



HELSINGIN YLIOPISTO

# **Konekääntimen ja generatiivisen tekoälyn vaatima jälkieditointityö urheilusääntöjen kääntämisessä**

Tapaustutkimus ultimaten sääntöjen kääntämisestä

Kääntämisen ja tulkkauksen maisteriohjelma, englant-suomi  
Maisterintutkielma

Laatija:  
Vilma Heinonen

Ohjaaja:  
FT Marjut Alho

27.4.2024  
Helsinki

**Tiedekunta:** Humanistinen tiedekunta

**Koulutusohjelma:** Kääntämisen ja tulkkauksen maisteriohjelma

**Opintosuunta:** Käännösteknologia, englanti–suomi

**Tekijä:** Vilma Heinonen

**Työn nimi:** Konekääntimen ja generatiivisen tekoälyn vaatima jälkieditointityö urheilusääntöjen kääntämisessä – tapaustutkimus ultimateen sääntöjen kääntämisestä.

**Työn laji:** Maisterintutkielma

**Kuukausi ja vuosi:** Huhtikuu 2024

**Sivumäärä:** 64 sivua, 15 liitesivua

**Avainsanat:** konekääntäminen, urheilusääntöjen kääntäminen, AI-kääntäminen, ChatGPT, tapaustutkimus

**Ohjaaja tai ohjaajat:** Marjut Alho

**Säilytyspaikka:** Helda – Helsingin yliopiston avoin julkaisuarkisto

**Muita tietoja:**

**Tiivistelmä:** Maisterintutkielmassa käsitellään urheilusääntöjen kääntämistä konekääntimen ja generatiivisen tekoälyn avulla. Kyseessä on tapaustutkimus, jossa tutkitaan yhden urheilulajin, ultimateen sääntöjen kääntämistä. Tutkimuksessa pyritään selvittämään, kuinka hyvään käännökseen konekääntäminen ja generatiivinen tekoäly pystyvät tässä tekstilajissa, vaativatko niiden raakakäännökset paljon jälkieditointia, ja mitkä ovat generatiivisen tekoälyn ja konekääntimen keskeisimmät, laatuun vaikuttavat erot.

Tutkimuksen aineisto koostuu näiden käännösten laadunarvioinneista, eli kymmenen vastaajan jälkieditoimista käännöksistä. Vastaajia tähän tutkimukseen on kerätty sosiaalisen median kanavien kautta lajin harrastajista ja kielen ja kääntämisen opiskelijoista. Aineistoa analysoidaan virheanalyysimenetelmän avulla ja tutkimuskysymyksiin vastataan tämän virheanalyysin pohjalta.

Virheanalyysin ja kielenulkoisten tekijöiden perusteella voidaan todeta, että tämän tutkimuksen tekstilajin kääntämisessä generatiivinen tekoäly onnistuu paremmin kuin konekääntäminen, vaikka molemmat raakakäännökset ovat toistaiseksi vielä kaukana luettavasta ja konkreettisesti hyödynnettävästä sääntöversiosta ja vaativat paljon jälkieditointia. Näiden kahden merkittävimpiä eroja, jotka eivät vaikuta käännösten laatuun, ovat generatiivisen tekoälyn käytettävyys ja saavutettavuus. Tässä tutkimuksessa hyödynnettävä ChatGPT3.5 on täysin ilmainen kaikille käyttäjille ja on siksi saavutettavampi pienille lajiorganisaatioille, kuin lisenssin ja maksumuurin takana oleva Phrasen konekääntäminen. Generatiivinen tekoäly on myös käytettävämpi, sillä se pystyy yhdellä promptilla hyödyntämään sille syötettyä termistöä käännöksessä, ilman jälkieditointia. Generatiivisen tekoälyn treenidata on paljon laajempi kuin yksittäisen konekääntimen ja tekoälyä kehitetään jatkuvasti, joten sujuvia käännöksiä urheilusääntöille voidaan odottaa tulevaisuudessa.

# Sisällysluettelo

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>JOHDANTO</b>                                      | <b>5</b>  |
| <b>2</b> | <b>TEOREETTINEN VIITEKEHYS</b>                       | <b>9</b>  |
| 2.1      | KONEKÄÄNTIMET JA NEUROVERKKOKÄÄNTIMET                | 9         |
| 2.2      | KONEKÄÄNNIN JA JÄLKIEDITOINTI                        | 11        |
| 2.3      | GENERATIIVINEN TEKOÄLY                               | 17        |
| 2.4      | URHEILU KÄÄNTÄMISEN ERIKOISALANA                     | 20        |
| <b>3</b> | <b>TUTKIMUSMENETELMÄ JA AINEISTO</b>                 | <b>24</b> |
| 3.1      | TUTKIMUSMENETELMÄ                                    | 24        |
| 3.2      | TYÖVAIHEET   | 29        |
| 3.3      | AINEISTO   | 30        |
| <b>4</b> | <b>ANALYYSI</b>                                      | <b>32</b> |
| 4.1      | TULOKSET   | 32        |
| 4.1.1    | KÄSITTEITÄ KOSKEVAT VIRHEET                          | 40        |
| 4.1.2    | TERMISTÖÄ KOSKEVAT VIRHEET                           | 42        |
| 4.1.3    | KIELIOPPIA KOSKEVAT VIRHEET                          | 47        |
| 4.2      | GENERATIIVISEN TEKOÄLYN JA KONEKÄÄNTIMEN EROISTA     | 49        |
| <b>5</b> | <b>POHDINTA</b>                                      | <b>52</b> |
| <b>6</b> | <b>JOHTOPÄÄTÖKSET</b>                                | <b>57</b> |
|          | <b>Lähteet</b>                                       | <b>61</b> |
|          | Aineisto   | 61        |
|          | Kirjallisuus   | 61        |
|          | <b>Liitteet</b>                                      | <b>65</b> |
|          | Liite 1. Lähtöteksti: "Marking infractions"          | 65        |
|          | Liite 2. Phrasen konekäännös                         | 67        |
|          | Liite 3. ChatGPT3.5:n käännös                        | 69        |
|          | Liite 4. Termistö                                    | 71        |
|          | Liite 5. Tieteellisen tutkimuksen tietosuojailmoitus | 73        |



# 1 JOHDANTO

Tässä maisterintutkielmassa tutkitaan urheilusääntöjen, ja tarkemmin ultimate sääntöjen kääntämistä konekääntimellä ja generatiivisella tekoälyllä. Tässä tutkielmassa pyritään selvittämään, onko urheilusääntöjen kääntäminen näiden käännösteknologisten työkalujen ja jälkieditoinnin avulla järkevää.

Tutkimusmenetelmänä tässä tutkimuksessa on virheanalyysi, ja sen avulla pyritään myös tutkimaan konekääntimen ja generatiivisen tekoälyn tyypillisesti tekemiä virheitä urheilusääntöjen kääntämisessä.

Ultimate tai liitokiekko on Suomessa hyvin marginaalinen urheilulaji, jolla on tällä hetkellä aktiivisia kansallisen lajiliiton jäseniä noin 450 (vuosina 2022-2023) (SLKL, 2023, s. 2). Ultimate on frisbeellä pelattava joukkuelaji, jota pelataan Suomessa talvisin sisätiloissa ja kesäisin ulkona. Pelikenttä on suorakulmanmuotoinen, noin jalkapallokentän mittainen alue, jonka kummassakin päässä on maalialueet. Maaleja tehdään heittämällä frisbeetä joukkuekaverilta toiselle, lopulta saavuttaen pelaajan, joka on sillä hetkellä maalialueella ja saa frisbeen kiinni. Puolustajat voivat katkaista syötön vain frisbeen ollessa ilmassa, mutta kädestä sitä ei saa ottaa, eikä toiseen pelaajaan saa koskea frisbeen haltuun ottamiseksi.

Ultimate vaatii pelaajaltaan kypsyyttä ja hyvää pelihenkeä, sillä se on täysin kontaktiton laji, jossa ei ole tuomareita. Tuomarittomuus johtaa siihen, että pelaajien tulee itse tulkita ja tuomita kyseenalaiset pelitilanteet. Tulkinta ja tuomitseminen vaativat pelaajilta vahvaa sääntöjen osaamista ja näiden osaaminen tulee jokaisen kilpapelajaajan todistaa lähes vuosittain läpäisemällä sääntötestin. Sääntötestit ovat kuitenkin toistaiseksi sääntöjen ohella olemassa vain englannin, ja muiden suurten lajimaiden, kuten ranskan ja japanin kielellä. Saavutettavuuden sekä sääntötietoisuuden lisäämiseksi kansallinen lajiliittomme on ottanut askeleita sääntötestin suomentamiseksi, mutta saavutettavuuden lisääminen vaatii myös sen, että ajankohtaiset säännöt olisivat olemassa suomen kielellä.

Aihe on erityisen lähellä sydäntäni oman lajitaustani vuoksi, mutta myös siksi, että haluan lisätä lajin saavutettavuutta ja auttaa lajinedistämistyössä omalla työpanoksellani. Siksi en ainoastaan tässä tutkimuksessa pyri selvittämään, onko

tekoälystä tai konekääntimestä kääntämään urheilulajien sääntöjä, vaan pyrin myös tuottamaan luettavan ja lähestyttävän version säännöistä.

Tämän tutkimuksen tarkoituksena on selvittää ja samalla tutkimuskysymyksiä toimivat:

1. Kuinka hyvään käännökseen konekäännin ja generatiivinen tekoäly valmiiksi luodun termipankin avulla pystyvät urheilusääntöjen kääntämisessä?
2. Kuinka paljon jälkieditointia konekäännös ja generatiivisen tekoälyn käännös vaativat, kun käännetään urheilusääntöjä?
3. Mitä ovat ChatGPT3.5:n ja Phrasen konekääntimen keskeisimmät käännöksen laatuun vaikuttavat erot?

ChatGPT on hyvin ajankohtainen ja keskustelua herättävä aihe kääntämisen alalla nykypäivänä ja sen hyödyntämistä käännösprosesseissa ei ole vielä tutkielman julkaisuhetkellä juuri tutkittu. Siksi olen päättänyt sisällyttää sen osaksi tutkimusta ja vertailla sen käännöstä konekääntimen tuotoksen kanssa.

Urheilusääntöjen kääntämistä ei vielä toistaiseksi ole merkittävässä määrin tutkittu, vaikka kuitenkin puhutaan sellaisesta tekstilajista, jota on saavutettavuuden ja kansainvälistymisen nimissä julkaistava monilla eri kielillä. Joitakin pienlajeja, kuten pesäpalloa tai hurlingia, lukuun ottamatta monia urheilulajeja harrastetaan ja pelataan ympäri maailman, joten on melko erikoista, että tätä tekstilajia tai sen kääntämistä ei vielä ole muutamaa tutkimusta lukuun ottamatta tutkittu. Tämä toisaalta osoittaa tämän tutkimuksen merkityksellisyyden kääntämisen tutkimuksen kannalta.

Tämä tutkimus on tapaustutkimus, jonka aineistona käytetään yhden lajin englanninkielisiä sääntöjä ja termistöä. Tässä tutkimuksessa ensin käännetään termipankin, konekääntimen ja tekoälyn avulla lajin säännöt englannista suomeen. Tämän jälkeen raakaversiot luetutetaan yleisöllä, joka koostuu lajin sekä kääntämisen asiantuntijoista, joilta kerätään palautetta käännösten kielestä ja sisällöstä. Tämän palautteen virheanalyysin perusteella vastaan tutkimuskysymyksiin.

Tutkielman seuraavassa luvussa esitellään tutkimuksen teoreettinen viitekehys, jossa käsitellään tutkimusta koskevia käännöstieteen tutkimussuuntauksia ja aiempaa tutkimusta aiheen osalta. Koska tutkimukseen sisältyy paljon uudenlaista tutkimustietoa ja monia tutkittavia kääntämisen osa-alueita, aloitetaan teoreettisessa viitekehyksessä laajemmin tutkituista, melko tutuista kääntämisen tutkimussuuntauksista ja siirrytään hiljalleen uudempiin, vähemmän tutkittuihin osa-alueisiin.

Teoreettisen viitekehysten jälkeen tutkielmassa esitellään tutkimusmenetelmä ja aineisto. Tämän tutkimuksen toteuttamiseen ja juuri sopivan tutkimusmenetelmän löytämiseen on ollut tarve soveltaa erilaisia tutkimus- ja aineistokeruumenetelmiä, ja tässä luvussa esitellään sovelletut menetelmät. Tutkimusmenetelmää ja aineistoa koskevassa luvussa sivutaan myös menetelmiä, joita olisi voitu tässä tutkimuksessa hyödyntää, mutta ovat esimerkiksi tutkimuksen rajauksen tai tutkimuskysymysten valinnan takia rajautuneet pois. Tämän luvun lopuksi esitellään myös lyhyesti tutkimusta varten kerätty aineisto.

Tutkimusmenetelmän ja aineiston kuvauksen jälkeen siirrytään aineiston analyysiin. Analyysiluvun aluksi esitellään virheanalyysin tulokset, jotka on saatu aineistona toimivista vastauksista. Analyysissä esitellään teoreettisen viitekehysten tapaan alkuun yleisluontoisemmat löydökset aineistosta ja siirrytään loppua kohden tarkemmin rajattuihin löydöksiin. Analyysiluvun lopussa käsitellään myös konekääntimen ja generatiivisen tekoälyn keskeisimpiä laatuun vaikuttavia eroja ja vastataan myös muihin tutkimuskysymyksiin.

Pohdintaluvussa esitetään kritiikkiä tutkimusta kohtaan ja pohditaan tutkimuksen merkitystä laajemmin käännösalan näkökulmasta. Tutkimuksen tuloksia vedetään yhteen lähdekirjallisuuden tutkimustulosten kanssa ja esitellään ristiriitaisuudet tämän tutkimuksen ja aiemman tutkimuksen välillä.

Lopuksi esitellään tutkimuksesta vedettävät johtopäätökset ja ideoidaan tutkimustulosten perusteella lisää aiheeseen liittyvää tutkimusta ja uusia

tutkimusnäkökulmia. Johtopäätöksissä käydään tutkielma myös vaihe vaiheelta läpi ja lopuksi todetaan, miksi tämä tutkimus on tärkeä käänntieteen tutkimusalalla.

## 2 TEOREETTINEN VIITEKEHYS

Tässä luvussa käyn läpi tutkielman lähdekirjallisuutta eri osa-alueittain. Koska tutkielma käsittelee aiemmin sellaisenaan tutkimatonta kääntämisen osa-aluetta, koen tarpeelliseksi sisällyttää monia eri näkökulmia tutkittavaan aiheeseen.

Teoreettista viitekehystä tarkastellaan ensin tutummista näkökulmista, kuten konekääntämistä, ja siirrytään loppua kohden vähemmän tutkittuihin ja uudempiin kääntämisen aluevaltauksiin, kuten urheilun erikoisalakääntämiseen ja uusien käännoistyökalujen tutkimukseen.

