Tampereen yliopisto

Elli Suomela

# MYÖTÄJÄISET

## Kulttuuriviittausten konekääntäminen ja jälkieditointi kaunokirjallisuudessa

Informaatioteknologian ja viestinnän tiedekunta
Pro gradu -tutkielma
huhtikuu 2026

# TIIVISTELMÄ

Elli Suomela: Myötäjäiset: Kulttuuriviittausten konekääntäminen ja jälkieditointi kaunokirjallisuudessa
Pro gradu -tutkielma
Tampereen yliopisto
Monikielisen viestinnän ja käännöstieteen maisteriohjelma, englannin kääntämisen ja tulkkauksen opintosuunta
huhtikuu 2026

---

Tutkielmassa tutkin sitä, miten konekäännin kääntää kulttuuriviittauksia kaunokirjallisessa tekstissä. Tutkielman aineistona toimii Siru Auran Myötäjäiset-teoksen kolme ensimmäistä lukua, niiden raakakonekäännös ja jälkieditoitu käännös. Myötäjäiset-teos on romaani, joka sijoittuu 2000-luvun alun Intiaan ja käsittelee naisten asemaa intialaisessa avioliittokulttuurissa. Raakakonekäännöksen ja jälkieditoidun käännöksen teoksen ensimmäisistä kappaleista on tehnyt Tampereen yliopiston lehtori Sari Hokkanen osana tutkimusprojektiaan. Tutkielman kohteena on lähdetekstistä ja sen käännöksistä valikoidut kulttuuriviittaukset. Tutkin sitä, miten konekäännin on kääntänyt kulttuuriviittaukset ja välittyykö lähdetekstin kulttuurinen merkitys myös raakakonekäännöksessä. Lisäksi tutkin, näkyykö englannin kielen ja länsimaisen kulttuurin valta-asema raakakonekäännöksessä.

Määrittelen kulttuurin ja kulttuuriviittauksen käsitteet, sekä esittelen kulttuurillisten tekstien kääntämisstrategioita ja -haasteita. Analyysin taustoittamiseksi esittelen neuroverkkokonekäänninten toimintaa ja kaunokirjallisuuden konekääntämisen nykytilannetta. Vaikka en tehnyt tutkielman aineistosta kattavaa laadullista analyysia, esittelen laadun käsitteen käännösalalla, koska se on silti merkityksellistä tutkielman analyysissa. Lisäksi esittelen konekääntämiseen liittyviä eettisiä kysymyksiä ja englannin kielen ja länsimaisen kulttuurin dominanssia käännösalalla ja konekääntämisessä.

Tutkielman tuloksista voidaan päätellä, että konekäännin ei tuota johdonmukaista kaunokirjallista käännöstä. Lähdetekstin kulttuurinen sisältö muuttui usein aineistosta valikoiduissa kulttuuriviittauksissa, ja jälkieditoija muokkasi käännöstä sisällyttääkseen alkuperäisen kulttuurisen merkityksen myös käännökseen. Aineiston esimerkeissä ilmeni myös merkittävä englannin kielen ja länsimaisen kulttuurin dominanssi. Lisäksi konekääntimen käännösvalinnoissa esiintyi maskuliininen sukupuolipreferenssi. En käsittele sukupuolipreferenssiä laajasti tässä tutkielmassa, mutta sen esiintyminen voi toimia pohjana jatkotutkimuksille. Myöhemmissä tutkimuksissa olisi myös hyödyllistä tutustua suurten kielimallien konekääntimillä tehtyihin kaunokirjallisiin käännöksiin, koska ne ovat nykyään johtavassa asemassa käännösteknologian alalla.

Avainsanat: konekääntäminen, jälkieditointi, kaunokirjallisuus, kulttuuriviittaus

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

# TEKOÄLYN KÄYTTÖ OPINNÄYTTEESSÄ

Opinnäytteessäni on käytetty tekoälysovelluksia:

- ☒ Ei
- ☐ Kyllä

# SISÄLLYS


1 JOHDANTO ..... 1

2 TEOREETTINEN VIITEKEHYS ..... 4

2.1 Kulttuuri ..... 4

2.1.1 Kulttuuriviittauksen määritelmä ..... 5

2.1.2 Kulttuurin kääntämisen perinteestä ..... 6

2.1.3 Strategiat ..... 8

2.1.4 Haasteet ..... 10

2.2 Kaunokirjallisuuden konekääntäminen ..... 11

2.2.1 Konekääntimet ja neuroverkkopohjainen konekääntäminen ..... 11

2.2.2 Jälkieditointi ..... 12

2.2.3 Konekääntäminen ja kaunokirjallisuus ..... 13

2.3 Laadun käsite käännösalalla ..... 14

2.4 Konekääntimien eettisyys ..... 19

2.5 Englannin kielen ja länsimaisen kulttuurin valta-asema käännösalalla ja konekääntimissä ..... 20

3 AINEISTO JA MENETELMÄ ..... 23

3.1 Aineisto ..... 23

3.2 Menetelmä ..... 24

4 ANALYYSI ..... 28

4.1 Erisnimet ..... 28

4.2 Implisiittiset kulttuuriviittaukset ..... 29

4.3 Reaalioiden käännökset ..... 32

4.4 Selittävät käännökset ..... 34

4.5 Amerikkalaisen kulttuurin vaikutus ..... 37

5 TULOKSET JA POHDINNAT ..... 40

LÄHTEET ..... 46

Aineistolähteet ..... 46

Kirjallisuuslähteet ..... 46

ENGLANNINKIELINEN LYHENNELMÄ ..... I

MACHINE TRANSLATION AND POST-EDITING IN A LITERARY TEXT ..... I

Introduction ..... I

Theoretical framework ..... II

Method and Data ..... VI

Analysis ..... VII

Conclusion ..... IX

# 1 JOHDANTO

Konekäänninteknologiaa on kehitetty jo 1950-luvulta asti, ja sen suosio on noussut viime vuosien aikana huomattavasti tekoälyteknologian kehityttyä ja yleistyttyä. Kiinnostus myös kaunokirjallisuuden kääntämisprosessin automatisointia kohtaan on noussut. (Rothwell ym. 2023; Taivalkoski-Shilov 2019.) Tämän tutkielman tavoitteena on tutkia kaunokirjallisen tekstin konekääntämistä ja jälkieditointia kaunokirjallisuudessa. Selvitän, miten konekäännin kääntää kulttuuriviittauksia kaunokirjallisessa tekstissä ja sitä, säilyykö lähdetekstin kulttuurinen merkitys konekäännöksessä. Tarkoituksena on myös tutkia, esiintyykö englannin kielen ja länsimaisen kulttuurin dominanssia konekäännöksessä. Tutkimuskysymysten tavoitteena on päätellä, luoko konekäännin julkaisukelpoista kaunokirjallista tekstiä itsenäisesti. Tutkimuksen kohteena on Myötäjäiset-romaani, josta on tehty englanninkielinen raakakonekäännös DeepL-konekääntimellä sekä viimeistelty käännös, jonka on luonut Tampereen yliopiston englannin kääntämisen lehtori Sari Hokkanen.

Myötäjäiset on suomalaisen kirjailija-antropologi Siru Auran romaani, joka sijoittuu Intiaan 2000-luvun alussa. Romaani käsittelee naisten asemaa ja myötäjäismurhia Intiassa. Myötäjäismurhalla tarkoitetaan nuoren naisen murhaa tai itsemurhaa 7 vuoden sisällä hänen avioitumisestaan, jonka syynä on naiseen kohdistuva jatkuva häirintä, kidutus ja väkivalta uuden aviomiehen tai hänen perheensä toimesta. Tekojen taustalla on tavoite kiristää vaimon vanhempia maksamaan suuremmat myötäjäiset tyttärensä uudelle puolisolle. Tällaiset murhat olivat yleisiä Intiassa, jossa perinteiseen avioliittosopimukseen ennen kuului morsiamen vanhempien uudelle sulhaselle maksamat myötäjäiset. Vuonna 2000 myötäjäismurhia tapahtui Intiassa 6995. Vaikka Intiassa kirjoitettiin laki, joka poisti myötäjäisten pakollisen maksamisen, jotkin miehet silti kokivat myötäjäismaksun oikeudekseen. (Mohanty ym. 2013.) Kun uuden vaimon vanhemmat eivät maksaneet myötäjäisiä, tilanne joissakin tapauksissa kärjistyi ja päättyi naisen kuolemaan. Siru Auran romaanin tarinassa päähenkilö Taarinin pikkusiskon uusi aviomies ei alun perin pyydä tyttöjen vanhemmilta myötäjäisiä kihlautuessaan ja avioituessaan pikkusiskon kanssa, mutta myöhemmin tuohtuu niiden puutteesta ja alkaa pahoinpidellä siskoa. Taarinin sisko kuolee lopulta epäselvissä olosuhteissa. Romaanin tarina seuraa Taarinia ja kahta muuta naista, joiden kaikkien elämään myötäjäismurhat ovat vaikuttaneet. Kahden intialaisen naisen lisäksi tarinassa tutustutaan suomalaiseen antropologiin, joka tekee etnografista tutkimusta naisten asemasta Intiassa. Teoksessa on siis läsnä viittauksia sekä intialaiseen että

suomalaiseen kulttuuriin. Siru Aura on suomalainen antropologi, joka kirjoitti myös oman väitöskirjansa naisten asemasta Intiassa suorittamansa etnografisen tutkimuksen pohjalta (Aura 2008). Tutkittavan romaanin tarina perustuukin kirjailijan omiin ja hänen tapaamiensa ihmisten kokemuksiin.

Sari Hokkanen toteuttaa tällä hetkellä tutkimusprojektia koneavusteisesta kaunokirjallisuuden kääntämisestä. Hokkanen tekee raakakonekäännöksen Myötäjäiset-teoksesta, jonka hän jälkieditoi. Tämän jälkeen käännös viimeistellään yhteistyössä kirjailijan kanssa. Hokkasen tutkimuksen lopullisesta aineistosta on tarkoitus tutkia, miten kulttuurinen tieto kulkee erilaisten tiedonjärjestelmien läpi (Hokkanen 2025). Tutkimus on tämän pro gradu -tutkielman aikaan jälkieditointi-vaiheessa, mutta Hokkanen on kirjailijan kanssa yhteistyössä tuottanut viimeistellyt käännökset teoksen kolmesta ensimmäisestä luvusta tutkimusprojektin pilottivaiheessa. Tämän tutkielman aineisto on kerätty näistä pilottivaiheen käännöksistä sekä Myötäjäiset-teoksen kolmesta ensimmäisestä luvusta.

Tutkielmassa perehdyn teoksen kolmessa ensimmäisessä luvussa esiintyvien kulttuuriviittausten konekäännöksiin sekä siihen, näkyykö englannin kielen ja länsimaisen kulttuurin dominanssi konekäännöksessä. Englannin kieli toimii globaalina lingua francana ja pitää siten valta-asemaa muihin kieliin nähden käännösalalla. Länsimaisen kulttuurin valta-asema juontaa juurensa kolonialismiin, ja käännösalallakin on perinteisesti suosittu länsimaista kulttuuria, ajoittain muita kulttuureja ja niiden viittauksia häivyttäen. (Bassnett & Lefevere 1998; Yoon 2021.) Koska suuri osa johtavasta käännösteknologiasta on kotoisin länsimaista ja niiden toimintajärjestelmien pohjana on toiminut englanti, tämä dominanssi on nähtävissä myös konekäännöksissä (Olojo ym. 2025). Kulttuuriviittaukset ovat erityisen relevantteja tämän tutkielman kohteena olevan romaanin kaltaisissa teoksissa, joissa romaanin tarina kulkee valtioiden, kielten ja kulttuurien rajojen yli. Hokkasen tutkimuksen tuloksena syntyvän englanninkielisen käännöksen tavoitteena on myös se, että romaani voidaan julkaista Intiassa. On siis erityisen tärkeää, että kulttuuriviittaukset ja kulttuurinen konteksti heijastavat intialaisten omaa elettyä todellisuutta.

Tässä tutkielmassa tutkittava teos tarjoaa monipuolisen pinnan kulttuurisen kontekstin välittymisen tutkimiselle käännöksessä. Teos sisältää sekä intialaisia että suomalaisia kulttuuriviittauksia. Lisäksi myös lähdeteos sisältää intialaisten puhumien hindin, kannadan ja tamilinkielisiä reaalioita, joita ei lähdetekstissäkään ole käännetty suomeksi. Lähdetekstikin sisältää siis translatorisia valintoja, vaikka teksti ei olekaan käännös. Lähdeteoksen kirjailija on

käyttänyt reaalioita, ja joissakin tapauksissa lisännyt reaalioiden yhteyteen selittävän käännöksen.

Myötäjäiset on tutkielman kannalta monipuolinen teos myös siksi, että käännöstä ei tehdä suoraan lähdekulttuurin kielestä kohdekulttuurin kieleen. Alkuperäinen teos on suomenkielinen romaani, joka käsittelee intialaista kulttuuria, ja sisältää intialaisia ja suomalaisia kulttuuriviittauksia. Teos käännetään englanniksi, joten teoksen kääntäminen lisää vielä kolmannen kielen tähän monikulttuuriseen ja -kieliseen teokseen. Aineiston analyysissa esitellään myös tapauksia, jolloin konekäännin on lisännyt kolmannen kulttuurin viittauksia käännökseen.

Tutkielman teoreettinen viitekehys rakennetaan määrittelemällä ensin kulttuuriviittauksen käsite ja esittelemällä kulttuurin kääntämisen perinteitä ja strategioita. Esittelen aineiston luontiin käytetyn konekääntimen toimintamallin ja kaunokirjallisuuden konekääntämisen nykytilanteen. Tutkielman tarkoituksena ei ole tehdä aineistosta kattavaa laatuarviota, mutta aineistoa kuitenkin tarkastellaan myös laadun näkökulmasta, joten laadun käsite käännösalalla ja siihen liittyvät haasteet esitellään. Myös konekääntimien etiikkaan liittyviä kysymyksiä esitellään, koska ne ovat kytköksissä englannin kielen ja länsimaisen kulttuurin dominanssiin käännösalalla ja konekääntimissä. Dominanssista keskustellaan omassa alaluvussa. Analysoitava materiaali kerättiin aineistosta Excel-taulukkoon, jotta kulttuuriviittausten käännöksiä voidaan helposti vertailla keskenään. Analyysiin valitut kulttuuriviittaukset jaettiin viiteen eri ryhmään ja jokaista ryhmää käsitellään omassa alaluvussaan. Ryhmien välisistä yhtäläisyyksistä keskustellaan tarkemmin tutkielman lopussa.

# 2 TEOREETTINEN VIITEKEHYS

Tutkielman teoreettisessa viitekehyksessä määritellään ensin kulttuurin ja kulttuuriviittauksen käsitteet. Ensimmäisessä alaluvussa esitellään määritelmät, tarkastellaan niitä ja esitellään kulttuurin ja kulttuuriviittausten kääntämiseen liittyviä yleisimpiä strategioita ja haasteita. Toisessa alaluvussa selitetään lyhyesti, miten konekääntimet toimivat, määritellään jälkieditoinnin käsite, sekä esitellään kaunokirjallisuuden konekääntämisen tämänhetkistä tilannetta. Kolmannessa alaluvussa keskustellaan laadun käsitteestä käännösalalla ja siitä, miten konekääntimet vaikuttavat käännösten laatuun. Laadun käsitteen ymmärtäminen käännösalalla on tärkeää muun teoreettisen taustan ymmärtämiseen. Neljännessä alaluvussa keskustellaan konekäänninten eettisyydestä, ja miten se liittyy erityisesti kaunokirjallisuuden kääntämiseen. Viidennessä alaluvussa avataan englannin kielen ja länsimaisen kulttuurin valta-asemaa muihin kieliin ja kulttuureihin nähden käännösalalla.

## 2.1 Kulttuuri

Kulttuuri on käsitteenä vaikea määritellä ja sille löytyykin useampia määritelmiä tutkimusalasta ja kontekstista riippuen. Esimerkiksi TEPA-termipankki määrittelee kulttuurin olevan ”tietyn yhteisön omaksumat ja jakamat arvot, merkityksellistämisen, viestimisen ja tekemisen tavat sekä todellisuutta jäsentävät käsitykset, uskomukset ja tietojärjestelmät”, mutta myös ”kyvyt, käsitykset ja käyttäytymisen tavat, jotka ihmiset ovat omaksuneet yhteiskunnan jäseninä” (TEPA 2026). Määritelmät ovat lähellä toisiaan, mutta tarkemmista selitteistä löytyy vivahde-eroja.

Kulttuurilla usein tarkoitetaan myös ihmiskunnan suurimpia saavutuksia, kuten taideteoksia (TEPA 2026), tai termiä käytetään ilmaisemaan historian tiettyä aikakautta (Katan 1999, 16). Katan (1999) käsittelee kulttuuria kirjassaan ja määrittelee sen yhtenäisenä uskomusten, arvojen ja kognitiivisten ympäristöjen systeeminä, joka ohjaa ja määrittää ihmisryhmien jaettua käyttäytymisperustaa. Systeemin jokainen osa linkittyy toisiinsa ja muodostaa yhtenäisen kulttuurisen kontekstin, jonka avulla ihmisen ja hänen kulttuurinsa voi tunnistaa. (Mts. 17.)

Baldwin (2005) keräsi kirjaansa tiivistelmän kulttuurin määritelmistä 1990-luvun antropologian kirjoista. Näissä määritelmissä kulttuuria kuvaillaan muun muassa kykynä käyttää sovittuja merkkejä, ihmisryhmien tavanomaisena tapana mukauttaa käytöstään

ympäristönsä mukaiseksi sekä jonkin ihmisryhmän sosiaalisesti jakamana tietona ja toimintatapoina. Kulttuuria määritellään myös yhteiskunnan jakamana, opittuna elämäntyylinä ja ajatuksena, sekä tietyn yhteiskunnan opittuina arvoina, uskomuksina, tapoina ja sääntöinä. (Mts. 35–37.) Näistäkin määritelmistä voi todeta, että vaikka yhtäläisyyksiäkin löytyy, kulttuurin käsitteelle ei löydy yhtä yhtenäistä ja tarkkaa määritelmää.

Tässä tutkielmassa kulttuuri määritellään tietyn ihmisryhmän jakamina uskomuksina, perinteinä, tapoina ja arvoina. Monet yllä mainituistakin lähteistä määrittelivät käsitteen tietyn ryhmän jakamana tieto- ja uskomuspohjana, johon ryhmän yhteiset toimintatavat ja perinteet perustuvat. Tämä on merkityksellistä, sillä tässä tutkielmassa analysoidussa lähdeteoksessakin käsitellään tietyn ryhmän (intialaisten) jakamia perinteitä, uskomuksia, tapoja ja arvoja (esimerkiksi avioliittoperinteitä). Lisäksi teoksessa nähdään jonkin verran eri ryhmien ja kulttuurien vastakkainasettelua, kun suomalaisesta kulttuurista kotoisin oleva antropologi alkaa tutustumaan intialaiseen kulttuuriin.

### 2.1.1 Kulttuuriviittauksen määritelmä

Kulttuuriviittauksen määritelmä voi vaihdella eri konteksteissa. Tässä tutkielmassa kulttuuriviittaus määritellään terminä, lausekkeena tai virkkeenä, joka ilmaisee tai kommentoi tiettyyn kulttuuriin sidonnaisia tapoja, asioita tai esineitä, käsityksiä ja arvoja. Tutkielman kontekstissa nämä kulttuurit ovat tutkittavassa teoksessa esiintyvät intialainen kulttuuri ja suomalainen kulttuuri. Kulttuuriviittausta kutsutaan myös kulttuurisidonnaiseksi viittaukseksi tai reaaliaksi, jonka TEPA-termipankki määrittelee ”sanaksi tai fraasiksi, joka kuuluu elimellisesti lähtökulttuurin käsitemaailmaan” (TEPA 2025).

Reaaliat ovat kulttuurisidonnaisia lingvistisiä osia tekstissä. Reaaliat ovat jollekin yhteiskunnalle erityisiä kulttuurisia ilmiöitä ja termejä, joihin voi kuulua materiaalisia asioita, kuten paikallisia ruokia, tietyn alueen luontoon kuuluvia asioita, kuten kasveja tai eläimiä, sekä ilmauksia, murteita ja tapoja. (Schiopu 2025.) Ritva Leppihalme (2011, 126) kertoo, että laajemmassa merkityksessä reaaliaksi voidaan katsoa kulttuurisidonnaiset ilmiöt, kuten esimerkiksi uskonnolliset käsitteet, arvot ja instituutiot. Reaalioita voidaan jakaa kategorioihin niiden tyyppien mukaan, mutta tarkkaa lajittelua ei voida kulttuurien välisten eroavaisuuksien takia tehdä. Siksi reaaliakategoriat ja niiden määritelmät voivat vaihdella tutkimusten välillä.

(Leppihalme 2011, 127.) Reaaliat ovat käännösprosessin kannalta hankalia, koska niille ei useimmiten löydy suoraa vastinetta kohdekielestä (Schiopu 2025).

