*leipä	T14, 234
tipp*restoranidele*	huippu*restoraanien*	huippu*ravintoloiden*	T14, 237
*vabriku*tööliste	*vabriku*työntekijöiden	*tehdas*työläisten	T3, 16
kirikukellad	kirikukellat	kirkonkellot	T3, 436

3.4.3. Poisto

Poistoja löytyi 26 kpl, eniten tekstistä T10. Monista konekäännetyistä lauseista oli jätetty pois esimerkiksi korostavat tai tarkentavat sanat, useimmiten adverbit kuten *praegu* (suom. ’nyt’; T10, 9), *siinsamas* (suom. ’täällä’; T10, 15), *ka* (suom. ’myös’; T10, 36; T14, 110), *pigem* (suom. ’pikemminkin’; T14, 17), *omavahel* (suom. ’keskenään’; T14, 215), *päris* (suom. ’aivan’; T14, 215), *eest* (suom. ’edestä’; T3, 233), *sinna* (suom. ’sinne’; T3, 310).

Joistakin lauseista oli jäänyt pois sanoja tai fraaseja, jotka olivat sisällön kannalta olennaisia. Poisto teki lauseiden merkityksestä epätäsmällisen. Taulukossa 11 on annettu kolme esimerkkiä siitä.

Taulukko 11. **Olennaisten sanojen/fraasien poisto.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Neis õpetati mõisa kärnereid, kuidas *pookida* puid ning kasvatada ananasse, sparglit, viigimarju jne.	Niissä opetettiin mantereen kärryjä, kuinka puuta ja kasvattaa ananassia, spargliä, viikunmarjoja ja niin edelleen.	Niissä opetettiin kartanon puutarhureille, kuinka puita *vartetaan* tai kasvatetaan ananasta, parsaa, viikunoita ym.	T10, 24

19. sajandi keskpaigas oli *vana-aastaõhtul* laual sealiha kaalika, naeri ja kapsaga.	1900-luvun puolivälissä pöydällä oli sianlihaa kaalika, naura ja kapsa.	1800-luvun puolivälissä pöydällä oli *uudenvuodenaattoiltana* sianlihaa, lanttua, naurista ja kaalia.	T14, 140
Hiljemalt 19. sajandil hakati Eesti *mõisaaedades* kahjurite tapmiseks kemikaale kasutama.	Viimeistään 1900-luvulla Viron mantereissa käytettiin kemikaaleja tuholaisten tappamiseen.	Viimeistään 1800-luvulla Viron *kartanoiden puutarhoissa* alettiin käyttää kemikaaleja tuholaisten tappamiseen.	T10, 61

Poisto ei välttämättä aiheuta ymmärrettävyysongelmia. Ihmiskääntäjäkin voi joskus päättää, että jonkin sanan voisi jättää esimerkiksi lauseen muotoilun korjaamisen takia käännöksestä pois. Seuraavasta taulukosta 12 löytyy kaksi esimerkkiä, jossa konekäännetystä segmentistä on jäänyt pois yksi sana, mutta se ei vaikuta ymmärrettävyyteen.

Taulukko 12. **Poisto, joka ei vaikuta ymmärrettävyyteen.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Jõukaid *inimesi* kutsuti üles loobuma eraldi saalist, söögitoast ja kabinetist kui tarbetust luksusest.	Varakkaita kehotettiin luopumaan erillisestä salista, ruokasalista ja toimistosta tarpeettomana ylellisyydenä.	Varakkaita *ihmisiä* kehotettiin luopumaan erillisestä salista, ruokailuhuoneesta ja työhuoneesta tarpeettomana ylellisyytenä.	T11, 69
Pääsuke tegi küll seeria fotosid Tartu *linnast*, kuid ka nende umbes 400 foto seas kujutavad paljud linna aguleid ja vaesema rahva elu.	Pääsuke teki sarjan kuvia Tartusta, mutta myös niiden noin 400 valokuvassa monet kaupungin aguleista ja köyhimpien ihmisten elämästä.	Pääsuke otti kyllä sarjan valokuvia Tarton *kaupungista*, mutta myös nämä noin 400 valokuvaa kuvaavat enimmäkseen laitakaupungin ja köyhemmän kansan elämää.	T3, 277

Poisto voi myös muuttaa lauseen merkityksen täysin vääräksi (taulukko 13). Myös lauserakenne voi muuttua sen takia.

Taulukko 13. **Merkitykseen vaikuttava poisto.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Raadile rajati suurte kasvuhoonetega *aed* ning mõisa tubades oli võimalik imetleda Balti provintside suurimat eksootiliste taimede kollektsiooni.	Raadille rakennettiin suuria kasvihuoneita, ja mantereen huoneissa oli mahdollista ihailla Baltian maakuntien suurinta eksoottisten kasvien kokoelmaa.	Raadille perustettiin *puutarha* suurine kasvihuoneineen ja kartanon rakennuksissa saattoi ihailla Baltian provinssien suurinta eksoottisten kasvien kokoelmaa.	T10, 20
Kliima soojenemisega 19. sajandil aga õnnestus *aiandusega* järjest rohkem raha teenida.	Ilmaston lämpeneminen 1900-luvulla onnistui kuitenkin ansaitsemaan yhä enemmän rahaa.	Ilmaston lämmetessä 1800-luvulla onnistuttiin puutarhanhoidolla kuitenkin ansaitsemaan yhä enemmän..	T10, 31
Sõjajärgsetel aastatel valitses suur korteri*kitsikus*.	Sodanjälkeisenä vuonna oli hieno asunto.	Sodanjälkeisinä vuosina vallitsi suuri asunto*pula*.	T11, 87

3.4.4. Alikääntäminen

Alikäännökseksi luokittelin sanat ja fraasit, joiden konekäännösvastineen merkitys oli liian yleinen tai käännös puutteellinen. Joissakin tapauksissa voi tuntua, että käännösvirheen voisi sijoittaa myös virheluokkaan väärä käännös. Tein virheluokan valintapäätöksen sen perusteella, että jos käännösvastineen sanavartalo tai yleiskäsite täsmäsi ihmiskäännöksen kanssa, niin sijoitin tällaisen tapauksen alikääntämisen luokkaan. Totesin yhteensä 11 alikäännöstapausta. Taulukossa 14 on annettu kolme esimerkkiä liian yleisestä sananmuodosta, jossa lähtöteksti ja ihmiskäännös ovat täsmällisempiä kuin konekäännös.

Taulukko 14. **Liian yleinen sananmuoto.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
sea*praad*	sian*ruoat*	sian*liha*	T14, 234
kevad*suvisesele*	kevään	kevät*kesään*	T14, 100
majutus*kohad*	majoitukset	majoitus*paikat*	T11, 169

Puutteelliseksi arvioin käännökset, josta puuttui jokin oleellinen sana tai sananosa (taulukko 15).

Taulukko 15. **Puutteellinen sananmuoto.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
täidetavad lõhnavee*pudelid*	täytettävät hajuvedet	täytettäviä hajuvesi*pulloja*	T3, 212
sada ja *rohkem* aastat	sata vuotta	*yli* satakin vuotta	T3, 200

Tekstin T10 segmenttien 18, 29 ja 33 lähtösana oli abstrakti (*aiandus*, suom ’puutarhanhoito’), mutta konekäännösvastine konkreettinen (*puutarha*). Päinvastainen tilanne oli tapahtunut tekstin T14 segmentissä 138, jossa lähtösanan *perenaised* (suom. ’perheen emännät’) konekäännösvastine oli *perheet*. Molemmissa esimerkeissä oli kone tunnistanut oikean sanavartalon.

Yleiskäsitteen tunnistamisen käännösesimerkiksi voisi nostaa tekstin T11 segmentin 118, jossa lähtösana *naps* (suom. ’alkoholijuoma’) oli konekäännetty yleiskäsitteellä *juoma* ja ihmisen käännetyssä tekstissä *vodka*. Sana *juoma* voi tarkoittaa mitä tahansa juomaa, mutta vironkielinen *naps* nimenomaan vahvaa alkoholia.

3.4.5. Liikakääntäminen

Vähiten tunnistin liikakääntämistä, vain neljä käännöstapausta (taulukko 16). Tekstistä T3 ei löytynyt liikakääntämistä. Yleinen ongelma tässä virhetyypissä on lähtötekstiä täsmällisempi sananmuoto. Esimerkiksi tekstin T10 segmentissä 20 esiintyvä termi *provints* (suom. ’provinssi’) tarkoittaa epätarkempaa hallintoyksikköä kuin konekäännösvastine *maakunta*.

Seuraavassa esimerkissä (T14, 100) lähteteksti ilmaisee mahdollisuutta, mutta konekäännös antaa kuvan varmasta tiedosta. Tekstin T11 segmentissä 56 vironkielisen sanan *vabanema* merkitys on ’päästä eroon’ (ihmiskäännöksessä *vapautua*), mutta konekäännöksessä on käytetty sanaa *hävittää*, jolla on vahvempi merkitys (’tuhota’).

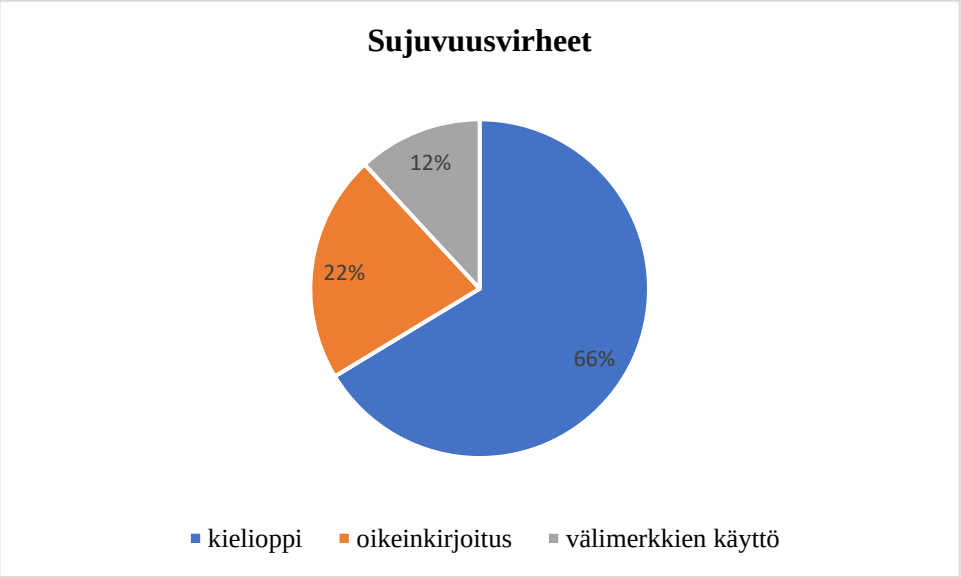
Neljännessä esimerkissä on lähtösanan *kabinet* (suom. ’työhuone’) konekäännösvastineena *toimisto*, jonka merkitys on täsmällisempi: viraston työhuoneet.

Taulukko 16. **Liikakäännetty sananmuoto.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Balti *provintside*	Baltian *maakuntien*	Baltian *provinssien*	T10, 20
pidu *sai pidada*	juhlat *pidettiin*	*voitiin* juhlia	T14, 100
soovitati *vabaneda* vanamoodsast tolmu koguvast mööblist	vanhanaikaiset pölyä keräävät huonekalut suositeltiin *hävittämään*	haluttiin *vapautua* vanhanaikaisista ja pölyä keräävistä huonekaluista	T11, 56
kabinetist	toimistosta	työhuoneesta	T11, 69

3.5. Sujuvuus

Toiseksi suurimman ryhmän muodostivat sujuvuusvirheet, joita esiintyi kaiken kaikkiaan 101 kpl. Eniten, 66 % sujuvuusvirheistä edustaa kielioppivirheitä. Oikeinkirjoitusvirheitä oli 22 % ja väärää välimerkkien käyttöä 12 %. Piirroksessa 2 on annettu sujuvuusvirheiden jakauma.



Piirros 2. **Sujuvuusvirheiden jakauma.**

3.5.1. Kielioppi

Kielioppivirheitä löytyi eniten tekstistä T11 (25 kpl) ja vähiten tekstistä T10 (7 kpl). Valtaosa kielioppivirheistä liittyi sijamuotojen käyttöön. Tässä virhetyypissä ilmeni säännönmukaisuus: jos konekäännetyssä lauseessa esiintyi tietty kielioppivirhe, esimerkiksi väärä sijamuoto, se esiintyi segmentin sisällä johdonmukaisesti. Ominaisena esimerkkinä tästä voi antaa tekstin T14 segmentin 7, jossa on käytetty virheellistä sijamuotoa: partitiivin sijaan nominatiivia. Tämän esimerkkisegmentin lähtötekstin nominatiivissa olevien korostettujen sanojen sijamuoto on konekäännöksessäkin nominatiivi. Siitä voisi päätellä, että kone ottaa mallia lähtötekstin sijamuodosta, mutta näin ei välttämättä ole. Sijamuotovirheistä nousi esiin juuri nominatiivin käyttö partitiivin sijaan (yli 20 tapausta) ja monesti myös lähtöteksti oli partitiivimuodossa (esimerkiksi T11, 56). Joissakin segmenteissä oli konekäännöksessä partitiivin sijasta genetiivi, inessiivi tai elatiivi. Esiintyi myös päinvastaisia käännöstapauksia, jossa kone oli käyttänyt nominatiivin sijasta partitiivia (T14, 17).

Virheellisen partitiivin käytön ohella erottui vielä nominatiivin käyttö genetiivin sijaan (T10, 23) ja illatiivin käyttö inessiivin sijaan (T11, 106). Edellä mainittujen tapausten ohella löytyi muitakin yksittäisiä sijamuotovirheitä: essiivin sijaan adessiivi, illatiivin sijaan allatiivi, inessiivin sijaan nominatiivi, nominatiivin sijaan genetiivi, elatiivin sijaan partitiivi ja adessiivin sijaan inessiivi. Esimerkkejä virheellisestä sijamuodon käytöstä on annettu taulukossa 17.

Taulukko 17. **Virheellinen sijamuodon käyttö.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Küsime, mis jääb püsima ja mis hääbub: on need *teadmised* ja *oskused*, *töövõtted* või *tehnoloogiad*, *retseptid* või *toorained*.	Kysymme, mitä pysyy ja mikä hämärtyy: ovatko nämä *osaaminen* ja *taidot*, *työskentely* tai *teknologia*, *reseptit* tai *raaka-aineet*.	Kysymme, mikä pysyy ja mikä katoaa. Ovatko ne *tietoja* ja *taitoja*, *työtapoja* vai *tekniikkaa*, *reseptejä* vai *raaka-aineita*.	T14, 7
Hinnati *valgusküllasust* ja *siledaid pindu*..	*Kirkkaus* ja *sileät pinnat* arvioitiin..	Arvostettiin *valoisuutta* ja *tasaisia pintoja*..	T11, 56

Kas need muutused on tehnoloogilised või muutuvad pigem *ideed* ja *toidueelistused*?	Ovatko nämä muutokset teknologisia vai muuttuvatko *ideoita* ja *ruokaetuja*?	Ovatko nämä muutokset teknisiä vai muuttuvatko pikemminkin *ideat* ja *ruokatottumukset*?	T14, 17
..18. sajandi lõpus ilmusid Tartus ka esimesed *eesti- ja lätikeelsed* aiandusraamatud.	..1800-luvun lopussa myös ensimmäiset *viro- ja latviakieliset* puutarhurikirjat ilmestyivät Tartussa.	..1700-lopulla Tartossa ilmestyivät myös ensimmäiset *viron- ja latviankieliset* puutarhanhoito-oppaat.	T10, 23
Argiellu ilmus juba arvukalt masinaid..	*Arjessa* on jo ilmestynyt lukuisia koneita..	*Arkielämään* ilmestyi myös jo paljon koneita..	T11, 106

Sijamuotovirheet eivät yleensä vaikuttaneet konekäännetyn tekstin ymmärrettävyyteen, mutta muutamassa tapauksessa aiheuttivat kuitenkin lauseen merkityksen muuttumisen. Taulukossa 18 on annettu tästä pari esimerkkiä. Tekstin T3 segmentissä 29 oli konekäännöksessä käytetty genetiivin sijasta nominatiivia ja myös possessiivisuffiksi puuttui. Lauseen alkuperäinen merkitys *ihmiset muuttivat nimensä* on muuttunut konekäännöksessä siten, että *nimi muutti ihmisiä*. Tekstin T10 segmentissä 36 taas on konekäännöksessä käytetty nominatiivin sijasta partitiivia, joten lähtötekstin merkitys *tuholaiseläimet tuhosivat kasveja*, on muuttunut konekäännetyssä lauseessa merkitykseksi *kylmät talvet ja sodat tuhosivat tuholaisia*.