### 2.1 KONEKÄÄNTIMET JA NEUROVERKKOKÄÄNTIMET

Koska tässä tutkielmassa arvioidaan neuroverkkokääntimen käännostä, on tärkeää avata konekääntimien ja etenkin neuroverkkokääntimien toimintatapaa ja syitä niiden käyttöön. Tässä luvussa pyritään hieman selittämään neuroverkkokääntimen toimintatapaa, mutta koska sen perinpohjainen ymmärtäminen ei ole tutkielman aiheen ymmärtämisen kannalta relevanttia, ei siihen tässä tutkielmassa perehdytä tarkemmin. Lukija voi itsenäisesti perehtyä aiheeseen syvemmin lähdekirjallisuuden avulla.

Konekääntimet ovat työkaluja kääntämiseen, joilla pyritään luomaan käännoksiä ilman ihmiskääntäjän työpanosta. Konekääntimille on kääntämisen mahdollistamiseksi syötetty monikielistä treenidataa, jota hyödyntäen ne pyrkivät luomaan mahdollisimman hyvän käännoksen. Konekääntimen tarkempi toimintatapa riippuu konekääntimen tyypistä, joita ovat tilasto-, sääntö- ja neuroverkkopohjainen konekääntäminen. Konekääntimet ovat alun perin luotu käänno prosessien nopeuttamiseen ja halventamiseen, ja jotta tavallisilla ihmisillä olisi mahdollisuus kääntää tietoa vieraalta kieleltä tuntemalleen kielelle (Koehn, 2020a, s. 19–20). Koehn toteaaakin, että tutkimusten mukaan konekääntimet ovat tähän mennessä nostaneet kääntämisen tehokkuutta 42–131 prosentilla (mts. 22, 23), mutta eivät silti yksinään pärjää ihmiskääntäjän aikaansaannoksille ja vaativat jälkieditointia (mts. 21). Konekääntimet eivät myöskään kykene nopeuttamaan käänno prosesseja kieleltään hyvin rikkaiden tekstien, kuten mainostekstien tai lauseiden, kaunokirjallisuuden tai runouden kääntämisessä. Koehnin mukaan konekääntimet tulisi myös optimoida tietyn erikoisalan tekstien kääntämiseen, jotta

ne toimisivat ihanteellisesti ja tuottaisivat parhaan mahdollisen käännöksen (mts. 22).

Konekääntimet ovat olleet olemassa jo ensimmäisten tietokoneiden synnystä, eli noin 1950-luvulta lähtien (Koehn, 2020b, s. 30, 33), mutta niiden kehitys ja käyttö ovat kokeneet monenlaisia vaiheita elinikänsä aikana (mts. 29). Alkutaipaleellaan konekääntimien kehitys ja tutkimus sai paljon rahoitusta sodan takia (mts. 34), jolloin kävi kuitenkin ilmi, että se ei ollut juuri halvempi tai tehokkaampi tapa tuottaa käännöksiä kuin ihmiskääntäjä ja tutkimuksen rahoitus väheni merkittävästi. Rahoituksen vähenemisestä huolimatta konekääntimien kehitys jatkui ja 2000-luvulla alettiin tutkia neuroverkkojen hyödyntämistä myös kääntämisessä erilaisin metodein, kuten *back-translation* ja *byte pair encoding* (mts. 39–40) ja pian tutkittiin ainoastaan enää neuroverkkokääntimiä tilasto- tai sääntöpohjaisten kääntimien sijaan.

Neuroverkkokääntimet toimivat Koehnin mukaan hyvin samalla tavalla kuin monet muutkin applikaatiot, jotka hyödyntävät koneoppimista treenimetodina (2020c, s. 67). Hyvin yksinkertaistettuna, neuroverkkokääntimet nojaavat lineaarimalliin, jossa lähdekielen eri sanat ovat arvotettuja ja lähdekieltä ja kohdekieltä yhdistävät nuolet ovat myös arvotettuja, jotta malli osaisi yhdistää lähdekielen sanan oikeaan kohdekielen sanaan (mts. 67–68). Lineaarimallin lisäksi neuroverkkokääntimissä on monia kerroksia, mukaan lukien piilotettuja kerroksia, joiden mekanismeja voidaan vain spekuloida syötetyn tiedon ja ulostulon perusteella (mts. 68). Tämä on vain hyvin yksinkertaistettu selitys neuroverkkokääntimien toiminnasta, ja todellisuudessa niiden kouluttamiseen ja toimintaan vaaditaan monia muitakin metodeja ja askeleita (ks. mts. 79–86).

Maučec ja Donaj (2019) listaavat artikkelissaan erilaisia tapoja tutkia konekäännösten tai konekääntimien laatua. He listaavat artikkelissaan erilaisiksi konekäännintyypeiksi sääntö-, esimerkki- ja tilastopohjaisen konekääntimen, ja myös tämän tutkimuksen kannalta olennaisen, neuroverkkokääntimen. Vaikka konekääntimet ovat viime vuosina kehittyneet hyvin nopeasti, on niissä Maučecin ja Donajin mukaan myös monenlaisia ongelmia (mts. 6–7), kuten se, että ne eivät toimi ihanteellisesti kaikissa kielipareissa tai edes kaikilla kielillä, etenkin jos puhutaan

morfologisesti rikkaista kielistä, kuten suomen kielestä. Kielet ovat myös rakenteiltaan hyvin erilaisia, ja näihin on Maučecin ja Donajin mukaan (mts. 6) pyritty hakemaan ratkaisua esimerkiksi sanan osien segmentoinnista (engl. *sub-word segmentation*), jossa sana jaetaan esimerkiksi morfeemeihin tai muihin sanaa pienempiin merkityksellisiin osiin, jotta merkitys välittyisi käännösprosessissa kohdekieleen (Song ym. 2023). Ongelmana on sanan osien segmentoinnissa kuitenkin se, jos käännetään vähemmän flekteeraavasta kielestä enemmän flekteeraavaan, jolloin kohdetekstiin täytyisi luoda merkityksiä ikään kuin tyhjästä (Maučec ja Donaj, 2019, s. 6).

Maučecin ja Donajin mukaan konekääntimen laadun arviointi voidaan jakaa kahteen kategoriaan: tietokoneen suorittamaan automatisoituun arviointiin ja ihmisen tekemään manuaaliseen arviointiin (2019, s. 7). Ihmisen tekemässä manuaalisessa arvioinnissa korostuu jälkieditoinnillekin tyypilliset keinot, kuten kohdetekstin vertaaminen lähdekieleen, kontekstin, tekstin sujuvuuden ja kieliopillisuuden arviointi (mts. 8), jotka tulevat tutummiksi seuraavassa alaluvussa.

## 2.2 KONEKÄÄNNIN JA JÄLKIEDITOINTI

Tässä tutkielmassa tullaan hyödyntämään jälkieditointia konekääntimen ja generatiivisen tekoälyn tuottamien käännösten arvioinnissa. Seuraavaksi selitän lähdekirjallisuuden avulla, mitä on konekäännöksen jälkieditointi ja miksi sitä hyödynnetään konekääntimen laadun tutkimuksissa.

Carl ym. (2011) ovat tutkineet kääntäjien tehokkuutta kääntää tietokoneavusteisesti verrattuna perinteiseen, kynä-paperi-menetelmän tehokkuuteen.

Tietokoneavusteisesta kääntämisestä puhutaan, kun hyödynnetään mitä tahansa tietokoneella käytettävää käännös menetelmää ihan tekstinmuokkausohjelmista lähtien jopa käännösmuisteihin ja projektinhallintatyökaluihin, mutta nämä ovat ainoastaan ihmiskääntäjän apuvälineinä käännösprosessissa, eikä pyritä täysin koneistettuun ja automatisoituun käännösprosessiin. Carlin ym. (2011) tutkimuksessa on tutkittu kolmea englanninkielistä uutistekstiä, jotka ovat käännetty tilastopohjaisen Google Translate -konekääntimen ja jälkieditoinnin avulla tanskan kieleen ja näitä verrattiin ihmisen perinteisellä tavalla kääntämiin teksteihin.

Tehokkuuden mittareina tutkimuksessa toimivat laadun mittaaminen, kääntämiseen kulunut aika, näppäimistön ja katseen seuranta (mts. 132).

Laatua mitattiin tutkimuksessa vertailemalla konekäännettyä ja jälkieditoitua virkettä ja manuaalisesti käännettyä ihmisen tuottamia virkevaihtoehtoja. Carl ym. tutkimuksen mukaan jälkieditoidut virkkeet saivat kokonaistuloksissa paremman tuloksen kuin manuaalisesti tuotetut virkkeet, vaikkakin pienellä marginaalilla, kun kaikki tehokkuuden mittarit otettiin huomioon (mts. 136).

Aikaa vertailtiin jälkieditoinnin keskimääräisen ajan ja manuaalisen käännöksen tuottamisen ajan välillä. Manuaalinen käännös vei vain noin puoli minuuttia enemmän aikaa, mutta Carlin ym. mukaan tuloksessa on hyvä ottaa huomioon se, että manuaalisia käännöksiä tutkimukseen tuottaneet henkilöt olivat pääosin kokeneita kääntäjiä, kun taas jälkieditoijista harvalla oli ennestään kokemusta jälkieditoinnista tai jälkieditointityökalujen käytöstä, mikä saattaa osaltaan vaikuttaa jälkieditointiin kuluvaan aikaan (mts. 137–138).

Katseen seurannassa ei Carlin ym. mukaan ollut ajallisesti merkittäviä eroja jälkieditoinnin ja manuaalisen kääntämisen välillä (mts. 138), mutta eroa syntyi enemmän katseenkiinnittymispisteiden suhteen. Kun tekstiä käännettiin manuaalisesti, Carlin ym. mukaan kääntäjän katse kiinnittyi pidemmiksi ajoiksi näytön ulkopuolelle, eli hyvin todennäköisesti näppäimistöön tai ”tyhjyyteen” kääntäjän etsiessään mielestään sopivaa käännösratkaisua (mts. 138).

Manuaalisesti käännettäessä Carlin ym. mukaan katse kiinnittyi tasaisesti lähtö- ja kohdetekstin välillä, kun taas jälkieditoidessa tekstiä kääntäjän katse kiinnittyi suurimman osan ajasta kohdetekstiin (mts. 140). Tämän perusteella Carl ym. tekee johtopäätöksen siitä, että manuaalisesti kääntävällä henkilöllä on todennäköisesti parempi käsitys lähtötekstin sisällöstä kuin jälkieditoijalla, ja että jälkieditoija tekee työtä kohdeteksti edellä, lukien sen ensin ja palaa lähtötekstiin tarkistamaan kohdetekstin oikeellisuutta (mts. 140).

Carl ym. tulevat tutkimuksessaan johtopäätelmään, että vaikka jälkieditoinnilla ja manuaalisesti kääntämisellä ei ajallisesti ole merkittävää eroa, on jälkieditointi hieman tehokkaampi prosessi ainakin tämän tutkimuksen tulosten perusteella (mts.

141). Carl ym. tulevat myös siihen johtopäätelmään katseen seurannan avulla, että jälkieditointi ja manuaalisesti kääntäminen ovat kognitiivisesti hyvin erilaisia prosesseja (mts. 141). Carlin ym. tutkimus oli kuitenkin otokseltaan hyvin pieni ja luonteeltaan tapaustutkimus, joten suurempia yleistyksiä jälkieditoinnin ja konekääntämisen tehokkuuden osalta ei tämän tutkimuksen perusteella voida vetää.

Koponen ym. (2012) ovat tutkineet erilaisia konekääntimen tekemiä virheitä ja pyrkineet selvittämään, mikä on tehokkain mittari mittaamaan jälkieditointiin kuluva työmäärä. Koponen ym. ovat pyrkineet tutkimuksessaan ottamaan huomioon kaikki erilaiset jälkieditointiin vaikuttavat näkökulmat, kuten jälkieditoinnin määrän ajallisesti, teknisesti ja kognitiivisesti. Koponen ym. käyttivät kahdeksaa eri konekääntintä ja kahdeksaa eri kääntäjää selvittääkseen, mikä on tehokkain tapa arvioida jälkieditoinnin määrää käännoistyössä. Koponen ym. tarkkailivat kääntäjien tekemiä muutoksia konekäännöksiin ja tarkastelivat muutoksia erityisesti neljällä eri mittarilla:

**TIME** – yhden virkkeen muokkaamiseen kulunut aika

**SPW** – keskimäärin kuinka monta sekuntia kääntäjä kulutti yhden sanan kääntämiseen

**KEYS** – kuinka monta näppäimistön painallusta kääntäjä teki yhden virkkeen kääntämisen aikana

**HTER** – kuinka kaukana raakakäännös ja jälkieditoitu versio ovat toisistaan automatisoidusti laskettuna

Koponen ym. tulevat tutkimuksessaan siihen tulokseen, että HTER ei käytännössä kerro jälkieditoinnin määrästä paljoakaan, se pystyy lähinnä spekuloidaan tätä määrää tuomalla esiin tekstiin jälkieditoinnissa tehdyt muutokset. Koponen ym. tulevat myös siihen johtopäätelmään, että eniten aikaa vievät virheet johtuvat vääristä sanoista tai termeistä tekstissä, ja vähiten aikaa vievät virheet johtuvat yleensä oikeista sanavalinnoista, jotka ovat väärässä muodossa.

Koponen ja Salmi (2015) ovat tutkineet jälkieditoinnin avulla konekääntimen käännöksen oikeellisuutta. Tutkimuksessaan Koponen ja Salmi käyttivät tutkimusryhmää, joka koostui eri kielten kääntäjäopiskelijoista Turun Yliopistolla.

Tutkimusryhmällä he luetuttivat kaksi eri uutistekstipätkää, jotka olivat konekääntimen kääntämiä ja pituudeltaan täysin samanpituisia, mutta tutkimusryhmälle ei annettu lähdetekstiä ollenkaan. Tutkimuksen vastaajat saivat ohjeistukseen arvioida jokaista lausetta tekstissä sekä sisällön oikeellisuuden, että kielen oikeellisuuden perusteella. Arviointiasteikko oli seuraavanlainen:

“- correct meaning – correct language: the sentence correctly conveys the meaning of the source text and is grammatically and idiomatically correct

- correct meaning – incorrect language: the sentence correctly conveys the meaning of the source text despite errors in spelling, punctuation, grammar or word choice

- incorrect meaning – correct language: the sentence fails to convey source text meaning or conveys a different meaning but is grammatically and idiomatically correct

- incorrect meaning – incorrect language: the sentence fails to convey the source text meaning and contains errors in spelling, punctuation, grammar or word choice” (Koponen ja Salmi, 2015, s. 123–124)

Arviointiasteikko toimii siis hyvin mustavalkoisesti, joko lause on kieliopillisesti oikein tai väärin, ja sisällöllisesti oikein tai väärin, eikä pyri selvittämään tarkemmin virheiden syytä. Harkitsin tätä luokittelua tätä tutkielmaa varten, mutta päädyin kuitenkin tarkempaan virheluokitteluun saadakseni tarkempaa tietoa virheiden tyypeistä, jotta tulevissa tutkimuksissa esimerkiksi tekstilajien tai -tyyppien välinen vertailu olisi mahdollista.

Jälkieditoinnin näkökulmasta Koposen ja Salmen (2015) tutkimuksessa erityisen hankalaksi todettiin pitkien ja monimutkaisten lauseiden tai virkkeiden jälkieditointi. Jälkieditointia ja analyysiä vaikeuttaa etenkin virheiden moninkertaistuminen pitkissä virkkeissä, mutta myös yksittäinen pieni virhe saattaa muuttaa ja kadottaa koko virkkeen oikean merkityksen, varsinkin kun jälkieditointia tässä tutkimuksessa tehtiin ilman lähdetekstiä (mts. 133).