Yksi Myötäjäiset-teoksen keskeisimmistä teemoista on naisten asema Intian avioliittomarkkinoilla, mikä on suoraan kytköksissä kulttuuriin, ja siitä käytävät keskustelut voidaan lukea kulttuuriviittauksiksi. Teoksessa käydään paljon keskustelua avioliitto- ja perhedynamiikoista Intiassa. Keskustelua käydään myös suomalaisen perheen ja parisuhteen dynamiikoista. Nämä katkelmat suoraan kommentoivat lähtökulttuurejaan, ja niitä tutkitaan kulttuuriviittausten kontekstissa tässä tutkielmassa. Kulttuureissa ilmeneviä valtasuhteita usein esitetään, vahvistetaan ja muodostetaan kaunokirjallisissa teksteissä tämänkaltaisten representaatioiden tai kulttuuristen käsitteiden avulla (Abrams & Harpham 2019, 245). Tutkittavassa teoksessa käsitellään muun muassa intialaista perinnettä, jossa vanhemmat järjestävät avioliiton tyttärilleen. Perinnettä koskevat katkelmat ovat keskeisessä roolissa teoksessa ja viittaavat suoraan lähtökulttuuriin. Siksi näitä katkelmia sisällytetään kulttuuriviittauksen määritelmään ja tutkitaan aineistossa. Myös esimerkiksi maininnat suomalaisesta kouluttautumisesta luetaan tässä tutkielmassa kulttuuriviittauksiksi, koska ne ovat suoraan yhteyksissä suomalaiseen koulutukseen ja korkeakoulukulttuuriin. Lisäksi analysoitavana on materiaalisia reaalioita, kuten ruokalajeja, vaatteita ja valuuttoja. Jotkin tutkittavat kulttuuriviittaukset liittyvät myös muihin kuin suomalaiseen tai intialaiseen kulttuuriin. Näitä viittauksia tarkastellaan englannin kielen ja länsimaisen kulttuurin dominanssin näkökulmasta.

### 2.1.2 Kulttuurin kääntämisen perinteestä

Kulttuuri ja kääntäminen kulkevat aina käsi kädessä. Kulttuuri voidaan nähdä viitekehyksenä, jonka sisällä kaikki kommunikaatio tapahtuu (Katan 1999, 241). Kääntämisen sen sijaan on ymmärretty olevan elintärkeää kulttuurien väliselle kommunikoinnille. Käännösten voidaan nähdä rakentavan kulttuureja, sillä ne siirtävät erilaisten strategioiden avulla kulttuurista sisältöä toiselle kielelle ja toiseen kulttuuriin niin, että lähtökulttuurista rakentuu ymmärrettävä ja toimiva käsitys myös kohdekielessä ja -kulttuurissa. Käännökset siis rakentavat maailman kulttuurillista pääomaa toimimalla siltana eri kulttuurien välillä. On myös tärkeää ymmärtää, että käännökset toimivat monille ihmisille ”alkuperäisinä” teoksina, koska tekstien lukijat harvoin lukevat saman teoksen sekä käännettynä että alkuperäisellä kielellä. (Bassnett & Lefevere 1998, 6–9.) Käännökset toimivat siis tärkeässä roolissa kulttuurien välisessä

kanssakäymisessä. Kääntäminen-termiä käytetään ajoittain myös laajemmassa, filosofisessa merkityksessä. Sillä viitataan semioottisten tai hermeneuttisten käsitteiden kääntämiseen, eikä vain lingvististen ongelmien ratkomiseen. (Carbonell 1996, 79.) Kääntäminen voidaan siis itsessään nähdä moniulotteisena kulttuurien välisenä kommunikointina perinteisen kielen kääntämisen lisäksi. Kääntäminen on kulttuurien välisen kommunikoinnin tärkeä työkalu, kun taas kulttuuri, ja ennen kaikkea sen ymmärtäminen sekä lähde- että kohdekielessä, on merkittävä osa käännösprosessia.

Kulttuurin ymmärtäminen on jo pitkään nähty yhtenä tärkeänä kompetenssina kääntäjän työssä. Käännettävä teksti ei ole yksitasoinen asia, jonka voi jakaa pienempiin osiin, vaan tekstin osat ovat kytköksissä toisiinsa sekä tekstin ulkopuolisiin tekijöihin. Kääntäminen on myös monitahoinen viestinnän tehtävä, jonka tavoitteena on viestiä lukijalle jotakin eri kielestä ja kulttuurista tämän omalla kielellä. Käännöstyöt vaativat siis tiettyjä ammattimaisia kompetensseja, jotta kääntäjä voi suoriutua tehtävästä onnistuneesti. Kääntäjien kompetensseja on määritelty ja mitattu esimerkiksi PACTE- ja EMT-mallien avulla. PACTE-malli tulee sanoista *Process in the Acquisition of Translation Competence and Evaluation*. PACTE-tutkimuksilla on pyritty tutkimaan ja kartoittamaan näitä vaadittavia kompetensseja, erityisesti liittyen kääntäjien työhön kulttuurien välisinä viestijöinä. (Albir 2017, 3–4.) PACTE-mallissa kulttuurien ymmärtäminen on jaoteltu osaksi kielen osaamisen ulkopuolisia kompetensseja (*extralinguistic competence*). Mallin mukaan nämä kompetenssit eivät liity suoraan kielitaitoon tai kielen osaamiseen, mutta ovat tärkeä osa kääntäjän kompetensseja, jotta kääntäjä voi tehdä onnistuneita käännöksiä. Näihin kompetensseihin kuuluu sekä yleinen tietämys että alakohtainen tietämys kahdesta eri kielestä ja kulttuurista, joiden välillä kääntäjä kommunikoi. (Albir 2017, 40.)

EMT (*European Master's in Translation*) taas on Euroopan komission luoma laatumerkintä maisteritason käännöskoulutusohjelmille. EMT-merkintä tarkoittaa sitä, että käännöskoulutusohjelma noudattaa sovittuja ammattivaatimuksia ja nykyisiä markkinatarpeita. Mallin tavoitteena on varmistaa, että koulutusohjelmista valmistuvilla kääntäjillä on mahdollisimman hyvät valmiudet ammattimaiseen toimintaan käännösalalla. (European Commission 2026.) Tällä hetkellä käytössä oleva EMT-malli astui voimaan vuonna 2022. Malli on ensimmäisen kerran luotu jo vuonna 2007, mutta sitä on päivitetty muun muassa vastaamaan kielialoilla tapahtuneiden teknologisten muutosten luomia uusia vaatimuksia ammattikääntäjille. EMT-mallissa on määritelty viisi kääntäjän kompetenssia, joiden oppimista

maisteritason kääntäjien koulutusohjelmien tulisi tukea. Kieli ja kulttuuri on lajiteltu mallissa yhdeksi kompetenssiksi, jonka sanotaan olevan jokaisen muun kompetenssin taustalla toimiva voima. Kääntäjällä tulisi olla kielen osaamisen ja tuntemuksen lisäksi kyky tunnistaa kulttuurisia elementtejä, arvoja ja viittauksia tekstistä, sekä kyky kirjoittaa kulttuuristen konventioiden mukaisesti. (EMT Competence Framework 2022, 6.)

### 2.1.3 Strategiat

Kääntäjät voivat hyödyntää kulttuuriviittausten kääntämiseen erilaisia strategioita, mutta strategian valintakin on kulttuurikohtaista. Esimerkiksi lähdekieliset reaaliat saatetaan poistaa kokonaan kääntäessä, mutta tätä strategiaa hyödynnetään yleensä vain kulttuureissa, joissa vahva adaptointi ja poisto ovat hyväksyttäviä normeja. (Leppihalme 2011, 128–129.) Poistostrategia voidaan nähdä myös ongelmallisena, koska sen koetaan ajoittain muuttavan lähdetekstin kulttuurista sisältöä liikaa (Sun 2012). Englannin kielellä on käännösalalla sekä konekäänninteknologiassa valta-asema muihin kieliin verrattuna. Lisäksi länsimaiset kulttuurit kantavat kulttuurista johtoasemaa. (Englannin ja länsimaisen kulttuurin valta-asemasta lisää luvussa 2.4.1.) Tämä näkyy käännösalalla siten, että esimerkiksi aasialaisia tekstejä on usein kotoutettu niin paljon, että lukija ei välttämättä enää koe niiden olevan aasialaista alkuperää. (Yoon 2021.) Jotkut kääntäjät ovatkin lähteneet dekolonisoimaan käännösprosessia hyödyntämällä käännösstrategioita, joiden avulla käännösten alkuperäistä kulttuuria ja kieltä ei tarvitse häivyttää yhtä paljon. Yoon (2021) tutkii artikkelissaan dekolonisoivaa kääntämistä ja mainitsee kääntäjien tällaisissa tapauksissa suosivan vieraannuttavia käännösstrategioita kotouttamisen sijaan. Vieraannuttavilla käännösstrategioilla tarkoitetaan strategioita, jotka tarkoituksellisesti etäännyttävät käännöstä sen kohdekielen kulttuurista ja pitävät sen kulttuurisen sisällön lähempänä lähdekulttuuria. Yoon nimeää esimerkiksi käännöslainat, selitysten lisäämisen ja lisätekstien käyttämisen strategioina, joita kääntäjät käyttävät halutessaan hyödyntää vieraannuttavia käännösstrategioita. Esimerkiksi käännöslainojen käyttö ja sanasanaiset käännökset ovat yleisiä kulttuuriviittausten käännösstrategioita (Yoon 2021; Leppihalme 2011).

Käännöslainojen käyttö on käännösstrategia, jossa lähdetekstin termejä käytetään myös käännöksessä. Näin lähdetekstin kulttuurista sävyä saadaan sisällytettyä myös kohdekielisessä käännöksessä. Jotkin vanhemmat käännöslainat ovat jo niin vakiintuneita, ettei niitä nähdä enää käännöslainoina ollenkaan. Tällaisia ovat esimerkiksi alun perin ranskan kielestä englannin

kieleen lainattu termi *menu*. (Vinay & Darbelnet 1995, 31–32.) Myötäjäiset-teoksessa esiintyviin käännöslainoihin voidaan luetella esimerkiksi hindin kielestä juontuva termi *bindi*. Koska termille ei löydy suoraa englanninkielistä vastinetta, termi nähdään nykyään myös englanninkielisenä terminä esineelle, vaikka sen juuret ovat hindin kielessä.

Sanasanainen käännösstrategia, eli *calquing* on hyödyllinen kulttuurillisesti rikkaiden tekstien kääntämisessä. *Calquing* tarkoittaa käännöslainaa, jolloin lähdetekstin ilmausta niin sanotusti lainataan käännökseen kääntämällä sen jokainen elementti suoraan. Lainat voivat olla joko leksikaalisia tai rakenteellisia. Leksikaaliset lainat noudattavat kohdekielen syntaktista rakennetta, mutta tuovat uuden ilmaisun kohdekieleen lähdekielestä, kun taas rakenteelliset lainat käyttävät kokonaan uutta rakennetta kohdekielessä. Sanasanaiset käännöslainat voivat myös muuttua osaksi kohdekieltä, kun niiden käyttö yleistyy ja vakiintuu. (Vinay & Darbelnet 1995, 32.)

Kääntäjät voivat myös suoraan lisätä kääntämiensä kaunokirjallisten tekstien yhteyteen esimerkiksi johdantotekstin, jossa he esittelevät käännösprosessiaan, lähdetekstin kulttuuria tai lähdekielen konventioita (Yoon 2021). Myös itse käännökseen voidaan lisätä selittäviä käännöksiä. Lähdekielisen reaalian käyttö sen alkuperäisessä muodossa on yleistä, ja esimerkiksi erisnimet usein kirjoitetaan alkuperäisessä muodossa myös käännökseen. Muitakin reaalioita voidaan käyttää käännöksessä lähdemuodossa, mutta ne saatetaan erottaa muusta tekstistä esimerkiksi kursiivilla. Jos käännöksessä käytetään lähdekielistä reaaliaa, sen yhteyteen saatetaan lisätä selitys, jotta reaalian konteksti selviää kohdekieliselle lukijalle. Leppihalme käyttää tästä esimerkkiä *Tuonela*, jonka yhteyteen on englanninkieliseen käännökseen lisätty selittävä käännös: *Tuonela, the Land of the Dead*. (Leppihalme 2011, 128–129.) Tällaisia esimerkkejä löytyy myös Myötäjäiset-teoksesta. Aura esimerkiksi käyttää tekstissään sanaa *akka*, joka tarkoittaa tamilin kielellä siskoa ja kannadan kielellä vanhempaa siskoa. *Akka*-huudahduksen perään on lähdetekstiin lisätty selittävä käännös *isosisko*, jolloin lukija ymmärtää, mitä sana tarkoittaa. Tämä on esimerkki siitä, miten Aura on jo lähdetekstiään tehdessä tehnyt translatorisia päätöksiä, vaikka hänen kirjoittama tekstinsä ei olekaan käännös vaan lähdeteksti.

### 2.1.4 Haasteet

Kulttuurin kääntäminen tuo mukanaan moniulotteisia haasteita kääntäjälle. Yksi näistä on se, että tietyt käsitteet tai reaaliat ovat niin tiukasti kiinni lähdekulttuurissa ja -kielessä, ettei niille löydy suoraa vastinetta kohdekulttuurista tai -kielestä (Fernández Guerra 2012). Aixelá (1996, 57) jopa totesi artikkelissaan, että kulttuurisidonnaisia käsitteitä ei itsessään löydy teksteistä, vaan niiden olemassaolo on sidonnainen käännösprosessissa syntyvään lingvistiseen ongelmaan, joka ilmenee niiden vastineettomuudesta kohdekielessä.

Kulttuurin kääntäminen on myös nostattanut kysymyksiä siitä, voiko tietynlaisia tekstejä kääntää ja vaikuttaako kääntäminen tekstien luettavuuteen. Perinteisen ymmärryksen mukaan käännösprosessi muuttuu haastavammaksi, mitä kauempana lähdetekstin ja kohdetekstin kulttuurit ovat toisistaan. Silloin käännöksen luoma silta kulttuurien välillä voidaan nähdä myös erottavana tekijänä, mikäli se vahvistaa haitallisia stereotypioita. (Carbonell 1996, 83.) Yifeng Sun (2012) pohti tutkimuksessaan, ovatko jotkut tekstit täysin kääntämättömissä, sekä sitä, miten laadukkaina käännökset voidaan nähdä, jos ne eroavat liikaa lähdetekstistä. Sun huomauttaakin, että usein tekstien käännettävyys ja täten luettavuus mahdollistetaan laadun kustannuksella, erityisesti kaunokirjallisuudessa (mts. 232). Tämä tarkoittaa sitä, että käännöksestä pyritään tekemään hyväksyttävä lukijalle, sen sijaan että varmistetaan laadukas käännös, joka sisältää kaiken tai lähes kaiken saman tiedon ja kulttuurisen sisällön kuin lähdeteksti. Sun käyttää artikkelissaan esimerkkinä kiinalaista romaania, josta kääntäjä on ensimmäisessä käännöksessä luettavuuden ja ymmärrettävyyden nimissä poistanut paljon tekstiä ja jopa kokonaisia kappaleita. Myös joitakin kiinalaisen kulttuurin termejä on täysin häivytetty tai Sunin mielestä käännetty harhaanjohtavasti, jolloin lukija ei voi saada tekstistä ja sen kulttuurisesta sisällöstä samanlaista ja yhtä kattavaa kuvaa, kuin alkuperäisen kiinankielisen tekstin lukija. Hänen mukaansa tekstin käännettävyys silloin on kyseenalaista, ja voitaisiin argumentoida, ettei teksti ole käännettävissä. Laadun käsite on kuitenkin käännösalalla hyvin vaihtelevaa. Laadun käsiteestä ja sen määrittelystä tässä tutkielmassa keskustellaan enemmän alaluvussa 2.3.

Vaikka kulttuurin ja kulttuuriviittausten kääntäminen on haasteellista, Sun (2012) mainitsee sen olevan myös kiehtovaa. Lisääntyvä kommunikointi kulttuurien välillä luo uusia strategioita ja mahdollisuuksia kääntää sellaisiakin tekstejä, joiden käännettävyys on aiemmin ollut kyseenalaista. Sun lisäksi huomauttaa, että kulttuurinen merkitys on kuitenkin siirrettävissä kulttuurista toiseen, vaikka se ei suoranaisesti olisi käännettävissä lähdekielestä kohdekieleen.

Ongelmallista siitä tulee vain, jos siirto tehdään umpimähkäisesti. (Mts. 246.) Lingvistisen ja viestinnällisen näkökulman mukaan mikä tahansa ihmismielen kuvittelema voidaan ilmaista millä tahansa kielellä. Kaikki tekstit ovat siis käännettävissä, vaikka joitakin muutoksia jouduttaisiin tekemään kielten erilaisuuksien takia, koska eri kielten evoluutiot eivät kulje yhtäaikaisesti. (Kade 1981, Fernández Guerran 2012 mukaan.)

## 2.2 Kaunokirjallisuuden konekääntäminen

### 2.2.1 Konekääntimet ja neuroverkkopohjainen konekääntäminen

Konekäänninteknologiaa on kehitetty jo 1950-luvulta asti, joten erilaisia järjestelmiä on ajan myötä kehittynyt useampia. Nykyään johtavassa asemassa ovat suurten kielimallien (LLM) konekääntimet, mutta neuroverkkokonekääntimet ovat myös yleisessä käytössä. Neuroverkkopohjaisten, eli Neural Machine Translation (NMT), konekääntimien idea perustuu jo 1940- ja 1950-luvulla syntyneeseen ideaan siitä, miten ihmisten aivot toimivat muodostaessaan neuroneista koostuvia verkkoja. Teknologia, jolla tämä idea pystyttiin toteuttamaan, syntyi kuitenkin huomattavasti myöhemmin, mutta 2010-lukuun mennessä markkinoille olivat saapuneet ensimmäiset keinotekoiset neuroverkot. Siitä eteenpäin teknologia alkoi kehittyä nopeammin ja tehokkaammin, ja esimerkiksi puhetta tekstiksi muuttavat teknologiat kehittyivät huomattavasti 2010-luvulla. Neuroverkkokonekäännin on tekoälypohjainen konekäännin, joka laskee todennäköisimmän käännöksen koulutusdatan pohjalta. NMT-kääntimiä koulutetaan syöttämällä niihin rinnakkaistekstejä, joiden pohjalta ne alkavat luoda keinotekoista neuroverkkoa. Nykyajan malleissa koulutusdatana toimii usein myös valmiiksi syötetyn datan lisäksi koneen internetistä imuroima tieto, kuvat ja videot. NMT-käännin laskee sanoille tai fraaseille numeraalisen arvon, jonka avulla se laskee sopivimman käännösvalinnan koulutusdatansa pohjalta muodostamastaan neuroverkosta. Koska NMT-kääntimet pystyvät laskelmoimaan lingvististen seikkojen lisäksi myös kontekstia rinnakkaisdatan perusteella, tuloksena voi olla hyvinkin sujuva käännös. (Rothwell ym. 2023, 102–104.) Vaikka konekääntimen tekemä käännös usein vaikuttaa sujuvalta, konekäännin on voinut luoda niin sanottuja hallusinaatioita käännökseen. Hallusinaatio on konekäännöksissä ajoittain esiintyvä ilmiö, jolloin kone on kääntänyt jonkin termin tai fraasin niin, että se voi vaikuttaa ensisilmäyksellä luonnolliselta vaikkei se ole uskollinen lähdetekstille ja sen kontekstille. Konekäännin siis pyrkii laskelmoimaan termille tai fraasille sopivan käännöksen

esimerkiksi jostain sen osasta, mutta päätyy lopulta käännökseen, jolla ei ole minkäänlaista yhteyttä lähdetekstiin ja voi olla jopa kieliopillisesti väärin (Rothwell ym. 2023, 106). Hallusinaatiot alentavat konekäännösten laatua muuttaessaan lähdetekstin kontekstia, ja ne voivat lisäksi aiheuttaa turvallisuusriskejä. Konekäännin voi esimerkiksi hallusinoida tietoturvan suojaamia henkilötietoja käännöksiin, joiden lähdeteksteissä tiedot eivät esiinny lainkaan. (Ji ym. 2023.)

On kuitenkin huomioitava, että NMT-kääntimien tuottamien käännösten laatu ja sujuvuus on pitkälti riippuvaista siitä, millaista koulutusdataa ne käyttävät (Rothwell ym. 2023, 103). Jos koulutusdata ei sisällä käännettävän tekstin tyylilajin kaltaisia tekstejä, voi käännöksen tyyli muuttua huomattavastikin. Myös hallusinaatioiden ilmeneminen konekäännöksissä johtuu usein koulutusdatasta johtuvista ongelmista, kuten sen suppeudesta tai sen sopimattomuudesta lähdetekstin kääntämiseen (Rothwell ym. 2023, 106). Myös kielten välistä epätasa-arvoa voi ilmentyä NMT-käännöksissä. Englannin kielellä on huomattava etulyöntiasema konekäännöksissä, sillä englanninkielistä koulutusdataa löytyy huomattavasti enemmän kuin monien muiden kielien dataa. Tämä johtaa usein myös kulttuurisiin ja kontekstuaalisiin haasteisiin käännöksissä, sillä kääntimet tunnistavat globaalin pohjoisen kulttuurien kontekstit globaalin etelän kulttuureja helpommin. Kulttuurien välistä epätasa-arvoa konekääntimissä käsitellään lisää luvussa 2.4.1.