Taulukko 18. **Sijamuotovirheistä johtuva merkitysmuutos.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
1930. aastate keskpaiku muutis *oma nime* ligi 210 000 inimest.	1930-luvun puolivälissä *nimi* muutti lähes 210 000 ihmistä.	1930-luvun puolivälissä *nimensä* muutti lähes 210 000 ihmistä.	T3, 29
Mitte ainult külmad talved ja sõjad ei hävitanud taimi, vaid ka *kahjurid*.	Kylmät talvet ja sodat eivät tuhonneet kasveja vaan *tuholaisia*.	Kasveja eivät tuhonneet pelkästään kylmät talvet ja sodat, vaan myös *tuholaiseläimet*.	T10, 36

Toistuvana kielioppivirheenä esiintyi aikamuotojen väärää käyttöä yhteensä viidessä segmentissä. Näistä neljä koski perfektin käyttöä oikean aikamuodon, pääasiassa imperfektin, mutta myös pluskvamperfektin sijasta. Yhdessä segmentissä on imperfektin

sijaan käytetty preesensiä. Taulukossa 19 on annettu kolme esimerkkiä virheellisestä aikamuodon käytöstä.

Taulukko 19. **Virheellinen aikamuodon käyttö.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Argiellu *ilmus* juba arvukalt masinaid..	Arjessa *on* jo *ilmestynyt* lukuisia koneita..	Arkielämään *ilmestyi* myös jo paljon koneita..	T11, 106
Eesti Vabariigis said naised põhiseadusega poliitilised õigused, mille eest sookaaslased mitmetes Euroopa riikides ning Ameerika Ühendriikides *olid võidelnud* sada ja rohkem aastat.	Viron tasavallassa naiset saivat perustuslailliset poliittiset oikeudet, joiden puolesta suomalaiset *ovat taistelleet* useissa Euroopan maissa ja Yhdysvalloissa sata vuotta.	Viron tasavallassa naiset saivat perustuslailla poliittiset oikeudet, joiden edestä naiset monissa Euroopan maissa ja USA:ssa *olivat taistelleet* yli satakin vuotta.	T3, 200
Suure osa Eesti merekuurortide külastajatest *moodustasid* venelased.	Venäläiset *muodostavat* suuren osan Viron merenkulkureittien vierailijoita.	Suuren osan Viron kylpyläkaupunkien vierailijoista *muodostivat* venäläiset.	T3, 301

Konekäännöksessä esiintyi toistuvasti myös aktiivimuotoa passiivin sijalla. Huomasin tällaisen virheen neljässä segmentissä, ja kaksi näistä näkyy esimerkkinä taulukossa 20. Tekstin T11 segmentti 147 erottui lisäksi siksi, että sijainnista (*lastenhuone*) on tullut tekijä.

Taulukko 20. **Aktiivimuodon käyttö passiivin sijalla.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Rand elavnes õhtupoolikul, kui sinna *tuldi* jalutama, muusikat kuulama ja päikeseloojangut imetlema.	Rand elvytti iltapäivällä, kun *he tulivat* kävelemään, kuuntelemaan musiikkia ja ihailemaan auringonlaskua.	Ranta vilkastui iltaisin, kun sinne *tultiin* kävelemään, kuuntelemaan musiikkia ja ihailemaan auringonlaskua.	T3, 310

Lastetoas *magatakse*, *mängitakse* ja *õpitakse*, aga kaasajal on see ka meelelahutus- ja meediakeskus.	Lastenhuone *nukkuu*, *leikkii* ja *oppii*, mutta nykyaikana se on myös viihde- ja mediakeskus.	Lastenhuoneessa *nukutaan*, *leikitään* ja *opiskellaan*, mutta se on nykyään myös viihde- ja mediakeskus.	T11, 147

Muutamassa segmentissä on konekäännöksessä käytetty monikkomuodon sijasta yksikköä (taulukko 21).

Taulukko 21. **Yksikön käyttö monikon sijalla.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
looduslikest materjalist *leludega*	luonnonmateriaalista *tehdyllä lelulla*	luonnonmateriaaleista *tehdyillä leluilla*	T11, 132
sõjajärgsetel aastatel	*sodanjälkeisenä vuonna*	*sodanjälkeisinä vuosina*	T11, 87

Tekstissä T10 segmenteissä 34 ja 55 esiintyi myös relatiivipronominin *joka* taivutusvirheitä: partitiivin sijaan on käytetty nominatiivia (taulukko 22). Kyseinen käännösongelma aiheutti segmentin 55 merkitysmuutoksen. Tämän lauseen alkuperäinen merkitys on, että *jyrsijät vahingoittivat olkipesiä*, mutta relatiivipronominin taivutusvirheen takia on merkitys muuttunut päinvastaiseksi (*olkipesät vahingoittivat jyrsijöitä*). Edellä mainittujen virhemallien lisäksi esiintyi muitakin yksittäisiä virheitä.

Taulukko 22. **Relatiivipronominin taivutusvirheet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
lemmikud olid maikellukesed, *mida* kasvatati palju ka Raadil	suosikkeja oli maikellukesia, *jotka* kasvatettiin paljon myös Radiossa	suosikkeja olivat kielot, *joita* kasvatettiin paljon myös Raadilla	T10, 34
õlgedest ja puidust valmistatud õlgtarud (2), *mida* aga tihti kahjustasid närilised	olkapäästä ja puusta valmistetut õlat (2), *jotka* kuitenkin usein vahingoittivat jyrsijöitä	oljista ja puusta valmistettuja olkipesiä (2), *joita* jyrsijät kuitenkin vahingoittivat usein	T10, 55

3.5.2. Oikeinkirjoitus

Oikeinkirjoitusvirheitä löytyi eniten teksteistä T14 (10 kpl) ja T10 (9 kpl), yhteensä 22 kpl. Ensiksi huomasin, että erisnimien suomentaminen on aiheuttanut käännösvirheitä. Teksteistä löytyi kolme paikannimeä, jotka oli kirjoitettu konekäännöksessä väärin. Nimeä *Tartu* esiintyi neljä ja *Raadi* kolme kertaa, *Põhjaka* yhden kerran. Erisnimiä oli teksteissä enemmänkin, mutta virheellisiä olivat juuri nämä. Kiinnostava seikka oli se, että nimi *Tartu*, jolle on olemassa laajalti levinnyt suomennos *Tartto*, oli jätetty lähtökieliseen muotoon *Tartu*, kun taas nimi *Raadi* oli tulkittu sanana *radio* isolla alkukirjaimellä. Nimen *Põhjaka* suomentaminen muotoon *Pohjoiska* on taas poikkeuksellista, koska toisaalta kone on yrittänyt kääntää nimen alkupuolta, toisaalta taas on nimen loppuun jätetty lähtökielisyys *-ka*. Erisnimien suomentamisesimerkit on annettu taulukossa 23.

Taulukko 23. **Erisnimien suomentamisvirheet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
kulp on ostetud ühest *Tartu* toidupoest	kulppu on ostettu yhdestä *Tartun* ruokakaupasta	kauha on ostettu eräästä *Tarton* ruokakaupasta	T14, 63
viiruste vastu peeti 18. sajandi lõpul oluline lahing just *Raadil*	viruksia vastaan pidettiin tärkeä taistelu 1800-luvun lopulla juuri *Radiossa*	viruksia vastaan pidettiin 1700-luvun lopulla olennainen taistelu juuri *Raadilla*	T10, 12
näiteks saabuvad värsked kukeseened suvel *Põhjakasse* postiga Lõuna-Eestist	esimerkiksi kesällä *Pohjoiskasse*-postilla Etelä-Eestistä saapuvat tuoreet keittiöt	esimerkiksi tuoreet kantarellit saapuvat kesällä *Põhjakan* kartanoon Etelä-Virosta	T14, 242

Virhemallina erottui vielä lauseen aloittaminen pienellä alkukirjaimellä (taulukko 24). Tämä ongelma esiintyi kolmessa segmentissä, joista kaksi (T10, 2 ja T10, 18) oli otsikkotekstejä. Yhteen segmenttiin (T10, 3) oli eksynyt englanninkielinen vastine (*estonian*), minkä voisi määritellä myös erisnimien suomentamisvirheiden alle, mutta yleisesti ottaen on joka tapauksessa kyse oikeinkirjoitusvirheestä.

Taulukko 24. **Lauseen pieni alkukirjain.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Inimene ja looduskeskkond	*ihminen* ja luonnonympäristö	*Ihminen* ja luonnonympäristö	T10, 2
Eesti looduskeskkond on kujunenud selliseks, nagu me seda praegu tunneme, pikaajalise inimmõju tulemusena.	*estonian* luonnonympäristö on muuttunut nykyisen kaltaiseksi pitkäaikaisen ihmisen vaikutuksen seurauksena.	*Viron* luonnonympäristö on muodostunut sellaiseksi, jona sen nyt tunnemme, ihmisen pitkäaikaisen vaikutuksen tuloksena.	T10, 3

Virhetyyppien määrittelyssä heräsi silloin tällöin kysymys, mihin tyyppiin tulisi jokin virhe luokitella, koska valintaperusteita oli monia. Kysymys ilmeni erityisesti oikeinkirjoitusvirheiden määrittelyssä. Tein niin, että jos konekäännösvastineen kirjoitusasu on ihmiskäännöksen tapainen, sitten luokittelin sen oikeinkirjoitusvirheiden alle.

Loput oikeinkirjoitusvirheet (taulukko 25) koskivat konekäännösvastineiden väärää kirjoitusasua. Tämän virhetyypin yhteinen piirre olikin juuri se, että konekäännöksessä oli käytetty ihmiskäännöksen tapaista sanavartaloa, mutta sanojen oikeinkirjoituksessa esiintyi virheitä.

Taulukko 25. **Väärä kirjoitusasu.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Kokkadel on peres ja kogukonnas oluline roll.	*Kokkailla* on tärkeä rooli perheessä ja yhteisössä.	*Kokeilla* on perheissä ja yhteisöissä tärkeä asema.	T14, 11
kabinetist kui tarbetust *luksusest*	toimistosta tarpeettomana *ylellisyydenä*	työhuoneesta tarpeettomana *ylellisyytenä*	T11, 69
„*koduseid*“ kaupu või teenuseid	"*kodiisia*" tavaroita tai palveluita	"*kodikkaita*" palveluita ja tuotteita	T11, 169

3.5.3. Välimerkkien käyttö

Havaitsin yhteensä 12 välimerkkivirhettä: kaikissa teksteissä saman verran paitsi tekstissä T3, josta ei löytynyt vastaavia virheitä. Valtaosa välimerkkivirheistä koski ajatusviivan virheellistä käyttöä. Konekäännöksessä on näissä tapauksissa otettu mallia lähtökielisestä lauseesta, jossa sivulause on erotettu päälauseesta ajatusviivoilla. Yhden kerran esiintyi konekäännöksessä virheellisesti myös yhdysviiva. Ihmiskäännöksissä on eri lauseissa käytetty esimerkiksi kaksois- ja puolipistettä tai sanaa *eli* ja pilkkua. Jos lähtölauseessa on vain yksi ajatusviiva, sitä seuraa esimerkiksi luettelo, niin konekäännöksessäkin on käytetty ajatusviivaa, mutta ihmiskäännöksessä kaksoispistettä, muissa tapauksissa pilkkua. Parissa konekäännetyssä segmentissä (T10, 20; T11, 169) esiintyy ylimääräinen pilkku. Yhdessä segmentissä (T11, 123) on konekäännöksessä yhdysviivalla muodostettu lyhenteestä ja sitä seuraavasta sanasta yhdyssana, vaikka genetiivin sijapääte olisi pitänyt yhdistää lyhenteeseen kaksoispisteellä. Tämän virhetyypin yleinen havainto on, että konekäännös tuntuu ottavan välimerkkien käytössä mallia lähtötekstistä. Taulukossa 26 on annettu välimerkkivirheistä esimerkkejä.

Taulukko 26. **Välimerkkivirheet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Alustame kohast, kus me praegu *oleme – Raadist* ja selle *lähiümbrusest – ja* uurime, kuidas peegelduvad selle paiga lugudes laiema maailma lood.	Aloitamme paikasta, jossa *olemme – Radiosta* ja sen *lähiympäristöstä – ja* tutkimme, miten tämän paikan tarinat heijastuvat laajemman maailman tarinoissa.	Aloitamme paikasta, jossa nyt olemme *eli* Raadin alueesta ja sen *lähiympäristöstä, ja* tutkimme, miten tämän paikan historia heijastelee maailman tapahtumia laajemmin.	T10, 9
Igapäevane oli võitlus tubaste *parasiitidega – täide*, kirpude ja lutikatega.	Päivittäin oli taistelu tilavaa loisia *vastaan - täitä*, kirppuja ja vikoja vastaan.	Syöpäläisten kanssa taisteltiin *päivittäin: riesana* olivat täit, kirput ja luteet.	T11, 24

Raadile rajati suurte kasvuhoonetega *aed ning* mõisa tubades oli võimalik imetleda Balti provintside suurimat eksootiliste taimede kollektsiooni.	Raadille rakennettiin suuria *kasvihuoneita, ja* mantereen huoneissa oli mahdollista ihailla Baltian maakuntien suurinta eksoottisten kasvien kokoelmaa.	Raadille perustettiin puutarha suurine *kasvihuoneineen ja* kartanon rakennuksissa saattoi ihailla Baltian provinssien suurinta eksoottisten kasvien kokoelmaa.	T10, 20
Eriti hinnati Jugoslaavia, Rumeenia või Saksa *DV komplekte*.	Erityisesti arvioitiin Jugoslavian, Romanian tai *Saksan DV-pakkauksia*.	Erityisesti arvostettiin Jugoslavian, Romanian ja *DDR:n kalusteita*.	T11, 123

3.6. Tyyli

Tyylivirheitä löysin yhteensä 30 kpl, joista puolet luokittelin kömpelön ja puolet epäidiomaattisen tyylin alle. Tyylivirheitä oli hankalampaa selvittää lähinnä kahdesta syystä. Ensinnäkin näiden toteaminen vaatii analysoijalta korkeinta kielitaitoa eli kohdekielen osaamista äidinkielen tasolla. Toiseksi tyylivirheitä esiintyy todennäköisemmin pidemmissä segmenteissä, joista voi taas löytyä monia muitakin virhetyyppejä. Segmentistä voi sen takia tulla sen verran ongelmallinen, että sen tyyliä on monimutkaista arvioida. Analyysituloksista voi siksi jäädä vaikutelma ikään kuin konekääntimellä ei olisikaan erityisemmin hankaluuksia tyylin kanssa.

Myös tyylivirheiden luokittelussa ilmeni kiistanalaisia tapauksia, joiden kohdalla tietyt virheet olisi voinut sijoittaa eri tyyppeihin. Esimerkiksi jos lauseen merkityksestä tulee käännetyn segmentin sanajärjestyksen takia väärä, onko tällöin kyse tyyli- vai kielioppivirheestä. Samanlainen kysymys herää myös kongruointivirheiden kohdalla.

3.6.1. Kömpelö tyyli

Tässä virhetyypissä nousi pääasiassa esiin seuraavat virhemallit: kömpelö sanajärjestys ja siihen liittyvät kongruointivirheet, sanan toisto, tyyliltään kehno sanavalinta ja sopimaton lauserakenne. Taulukossa 27 on annettu kömpelöstä tyylistä muutamaa esimerkki, ja seuraavaksi kuvailen näitä tapauksia tarkemmin.

Sanajärjestysongelmia oli kahta tyyppiä, josta toisessa oli kyse pelkästään kömpelöstä tyylistä, mutta toiseen liittyi väärinkäsityksiä aiheuttavia kongruointivirheitä. Esimerkiksi tekstin T11 segmentissä 57 sijaitsee verbi lauseen lopussa. Se ei vaikuta lauseen ymmärrettävyyteen, mutta tekee tyylistä kömpelön. Parempi olisi asettaa verbi lauseen alkuun kuten on tehty ihmiskäännöksessä.

Seuraavassa esimerkissä on kömpelö sanajärjestys aiheuttanut kongruointivirheen. Tekstin T10 segmentissä 3 alkuperäinen merkitys *pitkäaikainen vaikutus* on konekäännöksen kehnon sanajärjestyksen takia muuttunut vääräksi tulkinnaksi *pitkäaikainen ihminen*. Kongruointivirheitä ilmeni myös segmenteissä, joiden lauserakenne oli sopimaton. Esimerkiksi tekstin T3 konekäännetyn segmentin 460 lauserakenne aiheuttaa sanajärjestysmuutoksen ja sen takia myös väärän kongruenssin. Alkuperäisestä merkityksestä *järjestettiin juhlia* on tullut käännös *järjestettiin tuloja*.