Tutkimuksessaan Koponen ja Salmi tulivat siihen tulokseen, että vain 29,5 prosenttia konekääntimen kääntämistä lauseista oli täysin oikein (mts. 131). Johtopäätöksissä todetaan myös, että on hyvin tärkeää, että konekäännöksen oikeellisuutta

tarkastellaan sekä kieliopin, että sisällön näkökulmista toisistaan eriteltyinä ja tarkemman virheluokittelun luominen on tulevaisuuden tutkimuksissa tärkeää (mts. 133), kuten Koponen (2010) on omallakin tutkimuksellaan osoittanut. Koposen ja Salmen (2015) tutkimuksen mukaan siis konekäännin vaatii vielä pitkälti ihmisen työpanosta saavuttaakseen luettavan muodon jopa uutistekstien kaltaisissa yleisissä tekstilajeissa.

Koponen (2016) on tehnyt kirjallisuuskatsauksen erilaisiin jälkieditointia käsitteleviin artikkeleihin selvittääkseen, onko jälkieditointi kannattavaa. Tehdäkseen kattavat johtopäätökset kysymykseen, Koponen tutki laajasti konekääntimen ja jälkieditoinnin käyttömahdollisuuksia, sekä jälkieditointiin kuluvaan työpanosta. Kuten Koposen ym. (2012) tutkimuksessa, tässäkin kirjallisuuskatsauksessa on otettu huomioon jälkieditoinnin kolme eri tutkittavaa näkökulmaa: ajallinen, tekninen ja kognitiivinen työpanos (Koponen, 2016, s. 9). Näiden lisäksi Koponen on ottanut kirjallisuuskatsauksessaan huomioon yksikielisen jälkieditoinnin, eli konekäännöksen jälkieditoinnin ilman lähtötekstiä (mts. 7–8), joka mukailee Koposen ja Salmen (2015) tekemää tutkimusta, sekä jälkieditoinnin tehokkuuden (2016, s. 5–6) ja laadun (2016, s. 6–7).

Kaikkien näiden näkökulmien tutkimisen jälkeen, Koponen tulee siihen johtopäätökseen, että konekääntäminen ja jälkieditointi ovat ehdottomasti kannattavia kääntämisen muotoja, kunhan niitä hyödynnetään sopivanlaisissa olosuhteissa (2016, s. 12). Sopivanlaiset olosuhteet Koposen mukaan riippuvat muun muassa tarpeeksi kapeasta käännettävästä erikoisalasta (mts. 2), kyseessä olevan konekääntimen laadusta, kyseisestä tekstilajista ja myös käännettävästä kieliparista (mts. 12). Tehokkuuden lisääminen riippuu myös ennen kaikkea kyseessä olevasta kääntäjästä, tämän tietoteknisistä taidoista ja työn kulusta.

Koposen (2016) mukaan jälkieditointi ja konekääntäminen ovat nosteessa ja ovat matkalla käänösalan keskiöön (mts. 12–13). Tutkimukset, joissa on todistettu jälkieditoinnin parantavan tehokkuutta, ovat varmasti osaltaan nopeuttaneet jälkieditoinnin ja konekääntämisen nostetta käänösalan käytännöissä ja ne ovat matkalla osaksi yhä useamman kääntäjän työn kulkua ja työprosesseja (mts. 13).

Olen tutkielmaa varten tutkinut useampaa Koposen tekemää tutkimusta ja hyödynnän hänen luomaansa virhekategoriointia tutkielman metodeissa, vaikka tutkielman viitekehys saattaa sen takia vaikuttaa yksipuoliselta. Olen kuitenkin Koposen tutkimusten tarkastelun jälkeen todennut, että ne tarjoavat monipuolisen katsauksen tarkasti rajattuun tutkimusalueeseen, joka koskettaa myös tätä tutkielmaa, ja ovat siksi jokainen yhtä tärkeitä tiedonlähteitä tutkielman aiheeseen.

Daems ym. (2017) ovat tutkineet jälkieditointia hyvin samanlaisella näkökulmalla kuin aiemmin tässä osiossa mainitut tutkimukset, mutta tässä tutkimuksessa jälkieditoinnin erilaiset mittarit kuten TIME, SPW, KEYS ja HTER (Koponen, 2012) on jaettu kahteen eri kategoriaan; käännöksen lopputuloksen arvioinnin mittareihin (engl. *product effort indicator*) ja prosessin työpanoksen mittareihin (engl. *process effort indicator*) (Daems ym., 2017, s. 2). Ensimmäisessä kategoriassa siis keskeisempää on konekäännöksen ja jälkieditoidun tai manuaalisesti tuotetun käännöksen etäisyyden vertailu esimerkiksi HTER:in avulla ja jälkimmäisessä jälkieditointiprosessiin kuluva työmäärän mittausta esimerkiksi aikaa mittaamalla tai silmänliiketutkimuksen keinoin (mts. 2). Daemsin ym. tutkimuksessa pyritään selvittämään, vaikuttaako konekäännöksen huono laatu eri tavalla näihin kahteen eri kategoriaan lisäten jommankumman työpanoksen määrää. Daems ym. pyrkivät myös selvittämään, että minkälaiset virheet vaikuttavat jälkieditoinnin määrään ja vaihtelevatko tulokset tutkimusryhmien (kääntäjäopiskelijat ja ammattikäntäjät) välillä (mts. 3).

Daemsin ym. (2017) tutkimuksen tulosten mukaan konekäännöksen huono laatu vaikuttaa kaikkiin muihin jälkieditoinnin työpanoksen mittareihin, paitsi fiksaatioon (engl. *fixation*) ja taukoihin (mts. 9). Daemsin ym. tulosten mukaan virheet myös vaihtelevat kategorian mukaan; tulokset osoittavat, että käännösten lopputuloksia vertaillaessa käy ilmi enemmän kieliopillisia virheitä, kuten kirjoitus- ja kongruenssivirheitä kuin jälkieditoinnin prosessin arvioinnin mittareista (mts. 12).

Daemsin ym. (2017) mukaan opiskelijat tarvitsevat enemmän taukoja kuin kääntämistä työkseen tekevät ammattikäntäjät (mts. 11) ja heidän fiksaationsa on pidempi, kun lauseen merkitys tai tooni on muuttunut tai vaatii muutosta (mts. 12).

Opiskelijat myös kiinnittivät raan konekäännöksen ja viimeistellyn lopputuloksen vertailussa enemmän huomiota kirjoitusvirheisiin kuin ammattikäntäjät (mts. 13–14). Tällainen vastaajien erityispiirteiden tarkastelu rajautuu kuitenkin tämän maisterintutkielman ulkopuolelle, mutta voisi olla mielenkiintoinen näkökulma tarkasteltavaksi erikoisalojen kääntämisen tutkimuksessa.

Daemsin ym. tutkimuksen myötä voidaan todeta, että konekääntimen ja jälkieditoinnin tehokkuus manuaalisesti kääntämiseen verrattuna riippuu aiemmin mainittujen tekijöiden lisäksi myös kääntäjän tai jälkieditoijan kokemuksesta. Joskus kääntäjä on sen verran kokenut, että tekstipätkän manuaalisesti kääntäminen vie vähemmän aikaa kuin konekäännöksen jälkieditointi (mts. 14). Daemsin ym. (2017) tutkimus, kuten Carlin ym. (2011), on otokseltaan hyvin pieni, joten suuria yleistyksiä tämän tutkimuksen perusteella ei ole tehtävissä, mutta nämä tutkimukset voivat toimia viitekehyksenä tulevalle tutkimukselle.

Jälkieditointi toimii siis myös hyvin metodina konekääntimen laadun tutkimuksessa sen lisäksi, että on näiden tutkimusten perusteella hyvin olennainen osa konekääntämistä ja sen tekemistä luettavaksi.

Jälkieditointia ja tietokoneavusteisten työkalujen hyödyntämistä kääntämisessä on tutkittu myös kääntäjän ergonomian näkökulmasta, kuten Alvarez-Vidal ym. (2020) ja Ehrensberger-Dow (2019), ja se on hyvin tärkeä tutkimussuunta näiden työkalujen käytön yleistyessä käännösosalalla, mutta rajautuu tämän tutkimuksen ulkopuolelle sen tutkimuskohteen vuoksi. Tässä tutkielmassa tutkitaan käännösten laatua, eikä esimerkiksi kääntäjiä tai heidän kokemuksiaan jälkieditoinnista, mikä olisi ollut edellytys ergonomian näkökulman tutkimiselle.

### **2.3 GENERATIIVINEN TEKOÄLY**

Generatiivinen tekoäly ja laajat kielimallit toimivat molemmat tekoälypohjaisesti, mutta niiden suorittamat toiminnot eroavat toisistaan jonkin verran. Lyhyesti kuvailtuna: generatiivinen tekoäly on suunniteltu luomaan sisältöä, oli se mitä tahansa kuvista ja videoista teksteihin, kun taas laajat kielimallit on rajoitettu toimimaan kielen ja tekstien parissa (Toner, 2023). Tässä tutkimuksessa käytetty ChatGPT3.5 on luokiteltu teknisten rajoitustensa takia laajaksi kielimalliksi, mutta

tässä tutkimuksessa sitä nimitetään generatiiviseksi tekoälyksi, sillä aiheeseen liittyvä kirjallisuus ei toistaiseksi ole täysin yhteneväinen termien suhteen. Sekä Brown ym. (2020), että Siu (2023) ovat tutkineet laajojen kielimallien (englanniksi *large language models*, LLM) käyttöä kääntämisen apuna. Siu (2023) esittelee hiljattain julkaisemassaan artikkelissa laajojen kielimallien, kuten ChatGPT:n ja ChatGPT-3:n, toimintaperiaatetta, kuinka niitä opetetaan ja kuinka ne suoriutuvat käännöstehtävistä konekääntimiin verrattuina.

Brown ym. (2020) testaa ChatGPT-3:a kolmella eri tapaa, *zero-shot learning*, *one-shot learning* ja *few-shot learning* -metodeilla, joiden avulla pyrkii selvittämään, kuinka hyvin tekoäly voi selviytyä käännöstoista erilaisten promptien ja ohjeistusten avulla. *Promptilla* tarkoitetaan tehtävänantoa, joka annetaan generatiiviselle tekoälylle jonkin tietyn vastauksen saavuttamiseksi, ja tämän tehtävänannon antoa ilmaistaan verbillä *promptata*. Aiemmin mainituissa Brownin ym. metodeissa on siis kyse siitä, että kuinka paljon tekoälylle annetaan tietoa ennen kääntämistä käännettävästä aiheesta (mts. 5, 6–8). *Zero-shot learning* tarkoittaa sitä, että tekoälylle ei anneta taustatietoa laisinkaan, vaan annetaan yksinkertaisesti käännettävä sana, jolloin sen osaaminen riippuu täysin muiden sinne syöttämästä datasta, ja löytyykö sieltä kyseiseen sanaan tai aiheeseen liittyen mitään. *One-shot learning* puolestaan tarkoittaa sitä, että tekoälylle annetaan tieto, lähtökielinen sana ja kohdekielinen vastine: *dog = koira*, ja sen jälkeen pyydetään kääntämään jokin virke, jossa on muun muassa tämä sana. *Few-shot learning* tarkoittaa samaa kuin *one-shot learning*, mutta tässä metodissa tekoälylle tarjotaan useampi lähtökielen sana, joilla on kohdekielinen vastine ja sitten sitä pyydetään kääntämään jokin aineisto (mts. 7).

Sekä Brown ym. (2020), että Siu (2023) tulevat tutkimuksissaan siihen lopputulokseen, että laajoissa kielimalleissa on paljon potentiaalia kielellisiin tehtäviin, kuten kääntämiseen. Erityisen hyvin ne suoriutuvat kielentarkistamisesta, lauseiden täydentämisestä, luovasta kerronnallisuudesta, yksinkertaisiin (ja ajoittain monimutkaisiinkin) kysymyksiin vastaamisesta ja hakukoneiden kysymysasetteluiden tai esimerkiksi sähköpostien luomisesta (Brown ym., 2020, s. 34; Siu, 2023, s. 8). Käytännössä laajoilla kielimalleilla on mahdollisuus toimia sekä oikolukijoina, että myös luovemmissa tehtävissä.

Koska tekoälypohjaiset laajat kielimallit ovat jokseenkin uusia keksintöjä, löytyy niistä myös paljon kehitettävää monipuolisuudestaan ja hyödynnettävyydestään huolimatta. Molemmat Brown ym. ja Siu ovat tutkimuksissaan listanneet kehityskohtia laajoihin kielimalleihin, kuten: laajojen kielimallien helpon väärinkäytön mahdollisuus (Brown ym., 2020 s. 35), lähteistä ja taustatiedoista johtuva subjektiivisuus (mts. 36), suuri sähkönkulutus (mts. 39), pitkien käännettävien lauseiden haastavuus (Siu, 2023 s. 9), tyylin, kontekstin ja erikoisalojen kääntäminen (mts. 10–12), niiden vähäinen tietämys muista kuin englannin kielestä (mts. 11–12) ja hallusinaatio (engl. *hallucination*), joka tarkoittaa sitä, että laaja kielimalli mieluummin keksii täytettä tekstiin ilman faktapohjaa, kuin myöntäisi, ettei tiedä vastausta (mts. 12–14). Tämä on ongelma etenkin *zero-shot learning* -metodia käytettäessä, kun laajaa kielimallia hyödynnetään tehtävään, josta sillä ei ole aiempaa kokemusta tai dataa aiheeseen liittyen (mts. 12). Sekä Brown ym. (2020), että Siu (2023) ovat molemmat sitä mieltä, että suurin haaste laajoille kielimalleille on erikoisalojen kääntäminen, sillä niitä koulutetaan datalla, joka on niin yleisellä tasolla, että se pystyisi ikään kuin ”vastaamaan kaikenlaisiin kysymyksiin” (Brown ym., 2020, s. 3; Siu, 2023, s. 10). Tällöin sitä ei suoranaisesti pysty hyödyntämään hyvin spesifien käännosten, kuten erikoisalojen tekstien käännosten tuottamiseen ilman laajaa taustatyötä sen koulutukseen.

Peng ym. (2023) on tutkinut ChatGPT:n käännosten laatua kolmella eri mittarilla: lämpötilalla (engl. *temperature*), tehtävänannolla (engl. *task information*) ja vain ChatGPT:n sisältämällä tiedolla (engl. *domain information*). Lämpötilalla tarkoitetaan tässä kontekstissa sitä mittaria, jolla säädellään ChatGPT:n päämääräisyyttä, eli sen tuottaman tekstin informatiivisuutta. Mitä suurempi lämpötila käänökselle annetaan, sitä enemmän esimerkiksi metatietoa ja muuta käännoksen informatiivisuuden kannalta irrelevanttia tietoa käännos sisältää. Vain ChatGPT:n sisältämällä tiedolla tarkoitetaan sen treenidataa, mutta myös käyttäjien sille luovuttamaa tietoa. ChatGPT eroaa tyypillisestä neuroverkkokääntimestä sillä tavalla, että se pystyy myös hyödyntämään käyttäjien kanssa käytyä keskustelua hyväksi vastauksen saavuttamiseksi (mt.). ChatGPT:lle voi esimerkiksi syöttää tietyn erikoisalan termistön, jota hyödyntää käänöksessä, ja se hyödyntää tätä termistöä, vaikka sitä ei olisi sen treenidataan syötetty. Näiden mittarien avulla Peng ym. pyrkii

luomaan promptteja, joilla mahdollisimman hyvin saavuttaa halutun lopputuloksen, eli käännöksen.