### 2.2.2 Jälkieditointi

Jälkieditoinnilla viitataan siihen, että konekääntimellä luodun käännöksen oikolukee ja editoi ihmiskääntäjä. Termillä on alun perin eroteltu konekäännösten editointi ihmisen kääntämän tekstin editoinnista, koska niissä esiintyvät virheet ovat hyvin erilaisia. Jälkieditointi-termi on kuitenkin koettu ristiriitaiseksi ammattilaisten keskuudessa konekääntimien sekä CAT-työkalujen (Computer Assisted Translation) kehityttyä. Konekääntimiä hyödyntävää käännösprosessia voidaan kutsua myös *koneavusteiseksi kääntämiseksi*, mutta tämäkin voi olla harhaanjohtavaa. Termistö aiheuttaa kiistanalaisuutta, koska konekääntimien ja vastaavien työkalujen käyttöä käännösprosessissa käytetään usein syynä leikata ammattikääntäjien palkkioita. (O'Brien 2021.) Vaikka termi on hieman kiistanalainen, *jälkieditointi* on kuitenkin alalla vallitseva nimitys tällaiselle käännösprosessille, joten selvyyden säilyttämiseksi se on tässä tutkielmassa käytettävä nimitys.

Daems, Clerq ja Macken (2017) pohtivat artikkelissaan konekäännösten, ihmiskääntäjien tekemien käännösten ja jälkieditoitujen käännösten eroa. Jälkieditoinnilla pyritään yleensä varmistamaan konekäännösten laadukas lopputulos. Konekäännökset ovat edelleen usein ammattikääntäjän alusta asti tuottamia käännöksiä huonolaatuisempia. Artikkelissa kuitenkin huomautettiin, että jälkieditoidut konekäännökset eivät eroa laadullisesti ihmisten kääntämistä teksteistä juurikaan. Daems, Clerq ja Macken (2017) huomasivat lukijoiden kuitenkin suosivan ihmisten alusta asti kääntämiä tekstejä jälkieditoituja käännöksiä enemmän ja olettivat syynä olevan jälkieditoiduissa teksteissä esiintyvä omanlaisensa ääni, josta he käyttävät englanninkielistä nimitystä *post-editise*. Jälkieditoidun tekstin ääni voidaan rinnastaa teoriaan kirjailijan ja kääntäjän äänestä, jotka tulee huomioida muun muassa konekäännösten eettisiä seikkoja tutkiessa. Tekstien eri äänistä ja niiden eettisistä haasteista keskustellaan lisää alaluvussa 2.4.

### 2.2.3 Konekääntäminen ja kaunokirjallisuus

Kaunokirjallisuuden kääntäminen on yksi käännösalan haaroista, jossa konekääntimien käyttö ei ole vielä yleistynyt samalla tasolla kuin esimerkiksi asiatekstien kääntämisessä. Kolb ja Miller totesivat tutkimuksessaan, että heidän tutkimuksensa julkaisun aikaan konekääntimet eivät olleet soveltuvia kaunokirjallisuuden kaltaisten luovien tekstien kääntämiseen (Kolb & Miller 2022). Myös Guerberof-Arenasin ja Toralin (2020) artikkelissa todettiin, etteivät konekääntimet sovi kaunokirjallisuuden kääntämiseen, tosin he huomauttivat kiinnostuksen konekääntimien ja vastaavien käännösteknologioiden käyttämiseen luovassa kääntämisessä nousseen. Kiinnostus kaunokirjallisuuden konekääntämiseen nousi uudelleen pinnalle NMT-konekääntimien myötä. Niiden koettiin olevan aiempia konekääntimiä lupaavampia, koska ne pystyvät jossakin määrin huomioimaan myös kontekstin käännösvalinnoissaan. (Taivalkoski-Shilov 2019, 690.)

Jotkut kaunokirjallisuuden kääntäjät ovat todenneet konekääntimien ja CAT-työkalujen olevan kuitenkin hyödyllisiä käännösprosessissa myös kaunokirjallisuutta kääntäessä. Ne voivat auttaa esimerkiksi toistuvuuden välttelemisessä pitkissä teksteissä. Usein konekääntimet ja erityisesti käännösmuistiohjelmat edesauttavat myös tekstin johdonmukaisuuden ylläpitämisessä. Tässä on kuitenkin edellytyksenä se, että tekstin kääntämisprosessia varten on kääntäjälle annettu ja/tai luotu yksi yhtenäinen käännösmuisti työtä varten. (Taivalkoski-Shilov 2019.)

Kolb ja Miller tutkivat CAT-työkalujen soveltuvuutta luovien tekstien kääntämiseen tutkimalla sanaleikkien kääntämiseen luotua PunCAT-sovellusta. Englanninkielinen pun-termi tarkoittaa sanaleikkejä, joissa sanalla tai fraasilla viitataan niiden alkuperäisestä merkityksestä poikkeavaan samanlaiseen sanaan tai fraasiin. PunCAT on tutkimusta varten luotu käännöstyökalu, joka poimii tekstistä kaikki sanaleikit. Työkalu sitten tulkitsee sanaleikit ja tarjoaa sekä selityksen että erilaisia käännösvaihtoehtoja sanaleikeille. Tutkimuksessaan Kolb ja Miller totesivat, ettei sovellus toiminut varsinaiseen kääntämiseen kovinkaan hyvin. Ammattilaiset kuitenkin totesivat, että työkalua voi hyödyntää ideoimiseen ja luovaan työhön apuvälineenä, vaikkei se tarjonnut apua varsinaiseen kääntämiseen. (Kolb & Miller 2022.) Heidän tutkimuksensa tulokset osoittavat, että kääntäjät voivat hyödyntää CAT-työkaluja myös luovien tekstien kääntämiseen, mutta lopullisen käännöstyön tekemiseen tarvitaan kuitenkin ihmiskääntäjä, sillä konekääntimet ja CAT-työkalut eivät tuota toimivia luovia käännöksiä itsenäisesti.

Guerberof-Arenas ja Toral (2020) puolestaan vertailivat sitä, miten lukijat reagoivat konekäännettyihin käännöksiin verrattuna ihmisten tekemiin käännöksiin. Tutkimuksessa vertailtiin myös ihmisen jälkieditoimia konekäännöksiä. Tutkimuksen tulokset paljastivat, että lukijat pitivät huomattavasti enemmän käännöksistä, joissa ihmiskääntäjillä oli jokin osa. Kuten Daems, Clerq ja Macken (2017) artikkelissaan myös totesivat, lukijat vaikuttavat reagoivan paremmin ja suosivan käännöksiä, joiden tekemiseen ihmiskääntäjät ovat osallistuneet. Konekäännökset myös saivat alhaisemmat pisteet luovuudesta kuin ihmisen kääntämät tai jälkieditoimat käännökset. Tutkimuksessa siis todettiin, että ihmiskääntäjät tuovat erityisesti kaunokirjallisuuden kääntämisessä vaadittavaa luovuutta käännösprosessiin, jota nykyinen käännösteknologia ei pysty prosessille tarjoamaan. (Guerberof-Arenas & Toral 2020.)

## 2.3 Laadun käsite käännösalalla

Käännösten laatu ja niiden arviointi on suuresti kiistelty käsite käännöstieteen alalla. Ajoittain jopa yhteistyössä aiheeseen liittyvää tutkimusta toteuttavat tutkijat näkevät asian eri näkökulmista. Kuitenkin yksimielisyyttäkin löytyy, sillä kaikkien käännösten voidaan olettaa pitävän sisällään jonkin asteista täsmällisyyttä ja sujuvuutta. (Koby ym. 2014.) Laadun käsitettä voidaan tarkastella myös ekonomisesta lähtökohdasta. Gouadecin (2010) määritelmän mukaan käännöstransaktio on hyvä, kun sekä käännöksen tilaaja että kääntäjä ovat tyytyväisiä käännösprosessiin ja sen tulokseen. Tämän määritelmän syvällisempi olettamus on, että

kääntäjä noudattaa tarpeellisia toimintatapoja laadukkaan lopputuloksen varmistamiseksi. Määritelmän oletuksena on myös se, että käännöksen tilaaja on osallisena onnistuneeseen käännösprosessiin toimittamalla tarvittavat raakamateriaalit, lähdetiedot ja ohjeistukset käännöksen tekemistä varten. Hän kuitenkin toteaa tämän viimeisen kohdan olevan ennemmin toiveajattelua.

Käännöstuotteen laadun määrittelyä hankaloittaa se, että laadun käsitys voi olla erilainen riippuen siitä, tutkitaanko sitä esimerkiksi asiakkaan vai myyjän näkökulmasta. Tämä on erityisen hankalaa käännösalalla, jossa ”asiakasta” voi olla vaikea määritellä. Käännöksillä on sen tilaaja, mutta yleensä myös lopullinen käyttäjä, eli lukija tai useammat lukijat. (Koby ym. 2014.) Ideaalitilanteessa laadukas käännös siis tyydyttäisi niin kääntäjää, käännöksen tilaajaa kuin myös käännöksen lopullisia käyttäjiä.

Laatu ymmärretään myös monella eri tavalla, minkä vuoksi käännösalan tutkijoiden on ollut vaikeaa päästä yhteisymmärrykseen sen mittaamiseen sopivista parametreista. Yhteisymmärrykseen pääsemistä vaikeuttaa niin käännöksessä huomioon otettavat eri tekijät, kuten kulttuuriset ja sosiaaliset tekijät, kuin myös tutkijoiden eri näkemys kääntämisestä ja sen teorioista. Usein käännösten laadun arvioinneissa heijastuukin arvioijan omat teoriat kääntämisestä. (Colina 2011.) Kaiken kattavaa laadun arviointia on myös vaikea määritellä, koska käännösten vaatimukset ovat tilannekohtaisia, eli sama käännös voi olla laadukas siltä vaadittuun viestintätilanteeseen, mutta huonolaatuinen jossakin muussa tilanteessa (Gouadec 2010).

Käännösten laatua varten löytyy kuitenkin ohjeita ja säädöksiä, joilla pyritään takaamaan tietyn laatutason toteutuminen. Käännöspalveluita varten on muun muassa luotu ISO-standardit, jotka määrittelevät laatuvaatimukset käännöspalveluille ja joita suuri osa suomalaisistakin käännöspalveluiden harjoittajista noudattaa (SFS-EN ISO 17100:2015 + A1:2017). ISO-standardeja noudattamalla käännöspalvelun tarjoaja voi todistaa, että hänen resurssinsa ja prosessinsa ovat riittävät täyttämään asiakkaiden palvelumääritykset. Jälkieditoitujen konekäännösten käyttäminen ei kuitenkaan kuuluu käännöspalveluja koskevan standardin soveltamisalaan.

On kuitenkin huomattava, ettei käännöksissä voida koskaan taata sataprosenttista laatua ja virheettömyyttä. Täydellistä käännöstä ei voida taata, koska eri kielet eivät ole täydellisiä toistensa kopioita. Käännöksen kieliparista ei aina löydy täydellisiä vastakappaleita jokaiselle

sanalla ja fraasille, ja käännettäessä on lingvististen seikkojen lisäksi huomioitava muut asiat, kuten lähtö- ja kohdetekstin kulttuurit ja niiden erot. (Koby ym. 2014.) Lisäksi jopa ISO-standardeja noudattava ja kompetentti kääntäjä voi tehdä virheitä. Käännöksiä tilatessa myös usein huomioidaan ekonomiset tekijät ennen laatua. Käännös on tärkeää saada nopeasti ja kustannustehokkaasti, usein laadun kustannuksella. Mikäli kääntäjille annettaisiin aikaa hioa käännökset ”täydellisiksi”, käännösten hinnat nousisivat markkinatilanteeseen nähden kohtuuttomiksi. (Gouadec 2010.)

Käännösten laadun määritteleminen hankaloituu entisestään konekäännöksiä käsitellessä. Kuten ihmiskääntäjien tekemissä käännöksissä, konekäännöksiäkin tarkastellessa pitää ensimmäisenä määritellä, millaiseen käyttöön käännös on tarkoitettu ja minkälaisia vaateita tarkoitus mahdollisesti antaa konekäännökselle. Konekäännösten käyttötarkoituksen voi karkeasti jaotella kahteen ryhmään: konekäännös tekstin pääasiallisen viestin ymmärtämiseen ja konekäännös jälkieditointia varten. Ensimmäinen käyttötarkoitus viittaa siihen, että konekäännöksen on tarkoitus tarjota lukijalleen mahdollisuus ymmärtää jokin vieraskielinen teksti pääpiirteittäin. Tällöin konekäännöksen ei tarvitse olla ”täydellinen”, kunhan siitä ymmärtää tekstin tärkeimmät piirteet. Toinen tarkoitus taas on luoda konekäännös, jota ihmiskääntäjä voi käyttää pohjana tai raakakäännöksenä, josta hän sitten muokkaa julkaistavan käännöksen. (Rothwell ym. 2013, 97.) Näiden kummankin käyttötarkoituksen lopulliset käännökset siis ovat laadullisesti huomattavasti toisistaan eroavia, vaikka käännöksen lopullisen lukijan kannalta laatu voitaisiin nähdä samantasoisena. Jos tarkoitus on ymmärtää tekstistä vain pääpiirteet, suora konekäännös on sopiva, kun taas esimerkiksi julkaistavaan artikkeliin tai kirjaan käännös monessa tapauksessa olisi liian huonolaatuinen. Laadun käsite siis konekäännöksissä voi vaihdella huomattavastikin riippuen käännöksen käyttötarkoituksesta.

Konekääntimillä on yleisesti ymmärretty olevan negatiivinen vaikutus käännösten ja käännösalan laatuun. Kuten luvussa 2.2.1 selitettiin, sekä NMT-konekääntimet että LLM-konekääntimet tekevät käännösvalintoja niiden saatavilla olevan koulutusdatan pohjalta. Tämä voi olla ongelmallista erityisesti kaunokirjallisia tekstejä kääntäessä, sillä koulutusdata ei välttämättä sisällä rinnakkaistekstejä, joista ammentaa lähdetekstin tyylille sopivia käännösvalintoja. On myös huomioitava, että koneen käyttämä data ei välttämättä ole itsessään laadukasta. Konekäänninten käyttämät rinnakkaistekstit voivat yhä useammin olla konekäännöksiä, jolloin kone imitoi itseään. Vaikka rinnakkaistekstit olisivat ihmiskääntäjien

luomia käännöksiä, ei niidenkään laatua voida taata. Ihminenkin voi tehdä virheitä, eivätkä kaikki saatavilla olevat käännökset ole virheettömiä, vaikka ne olisivat ihmiskääntäjän käsialaa. (Rothwell ym. 2023, 106.) Ideaalitilanteessa siis koulutusdataa kootessa pystyttäisiin keskittymään sekä laatuun että määrään.

Toisaalta koulutusdatan laajuus voi olla kaunokirjallisuutta kääntäessä myös haitaksi. Kun konekäännin ammentaa useamman eri kirjoittajan teksteistä tietoa yhden kirjailijan tekemään teokseen, kirjailijan oma ääni sumentuu. Taivalkoski-Shilov (2019) esittelee artikkelissaan Folkartin (1996) määritelmän ”äänestä” tekstuaalisten ominaisuuksien joukkona, joiden voidaan nähdä tulevan yhdestä lähteestä. Jokaisella kirjailijalla on siis oma ääni, joka on näkyvissä heidän kirjoittamissaan teksteissä. Kun konekääntimet käyttävät lähteenään tekstejä eri kirjailijoilta ja niin sanotusti sekoittavat niitä, eri kirjailijoiden teokset muuttuvat pikkuhiljaa enemmän toisiaan muistuttavammiksi ja kirjailijoiden ääni muuttuu homogeenisemmäksi. Toisaalta yhden kirjailijan ääni muuttuu heterogeenisemmäksi, kun hänen ääneensä sekoittuu käännöksessä piirteitä useamman kirjailijan äänestä. (Taivalkoski-Shilov 2019.)

Lisäksi konekäänninten on todettu vähentävän kääntäjien luovuutta. Luodessaan käännöstä suoraan lähdeteoksesta, kääntäjät tekevät omia luovia valintoja käännösprosessissa. Raakakonekäännöstä luodessa taas useammin kääntäjä valitsee koneen ehdottaman käännösvalinnan. Tämä pienentää kääntäjien ääntä heidän tekemissään käännöksissä. Vaikka kääntäjän ääni on usein tekstissä ”hiljaisempi” kuin alkuperäisen kirjailijan ääni, voidaan saman kääntäjän tekemistä käännöksistä löytää merkkejä kääntäjän äänestä. Tietyt kääntäjät esimerkiksi suosivat tietynlaisia kieliopillisia valintoja käännöksissään. Kääntäjän ääni muodostuukin käännösprosessin aikana heidän tekemistään käännösvalinnoista. Mikäli konekäännin tekee käännösvalinnat, kääntäjän ääni vaimenee. Tämä myös usein vaikuttaa tekstin luovaan ilmeeseen. Mikäli kaunokirjallisuuden kääntämisessä aletaan suosimaan täysin automatisoituja käännösprosesseja, tämä käännösalan haara luovana alana on vaarassa kadota kokonaan. (Taivalkoski-Shilov 2019.)

On myös huomionarvoista, että konekäännösten ja koneavusteisten käännösten laatu on vaihtelevaa myös siksi, että niitä luovat muutkin kuin ammattikääntäjät. Yhä useammin konekääntimiä ja muita CAT-työkaluja käyttävät paraprofessionaalit ja ei-ammatilliset kääntäjät. Vaikka heiltä löytyisi kielitaitoa, heiltä ei voida olettaa löytyvän samoja kompetensseja käännösten arviointiin ja muokkaamiseen kuin koulutetuilla ammattikääntäjillä

on. Tällöin käännösten laatu heikkenee ja markkinat täyttyvät entisestään huonolaatuisista käännöksistä (Lubna & Shraideh 2025, 303).

Konekääntimien kehityksen myötä on kehitetty myös tapoja mitata konekäännösten laatua. Konekäännösten laatua on mitattu ihmisarvioijien avulla, jotka tyypillisesti tarkastelevat käännöksistä vastaavuutta ja sujuvuutta, eli kuinka hyvin ne vastaavat lähdetekstiä ja kuinka sujuvaa käännöksen kieli on. Joissakin arvioinneissa tutkitaan myös muun muassa käännöksen luettavuutta ja käytettävyyttä. (Castilho ym. 2018, 29.) Koska ihmisten tekemät arvioinnit kuitenkin usein vievät paljon aikaa ja rahaa, konekäännösten laadun mittaamiseen on kehitetty automaattisia keinoja. Ensimmäiset laadun mittaamisen mallit katsoivat lähinnä sitä, vastaavatko käännetyt sanat lähdetekstiä. Nykyisin jotkin mallit kuitenkin mittaavat käännöksiä jo laajemmin. Yleisimmässä käytössä tällä hetkellä on BLEU-arvioinnit. BLEU-arvioinnit (*Bilingual Evaluation Understudy*) mittaavat käännöksen sujuvuutta vertaamalla konekäännöstä ihmisen samasta lähdetekstistä tekemään käännökseen. (Castilho ym. 2018, 30–40.) BLEU pisteyttää konekäännökset niin, että korkeammat pisteet merkkaavat suurempaa samankaltaisuutta rinnakkaiskäännöksiin ja siten parempaa laatua. Vaikka muitakin laadun arviointimenetelmiä on kehitetty, BLEU on säilyttänyt paikkansa alalla eniten käytettynä laadun mittarina. (Kong & Lu 2025.)

Laadun käsitteen voidaan siis todeta vaihtelevan käännösalalla huomattavasti näkökulmasta, tarkoituksesta ja teoriasta riippuen. On myös huomattavaa muistaa, että kaunokirjallisuuden käännösten laatuvaatimukset ovat yleensä muita tekstejä korkeammat, koska ne ovat pitkään säilyviä ja yleensä tekijänoikeuslakien suojelemia tekstejä (Taivalkoski-Shilov 2019). Tässäkin luvussa mainitut tutkijat useimmiten keskittyivät ennemmin käännöksen hyväksyttävyyteen, eli siihen, onko käännös hyväksyttävä siltä vaadittuun viestintätilanteeseen. Tämänkään tutkielman tavoitteena ei ole tehdä kattavaa laatuarviota konekäännöksen ja jälkieditoidun käännöksen välillä. Laatua käsitellään tutkimuksessa lähinnä kulttuurillisen merkityksen välittymisessä, eli raakakonekäännöksestä tarkastellaan sitä, onko konekäännin tehnyt sellaisia valintoja kulttuuriviittausten käännöksissä, että niiden kulttuurillinen sisältö välittyy lukijalle. Toisin sanoen, onko raakakonekäännös hyväksyttävä käännös itsessään, vai vaatiiko se jälkieditointia, jotta julkaisua varten vaadittu hyväksyttävyys tavoitetaan. Lisäksi pohditaan sitä, miten konekäännin on mahdollisesti päätynyt tiettyihin valintoihin, jos niiden voidaan katsoa olevan käännökseen sopimattomia.