Kömpelöön tyyliin viittasi myös sanan toisto, jota esiintyi monessa käännössegmentissä. Konekäännöksessä toistui esimerkiksi sana (T11, 24) tai sanavartalo (lauseessa T11, 69: *erillinen sali – ruokasali*). Erikseen voi nostaa esiin tekstin T3 segmentin 15, jossa lähtötekstissä on käytetty sanoja *tehased* ja *vabrikud*, joiden merkitysero ei ole iso. Ihmiskääntäjä on ratkaissut asian käyttämällä käännöksessään vain sanaa *tehtaat* ja jättänyt *vabrik* -sanan pois. Tämä tapaus on ollut konekääntimelle ilmeisesti sekava, joten se on toistanut sanan *tehtaat*.

Taulukko 27. **Kömpelö tyyli.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Soositi heledaid tapeete ja tekstiile.	Valotaustakuvia ja tekstiilejä *suosittiin*.	*Suosittiin* vaaleita tapetteja ja tekstiilejä.	T11, 57
Eesti looduskeskkond on kujunenud selliseks, nagu me seda praegu tunneme, *pikaajalise inimmõju* tulemusena.	estonian luonnonympäristö on muuttunut nykyisen kaltaiseksi *pitkäaikaisen ihmisen vaikutuksen* seurauksena.	Viron luonnonympäristö on muodostunut sellaiseksi, jona sen nyt tunnemme, *ihmisen pitkäaikaisen vaikutuksen* tuloksena.	T10, 3

Laen tasuti pasunakoori heaks *korraldatud pidude* sissetulekutest.	Laina maksettiin juhlien *tuloista, jotka järjestettiin* pasunakooren hyväksi.	Laina maksettiin takaisin puhallinorkesterin hyväksi *järjestettyjen juhlien* pääsymaksuista.	T3, 460
Igapäevane oli võitlus tubaste parasiitidega – täide, kirpude ja lutikatega.	Päivittäin oli taistelu tilavaa loisia *vastaan* - täitä, kirppuja ja vikoja *vastaan*.	Syöpäläisten kanssa taisteltiin päivittäin: riesana olivat täit, kirput ja luteet.	T11, 24
1820.–1830. aastatel rajati Eestis esimesed *tehased* ja *vabrikud*.	Virossa perustettiin ensimmäiset *tehtaat* ja *tehtaat* vuosina 1820–1830.	Ensimmäiset *tehtaat* perustettiin 1820–1830-luvulla.	T3, 15
aiandusraamatud	puutarhurikirjat	puutarhanhoito-oppaat	T10, 23
üheaastase köögiviljapeenra	yhden vuoden vihannespeenra	yksivuotista vihannespenkkiä	T10, 28
elektriline õmblusmasin	sähkö ompelukone	sähkökäyttöinen ompelukone	T11, 106

3.6.2. Epäidiomaattinen tyyli

Idiomaattisella kääntämisellä tarkoitetaan ”kääntämisen tapa, jolla pyritään muodollisen vastaavuuden sijaan välittämään lähtötekstin merkitys luontevalla, ns. keskimääräisen kielenkäyttäjän yleiskielellä” (Tieteen termipankki 2023). Epäidiomaattisuuksista nousivat esiin suorat käännökset ja kohdekieleen sopimattomat sanavalinnat (taulukko 28), joita löytyi sekä sana- että fraasitasolta. Suorat käännökset vaikuttivat ilmeisesti osittain myös lauseiden ymmärrettävyyteen, mutta pääasiassa idiomaattisuuteen. Epäidiomaattisuuksia esiintyi eniten tekstissä T14 ja vähiten tekstissä T3.

Taulukko 28. **Suorat käännökset.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
kuidas vanaema *süüa tegi*	miten isoäiti *teki ruokaa*	miten isoäiti *laittoi ruokaa*	T14, 110
kui pidu sai pidada nii *toas* kui õues	kun juhlat pidettiin *huoneessa* ja ovessa	jolloin voitiin juhlia sekä *sisällä* että ulkona	T14, 100
kultuurviljapuude *kõrgaeg*	kulttuuriviljapuudet *korkea aika*	hedelmäpuutarhojen *huippuaikaa*	T10, 46

Sageli magati mitmekesi *ühes* voodis.	Usein sukeltajat nukkuivat *yhdessä* sängyssä.	Usein *samassa* vuoteessa nukkui monta henkilöä.	T11, 20

Kiinnostavana ilmiönä esiintyi konekäännöksessä vaikeuksia nollapersoonan kääntämisessä (taulukko 29). Ihmiskääntäjä on käyttänyt passiivimuotoa ilmeisesti sen takia, että se korostaa monikollista ihmistarkoitetta tekemisen taustalla, kun taas nollapersoona implikoi yhtä ihmistä (VISK § 1363). Tekstin T10 segmentin 28 konekäännösvastine antaa aktiivimuodon takia mahdollisuuden tulkita subjektina sanan *vihannespenkki*. Luokittelin tämänkaltaiset tapaukset epäidiomaattisen tyylin alle.

Taulukko 29. **Nollapersoonan kääntämisen vaikeudet.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Üheaastase köögiviljapeenra jaoks *ei olnud* isegi enda maad *vaja*.	Yhden vuoden vihannespeenra *ei tarvinnut* edes omaa maata.	Yksivuotista vihannespenkkiä varten *ei tarvittu* edes omaa maata.	T10, 28
Kliima soojenemisega 19. sajandil aga *õnnestus* aiandusega järjest rohkem raha *teenida*.	Ilmaston lämpeneminen 1900-luvulla *onnistui* kuitenkin *ansaitsemaan* yhä enemmän rahaa.	Ilmaston lämmetessä 1800-luvulla *onnistuttiin* puutarhanhoidolla kuitenkin *ansaitsemaan* yhä enemmän..	T10, 31

3.7. Muu

Virhetyyppiin muu sijoitin käännöstapaukset, joissa konekäännöksessä esiintyi väärä kohdekieli eli suomen kielen sijaan englanti (taulukko 30). Nämä virheet voivat johtua siitä, että kyseinen Tartu NLP konekäännin on monikielinen ja englanti on yksi sen työkielistä. Havaitsin parissa vastaavassa segmentissä, että vaikka kohdekieli onkin väärä, niin sanaan on lisätty suomen kielen mallin mukainen sijapääte (T10, 3; T3, 212). Erikoistapauksena voi nostaa esiin tekstin T14 segmentin 238, jossa konekäännin on tulkinnut englanninkieliseksi viron sanan *road* ja tarjonnut suomen kielellä sen käännösvastineeksi *tie*, vaikka oikea tulkinta olisi suomeksi *ruoat*.

Taulukko 30. **Väärä kohdekieli.**

Lähtöteksti	Konekäännös	Ihmiskäännös	Viite
Kodus olid *daami* tualettlaual Lorupi vabrikus valmistatud puudritoos ja täidetavad lõhnaveepudelid.	Kotona oli *Ladyn* vessapöydällä Lorupin tehtaalla valmistettu puuritos ja täytettävät hajuvedet.	Kotona meikkauspöydällä oli Lorupin tehtaalla valmistettu puuterirasia ja täytettäviä hajuvesipulloja.	T3, 212
Road valmistatakse valdavalt algusest lõpuni kohapeal	*Tie* valmistetaan enimmäkseen alusta loppuun paikan päällä	*Ruoat* valmistetaan enimmäkseen alusta asti paikan päällä	T14, 238

3.8. Konekäännöksen ja ihmiskäännöksen samankaltaisuus

Aloitin kvantitatiivisen analyysin mittaamalla ensin ihmisen tekemien käännösten keskipituuden. Lyhyin ihmisen toimesta käännetty segmentti on 6 ja pisin 247 sekä keskimääräinen 96,3 merkkiä. Laskin Levenshteinin editointietäisyyden prinsiipillä: montako merkkiä (ilman välilyöntejä) tulisi konekäännetyssä segmentissä muuttaa, jotta se olisi yhdenvertainen ihmiskäännökseen. Laskin LE:n sekä lähtö- ja kohdetekstien pituuksien avulla myös konekäännöksen prosentuaalisen samanlaisuuden ihmiskäännökseen. (AWSM 2022) Yksinkertaistetusti sanottuna, mitä lyhyempi on LE, sitä lähempänä on konekäännös ihmisen tekemään käännökseen.

Vertailin keskenään vuosina 2021 ja 2022 Tartu NLP -konekääntimellä tehtyjen käännösten LE:tä. Vertailun tavoitteena oli selvittää, minkä verran saman konekääntimen tuotos on noin vuodessa muuttunut. Huhtikuussa 2021 konekäännettyjen segmenttien keskimääräinen LE on 45,3 ja 2022 helmikuussa 43,1 merkkiä. Taulukosta 31 näkyy Tartu NLP:n konekäännöstuotoksen LE teksteittäin.

Keskimääräinen samankaltaisuus ihmiskäännöksen kanssa (taulukko 32) on vastaavasti 58,6 % (2021) ja 60,8 % (2022). Tekstierittelystä käy ilmi, että isoin muutos koskee tekstejä T11 ja T10, jonka konekäännöstuotos on vuodessa muuttunut vastaavasti 8,6 ja 4 prosenttiyksikön verran samankaltaisemmaksi ihmiskäännöksen kanssa. Tekstin T14 konekäännöstuotos poikkeaa ihmiskäännöksestä vuoden 2021 käännökseen verrattuna 3,1 prosenttiyksikön verran. Tuloksista päätellen on Tartu NLP:n konekäännöstuotos

vuodessa muuttunut keskimäärin 2,2 prosenttiyksikön verran samankaltaisemmaksi ihmiskäännöksen kanssa.

Taulukko 31. **Tartu NLP:n konekäännöstuotoksen LE teksteittäin vuosina 2021 ja 2022.**

Teksti	Ihmiskäännöksen keskipituus (merkkiä)	Keskimääräinen LE (merkkiä)	
		2021	**2022**
T10	105	49,4	46,1
T14	98,4	41,9	43,8
T11	94	50,7	41,6
T3	87	39,1	40,7
kesk. yhteensä	**96,1**	**45,3**	**43,1**

Taulukko 32. **Tartu NLP:n konekäännöstuotoksen samankaltaisuus ihmiskäännöksen kanssa teksteittäin vuosina 2021 ja 2022.**

Teksti	Keskimääräinen samankaltaisuus ihmiskäännöksen kanssa (%)		Samankaltaisuuden muutos (%-yksikkö)
	2021	**2022**	
T10	59,9	63,9	4
T14	60,9	57,8	-3,1
T11	52	60,6	8,6
T3	61,6	60,8	-0,8
kesk. yhteensä	**58,6**	**60,8**	**2,2**

Seuraavaksi vertailin samoja tietoja Tilde NMT:n ja Google NMT:n konekäännöstuotoksiin. Tilde NMT:llä konekäännettyjen segmenttien keskimääräinen LE on 39,9 ja Google NMT:llä käännettyjen 46,5 merkkiä. Taulukossa 33 on vertailevasti Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotosten LE teksteittäin.

Taulukko 33. **Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotosten LE teksteittäin.**

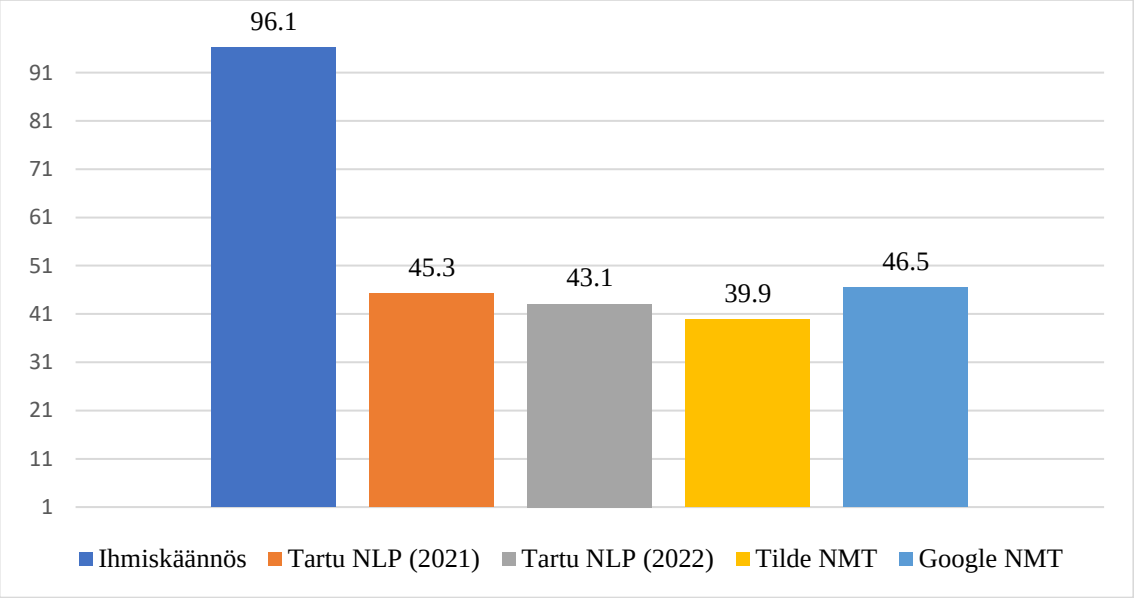
Teksti	Keskimääräinen LE (merkkiä)			
	Tartu NLP (2021)	**Tartu NLP (2022)**	**Tilde NMT**	**Google NMT**
T10	49,4	46,1	48,4	56,1
T14	41,9	43,8	37	42,7
T11	50,7	41,6	39,6	49,3
T3	39,1	40,7	34,4	37,9
kesk. yhteensä	**45,3**	**43,1**	**39,9**	**46,5**

Keskimääräinen samankaltaisuus ihmiskäännökseen (taulukko 34) on vastaavasti 64 % (Tilde NMT) ja 58 % (Google NMT). Tekstierittelystä ilmenee, että Tilde NMT:n konekäännöstuotos on tekstiä T10 lukuun ottamatta samankaltaisin ihmiskäännöksen kanssa. Tekstin T10 käännöksistä samankaltaisin ihmiskäännökseen on Tartu NLP 2022 konekäännöstuotos.

Taulukko 34. **Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotoksen samankaltaisuus ihmiskäännöksen kanssa teksteittäin.**

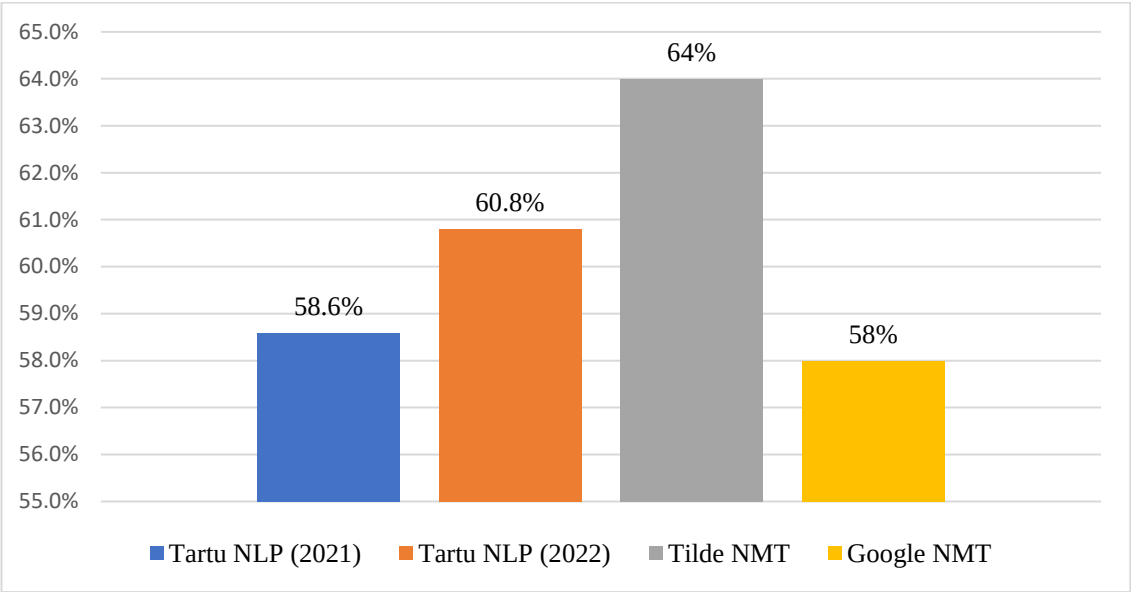
Teksti	Keskimääräinen samankaltaisuus ihmiskäännöksen kanssa (%)			
	Tartu NLP (2021)	**Tartu NLP (2022)**	**Tilde NMT**	**Google NMT**
T10	59,9	63,9	61,7	54,8
T14	60,9	57,8	64,8	61,9
T11	52	60,6	62,1	51,7
T3	61,6	60,8	67,2	63,5
kesk. yhteensä	**58,6**	**60,8**	**64**	**58**

Piirroksessa 3 on ihmiskäännöksen keskipituus verrattuna eri konekääntimien käännöstuotosten keskimääräiseen LE:en. Siitä voi päätellä, että tässä tutkimuksessa mukana olevista konekäännöstuotoksista samankaltaisin ihmiskäännöksen kanssa on Tilde NMT, jonka LE on 39,9 merkkiä. Google NMT eroaa ihmiskäännöksestä eniten, sen LE on 46,5 merkkiä. Tartu NLP:n 2021 ja 2022 konekäännöstuotokset jäävät Tilde NMT:n ja Google NMT:n väliin.