Tässä kohtaa lienee hyvä huomata, että generatiiviset tekoälyt ovat jokseenkin uusi innovaatio ja parin viime vuoden aikana tehneet nousua. Tämä on yksi syy, miksi niiden hyödyntämistä käännösteknologisina apuvälineinä ei ole juuri tutkittu ja vertaisarvioitua tutkimusta aiheesta ei vielä tämän maisterintutkielman kirjoitushetkellä löydy. Olen kuitenkin halunnut sisällyttää tässä osiossa mainitut lähteet osaksi tutkielmaani, jotta voitaisiin saada parempi käsitys tästä ilmiöstä ja miksi sitä on syytä tutkia myös jatkossa.

## 2.4 URHEILU KÄÄNTÄMISEN ERIKOISALANA

Koska tässä tutkimuksessa on kyse urheilulajin sääntöjen kääntämisestä, on tarpeellista tutkia myös kyseisen erikoisalan kääntämistä. Tutkielmaa suunnitellessani huomasin, ettei urheilun erikoisalan kääntämistä ole vielä juuri tutkittu, mutta löysin kaksi tutkimusta aiheesta, jotka esittelen tässä alaluvussa. Käytän teoriapohjana urheilun erikoisalan kääntämiseen Gözde Begüm Uyanikin (2017) artikkelia kääntämisestä ja tulkkauksesta urheilukonteksteissa ja Teija Alasalmen (2014) pro gradu -tutkielmaa rugbyin urheilusääntöjen kääntämisestä. Uyanik jakaa kääntämisen ja tulkkauksen urheilussa tyyllillisesti kahteen suureen kategoriaan: kirjoitettuun ja puhuttuun aineistoon (Uyanik, 2017, s. 102). Tämän tutkimuksen kannalta tärkeää on erityisesti kirjoitettu aineisto. Vaikka puhuttu aineisto on myös tärkeässä roolissa tällaisen tutkimusaiheen kannalta, on sen näkökulma tässä tutkielmassa rajautunut tutkimuskysymysten takia ulkopuolelle.

Uyanik nostaa tutkimuksessaan esiin urheilukonteksteissa kääntämisessä ja tulkkauksessa erityisesti yhden haastavaksi koetun asian; terminologian puutteen (mts. 105). Uyanikin mukaan urheilukonteksteissa termien käännösstrategioina käytetään joko suoria lainoja muista lajin valtakielistä (mts. 107) tai suoria käännöksiä, joko sisällöllisesti tai rakenteellisesti. Sisällöllisellä suoralla käännöksellä tarkoitetaan sitä, että käännöksessä pyritään kunnioittamaan lähdekielen ilmausta sisällön osalta, mutta ei esimerkiksi kiinnitetä huomiota sanajärjestykseen pitkien ilmausten kohdalla (mts. 108). Semantiikka menee siis sisällöllisessä suorassa

käännöksessä syntaksin edelle, jolloin lähdekielen ilmaus ei välttämättä niin suoraan paista käännöksen läpi, mutta kohdekielinen ilmaus voi olla vaikeammin tulkittavissa, jos lähde- ja kohdekielien kulttuurit ovat kovin kaukana toisistaan ja käännetty termi on hyvin kulttuuri- tai kontekstisidonnainen.

Rakenteellisella suoralla käännöksellä tarkoitetaan puolestaan sellaista kääntämistä, jossa syntaksi menee semantiikan edelle ja pyritään niin kutsuttuun ”sana-sanaiseen” suoraan käännökseen (mts. 108). Tällaisessa tapauksessa usein lähdekieli on käännöksestä helposti tulkittavissa, eikä käännös ole välttämättä kovin kansankielinen tai helposti käytettävä, tai tällaisessa urheilumaailman kontekstissa kovin käytännöllinen. Minä pyrin tässä tutkimuksessa ja termistössä välttämään sekä rakenteellista, että sisällöllistä suoraa käännösstrategiaa, sillä tutkimuksen tarkoituksena on luoda saavutettava sääntökirja, jonka termit voisivat mahdollisesti vakiintua käyttöön lajiyhteisössä. Ultimatussa lajin sanastoa käytetään etenkin suullisesti pelaajien kesken, ja silloin termienkin tulisi olla ”suuhun sopivia”, helposti käytettäviä ja mieleenpainuvia, jotta ne vakiintuisivat käyttöön.

Uyanik mukaan, että urheilutekstien kääntämiseen tarvitaan ammattilaiskääntäjä, jolla on myös vahva tietämys ja ymmärrys kyseisestä lajista (mts. 109, 112). Kääntäjältä vaaditaan sääntökirjojen kääntämiseen jopa lakitekstien kääntämiseen tarvittavaa otetta (mts. 111), eli sanan ja sanamuodon valinnalla voi olla hyvin perustavanlaatuinen merkitys tekstiä tulkittaessa ja siksi sääntökirjojen kääntäminen vaatii kääntäjältä tarkkuutta. Tästä voitaisiin päätellä, että kääntäjäksi ei sovellu pelkästään kouluttautunut kääntäjä, eikä myöskään erikoisalan tietämystä omaava maallikkokääntäjä, vaan molemmat taidot omaava kääntäjä.

Alasalmi (2014) on pro gradu -tutkielmassaan tutkinut urheilun erikoisalakääntämistä ja pyrkinyt luomaan sille tutkimuksellista viitekehystä. Alasalmen tutkielman tutkimuskohteena ovat rugby säännöt, joiden käännösongelmia tutkimalla Alasalmi pyrkii muodostamaan viitekehyksen urheilusääntöjen kääntämiselle.

Alasalmi määrittelee urheilusäännöissä esiintyvän kielen ja ylipäänsä urheilun ympärillä puhutun ja kirjoitetun kielen jonkinasteiseksi jargoniksi (mts. 10) ja tiettyä

urheilulajia koskevan kielen taas lajijargoniksi (engl. *sport jargon*). Urheilun kieli (engl. *sport language*) taas kattaa sekä kirjoitetun, että puhutun kielen, mutta myös eleet, kuvat, kielikuvat ja metaforat, joita käytetään urheilutilanteissa tai urheilun yhteydessä käytävässä sosiaalisessa kanssakäymisessä (Alasalmi, 2014, s. 11).

Alasalmi toteaa yhtyen Uyanikin tutkimukseen, että urheilusäännöissä on paljon lakitekstiä muistuttavia elementtejä, kuten että urheilusäännöistä puhutaan englanninkielisessä versiossa lakeina (2014, s. 76), ja niissä käytetään tyyllisesti lakiteksteistä tuttuja kielellisiä ilmauksia, kuten verbin *must* käyttö (mts. 77). Alasalmi selittää tätä tyyliä sillä, että urheilusääntöjen on, kuten lakitekstien, vältettävä tulkinnanvaraisuutta ja oltava kaikin puolin selkeitä (mts. 77).

Alasalmen tutkimuksen mukaan suurin osa (42 prosenttia kaikista tunnistetuista käännösongelmista) käännösongelmista liittyi kyseiseen urheilulajiin tai sen termistöön ja toiseksi suurinta käännösongelmakategoriaa edustivat kieliopilliset virheet (yhteensä 31 prosenttia kaikista ongelmista) (mts. 81). Tämä osoittaa yhdessä Alasalmen tekemän kirjallisuuskatsauksen (mts. 12–16) kanssa, että termityö on hyvin keskeisessä asemassa urheilun ja muiden erikoisalojen kääntämisessä (mts. 116). Alasalmi toteaa myös, että termiongelmiin tai urheilulajiin spesifisti liittyvien käännösongelmien kohdalla monien eri käännösstrategioiden hyödyntäminen on tarpeellista (mts. 96).

Käännösongelmissa Alasalmi ei lainkaan turvautunut lajin suomalaisten asiantuntijoiden, eli pelaajien, valmentajien tai tuomarien puoleen, vaan nojasi aiempiin sääntöversioihin, suomenkieliseen raakakäännökseen ja eri kielisiin rinnakkaisteksteihin (mts. 116). Alasalmi on kuitenkin yhtä mieltä Uyanikin kanssa siitä, että kääntäjän tulisi tuntea käännettävien urheilusääntöjen laji hyvin (mts. 116).

Koska Alasalmen tutkimus keskittyy vain yhden lajin sääntöihin, ei tämän tutkimuksen perusteella voida tehdä yleistyksiä urheilusääntöjen kääntämisestä tai rakentaa kokonaisvaltaista viitekehystä urheilusääntöjen kääntämiseen. Alasalmi on kuitenkin tehnyt sellaisia löydöksiä, joita voidaan hyödyntää muiden lajien urheilusääntöjen kääntämisen tutkimuksessa jatkossa. Alasalmen mukaan urheilusääntöjen kääntäminen myös jatkossa on tärkeää neljästä eri syystä (mts. 63):

1. Urheilusääntöjen kääntäminen tekee lajista saavutettavamman uusille harrastajille
2. Pelaajat, valmentajat ja tuomarit hyötyvät omalla äidinkielellään kirjoitetuista säännöistä
3. Urheilusääntöjen kääntäminen tuo urheilulajin lähemmäs ihmistä, jolla ei ole muuta tietoa lajista
4. Panostus lajin saavutettavuuteen tuo myös kansainvälistä tunnustusta

Koska urheilun erikoisalakääntämistä on toistaiseksi tutkittu hyvin vähän, ja Uyanikin ja Alasalmen tutkimukset eroavat näkökulmiltaan merkittävästi toisistaan, ei niiden tulosten pohjalta voida tehdä yleistyksiä kyseisen erikoisalakääntämisen konventioista. Näiden tutkimusten tuloksia voidaan kuitenkin peilata muihin tuleviin urheilun erikoisalalla tehtyihin käännostutkimuksiin, ja siten mahdollisesti pyrkiä rakentamaan jonkinlaista viitekehystä erikoisalan kääntämiseen ja se on myös tämän maisterintutkielman tavoite.

### 3 TUTKIMUSMENETELMÄ JA AINEISTO

Tässä luvussa esittelen tutkimusmenetelmäni varten lukemaani kirjallisuutta ja selitän, miksi olen valinnut juuri kyseisen tutkimusmenetelmän.

Tutkimusmenetelmän kuvauksen jälkeen selitän vaihe vaiheelta, kuinka aineisto on kerätty ja mitä haasteita aineistonkeruussa on ilmennyt, minkä jälkeen kuvailen lyhyesti keräämäni aineiston.

#### 3.1 TUTKIMUSMENETELMÄ

Tässä tutkimuksessa on kyse tapaustutkimuksesta, sillä konekääntimen tai ChatGPT:n hyödyntämistä urheilusääntöjen kääntämisessä ei ole aiemmin tutkittu. Tämän lisäksi tutkimuksen kohteena toimii vain yhden urheilulajin yksi sääntöversio. Tämä tutkimus on tapaustutkimus siis sen vuoksi, että sen tulosten perusteella ei voida tehdä yleistyksiä urheilusääntöjen kääntämisestä tai konekääntimen tai laajojen kielimallien hyödyntämisestä niiden kääntämisessä (Vilkkä, Saarela ja Eskola, 2018, s. 196–197). Tarkoitus on kuitenkin edistää tekstilajin ja sen kääntämisen tutkimusta ja teoriassa siihen soveltuvien työkalujen avulla kääntämisen tutkimusta, jolloin tapaustutkimus on myös yleisesti tärkeä. Koska kyse on tapaustutkimuksesta, on tärkeää hyödyntää monimetodisuutta (mts. 198), jotta päätelmissä voitaisiin lähestyä yleistämistä esimerkiksi useamman eri urheilulajin sääntöjen kääntämiseen ja siten havainnollistamaan mahdollisimman hyvin tutkimuksen kohteena olevaa ilmiötä.

Tutkimusmetodin pohjana tässä tutkimuksessa käytetään Maarit Koposen (2010) pilottitutkimuksessaan käyttämää virheanalyysiä ja siinä luotua virheanalyysiluokittelua konekääntimien tekemille virheille. Koponen (2010) on tutkimuksessaan vertaillut sääntö- ja tilastopohjaista konekäännintä ja näiden tekemiä virheitä kolmen eri tekstilajin käännösten perusteella. Minun tutkimukseni eroaa Koposen (2010) tutkimuksesta sillä tavalla, että tässä tutkimuksessa vertaillaan neuroverkkokäännintä ja laajaa kielimallia ja niiden tekemiä virheitä yhden tekstilajin yhteen tekstipätkään. Alla Koposen virheanalyysiluokittelut:

Taulukko 1: "concept error categories" (2010, s. 4–5)

| <b>ERROR CATEGORY</b> | <b>EXPLANATION</b>   | <b>EXAMPLE</b>   |
|-----------------------|--|--|
| Omitted concept       | ST concept that is not conveyed by the TT  | Example 1: <i>places</i>   |
| Added concept         | TT concept that is not present in the ST   | Example 1: <i>henkilöllä</i> 'person'  |
| Untranslated concept  | SL words that appear in TT   | Example 1: <i>locale</i>   |
| Mistranslated concept | A TT concept has the wrong meaning for the context   | Example 1: <i>inhimillinen kehitys</i> 'humane development' for <i>human evolution</i> |
| Substituted concept   | TT concept is not a direct lexical equivalent for ST concept but can be considered a valid replacement for the context | <i>valitse</i> 'select' for <i>click</i>   |
| Explicitated concept  | TT concept explicitly states information left implicit in ST without adding information                                | addition of <i>ohjelma</i> 'program' to <i>Norton AntiVirus</i>                        |

Taulukko 2: "relation error categories" (2010, s. 6)

| <b>ERROR CATEGORY</b>   | <b>EXPLANATION</b>   | <b>EXAMPLE</b>   |
|-------------------------|--|--|
| Omitted participant     | ST relation not conveyed by the TT due to an omitted head or dependent   | Example 1: all relations involving the omitted concept <i>places</i>   |
| Omitted relation        | ST relation not conveyed by the TT due to morpho-syntactic errors that prevent parsing the relation although both concepts are present in TT | Example 1: relation between <i>the center</i> and <i>one</i> cannot be parsed in Finnish   |
| Added participant       | TT relation not present in ST introducing an added concept   | Example 1: addition of <i>henkilöllä</i> 'person' and <i>henkilöllä on yksi</i> 'the person has one'                             |
| Added relation          | TT relation not present in ST arises due to morpho-syntactic errors  | Example 2: <i>Norton AntiVirus</i> as a modifier of <i>CD-levyiltä</i>   |
| Mistaken participant    | Head or dependent of the relation different in ST and TT, not same entity  | Example 3: <i>tilanneilmmaisimen</i> 'progress bar' as subject of <i>napsahtavan</i> 'to click'                                  |
| Mistaken relation       | Relation between two concepts different in ST and TT, changed role   | Example 3: temporal relation of <i>seuraavaksi</i> 'next' and <i>napsahtavan</i> 'to click'                                      |
| Substituted participant | Head or dependent of the relation different in ST and TT, same entity  | one TT concept, e.g. <i>kansalaista</i> 'citizens' or 'nationals', replaces two in ST, e.g. <i>citizens</i> and <i>nationals</i> |
| Substituted relation    | Relation between two concepts different in ST and TT, same semantic roles  | <i>Lissabonin sopimuksella muutetaan</i> 'with the Treaty of Lisbon is amended' for Treaty of Lisbon amends                      |

Virheanalyysin lisäksi sovelsin tässä tutkimuksessa asiantuntijakyselyn aineistonkeruumenetelmää ja yhdistän nämä kaksi saadakseni enemmän dataa ja useamman näkökulman käännöksiin. Asiantuntijoina tässä tutkimuksessa toimivat kielen ja kääntämisen opiskelijat sekä lajin asiantuntijat, joilla on näkemystä tekstin sisällölliseen oikeellisuuteen.