## 2.4 Konekääntimien eettisyys

Konekääntimien käytön yleistyessä käännösalan ovat vallanneet myös uudet eettiset pohdinnat ja ongelmat. Taivalkoski-Shilov (2019) avaa näitä artikkelissaan, jossa todetaan, että näistä yhtenä suurimpana ongelmana esiin nousee kääntäjien oikeudet. Kääntäjillä on oikeus saada palkkio työstään, samoin kuin muillakin kouluttautuneilla ammattilaisilla. Kuitenkin monet asiakkaat haluavat maksaa käännöstöistä vähemmän, mikäli niiden tekemiseen on käytetty konekääntimiä tai CAT-työkaluja. Alan jatkuva teknologisoituminen siis vaarantaa kääntäjien jo valmiiksi suhteellisen matalat palkkiot. Samalla kun palkkiot madaltuvat, toisaalta työn nopeuden odotukset nousevat. Kun koneet tehostavat työskentelyä, asiakkaat myös odottavat käännösten valmistuvan nopeammin, mikä lisää työn henkistä kuormaa. Toisaalta teknologisoituminen myös vaatii alan ammattilaisilta uudenlaisia kompetensseja, jotta työn voi suorittaa ammattitaitoisesti. Vallitseva näkemys tuntuukin olevan, että kääntäjän pärjääminen alalla on kytköksissä hänen teknologisiin taitoihinsa sekä kykyyn sopeutua alati muuttuvaan ja kehittyvään markkinatilanteeseen. (Taivalkoski-Shilov 2019.)

Tekijänoikeuskysymykset ovat myös konekääntimien käytön eettisyyden kysymysten kärjessä. Ensinnäkin on pohdittava, kenellä on tekijänoikeus koneavusteisesti tehtyyn käännökseen. Euroopan Unionin julkaiseman tutkimuksen mukaan konekäännin ei voi olla käännöksen kirjailija tekijänoikeuslain määritelmän perusteella, koska kirjailija on määritelty ihmisenä, joka on luonut työn. Täten konekääntimen luoma käännös ei ole tekijänoikeuslain alainen, jos ihmiskääntäjä ei ole jälkieditoinut käännöstä. Koneavusteinen käännös on kääntäjän omaisuutta tekijänoikeuslain alla, mikäli kääntäjä on tehnyt siihen muutoksia, koska silloin käännös voidaan nähdä kääntäjän tekemänä luovana työnä ja kääntäjä käännöksen kirjoittajana. (Debusshe & Troussel 2014.) Toisaalta ongelman tuottaa se, että konekääntimet käyttävät kääntäjien jo tekemiä käännöksiä koulutusdatassaan. Esimerkiksi e-kirjat ovat hyödyllisiä lähteitä kaunokirjallisuutta kääntävien koneiden kehittämiseen ja kouluttamiseen, mutta niitä käytettäessä pitäisi varmistaa, että kääntäjät saavat asianmukaisen kompensaation, mikäli heidän käännöksiään käytetään koulutusmateriaalina (Toral & Way 2015, 240).

Kuten jo alaluvussa 2.3 todettiin, konekäänninten käytön riskinä kaunokirjallisuuden kääntämisessä on myös se, että yhden kirjailijan ääni heterogenisoituu tai useamman kirjailijan ääni homogenisoituu (Taivalkoski-Shilov 2019). Tämä on suora uhka alkuperäisen tekstin kirjoittajan oikeuksille. Huonolaatuinen käännös, niin ihmisen kuin koneen tekemänä, voidaan

aina nähdä uhkana kirjailijan oikeuksille, koska se voi antaa vääränlaisen kuvan käännöksen kohdelukijoille kirjailijasta tai hänen töistään. Toisaalta huonolaatuinen käännös on myös eettinen haitta lukijoille. Kuten Guerberof-Arenas ja Toral (2020) totesivat tutkimuksessaan, lukijat yleisesti reagoivat paremmin käännöksiin, joiden tekemisessä oli ihmiskääntäjiä mukana. Konekäännökset siis voivat vaikuttaa tekstin luettavuuteen ja täten nautittavuuteen negatiivisesti. Toisaalta huonolaatuiset käännökset voivat vaikuttaa kohdeyleisön kielitaitoon negatiivisesti, erityisesti kun kohdeyleisönä ovat lapset. Myös kirjallisuusideoiden ja -valikoiman näkyvyys ja siirtyminen kulttuurien välillä voi vaarantua, mikäli konekäännin tuottaa huonolaatuisia käännöksiä, mikä on suora haitta kohdeyleisölle ja tiedon tasa-arvoiselle levittymiselle. (Taivalkoski-Shilov 2019.)

Konekäänninten eettiset ongelmat alkavat jo koneiden kehittämisvaiheessa. Useimmiten käännöstyökalujen kehitysprosesseissa ei ole mukana käännösalan ammattilaisia, joten edes ammattilaisille suunnatut työkalut eivät aina vastaa kääntäjien tarpeita ja toiveita konekääntimiltä ja CAT-työkaluilta. Kaunokirjallisuutta kääntäviä konekääntimiä on erityisen vaikea luoda, sillä suuri osa kaunokirjallisuuden kääntämisestä tapahtuu ”rivien välissä”. Kaunokirjallisen tekstin tyyli ja merkitykset muotoutuvat yhtä paljon tekstin välissä ja siinä mitä ei ole sanottu, kuin myös itse tekstissä. Tämä vaatii kaunokirjallisuuden kääntäjiltä taitoa, tietoa ja ymmärrystä kirjallisuudesta niin lähde- kuin kohdekielessäkin. Vaikka teknologia on koko ajan kehittynyt, konekääntimet harvoin pystyvät lukemaan tekstiä myös ”rivien välistä”, mikä tekee kaunokirjallisuuden tekstien kääntämisestä konekääntimillä erittäin riskialtista muille mainituille eettisille ongelmille. (Taivalkoski-Shilov 2019.)

## 2.5 Englannin kielen ja länsimaisen kulttuurin valta-asema käännösalalla ja konekääntimissä

Englannin kielen valta-asema globaalina lingua francana on tunnistettu jo pitkään. Bassnett ja Lefevere (1998) sanoivat kirjassaan englannin kielellä olevan samanlainen valta-asema kuin latinan kielellä aikanaan oli, ja siksi myös käännökset muista kielistä englantiin usein nojaavat enemmän englannin kielen konventioiden suuntaan ja monia eksoottiseksi leimattujen kulttuurien ilmiöitä standardisoidaan. EMT-kompetenssimallin viitekehyksessä englannin kasvava asema lingua francana mainitaan myös syynä sille, miksi käännösalan markkinoiden tarpeet ovat muuttuneet ja sitä myötä myös EMT-mallissa mainittuja kompetensseja on pitänyt päivittää (EMT Competence Framework 2022). Englannin kielellä on myös johtoasema

globaaleissa instituutioissa ja organisaatioissa, ja se on myös tieteellisten alojen johtava kieli (Olojo ym. 2025).

Englannin kielen valta-asema juontaa juurensa länsimaiseen kolonialismiin ja sen aiheuttamaan länsimaisen kulttuurin valta-asemaan. Kieli toimi yhtenä kolonialististen voimien vallan käytön aseena, jolla kolonisoituja yhteisöjä sorrettiin. Kyky puhua ja ymmärtää kolonisoidun valtion uutta lingua francaa (usein englantia) tarkoitti myös lähempää asemaa valtaapitäviin voimiin ja oikeudenmukaiseen kohteluun. (Olojo ym. 2025.) Myös kielen kääntäminen toimi kolonisoinnin työkaluna. Esimerkiksi Idästä tulleita tekstejä kääntäessä kääntäjät käyttivät käännösstrategioita, jotka eksotisoivat ja toiseuttivat lähdetekstien kulttuuria ja ihmisiä entisestään. Toisaalta myös lähdetekstejä kotoutettiin länsimaiseen kulttuuriin ja kieliin niin vahvasti, ettei niitä välttämättä tunnistaisi niiden todellisen lähtökulttuurin teksteiksi lainkaan. Näitä strategioita ja käännöksiä on myöhemmin alettu kritisoimaan käännösalalla, ja jotkut kääntäjät ovat lähteneet dekolonisoimaan käännösprosessia valitsemalla käännösstrategioita, jotka taistelevat näitä kolonialistisia perinteitä vastaan. (Yoon 2021.)

Englannin kielen ja länsimaisten kulttuurien dominanssi on myös nähtävissä konekäänninteknologiassa. Myös useat sellaiset kielet, joilla on paljon puhujia, ovat usein huomattavasti huonommin edustettuina konekäänninteknologiassa englannin kieleen verrattuna. Ongelma juontaa juurensa koulutusdatan puutteeseen. (Olojo ym. 2025.) Siinä missä englanninkielisiä rinnakkaistekstejä löytyy hyvinkin paljon ja eri tyylilajeissa, monet varsinkin globaalin etelän kielet lasketaan matalien resurssien kieliksi, koska niistä ei ole tarjolla yhtä laajaa koulutusdataa hyödynnettäväksi konekäänninteknologiassa. Monet konekääntimet suoriutuvatkin käännöstehtävistä paremmin, mikäli tehtävän kieliparin toinen kieli on englanti, koska englannin kieli toimii useimmiten pohjana käännösteknologian ohjelmille. Koska englannin kielellä ja länsimaisilla kulttuureilla on niin pitkälle historiaan yltävä valta-asema, niiden jälkiä näkyy väkisinkin myös konekääntimien luomissa käännöksissä. (Olojo ym. 2025.)

Konekääntimissä esiintyvä kulttuurien välinen epätasa-arvo tuottaa ylimääräistä työtä kääntäjille, joiden työtä sen olisi tarkoitus helpottaa, sekä pahimmillaan vahingoittaa käännöksen kohdeyleisöä. Esimerkiksi Jordaniassa suoritetun tutkimuksen mukaan tekoälypohjaisia käännöstyökaluja käyttävät kansalaisjärjestöjen työntekijät huomauttivat, miten ohjelmistot tekivät käännösvalintoja selvästi globaalin pohjoisen kulttuuripainotuksella, vaikka käytetyt kieliparit sijoittuivat Lähi-idän kielialueisiin. Työntekijät sanoivat

huomanneensa, että joidenkin kulttuuristen viittausten merkitys oli muuttunut tai jopa poistettu käännöksistä kokonaan. Tällaiset käännösongelmat voivat vahingoittaa järjestöjen apua tarvitsevia henkilöitä tai antaa järjestöistä epäluotettavan kuvan. (Abdul-Hadi & Shraideh 2025, 303–304.)

Myös nigerialaisiin kieliin keskittyneessä tutkimuksessa esiintyi vastaavanlaisia tuloksia. Tutkimukseen osallistuneet kääntäjät raportoivat konekääntimien olevan kykenemättömiä ymmärtämään tekstiin liittyvää ulkoista kontekstia, joten käännöksissä kulttuurinen konteksti oli usein muuttunut huomattavasti. Yksi osallistuja huomautti tämän olevan erittäin ongelmallista, koska hänen äidinkielensä yoruba on myös kulttuuri, joten kieltä ja kulttuuria ei voi erottaa toisistaan. Kun tekstien kulttuuriset merkitykset muutetaan tai poistetaan käännöksistä, sillä on myös pitempikantoinen vaikutus kieleen ja kulttuuriin. Kun englannin kieli pysyy yhteisöjen välisen kommunikaation valtakielenä, yorubalaiset sanonnat, sanat ja jopa jotkut perinteet alkavat kadota kokonaan käytöstä. (Olojo ym. 2025.)

# 3 AINEISTO JA MENETELMÄ

Tässä luvussa tutkielman aineisto esitellään tarkemmin ja käydään läpi prosessi, jolla aineisto on luotu. Lisäksi keskustellaan siitä, millaisia translatorisia ilmiöitä jo lähdeteoksessa esiintyy. Aineiston esittelyn jälkeen tutkielman menetelmä käydään läpi ja esitellään ryhmitykset, joihin esimerkit kulttuuriviittauksista jaoteltiin analyysia varten.

## 3.1 Aineisto

Tutkielman aineisto on kerätty Siru Auran Myötäjäiset-teoksen kolmesta ensimmäisestä luvusta. Sari Hokkanen on tutkimuksensa pilottivaihetta varten tehnyt luvuista raakakonekäännöksen, jonka hän on jälkieditoinut. Jälkieditoitu käännös on myös viimeistelty lähdeteoksen kirjailijan kanssa yhteistyössä. Teoksesta ei ole vielä olemassa julkaistua käännöstä. Hokkasen tutkimuksen osana tehty englanninkielinen käännös julkaistaan käännöksen valmistuttua. Tässä tutkielmassa käytetty viimeistelty käännös voi siis vielä muuttua ennen julkaisua.

Konekäännöksen tekemiseen käytettiin Wordfast-käännösmuistiohjelmaan integroitua DeepL-konekäännintä. Tutkimuksessa käytetty versio on perinteinen neuroverkkokonekääntimen. DeepL-ohjelmaa on tutkimuksen jälkeen päivitetty suuren kielimallin (LLM) toimintajärjestelmään, mutta Hokkasen tutkimuksessa on käytetty aiempaa, neuroverkkotoimintaista versiota. Teoksen ensimmäiset kolme lukua on syötetty ohjelmaan pdf-versiona, ja käännetty Intian englannin varianttia käyttäen englanniksi.

Hokkanen on jälkieditoinut raakakonekäännöksen lähdetekstin pohjalta. Käännösprosessin päätyttyä teoksen kirjailija on lukenut Hokkasen käännöksen, ja antanut omat huomionsa tekstiin. Viimeistellyssä käännöksessä esiintyy siis kääntäjän omat käännösvalinnat, sekä joitakin kirjailijan huomioiden pohjalta tehtyjä termivalintoja ja muutoksia. Aura on esimerkiksi keskusteluissa Hokkasen kanssa pyytänyt lisäystä käännökseen tekstin kohtaan, jossa mainitaan Intian myötäjäisiä koskeva laki. Aura on myöhemmin huomannut lain olevan väärin lähdeteoksessa, joten laki on korjattu oikeaksi jälkieditoituun käännökseen. Myös myötäjäismurha-termin englanninkielinen käännös, *dowry death*, on Siru Auran mukaan vakiintunut termi ja siksi valittu jälkieditoituun käännökseen konekääntimen ehdottaman dowry murder -ilmaisun sijaan.

Tässä tutkimuksessa analysoitu aineisto on kerätty lähdetekstistä, raakakonekäännöksestä ja viimeistellystä käännöksestä. Eri versioista on poimittu kulttuuriviittauksia tässä tutkielmassa käytettävän kulttuuriviittauksen määritelmän pohjalta. Viittaukset sisältävät yksittäisiä termejä, lausahduksia, lauseen osia ja kokonaisia virkkeitä. Tutkielmaan kerätyt kulttuuriviittaukset liittyvät pääosin teoksen kontekstin valtakulttuureihin, eli Intian ja Suomen kulttuureihin. Aineistosta on kuitenkin poimittu myös muutama muita kulttuureja koskeva esimerkki, joita tutkitaan englannin kielen ja länsimaisen kulttuurin näkökulmasta.

Siru Auran teos on suomenkielinen, mutta siinä esiintyy myös intialaisista kielistä peräisin olevia reaalioita. Kirjan hahmot käyttävät hindin-, kannadan- ja tamilinkielisiä ilmaisuja. Hindi on Intian virallinen kieli, ja sitä puhuu äidinkielenään tai toisena kielenään satoja miljoonia ihmisiä. Kannada taas kuuluu Intian kansallisiin kieliin ja silläkin on monia kymmeniä miljoonia puhujia. Kannadaa puhutaan lähinnä Karnatakan osavaltiossa Intiassa. Myötäjäiset-teoksessa päähenkilö Taarinin perhe on tamileja ja he puhuvat tamilia kotonaan. Tästä syystä myös tamilinkielisiä reaalioita löytyy tekstistä. Tamili on myös yksi Intian kansallisista kielistä ja silläkin on kymmeniä miljoonia puhujia. Jotkin kirjassa esiintyvät reaaliat voivat olla peräisin useammasta kuin yhdestä näistä intialaisista kielistä. Esimerkiksi aiemminkin mainittu termi *akka* tarkoittaa tamilin kielellä siskoa ja kannadan kielellä vanhempaa siskoa.

Koska teoksessa esiintyy vieraskielisiä termejä, Siru Aura on tehnyt translatorisia päätöksiä jo lähdeteoksen kirjoitusvaiheessa, vaikka teos ei varsinaisesti ole käännös. Koska suomenkielisessä teoksessa esiintyy myös intialaisten kielten ja kulttuurien reaalioita ja ilmiöitä, Auran on pitänyt tehdä translatorisia päätöksiä siitä, miten ilmaista asiat tekstissään. Esimerkiksi monet päähenkilöiden uskontoon liittyvät termit on jätetty alkuperäisessä muodossaan lähdetekstiin, eikä niille ole yritetty keksiä suomenkielistä vastinetta. Aura on kuitenkin käyttänyt sanojen yhteydessä muita käännösstrategioita, kuten esimerkiksi lisännyt termin yhteyteen selityksen sen merkityksestä. Nämä valinnat huomioitiin myös aineistoa kerätessä.

## 3.2 Menetelmä

Tutkielman analyysia ohjaa kaksi tutkimuskysymystä: 1. Miten konekäännin kääntää kulttuuriviittauksia kaunokirjallisessa tekstissä? 2. Esiintyykö konekäännöksessä englannin kielen ja länsimaisen kulttuurin dominanssia? Ensimmäisen kysymyksen tavoitteena on

selvittää, millaisia käännösvalintoja tekee ja välittyykö lähdetekstin kulttuurinen merkitys näiden käännösvalintojen avulla myös käännöksessä. Näin voidaan luoda käsitys siitä, vaatiiko kulttuurisen merkityksen siirtyminen kielien ja kulttuurien välillä jälkieditoijaa, vai pystyykö kone siihen itsenäisesti. Toisen tutkimuskysymyksen tavoitteena on tutkia, tuottaako konekäännin ratkaisuja, joissa korostuu englannin kieli tai länsimainen kulttuuri. Mikäli dominanssi on vahvasti havaittavissa konekääntimen koulutusdatassa, se luo haasteita erityisesti analyysin kohteena olevan teoksen kaltaisille teksteille, jotka ovat kulttuuririkkaita ja joiden painopiste on muissa kuin englannin kielessä tai länsimaisessa kulttuurissa. Molempien tutkimuskysymysten avulla pyritään päättelemään, luoko konekäännin julkaisukelpoista kaunokirjallista tekstiä itsenäisesti.

Tutkielmassa tarkasteltavat kulttuuriviittaukset on kerätty aineistosta kahdessa eri osassa. Luin käännösversioita rinnakkain ja merkitsin kaikki mahdolliset tutkimukseen soveltuvat osat teksteistä. Raakakeruun jälkeen kävin merkityt kohdat uudelleen läpi ja keräsin ne Excel-taulukkoon. Ensimmäisessä vaiheessa merkatuista tekstinosista sulkeutui pois muun muassa kohtia, jotka ovat liian pitkiä tarkkaan analyysiin, tai jotka eivät sovi tutkielman kulttuuriviittauksen määritelmän piiriin. Alun perin tässä vaiheessa aineistosta myös rajattiin pois kohdat, joissa konekäännin oli lisännyt kokonaan lähdetekstin kontekstin ulkopuolista kolmatta kulttuuria tekstiin. Sisällytin näitä esimerkkejä kuitenkin aineistoon ja tutkielmaan, koska ne ovat tärkeitä esimerkkejä konekääntimen koulutusdatasta ja kielellisestä hierarkiasta. Kun kaikki analysoitavat esimerkit oli kirjoitettu Excel-taulukkoon, samankaltaiset ilmiöt ryhmiteltiin väreillä. Tässä vaiheessa esimerkeistä poissuljettiin vielä osa, joissa ei esiintynyt tutkielman tutkimuskysymysten kannalta relevantteja ilmiöitä. Joitakin kulttuuriviittauksia oli esimerkiksi muokattu jälkieditoituun käännökseen, mutta muutokset liittyivät kielenhuollollisiin seikkoihin eikä kulttuurisisältöön.

Kun olin kerännyt kaikki tutkimukseen sopivat kulttuuriviittaukset Excel-taulukkoon, aloin ryhmitellä niitä. Pyrin luomaan ryhmiä sen mukaan, millaisia haasteita kulttuuriviittausten kääntämisessä on ollut, millaista termistöä niissä esiintyy tai muihin erikoisuuksiin, joita niissä ilmeni. Ryhmiä oli alkuun useampia, mutta lopulliseen analyysiin niitä valikoitui viisi. Ensimmäiset kolme ryhmää käsittelevät tietynlaisia kulttuuriviittauksia ja sitä, miten konekäännin on kääntänyt. Neljäs ryhmä eroaa kolmesta ensimmäisestä ryhmästä, sillä siinä tutkitaan selittäviä käännöksiä sekä käännösstrategiana että tekstin käännettävänä osana. Viidennen ryhmän esimerkit taas eivät lähdetekstin valtakulttuurien (intialaisen tai

suomalaisen) kulttuuriviittauksia, mutta konekäännin on lisännyt tekstin ulkopuolisen kulttuurin viittauksia käännökseen, josta syystä niitä tutkitaan erikseen esimerkkeinä amerikkalaisen kulttuurin valta-asemasta konekäänninteknologiassa.