Piirros 3. **Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotosten keskimääräinen LE merkeissä verrattuna ihmiskäännökseen.**

Keskimääräistä samankaltaisuutta ihmiskäännöksen kanssa kuvaa edellisten tietojen rinnalla piirros 4. Sen mukaan Tilde NMT:n samankaltaisuus ihmiskäännöksen kanssa on 64 %, Tartu NLP:n vastaavasti 60,8 % (2022) ja 58,6 % (2021) sekä Google NMT:n 58 %.



Piirros 4. **Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotosten keskimääräinen samankaltaisuus ihmiskäännöksen kanssa.**

Tiivistelmä

Tämän maisterintutkielman tavoitteena oli testata yleisessä käytössä olevien neuroverkkokonekääntimien toimivuutta viro–suomi-kieliparissa tutkimalla niiden käännöstuotosta. Tutkimuksen fokuksessa oli Tartu NLP konekäännöstuotoksen arviointi yhdistämällä kvalitatiivista ja kvantitatiivista menetelmää.

Konekääntämisen historia ulottuu 1950-luvulle asti ja se on ajan mittaan kehittynyt primitiivisistä suorakäännösjärjestelmistä monimutkaisiin syväoppimiseen perustuviin neuroverkkokonekääntimiin. Vaikka konekääntäminen on edennyt suurin askelin, on yhä syytä arvioida ja tutkia sen tuloksia. Konekäännöksen arvioinnissa on useimmiten vertauskohteena ihmisen tekemä käännös. Se ei välttämättä tarkoita, että ihmiskäännös on täysin viaton, vaan sitä käytetään arviointiprosessissa referenssinä.

Havaitsin konekääntämisen ongelmakohtia tutkiessani, että alkuvaiheen konekääntimien isoimmaksi ongelmaksi osoittautui sanojen kaikkien käyttökontekstien huomioon ottaminen eli kielen monimerkityksisyys. Suorakäännösjärjestelmät eivät tulleet toimeen polysemian kanssa. Rinnakkaiskorpuksiin perustuvien konekääntimien tuloksellisuus riippuu saatavissa olevasta datamäärästä. Konekääntimien kouluttamiseen käytettävät rinnakkaiskorpukset eivät useimmiten isosta koosta huolimatta ole riittäviä kattamaan harvinaisia tai monimutkaisia kieli-ilmiöitä. NMT:n ongelmallisuus johtuu siitä, että neuroverkkojen toimimista on hankala ymmärtää ja analysoida. Sen datan sisäinen kuvaus on vain numeerista eikä ihmiselle suoraan luettavaa.

Konekäännöksen ensisijaisessa arvioinnissa käytetään lähinnä automaattisia arviointimenetelmiä, koska ne vaativat vähemmän henkilöresurssia ja niiden tulokset selviävät nopeasti. Käytetyin näistä on sanaperustainen BLEU-menetelmä. Automaattisiin mittareihin perustuva arviointi ei kuitenkaan kerro varmaa tietoa käännöslaadusta ja on tarkoitettu ensinnäkin järjestelmien järjestykseen asettamiseksi. Siksi käytetään niiden ohella myös ihmisen tekemää kvalitatiivista arviointia. Vaikka ihmisarviointia pidetään subjektiivisena, se kuitenkin antaa konekäännöstuotoksesta

perusteellisempaa sisällöllistä tietoa. Konekäännöksen kvalitatiivista virheanalyysiä pidetään objektiivisempana arviointimenetelmänä, sillä sen tulokset ovat johdonmukaisempia.

Tämän tutkielman analyysiaineisto koostui 100 kone- ja ihmiskäännetystä segmentistä, jotka oli koottu neljästä näyttelytekstistä, joiden aihepiirit olivat: yhteiskuntaelämä, ympäristö, kotisisustus ja elinympäristö ja ruoka. Tutkielman kvalitatiivinen analyysi perustui TAUSin DQF & QT21-virheluokitteluun. Käännösvirheet jakautuivat neljään päävirhetyyppiin: tarkkuus, sujuvuus, tyyli ja muu, jotka jakautuivat alatyyppeihin, paitsi muun tyyppi. Virheanalyysin fokuksessa olivat vuonna 2021 Tartu NLP:llä konekäännetyt segmentit. Kvantitatiivisessa analyysissa laskin konekäännöstuotosten Levenshteinin editointietäisyyden. Vertailin Tartu NLP:n ohella myös Tilden NMT:n ja Google NMT:n konekäännöstuotosta. Lisäksi tutkin, miten Tartu NLP:n konekäännöstuotos on noin vuodessa muuttunut, sillä olin konekääntänyt analysoitavia segmenttejä kahtena vuotena (2021 ja 2022).

Havaitsin 100 analysoidusta segmentistä yhteensä 333 käännösvirhettä, joten yksi segmentti sisälsi keskimäärin 3,3 käännösvirhettä. 78 % segmenteistä sisälsi 1–4 käännösvirhettä. Vironkielisten lähtösegmenttien keskipituus oli 84,3 merkkiä. Enintään neljä virhettä sisältävien lähtösegmenttien keskipituus oli 61,6–94,6 merkkiä. Yli neljä virhettä sisältävät segmentit olivat pitempiä (97,5–161 merkkiä). Näin ollen lyhyemmät segmentit sisälsivät yleensä vähemmän ja pidemmät enemmän käännösvirheitä. 59,5 % kaikkien tekstien virheistä oli tarkkuusvirheitä. Toiseksi suurin virhetyyppi oli sujuvuus (30,3 %). Tyylivirheiden osuus oli 9 % ja muiden virheiden 1,2 %. Teksteittäin oli jonkun verran eroavaisuuksia virhejakaumassa, erityisesti alatyypeissä. Virhejakauma teksteittäin on annettu taustatietona.

Tarkkuusvirheet jakautuivat viiteen alatyyppiin: väärä käännös (61 %), kääntämätön (18 %), poisto (13 %), alikääntäminen (6 %) ja liikakääntäminen (2 %). Tuon esiin tärkeimpiä havaintoja. Vääristä käännöksistä suurin osa oli sananvalintavirheitä. Havaitsin virheiden luokittelussa tiettyjä toistuvia malleja. Esiintyi samankaltaista äänteellistä kirjoitusasua lähtötekstissä ja konekäännöksessä sekä toistuvia sananvalintavirheitä, esimerkiksi vuosilukujen kääntämisessä. Tuntemattomien sanojen

kääntäminen aiheutti konekääntimelle ongelmia, kuten voisikin olettaa. Konekäännin käänsi tuntemattoman sanan joko osittain, jätti jonkun osan sanasta kääntämättä tai tarjosi vastineen, jota ei kohdekielessä ole olemassakaan. Tämä koski pääasiassa yhdyssanoja. Kääntämättömyyksistä voi tuoda esiin erikoisalan sanoja ja kulttuuriviittauksia, jonka konekäännin oli jättänyt kääntämättä. Löytyi myös yhdyssanoja, joiden jokin yhdysosa tai koko yhdyssana oli kääntämättä. Monissa tapauksissa oli kääntämättömään sanaan lisätty suomen kielen mallin mukainen sijapääte tai monikon tunnus. Monista konekäännetyistä sanoista oli jätetty pois korostavat tai tarkentavat sanat, useimmiten adverbit. Löytyi myös lauseen sisällön kannalta olennaisia poistoja, sanoja tai fraaseja. Toisin sanoen, jotkut poistot eivät vaikuttaneet lauseen ymmärrettävyyteen, kun taas jotkut muuttivat merkityksen täysin vääräksi. Alikääntämisen alatyypissä esiintyi liian yleistä tai päinvastoin, puutteellista sananmuotoa. Kun taas liikakääntämisessä oli yleisenä ongelmana lähtötekstiä täsmällisempi sananmuoto.

Sujuvuusvirheet jakautuivat kolmeen alatyyppiin: kielioppi (66 %), oikeinkirjoitus (22 %) ja välimerkkien käyttö (12 %). Valtaosa kielioppivirheistä liittyi sijamuotojen käyttöön. Tässä alatyypissä ilmeni säännönmukaisuus: jos konekäännetyssä esiintyi tietty kielioppivirhe, esimerkiksi väärä sijamuoto, se esiintyi segmentin sisällä johdonmukaisesti. Sijamuotovirheistä nousi esiin nominatiivin käyttö partitiivin sijaan, mutta sen ohella löytyi muitakin sijamuotovirheitä. Havaitsin, että sijamuotovirheet eivät yleensä vaikuttaneet konekäännetyn tekstin ymmärrettävyyteen, mutta muutamassa tapauksessa aiheuttivat lauseen merkityksen muuttumisen. Toistuvana kielioppivirheenä esiintyi aikamuotojen väärää käyttöä, aktiivimuotoa passiivin sijalla sekä yksikön käyttöä monikon sijalla. Lisäksi voi tuoda esiin relatiivipronominin *joka* taivutusvirheet. Löytyi muitakin yksittäisiä kielioppivirheitä. Oikeinkirjoitusvirheistä huomasin, että erisnimien suomentaminen on aiheuttanut käännösvirheitä. Virhemallina erottui lisäksi lauseen aloittaminen pienellä alkukirjaimellä. Esiintyi myös konekäännösvastineiden väärää kirjoitusasua. Tämän virhetyypin yhteinen piirre olikin juuri se, että konekäännöksessä oli käytetty ihmiskäännöksen tapaista sanavartaloa, mutta sanojen oikeinkirjoituksessa esiintyi virheitä. Valtaosa välimerkkivirheistä koski ajatusviivan virheellistä käyttöä, jossa on siten otettu mallia lähtökielisestä lauseesta.

Tyylivirheistä puolet luokittelin kömpelön ja puolet epäidiomaattisen tyylin alle. Kömpelön tyylin alaluokassa nousi pääasiassa esiin virhemallit: kömpelö sanajärjestys ja siihen liittyvät kongruointivirheet, sanan toisto, kehno sanavalinta ja sopimaton lauserakenne. Epäidiomaattisuuksista löytyi suoria käännöksiä ja kohdekieleen sopimattomia sanavalintoja. Suorat käännökset vaikuttivat ilmeisesti osittain myös lauseiden ymmärrettävyyteen, mutta pääasiassa idiomaattisuuteen. Virhetyypin muu alle sijoitin käännöstapaukset, joissa konekäännöksessä esiintyi väärä kohdekieli eli suomen sijaan englanti. Joissakin tapauksissa oli näihin englanninkielisiin sanoihin lisätty suomen kielen mallin mukainen sijapääte.

Kvantitatiivisessa arvioinnissa laskin kolmen eri konekääntimen, Tartu NLP:n, Tilde NMT:n ja Google NMT:n konekäännöstuotoksen Levenshteinin editointietäisyyden ja prosentuaalisen samanlaisuuden ihmiskäännöksen kanssa. Aluksi vertailin keskenään vuosina 2021 ja 2022 Tartu NLP -konekääntimellä tehtyjen käännösten LE:tä. Tartu NLP:n keskimääräinen LE oli 45,3 merkkiä (2021) ja 43,1 merkkiä (2022) ja keskimääräinen samankaltaisuus ihmiskäännöksen kanssa 58,6 % (2021) ja 60,8 % (2022). Tuloksista päätellen on Tartu NLP:n konekäännöstuotos noin vuodessa muuttunut keskimäärin 2,2 prosenttiyksikön verran samankaltaisemmaksi ihmiskäännöksen kanssa. Kolmen eri konekääntimen vastaavassa vertailussa ilmeni, että keskimääräinen LE oli 39,9 merkkiä (Tilde NMT) ja 46,5 merkkiä (Google NMT) ja keskimääräinen samankaltaisuus ihmiskäännöksen kanssa 64 % (Tilde NMT) ja 58 % (Google NMT). Levenshteinin etäisyyden mittaamisen perusteella voi päätellä, että tutkituista konekääntimistä tuottaa Tilde NMT käännöksiä, jotka muistuttavat eniten ihmiskäännöksiä ja Google NMT eroaa niistä eniten.

Kvantitatiivisesta arvioinnista ei kannata tehdä perustavanlaatuisia johtopäätöksiä konekäännösten laadusta. Todellisemman kuvan antaa kuitenkin kvalitatiivinen virheanalyysi. Tutkielmani on toivottavasti antanut katsauksen viro–suomi-konekääntämisen kielivivahteisiin, mutta konekääntimet tuottavat yhä uutta tutkimusaineistoa.

Lyhenteet

BLEU – engl. *Bilingual Evaluation Understudy*

DQF – engl. *Dynamic Quality Framework*

Google NMT – Googlen neuroverkkokonekäännin

IK – ihmisen tekemä käännös/ihmiskäännös

KK – konekäännös

LE – Levenshteinin editointietäisyys

LeBLEU – engl. *Letter-edit-BLEU*

LT – lähtöteksti

MQM – engl. *Multidimensional Quality Metrics*

NMT – neuroverkkokonekäännin (engl. *Neural Machine Translation*)

RBMT – sääntöpohjainen konekäännin (engl. *Rules Based Machine Translation*)

SMT – tilastollinen konekäännin (engl. *Statistical Machine Translation*)

Tartu NLP – Tarton yliopiston kieliteknologian tutkimusryhmän kehittämä neuroverkkokonekäännin

Tilde NMT – Tilden neuroverkkokonekäännin

TAUS – engl. *Translation Automation User Society*

WMT18 – engl. *Conference on Machine Translation*

Lähteet

Arnold et al. 1994 = Arnold, Doug, Lorna Balkan, Siety Meijer, R. Lee Humphreys, Louisa Sadler 1994. Machine Translation: An Introductory Guide, London: Blackwells-NCC.

AWSM 2022 = Levenshtein distance; https://awsm-tools.com/text/levenshtein-distance. Viitattu 16.3.2022.

Chatzikoumi, Eirini 2020. How to evaluate machine translation: A review of automated and human metrics. – Cambridge University Press 2019, Natural Language Engineering 26, s. 137–161; https://doi.org/10.1017/S1351324919000469. Viitattu 12.5.2023.

eTranslation 2022 = Euroopan Komission konekäännin; https://ec.europa.eu/digital-building-blocks/wikis/display/CEFDIGITAL/eTranslation. Viitattu 16.3.2022.

Fišel, Mark 2016. Milline on hea masintõlge? – Sirp, 11. maaliskuuta; https://www.sirp.ee/s1-artiklid/varia/milline-on-hea-masintolge/. Viitattu 25.4.2021.

Flanagan, Mary 1994. Error Classification for MT Evaluation. – Proceedings of the 1st Conference of the Association for Machine Translation in the Americas: Technology Partnerships for Crossing the Language Barrier (AMTA-94), Columbia, Maryland, USA: 65−72.

Fomicheva ym. 2020 = Fomicheva, Marina, Shuo Sun, Lisa Yankovskaya, Frédéric Blain, Francisco Guzmán, Mark Fišel, Nikolaos Aletras, Vishrav Chaudhary, Lucia Specia 2020. Unsupervised Quality Estimation for Neural Machine Translation. – Transactions of the Association for Computational Linguistics 8, s. 539–555; https://aclanthology.org/2020.tacl-1.35.pdf. Viitattu 30.4.2021.

Forcada, Mikel L. 2017. Making sense of neural machine translation. Translation Spaces 6:2 (2017), s. 291–309; https://benjamins.com/catalog/getpdf?webfile=a996243603. Viitattu 8.5.2023.

Garcia, Ignacio 2015. Computer-aided translation: systems. – The Routledge Encyclopedia of Translation Technology. Ed. by Chan Sin-wai. London and New

York: Routledge Taylor & Francis Group, s. 72; http://dspace.vnbrims.org:13000/xmlui/bitstream/handle/123456789/4402/Routledge%20Encyclopedia%20of%20Translation%20Technology.pdf?sequence=1&isAllowed=y. Viitattu 16.3.2022.

Google NMT = Googlen kehittämä neuroverkkokonekäännin; https://translate.google.ee/?hl=et. Viitattu Viitattu 16.3.2022.