Käännätin saman tekstipätkän säännöistä sekä Phrasen konekääntimellä, että ChatGPT:llä ja annoin molemmille samat lähtötiedot (termistö, Liite 4.). Tämän

jälkeen annoin lähdetekstin ja käännökset asiantuntijayleisölle arvioitaviksi tekstityökaluun ja pyysin heitä korostamaan virhekohdat tekstissä ja lisäämään omat korjausehdotuksensa virheanalyysin luokittelua helpottaakseni, eli vastaajat pyrkivät parhaansa mukaan jälkieditoimaan käännöksiä. Seuraavaksi näiden vastausten perusteella luokittelin konekääntimen ja ChatGPT:n tekemät virheet ja laskin tuloksen prosentteina ja vertasin sitä koko tekstipätkän sanamäärään, kuten Koponen on tutkimuksessaan tehnyt (2010, s. 7).

Hyödynsin siis tutkimuksessani kyselytutkimusta tutkimusmetodina, ja tarkemmin avoimen kysymyksen kyselytutkimusta, sillä se on näkemykseni mukaan paras tutkimuksen toteutustapa, joka ei johdattele asiantuntijalukijaa liikaa. Avoin kysymys mahdollistaa myös aineiston kahdenlaisen arvioinnin, kvalitatiivisen ja kvantitatiivisen (Valli, 2018, s. 114). Asiantuntijat joutuvat pohtimaan, miksi jokin asia on väärin, koska heidän tulee esittää korjausehdotus korostamansa virheen kohdalla tekstissä, joka edellyttää kvalitatiivista arviointia. Koposen (2010) virheanalyysin kategorisointi taas auttaa hahmottamaan löydöksiä tilastojen valossa ja siten tekemään päätelmiä jälkieditoinnin tarpeen määrästä.

Havainnointi on myös yksi tutkimuksen metodeista, jota asiantuntijat joutuvat harjoittamaan etsiessään teksteistä virheitä ja esittäessään niihin korjausehdotuksia. Tässä tutkimuksessa käytetty havainnoinnin muoto on näkemisen avulla tehty havainnointi tutkimuskohteesta eli kolmesta lyhyestä tekstipätkästä (Vilkkä, 2018, s. 162). Havainnointi tutkimusmetodina tarjoaa myös mahdollisuuden tulevaisuudessa tutkia erilaisuuksia lajin asiantuntijoiden ja kielen asiantuntijoiden tekemien havaintojen välillä, sillä lukijaa ei johdatella päätelmiin tavanomaisin kyselytutkimuksen toteutuksin, kuten erilaisten mittarien tai valmiiden vastausvaihtoehtojen avulla (mts. 106–113). Mittarit ovat myös ristiriidassa virheanalyysin tutkimusmenetelmän kanssa, joten ne rajautuvat kyselytutkimuksen toteutustapana tämän tutkimuksen ulkopuolelle.

Sosiaalinen media on yksi tutkimusmenetelmäni kulmakivistä ja mahdollisti suunnittelemani aineistonkeruun. Sosiaalisen median alustat tarjosivat minulle mahdollisuuden löytää enemmän mahdollisia asiantuntijavastaajia, kuin kynällä ja paperilla toimitettu tutkimus olisi saanut aikaan. Sosiaalinen media on ennen

kaikkea tutkimukseni väline, eikä sen kohde tai paikka, minkä takia voin sivuuttaa sosiaaliseen mediaan tutkimuksen aiheena tai tutkimuspaikkana kohdistuvat tutkimuseettiset kysymykset (Valli ja Perkkilä, 2018, s. 124, 126). Hyödynsin seuraavia sosiaalisen median tarjoamia työkaluja tutkimuksessani: Facebook ja sähköpostilista potentiaalisten vastaajien keräämiseen, Google Forms tutkimusluvan ja yhteystietojen keräämiseen ja Google Docs virheanalyysin aineistonkeruuseen. Näiden sovellusten avulla käsittelin keräämääni aineistoa anonymisti.

Konekääntimien laadun arvioinnissa käytetään usein BLEU-mittaa (Koponen, 2010, s. 1), mutta se vaatisi paljon enemmän dataa hyvälaatuisista käännöksistä ja etenkin lähdetekstiin lähemmin liittyvistä hyvistä käännöksistä, kuin tämän tutkimuksen ja sen aiheen puitteissa on mahdollisuutta kerätä, joten olen rajannut sen arviointimenetelmänä pois.

Perehdyin menetelmän valinnassa kirjallisuuteen valintaverkkomenetelmään liittyen, mutta se rajautuu tutkimuskysymyksien vuoksi tutkielman ulkopuolelle. Salmen ja Koposen (2020) tutkimuksessa tutkittiin kuinka erilaiset kääntäjät tekevät muokkauksia eri konekääntimillä konekäännettyyn tekstiin valintaverkkomenetelmän avulla (Salmi ja Koponen, 2020). Valintaverkkoanalyysi voi myös olla hyvä apuväline samanlaisen materiaalin tutkimisessa, mutta se tarjoaa enemmän kääntäjään keskittyvän näkökulman ja kertoo enemmän siitä, mitkä konekääntimen tekemät ratkaisut kuormittavat kääntäjää enemmän kuin toiset. Valintaverkkoanalyysi menetelmänä rajautuu tutkimuksen ulkopuolelle myös työläämmän luonteensa ja menetelmän toteuttamiseen vaadittavien työkalujen, kuten silmänliikemittauksen ja käännösprosessien tutkimusryhmien vuoksi.

Tutkimukseni ulkopuolelle rajautuu myös jälkieditoinnin kognitiivisen työn mittaaminen, joka mainitaan myös Koposen ym. (2012) tutkimuksessa. Kognitiivista työmäärää olisi hyvä erikseen myös tutkia, sillä se ei Koposen ym. (2012) mukaan aina korreloidu ajallisen ja teknisen työpanoksen kanssa. Kognitiivista työmäärää olisi tässä tutkimuksessa voinut tutkia esimerkiksi aiemmin mainitun valintaverkkomenetelmän tai vastaajien haastattelun avulla, mutta sekin rajautuu tutkimuksen ulkopuolelle resurssipulan ja tutkimuksen laajuuden takia.

## 3.2 TYÖVAIHEET

Ensimmäisenä työvaiheena oli ultimateen liittyvän sanaston kerääminen, vastineiden löytäminen ja niiden muokkaaminen tekstiin soveltuviksi. Tämän työvaiheen olen tehnyt osana terminologista kurssityötä, joten en siitä jaa tähän tutkimukseen tämän enempää.

Seuraavaksi etsin kääntämisen kannalta kiinnostavan, noin 400-sanaisen tekstipätkän säännöistä. Kiinnitin tekstipätkän valikoinnissa huomiota myös siihen, että se sisältäisi mahdollisimman monia termistön termeistä, jotta konekääntimellä ja ChatGPT:llä olisi mahdollisimman hyvät onnistumismahdollisuudet.

Seuraavaksi vein tekstin tekstinmuokkausohjelmaan, jossa segmentoin, eli jaoin tekstiä pienempiin osiin käännöstyökalua ja konekäännintä varten. Segmentointiin kuuluu muun muassa kappalejako ja rivinvaihdosten sekä sääntöpykälänumeroiden siirtämistä oikealle paikalleen. Segmentoinnin ohessa poistin tekstipätkästä ylimääräistä tietoa, kuten sivunumerot, jotka olivat tekstiä kopioidessa eksyneet tekstiin mukaan, jaottelin pitkiä tekstipötköjä sanoihin, jotka olivat tekstiä kopioidessa siis menettäneet kaikki välilyönnit ja muuta, joka tekee tekstistä täysin luettavan version. Tekstin esikäsittelyyn minulla kului tovi, kuten yleensä käännösprojektien managereilla kuluu, mutta en aio tätä aikaa ottaa huomioon tutkimuksessani, sillä tarkoitus on tutkia tekstin jälkieditointiin kuluva työpanosta.

Vein seuraavaksi käännettävän tekstin ja termit ChatGPT-sovellukseen. ChatGPT3.5 ei vaatinut paljolti ohjeistusta ja sille riitti tehtävänannoksi: ”käännä teksti englannista suomeen käyttäen annettuja termejä”. Huomasin GPT3.5:n käännösversioita lukiessani, että termistöstäni puuttuu hyvin olennaisia termejä, joten palasin termityökaluun lisäämään termejä ja niiden vastineita, ja sitten syötin käännettävän tekstin ja termistön GPT3.5:lle uudestaan. Syy, miksi korjasin välissä termityökaluun syöttämiäni termejä, oli se, että käännöstyökaluun syöttämäni termit olisivat linjassa GPT3.5:lle ilmoittamieni termien kanssa ja täten Phrase saisi myös saman termistön kuin GPT3.5.

Seuraavaksi loin projektin Phrasen, jossa määrittelin lähtö- ja kohdekielet, käytettävän konekääntimen ja termistön. Projektin luomisen jälkeen lähdin itse ”kääntämään” käännettävää tekstiä, eli oikeastaan hyväksymään ja vahvistamaan konekääntimen ehdotuksia. Tutkimukseen oli tarkoitus käyttää kummankin työkalun raakakäännöksiä, eli täysin muokkaamattomia käännöksiä ja siksi en voinut muokata konekäännöstä niin, että siinä olisi otettu huomioon termikannan termistö. Eli termistön aiemmasta päivittämisestä huolimatta se ei tullut Phrasessa käyttöön, sillä konekäännöksen raakakäännöstä olisi tullut jälkieditoida, jotta termistön termit olisi voitu ottaa käännöksessä käyttöön.

### **3.3 AINEISTO**

Keräsin aineistoni asiantuntijoilta, joita pyysin osallistumaan tutkimukseeni erilaisten sosiaalisen median kanavien kautta. Asiantuntijoilla tarkoitan tässä tutkimuksessa englannin kielen ja kääntämisen opiskelijoita, sekä ultimatien asiantuntijoita, eli lajin harrastajia. Pyysin potentiaalisia vastaajia ensin täyttämään yhteystietonsa ja antamaan luvan tutkimukseen erillisellä lomakkeella, jonka jälkeen lähetin jokaiselle vastaajalle oman korjattavan ja kommentoitavan dokumentin Google Docs -alustalla. Google Docs -alustalla vastaajat saavat käsiinsä aineiston, eli lähtötekstin, sekä GPT3.5:n ja Phrasen käännökset lähtötekstistä (Liitteet 1.–3.). Vastaajilla ei ollut pääsyä termistöön, joka on ollut kääntämisen apuna konekääntimellä ja generatiivisella tekoälyllä, sillä en halunnut sillä rajoittaa vastaajien tulkintaa termeistä tai niiden vastineista. Tässä vaiheessa vastaajia oli ilmoittautunut yhteensä 17.

Kahden viikon muokkaus- ja kommentointiajan jälkeen latasin tiedostot omalle laitteelleni, joista huomasin, että muokkauksia oli tehty yhteensä kymmeneen tiedostoon, mutta yhdessä tiedostossa oli muokkauksia tehty vain toiseen käännöksistä, jolloin tämä vastaus jää osittain tutkimukseni ulkopuolelle. Tämä vastaus on otettu huomioon keskiarvoissa, mutta ei kokonaistuloksissa, jotta voidaan taata tulosten objektiivisuus ja luotettavuus.

Aineiston tutkimukseen olisin voinut kerätä vain oman arvioni pohjalta, mutta koin virheanalyysin kannalta paremmaksi, että arvioita olisi useammasta näkökulmasta,

jolloin analysoitavaa dataa on myös enemmän. Halusin myös ennen kaikkea lisätä näkökulmia sekä kielelliseen, että käännosten sisällölliseen puoleen, minkä takia useampi arvioija voisi olla parempi. Oletus tässä kohtaa minulla oli se, että lajin asiantuntijat pystyvät tarjoamaan parempaa näkemystä käännosten sisällöstä ja sen oikeellisuudesta, kun taas kielen asiantuntijat tulisivat todennäköisesti keskittymään kielellisiin asioihin, tekstin sisällön hyvin ymmärrettävästi olevan hieman hankalaa hahmottaa yhden lyhyen tekstipätkän avulla. Tätä oletusta ei voida kuitenkaan tässä tutkimuksessa tutkia, sillä käsittelin vastauksia anonyymisti, jolloin aineistosta ei käy ilmi, kumpaan kohderyhmään kukin vastaaja kuuluu.

Seuraavassa luvussa esittelen aineiston pohjalta tehdyn virheanalyysin tulokset ja analysoin tuloksia yhdessä teoreettisen viitekehyksen kanssa.

## 4 ANALYYSI

Tämän tutkimuksen tarkoituksena oli selvittää, kuinka paljon jälkieditointityötä konekääntimen ja generatiivisen tekoälyn käännökset määrällisesti vaativat, kun käännetään urheilusääntöjä. Tarkoituksena oli samalla arvioida konekääntimen ja generatiivisen tekoälyn tekemien käännösten laatua, jota tässä tutkimuksessa tutkittiin tutkimusmenetelmästä kertovassa luvussa tutustetun virheanalyysimetodin avulla.

Hypoteesi tässä tutkimuksessa oli se, että konekäännin tekee generatiivista tekoälyä vähemmän virheitä kääntäessään tekstiä. Konekäännin ei ainoastaan ole hyvä kääntämään tekstejä, joissa on paljon toisteisuutta, mutta saa etulyöntiaseman kokemuksestaan ja pitkäjänteisemmästä käytöstään generatiiviseen tekoölyyn verrattuna, joka on ollut käytössä kielenkääntämisessä vasta muutaman viime vuoden ajan. Konekäännin on todettu hyväksi työkaluksi asiatekstien, kuten lakitekstien kääntämisessä, jotka puolestaan Uyanikin (2017) mukaan muistuttavat urheilusääntöjen pykälämäistä tyyliä, joten hypoteesin voitaisiin lähdekirjallisuuden perusteella olettaa toteutuvan.

Seuraavaksi esittelen aineiston virheanalyysin tulokset erilaisten taulukoiden avulla. Tarkastelen ensin tuloksia yleisesti ja myöhemmin tässä luvussa tarkastelen tarkemmin jaotteleamalla tuloksia sisällöllisiin virheisiin ja kieliopillisiin virheisiin ja erinäisiin ilmiöihin, joita on ilmennyt aineiston analyysin aikana.