Ensimmäinen ryhmä käsittelee kulttuuriviittauksia, joissa esiintyy erisnimiä. Aineistossa esiintyy sekä intialaisia että suomalaisia henkilönnimiä ja konekäännöksessä ilmeni erilaisia strategioita niiden kääntämiseen. Jotkin koneen tekemistä nimien käännöksistä voidaan myös nähdä osoituksena länsimaisen kulttuurin dominanssista. Lisäksi ryhmän sisältyy erisnimiä, kuten tunnettujen kirjojen nimiä, joiden käännöksistä ilmenee englannin kielen valta-asema konekäännöksessä.

Toinen ryhmä kattaa implisiittiset kulttuuriviittaukset. Implisiittisillä kulttuuriviittauksilla tarkoitetaan tässä tutkielmassa kulttuuriviittauksia, joissa kulttuurinen sisältö on tunnistettavissa lauseen ulkopuolisesta kontekstista. Viittauksessa ei siis esimerkiksi esiinny reaaliaa tai muuta selkeää kulttuuriviittauksen lingvististä merkkiä, vaan pitemmässä lauseessa tai virkkeessä käsitellään jotakin kulttuurista ilmiötä. Tällaisiksi viittauksiksi lasketaan esimerkiksi avioliittokulttuuriin liittyvät keskustelut ja kommentit. Vaikka niissä ei esiinny suoranaisia reaalioita, niiden kulttuurinen konteksti on tiukasti kiinni tietyssä kulttuurissa. Implisiittiset kulttuuriviittaukset toimivat hyvänä pohjana koneen kontekstin tunnistamisen kyvyn mittaamiseen, koska niiden kääntäminen tapahtuu myös ”rivien välissä”.

Kolmantena ryhmänä käsitellään reaalioita. Reaaliat ovat kulttuurisidonnaisia termejä ja fraaseja, joille ei löydy suoranaista vastinetta kohdekielestä. Tällaisia esimerkkejä esiintyy aineiston esimerkeissä useita. Reaaliat ovat sekä suomenkielisiä että intialaisista kielistä peräisin olevia termejä. Reaalioita myös löytyy sekä suomalaiseen kulttuuriin liittyen että intialaiseen kulttuuriin liittyen. Reaalioiden käännösstrategioissa esiintyy myös paljon vaihtelevuutta niin raakakonekäännöksessä kuin jälkieditoidussa käännöksessä.

Neljännessä ryhmässä tutkitaan selittäviä käännöksiä. Selittävä käännösstrategia toimii omana tutkittavana ryhmänään, koska siitä löytyy hyvin erilaisia esimerkkejä ja sitä tutkitaan kattavammin sekä tekstin käännettävänä osana että käännösstrategiana. Lähdeteoksessa esiintyy selittäviä käännöksiä jo valmiiksi. Toisaalta jälkieditoija on lisännyt selittäviä käännöksiä joihinkin kohtiin jälkieditoidussa käännöksessä. Konekäännin taas ei lisännyt selityksiä itsenäisesti, mutta sen strategiat jo lähdetekstistä löytyvien selittävien käännösten kääntämiseen vaihtelivat. Selittävän käännöksen lisääminen käännökseen on myös

dekolonisoivana käännösstrategiana merkittävä, koska tutkielman toinen tutkimuskysymys koskee englannin kielen ja länsimaisen kulttuurin dominanssia konekäännöksissä.

Viidennessä ryhmässä käsitellään sellaisia konekääntimen tekemiä käännösratkaisuja, joissa voidaan nähdä amerikkalaisen kulttuurin vaikutteita käännöksessä. Raakakonekäännöksestä löytyi tapauksia, joissa konekäännin on ikään kuin lisännyt kolmannen, eli amerikkalaisen, kulttuurin viittauksia tekstiin. Nämä käännösvalinnat ovat myös toisen tutkimuskysymyksen kannalta merkittäviä, joten niihin paneudutaan erikseen.

Vaikka kulttuuriviittaukset on analyysia varten jaoteltu viiteen ryhmään, on ryhmissä myös päällekkäisyyksiä. Selittävää käännösstrategiaa on hyödynnetty myös muiden analysoitavien ryhmien kulttuuriviittauksissa. Esimerkiksi reaalioiden yhteyteen on lisätty selittävä käännös usein jo lähdeteoksessa. Jokaisessa ryhmässä myös esiintyy kulttuuriviittausten käännöksiä, jotka voidaan nähdä esimerkkeinä englannin kielen ja/tai länsimaisen kulttuurin valta-asemasta käännösalalla. Viidennessä analysoitavassa ryhmässä käsitelläänkin erikseen vain tapauksia, jossa konekäännin on tehnyt valinnan, joka voidaan nähdä suorana amerikkalaisen kulttuurin lisäyksenä lähdetekstin kulttuuriseen merkitykseen.

# 4 ANALYYSI

Tässä luvussa esittelen tämän tutkielman aineistosta nousseita esimerkkejä siitä, miten konekäännin on kääntänyt erilaisia kulttuuriviittauksia. Jokaisessa alaluvussa käsitellään yhtä tiettyä ryhmää kulttuuriviittauksista (ryhmät esitellään luvussa 3.3).

## 4.1 Erisnimet

Tämän tutkielman aineistosta ilmenee, että konekäännin ei aina tunnista henkilönnimiä oikein. Aineistossa esiintyi useampi tapaus, jossa konekäännin joko sekoittaa henkilönnimen johonkin toiseen asiaan tai muokkaa henkilönnimeä. Tämä virhe myös nousi esiin tapauksissa, joissa kyseistä henkilönnimeä on käytetty jo aiemmin tekstissä, mutta konekäännin ei osannut tunnistaa sitä aiemman tekstin kontekstista.

(1) (ST) Tunnin päästä Agiran koulu on ohi tältä päivältä […]
(MT)In an hour, the school in Agira will be over […]
(PE) Her school will be done in an hour for the day […][1]

Yllä olevasta esimerkistä (1) ilmenee, että konekäännin on ymmärtänyt Agiran olevan kaupunginnimi eikä henkilönnimi. Agira on päähenkilö Taarinin pikkusisko, ja hänet mainitaan tekstissä useasti jo ennen tätä esimerkkiä. Huomattavan tästä esimerkistä tekee se, että Agira mainitaan myös esimerkkiä edeltävässä virkkeessä. Konekäännin ei tästä huolimatta ole kääntänyt virkettä oikeassa kontekstissa. Jälkieditoija sen sijaan on poistanut nimen lauseesta kokonaan ja käyttänyt tilalla pronominia *her*. Jälkieditoinnissa tulee esiin siis kattavampi kontekstin ymmärrys.

Sekä suomalaiset että intialaiset henkilönnimet tuottivat konekääntimelle haasteita. Tekstissä mainitaan suomalaisen päähenkilön poikaystävä Markus, jonka nimen konekäännin on englanninkieliseen versioon kääntänyt *Mark*. Konekäännin on voinut mahdollisesti lukea nimen uskonnollisesta näkökulmasta. Englanninkielisessä Raamatussa evankelista Markus on Mark. Konekäännin on siis voinut laskelmoida nimen vaativan vastaavaa käännöstä myös tässä teoksessa. Kristinuskoon pohjautuvan käännösratkaisun tekeminen myös kielii konekääntimen koulutusdatassa vallitsevasta länsimaisen valtauskonnon valta-asemasta.

(2) (ST) Muutin Markuksen perässä […]

[1] ST= source text, ’lähdeteksti’. MT= machine translation, ’konekäännös’. PE= post-edited, ’jälkieditoitu’.

(MT) Following Mark […]
(PE) I followed Markus […]

Hindin kielestä peräisin oleva *sari*-termi viittaa intialaiseen naisten vaatetukseen. *Sari*-termi esiintyy aineistossa muutaman kerran ja konekääntimen käännösratkaisut sen ympärillä ovat vaihtelevia.

(3) (ST) Sari vaikuttaisi liian perinteiseltä.
(MT) Sari would seem too traditional.
(PE) A sari would seem too traditional.

Esimerkissä 3 konekäännin on päättänyt jättää *sari*-termin kääntämättä, mutta koko virkkeen käännöstä on muokattu jälkieditoinnissa. Toisin kuin jälkieditoinnin suorittanut kääntäjä, konekäännin ei ole laittanut artikkelia termin eteen. Koska konekääntimen kieliparina on toiminut suomi ja englanti, on mahdollista, että konekäännin on ymmärtänyt sanan suomalaisena erisnimenä, eikä intialaisena naistenvaatteena, mihin tekstissä viitataan. Kääntäjän jälkieditoituun käännökseen lisäämä artikkeli tuo käännökseen kielenhuollollista puhtautta ja varmistaa, ettei lukija voi erehtyä virkkeen kulttuurisesta sisällöstä.

Vaikka konekääntimen useat erisnimien käännösvalinnat viittasivat ja länsimaisen kulttuurin dominanssiin, kone ei tunnistanut kaikkia siihen kuuluvia viittauksia. Humiseva harju -teos (englanniksi *Wuthering Heights*) on merkittävä osa englanninkielistä kirjallisuuden kaanonia, jolle on vakiintunut suomennettu nimi.

(4) (ST) […] luettuani Humisevan harjun, […]
(MT) […] when I read Humiseva Ridge, […]
(PE) […] after having read *Wuthering Heights*, […]

Konekäännin ei tunnistanut teoksen suomennettua nimeä, vaan pyrki kääntämään nimen sanasanaisesti (esimerkki 4). Voidaan siis olettaa, että konekääntimen koulutusdatasta on puuttunut tarvittavat rinnakkaismateriaalit, jotta konekäännin olisi tunnistanut suomennetun nimen. Konekääntimellä siis esiintyi haasteita myös länsimaiseen kulttuuriin kytköksissä olevien kulttuuriviittausten kääntämisessä.

## 4.2 Implisiittiset kulttuuriviittaukset

Raakakonekäännöksen luomiseen käytetyllä konekääntimellä ilmeni haasteita myös implisiittisten kulttuuriviittausten kääntämisessä. Näissä tapauksissa haasteet olivat useimmiten kontekstiin liittyviä. Länsimaisella kulttuurilla on selkeä valta-asema konekääntimien

koulutusdatassa, mikä voi olla tämänkin tutkielman aineistossa esiintyneiden haasteiden taustalla.

(5) (ST) Asemani on kohonnut.
(MT) My position has been elevated.
(PE) My status has been elevated.

Esimerkissä 5 implisiittisen kulttuuriviittauksen kulttuurinen merkitys on muuttunut raakakonekäännöksessä, ja kääntäjä on muuttanut viittausta viimeisteltyyn käännökseen. Intian kastilaitos tarkoittaa vanhaa järjestelmää, jonka perusteella jokainen ihminen jaotellaan arvojärjestykseen. Kastilaitoksen arvojärjestys perustuu vanhoihin uskonnollisiin uskomuksiin. Vaikka nykyään kastiin liittyvä syrjintä on kielletty Intian laissa, kastijärjestelmä on silti valloillaan erityisesti Intian maaseuduilla. (Mayell 2003.) Vaikka konekääntimen käyttämä position-termi (MT) on teknisesti ottaen oikea käännös asema-sanalle, sillä ei englannin kielessä ole samaa kulttuurista konnotaatiota, kuin mihin lähdetekstissä viitataan. Position-termillä viitataan useimmiten henkilön työpaikkaan tai -titteliin, eikä yhteiskunnalliseen asemaan (TEPA 2026). Kääntäjän käyttämä käännös, *status* (PE), sen sijaan tarkoittaa Oxford Dictionary of Englishin (OED 2026) mukaan henkilön sosiaalista tai yhteiskunnallista asemaa. Tässä tapauksessa siis kääntäjän käyttämä käännös välittää lähdeteoksen kulttuurisen kontekstin ilman väärinymmärtämisen riskiä, kun taas konekääntimen käyttämä käännös voi harhaanjohtaa lukijaa alkuperäisteoksen sisällöstä.

Samankaltainen haaste ilmeni myös lähdeteoksen kohdissa, joissa keskusteltiin intialaisista avioliittoperinteistä. Intialaisessa kulttuurissa on edelleen yleistä, että vanhemmat tai sukulaiset järjestävät lapsilleen avioliiton, eli he etsivät ja valitsevat lapsensa tulevan aviopuolison. Puolisoilla on kuitenkin ainakin teoriassa mahdollisuus torjua toisensa, eli vanhempien suosituksia ei ole pakko noudattaa. Myötäjäiset-teoksessa järjestettyjen avioliittojen yleinen asema kulttuurissa nousee esiin hyvin nopeasti.

(6) (ST) Vanhempani eivät järjestä mitään.
(MT) My parents don’t organise anything.
(PE) My parents are not making arrangements.

Kirjan päähenkilö Taarini kertoo työhaastattelussa haastattelijoille, ettei hänen vanhemmillaan ole suunnitelmia järjestää hänelle avioliittoa (esimerkki 6). Konekäännin on kääntänyt lauseen niin, että sen tarkoitus viittaa ennemmin yleisesti asioiden järjestämiseen. Järjestetyn avioliiton vakiintunut nimitys englanniksi on *arranged marriage*, ja siitä puhuttaessa käytetään useimmiten jotakin nimityksestä johdettua variaatiota. Konekääntimen *organise*-termin käyttö

siis häivyttää lähdetekstin kulttuurillista sisältöä, vaikka jossakin muussa tapauksessa käännös olisikin täysin hyväksyttävä.

Aineistosta nousi esiin myös useita esimerkkejä implisiittisistä kulttuuriviittauksista, jotka olivat oikein jo raakakonekäännöksessä eikä niitä muutettu jälkieditoinnissa ollenkaan. Esimerkiksi Suomen luontoa kuvailevassa kulttuuriviittauksessa ei ilmennyt eroja ollenkaan raakakonekäännöksen ja kääntäjän jälkieditoiman version välillä. Joissakin muissa tapauksissa kulttuuriviittauksiin oli tehty kielenhuollollisia muutoksia tai korjauksia jälkieditoinnissa, mutta itse kulttuuriviittaus säilyi samanlaisena eri versioiden välillä. Tällaisissa esimerkeissä ei useimmiten esiintynyt esimerkiksi lainasanoja ollenkaan. Konekääntimen voikin olla helpompi kääntää implisiittisiä kulttuuriviittauksia, joissa ei käytetä lainasanoja, koska oikeiden käännösten laskeminen on todennäköisempää eivätkä ne välttämättä vaadi kulttuurisen kontekstin ymmärtämistä. Toisaalta voidaan olettaa länsimaisen kulttuurin valta-aseman vaikuttavan myös implisiittisiin kulttuuriviittauksiin. Suomalaisen luonnon kuvailuun ei liity samanlaisia rivien välistä luettavia kulttuurisia konnotaatioita, kuten intialaisen kastilaitoksen tai avioliittokulttuurin ymmärtämiseen liittyy. Vaikka ne ovat myös länsimaisessa kulttuurissa tunnettuja perinteitä, konekääntimen koulutusdatan ei oletettavasti ole sisältänyt tarpeellista rinnakkaismateriaalia intialaisen kulttuurin ymmärtämiseen.

Implisiittiset kulttuuriviittaukset paljastivat konekääntimellä olevan haasteita kontekstin ymmärtämisessä muutenkin, kuin kulttuurin ymmärtämisen osalta.

(7) (ST) Hän istuu keittiön lattialla oranssi sari repsottaen […]
(MT) He sits on the kitchen floor, orange sari flapping […]
(PE) She sits on the kitchen floor, orange sari hanging limply, […]

Esimerkissä 7 nähdään, että vaikka konekäännin on tällä kertaa tunnistanut *sari*-sanan ja jättänyt sen kääntämättä, tekstin laajempi konteksti on jäänyt koneelta tunnistamatta. Pelkästään esimerkissä esiintyvän virkkeen kontekstista voidaan päätellä henkilön olevan nainen, koska sari on perinteisesti naisten vaate. Lisäksi esimerkkiä edeltäneessä virkkeessä kyseessä olevan henkilön mainitaan olevan perheen äiti. Konekäännin on kuitenkin käyttänyt englannin maskuliinista kolmannen persoonan pronominia *he*. Kone ei siis ole tulkinnut käännettävän tekstin kontekstia kulttuurin eikä kieliopin kannalta. On kuitenkin myös huomattava, että käännökseen käytetyllä konekääntimellä on aineiston perusteella selkeä maskuliininen preferenssi, koska aineiston jokainen hän-pronomini oli käännetty maskuliiniseksi (*he*) vaikka pronominia ympäröivästä kontekstista pystyisi helposti päätellä feminiinisen pronominin olevan parempi valinta.

## 4.3 Reaalioiden käännökset

Vaikka konekäännin käytti Intian englannin varianttia, lähdeteoksesta löytyneet intialaisista kielistä kotoisin olevat reaaliat tuottivat haasteita konekääntimelle. Jotkin suomenkieliset reaaliat tuottivat myös haasteita, tosin konekäännin vaikutti ymmärtäneen suomalaisia reaalioita intialaisia reaalioita paremmin. Tästä ei kuitenkaan voida tehdä lopullista johtopäätöstä tämän tutkielman aineiston perusteella, koska aineistossa esiintyi suomenkielisiä reaalioita intialaisten kielten reaalioita harvemmin. Reaalioiden käännöksissä oli myös käytetty eniten vaihtelevia käännösstrategioita sekä raakakonekäännöksessä että jälkieditoidussa käännöksessä.

(8) (ST) […] pieni, punareunainen helmibindi […]
(MT) A small, reddish-red pearl band […]
(PE) […] a small, red-rimmed pearl bindi […]

*Bindi* on intialaisten naisten käyttämä otsakoriste, jonka nimi on peräisin hindin kielestä. *Bindi*-sana on intialainen reaalia, jota on adoptoitu lainasanaksi myös monessa muussa kielessä, eikä sille löydy esimerkiksi suomen kielessä tai englannin kielessä muuta vastinetta. Yllä olevasta esimerkistä kuitenkin käy ilmi, että konekäännin on pyrkinyt kääntämään reaalian. Yritys kääntää reaalia kuitenkin muuttaa käännöksen kulttuurista sisältöä huomattavasti. Konekääntimen käännös esimerkissä 8 viittaa helmipantaan, joka poikkeaa bindistä. Bindi on keskelle otsaa liimattava tai piirrettävä koriste, kun taas helmipanta on erillinen asuste. Kääntäjä on käyttänyt lähdetekstissä esiintyvää reaaliaa viimeistellyssä käännöksessä, palauttaen täten kulttuurisen kontekstin käännökseen.

Kuten jo aiemmin mainittiin, *sari*-termi on aineistosta löydettyjen esimerkkien perusteella tuottanut erityisesti hankaluuksia konekääntimelle.

(9) (ST) Silkkisarit
(MT) Silk trousers
(PE) Silk saris

Edellä mainitussa esimerkissä konekäännin ei ole tunnistanut *sari*-reaaliaa, vaan on kääntänyt sen englannin kielen sanaan *trousers*, joka tarkoittaa housuja. Sari on perinteisesti naisten asu, johon sisältyy hame, joten konekääntimen valitsema housuihin viittaava käännös on harhaanjohtava. Kääntäjä on jälleen muokannut käännöstä ja käyttänyt alkuperäistä reaaliaa viimeistellyssä käännöksessä. Näin kulttuuriviittaus välittyy lukijalle.

Joissakin tapauksissa konekäännin ei myöskään joka kerta noudattanut samaa käännösstrategiaa useammin esiintyvälle reaalialle. *Myötäjäiset*-termi esiintyy aineistossa useasti. Ensimmäisellä kerralla termin esiinnyttyä tekstissä konekäännin käänsi termin englanninkielisenä terminä *dowry*. Myötäjäisten maksaminen ei ole pelkästään intialaisessa kulttuurissa esiintyvä perinne, joten konekäännin on mahdollisesti tunnistanut termin tästä syystä muita reaalioita helpommin. Kun *myötäjäiset*-termi myöhemmin esiintyi omana lauseenaan (esimerkki 10), konekäännin ei kuitenkaan enää kääntänytkään termiä aiemmalla tavalla. Konekäännin vaikuttaa luoneen hallusinaation termin käännöksessä.

(10) (ST) Myötäjäisinä?
(MT) Forever and ever?
(PE) As dowry?

Konekäännin on mahdollisesti olettanut termin alkuosan perusteella sen vastaavan länsimaisessa hääkulttuurissa käytettävää yleistä sanontaa ”myötä- ja vastoinkäymisissä”, ja pyrkinyt luomaan käännöksen sanonnan kontekstin avulla. Ratkaisu kuitenkin muuttaa tekstin kulttuurisen sisällön täysin, koska myötäjäiset ovat merkittävä osa järjestettyjen avioliittojen kulttuuria, erityisesti Intiassa, jossa ne ovat myös aiheuttaneet huomattavan määrän epätasa-arvoa naisten ja miesten välille. Tässä tapauksessa konekäännöksessä esiintyvä hallusinaatio siis paitsi häivyttää kulttuurista sisältöä, se voi myös aiheuttaa lukijalle romantisoidumman kuvan aiheesta, joka on lähdetekstissä suuri konfliktin lähde.

Samankaltaisia johdonmukaisuusongelmia esiintyi muuallakin aineistossa. Intiassa käytettävä valuutta rupiat mainitaan aineiston lähdetekstissä muutamassa eri kohdassa. Kun rupiat mainitaan lähdetekstissä ensimmäisen kerran (esimerkki 11), konekäännin ei tunnistanut termiä valuutaksi, vaan käänsi sen krutongeiksi.