Gröhn, Ari 2019. Suitability of Neural Machine Translation for Different Types of Texts. A study on potential predictors. Pro Gradu Thesis. University of Helsinki. https://helda.helsinki.fi/bitstream/handle/10138/301912/Grohn_Ari_Pro_gradu_2019.pdf?isAllowed=y&sequence=2. Viitattu 1.2.2022.

Görög, Attila 2014. Dynamic Quality Framework: quantifying and benchmarking quality. – Tradumàtica: tecnologies de la traducció, No. 12, s. 453; https://revistes.uab.cat/tradumatica/article/view/n12-gorog2/pdf. Viitattu 1.2.2022.

IATE 2023 = Interactive Terminology for Europe; https://iate.europa.eu/. Viitattu 12.5.2023.

Kaalep, Heiki-Jaan, Mare Koit 2010. Kuidas masin tõlgib. – Keel ja Kirjandus, nro 10, s. 726–738; https://keeljakirjandus.eki.ee/726-738.pdf. Viitattu 25.4.2021.

Kit, Chunyu, Billy Wong Tak-ming 2015. Evaluation in machine translation and computer-aided translation – The Routledge Encyclopedia of Translation Technology. Ed. by Chan Sin-wai. London and New York: Routledge Taylor & Francis Group; http://dspace.vnbrims.org:13000/xmlui/bitstream/handle/123456789/4402/Routledge%20Encyclopedia%20of%20Translation%20Technology.pdf?sequence=1&isAllowed=y. Viitattu 12.5.2023.

Kolk, Mariliis 2023. Tartu Ülikooli masintõlkemootori arsenalis on 23 soome-ugri keelt. – Postimees, 15. maaliskuuta; https://teadus.postimees.ee/7732337/tartu-ulikooli-masintolkemootori-arsenalis-on-23-soome-ugri-keelt. Viitattu 29.3.2023.

Koponen, Maarit 2010. Assessing Machine Translation Quality with Error Analysis. – MikaEL: Electronic proceedings of the KäTu symposium on translation and interpreting studies Vol. 4, Suomen kääntäjien ja tulkkien liitto, s. 1; https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf. Viitattu 16.3.2022.

Liin ym. 2012 = Liin, Krista, Kadri Muischnek, Kaili Müürisep, Kadri Vider 2012. Eesti keel digiajastul = The Estonian Language in the Digital Age. Toim. Georg Rehm, Hans Uszkoreit. Springer: White paper series = Valge raamatu sari, s. 21; http://www.digar.ee/id/nlib-digar:121285. Viitattu 25.4.2021.

Liu Yang, Min Zhang 2015. Statistical machine translation. – The Routledge Encyclopedia of Translation Technology. Ed. by Chan Sin-wai. London and New York: Routledge Taylor & Francis Group; s. 201; http://dspace.vnbrims.org:13000/xmlui/bitstream/handle/123456789/4402/Routledge%20Encyclopedia%20of%20Translation%20Technology.pdf?sequence=1&isAllowed=y. Viitattu 12.5.2023.

Lommel, Arle ym. 2015 = Lommel, Arle, Attila Görög, Alan Melby, Hans Uszkoreit, Aljoscha Burchardt, Maja Popović 2015. QT21. Deliverable 3.1. Harmonised Metric; http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf. Viitattu 1.2.2022.

Nieminen, Tommi 2020. Konekäännös suomalaisen kääntäjän näkökulmasta. Euroopan Komission #TranslatingEurope työpajan „Machine translation and the human translator“ esitelmä, Tampere, Suomi, 31.1.2020; https://puolukka.uta.fi/~textmine/events/tew-tampere/TNieminenTre.pdf, https://tuni.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=1e559295-a1dc-4fd8-9747-acc3005ebf51. Viitattu 26.4.2021.

Papineni ym. 2002 = Papineni, Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. – Philadelphia: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), s. 311–318; https://www.aclweb.org/anthology/P02-1040/. Viitattu 16.3.2022.

Poibeau, Thierry 2017. Machine Translation. The MIT Press. Cambridge, Massachusetts; London, England; https://ia802801.us.archive.org/9/items/ThierryPoibeauMachineTranslation/Thierry%20Poibeau_Machine%20Translation.pdf. Viitattu 2.5.2023.

Quah, Chiew Kin 2006. Translation and technology. Palgrave textbooks in translating and interpreting. Basingstoke, New York: Palgrave Macmillan.

Sibul, Ingrid 2018. Masintõlge ja eestikeelsed õigustekstid – ühe tõlkija senised kogemused Euroopa Komisjonis. – Õiguskeel nro 2, s. 5; https://www.just.ee/sites/www.just.ee/files/ingrid_sibul._masintolge_ja_eestikeelsed_oigustekstid_-_uhe_tolkija_senised_kogemused_euroopa_komisjonis.pdf. Viitattu 16.3.2022.

Schwarzl, Anja 2001. The (Im)Possibilities of Machine Translation, Frankfurt: Peter Lang Publishing.

Tartu NLP = Tarton yliopiston kieliteknologian tutkimusryhmän kehitetty konekäännin; https://www.neurotolge.ee/#info. Viitattu 16.3.2022.

TAUS 2022 = Harmonized DQF-MQM Error Typology; https://www.taus.net/qt21-project#harmonized-error-typology. Viitattu 1.2.2022.

TEPA-termipankki 2023 = Erikoisalojen sanastojen ja sanakirjojen kokoelma, Sanastokeskus; https://termipankki.fi/tepa/fi/. Viitattu 12.5.2023.

Tieteen termipankki 2023 = The Helsinki Term Bank for the Arts and Sciences, Käännöstieteen termistö; https://tieteentermipankki.fi/wiki/K%C3%A4%C3%A4nn%C3%B6stiede. Viitattu 12.5.2023.

Tilde NMT = Tilden kehittämä neuroverkkokonekäännin; https://translate.tilde.com/?utm_source=from_mt_page&utm_medium=inner_click#/. Viitattu 16.3.2022.

Torné, Anna, Anna Matamala 2021. Human evaluation of three machine translation systems: from quality to attitudes by professional translators. – Vigo International Journal of Applied Linguistics, No. 18, s. 97–121; https://revistas.webs.uvigo.es/index.php/vial/issue/view/208/78. Viitattu 16.3.2022.

Virpioja, Sami, Stig-Arne Grönroos 2015. LeBLEU: N-gram-based Translation Evaluation Score for Morphologically Complex Languages. – Proceedings of the Tenth Workshop on Statistical Machine Translation. Lisbon: Association for

Computational Linguistics, s. 412; https://www.aclweb.org/anthology/W15-3052/. Viitattu 16.3.2022.

VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004. Iso suomen kielioppi. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio; http://scripta.kotus.fi/visk URN:ISBN:978-952-5446-35-7 Viitattu 16.5.2023.

Yu Shiwen, Bai Xiaojing 2015. Rule-based machine translation. – The Routledge Encyclopedia of Translation Technology. Ed. by Chan Sin-wai. London and New York: Routledge Taylor & Francis Group; http://dspace.vnbrims.org:13000/xmlui/bitstream/handle/123456789/4402/Routledge%20Encyclopedia%20of%20Translation%20Technology.pdf?sequence=1&isAllowed=y. Viitattu 12.5.2023.

Eesti-soome inim- ja masintõlke võrdlus ERMi näituse ”Kohtumised” tekstide põhjal. Resümee

Siinses magistritöös uuritakse masintõlke hindamist. Uurimuse eesmärk oli võrrelda eesti-soome keelesuuna masintõlkeväljundit inimese tehtud tõlkega ja selgitada välja, mis tüüpi vigu teevad üldkasutuses olevad neurotõlkesüsteemid. Selle taustal käsitleti ka eri masintõlkesüsteemide probleemkohti ja masintõlke hindamise võimalusi. Kuna masintõlge on üha rohkem osa meie igapäevaelust, on selle hindamine suure tähtsusega. Just väiksemate keelte puhul on masintõlke tulemuslikkust põhjalikumalt vähe uuritud.

Masintõlge on aja jooksul arenenud algsetest otsetõlkesüsteemidest keeruka sügavõppel põhineva neurotõlkeni välja. Esimeste masintõlkesüsteemide probleemiks kujunes asjaolu, et need ei tulnud toime mitmetähenduslikkusega. Paralleelkorpustel põhinevate süsteemide tulemuslikkus sõltus masintõlkemudeli koolitamiseks saada olevast andmemahust. Neurotõlke probleemsus on seotud sellega, et tehisnärvivõrkude toimimist on raske mõista ja analüüsida, sest andmed on vaid numbrilised.

Masintõlke eri hindamismeetodite võrdlemisel selgus, et automaatsete meetodite abil saadakse küll kiireid tulemusi, kuid need on mõeldud pigem süsteemide järjestamiseks ja nende arengu jälgimiseks. Masintõlkeväljundist sisulise ja põhjalikuma keelelise ülevaate saamiseks tuleks rakendada ikkagi kvalitatiivset analüüsi, mida teeb inimene. See on küll aeganõudvam ja nõuab rohkem ressurssi, kuid annab paremini aimu masintõlkeväljundi kvaliteedist. Nii kvalitatiivsete kui ka kvantitatiivsete hindamismeetodite puhul on masintõlkeväljundi võrdlusobjektiks inimtõlge.

Magistritöö sajast segmendist koosnev analüüsimaterjal pärineb eestikeelsetest näitusetekstidest, mis masintõlgiti kolme erineva neurotõlkemootori Tartu NLP, Tilde NMT ja Google NMT-ga. Uurimuses rakendati kombineeritult kvalitatiivset ja kvantitatiivset masintõlke hindamismeetodit: subjektiivset vigade analüüsi ja Levenshteini teisenduskauguse mõõtmist. Veaanalüüsi uurimisobjektiks olid 2021. a Tartu NLP-ga masintõlkemootoriga tehtud tõlked ja vealiigitus tugines TAUSi DQF &

QT21 klassifikatsioonil. Automaatse ja manuaalse hindamismeetodi kombineerimist on kasutatud ka varasemates analoogsetes uurimustes.

Subjektiivse veaanalüüsi käigus tuvastatud vead jagunesid nelja põhikategooriasse: täpsus (59,5%), soravus (30,3%), stiil (9%) ja muu (1,2%), mis jagunesid omakorda alamkategooriatesse. Üks segment sisaldas keskmiselt 3,3 tõlkeviga. Täpsusvead jagunesid viide alamkategooriasse: valetõlge (61%), tõlkimata (18%), väljajätt (13%), alatõlkimine (6%) ja ületõlkimine (2%). Soravusvead jagunesid kolme alamkategooriasse: grammatika (66%), õigekiri (22%) ja kirjavahemärgistus (12%). Pooled stiililivigadest paigutusid kohmaka ja pooled ebaidiomaatilise stiili alla.

Kvantitatiivse hindamise käigus arvutati kolme neurotõlkemootori väljundi keskmine Levenshteini teisenduskaugus (LE) ja protsentuaalne sarnasus inimtõlkega: 39,9 tähemärki (Tilde NMT) ja 46,5 tähemärki (Google NMT) ning keskmine sarnasus inimtõlkega 64% (Tilde NMT) ja 58% (Google NMT). Tartu NLP masintõlkeväljundi puhul oli lisaks võimalik võrrelda umbes üheaastase vahega (2021 ja 2022) tehtud sama lähteteksti tõlkeid. Tartu NLP väljundi keskmine LE oli 45,3 tähemärki (2021) ja 43,1 tähemärki (2022) ja keskmine sarnasus inimtõlkega 58,6% (2021) ja 60,8% (2022). Arvutustest selgus, et Tartu NLP masintõlkeväljund oli muutunud 2,2% võrra sarnasemaks inimtõlkega. Levenshteini teisenduskauguse mõõtmise põhjal võib järeldada, et uuritud masintõlkemootoritest loob Tilde NMT tõlkeid, mis on inimtõlkega kõige sarnasemad. Google NMT samas on inimtõlkest kõige erinevam. Kvantitatiivsete hindamistulemuste põhjal ei saa teha põhjapanevaid järeldusi masintõlgete kvaliteedi kohta. Loodetavasti on siinne uurimus andnud väärtuslikke andmeid eesti-soome masintõlkimise hetkeseisu kohta.

Liite 1: analyysimateriaalin virheluokittelu

Viite	Lähtöteksti	Konekäännös	Ihmiskäännös	Virhetyyppi
T10–2	***Inimene*** ja looduskeskkond	***ihminen*** ja luonnonympäristö	***Ihminen*** ja luonnonympäristö	sujuvuus – oikeinkirjoitus
T10–3	***Eesti*** looduskeskkond on kujunenud selliseks, nagu me seda praegu tunneme, ***pikaajalise inimmõju*** tulemusena.	***estonian*** luonnonympäristö on muuttunut nykyisen kaltaiseksi ***pitkäaikaisen ihmisen vaikutuksen*** seurauksena.	***Viron*** luonnonympäristö on muodostunut sellaiseksi, jona sen nyt tunnemme, ***ihmisen pitkäaikaisen vaikutuksen*** tuloksena.	sujuvuus – oikeinkirjoitus muu kömpelö tyyli
T10–6	Kui tänapäeval räägitakse keskkonnast, siis sageli seostub see globaalsete probleemidega – reostuse, kasvuhoonegaaside või kliimamuutusega, teisalt ka keskkonnakaitsega – looma- ja taimeliikide ning nende elupaikade hoidmisega.	Kun nykyään puhutaan ympäristöstä, se liittyy usein maailmanlaajuisiin ongelmiin – saastumiseen, kasvihuonekaasuihin tai ilmastonmuutokseen, toisaalta ympäristönsuojeluun – eläin- ja kasvilajien ja niiden luontotyyppien säilyttämiseen.	Kun nykyään puhutaan ympäristöstä, se liittyy usein maailmanlaajuisiin ongelmiin**:** saasteisiin, kasvihuonekaasuihin tai ilmaston muutokseen**;** tai toisaalta myös ympäristönsuojeluun eläin- ja kasvilajien ja niiden asuinpaikkojen säilyttämiseen.	sujuvuus – välimerkkien käyttö
T10–7	Teadlased vaidlevad, mil määral on inimesed ***praegu toimuvaid muutusi*** põhjustanud ning kuhu see viib.	Tutkijat väittävät, missä määrin ihmiset ovat aiheuttaneet ***muutoksia*** ja mihin se johtaa.	Tutkijat kiistelevät, missä määrin ihminen on aiheuttanut ***nykyiset muutokset*** ja mihin ne johtavat.	tarkkuus – poisto
T10–9	Alustame kohast, kus me ***praegu*** oleme **–** ***Raadist*** ja selle lähiümbrusest – ja uurime, kuidas peegelduvad selle paiga lugudes laiema maailma lood.	Aloitamme paikasta, jossa olemme – ***Radiosta*** ja sen lähiympäristöstä – ja tutkimme, miten tämän paikan tarinat heijastuvat laajemman maailman tarinoissa.	Aloitamme paikasta, jossa ***nyt*** olemme ***eli Raadin*** alueesta ja sen lähiympäristöstä, ja tutkimme, miten tämän paikan historia heijastelee maailman tapahtumia laajemmin.	sujuvuus – välimerkkien käyttö, oikeinkirjoitus tarkkuus – poisto

T10–12	Inimeste ***igipõlises*** võitluses haigustekitajate **–** viiruste ***vastu*** peeti ***18. sajandi*** lõpul oluline lahing just ***Raadil***, kui esmakordselt Eestis õnnestus kaitsepookimine ***rõugete*** vastu.	Ihmisten ***igipölisessä*** taistelussa taudinaiheuttajia ***vastaan*** **–** viruksia ***vastaan*** pidettiin tärkeä taistelu ***1800-luvun*** lopulla juuri ***Radiossa***, kun Virossa ensimmäistä kertaa onnistuttiin puolustamaan ***rõugeita***.	Ihmisten ***ikuisessa*** taistelussa tautien aiheuttajia ***eli*** viruksia ***vastaan*** pidettiin ***1700-luvun*** lopulla olennainen taistelu juuri ***Raadilla***, jossa ensimmäistä kertaa Virossa onnistui ***isorokkorokotus***.	tarkkuus – kääntämätön (2) tyyli – kömpelö tarkkuus – väärä käännös (2) sujuvuus – oikeinkirjoitus, välimerkkien käyttö
T10–15	***Siinsamas*** paiknes Nõukogude Liidu ajal üks Ida-Euroopa suuremaid ***sõjaväelennuvälju***, mille keskkonnamõjusid tunneme veel tänapäevalgi.	Neuvostoliiton aikana oli yksi Itä-Euroopan suurimmista ***armeijan lentoväylistä***, jonka ympäristövaikutukset tunnemme vielä nykyäänkin.	***Täällä*** sijaitsi myös neuvostoaikana yksi Itä-Euroopan suurimmista ***sotilaslentokentistä***, jonka ympäristövaikutukset tunnemme yhä.	tarkkuus – poisto tarkkuus – väärä käännös
T10–18	***Tartu*** – ***aianduse*** pealinn	***tartu*** – ***puutarhan*** pääkaupunki	***Tartto*** – ***puutarhanhoidon*** pääkaupunki	sujuvuus – oikeinkirjoitus (2) tarkkuus – alikääntäminen
T10–20	Raadile rajati suurte kasvuhoonetega ***aed*** ning ***mõisa*** tubades oli võimalik imetleda Balti ***provintside*** suurimat eksootiliste taimede kollektsiooni.	Raadille rakennettiin suuria kasvihuoneita**,** ja ***mantereen*** huoneissa oli mahdollista ihailla Baltian ***maakuntien*** suurinta eksoottisten kasvien kokoelmaa.	Raadille perustettiin ***puutarha*** suurine kasvihuoneineen ja ***kartanon*** rakennuksissa saattoi ihailla Baltian ***provinssien*** suurinta eksoottisten kasvien kokoelmaa.	tarkkuus – poisto, väärä käännös, liikakääntäminen sujuvuus – välimerkkien käyttö