### 4.1 TULOKSET

Ensimmäisenä lähdin karkeasti laskemaan sekä konekääntimen, että generatiivisen tekoälyn tekemiä virheitä sen perusteella, mitä vastaajat olivat teksteistä korjanneet. Laskin virheiksi kaikki korjaukset/muokkaukset, mitä vastaajat olivat dokumentteihin tehneet. Taulukossa 3 esitetään, kuinka paljon sanoja kussakin tekstissä on, kuinka monta virkettä kussakin tekstissä on, kuinka monta virhettä kussakin tekstissä on keskimäärin ja kuinka monta virhettä on keskimäärin yhdessä virkkeessä.

Taulukko 3: tietoa tekstipätkistä ja virheanalyysin tulokset

|                                | <b>LÄHTÖTEKSTI</b> | <b>CHATGPT3.5</b> | <b>PHRASEN<br/>KONEKÄÄNNIN</b> |
|--------------------------------|--------------------|-------------------|--------------------------------|
| <b>SANAMÄÄRÄ</b>               | 417                | 297               | 313                            |
| <b>VIRKEMÄÄRÄ</b>              | 27                 | 27                | 27                             |
| <b>SANAMÄÄRÄ<br/>PER VIRKE</b> | 15,44444444        | 11                | 11,59259259                    |
| <b>VIRHEET YHT.</b>            |                    | 420               | 472                            |
| <b>VIRHEET PER<br/>VASTAUS</b> |                    | 46,66666667       | 52,44444444                    |
| <b>VIRHEET PER<br/>VIRKE</b>   |                    | 1,728395062       | 1,942386831                    |

Taulukosta 3 käy jo yksinään ilmi, että hypoteesi tämän tutkimuksen ja aineiston puitteissa olettaa väärin ja konekääntimen tekemästä käännöksestä löytyi yhteensä enemmän virheitä kuin generatiivisen tekoälyn tekemästä käännöksestä. Tämä kokonaistulos heijastuu myös virheiden keskimäärään yhdessä tekstipätkässä ja virkkeessä, jotka myös käyvät taulukosta 3 ilmi. Vaikka taulukko 3 antaa olettaa keskiarvoillaan sen, että jokaisessa virkkeestä löytyisi virheitä kummaltakin, konekääntimeltä ja generatiiviselta tekoälyltä, kummastakin tekstipätkästä löytyi myös sellaisia lauseita ja virkeitä, joissa vastausten mukaan ei ollut ollenkaan virheitä.

Aineistossa vähiten korjattu virke on lähtötekstin

Esimerkki 1: *18.1.1.1.3. counts in less than one second intervals*, (Liite 1.),

jossa tarkoitetaan markkerin suorittamaan pysäytyslaskentaa sille pelaajalle, jolla on frisbee sillä hetkellä hallussaan. Pysäytyslaskentaa suoritetaan siksi, että kukaan pelaaja ei voisi ikuisesti pitää pelivälinettä käsissään, jolloin peli ei välttämättä etenisi laisinkaan. Jos pysäytyslaskennassa saavutetaan aikaraja, joko 10 sekuntia tai 8 sekuntia (tämä riippuu peliformaatista), hyökkäysvuoro vaihtuu aiemmin puolustaneelle joukkueelle.

Generatiivinen tekoäly on kääntänyt tämän:

Esimerkki 2: *18.1.1.1.3. laskee alle yhden sekunnin välein, (Liite 3.)*

ja konekäännin on kääntänyt:

Esimerkki 3: *18.1.1.1.1.3. laskee alle sekunnin välein, (Liite 2.),*

jotka voidaan tässä todeta täysin kieliopillisesti oikeiksi ja sisällöltään totuuden- ja lähtötekstinmukaisiksi ja ei siksi ole yllätys, että niitä ei ole vastauksissa muokattu.

Aineistosta löytyi pari virkettä, joita on korjattu huomattavasti muita virkkeitä enemmän. Ensimmäinen virke on lähtötekstissä:

Esimerkki 4: *18.1.3. After all marking infractions listed in 18.1.1 that are not contested, the marker must resume the stall count with the number last fully uttered before the call, minus one (1). (Liite 1.)*

Tässä esimerkissä puhutaan markkerin virheistä tai rikkomuksista. Koska ultimatessa kaikki virhetilanteet tuomitaan pelaajien toimesta, on näihin virhehuutoihin mahdollista osoittaa epätyytyväisyytensä kiistämällä huuto ja toisaalta myöntämällä huuto, jos on virhetilanteesta samaa mieltä virheen huutaneen pelaajan kanssa.

Konekäännin käänsi tämän virkkeen seuraavanlaisesti:

Esimerkki 5: *18.1.3. Kaikkien 18.1.1 kohdassa lueteltujen merkintärikkomusten jälkeen, joita ei ole kiistetty, merkitsijän on jatkettava karsintalaskentaa sillä numerolla, joka oli viimeksi kokonaan lausuttu ennen kutsua, vähennettynä yhdellä (1). (Liite 2.)*

ja generatiivinen tekoäly:

Esimerkki 6: 18.1.3. Kaikkien 18.1.1 kohdissa lueteltujen merkkauseiden jälkeen, jotka eivät ole kiistettyjä, markkerin on jatkettava pysäytyslaskentaa viimeksi täysin lausutulla numerolla ennen huutoa, vähennettynä yhdellä (1). (Liite 3.)

Esimerkeissä 5 ja 6 on korjattu sekä virkkeen termejä, että myös lauserakennetta. Näissä esimerkeissä valmiiksi syötettyjä termejä ovat *rike*, *markkeri*, *huuto* ja *pysäytyslaskenta*, joita generatiivinen tekoäly (esimerkki 6) on onnistuneesti käyttänyt, mutta konekäännin (esimerkki 5) on käyttänyt *rikkomus*, *merkitsijä*, *kutsu* ja *karsintalaskenta*. Näistä *rikkomus* ja *kutsu* ovat lähtötekstin perusteella ymmärrettäviä ja pelaaja tulisi varmasti ymmärretyksi näitä käyttäessään pelitilanteessa, joten kyse on enemmän tyylivirheestä kuin termin vääryydestä. *Merkitsijä* ja *karsintalaskenta* taas ovat niin kaukana tavoitelluista termeistä, että nämä termit haittaavat ymmärrettävyyttä.

Esimerkit 4–6 osoittavat myös hyvin sen, että vaikka konekäännin ja generatiivinen tekoäly ovat kummatkin syntaktisesti taitavia ja kykeneviä oppimaan kunkin kielen kielioppisäännöt, ovat oikeastaan kykenemättömiä kotouttamaan lähtötekstin rakennetta kohdekieleen sopivaksi. Esimerkeissä 5 ja 6 nähdään sivulauseita erottelevia pilkkuja myöten, että kummassakaan käännöksistä ei ole osattu yhdistää lauseita tai jakaa pitkää ja monimutkaista virkettä osiin, mikä olisi suomen kielen konventioille tyypillisempää.

Toinen eniten korjatuista virkkeistä on heti lähtötekstin seuraava pykälä:

Esimerkki 7: 18.1.4. *The marker may not resume counting until any illegal positioning has been corrected. To do otherwise is a subsequent marking infraction.* (Liite 1.),

generatiivinen tekoäly käänsi sen:

Esimerkki 8: 18.1.4. *Markkeri ei saa jatkaa laskentaa ennen kuin kaikki laittomat asennot on korjattu. Toimiminen toisin on seuraava merkkause.* (Liite 3.)

ja konekäännin:

Esimerkki 9: 18.1.4. *Merkkiainetta ei saa laskea uudelleen ennen kuin laitton sijainti on korjattu. Muussa tapauksessa kyseessä on myöhempi merkintärikkomus.* (Liite 2.)

Käännöksiä tarkastellessa voidaan todeta, että vaikka kyseessä ei olekaan pitkä ja monimutkainen virke, voivat käännökset olla melko epäonnistuneita. Esimerkeissä 5 ja 9 olennaista on kuitenkin se, että konekäännin käyttää termejä hyvin epäjohdonmukaisesti, koska se ei ole voinut hyödyntää termistöä samalla tavalla kuin generatiivinen tekoäly ja siksi koko lauseen merkitys jää hyvin vaikeaksi ymmärtää ilman lähtötekstiä. Generatiiviselle tekoälylle syötettyjä termejä esimerkissä 8 ovat *markkeri* ja *rike*. Esimerkissä 8 taas voidaan huomata hyvin yksinkertainen kirjoitusvirhe *merkkausrikke*, joka on hieman yllättävä virhe generatiivisen tekoälyn käännöksessä.

Taulukon 3 perusteella voidaan myös tehdä yleislaatuinen huomio konekääntämisestä tai generatiivisella tekoälyllä kääntämisestä; virkkeiden määrä ei ole lisääntynyt tai vähentynyt käännösprosessin aikana kummassakaan käännöksessä alkuperäistekstiin verraten, vaan on kaikissa sama 27 virkettä. Tästä puolestaan voitaisiin päätellä, että konekäännin tai generatiivinen tekoäly ei kumpikaan omista luovia valmiuksia muokata lähtötekstissä ilmaistuja lauserakenteita, vaan pyrkii noudattamaan niitä myös kohdetekstissä siitä huolimatta, istuvatko ne molempiin kieliin vai eivät. Ihmiskääntäjä voisi tässä kohtaa esimerkiksi pyrkiä kotouttamaan lähdekielen lauserakenteen kohdekieleen sujuvoittaakseen tekstiä ja sen saavuttamiseksi joutuisi esimerkiksi pätkimään lauseita lyhyemmiksi tai vaihtoehtoisesti yhdistämään lauseita.

Virkkeiden määrästä huolimatta suomenkielisissä käännöksissä on hieman odotetustikin vähemmän sanoja kuin englanninkielisessä alkuperäistekstissä; konekäännöksessä 104 sanaa vähemmän kuin alkuperäistekstissä ja generatiivisen tekoälyn käännöksessä jopa 120 sanaa vähemmän kuin alkuperäistekstissä. Suomen kielessä ei käytetä artikkeleja ja suffikseilla voidaan ilmaista asioita, joihin englannin kielessä tarvitaan kokonaisia sanoja, kuten prepositiot, ja muun muassa siksi teksti on varmasti suomenkielisissä käännöksissä jopa 28 prosenttia lyhyempi. Tämä ilmiö

saattaa osittain selittää myös virheiden määrän eroa generatiivisen tekoälyn käännökseen ja konekäännökseen välillä, sillä sanamäärän kasvaessa lähtö- ja kohdetekstien välillä varmasti myös virheiden määrä kasvaa.

Alla vielä virheanalyysin tulokset eriteltyinä käännökseen, vastaajan ja virhekattegorian mukaan.

Taulukko 4: konekääntimen virheet vastaajittain ja kategorioittain järjesteltyinä

| <b>VIRHEANALYYSI-KATEGORIA</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> |
|--------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| POISJÄTETTY KÄSITE             | 2        | 6        | 3        |          | 1        | 4        | 4        | 2        | 3        | 7         |
| LISÄTTY KÄSITE                 | 1        |          |          | 3        | 1        |          |          |          |          |           |
| KÄÄNTÄMÄTÖN KÄSITE             | 1        | 1        | 1        |          |          |          |          |          |          | 1         |
| VÄÄRINKÄÄNNETT Y KÄSITE        | 11       | 26       | 32       | 7        | 33       | 39       | 38       | 10       | 20       | 37        |
| KORVATTU KÄSITE                | 17       | 14       | 19       | 13       | 19       | 14       | 17       | 8        | 9        | 22        |
| SELKEYTETTY KÄSITE             | 1        |          |          |          |          |          |          |          |          |           |
| POISJÄTETTY LAUSEENJÄSEN       |          | 1        |          |          |          |          | 1        |          | 1        | 1         |
| POISJÄTETTY MUU                |          |          |          |          |          |          |          |          |          |           |
| LISÄTTY LAUSEENJÄSEN           |          |          |          |          |          |          |          |          |          |           |
| LISÄTTY MUU                    | 1        |          |          |          |          |          |          |          |          |           |
| VÄÄRÄ LAUSEENJÄSEN             |          | 1        |          |          |          |          |          |          |          |           |
| VÄÄRÄ MUU                      | 8        | 5        | 6        |          | 5        | 4        | 1        | 6        | 1        | 6         |
| KORVATTU LAUSEENJÄSEN          |          |          |          |          |          |          |          |          |          |           |

Taulukoista 4 ja 5 käy selvästi ilmi, että kaikki Koposen (2010) virhekattegorioista eivät ole tässä tutkimuksessa tulleet tarpeeseen. Tämä saattaa johtua tekstilajista, käännettävästä kielestä tai konekääntimestä tai generatiivisesta tekoälystä, joilla on

aineistoa käännetty. Koska tutkimuksessa ei vertailtu esimerkiksi kahden eri tekstilajin virheanalyysin tuloksia tai muutaman eri konekääntimen käynnösten virheanalyysin tuloksia, ei tästä voida tehdä sen tarkempia päätelmiä. Jatkossa kuitenkin Koposen virheanalyysiluokittelun käyttö kokonaisuudessaan on tärkeää, sillä sen avulla voidaan kuvailla virheiden laatua hyvinkin tarkasti ja toisaalta aiempiin tutkimuksiin verraten voidaan tehdä johtopäätelmiä kussakin tekstilajissa ja kieliparissa tarvittavista virheanalyysikategorioista.

Taulukko 5: generatiivisen tekoälyn virheet vastaajittain ja kategorioittain eriteltyinä

| <b>VIRHEANALYYSI-KATEGORIA</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b> | <b>8</b> | <b>9</b> | <b>10</b> |
|--------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| POISJÄTETTY KÄSITE             | 2        | 5        | 4        |          | 1        | 5        | 5        | 2        | 4        | 10        |
| LISÄTTY KÄSITE                 | 1        |          | 4        |          |          | 2        |          | 2        |          |           |
| KÄÄNTÄMÄTÖN KÄSITE             |          |          |          |          |          |          |          |          |          |           |
| VÄÄRINKÄÄNNETTY KÄSITE         | 29       | 19       | 16       |          | 20       | 21       | 21       | 3        | 11       | 36        |
| KORVATTU KÄSITE                | 15       | 14       | 17       |          | 11       | 16       | 11       | 9        | 6        | 23        |
| SELKEYTETTY KÄSITE             |          | 2        |          |          |          |          |          |          |          |           |
| POISJÄTETTY LAUSEENJÄSEN       | 3        | 1        | 2        |          | 2        |          | 1        |          |          | 1         |
| POISJÄTETTY MUU                | 1        |          |          |          |          |          | 1        |          |          |           |
| LISÄTTY LAUSEENJÄSEN           |          |          |          |          |          |          |          |          |          |           |
| LISÄTTY MUU                    | 3        |          |          |          |          |          |          |          |          |           |
| VÄÄRÄ LAUSEENJÄSEN             |          |          |          |          |          |          |          |          |          |           |
| VÄÄRÄ MUU                      | 15       | 13       | 11       |          | 3        | 9        | 14       | 16       | 9        | 11        |
| KORVATTU LAUSEENJÄSEN          |          |          |          |          |          |          |          |          |          |           |

Taulukosta 5 käy ilmi, että ainakin yhden vastaajan (numero 4) kohdalla on tullut esiin ilmiö, josta puhutaan Koposen ja Salmen tutkimuksessa (2015, s. 132). Tämä

ilmiö on vapaasti suomennettuna nimeltään editointiväsymys (engl. *editing fatigue*), ja pätee Salmen ja Koposen mukaan erityisesti pitkissä editointiprojekteissa. Ilmiössä on kyse siitä, että ihmisen keskittymiskyky herpaantuu esimerkiksi pitkiä konekääntimen tekemiä käännöksiä jälkieditoissa, ja siksi jälkieditoija joko sokaistuu konekääntimen tekemille virheille tai arvostelu muuttuu prosessin aikana eikä joitakin virheitä merkitä, vaikka niitä olisi aiemmin saman tekstin sisällä merkitty.