(11) (ST) Lataan rupiat pöytään niin nopeasti, etten ehdi katua.
(MT) I put the croutons on the table so fast, I don’t have time to regret it.
(PE) I stack the rupees on the table so fast I don’t have time for regrets.

Konekäännin kuitenkin käänsi saman termin myöhemmin tekstissä termillä *rupees*, joka on englanninkielinen termi valuutalle. Konekäännin siis tunnisti termin myös oikeassa kulttuurisessa kontekstissa, mutta konekäännin ei pysynyt johdonmukaisesti käännösvalinnassaan.

Vaikka konekäännin käänsi muissa tapauksissa rupiat myös oikein, koneella oli haasteita muidenkin intialaisten valuuttojen käännöksissä. Lähdetekstissä käytetään myös termiä *paisa*,

joka on rupian sadasosa. Termi myös tarkoittaa hindin kielellä rahaa tai käteistä. Konekäännin ei tunnistanut termiä, vaan poisti sen käännöksestä kokonaan.

(12) (ST) […] joka ikisen paisan.
(MT) […] every last bit of it.
(PE) […] every single paisa.

Esimerkissä 12 Taarinin isä puhuu vihaisena siitä, että hän maksoi nuoremman tyttärensä häistä ja nyt tyttären uusia aviomies pyytääkin heidän sopimuksensa vastaisesti myötäjäisiä uudesta vaimostaan. Konekääntimen käännös tarkoittaa ennemminkin ”maksoin kaiken sen”. Vaikka alkuperäinen merkitys ei varsinaisesti muutu raakakonekäännöksessä, häivyttää se silti intialaista kulttuuria tekstistä. Tekstissä myös mainitaan aiemmin se, miten Taarinin isä rohkaisee tyttäriään käyttämään äidinkieltään tamilia ja hän kokee suurta ylpeyttä kulttuurisista juuristaan. Kulttuurin häivytys siis vaikuttaa tekstin sävyyn, vaikkei yleinen merkitys muuttuisikaan. Konekääntimen haasteet intialaisten valuuttojen kääntämisessä voidaan myös nähdä esimerkkinä länsimaisen kulttuurin dominanssista konekääntimessä.

Raakakonekäännökseen käytetty konekäännin myös jätti reaalioita usein kääntämättä. Kone esimerkiksi tunnisti intialaisia ruokia kuvaavat termit, ja jätti ne alkuperäiseen muotoonsa. Joissakin tällaisissa tapauksissa konekäännin ei kuitenkaan ollut taivuttanut termejä oikein, joten kääntäjä on jälkieditointi vaiheessa lisännyt termeille oikeat taivutusmuodot. Esimerkiksi lähdetekstissä esiintyvä *pakoroita*-muodossa oleva termi oli käännetty raakakonekäännökseen *pakoroi*. Kääntäjä korjasi termille monikkomuodon *pakoras* jälkieditoituun käännökseen.

## 4.4 Selittävät käännökset

Kulttuuriviittauksia kääntäessä yksi yleinen strategia on lisätä käännökseen selitys, eli esimerkiksi lähdekielistä reaaliaa ei käännetä käännökseen laisinkaan, mutta kääntäjä lisää käännöksen tekstiä, joka selittää kohdekieliselle lukijalle mitä sana tarkoittaa. Selittävän käännökseen strategiaa oli hyödynnetty tämän tutkielman aineistossa useasti. Useimmiten konekäännin jäljitteli lähdeteoksen muotoilua, eikä lisännyt selitystä, mikäli sitä ei löytynyt lähdeteoksesta. Jälkieditoidusta käännöksestä kuitenkin löytyi kääntäjän lisäämiä selityksiä myös sellaisissa kohdissa, joissa sitä ei löytynyt lähdetekstissä.

Selittävä käännösstrategia on tämän tutkielman kontekstissa erityisen silmiinpistävä, koska Aura on hyödyntänyt sitä strategiaa jo lähdeteoksessa. Lähdeteos on suomenkielinen, mutta

koska teoksessa käsitellään intialaista kulttuuria ja kaksi päähahmoista ovat intialaisia, lähdeteoksessakin esiintyy huomattava määrä intialaisista kielistä peräisin olevia reaalioita. Joissakin tapauksissa lähdeteoksen kirjailija on tehnyt translatorisen päätöksen sisällyttää lähdetekstiin selityksen tällaisiin kohtiin.

(13) (ST) […] Mangala snanam -seremonian, […]
(MT) […] a ceremony of Mangala snanam […]
(PE) […] a ceremony of Mangala snanam […]

Esimerkissä 13 kirjailija on maininnut Mangala snanam -seremonian lähdetekstissä. Seremonia on intialaiseen hääkulttuuriin kuuluva ennen häitä suoritettava seremonia, joka tarkoittaa pyhää kylpyä ja jonka uskotaan tuovan suotuisuutta uuteen avioliittoon. Aura on lähdeteoksessa päättänyt käyttää seremonian alkuperäistä nimitystä, eikä ole pyrkinyt kääntämään sitä. Merkitystä selittääkseen hän nimeää *Mangala snanamin* seremoniaksi ja jopa selittää lyhyesti, mitä seremoniassa tapahtuu. Konekäännin on myös jättänyt reaalian kääntämättä ja sisällyttänyt selittävän seremonia-termin konekäännökseen. Jälkieditointivaiheessa on käytetty konekääntimen käännöstä, eli molemmissa käännösvaiheissa alkuperäinen reaalia on jätetty kääntämättä, käytettiin selittävää käännösstrategiaa ja kulttuurinen sisältö välittyy samanlaisena molemmissa käännöksissä.

Joissain tapauksissa konekäännin pyrki kääntämään myös reaalian. Aineiston perusteella intialaisista kielistä peräisin olevat reaaliat yleisemmin jätettiin kääntämättä ja niitä käytettiin sellaisenaan sekä lähdetekstissä että lopullisessa käännöksessä, ja joskus niiden yhteyteen lisättiin selittävä käännös. Raakakonekäännöksessä kuitenkin joitakin reaalioita oli pyritty kääntämään, vaikka niiden yhteydessä olisi jo ollut selittävä käännös.

(14) (ST) Oletko syönyt? Oota aaita?
(MT) Have you eaten? Wait a minute?
(PE) Have you eaten? Oota aaita?

Lähdeteoksessa on käytetty selittävää käännöstä kannadankielisen lausahduksen lisäksi (esimerkki 14). Konekäännin on kuitenkin pyrkinyt kääntämään lausahduksen englanniksi. Tämä kuitenkin lisää käännökseen kokonaan uuden merkityksen, jota ei esiinny lähdeteoksessa. Lisäksi se häivyttää kulttuurista sisältöä, sillä reaalian poisto antaa vaikutelman, että keskustelussa mukana olevat hahmot puhuisivat keskenään pelkästään englantia. Lopulliseen käännökseen reaalia on muutettu takaisin alkuperäiseen muotoonsa, mutta konekääntimen tekemä käännös fraasin selityksestä on säilytetty samanlaisena kuin se on raakakonekäännöksessä.

On myös mahdollista, että esimerkissä 14 esiintyvä konekäännös on hallusinaatio. Konekäännin on mahdollisesti ymmärtänyt kannadankielisen termin *oota* suomenkielisen odota-termin puhekielisenä versiona. Konekäännin on siis voinut olettaa kyseessä olevan suomenkielinen lausahdus ja luonut käännöksen sen perusteella. Hallusinaatio kuitenkin muuttaa lähdetekstin merkitystä huomattavasti.

Vastapainoisesti lähdetekstissä esiintyvät suomalaiset kulttuuriviittaukset ovat vaatineet kääntäjältä selittäviä käännöksiä englanninkieliseen käännökseen, vaikka selityksiä ei ole lähdetekstissä. Kääntäjä on esimerkiksi lisännyt selittävät käännökset *Domman lukusalien* ja *eximian* yhteyteen (esimerkit 15 ja 16), koska ne ovat suoraan suomalaiseen opiskelijakulttuuriin viittaavia termejä, joille ei löydy englanninkielistä vastinetta.

(15) (ST) […] Domman lukusaleissa.
(MT) […] in Doma's reading rooms.
(PE) […] in the reading rooms of the student housing and library complex, Domma.

(16) (ST) Sain arvosanaksi eximian.
(MT) I got a grade of eximia.
(PE) My grade was eximia, the second highest on a scale of seven.

Konekäännin ei tehnyt samanlaisia lisäyksiä raakakonekäännökseen. Sen sijaan konekäännin käytti calquing eli sanasta sanaan -käännösstrategiaa näiden reaalioiden kääntämisessä. Konekääntimen tekemät käännökset eivät häivyttäneet lähdetekstin kulttuurista sisältöä, mutta kääntäjän valitsemat selittävät käännökset muuttivat käännöstä helpommin luettavaksi ja kielenhuollollisesti puhtaammaksi.

Toisaalta aineistosta nousi esiin tapauksia, joissa lähdeteoksessa on ollut selittävä käännös, jonka jälkieditoija on poistanut lopullisesta käännöksestä. Tällaisia tapauksia esiintyi, kun lähdetekstissä mainittiin englanninkielisiä lausahduksia, jotka olivat lähdetekstissä sekä englanniksi että suomeksi.

(17) (ST) *God is everywhere.* Jumala on kaikkialla.
(MT) *God is everywhere*. God is everywhere.
(PE) God is everywhere.

Esimerkissä 17 puhutaan autoriksan ikkunaan kiinnitetystä englanninkielisestä tarrasta. Aura on lähdetekstiin kirjoittanut tarran tekstin englanniksi ja lisännyt perään selittävän käännöksen. Konekäännin on tunnistanut alkuperäisen englanninkielisen tekstin ja jättänyt sen sellaisenaan konekäännökseen. Se on myös kääntänyt suomenkielisen selittävän käännöksen englanniksi, jolloin käännöksessä toistuu sama englanninkielinen lause kahdesti. Jälkieditoija sen sijaan on

poistanut selittävän käännöksen kokonaan lopullisesta käännöksestä, koska se ei ole englanninkielisessä käännöksessä tarpeellinen.

Jälkieditoija myös lisäsi selittäviä osia käännökseen säilyttääkseen lähdeteoksen kielellisen ja kulttuurisen moninaisuuden. Vaikka käännöksessä ei ymmärrettävyyden vuoksi tarvita selittäviä käännöksiä englanninkielisille fraaseille, kuten esimerkistä 17 huomattiin, aineistossa esiintyi tapaus, jossa jälkieditoija on kuitenkin tehnyt käännöksestä eksplisiittisemmän.

(18) (ST) He ryhmittyvät odottamaan liikennevaloihin pysähtyvää bussia ja huutavat rytmikkäästi: – *Puffed rice, puffed rice, puffed rice!*
(MT) They cluster around the traffic lights waiting for a bus to stop and shout rhythmically: - *Puffed rice, puffed rice, puffed rice!*
(PE) They cluster around the bus at the traffic lights and shout rhythmically in English: "Puffed rice, puffed rice, puffed rice!"

Konekääntimen luoma käännös esimerkissä 18 on täysin toimiva. Jälkieditoija on kuitenkin lisännyt lopulliseen käännökseen tarkennuksen siitä, että kyseessä olevat henkilöt huutavat englanniksi. Tämä varmistaa sen, että lukija ymmärtää tilanteen monikielisyyden ja vahvistaa sitä, että huudot on suunnattu englantia puhuville turisteille. Huutelijat tilanteessa ovat intialaisia, eli lukija voisi olettaa huutojen olevan jollakin intialaisella kielellä, jos jälkieditoija ei selventäisi sitä käännöksessä.

Lähdetekstistä ja käännöksistä löytyi myös tapauksia, joissa reaalia on jätetty kääntämättä, mutta selittävää käännöstä ei ole lisätty. Esimerkiksi kannadan- ja hindinkielinen sana *pujari*, joka tarkoittaa pappia, on jätetty kääntämättä sekä raakakonekäännöksessä että kääntäjän jälkieditoimassa käännöksessä. Sanan yhteydessä ei myöskään käytetty selittäviä käännösstrategioita. Kääntäjän ja kirjailijan päätös voi kuitenkin johtua siitä, että sanan tarkoitus ilmenee sitä edeltävästä virkkeestä, missä keskustellaan hääpapin valikoinnista. Lukija siis ymmärtää mistä on kyse, vaikka hän ei tietäisi pujari-sanan merkitystä.

## 4.5 Amerikkalaisen kulttuurin vaikutus

Vaikka konekäännin ei lähtökohtaisesti tehnyt raakakonekäännökseen ylimääräisiä lisäyksiä, länsimaisen kulttuurin valta-asema konekääntimen koulutusdatassa tuotti joitakin kulttuurillisia lisäyksiä käännökseen. Käännettävän tekstin käsittelemät kulttuurit ovat intialainen kulttuuri ja suomalainen kulttuuri ja käännösprosessissa käytettiin Intian englannin varianttia. Konekäännin oli kuitenkin tulkinnut joitakin ilmaisuja muun muassa Yhdysvaltojen kulttuurikontekstissa.

(19) (ST) Kiitosjuhlien jälkeen en enää tuhlaa, vaan kerrytän säästöjä säntillisesti.
(MT) After Thanksgiving, I no longer spend, but build up my savings in a prudent way.
(PE) After the thank-you party, I stop spending and instead focus on building up my savings.

Lähdetekstissä mainitut *kiitosjuhlat* viittaavat päähenkilön ystävilleen järjestämiin juhliin, joilla hän kiitti ystäviään avusta uuden työn saamisessa (esimerkki 19). Konekäännin on kuitenkin ymmärtänyt *kiitosjuhlan* tarkoittavan kiitospäivää (*Thanksgiving*), joka on kansallinen juhlapyhä, jota juhlitaan esimerkiksi Yhdysvalloissa ja Kanadassa. Intiassa kiitospäivää ei juhlita. Konekääntimen käännösvalinta siis paitsi häivyttää lähdetekstin alkuperäistä merkitystä, mutta myös lisää tekstiin kulttuuriviittauksia kokonaan uudesta kulttuurista, johon ei löydy viittauksia lähdetekstistä ollenkaan.

Lähdeteoksessa mainitaan myös muutaman kerran ihmisten pituuksia. Lähdeteoksessa niistä puhutaan aina senttimetreissä. Yhdessä tapauksessa konekäännin oli kuitenkin muuttanut mainitun pituuden jalkoihin. Jälkieditoija on muuttanut yksikön takaisin senttimetreihin lopullisessa käännöksessä, mutta on käyttänyt yksikön ilmaisuun lyhennettä *cm*, toisin kuin lähdetekstissä.

(20) (ST) Hän on keskimittainen, ehkä 165 senttiä, ja hieman tukeva.
(MT) He is of average height, maybe 5'5", and a little stout.
(PE) He’s of average height, maybe 165 cm and a little heavy.

Intiassa pääosin noudatetaan metrijärjestelmää. Konekääntimen päätös muuttaa pituusyksikköä voidaan siis nähdä esimerkkinä länsimaisen, ja erityisesti pohjoisamerikkalaisen, kulttuurin dominanssista. Intiassakin ihmisten mitoista voidaan käyttää jalkoja, joten tämä esimerkki ei yksinään toimi todisteena dominanssille. Esimerkki 20 kuitenkin osoittaa konekääntimen johdonmukaisuuden puutteen sen käännösstrategioissa. Kun lähdetekstissä muualla mainitaan ihmisten pituuksia senttimetreissä, konekäännin on käyttänyt samaa yksikköä kuin lähdetekstissä. On huomionarvoista, että konekäännin noudatti samaa yksikköä tapauksissa, joissa pituuden ilmaisuun oli käytetty lyhennettä *cm*. Yllä mainitussa esimerkissä 20 yksikkö taas on kirjoitettu auki, jolloin konekäännin on muuttanut yksikköä.

Aineistosta nousi esiin muitakin tilanteita, jotka voidaan nähdä mahdollisena englannin kielen tai länsimaisen kulttuurin dominanssin esimerkkeinä, vaikkeivat ne yksinään toimisi kovin vahvoina todisteina siitä. Lähdetekstissä päähenkilö Taarini kuvailee häntä haastattelevia miehiä ja mainitsee toisen miehen jenkkiaksentin. Hän tarkoittaa sitä, että mies puhuu amerikkalaisella aksentilla.

(21) (ST) […] nasaali jenkkiaksentti.
(MT) […] a nasal Yankee accent.
(PE) […] a nasal American accent.

Konekäännin käänsi kirjailijan käyttämän jenkki-termin suoraan sanaksi *Yankee* (esimerkki 21). *Yankee* tai *Yank* on yleinen termi, jota käytetään Yhdysvaltojen ulkopuolella ja jolla viitataan Yhdysvalloista kotoisin olevaan henkilöön. Konekääntimen tekemä käännösvalinta on siis sopiva. Termillä *yankee* on kuitenkin myös muita merkityksiä. Yhdysvalloissa termiä on käytetty henkilöistä, jotka ovat kotoisin valtion pohjoisemmista osavaltioista, erityisesti Uudesta-Englannista. Suomen kielessä termillä *jenkki* kuitenkin viitataan kaikkiin amerikkalaisiin heidän kotiosavaltiostaan huolimatta. Yankee-termin käyttö voi siis mahdollisesti liittää tekstiin ylimääräisiä konnotaatioita, joita ei ole lähdetekstissä. Alkuperäisessä tarinassa Taarini kuvailee miestä, joka on kotoisin Intiasta, mutta on asunut Yhdysvalloissa, josta syystä hän puhuu amerikkalaisella aksentilla. Miehen kotiosavaltiota ei mainita. Konekäännöksen lukija voisi kuitenkin tehdä olettamuksen siitä, mistä päin Yhdysvaltoja mies on. Jälkieditoija onkin muuttanut termin lopulliseen käännökseen muotoon *American accent*, jolloin lähdetekstin alkuperäisestä tarkoituksesta ei voi erehtyä.

# 5 TULOKSET JA POHDINNAT

Tutkielman tavoitteena on tutkia, miten konekäännin kääntää kulttuuriviittauksia kaunokirjallisuudessa ja millaisia ongelmia viittausten käännöksissä mahdollisesti esiintyy. Toisena tavoitteena on tarkastella, näkyykö englannin kielen ja länsimaisen kulttuurin dominanssi konekäännöksessä.

Aineiston analyysista ilmenee, että konekäännin ei luo johdonmukaista käännöstä. Se kääntää samoja termejä usealla eri tavalla läpi tekstin sen sijaan, että se käyttäisi termeille jo aiemmin valitsemiaan käännöksiä. Esimerkiksi *myötäjäiset* ja *rupiat* on käännetty muualla aineistossa termeiksi *dowry* ja *rupees*. Esimerkeissä 10 ja 11 kuitenkin nähdään, miten konekäännin poikkeaa omista käännösvalinnoistaan ja on kääntänyt termit eri tavalla. Tämä todistaa konekääntimen johdonmukaisuuden puutteen. Tämä on ongelmallista myös siksi, että konekäännin muuttaa esimerkkien kulttuurillista sisältöä ja häivyttää lähdetekstin merkitystä. Näistä esimerkeistä voidaan myös päätellä, että konekäännin ei näe kääntämäänsä tekstiä kokonaisuutena, vaan kääntää tekstin pienempinä segmentteinä yhden kokonaisuuden sijaan.

Konekäännin ei myöskään tee lisäyksiä tai poistoja käännökseen, vaikka ne olisivat tarpeellisia. Esimerkissä 17 nähdään, miten Aura on alkuperäisessä teoksessaan lisännyt selittävän käännöksen englanninkieliselle virkkeelle ”God is everywhere”. Jälkieditoija on poistanut alkuperäisessä tekstissä olevan selittävän käännöksen, koska se ei ole englanninkielisessä käännöksessä tarpeellinen. Tämä valinta poistaa virkkeen turhan toiston käännöksestä ja tekee käännetystä tekstistä luettavuuden kannalta toimivamman. Konekäännin taas ei poistanut selittävää käännöstä, joten sama lause esiintyi raakakonekäännöksessä kaksi kertaa peräkkäin.

Jälkieditoinnissa käännökseen on myös lisätty selittäviä käännöksiä esimerkiksi suomalaisten kulttuuriviittausten yhteyteen, jota ei suomenkielisessä lähdeteoksessa esiinny. Esimerkeissä 15 ja 16 nähtiin, miten jälkieditoija on lisännyt selittävät käännökset suomalaiseen korkeakoulukulttuuriin liittyviin kulttuuriviittauksiin. Konekäännin ei ole tehnyt selittäviä lisäyksiä raakakonekäännökseen. Tämä voi vaarantaa merkityksen välittymisen, sillä englanninkielinen lukija ei välttämättä tunne suomalaisten viittausten, kuten eximia-arvosanan, merkitystä. Selittävä käännösstrategia siis varmistaa sen, että lähdetekstin kulttuurinen konteksti välittyy myös käännöksen lukijalle. Konekäännin käytti selittävän käännösstrategian sijaan sanasanaista käännösstrategiaa.