T10–23	Aednikuametit pidasid pikka aega sakslased ja venelased, kuid ***18. sajandi*** lõpus ilmusid ***Tartus*** ka esimesed ***eesti- ja lätikeelsed aiandusraamatud***.	Saksalaiset ja venäläiset pitivät puutarhurivirastoa pitkään, mutta ***1800-luvun*** lopussa myös ensimmäiset ***viro- ja latviakieliset puutarhurikirjat*** ilmestyivät ***Tartussa***.	Puutarhurin ammattia harjoittivat pitkään saksalaiset ja venäläiset, mutta ***1700-***lopulla ***Tartossa*** ilmestyivät myös ensimmäiset ***viron- ja latviankieliset puutarhanhoito-oppaat***.	tarkkuus – väärä käännös sujuvuus – oikeinkirjoitus, kielioppi tyyli – kömpelö
T10–24	Neis õpetati ***mõisa kärnereid***, kuidas ***pookida*** puid ning kasvatada ***ananasse***, ***sparglit***, ***viigimarju*** jne.	Niissä opetettiin ***mantereen kärryjä***, kuinka ***puuta*** ja kasvattaa ***ananassia***, ***spargliä***, ***viikunmarjoja*** ja niin edelleen.	Niissä opetettiin ***kartanon puutarhureille***, kuinka puita ***vartetaan*** tai kasvatetaan ***ananasta***, ***parsaa***, ***viikunoita*** ym.	tarkkuus – väärä käännös (3), poisto, kääntämätön sujuvuus – oikeinkirjoitus
T10–28	***Üheaastase köögiviljapeenra jaoks ei olnud*** isegi enda maad ***vaja***.	***Yhden vuoden vihannespeenra ei tarvinnut*** edes omaa maata.	***Yksivuotista vihannespenkkiä varten ei tarvittu*** edes omaa maata.	tarkkuus – kääntämätön, poisto tyyli – kömpelö, epäidiomaattinen
T10–29	Samas oli ***aiandus riskantne*** ettevõtmine.	***Puutarha*** oli kuitenkin ***riskillinen*** yritys.	Samalla ***puutarhanhoidossa*** oli kuitenkin ***riskinsä***.	tarkkuus – alikääntäminen, väärä käännös

T10–31	Kliima ***soojenemisega 19. sajandil*** aga ***õnnestus aiandusega*** järjest rohkem raha ***teenida***.	Ilmaston ***lämpeneminen 1900-luvulla onnistui*** kuitenkin ***ansaitsemaan*** yhä enemmän rahaa.	Ilmaston ***lämmetessä 1800-luvulla onnistuttiin puutarhanhoidolla*** kuitenkin ***ansaitsemaan*** yhä enemmän, mistä kertoivat Tartossa suuret Daugullin ja Tõnissonin puutarhat.	sujuvuus – kielioppi tarkkuus – väärä käännös, poisto tyyli – epäidiomaattinen
T10–33	Nagu majanduses ikka, saame ka ***aianduses*** jälgida ***moesuundi***.	Kuten taloudessakin, voimme seurata ***puutarhassa modesuuntia***.	Kuten taloudessa yleensäkin, voidaan myös ***puutarhanhoidossa*** seurata ***muotia***.	tarkkuus – väärä käännös, alikääntäminen
T10–34	Nii ***kaubeldi 17. sajandil tulbisibulatega*** nagu väärtpaberitega, ***18. sajandil olid moes nelgid*** ning ***19. sajandi*** vaieldamatud lemmikud ***olid maikellukesed***, ***mida*** kasvatati palju ka ***Raadil***.	***1700-luvulla oli myyty tulbisibulaa*** kuten arvopapereita, ***1800-luvulla moes nelgia*** ja ***1900-luvun*** kiistattomia suosikkeja ***oli maikellukesia***, ***jotka*** kasvatettiin paljon myös ***Radiossa***.	***1600-luvulla käytiin kauppaa tulppaanisipuleilla*** aivan arvopaperien tapaan, ***1700-luvulla olivat muodissa neilikat*** ja ***1800-luvun*** kiistattomia suosikkeja ***olivat kielot***, ***joita*** kasvatettiin paljon myös ***Raadilla***.	tarkkuus – väärä käännös (4), kääntämätön (4), sujuvuus – kielioppi (2), oikeinkirjoitus
T10–35	Igas paradiisis on ***oma madu***.	Jokaisessa paratiisissa on ***oma madunsa***.	Jokaisessa paratiisissa on ***käärmeensä***.	tarkkuus – kääntämätön tyyli – epäidiomaattinen
T10–36	Mitte ainult külmad talved ja sõjad ei hävitanud taimi, vaid ***ka kahjurid***.	Kylmät talvet ja sodat eivät tuhonneet kasveja vaan ***tuholaisia***.	Kasveja eivät tuhonneet pelkästään kylmät talvet ja sodat, vaan ***myös tuholaiseläimet***.	tarkkuus – poisto sujuvuus – kielioppi

T10–37	Nende vastu võideldi ***19. sajandist*** alates järjest rohkem ***keemia abil***, kasutades mitmesuguseid ***pritse*** ja tolmuteid.	Heitä vastaan taisteltiin ***1900-luvulta*** lähtien yhä enemmän ***kemian avulla***, käyttäen erilaisia ***pritseja*** ja pölytteitä.	Niitä vastaan taisteltiin ***1800-luvulta*** lähtien yhä enemmän ***kemiallisilla aineilla*** erilaisia ***suihkeita*** ja pölyttäjiä käyttäen.	tarkkuus – väärä käännös (2), kääntämätön
T10–46	20. sajandi algul oli Eestis ***kultuurviljapuude kõrgaeg***.	1900-luvun alussa Virossa oli ***kulttuuriviljapuudet korkea aika***.	1900-luvun alku Virossa oli ***hedelmäpuutarhojen huippuaikaa***.	tarkkuus – kääntämätön tyyli – epäidiomaattinen
T10–47	Need ***jõudsid eestlaste*** aedadesse suhteliselt hilja ja tõrjusid järkjärgult välja vähemnõudlikud ***metsviljapuud***, mis vajasid küll vähem hoolt, aga andsid ka kehvemat saaki.	Ne ***pääsivät espanjalaisiin*** puutarhoihin suhteellisen myöhään ja jättivät vähitellen vähemmän vaativia ***viljapuita pois***, jotka tarvitsivat vähemmän huolta, mutta antoivat myös huonomman sadon.	Puut ***ilmestyivät virolaisten*** puutarhoihin melko myöhään ja niille saivat antaa tilaa vähempään tyytyvät ***luonnonpensaat***, joiden sato oli ***vatsaavasti pinempi***.	tarkkuus – väärä käännös (2), kääntämätön tyyli – kömpelö
T10–53	***Pakktarudest*** ei olnud võimalik kõiki ***kärgi*** kätte saada.	***Pakkitaruista*** ei ollut mahdollista saada kaikkia ***kääpiöitä***.	***Onkalopesistä*** ei pystytty aina saamaan kaikkia ***kennoja***.	tarkkuus – kääntämätön, väärä käännös
T10–55	Hiljem võeti eelkõige saksa eeskujul kasutusele ***õlgedest*** ja puidust valmistatud ***õlgtarud*** (2), ***mida*** aga tihti kahjustasid ***närilised***.	Myöhemmin erityisesti saksalaisen mallin mukaan otettiin käyttöön ***olkapäästä*** ja puusta valmistetut ***õlat*** (2), ***jotka*** kuitenkin usein vahingoittivat ***jyrsijöitä***.	Myöhemmin otettiin käyttöön ennen muuta saksalaisten esimerkin mukaan ***oljista*** ja puusta valmistettuja ***olkipesiä*** (2), ***joita jyrsijät*** kuitenkin vahingoittivat usein.	tarkkuus – väärä käännös, kääntämätön sujuvuus – kielioppi (2)
T10–61	Hiljemalt ***19. sajandil hakati*** Eesti ***mõisaaedades*** kahjurite tapmiseks kemikaale ***kasutama***.	Viimeistään ***1900-luvulla*** Viron ***mantereissa käytettiin*** kemikaaleja tuholaisten tappamiseen.	Viimeistään ***1800-luvulla*** Viron ***kartanoiden puutarhoissa alettiin käyttää*** kemikaaleja tuholaisten tappamiseen.	tarkkuus – väärä käännös (2), poisto (2)

T14–1	***Süüa*** me teeme	***Syödä*** me teemme	Valmistamme ***ruokaa***	tarkkuus – väärä käännös
T14–3	Toidukultuur on oluline osa rahvuskultuurist, ***millel*** on üha suurem majanduslik ja poliitiline tähtsus.	Ruokakulttuuri on tärkeä osa kansalliskulttuuria, ***jolla*** on yhä suurempi taloudellinen ja poliittinen merkitys.	Ruokakulttuuri on tärkeä osa kansankulttuuria, ja ***sillä*** on yhä suurempi taloudellinen ja poliittinen merkitys.	sujuvuus – kielioppi
T14–4	Toidutraditsioonide ja gastronoomiliste saavutustega loodetakse ***võita*** nii turistide kui kohalike ***südameid***.	Ruokaperinteillä ja gastronomisilla saavutuksilla toivotaan ***voittavan*** sekä turistien että paikallisten ***sydänten***.	Ruokaperinteillä ja gastronomisilla saavutuksilla toivotaan ***voitettavan*** niin matkailijoiden kuin paikallisten asukkaidenkin ***sydämiä***.	tyyli – epäidiomaattinen sujuvuus – kielioppi
T14–7	Küsime, mis jääb püsima ja mis ***hääbub***: on need ***teadmised*** ja ***oskused***, ***töövõtted või tehnoloogiad***, ***retseptid või toorained***.	Kysymme, mitä pysyy ja mikä ***hämärtyy***: ovatko nämä ***osaaminen*** ja ***taidot***, ***työskentely tai teknologia***, ***reseptit tai raaka-aineet***.	Kysymme, mikä pysyy ja mikä ***katoaa***. Ovatko ne ***tietoja*** ja ***taitoja***, ***työtapoja vai tekniikkaa***, ***reseptejä vai raaka-aineita***.	tarkkuus – väärä käännös (2) sujuvuus – kielioppi (6)
T14–9	Lihtne on märgata kohaliku toidukultuuri ***hooajalisust***, seotust ***perekondlike või kalendritähtpäevadega***: kevadel toidulauale ilmuv värske ***räim***, ***jõulu ajal poelette*** täitev ***verivorst*** või iga-aastane rosoljekauss vanaema sünnipäevalaual.	On helppo huomata paikallisen ruokakulttuurin ***kausillisuutta***, yhteyttä ***perhe- tai kalenteripäivään***: keväällä ruokapöydälle ilmestyvä tuore ***räim***, ***joulun aikana poelette*** täyttävä ***verivorst*** tai vuotuinen rosoljekauss isoäidin syntymäpäiväpöydässä.	On helppo havaita paikallisen ruokakulttuurin ***kausittaisuus*** ja sen liittyminen ***perhetapahtumiin tai juhlapyhiin***: keväällä ruokapöytään ilmestyvät tuoreet ***silakat***, ***joulun aikaan kauppojen hyllyt*** täyttyvät ***verimakkaroista*** ja joka vuonna on rosollikulho isoäidin syntymäpäivillä.	sujuvuus – oikeinkirjoitus tarkkuus – väärä käännös, kääntämätön (4) tyyli – kömpelö
T14–11	***Kokkadel*** on peres ja kogukonnas oluline roll.	***Kokkailla*** on tärkeä rooli perheessä ja yhteisössä.	***Kokeilla*** on perheissä ja yhteisöissä tärkeä asema.	sujuvuus – oikeinkirjoitus

T14–15	***Silda*** traditsiooni ja uuenemise ***vahele*** ehitab ***gastronoomia –*** kaasaegne professionaalne ***kokakunst***.	***Sillan*** perinteen ja uudistuksen ***välillä*** rakentaa ***gastronoomia –*** nykyaikainen ammattimainen ***kokataito***.	***Siltaa*** perinteen ja uudistumisen ***välille*** rakentaa ***gastronomia,*** nykyaikainen ammattimainen ***ruoanlaittotaide***.	sujuvuus – oikeinkirjoitus (2), välimerkkien käyttö, kielioppi tarkkuus – väärä käännös
T14–17	Kas need muutused on tehnoloogilised või muutuvad ***pigem ideed*** ja ***toidueelistused***?	Ovatko nämä muutokset teknologisia vai muuttuvatko ***ideoita*** ja ***ruokaetuja***?	Ovatko nämä muutokset teknisiä vai muuttuvatko ***pikemminkin ideat*** ja ***ruokatottumukset***?	sujuvuus – kielioppi (2) tarkkuus – poisto, väärä käännös
T14–18	Mida hindavad Eesti toidukultuuris ***professionaalsed kokad*** ja kuidas ***nad*** seda uuendavad?	Mitä ***ammattimaiset kokat*** arvostavat Viron ruokakulttuurissa ja miten ***ne*** uudistavat sitä?	Mitä ***ammattikokit*** arvostavat Viron ruokakulttuurissa ja miten ***he*** uudistavat sitä?	sujuvuus – oikeinkirjoitus, kielioppi
T14–21	***Kulbid***	***Kulmat***	***Kauhat***	tarkkuus – väärä käännös
T14–22	***Kulp*** on ***üks vanemaid*** ja ***argisemaid toiduvalmistamise riistu***, millega on ***sööki*** segatud ja ***tõstetud***.	***Kulp*** on ***yksi vanhimmista*** ja ***arkistoista ruokailuvälineistä***, jolla ***ruoka*** on sekoitettu ja ***nostettu***.	***Kauha*** on ***vanhimpia*** ja ***arkisimpia ruoanvalmistusvälineitä***, jolla on sekoitettu ja ***otettu ruokaa***.	tarkkuus – kääntämätön, väärä käännös (3) sujuvuus – kielioppi tyyli – epäidiomaattinen (2)

T14–102	***Oli kombeks***, et ***teisel pulmapäeval*** pärast sööki tantsisid ***kokad omavahel***, vahel ka ***peiupoistega***.	***Toisella hääpäivällä*** aterian jälkeen ***kokat*** tanssivat ***toistensa kanssa***, joskus ***peukkupojien*** kanssa.	***Oli tapana***, että ***toisena hääpäivänä*** ruokailun jälkeen ***kokit*** tanssivat ***keskenään***, välillä myös ***sulhaspoikien*** kanssa.	tarkkuus – poisto, väärä käännös sujuvuus – kielioppi, oikeinkirjoitus tyyli – kömpelö
T14–68	Minu ***supid*** kipuvad olema ka ***paksud***, pigem nagu ***ühepajatoidud***.	***Keittiöni*** ovat myös ***paksuja***, mieluummin kuin ***yksipajatoidut***.	Minun ***keittoni*** ovat yleensä melko ***paksuja***, pikemminkin ***patoja***.	tarkkuus – väärä käännös, kääntämätön tyyli – epäidiomaattinen
T14–63	Roheline ***kulp*** on ostetud ühest ***Tartu*** toidupoest 2006. aastal just värvi pärast.	Vihreä ***kulppu*** on ostettu yhdestä ***Tartun*** ruokakaupasta vuonna 2006 juuri värin vuoksi.	Vihreä ***kauha*** on ostettu eräästä ***Tarton*** ruokakaupasta vuonna 2006 juuri sen värin vuoksi.	tarkkuus – väärä käännös sujuvuus – oikeinkirjoitus
T14–100	Tasapisi hakkasid pulmad nihkuma ***kaunile kevadsuvisesele ajale***, kui ***pidu sai pidada*** nii ***toas*** kui ***õues***.	Häät alkoivat vähitellen siirtyä ***kauniille kevään aikoille***, kun ***juhlat pidettiin huoneessa*** ja ***ovessa***.	Pikku hiljaa häät alkoivat siirtyä ***kauniiseen kevätkesään***, jolloin ***voitiin juhlia*** sekä ***sisällä*** että ***ulkona***.	sujuvuus – kielioppi tarkkuus – alikääntäminen, liikakääntäminen, väärä käännös tyyli – epäidiomaattinen