Sekä konekääntimen, että generatiivisen tekoälyn käännöksissä on vaihtelevasti käytetty termejä *rikkomus* ja *rike*, joilla ei suomen kielessä ole merkittävää semanttista eroa ja niiden erottamattomuus näkyi käännöksissä selkeinä epäloogisuuksina ja epäjohdonmukaisuutena niiden käytössä. Moni vastaajista olikin tekstissä kommentoinut tätä epäjohdonmukaisuutta ja pyrkinyt välttämään kummankaan käyttöä luoden uusia termiehdotuksia, kuten *rike* oli vaihdettu jossain kontekstissa *markkerointirikkomukseksi*. Näitä uusia termiehdotuksia ei ollut kuitenkaan systemaattisesti käytetty vastaajien vastauksissa, vaan tarkoituksena on mahdollisesti ollut ehdottaa uutta termiä, mitä voitaisiin mahdollisesti käyttää jatkossa tai koko sääntökirjan kääntämisessä.

Epäjohdonmukaisuuden lisäksi vastauksissa ilmeni myös toinen editointiväsymyksestä kielivä ilmiö, jossa molempiin käännöksiin oli ehdotettu täysin samanlaisia käännöksiä siitä huolimatta, että raakakäännökset erosivat toisistaan melko paljon. Esimerkkinä tästä ilmiöstä: vastauksessa on käytetty termiä *rike* siitä huolimatta, onko kyseessä lähtötekstin *infracation* vai *violation*, ja käännöksissä käytetyistä termeistä huolimatta. Toisaalta tämä saattaa olla merkki siitä, että vastaaja on pitänyt omaa ehdotustaan hyvänä molemmissa käännöksissä, tai on pyrkinyt yhtenäistämään termejä, joilla ei kontekstista irrotettuina ole merkittävää semanttista eroa.

Mikäli on kuitenkin kyse editointiväsymyksestä, se ilmeni tässä jälkieditointiprosessissa sillä tavalla, että virhe todettiin kummassakin tekstissä ja lähtötekstiin verraten molempiin teksteihin pätisi samanlainen käännös. Siksi on pyritty nopeuttamaan jälkieditointiprosessia kopioimalla korjaus ensimmäisestä käännöksestä toiseen.

Editointiväsmyksen takia olisi syytä pohtia, olisiko käännösten järjestystä dokumentissa pitänyt vaihdella, tässä tutkimuksissa ne olivat kaikille vastaajille samassa järjestyksessä: Phrasen konekäännös ensin ja toisena ChatGPT3.5:n käännös, jolloin editointiväsmyks on voinut vaikuttaa jälkieditointiin siten, että jälkimmäisessä käännöksessä on merkitty vähemmän virheitä, eli tässä tapauksessa ChatGPT3.5:n käännöksessä. Samoin jatkossa tulisi myös pohtia, olisiko vertailussa hyödyllisempää ottaa kaksi eri tekstipätkää käännettäväksi, tai tulisiko eri käännöstyökalulla käännettyt tekstiin jakaa kahdelle eri fokusryhmälle niin, että toinen ryhmä tutkisi kääntimen A kääntämää tekstiä ja toinen tutkisi kääntimen B kääntämää tekstiä.

Vaikka viitteelliseksi aikamääreeksi jälkieditoinnille ohjeistuksessa annettiin 20 minuuttia, sain kommentteja vastaajilta, että jälkieditointiin oli saattanut kulua jopa tunnista puoleentoista tuntia. Tämä on merkittävän pitkä aika, kun kyseessä on vapaaehtoinen jälkieditointitehtävä, ja sillä perusteella voitaisiin olettaa, että se on vastaajan keskittymiskykyä koetteleva pituus.

#### 4.1.1 KÄSITTEITÄ KOSKEVAT VIRHEET

Käsitteitä koskevat virheet tarkoittavat Koposen (2010) virheanalyysiluokittelun mukaan tekstin sisältöä koskevia virheitä, kun taas kielioppia koskevat virheet (engl. *relation errors*) nimensä mukaisesti pätevät kieleen liittyviin virheisiin. Päätin pitää Koposen tutkimuksessa esiteltyt kategorisoinnit omassa tutkimuksessani, sillä halusin ennen kaikkea erotella nämä kaksi erilaista virhetyyppiä ja käsitellä niitä omina kokonaisuuksinaan, mutta myös hyödyntää niiden sisäistä yksityiskohtaisempaa jaottelua.

Seuraavassa taulukossa käsitevirheet on jaoteltu käännöksen ja tarkemman jaottelun mukaan virhekategoriioihin.

Taulukko 6: käsitteiden virheanalyysi

| <b>KÄSITTEET</b>           | <b>CHATGPT3.5</b> | <b>PHRASEN<br/>KONEKÄÄNNIN</b> |
|----------------------------|-------------------|--------------------------------|
| POISJÄTETTY KÄSITE         | 38                | 32                             |
| LISÄTTY KÄSITE             | 9                 | 2                              |
| KÄÄNTÄMÄTÖN KÄSITE         | 0                 | 4                              |
| VÄÄRIN KÄÄNNETTY<br>KÄSITE | 176               | 246                            |
| KORVATTU KÄSITE            | 122               | 139                            |
| SELKEYTETTY KÄSITE         | 2                 | 1                              |
| <b>YHT.</b>                | <b>347</b>        | <b>424</b>                     |

Taulukon 6 mukaan kaikki käsitevirheitä koskevat luokittelut ovat tässä tutkimuksessa osoittautuneet tarpeellisiksi, sillä jokaisesta luokittelusta löytyy ainakin yksi virhe. Väärin käännetty käsite -kategoria on taulukon mukaan lukumäärältään suurin, ja sitä selittäisin osaltaan sillä tavalla, että siihen kategoriaan olen luokitellut korjaamattomat, mutta virheiksi luokitellut vastaukset teksteistä. Näiden osuus kaikista kategorian virheistä ei ole suuri, ja siihen törmäsin noin kymmenen kertaa aineistoa läpi käydessäni.

Toiseksi lukumäärältään suurimpana virhekategoriassa on korvatut käsitteet, eli sisällölliset virheet, jotka on korjattu käännoksissä, mutta korjaus ei ole muuttanut kyseisen kohdan tai lauseen merkitystä. Tämä on sellainen kategoria, jonka takia jouduin palaamaan uudelleen virhekategoriassoinnin ääreen ja luokittelemaan uudelleen konekääntimen ja generatiivisen tekoälyn virheitä. En ollut alun perin luokitellut tähän kyseiseen kategoriaan ollenkaan virheitä, mutta Koposen (2010) virheluokittelun perusteita konsultoituani totesin uudelleenkategorisoinnin olevan tarpeellista.

Tämän taulukon perusteella voitaisiin väittää, että raakakäännös voisi johtaa lukijaa harhaan, sillä noin puolet ChatGPT3.5:n sisältöä koskevista virheistä luokittelin vääriksi käännoksiksi ja yli puolet Phrasen konekäännöksen virheistä (246/424) on luokiteltu vääriksi käännoksiksi. Tämän lisäksi, noin kymmenen prosenttia kummankin virheistä luokittelin poisjätöksi, eli asioiksi, jotka on raakakäännöksissä

jätetty pois, mutta korjauksissa on lisätty jotakin asioita käännökseen. Tämän perusteella voitaisiin todeta, että käännökset ovat sisällöllisesti vaatineet oikaisun lisäksi myös eksplikointia ja asioiden selventämistä lukijalle.

#### 4.1.2 TERMISTÖÄ KOSKEVAT VIRHEET

Palasin analyysivaiheessa laskentataulukoihin ja virheanalyysiin saadakseni lisää tietoa termistöön liittyvistä virheistä. Alkuperäisessä luokittelussa ei ole erikseen tarkennettu termejä koskevia virheitä, mutta koska Alasalmi (2014, s. 83–98) on pro gradu -tutkielmassaan tarkastellut tarkemmin termejä urheilun kääntämisessä, koin tässäkin tutkimuksessa termien erittelyn muusta tutkittavasta aineistosta tärkeäksi. Alasalmen tutkimus, kuten tämäkin tutkimus, käsittelevät erikoisalakääntämistä, jossa yleensä sanaston ja termien merkitys on valtava. Koposen artikkelissa (2010), jossa virheanalyysi on esitelty, on käsitelty uutistekstejä, joissa ei tyypillisesti termejä ole niiden ymmärrettävyyden ja selkeyden maksimoimisen vuoksi. Tämän vuoksi alkuperäisessä virhekategoriisoinnissa ei ole osattu termejä omana ryhmänään tai kategoriana huomioon.

Taulukossa 7 on tietoa termien osuudesta sekä generatiivisen tekoälyn käännökseen virheissä, että konekääntimen käännökseen virheissä. Huomionarvoista on se, että termit luonnollisesti jakautuvat vain sisältöä koskevien virheiden alle, eikä niitä löydy kieltä koskevien virheiden kategorioista sen takia, että termejä koskevat virheet välittömästi muuttavat tekstin sisältöä ja merkitystä. Terminologia on yksi olennaisimmista semantiikan osa-alueista ja Vehmas-Lehdon mukaan ”Termit ovat erikoiskielisen tekstin avain, sillä tekstin kielenulkoisen informaatio keskittyy niihin.” (2010, s. 362) Kaikki käsitteitä koskevat virheet eivät kuitenkaan ole termistöön liittyviä virheitä, sillä esimerkiksi tietyt tyyliä määrittävät sanavalinnat, jotka olen luokitellut käsitevirheiksi, eivät ole termejä vaan yleiskielen sanoja. Tästä esimerkkeinä:

Phrasen käännös:

Esimerkki 10: ”*puolustavan pelaajan jalkojen välinen linja tulee yhden kiekon halkaisijan päähän heittäjän kääntöpisteestä*” (Liite 2.)

ja esimerkki halkaisijan muokkauksesta, joka esiintyi useammassa eri vastauksessa:

Esimerkki 11: ”*puolustavan pelaajan jalkojen välinen linja tulee yhden kiekon **mitan** päähän heittäjän tukipisteestä*”

Tässä esimerkissä on konekäännöksen *halkaisija* vaihdettu vastauksessa *mittaan*, joka ilmaisee samaa asiaa, mutta on vastaajan mielestä mahdollisesti ollut tyyllisesti sopivampi tähän kontekstiin.

ChatGPT3.5:n käännös:

Esimerkki 12: ”*Markkeri ei saa jatkaa laskentaa ennen kuin kaikki **laittomat asennot** on korjattu.*” (Liite 3.)

esimerkkikorjaus:

Esimerkki 13: ”*Markkeri ei saa jatkaa laskentaa ennen kuin kaikki **sääntöjen vastaiset sijoittumiset** on korjattu.*”

Tässäkin esimerkissä, kuten edellisessä, on korjattu virhe, joka ei ole harhaanjohtava, vaan enemmänkin tyylivirhe, joka on mahdollisesti vastaajan mukaan ollut parempi kuvailemaan tapahtunutta. Kun puhutaan urheilusäännöistä, *laiton* ja *laittomuus* kuulostavat jokseenkin vahvoilta ilmauksilta, kun taas *sääntöjen vastainen* kuvaa paremmin sitä, miksi sijoittautuminen tai asento on väärin.

Esimerkkien 10 ja 12 lihavoidut ilmaukset olen luokitellut aineistossa käsitevirheinä, mutten termivirheinä, sillä nämä ilmaukset eivät ole olleet osana ultimatien termistöä, jonka olen luonut kääntämistä varten (Liite 4.).

Taulukko 7: termien osuus käsitteiden virheluokittelusta

| <b>VIRHEEN SYY</b> | <b>CHATGPT3.5</b> | <b>PHRASEN<br/>KONEKÄÄNNIN</b> |
|--------------------|-------------------|--------------------------------|
| POISJÄTETTY        | 1                 | 0                              |
| KÄÄNTÄMÄTÖN        | 0                 | 8                              |
| VÄÄRIN KÄÄNNETTY   | 107               | 135                            |
| KORVATTU           | 52                | 88                             |

Tässä kohtaa lienee kerrata konekääntimellä ja generatiivisella tekoälyllä kääntämisen ero. Käännösprosessissa kävi ilmi, että konekäännintä ei voi samalla tavalla pakottaa käännöksessä omaksumaan tiettyjä termejä, kun taas generatiivisen tekoälyn voi ja generatiivinen tekoäly tunnollisesti käytti näitä termejä käännöksessä. Siksi generatiivisen tekoälyn ja konekääntimen käännöksissään käyttämät termit eroavat niin paljon toisistaan, ja konekäännöksessä termejä koskevat virheet ovat vastaajilta keränneet suuremman osuuden kaikista virheistä kuin generatiivisen tekoälyn käännöksessä.

Generatiivinen tekoäly toimii ihmisen antamien ohjeiden perusteella ja pyrkii noudattamaan promptia parhaan mahdollisen osaamisensa mukaan, siksi suositeltuja termejä ilmeni GPT3.5:n käännöksessä enemmän kuin Phrasen käännöksessä. Phrasen CAT-työkalussa kääntäjä sen sijaan valitsee konekääntimen, käännösmuistin ja termipankin ehdotusten väliltä sopivimman käännöksen ja tutkimuskysymyksen vuoksi valitsin aina konekääntimen ehdotuksen. Tällöin termipankin ehdotus jäi käännöksestä pois ja saadakseni käännöksen raakaversioon, en muokannut konekäännöstä edes termipankin ehdotuksen mukaiseksi. Tämän ongelman olisi voinut mahdollisesti välttää valitsemalla toisen konekääntimen tai CAT-työkalun, mutta Phrase valikoitui tutkimukseen sen saavutettavuuden vuoksi.

Kuten taulukosta 7 käy ilmi, Phrasen konekäännin teki noin 25 prosenttia enemmän termeihin liittyviä virheitä, kuin GPT3.5. Tämä selittyy osittain sillä, että Phrase ei ole voinut hyödyntää sille syötettyä sanastoa samalla tavalla kuin GPT3.5, mutta sen selityksen perusteella GPT3.5:n termeihin liittyviä virheitä ei olisi pitänyt löytyä aineistosta ollenkaan. Vastausten perusteella niitä kuitenkin löytyi ja esimerkit näistä selityksineen ovat myöhemmin tässä alaluvussa.