Sanasanainen käännösstrategia vaikuttaa aineiston perusteella olevan konekääntimen oletusstrategia. Useammassa esimerkissä konekääntimen valitsema käännös vaikuttaa olevan pyrkimys sanasanaiseen käännökseen silloin, kun termille ei ole löytynyt suoraa vastinetta. Esimerkiksi hindin kielestä peräisin olevalle bindi-termille (esimerkki 8) ei löydy englanninkielistä vastinetta, joten konekäännin todennäköisesti pyrki kääntämään termin sanasanaisesti. Myös esimerkissä 10 konekäännin on ilmeisimmin pyrkinyt kääntämään termin *myötäjäiset* sanasanaisesti. Sanasanainen käännösstrategia on haasteellinen tällaisissa tapauksissa, koska sen sijaan että se kääntäisi lähdetekstin ymmärrettävästi ja kulttuurisen sisällön säilyttäen, konekäännin on päätynyt häivyttämään tai muuttamaan lähdeteoksen kulttuurista merkitystä käännöksessä. Erityisesti esimerkissä 10 käännöksen lukijalle voi välittyä romantisoitu ymmärrys keskustelun aiheesta. Tämä on erityisen haitallista, koska lähdetekstin todellisena keskustelun aiheena on se, että Taarinin pikkusiskon uusi mies on paljastunut väkivaltaiseksi ja alkanut vaatimaan myötäjäisiä takautuvasti. Sanasanainen käännös siis paitsi häivyttää lähdetekstin kulttuurista sisältöä, mutta myös muuttaa sitä merkittävästi.

Aineistosta myös ilmeni se, että englannin kielellä ja länsimaisella kulttuurilla on merkittävä dominanssi konekääntämisessä. Dominanssi esiintyi usealla eri tavalla aineistossa. Konekääntimellä oli haasteita tunnistaa henkilöiden nimiä. Kone oli esimerkiksi luullut naisen nimeä kaupungin nimeksi (esimerkki 1). Tämä viestii länsimaisen kulttuurin dominanssista, koska nimi on intialainen. Vastakohtaisesti konekäännin on mahdollisesti luullut sari-termiä suomalaiseksi naisen nimeksi, eikä intialaiseksi naisen vaatteeksi, johon lähdetekstissä viitataan (esimerkki 3). Näissä tapauksissa konekäännin on siis muuttanut tai häivyttänyt lähdetekstissä esiintyvien intialaisten reaalioiden merkitystä huomattavasti.

Toisaalta raakakonekäännöksessä esiintyy tapauksia, joissa konekääntimellä oli haasteita myös länsimaiseen kulttuuriin kytköksissä olevien erisnimien kääntämisessä. Suomalainen miehen nimi *Markus* on raakakonekäännöksessä käännetty englanniksi *Mark* (esimerkki 2). Konekäännin on mahdollisesti tehnyt käännösvalinnan sen perusteella, että evankelisessa uskomuksessa *Markus* kääntyy raamatullisissa teksteissä nimeksi *Mark*. Koska lähdetekstissä ei kuitenkaan ole kyse Raamatun henkilöistä eikä kappaleessa keskustella kristinuskoon liittyvistä aiheista, käännösvalinta lisää käännökseen kuulumatonta kulttuurista sisältöä. Konekäännin ei myöskään tunnistanut Humiseva harju -kirjan (esimerkki 4) tunnettua käännöstä vaan pyrki kääntämään suomennetun nimen sanasanaisesti. Tämä saattaa viitata

siihen, että konekääntimen suomenkielinen koulutusdata ei ole tarpeeksi kattava tällaisten kulttuuriviittausten kääntämiseen.

Länsimaisen kulttuurin dominanssi välittyy myös implisiittisissä kulttuuriviittauksissa. Vaikka konekääntimen käännösvalinnat näissä viittauksissa olivat useimmiten kielenhuollollisesti oikein, ne häivyttivät lähdeteoksen kulttuurillista sisältöä. Konekäännin ei käyttänyt järjestettyjä avioliittoja käsittelevässä virkkeessä käännöstä *arrange*, vaan valitsi termin *organise* (esimerkki 6). Vaikka sanat ovat synonyymeja, arrange-termiä käytetään yleisemmin järjestetyistä avioliitoista puhuttaessa. Termeillä on siis erilainen kulttuurinen konnotaatio. Jälkieditoijan tekemä muutos termivalinnassa siis vahvistaa käännöksen kulttuurista kontekstia, kun raakakonekäännöksessä se on häivytetty. Vastaavanlaisesti konekäännin oli valinnut kääntää termin asemaa englanniksi *position*, kun lähdetekstissä viitataan Intian kastilaitokseen (esimerkki 5). Vaikka termi voi olla sujuva käännös sanalle, tässä tapauksessa se häivyttää lähdeteoksen kulttuurista merkitystä. Kastilaitokseen viitatessa käytetään yleensä englanninkielistä termiä *status*. Koneen käyttämällä position-termillä taas on erilainen konnotaatio, koska sillä viitataan yleensä ihmisen asemaan esimerkiksi työpaikalla. Konekääntimen käännösvalinta siis häivyttää jälleen käännöksestä intialaisen kulttuurin kontekstia. Nämä esimerkit myös todistavat, että konekääntimen on vaikea ymmärtää tekstin ulkopuolista kontekstia, eli konekäännin kääntää vain tekstin, mutta ei siihen liittyvää kontekstia.

Länsimaisen kulttuurin dominanssi näkyy myös konekääntimen käännösvalinnoissa reaalioita kääntäessä. Konekääntimellä oli haasteita tunnistaa intialaisten kielten reaalioita, jolloin kulttuurinen sisältö joko poistui kokonaan raakakonekäännöksestä tai muuttui merkittävästi. Konekäännin ei myöskään esimerkiksi tunnistanut intialaisia valuuttoja ja käänsi ne täysin eri merkityksellä tai poisti ne tekstistä kokonaan (esimerkit 11 ja 12). Joidenkin intialaisten reaalioiden kohdalla konekäännin loi hallusinaatioita, eli täysin lähdetekstin kontekstista irrallisia käännöksiä. Esimerkissä 10 konekäännin näyttää yhdistäneen termin länsimaisessa hääkulttuurissa yleiseen sanontaan ja luoneen käännöksen sen pohjalta. Esimerkissä 14 konekäännin taas vaikuttaa ymmärtäneen kannadankielisen termin olevan suomenkielisen termin puhekielinen versio, jonka se on pyrkinyt kääntämään. Molemmat hallusinaatiot muuttavat lähdetekstin kulttuurillista merkitystä huomattavasti.

Jotkin konekääntimen käännösvalinnat viittaavat myös siihen, että konekääntimessä on erityisesti amerikkalaisen kulttuurin dominanssi. Esimerkissä 19 nähdään, miten konekäännin

on lisännyt pohjoisamerikkalaista kulttuuria raakakonekäännökseen kääntämällä kiitosjuhla-termin englanniksi *Thanksgiving*, joka on pohjoisamerikkalainen juhlapyhä. Tämä lisää tekstiin ylimääräistä kulttuurista sisältöä, jota ei löydy lähdeteoksesta. Konekäännin oli myös kääntänyt lähdeteoksessa mainitut sentit jalkoihin ja tuumiin. Konekääntimen käyttämät mittayksiköt ovat käytössä Yhdysvalloissa, mikä voi viitata amerikkalaisen kulttuurin dominanssiin. Tämä ei yksinään kuitenkaan toimi todisteena dominanssille, koska jalat ja tuumat ovat toisinaan käytössä myös Intiassa, vaikka maassa virallisesti käytetään metrijärjestelmää. Kun tapaus yhdistetään aineistosta muihin löytyviin esimerkkeihin, voidaan todeta, että konekääntimen koulutusdatan olevan todennäköisimmin amerikkalaispainotteista.

Aineistosta löytyneiden esimerkkien nojalla voidaan todeta, että konekääntimen koulutusdatassa on merkittävä englannin kielen ja länsimaisen kulttuurin dominanssi. Tämä vaikuttaa erityisesti muita kulttuureja sisältävien tekstien kääntämiseen. Tässä tutkielmassa tutkittu lähdeteos käsittelee intialaista ja suomalaista kulttuuria, eikä lähdetekstissä juurikaan käsitellä englanninkielisiä kulttuureja. Lisäksi raakakonekäännös on luotu käyttäen Intian englannin varianttia. Konekäännin on silti useasti tehnyt käännösvalintoja, jotka vaikuttavat juontavan juurensa länsimaiseen, ja erityisesti amerikkalaiseen, kulttuuriin. Tuloksena on raakakonekäännös, jossa lähdeteoksen kulttuurinen sisältö on häivyttynyt, muuttunut tai ajoittain poistettu kokonaan. Tämä ongelma voitaisiin pyrkiä korjaamaan esimerkiksi suuremmalla koulutusdatalla. Konekääntimen koulutusdataan tulisi sisällyttää enemmän erityisesti intialaisia kulttuureja käsitteleviä rinnakkaistekstejä, jolloin konekääntimen olisi helpompi valikoida myös kulttuurillisesti sopivat käännökset termeille.

Raakakonekäännöksen voidaan myös aineistosta esiin tulleiden esimerkkien perusteella todeta olevan liian huonolaatuista ollakseen julkaistavaa kaunokirjallista tekstiä. Erityisesti tällaisessa teoksessa, jossa kulttuurinen sisältö on valtavassa asemassa tekstissä, tulisi lähdeteoksen merkityksen välittyä käännöksessä vahvemmin, kuin mitä tutkitussa raakakonekäännöksessä se välittyi. Jälkieditoija on ajoittain tehnyt merkittäviäkin muutoksia kulttuuriviittauksiin, jotta lähdetekstin merkitys on välittynyt käännöksessä. Toisaalta jälkieditoija on tehnyt myös luettavuuteen liittyviä muutoksia, kuten lisännyt selittäviä käännöksiä tai muuttanut intialaisten kielien reaalioita takaisin alkuperäiseen muotoonsa. Myös tapauksissa, joissa konekäännin oli säilyttänyt kulttuurisen kontekstin, on jälkieditointivaiheessa tehty kielenhuollollisia korjauksia, jotka tekevät käännöksestä sujuvamman ja luettavamman.

Aineistosta myös ilmeni konekääntimen vahva sukupuolipreferenssi. Lähes jokainen lähdetekstissä esiintyvä hän-pronomini on raakakonekäännöksessä käännetty englannin kielen maskuliiniseksi pronominiksi *he*. Konekäännin käytti maskuliinista pronominia jopa sellaisissa tapauksissa, joissa virkkeen muusta kontekstista pystyy päättelemään kyseessä olevan henkilön olevan nainen. Tästä voidaan päätellä, että konekäännin kääntää hän-pronominin oletusarvollisesti maskuliiniseksi. Tämä on erityisen haitallista Myötäjäiset-teoksen käännöksessä, koska teoksen suurin osa hahmoista on naisia ja teos käsittelee nimenomaan naisten asemaa Intiassa. Sukupuolittunut ja maskuliinisuutta suosiva käännös siis myös häivyttää alkuperäisen teoksen merkitystä. Ilmiö toimii myös lisätodisteena sille, ettei konekäännin ymmärrä kääntämäänsä tekstiä kokonaisuutena, vaan tekee käännösvalinnat katsoen hyvin pientä segmenttiä tekstistä kerrallaan.

Tutkielman tulosten perusteella voidaan siis todeta, että konekäännin ei tuota julkaisuun kelpaavaa kaunokirjallista tekstiä ilman jälkieditoijan osallistumista. Aineistossa esiintyy myös vahva englannin kielen ja länsimaisen kulttuurin dominanssi. Tutkielmassa tarkasteltavat esimerkit luovat melko kattavan kuvan siitä, millä tasolla NMT-konekäänninten käännösten laatu on. Erityisesti kulttuuririkasta tekstiä kääntäessä konekääntimellä on haasteita kulttuurisen sisällön kääntämisessä ymmärrettävästi, varsinkin kun on kyse kulttuuriviittauksista, jotka eivät kuulu länsimaiseen tai englanninkieliseen valtavirtaan. Näitä ongelmia voidaan tietyissä määrin korjata käyttämällä laajempaa koulutusdataa, joka sisältää laajemmin rinnakkaismateriaalia tekstin käsittelemistä kulttuureista. Aineistosta poimitut esimerkit myös todistavat sen, että konekäännin ei pysty tuottamaan luovaa tekstiä yhtä laadukkaasti kuin ihmiskääntäjä, joten luovien tekstien, kuten kaunokirjallisuuden, kääntämiseen tarvitaan edelleen ihmiskääntäjiä, mikäli halutaan tuottaa laadukasta ja luettavuuden kannalta miellyttävää tekstiä.

Tämän tutkielman tulokset luovat esimerkin NMT-kääntimien luomien kaunokirjallisten käännösten laadusta. Tutkielmassa ei hyödynnetty virallisia laadunmittareita, kuten BLEU-arviointia, joka voisi antaa laajemman käsityksen käännösten kokonaisvaltaisesta laadusta. Kattavamman käsityksen saamiseksi tulisi kuitenkin tutkia useamman kaunokirjallisen tekstin raakakonekäännöksiä. Tämä antaisi muun muassa kattavamman käsityksen siitä, miten paljon konekäännökset muistuttavat toisiaan ja välittyykö niistä alkuperäiskirjailijan ääni. Kattavampaa kuvaa konekäänninten tämänhetkisestä tehokkuudesta edistäisi myös se, jos konekäännöksistä tutkittaisiin kulttuuriviittausten lisäksi muitakin tekstin osia. Tämänkin

tutkielman aineistosta nousi esiin konekääntimen merkittävä maskuliinisuuspreferenssi, joka itsessään vaatii laajempaa tutkimusta.

Jatkotutkimuksissa olisi myös eduksi tarkastella useamman NMT-konekääntimen käännöksiä. Tässä tutkielmassa käytettiin esimerkkinä yhtä konekäännöstä, joka on luotu yhdellä NMT-konekääntimellä. Jotta saataisiin kattava käsitys NMT-teknologian nykyisestä kehityksestä ja tilanteesta, tulisi tutkia myös muita NMT-kääntimiä DeepL-kääntimen lisäksi. Jatkotutkimuksissa olisi myös hyödyllistä tarkastella LLM-järjestelmillä tehtyjä konekäännöksiä, koska ne ovat tällä hetkellä johtavassa asemassa konekäänninteknologiassa.

# LÄHTEET

## Aineistolähteet

Aura, Siru, 2022. *Myötäjäiset*. Helsinki: Stresa Kustannus

## Kirjallisuuslähteet

Abdul-Hadi, Lubna A. & Khetam Shraideh, 2025. Translation technologies in humanitarian work: expectations and challenges among NGOs in Jordan. *Tradumàtica* 2025:12, 286–307. https://doi.org/10.5565/rev/tradumatica.517.

Abrams, M.H. & Harpham, Geoffrey Galt, 2019. *A Glossary of Literary Terms*. 11. painos. Stamford: Cengage Learning.

Aixelá, Javier Franco, 1996. Culture-specific Items in Translation. Teoksessa Román Álvarez & M. Carmen-África Vidal, *Translation, Power, Subversion.* Clevedon: Multilingual Matters, 52–78.

Albir, Amparo Hurtado, 2017. Researching Translation Competence by PACTE Group. John Benjamins Publishing Company. https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=4791152.

Aura, Siru, 2008. Women and Marital Breakdown in South India: Reconstructing Homes, Bonds and Persons. Helsingin yliopisto, väitöskirja. https://helda.helsinki.fi/server/api/core/bitstreams/a915be90-fc4c-44b8-b7b7-247acf0e2063/content.

Baldwin, John R., 2005. *Redefining Culture: Perspectives Across the Disciplines*. 1. painos. Oxford: Taylor & Francis Group. https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=274525.

Bassnett, Susan & André Lefevere, 1998. *Constructing cultures: essays on literary translation*. Clevedon: Multilingual Matters.

Carbonell, Ovidio, 1996. The Exotic Space of Cultural Translation. Teoksessa Román Álvarez & M. Carmen-África Vidal, *Translation, Power, Subversion.* Clevedon: Multilingual Matters, 79–98.

Castilho, Sheila, Stephen Doherty, Federico Gaspari & Joss Moorkens, 2018. Approaches to Human and Machine Translation Quality Assessment. Teoksessa *Translation Quality Assessment: From Principles to Practice*. Springer International Publishing AG, 9–38. https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=5452404.

Colina, Sonia, 2011. Evaluation/Assessment. Teoksessa Luc van Doorslaer & Yves Gambier, *Handbook of Translation Studies online*. Alankomaat: John Benjamins Publishing Company. https://doi.org/10.1075/hts.2.eva1.

Daems, Joke, Orphée De Clercq & Lieve Macken, 2017. Translationese and Post-editise: How comparable is comparable quality? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 89–103. https://doi.org/10.52034/lanstts.v16i0.434.

Debussche, Julien & Jean-Christophe Troussel, 2014. Machine-aided translations. *Studies on translation and multilingualism: Translation and intellectual property rights*. European Union. 102–103. https://op.europa.eu/en/publication-detail/-/publication/e079e290-e250-482d-9d4f-dae566ba67ff/language-en.

European Commission, 2026. European Master's in Translation. Verkkojulkaisu. https://translation.ec.europa.eu/get-involved-european-language-activities-and-initiatives/european-masters-translation_en.

Folkart, Barbara, 1996. Polylogie et registres de traduction: Le cas d'Ulysses. *Palimpsestes* 10, 125–140. https://journals.openedition.org/palimpsestes/1513.

Fernández Guerra, Ana, 2012. Translating culture: problems, strategies and practical realities. *SIC – A Journal of Literature, Culture and Literary Translation* 3:1. https://doi.org/10.15291/sic/1.3.lt.1.

Goudaec, Daniel, 2010. Quality in translation. Teoksessa Luc van Doorslaer & Yves Gambier, *Handbook of Translation Studies online*. Alankomaat: John Benjamins Publishing Company. https://doi.org/10.1075/hts.1.qua1.

Guerberof-Arenas, Ana & Antonio Toral, 2020. The impact of post-editing and machine translation on creativity and reading experience. *Translation Spaces* 9:2, 255–282. https://doi-org.libproxy.tuni.fi/10.1075/ts.20035.guev.

Hokkanen, Sari, 2025. What does meaning move through? Identifying epistemic domains in an AI-assisted literary translation project. SKY Symposium 2025. Tampereen yliopisto 2.10.2025.

Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto & Pascale Fung, 2023. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys* 55:12, 1–38. https://doi-org.libproxy.tuni.fi/10.1145/3571730.

Kade, Otto, 1981. *Probleme des übersetzungswissenschaftlichen Textvergleichs*. Leipzig: VEB Verlag Enzyklopädie.

Katan, David, 1999. *Translating cultures: an introduction for translators, interpreters and mediators*. Manchester: St. Jerome.

Koby, Geoffrey S., Paul Fields, Daryl R. Hague, Arle Lommel & Alan Melby, 2014. Defining Translation Quality. *Tradumàtica* 2014:12, 413–423. https://revistes.uab.cat/tradumatica/article/view/n12-koby-fields-hague-etal.

Kolb, Waltraub & Tristan Miller, 2022. Human-computer interaction in pun translation. Teoksessa Carlos S.C. Teixeira, Kristiina Taivalkoski-Shilov, James Luke Hadley & Antonio Toral, *Using Technologies for Creative-Text Translation*. Yhdistynyt kuningaskunta: Routledge, 66–88. https://doi-org.libproxy.tuni.fi/10.4324/9781003094159.

Kong, JinQiu & Wenbo Lu, 2025. BLEU-Based Evaluation of Papago Translation Quality. *Proceedings of The 2nd International Conference on Intelligent Education and Computer Technology, IECT 2025,* 709–716. https://doi-org.libproxy.tuni.fi/10.1145/3764206.3764317.

Leppihalme, Ritva, 2011. Realia. Teoksessa Luc van Doorslaer & Yves Gambier, *Handbook of Translation Studies online*. Alankomaat: John Benjamins Publishing Company. https://doi.org/10.1075/hts.2016.rea1.

Mayell, Hillary, 2003. India's "Untouchables" Face Violence, Discrimination. *National Geographic News*. https://www.nationalgeographic.com/pages/article/indias-untouchables-face-violence-discrimination.

Mohanty, Sachidananda, Mousumi Sen & Geeta Sau 2013. Analysis of risk factors of dowry death – A south Indian study. *Journal of Forensic and Legal Medicine,* 20:4, 316–320. https://doi-org.libproxy.tuni.fi/10.1016/j.jflm.2012.09.027.

O'Brien, Sharon, 2021. Post-editing. Teoksessa Luc van Doorslaer & Yves Gambier, *Handbook of Translation Studies online*. Amsterdam: John Benjamins Publishing Company, 177–183. https://doi.org/10.1075/hts.5.pos3.

OED = MOT Oxford Dictionary of English 2026. status. https://www.sanakirja.fi/l_oxford_english/english-english/status. (28.2.2026)

Olojo, Seyi, Janina Zakrzewski, Andrew Smart, Erin van Liemt, Milagros Miceli, Amber Ebinama & Lameck Mbangula Amugongo, 2025. Lost in Machine Translation: The Sociocultural Implications of Technologies in Nigeria. *FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency,* 2249–2259. https://doi.org/10.1145/3715275.3732153.