T14–110	Toidu valmistamisega seotud köögiriistade kadumisel on sümboolne tähendus – nende kadumisega hääbub tasapisi ***ka*** mälestus sellest, kuidas vanaema ***süüa tegi***.	Ruoan valmistamiseen liittyvien keittiövälineiden katoamisella on symbolinen merkitys – muisto siitä, miten isoäiti ***teki ruokaa***, katoaa vähitellen.	Ruoanlaittoastioiden katoamisella on symbolinen merkitys***:*** samalla katoaa vähitellen ***myös*** muisto siitä, miten isoäiti ***laittoi ruokaa***.	sujuvuus – välimerkkien käyttö tarkkuus – poisto tyyli – epäidiomaattinen
T14–138	Nii jõulu- kui ka ***vana-aastaõhtul*** on ***perenaised*** teinud oma parima, et laud oleks ***rikkalikult kaetud***.	Sekä joulu- että ***vanha-iltana perheet*** ovat tehneet parhaansa, jotta pöytä olisi ***rikkaasti peitetty***.	Sekä joulu- että ***uudenvuodenaattoiltana perheen emännät*** ovat tehneet parhaansa, jotta pöytä olisi ***runsas***.	tyyli – epäidiomaattinen tarkkuus – väärä käännös, alikääntäminen
T14–140	***19. sajandi*** keskpaigas oli ***vana-aastaõhtul*** laual sealiha ***kaalika***, ***naeri*** ja ***kapsaga***.	***1900-luvun*** puolivälissä pöydällä oli sianlihaa ***kaalika***, ***naura*** ja ***kapsa***.	***1800-luvun*** puolivälissä pöydällä oli ***uudenvuodenaattoiltana*** sianlihaa, ***lanttua***, ***naurista*** ja ***kaalia***.	tarkkuus – väärä käännös, poisto, kääntämätön (2) sujuvuus – oikeinkirjoitus
T14–144	Aastalõpu ***pühademenüü*** uueneb – tihti tahetakse lauale panna midagi ***kergemat***.	Vuoden lopun ***pyhimys*** uusiutuu – usein halutaan laittaa jotain ***helpompaa*** pöydälle.	Vuodenvaihteen ***menu*** on nyt uudistumassa***:*** usein pöydälle halutaan laittaa jotain ***kevyempää***.	tarkkuus – väärä käännös (2) sujuvuus – välimerkkien käyttö
T14–147	Ehk aitab see rituaal meidki valmistuda ees ootavaks ***uueks aastaringiks***?	Ehkä tämä rituaali auttaa meitäkin valmistautumaan tulevaan ***uuteen vuosisataaseen***?	Ehkäpä tämä rituaali auttaa meitäkin valmistautumaan edessä odottavaan ***uuteen vuoteen***?	tarkkuus – väärä käännös

T14–177	Nii koolitatud kui koolitamata ***kokkadel*** on kogukonnas oluline roll.	Sekä koulutetuilla että kouluttamattomilla ***kokkailla*** on tärkeä rooli yhteisössä.	Sekä koulutetuilla että kouluttamattomilla ***kokeilla*** on olennainen rooli yhteisössä.	sujuvuus – oikeinkirjoitus
T14–214	Toiduvalmistamise kaudu luuakse side ***oma minevikuga*** ja tunnustatakse eelnevate põlvede ***loovust*** ja ***leidlikkust köögis rügamisel***.	Ruoan valmistamisen avulla luodaan yhteys ***menneisyyteensä*** ja tunnustetaan aiempien sukupolvien ***luovuus*** ja ***kekseliäisyys keittiössä***.	Ruoanlaiton kautta luodaan side ***omaan menneisyyteen*** ja tunnustetaan aikaisempien sukupolvien ***luovuutta*** ja ***kekseliäisyyttä keittiöaherruksessa***.	sujuvuus – kielioppi (3) tarkkuus – alikääntäminen
T14–215	Toiduvalmistamine annab ***kokale*** võimaluse kombineerida ***omavahel*** erinevaid, vahel ehk isegi vastandlikke traditsioone, kontrollida toiduga seotud rituaale või luua ***päris*** uusi.	Ruoanvalmistus tarjoaa ***kokaiinille*** mahdollisuuden yhdistää erilaisia, joskus jopa ristiriitaisia perinteitä, hallita ruokaan liittyviä rituaaleja tai luoda uusia.	Ruoanvalmistus antaa ***kokille*** mahdollisuuden yhdistellä ***keskenään*** erilaisia, välillä ehkä vastakkaisiakin perinteitä, valvoa ruokaan liittyviä rituaaleja tai jopa luoda ***aivan*** uusia.	tarkkuus – väärä käännös, poisto (2)
T14–234	Kümne aasta eest toimunud uuringu järgi kuulusid eelistatumate roogade hulka ***seapraad*** ja verivorst, ***mulgi kapsad***, ***mulgipuder*** ja ***rukkileib***.	Kymmenen vuotta sitten tehdyn tutkimuksen mukaan suosituimpia ruokia olivat ***sianruoat*** ja ***verivorsti***, ***mulgikapsat***, ***mulgipuder*** ja ***rukkileipä***.	Kymmenen vuotta sitten tehdyn tutkimuksen mukaan suosituimpien ruokien joukkoon kuuluivat ***sianliha*** ja ***verimakkara***, ***sianliha-ohrasuurimohapankaali***, ***peruna-ohraryynipuuro*** ja ***ruisleipä***.	tarkkuus – kääntämätön (4), alikääntäminen
T14–237	Sarnaselt Põhjamaade ***tipprestoranidele*** armastatakse ka Eesti ***restoraniköökides*** kohalikku toorainet.	Pohjoismaiden ***huippuprestoraanien*** tavoin paikallisia raaka-aineita rakastetaan myös Viron ***ravintolakeissa***.	Pohjoismaiden ***huippuravintoloiden*** tapaan myös Viron ***ravintoloiden keittiöissä*** arvostetaan kotimaisia raaka-aineita.	tarkkuus – kääntämätön, väärä käännös

T14–238	***Road*** valmistatakse valdavalt algusest lõpuni kohapeal, sealhulgas ***rakendatakse traditsioonilisi tehnoloogiaid***, nagu tooraine ***marineerimine***, ***soolamine***, ***suitsutamine*** ja ***moorimine***.	***Tie*** valmistetaan enimmäkseen alusta loppuun paikan päällä, mukaan lukien ***perinteiset tekniikat***, kuten raaka-aineiden ***marinointi***, ***suolaaminen***, ***tupakointi*** ja ***polttaminen***.	***Ruoat*** valmistetaan enimmäkseen alusta asti paikan päällä ***käyttäen*** perinteisiä menetelmiä, kuten raaka-aineen ***marinointia, suolausta, savustamista*** ja ***hauduttamista***.	muu tarkkuus – poisto, väärä käännös (2) sujuvuus – kielioppi
T14–242	Tooraine jõuab kööki Eesti eri paigust – näiteks saabuvad värsked ***kukeseened*** suvel ***Põhjakasse*** postiga Lõuna-Eestist.	Raaka-aineet tulevat keittiöön Viron eri paikoista – esimerkiksi kesällä ***Pohjoiskasse***-postilla Etelä-Eestistä saapuvat tuoreet ***keittiöt***.	Raaka-aineet tulevat keittiöihin eri puolilta Viroa***:*** esimerkiksi tuoreet ***kantarellit*** saapuvat kesällä ***Põhjakan kartanoon*** Etelä-Virosta.	sujuvuus – välimerkkien käyttö, oikeinkirjoitus tarkkuus – väärä käännös
T11–20	Sageli magati ***mitmekesi ühes*** voodis.	Usein ***sukeltajat*** nukkuivat ***yhdessä*** sängyssä.	Usein ***samassa*** vuoteessa nukkui ***monta henkilöä***.	tarkkuus – väärä käännös tyyli – epäidiomaattinen
T11–21	Magamis***asemeks võis olla*** seina külge kinnitatud ***lavats***, toolidele tõstetav raam, ***riide***kirstu kaas, ***reheahi*** või ***partepealne***.	Nukkuma-***asento olisi voinut olla*** seinään kiinnitetty ***penkki***, tuoleihin nostettu runko, ***kangas***arkun kansi, ***reheahi*** tai ***ankan yläosa***.	Makuu***sija saattoi olla*** seinään kiinnitetty ***laveri***, tuolille nostettava kehys, ***vaate***arkun kansi, ***riihen uunin*** tai ***parsien ylinen***.	tarkkuus – väärä käännös (4), kääntämätön sujuvuus – kielioppi
T11–23	***Põhjalikumalt*** koristati kodu ***vaid*** suurte ***pühade*** eel, maarahvas käis kord nädalas saunas.	***Tarkemmin*** koti siivottiin ***vasta*** ennen suuria ***lomia***, maaseudulla saunottiin kerran viikossa.	***Perusteellisemmin*** kotona siivottiin ***vain*** suurien ***juhlapyhien*** alla ja maalla käytiin saunassa kerran viikossa.	tarkkuus – väärä käännös (3)

T11–24	Igapäevane oli võitlus ***tubaste*** parasiitidega **–** täide, kirpude ja ***lutikatega***.	Päivittäin oli taistelu ***tilavaa*** loisia ***vastaan -*** täitä, kirppuja ja ***vikoja vastaan***.	Syöpäläisten kanssa taisteltiin päivittäin***:*** riesana olivat täit, kirput ja ***luteet***.	tarkkuus – väärä käännös (2) tyyli – kömpelö sujuvuus – välimerkkien käyttö
T11–52	Linnades ***nägi hea toon ette*** rangemat hügieeninormide järgimist kui maal.	Kaupungeissa ***hyvä sävy ennakoi*** tiukempaa hygieniastandardien noudattamista kuin maaseudulla.	Kaupungeissa ***noudatettiin*** hygieniavaatimuksia täsmällisemmin kuin maalla.	tyyli – epäidiomaattinen tarkkuus – väärä käännös
T11–56	***Hinnati valgusküllasust ja siledaid pindu*** ning ***soovitati vabaneda vanamoodsast tolmu koguvast mööblist***.	***Kirkkaus ja sileät pinnat arvioitiin*** ja ***vanhanaikaiset pölyä keräävät huonekalut suositeltiin hävittämään***.	***Arvostettiin valoisuutta ja tasaisia pintoja*** ja ***haluttiin vapautua vanhanaikaisista ja pölyä keräävistä huonekaluista***.	sujuvuus – kielioppi (4) tarkkuus – väärä käännös, liikakääntäminen
T11–57	***Soositi heledaid tapeete*** ja tekstiile.	***Valotaustakuvia*** ja tekstiilejä ***suosittiin***.	***Suosittiin vaaleita tapetteja*** ja tekstiilejä.	tyyli – kömpelö tarkkuus – väärä käännös
T11–69	***Jõukaid inimesi*** kutsuti üles loobuma eraldi ***saalist, söögitoast*** ja ***kabinetist*** kui tarbetust ***luksusest***.	***Varakkaita*** kehotettiin luopumaan erillisestä ***salista, ruokasalista*** ja ***toimistosta*** tarpeettomana ***ylellisyydenä***.	***Varakkaita ihmisiä*** kehotettiin luopumaan erillisestä ***salista***, ***ruokailuhuoneesta*** ja ***työhuoneesta*** tarpeettomana ***ylellisyytenä***.	tarkkuus – poisto, liikakääntäminen tyyli – kömpelö sujuvuus – oikeinkirjoitus

T11–71	***Linnade kallimates korterites olid lisaks*** elektrivalgustusele, keskküttele ja ***vesiklosetile*** ka vannituba ja elektripliit.	Sähkövalaistuksen, keskuslämmityksen ja ***vesivaatteen lisäksi kaupunkien kalliimpiin asuntoihin kuuluivat*** myös kylpyhuone ja sähköliesi.	***Kaupunkien kalliimmissa asunnoissa oli*** sähkövalaistuksen, keskuslämmityksen ja ***wc:n*** lisäksi myös kylpyhuone ja sähköliesi.	tyyli – kömpelö tarkkuus – väärä käännös
T11–87	***Sõjajärgsetel aastatel*** valitses ***suur*** korteri***kitsikus***.	***Sodanjälkeisenä vuonna*** oli ***hieno*** asunto.	***Sodanjälkeisinä*** vuosina vallitsi ***suuri*** asunto***pula***.	sujuvuus – kielioppi tarkkuus – väärä käännös, poisto
T11–91	Väikseimad köögid olid vaid 4,5 ruutmeetri suurused, sest ***eeldati***, et ***sööklad*** asendavad ***kodust söögitegemist***.	Pienimmät keittiöt olivat kooltaan vain 4,5 neliömetriä, sillä ***ruokailijoiden odotettiin*** korvaavan ***kotiruokailijat***.	Pienimmät keittiöt olivat vain 4,5 neliömetrin kokoisia, koska ***edellytettiin***, että ***yhteisruokalat*** korvaavat ***ruoanvalmistamisen kotona***.	tarkkuus – väärä käännös (3)
T11–97	***Ajakirjas Kunst ja Kodu*** jagati väikekorterite omanikele ***näpunäiteid***, ***kuidas interjööre sisustada***, ise ***mööblit*** ja ***dekoratiivesemeid valmistada***.	***Kunst and Home -lehdessä*** annettiin ***vinkkejä*** pienten asuntojen omistajille siitä, ***miten sisustetut sisätilat***, ***huonekalut*** ja ***koriste-esineet koristellaan*** itse.	***Aikakauslehdessä Kunst ja Kodu*** jaettiin pienasuntojen omistajille ***vinkkejä***, ***kuinka tiloja voisi sisustaa*** ja ***valmistaa itse kalusteita*** ja ***koriste-esineitä***.	muu tyyli – kömpelö sujuvuus – kielioppi (4) tarkkuus – väärä käännös (2)

T11–169	Samas pakuvad mitmesugused avalikud kohad nagu poed, restoranid või ***majutuskohad*** „***koduseid***“ kaupu või teenuseid.	Samaan aikaan erilaiset julkiset paikat**,** kuten kaupat, ravintolat tai ***majoitukset***, tarjoavat "***kodiisia***" tavaroita tai palveluita.	Myös useat julkiset paikat kuten kaupat, ravintolat tai ***majoituspaikat*** tarjoavat "***kodikkaita***" palveluita ja tuotteita.	sujuvuus – välimerkkien käyttö, oikeinkirjoitus tarkkuus – alikääntäminen
T11–100	Massiivsete voodite ***kõrvale*** või ***asemele ilmusid*** minimalistlikud ***kušetid***.	Massiivisten sänkyjen ***vieressä*** tai ***vieressä näytti*** minimalistisia ***sohvia***.	Massiivisten vuoteiden ***rinnalle*** tai ***sijalle ilmestyivät*** minimalistiset ***runkosängyt***.	sujuvuus – kielioppi tarkkuus – väärä käännös (3)
T11–103	Interjööri tulid plastmassist lambid, õhukesed ja ***heledad*** kardinad.	Sisätiloihin tuli muovilamppuja, ohuita ja ***kevyitä*** verhoja.	Sisustukseen tuli muovisia lamppuja sekä ohuita ja ***vaaleita*** verhoja.	tarkkuus – väärä käännös
T11–104	Kümnendi keskpaigas soovitati juba ***toekamaid*** ja ***kandilisemaid*** vorme ning lõpus ***sai ihaldusväärseks suur sektsioonkapp***.	Vuosikymmenen puolivälissä suositeltiin jo ***kannustavampia*** ja ***neliömäimpiä*** muotoja, ja lopussa ***suuren osan kaapista tuli toivottava***.	Vuosikymmenen puolivälissä suosittiin jo ***tukevampia*** ja ***kulmikkaita*** muotoja, ja lopulta ***suuret seinäkaapistot saivat suuren suosion***.	tarkkuus – väärä käännös (4)
T11–106	***Argiellu ilmus*** juba arvukalt masinaid – tolmuimeja, külmutuskapp, pesumasin, ***elektriline õmblusmasin***, ***triikraud***, kohviveski jm.	***Arjessa on*** jo ***ilmestynyt*** lukuisia koneita – pölynimuri, jääkaappi, pesukone, ***sähkö ompelukone***, ***rauta***, kahvimylly jne.	***Arkielämään ilmestyi*** myös jo paljon koneita***:*** pölynimuri, jääkaappi, pyykinpesukone, ***sähkökäyttöinen ompelukone***, ***silitysrauta***, kahvimylly ym.	sujuvuus – kielioppi (2), välimerkkien käyttö tyyli – kömpelö tarkkuus – väärä käännös