Rothwell, Andrew, Joss Moorkens, María Fernández-Parra, Joanna Drugan & Frank Austermuehl, 2023. Current Machine Translation Technologies. Teoksessa *Translation Tools and Technologies*. 1. painos. Yhdistynyt kuningaskunta: Routledge, 97–113. https://www.taylorfrancis.com/reader/read-online/c73c6013-7a87-43a0-a696-7e5a545ab39e/chapter/pdf?context=ubx.

Schiopu, Lucia, 2025. Teaching Terms, Culturemes and Realia in Teacher Education Language Programs. *ACROSS Journal of of Interdisciplinary Cross-Border Studies,* 9:3, 75–84. https://doi.org/https://doi.org/10.35219/across.2025.3.07.

SFS-EN ISO 17100:2015 + A1:2017. Käännöspalveluja koskevat vaatimukset.

Sun, Yifeng, 2012. (Un)translatability and cross-cultural readability. *Perspectives: Studies in Translatology* 20:2, 231–247. https://andor.tuni.fi/view/action/uresolver.do?operation=resolveService&package_service_id=26398420900005973&institutionId=5973&customerId=5965&VE=true.

Taivalkoski-Shilov, Kristiina, 2019. Ethical issues regarding machine(-assisted) translation of literary texts. *Perspectives* 27:5, 689–703. https://doi.org/10.1080/0907676X.2018.1520907.

TEPA = TEPA-termipankki 2026. kulttuuri. https://termipankki.fi/tepa/fi/haku/kulttuuri. (21.1.2026)

TEPA = TEPA-termipankki 2026. position. https://termipankki.fi/tepa/fi/haku/position. (28.2.2026)

TEPA = TEPA-termipankki 2025. reaalia. https://termipankki.fi/tepa/fi/haku/reaalia. (14.10.2025)

Toral, Antonio, & Way, Andy, 2015. Machine-assisted translation of literary text: A case study. *Translation Spaces*, 4:2, 240–267. https://doi.org/10.1075/ts.4.2.04tor.

Vinay, Jean-Paul & Jean Darbelnet, 1995. *Comparative stylistics of French and English: methodology for translation*. Amsterdam: John Benjamins Publishing Company. https://ebookcentral.proquest.com/lib/tampere/detail.action?docID=622331.

Yoon, Sun Kyoung, 2021. Decolonising Acts: An Analysis of Deborah Smith's English Translation, *Human Acts*. *Acta Koreana* 24:2, 75–92. https://doi.org/10.18399/acta.2021.24.2.004.

# ENGLANNINKIELINEN LYHENNELMÄ

# MACHINE TRANSLATION AND POST-EDITING IN A LITERARY TEXT

## Introduction

Machine translation technology has been developed since the 1950s, and it is used increasingly across the field of translation (Rothwell et al. 2023). The interest towards using machine translation in literature has been growing, even though its use in literary translation is not as frequent as in other areas (Taivalkoski-Shilov 2019). In this thesis, I am studying the use of machine translation technology in a literary text and specifically looking at translating cultural references in a literary text with a machine translation tool. The data of this thesis is collected from the first three chapters of Siru Aura's novel *Myötäjäiset*, a raw machine translation of the chapters and a finalized post-edited translation of the chapters. The machine translation and the post-edited translation were made by Sari Hokkanen, an English language lecturer from Tampere University. Hokkanen is translating the novel from Finnish to English by using a raw machine translation which she then post-edits. The translation is then used for her research project, in which she studies the transfer of cultural knowledge through different epistemic domains. Another goal of the project is to publish the English translation of the novel and bring in "back to India" where the story of the novel takes place. Hokkanen's translation process and research project are still on going at the time this thesis was written, so the finalized post-edited translation used for this thesis' data may still change by the time of publication of the English translation of the novel.

*Myötäjäiset* is Aura's fictional novel that follows three women in India in the early 2000s. The novel's central themes are women's rights and dowry deaths in India. Dowry death is a term used for the death of a young woman, who dies within 7 years of her nuptials from murder or suicide due to continuous harassment from her new husband and/or in-laws (Mohanty et al. 2013). The harassment is motivated by her new family's wish to receive a better dowry payment from her parents (ibid.). The giving and receiving of a dowry was a common practice in Indian

marriage tradition, where the parents typically arranged marriages for their children and the woman's parents paid her new husband a dowry. The payment and receiving of dowries have been made illegal in India, but some husbands and in-laws still felt the dowry would be their right, which in severe cases can lead to a dowry death. In 2000, there were over 6000 dowry deaths recorded in India (Mohanty et al. 2013). Siru Aura's novel follows three women whose lives have been affected by a dowry death. Aura is a Finnish author and anthropologist who wrote her dissertation on the same subject based on an ethnographic field study she conducted in India (Aura 2008). While the events and characters in her book are fictional, they are inspired by her own experiences during her time in India. The novel portrays both Indian and Finnish cultural references. Two of the main characters are Indian women and one is a Finnish anthropologist doing a study on women's rights and the Indian marriage culture in India. Because of the multicultural and multilingual nature of the novel, it offers a rich foundation for this thesis to study the translation of cultural references. In the case of this novel, the text is not simply translated from the source culture and language to a singular target culture and language, but there are two cultures and multiple languages already present in the source text. The source text is translated into English which brings in a third language culture into the process, which is why this novel is a good subject for studying the English language dominance within machine translation.

In this thesis I will attempt to answer two questions: 1. How does the machine translate cultural references? And 2. Are there instances of English language or Western culture dominance in the machine translation? The aim of the first question is to study whether the cultural meaning of the original text transfers understandably in the raw machine translation or if input from the post-editor is needed. The aim of the second research question is to see whether the dominant position of English language and Western culture is evident in the machine translation. The overall goal of this thesis is to study whether the machine translation tool creates publishable quality translations of culturally rich literary text independently.

## Theoretical framework

To identify cultural references in the text, we must first define what is meant by culture and cultural reference. Giving the concept of culture a finite definition has historically been difficult as it can be understood to mean a variety of things depending on the circumstances (Baldwin 2005). In this thesis, culture is understood as a set of beliefs, tradition, manners, and values

shared by a group of people. Cultural reference as a term is similarly difficult to finitely define, but in the context of this thesis it is understood as a term, phrase or sentence, that is expressing or commenting on an action, tradition, material things, values or other aspects that are directly connected to a specific culture. An example of a cultural reference from the data of this thesis would be the comments and terms used to describe the Indian marriage traditions that are also prevalent to the theme of the novel.

Translating culture and cultural references is a key part of the translation process especially when translating a culturally rich text such as the novel that this thesis studies. It has been agreed that cultural understanding and knowledge is part of professional translator's competences (EMT Competence Framework 2022, 6). To produce a good quality translation and to transfer the cultural meaning of the original text into another language and culture means that the translator must have knowledge of cultural phenomena of the source culture along with the linguistic skills (Albir 2017, 40).

Translators may use several different strategies when translating cultural references. Some of the most common ones are calquing, explicitation, and loan words (Yoon 2021; Leppihalme 2011). Also changing the cultural references into ones understood by the target audience may be used, however, this strategy is typically only used in cultures where this is an accepted norm for a translation (Leppihalme 2011, 128–129). This strategy can be seen as a colonizing strategy, because it has been used for example when translating Chinese texts into European languages, where the original text's meaning changed so significantly that the reader may not even realize the text originated from China (Sun 2012). Nowadays, some translators utilize other, decolonizing translation strategies in order to bridge the cultural gap that has developed between the dominant Western cultures and other more exoticized cultures (Yoon 2021). This is also where the English language dominance derives from. Since English holds a position as a global lingua franca, many translations tend to lean towards English traditions and linguistic styles (Bassnett & Lefevere 1998). This can be harmful since the dominant position of English threatens smaller resource languages and their traditions (Olojo et al. 2025).

The raw machine translation used in this thesis was done by using a DeepL translation tool that was integrated into Wordfast translation memory program. While the current version of DeepL uses the large language model (LLM) system, the version used for this translation is a traditional neural machine translation (NMT) system. Neural machine translation systems work by calculating the most likely translation for a word or a phrase based on an artificial neural

network that they create based on a training data (Rothwell et al. 2023, 102–104). NMT training datasets are comprised of reference texts and translations and now with the new developments in artificial intelligence the data can also include pictures, videos, and texts the machine can vacuum from the internet. The NMT system can also calculate context to a certain extent which is why it is considered to produce more fluent sounding text to previous machine translation systems and has gained interest in the literary translation field as well (Taivalkoski-Shilov 2019, 690).

While this thesis does not aim to make a comprehensive qualitative study of the material, understanding the concept of quality within the translation field is helpful in understanding this thesis. The definition of quality has raised controversy within the translation field, although the common shared understanding is that all translations must contain some level of adequacy and fluency (Koby et al. 2014). When looking at the translation as a product, the transaction is good when both the translator and the customer are satisfied (Gouadec 2010). However, pinpointing the product quality in a translation can be difficult because it is difficult to define who the customer is in a translation transaction. Translations typically have both the person who commissioned the translations and the person/people who use them, the readers. Ideally, all three parties would be satisfied with the translation (Koby et al. 2014).

Translation quality is also difficult to determine because different people view quality differently and there is no finite agreement on the parameters that make up a good quality translation (Colina 2011). Often the parameters in a given evaluation tend to showcase the values and opinions of the evaluator (ibid.). Additionally, the quality of a translation is dependent on the circumstance in which it is used, which further complicates the matter of narrowing down quality markers. A translation could be adequate quality for the communication situation in which it is needed but not good enough quality in another situation (Gouadec 2010).

Defining quality in translation is equally difficult with machine translations. This is also due to the circumstantial nature of the translations. Most machine translations can largely be divided into two categories: MT for gisting and MT for post-editing (Rothwell et al. 2013, 97). Gisting means that the reader simply needs to understand the main points of the translated text, which also means that the translation does not need to be “perfect”. MT for post-editing means that the machine translation is used as a foundation for a text that a translator further edits until it becomes of publishable quality.

What tends to be the issue regarding machine translation quality is the way it is received. Readers tend to respond better to translations that are either entirely translated by human translators or at the very least post-edited by a human translator (Guerberof-Arenas & Toral 2020). Machine translations tend to have a specific "voice" which can also change the voice of the original author, which creates another level of issues both ethically and quality-wise in machine translations (Daems, Clerq & Macken 2017). Machine translations have in general created new ethical issues regarding translations. One of the main ethical issues with machine translations is the fact that customers and businesses tend to want to pay less for translations if an MT tool or a CAT tool was used in the creation of the translation (Taivalkoski-Shilov 2019). With the ever-growing time pressure and the need to get the translations as fast as possible, this jeopardizes the already rather low compensations translator get while increasing the mental load of the work. The need to create translations at a fast pace also inevitably causes the quality of the translations to decrease since translators do not have time to proof-read and finalize the translations and more mistakes may slip through (ibid.).

Machine translation tools are also increasingly used by people who are not professional translators (Lubna & Shraideh 2025, 303). Since it cannot be expected for non-professional translators to have the same competences as professional translators, the market gets flooded with a new wave of lower quality translations. With the rise of NMT and LLM tools, these lower quality translations are fed back into the systems and used further for new translations. The machine eventually starts imitating itself, further erasing the voice of human authors and translators (Rothwell ym. 2023, 106).

As mentioned before, the English language and Western cultures hold a dominant position in the translation field. This is also true within machine translations, which creates another issue that affects the translations both ethically and in terms of quality. Since the machine translation tools tend to have more training data in English and based on Western cultures, they make translation choices that lean towards the traditions and linguistic styles of them, even at the expense of the languages and cultures in question (Olojo et al. 2025). This is especially harmful in culturally rich literary text, where a lot of the translation happens "in between the lines" (Taivalkoski-Shilov 2019)

## Method and Data

The data for this thesis was collected from the three first chapters of *Myötäjäiset*, a raw machine translation of the chapters, and a finalized, post-edited translation of the chapters. The machine translation was created using a DeepL machine translation tool that was integrated into Wordfast translation memory tool. The English variant used for the translation was Indian English. The finalized translation was post-edited by Hokkanen and approved by Siru Aura. The final translation includes some additions from the author herself, like an addition to the 1961 Dowry Prohibition Act.

The segments picked for analysis all fall under the definition of cultural reference as it is defined in this thesis. The cultures referenced are Indian and Finnish, although a few examples were picked for analysis where the machine added a third culture outside of the Indian and Finnish cultural environment of the novel. The original novel is written in Finnish, but some terms and phrases from Indian languages appear. The novel includes terms and phrases that originate from Tamil, Hindi, and Kannada. These are Indian languages spoken by characters of the book.

Since the original novel already has a multilingual aspect to it, Aura has had to make some translatory decisions while writing it. For example, she has at times chosen to include an explanatory phrase to accompany some of the terms originating from one of the Indian languages that appear in the story. These decisions were also taken into account when collecting data for the analysis of this thesis.

This thesis attempts to answer two research questions. How does the machine translation tool translate cultural references in a literary text? And does the English language and Western culture dominance appear in the machine translation? The goal of the first question is to identify what kind of translation decision the machine makes and whether the cultural meaning of the original text translates into the translation. The second question studies if the machine makes solutions that highlight the English language and Western culture dominance. The overall goal of the thesis is to study whether the machine translation tool creates publishable literary text independently.

The examples for analysis were collected in two phases. I first read the translations side by side and marked down all possible cultural references that could be used for analysis. I then went over the marked segments and collected the relevant ones into an Excel sheet. Originally, the

third culture references were dropped out of the analysis at this point, but they were later brought back for analysis. Once the cultural references were collected into the Excel sheet, they were divided into several groups based on style, translatory decisions, and other factors.

While there were various groups at first, they were then narrowed down to five for the final analysis. The five groups are: 1. Cultural references with proper names, 2. Implicit cultural references, 3. Realia, 4. Explanatory translations, and 5. Cultural references where influences of American culture are evident. The first group involves cultural references where a proper name is presented. The second group involves cultural references, that do not have specific realia or terminology to indicate that it is a cultural reference, but that involves cultural meaning that is evident "in between" the linguistic markers of the phrase/sentence. The third group discusses cases where the reference includes specific cultural realia and how they are treated in both the machine translation and the finalized translation. The first three groups are analyzed from the perspective of how the machine translated cases where the mentioned cultural references were involved. The fourth category discusses the explanatory translation strategy more broadly, how it was used in the original, how the machine translated these cases, and how the human translator utilized the strategy in the finalized translation. The fifth group focuses more on the English language/ Western culture dominance and discusses cases that could be viewed as evidence of it in the text.

## Analysis

The analysis of the first three groups of cultural references revealed that the machine translation tool struggled with context while translating. Even though the context of the specific cultural references was evident either within the sentence where the cultural reference appeared, or in the text right before the cultural reference, the machine translation tool made translation choices that suggest that it did not understand the context and therefore the meaning of the original text became unclear. These choices included, for example, instances such as choosing a masculine pronoun for a female character and mistaking a person's name for a city. The machine also did not remain consistent in its own translation choices but instead translated the same term in different ways throughout the text. These instances included for example the term *myötäjäiset*, 'dowry', which the machine translated as such a few times, but once turned into a phrase *Forever and ever* (example 10). This could indicate that the machine attempted calquing and understood the term to mean the traditional Finnish wedding vow *myötä-ja vastoinkäymisissä*,

‘in sickness and in health’. However, this is problematic since this erases the true meaning of the original text, where a person’s dowry was discussed.

The English language or Western culture dominance can be seen as a reason for some of the translation issues in the first three groups. The machine did not recognize an Indian woman’s name and also translated a Finnish man’s name, *Markus* into ‘Mark’. This could indicate an evangelical reading of the name, since *Mark* is typically translated into ‘Markus’ in the Finnish translation of the Bible. The machine translation tool also mistranslated terms for Indian currency in examples 11 and 12, which could be seen as evidence toward the Western culture dominance.

The implicit cultural references appeared to have been easier for the machine translation tool to translate, especially when they did not include any culture specific terms or realia. It may be that it was easier for the machine to calculate the correct translation for the references when it did not have to consider the context as much. However, some implicit references challenged the machine, when they carried heavy cultural meaning “in between the lines”. For example, the machine used the term *organize* instead of *arrange* when discussing the Indian tradition of arranged marriages. While *organize* is an acceptable translation for the Finnish term ‘järjestää’ in a lot of cases, it does not carry the same cultural meaning in the context of the original text. Therefore, the cultural meaning does not transfer entirely in the raw machine translation. The human translator changed the term in the post-edited translation to ensure that the cultural meaning is transferred to the reader.

The explanatory translation strategy was already utilized in the source text and it was also utilized by the translator in the post-editing phase of the translation process. The machine translation tool did not add any explanatory translations independently but it did translate some of the existing explanations from the source text. In example 13, Aura has included the explanatory term *ceremony* in her original text to accompany the phrase *mangala snanam*, so the reader will understand the context of the phrase. The machine similarly used the explanatory term, as did the translator in the post-edited translation.

While the machine translation tool did not independently add explanatory translations to the raw machine translation, the human translator did add some to the finalized translation. Since the original text was written in Finnish, most of the Finnish cultural references did not include explanatory phrases within the original text. The translator did add explanations to some of

these in the finalized translation. In examples 15 and 16, the translator added explanations to ensure that the reader of the translation will understand the Finnish terms referencing the Finnish higher education system. The machine translation simply translated these without adding the explanations.

The original text also included some English phrases. In most cases, these phrases were simply left as is in the raw machine translation. In example 17, the English phrase was accompanied by a Finnish explanatory translation in the original text. While the machine repeated the same phrase twice, the human translator deleted the repetition, as the explanation is not necessary in the English translation.

While the explanatory translation strategy was utilized multiple times both by the original author and the translator, there were cases when realia from one of the Indian languages were not accompanied by a specific explanation. In these cases, the meaning of the realia was typically determinable from the surrounding context, and no explanation was added by the machine translation tool or the human translator.

Although some of the earlier examples from the first four groups work as evidence of the dominance of English language and Western culture, there were a few cases when the machine specifically added a third culture meaning to the machine translation. These were typically cases of adding cultural references to North America. For example, a *thank you party* thrown by the main character for her friends was translated into 'Thanksgiving', the North American holiday, which is not celebrated in India. Examples such as this would suggest a significant Western culture dominance in the machine translation.

## Conclusion

From the analysis of the data, we can conclude that the machine translation tool does not create consistent translations. The machine uses different translations for the same term throughout the text. Simultaneously, it does not read the context of the words, sometimes even if the context appears within the same sentence. This becomes problematic not only because of the lack of consistency but also because at times the machine chose translations that also alter the cultural meaning of the original text. It is also evident from the translations of implicit cultural references that the machine cannot translate the cultural meaning of the text when it appears "in between the lines" and not explicitly in the text.

Simultaneously, the machine does not use additions or removals to the translation, even if they are necessary or would make the text more pleasing to read. Most of the time the machine included everything from the original text into the translation and the human translator made the necessary additions or removals to the translation in the post-editing phase of the translation process to ensure that the translation is both understandable and readable.

Based on the data, the machine translation tool's primary translation strategy seems to be calquing. There appeared several instances where realia without a specific English translation appeared in the original text. Instead of using the realia, the machine attempted a calquing translation. Unfortunately, this often resulted in a translation that altered or completely erased the cultural meaning of the original text. This is problematic, especially in a culturally rich literary text such as the novel that this thesis analyzes.

The machine translation tool struggled with other cultural references too, especially with references specific to Indian culture or languages. This suggests a significant Western culture dominance in the training data of the machine translation tool. Some of the translation decisions made by the machine specifically suggest an American culture dominance in the machine translation. While most of these cases such as switching measuring units from centimeters to feet are not strong enough evidence of the dominance on their own, when added along to the other instances found in the data, it can be concluded that the dominance is significant.

The data also showed that the machine translation tool showcased a significant male preference. In almost all cases appearing in the data of this thesis, the Finnish gender-neutral pronoun *hän* was translated into the English masculine pronoun 'he'. This happened even in cases where the gender of the person in question was evident within the same sentence. For example, the machine used the masculine *he* to refer to a person, who was earlier in the sentence referred to as *mother*. The masculine gender preference in the translation is harmful, because the main theme of the original novel is women's rights and position in India. The masculine preference therefore alters the cultural meaning of the original text, as well as proves that the machine does not understand the context of the entire text it is translating.

The cultural meaning of the original text was often diminished or even completely eliminated in the raw machine translation. This is especially problematic when translating culturally rich text such as *Myötäjäiset*, and literary texts in general. This problem could be corrected by creating a more diverse training data for the machine translation tool. Even in cases where the

machine translation tool did transfer the cultural meaning of the original text into the translation, the human translator did make significant linguistic and grammar corrections to the translation. Therefore, it can be concluded that the machine translation tool does not create publishable quality translations of literary texts independently.

While this thesis provides a general example of the status of machine translation tools within the literary field, it is only an example based on a singular novel. For a more comprehensive study of the subject, it would be necessary to use multiple novels or other literary texts in the study. In order to gain a complete understanding of the overall quality of the machine translations, it would also be beneficial to look at other translation decisions made by the machine and not only the decisions when translating cultural references.

This study was conducted based on one machine translation by one NMT system. To assess the general quality of NMT systems, a study would need to be conducted using raw machine translations of the text from multiple NMT tools. However, as the LLM tools are becoming the current norm of the field, it would also be beneficial to study the translation decisions and capabilities of LLM tools rather than NMT tools.