T11–110	1970. aastatel algas massiline paneelehitus, ***kerkisid*** suured ***üheülbaliste*** kortermajade linnaosad.	Joukkopaneelirakentaminen alkoi 1970-luvulla, ***syntyi*** suuria ***yksinäisten*** kerrostalojen kaupunginosia.	1970-luvulla alkoi massiivinen elementtirakentaminen ja ***tehtiin*** runsaasti suuria ***toistensa kaltaisia*** kerrostaloalueita.	tarkkuus – väärä käännös (2)
T11–111	***Korterikitsikust*** see ei vähendanud, sest aastas tuli juurde umbes 10 000 sisserändajat.	Se ei vähentänyt ***kerrostaloasuntojen määrää***, sillä maahanmuuttajia tuli vuosittain noin 10 000.	***Asuntopulaa*** se ei vähentänyt, koska joka vuosi tuli noin 10 000 maahanmuuttajaa.	tarkkuus – väärä käännös (2)
T11–112	Neis ***vabaplaneeringuga rajoonides*** ehk ***magalates*** elas mitmerahvuseline ja sotsiaalselt ***kirju seltskond***, kelle arusaamad kodust ja ***kodutundest olid erinevad***.	Näissä ***vapaasuunnittelupiireissä*** eli ***nukkujissa*** oli monietninen ja sosiaalisesti ***kirjoitettu yhteiskunta***, jolla oli erilaiset käsitykset kodista ja ***kodista***.	Näissä ***saman kaavan mukaan rakennetuissa nukkumalähiöissä*** asui monikansallista ja sosiaalisesti ***värikästä väkeä***, jonka näkemykset kodista ja ***kodikkuudesta poikkesivat*** toisistaan.	tarkkuus – väärä käännös (5) sujuvuus – kielioppi
T11–117	Kodud muutusid omamoodi ladudeks: tagavaraks osteti ***kõike***, mida ***saada oli***.	Kodeista tuli eräänlainen varasto: ***kaikki***, mitä ***he pystyivät saamaan***, ostettiin varasijalle.	Kodit muuttuivat tavallaan varastoiksi: varmuuden vuoksi ostettiin ***kaikkea***, mitä ***oli saatavilla***.	sujuvuus – kielioppi (2)
T11–118	Kindlasti ***pidi*** varuks ***olema kuivaineid***, kohvi ja ***napsi***, ***aga ka voodilinu***, käterätte, ***serviise*** ja seepi.	Varmasti ***on täytynyt olla kuiva-ainetta***, kahvia ja ***juomaa sekä kääpiöitä***, pyyhkeitä, ***serererejä*** ja saippuaa.	Varastossa ***piti*** ehdottomasti ***olla kuivatuotteita***, kahvia ja ***vodkaa***, ***mutta*** ostettiin ***myös liinavaatteita***, pyyhkeitä, ***astiastoja*** ja saippuaa.	sujuvuus – kielioppi (2) tarkkuus – väärä käännös (3), alikääntäminen

T11–123	Eriti ***hinnati Jugoslaavia***, ***Rumeenia*** või ***Saksa DV komplekte***.	Erityisesti ***arvioitiin Jugoslavian***, ***Romanian*** tai ***Saksan DV-pakkauksia***.	Erityisesti ***arvostettiin Jugoslavian***, ***Romanian*** ja ***DDR:n kalusteita***.	tarkkuus – väärä käännös (3) sujuvuus – välimerkkien käyttö tyyli – epäidiomaattinen
T11–131	Laste***mööblit*** ja ***mänguasju*** valmistati siiski juba ammu.	Lasten ***huonekalut*** ja ***lelut*** tehtiin kuitenkin jo kauan sitten.	Lasten ***huonekaluja*** ja ***leikkikaluja*** valmistettiin kuitenkin jo aiemmin.	sujuvuus – kielioppi (2)
T11–132	Talulapsed mängisid enamasti looduslikest materjalist ***leludega***.	Maatilan lapset leikkivät enimmäkseen luonnonmateriaalista ***tehdyllä lelulla***.	Maalaislapset leikkivät enimmäkseen luonnonmateriaaleista ***tehdyillä leluilla***.	sujuvuus – kielioppi
T11–138	Mänguasjad ja raamatud pidid aitama ***kujundada*** lapse isiksust.	Lelujen ja kirjojen piti auttaa ***muokkaamaan*** lapsen persoonallisuutta.	Leikkikalujen ja kirjojen piti auttaa ***kehittämään*** lapsen persoonallisuutta.	tarkkuus – väärä käännös
T11–147	***Lastetoas magatakse, mängitakse*** ja ***õpitakse***, aga kaasajal on see ka meelelahutus- ja meediakeskus.	***Lastenhuone nukkuu, leikkii*** ja ***oppii***, mutta nykyaikana se on myös viihde- ja mediakeskus.	***Lastenhuoneessa nukutaan, leikitään*** ja ***opiskellaan***, mutta se on nykyään myös viihde- ja mediakeskus.	sujuvuus – kielioppi (4)
T3–15	1820.–1830. aastatel rajati Eestis esimesed tehased ja ***vabrikud***.	Virossa perustettiin ensimmäiset ***tehtaat*** ja ***tehtaat*** vuosina 1820–1830.	Ensimmäiset ***tehtaat*** perustettiin 1820–1830-luvulla.	tyyli – kömpelö
T3–16	***Vabrikutööliste arvel*** kasvasid plahvatuslikult linnad, kättesaadavaks muutus odav tarbekaup.	***Vabrikutyöntekijöiden kustannuksella*** kaupungit kasvoivat räjähdysmäisesti, ja halpoja kulutustuotteita tuli saataville.	***Tehdastyöläisten ansiosta*** kaupungit kasvoivat räjähdysmäisesti ja halvat kulutustavarat tulivat kaikkien saataville.	tarkkuus – kääntämätön, väärä käännös

T3–23	Laguneva Vene ***keisririigi nurgakesest*** sai 1918. aastal iseseisev Eesti Vabariik.	Hajoavan Venäjän ***keisarivaltion kulmakivistä*** tuli vuonna 1918 itsenäinen Viron tasavalta.	Murenevan Venäjän ***keisarikunnan äärialueesta*** tulikin vuonna 1918 itsenäinen Viron tasavalta.	tarkkuus – väärä käännös (2)
T3–29	1930. aastate keskpaiku muutis oma ***nime*** ligi 210 000 inimest.	1930-luvun puolivälissä ***nimi*** muutti lähes 210 000 ihmistä.	1930-luvun puolivälissä ***nimensä*** muutti lähes 210 000 ihmistä.	sujuvuus – kielioppi
T3–33	***Raha saadi alkoholi- ja tubakaaktsiisilt***.	***Rahat tulivat alkoholi- ja tupakka-aktiivisista veroista.***	***Rahaa saatiin alkoholi- ja tupakkaverosta.***	sujuvuus – kielioppi tarkkuus – väärä käännös
T3–460	Laen tasuti ***pasunakoori heaks korraldatud pidude*** sissetulekutest.	Laina maksettiin ***juhlien tuloista, jotka järjestettiin pasunakooren*** hyväksi.	Laina maksettiin takaisin ***puhallinorkesterin hyväksi järjestettyjen juhlien*** pääsymaksuista.	tyyli – kömpelö tarkkuus – kääntämätön
T3–183	***Moekas*** naine oli ***nooruslik***, sportlik ja ***poisiliku väljanägemisega***.	***Moekas*** nainen oli ***nuorempi***, urheilullinen ja ***poikaystävällinen***.	***Muodikas*** nainen oli ***nuorekas***, urheilullinen ja ***poikamainen***.	tarkkuus – kääntämätön, väärä käännös sujuvuus – kielioppi
T3–184	***Poisilikkust*** rõhutas ajastu ***moesoeng poisipea***.	***Poikaystävällisyyttä*** korosti aikakauden ***moesoeng poispää***.	***Poikamaisuutta*** korosti aikakauden ***muotileikkaus poikatukka***.	tarkkuus – väärä käännös (2), kääntämätön
T3–187	***Meeleolukad tantsuõhtud*** olid populaarsed nii linnas kui maal.	***Mielenkiintoiset tanssipohjat*** olivat suosittuja sekä kaupungissa että maakunnassa.	***Viihteelliset tanssi-illat*** olivat suosittuja sekä maalla että kaupungissa.	tarkkuus – väärä käännös (2)

T3–198	Meeste äraolekul olid naised teinud kõik selleks, et elu läheks edasi: töötanud vabrikutes, ***harinud*** põldu ja ***põetanud*** haavatuid.	Miesten poissa ollessa naiset olivat tehneet kaikkensa elämän jatkamiseksi: työskennelleet tehtaissa, ***kouluttaneet*** maata ja ***sairastuneet*** haavoittuneita.	Miesten ollessa poissa naiset olivat tehneet kaikkensa, jotta elämä kulkisi eteenpäin: työskennelleet tehtaissa, ***viljelleet*** maata ja ***hoitaneet*** haavoittuneita.	tarkkuus – väärä käännös (2)
T3–200	Eesti Vabariigis said naised põhiseadusega poliitilised õigused, mille eest ***sookaaslased*** mitmetes Euroopa riikides ning Ameerika Ühendriikides ***olid võidelnud sada ja rohkem aastat***.	Viron tasavallassa naiset saivat perustuslailliset poliittiset oikeudet, joiden puolesta ***suomalaiset ovat taistelleet*** useissa Euroopan maissa ja Yhdysvalloissa ***sata vuotta***.	Viron tasavallassa naiset saivat perustuslailla poliittiset oikeudet, joiden edestä ***naiset*** monissa Euroopan maissa ja USA:ssa ***olivat taistelleet yli satakin vuotta***.	tarkkuus – väärä käännös, alikääntäminen sujuvuus – kielioppi
T3–212	Kodus olid ***daami tualettlaual*** Lorupi vabrikus valmistatud ***puudritoos*** ja täidetavad ***lõhnaveepudelid***.	Kotona oli ***Ladyn vessapöydällä*** Lorupin tehtaalla valmistettu ***puuritos*** ja täytettävät ***hajuvedet***.	Kotona ***meikkauspöydällä*** oli Lorupin tehtaalla valmistettu ***puuterirasia*** ja täytettäviä ***hajuvesipulloja***.	muu tarkkuus – väärä käännös (2), alikääntäminen
T3–233	Selgitati, et 1,5 ***kg saia küpsetamine*** elektriahjus maksab 2–4 senti, et 30 ***kg pesu pesemiseks*** ja ***triikimiseks*** kulub ***23 sendi eest elektrit***.	Selitettiin, että 1,5 ***kg leivän keittäminen*** sähköuunissa maksaa 2–4 senttiä, että 30 ***kg pesuun*** ja ***pyyhkimiseen*** kuluu ***23 senttiä sähköä***.	Kerrottiin, että 1,5 ***kg:n pullan paistaminen*** sähköuunissa maksaa 2–4 senttiä, ja että 30 ***kg:n pyykin pesemiseen*** ja ***silittämiseen*** kuluu ***23 sentin edestä sähköä***.	tarkkuus – väärä käännös (3), poisto (2) sujuvuus – kielioppi (2)
T3–236	Levinuim elektriline ***majapidamisriist*** oli ***triikrauad***.	Yleisin sähköinen ***kodinkoneet*** oli ***triikraat***.	Yleisin ***sähkölaite*** oli ***silitysrauta***.	sujuvuus – kielioppi tarkkuus – kääntämätön

T3–277	Pääsuke ***tegi*** küll seeria ***fotosid Tartu linnast***, kuid ka ***nende*** umbes 400 ***foto*** seas ***kujutavad*** paljud linna ***aguleid ja vaesema rahva elu***.	Pääsuke ***teki*** sarjan ***kuvia Tartusta***, mutta myös ***niiden*** noin 400 ***valokuvassa*** monet kaupungin ***aguleista ja köyhimpien ihmisten elämästä***.	Pääsuke ***otti*** kyllä sarjan ***valokuvia Tarton kaupungista***, mutta myös ***nämä*** noin 400 ***valokuvaa kuvaavat*** enimmäkseen ***laitakaupungin ja köyhemmän kansan elämää***.	tyyli – epäidiomaattinen sujuvuus – oikeinkirjoitus, kielioppi (3) tarkkuus – poisto (2), kääntämätön
T3–301	Suure osa Eesti ***merekuurortide külastajatest moodustasid*** venelased.	Venäläiset ***muodostavat*** suuren osan Viron ***merenkulkureittien vierailijoita***.	Suuren osan Viron ***kylpyläkaupunkien vierailijoista muodostivat*** venäläiset.	sujuvuus – kielioppi (2) tarkkuus – väärä käännös
T3–309	Päeval ***veedeti*** aega pigem ***varjulistes*** aedades ja parkides.	Päivällä ***vietiin*** enemmän aikaa ***varjoissa*** puutarhoissa ja puistoissa.	Päivällä ***vietettiin*** aikaa pikemminkin ***varjoisissa*** puutarhoissa ja puistoissa.	tarkkuus – väärä käännös sujuvuus – kielioppi
T3–310	Rand ***elavnes*** õhtupoolikul, kui ***sinna tuldi jalutama***, muusikat kuulama ja päikeseloojangut imetlema.	Rand ***elvytti*** iltapäivällä, kun ***he tulivat kävelemään***, kuuntelemaan musiikkia ja ihailemaan auringonlaskua.	Ranta ***vilkastui*** iltaisin, kun ***sinne tultiin kävelemään***, kuuntelemaan musiikkia ja ihailemaan auringonlaskua.	tarkkuus – väärä käännös, poisto sujuvuus – kielioppi
T3–326	Kui varem kaheldi ***meresupluse sobilikkuses*** lastele, ***siis*** nüüd sai rannast laste paradiis.	Kun aiemmin kyseenalaistettiin ***merisuplusen sopivuus*** lapsille, nyt rannasta tuli lasten paratiisi.	Aiemmin epäiltiin ***meressä uimisen sopivuutta*** lapsille, ***mutta*** nyt rannasta tuli lasten paratiisi.	tarkkuus – kääntämätön, poisto sujuvuus – kielioppi

T3–428	Sõitjatele pakkusid huvi ***võõrad*** paigad ja kaunis ***varasuvine loodus***, sest ***talurahvas*** pääses harva kodust kaugemale.	Matkustajat olivat kiinnostuneita ***ulkomaisista*** paikoista ja kauniista ***varhaisesta luonteesta***, koska ***maanviljelijät*** pääsivät harvoin kotia kauemmas.	Matkaajia kiehtoivat ***uudet*** seudut ja kaunis ***varhaiskevään luonto***, koska ***talonpoikaiskansa*** oli päässyt harvoin kotitilan ulkopuolelle.	tarkkuus – väärä käännös (3), poisto
T3–436	Kell üheksa andsid ***kirikukellad*** lauljatele märku koguneda ***rongkäigule***.	Kello yhdeksältä ***kirikukellat*** antoivat laulajille merkin kokoontua ***junakävelylle***.	Yhdeksältä ***kirkonkellot*** antoivat laulajille merkin kerääntyä ***kulkueeseen***.	tarkkuus – kääntämätön, väärä käännös
T3–437	***Maarahva rongkäik*** kulges läbi ***ülikoolilinna*** tänavate, ***kus*** seni oli liikumas nähtud vaid ***saksa seltside*** või ***üliõpilaskorporatsioonide*** rühmi.	***Maaväen juna*** kulki ***yliopistolinn*** kaduilla, ***joissa*** tähän mennessä oli nähty vain ***saksalaisten konsernien*** tai ***opiskelijayritysten*** ryhmiä.	***Maalaiskansan kulkue*** kulki pitkin ***yliopistokaupungin*** katuja, ***joilla*** oli siihen saakka nähty vain ***saksalaisyhdistysten*** tai ***ylioppilasosakuntien*** ryhmiä.	tarkkuus – väärä käännös (5) sujuvuus – kielioppi

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Krista Kotkamäe,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Viro-suomi ihmis- ja konekäännöksen vertailu Viron kansallismuseon Kohtaamisia-näyttelyn tekstien perusteella“, mille juhendaja on Juha-Matti Aronen, 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 Commonsi 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.

Krista Kotkamäe

26.05.2023