Alasalmi (2014, s. 83) on tutkimuksessaan todennut termien liittyvän 42 prosenttiin kaikista tutkimuksessaan löytämistään urheilun kääntämiseen liittyvistä käännösongelmista. Tässä tutkimuksessa termeihin liittyvien virheiden osuus kaikista sisältöä koskevista virheistä on 46 prosenttia GPT3.5:n käännöksessä ja 54 prosenttia Phrasen käännöksessä. Näiden tulosten perusteella voidaan siis tehdä johtopäätös siitä, että urheilusääntöjen kääntämisessä sisällölliset virheet ilmenevät pääosin termien virheellisyytenä ja Alasalmi onkin tämän ilmiön tutkimuksessaan jo nimennyt ”urheilun erikoisalakääntämiselle tyypilliseksi ongelmaksi” (engl. *sport-specific translation problem*) (mts. 83).

Alasalmi tutkimuksessaan totesi, että termien kääntäminen on erityisen hankalaa siksi, että termistöä ei ole suomeksi (mts. 83). Tämä kyseinen ongelma ei päde ultimaten sääntöihin kokonaisuudessaan, sillä jotkin tietyt termit ovat jo ennalta olemassa, mutta niitä ei välttämättä käytetä käytännössä. Tämä saattaa olla yksi syy siihen, miksi termejä muokattiin kummassakin käännöksessä. Esimerkkinä ChatGPT3.5:n käännöksestä, jossa käytettiin aiemman sääntöversion mukaisesti termiä *Haarukointi*. Aiemman sääntöversion termit ja säännöt on hyväksytty vuonna 2017 kansallisessa lajiliitossa, mutta käännöksen *Haarukointi* oli kuitenkin useamman vastaajan toimesta muutettu *Liian lähellä*, mikä puolestaan kielii siitä, että hyväksytyt termit eivät ole käytännön käytössä, vaan on valittu joku tapahtumaa yleisemmin kuvaava ilmaus.

Vaikka kaikkia ultimaten termejä tarvinnut luoda tätä projektia tai uusinta sääntöversiota varten tyhjästä, on terminologista työtä kuitenkin jossain määrin tehtävä aina, kun säännöistä julkaistaan uusi versio, jossa täsmennetään ja päivitetään vanhaa tietoa.

Alasalmi on tutkimuksessaan todennut, että termien kääntämisen hankaluus havaittiin myös jo käyttöön vakiintuneissa, englanninkielisissä termeissä, jotka olivat alkaneet muistuttaa lausumisasultaan suomenkielisiä termejä (2014, s. 83) ja ajattelin ennen tutkimusta tämän olevan ongelma myös ultimaten sääntöjä käännettäessä. Ultimatussa pelitilanteissa ja puhekielessä pitkälti hyödynnetään englanninkielisiä termejä, jotka ovat vakiintuneet käyttöön siitä huolimatta, että

niille löytyy täysin suomenkielinen vastine. Vastauksista kuitenkin kävi ilmi, että kaikki vieraskieliset termit on pyritty korjaamaan suomenkielisiksi, kuten konekääntimen kääntämättä jättämä *Straddle*, joka on kahta vastausta lukuun ottamatta pyritty kääntämään etsimällä tai luomalla sille suomenkielisen termivastineen. Samoin käännettyjä termejä ei ole pyritty vastauksissa vaihtamaan englanninkielisiksi, vaan niiden ollessa epäsoivia, on tilalle pyritty ehdottamaan suomenkielistä uutta termivastinetta, kuten *Vision* on generatiivisen tekoälyn käännoksessä *Näkö* ja konekäännöksessä *Näkeminen*, ja niille on vastaajien toimesta ehdotettu uusiksi termivastineiksi esimerkiksi *Näkyvyys* ja *näkökenttä*.

Alasalmi pyrki tutkimuksessaan välttämään lähdekielisten termien viljelyä luomalla uusia termejä, jotka muistuttavat ulkoasultaan lähdekieltä, mutta ovat kieliopillisesti täysin suomenkielisiä termejä, kuten: ”*huukkeri, flänkkeri and proppi*” (2014, s. 84). Harkitsin tällaista terminmuodostusmenetelmää, sillä tämän kaltaiset termit ovat osittain käytössä lajin puhekielessä. Tuntui kuitenkin luontevammalta pyrkiä luomaan täysin suomenkielistä termistöä, ilman sen suurempia vaikutteita lähdekielen termeistä, jotta termin tarkoite olisi paremmin pääteltävissä pelkästään termistä, ja tätä näkökulmaa myös vastaukset vaikuttavat edustavan (ks. aiemmat esimerkit *Straddle* ja *Vision*).

Termistön luonnissa nousi haasteena myös esiin termien ”suuhunsopivuus”, eli kuinka helppoa termien käyttö pelitilanteissa olisi ja että termi pystyisi antamaan tarpeeksi informaatiota ilman esimerkiksi useamman sanan lausetta, vaan tarkoitteen pystyisi välittämään vain yhdellä käsitteellä. Tähän pyrin termistön luontiprosessissa ja vastaukset heijastelivat samanlaista ideaa, eikä niissä pyritty esimerkiksi selittämään termejä, vaan pyrittiin löytämään mahdollisimman kompakti vastine. Tästä hyvä esimerkki on kummastakin käännoksestä löytyvä *liikaa pelaajia* tai *kaksoisjoukkue*-termit. *Double teaming* -huutoa käytetään, kun heittäjää puolustaa kolmen metrin säteellä enemmän kuin yksi pelaaja, ja vain yksi pelaaja, markkeri, on sallittu sääntöjen mukaan tämän kolmen metrin säteelle. Kun katsotaan alkuperäistekstiä, on *kaksoisjoukkue*-termi hyvin suora käännosvastine, mutta toisaalta termin tarkoitteeseen nähden hieman harhaanjohtava ja luo kuvitelman siitä, että tilanteesta olisivat vastuussa molemmat pelaavat joukkueet, vaikka todellisuudessa on kyse pelkästään puolustuksen rikkomuksesta. Tähän termiin

ehdotettiin useassa vastauksessa vastinetta *tuplaus*, joka mielestäni kuvaa tarkoitetta hyvin, ei aiheuta väärinymmärryksiä, on kompakti ja lähtötekstille ulkoasultaan uskollinen.

Termeille on vastauksissa useita erilaisia ehdotuksia, joita voidaan selittää näkemyseroilla, sääntöjen tulkintaeroilla, kielen osaamisen tasoeroilla, lajista kertyneellä kokemuksella tai sillä, että termeille tai säännöille ei ole aiemmin juuri ollut suomenkielisiä vastineita.

#### 4.1.3 KIELIOPPIA KOSKEVAT VIRHEET

Aiemmin analyysiosiossa mainitsin, että jouduin vielä analyysivaiheessa palaamaan virheluokitteluun ja luokittelemaan virheitä uudelleen, mutta se ei aiheuttanut muutosta seuraavan taulukon tietoihin, eli kielioppia koskeviin virheisiin.

Taulukko 8: kieliopin virheanalyysi

| <b>KIELIOPPIVIRHEET</b>     | <b>CHATGPT3.5</b> | <b>PHRASEN<br/>KONEKÄÄNNIN</b> |
|-----------------------------|-------------------|--------------------------------|
| POISJÄTETTY<br>LAUSEENJÄSEN | 10                | 4                              |
| POISJÄTETTY MUU             | 2                 | 0                              |
| LISÄTTY LAUSEENJÄSEN        | 0                 | 0                              |
| LISÄTTY MUU                 | 3                 | 1                              |
| VÄÄRÄ LAUSEENJÄSEN          | 0                 | 1                              |
| VÄÄRÄ MUU                   | 101               | 42                             |
| KORVATTU<br>LAUSEENJÄSEN    | 0                 | 0                              |
| <b>YHT.</b>                 | <b>116</b>        | <b>48</b>                      |

Taulukon 8 mukaan kielioppia ja kieltä koskevat virheet vastaavat lukumäärältään GPT3.5:n kohdalla noin viidesosaa kaikista sen tekemistä virheistä ja Phrasen kohdalla noin kymmenesosaa kaikista sen tekemistä virheistä. Kun verrataan GPT3.5:n ja Phrasen kielioppia koskevia virheiden lukumääriä keskenään, käy ilmi, että näiden kahden välinen hierarkia on kääntynyt pääläelleen 4.1.1 osiossa esiteltyihin tuloksiin nähden ja GPT3.5 onkin tässä näkökulmassa altavastaajana. Seuraavaksi esimerkkejä taulukon 8 kategorioista:

Esimerkki 14: kieliopilliset virheet ChatGPT3.5:n käännöksessä

*”puolustavan pelaajan jalkojen väliin tulee **yksi** kiekon halkaisijan **etäisyydellä** heittäjän tukipisteestä”* (Liite 3.)

Esimerkki 15: korjausehdotus vastauksesta esimerkkiin 14

*”puolustavan pelaajan jalkojen välinen linja tulee kiekon halkaisijan **etäisyydelle** heittäjän tukipisteestä”*

Esimerkeissä 14 ja 15 esimerkit ”väärä muu” -kategorian virheestä, ja kuinka se on vastauksessa pyritty korjaamaan. Kyseessä on siis väärässä sijamuodossa oleva sana. Esimerkissä 14 on myös muita kieliopillisiä virheitä, kuten sana *yksi* on väärässä sijamuodossa, mutta se on esimerkissä 15 poistettu kokonaan lauseesta, ja siksi se on luokiteltu ”lisätty muu” -kategoriaan. Tämä sana on osa lähtötekstiä (esimerkki 16), mutta käännös on ymmärrettävä ilman sitä, minkä vuoksi se voidaan luokitella lisäyksenä.

Esimerkki 16: *”a line between a defensive player’s feet comes within one disc diameter of the thrower’s pivot point”* (Liite 1.)

Esimerkki 17: ChatGPT3.5:n käännös

*”Jos merkkaurikkeesta tai merkkaurikkomuksesta huudetaan, ja heittäjä yrittää myös syöttää ennen,”* (Liite 3.)

Esimerkki 18: korjausehdotus esimerkkiin 17

*”Jos markkerointirike tai -virhe huudetaan, ja heittäjä yrittää myös syöttää **sitä** ennen,”*

Esimerkissä 18 on esimerkki ”poisjätetty muu” -kategoriasta. Tässä korjausehdotuksessa raakakäännöstä täydentämään lisätty tarkentava demonstratiivipronomini, jossa viitataan edellisessä lauseessa mainittuun *rikkeeseen* tai *virheeseen*. Tämä *sitä*-sana ei ole siis läsnä raakakäännöksessä (esimerkki 17).

Kun tarkastellaan esimerkkiä 16 tarkemmin käännöksen (esimerkki 14) ja korjausehdotuksen (esimerkki 15) kanssa, käy ilmi vielä yksi esimerkki ”poisjätetty lauseenjäsen” -kategoriasta. Lähtötekstissä (esimerkki 16) on sana *a line*, kun taas raakakäännöksestä (esimerkki 14) tämä puuttuu kokonaan, joka aiheuttaa osittain lauseen kielellisen kömpelyyden. Tämä on korjattu vastauksessa (esimerkki 15) lisäämällä muokkaamalla alkuperäistekstin sijapäätteitä ja lisäämällä sana *linja*.

Taulukon 8 mukaan siis GPT3.5 on merkittävästi huonompi pitämään hyvästä kielestä kiinni ja tämä korostuu etenkin virheanalyysin kategoriassa ”väärä muu”, joka tarkoittaa kaikkia niitä virheitä, jotka eivät muihin virhekategorioidiin sopineet, mutta kyseessä on kuitenkin kieliopillinen virhe. Tämä on toisaalta hieman yllättävä tulos, sillä yleisesti generatiivista tekoälyä pidetään jopa kielenhuollossa ansioituneena keksintönä (Siu, 2023, s. 8), jolla luodaan sähköpostiviestejä, julkaisuja ja oikoluetaan omia tuotoksia. Ehkä suomen kieli luo otolliset olosuhteet virheille, ja sen takia generatiivisen tekoälyn huono menestys korostuu tämän taulukon tuloksissa.

Taulukosta 8 on myös huomionarvoista se, että kaikki sen virhekategorioidit eivät ole tulleet käytetyiksi. Tämä saattaa johtua siitä, että tekstilaji ei anna mahdollisuutta sellaisiin virheisiin, editointiväsymyksestä, tai allekirjoittaneen osaamattomuudesta hyödyntää kaikkia kategorioita analyysissä. Näitä tyhjäksi jääneitä kategorioita on kuitenkin tässä tutkimuksessa vain kaksi, joten voitaisiin olettaa, että mahdollisesti nämä kategoriat eivät sovellu tämän tekstilajin analyysiin. Suurinta osaa kategorioista on käytetty analyysissä, joten on myös perusteltua hyödyntää tällaista tarkkaa virhekategoriointia tulevissa, samansuuntaisissa tutkimuksissa.

## 4.2 GENERATIIVISEN TEKOÄLYN JA KONEKÄÄNTIMEN EROISTA

Vaikka generatiivinen tekoäly ja konekäännin hyödyntävät samaa toimintamallia, neuroverkkoja, on niiden treenidatassa eroja, jotka aiheuttavat myös eroja käännösten välillä. Generatiivinen tekoäly on treenattu toimittamaan myös monia muita tehtäviä kuin kääntämistä, kuten erilaisten sisältöjen luomista, kun taas konekäännin on viilattu tuottamaan ainoastaan käännöksiä neuroverkkojen avulla.

Kuten jo teoreettisessa viitekehyksessä kävi ilmi, on konekäännöksen jälkieditointia tutkittu paljon. Generatiivisen tekoälyn tuotoksia ei ole niinkään vielä edes hyödynnetty käännösosalalla, saati sitten tutkittu samanlaisesta näkökulmasta. Voidaan silti tämän tutkimuksen perusteella todeta, että generatiivisen tekoälyn tuotoksia voidaan myös jälkieditoimalla hioa luettavaksi, vaikka se ei niinkään onnistu käännöstyökalun tai -ohjelman avulla. Generatiivisella tekoälyllä käännettäessä joudutaan turvautumaan uudenlaisiin keinoihin. Tekoälyä voidaan ohjastaa muokkaamaan jotakin asiaa erinäisten käskyjen avulla, tai toisaalta uudelleenpromptaamalla koko tehtävänannon. Toistaiseksi generatiivisessa tekoälyssä on se huono puoli, ettei se pysty täysin toistamaan tekemäänsä kahdesta täysin samanlaisesta promptista huolimatta ja on siksi hieman epäluotettava. Samoin se saattaa lisäohjeiden kanssa muokata sellaisia asioita tekstissä, jotka olivat jo hyviä ja julkaisukelpoisia, eikä niihin olisi annettu korjauspyyntöä. Tulin kosketuksiin tämän asian kanssa, kun harjoittelin ChatGPT3.5:n käyttöä tutkimusta varten.

Jos generatiivisen tekoälyn käännöstä ei halua promptien avulla lähteä ”jälkieditoimaan”, on sitä jälkieditoitava tekstimuokkaustyökalussa, kuten on tämän tutkimuksen aineistonkeruussa tehty, tai erikseen syöttämällä sen käännöstyökaluun, jossa jälkieditoida käännöstä. Tämä on hieman työläämpää ammattilaiskääntäjälle, joka on todennäköisesti tottunut hyödyntämään käännöstyökaluja, jotka ovat ammattitarkoitukseen luotuja välineitä ja siksi käteviä. Toisaalta generatiivinen tekoäly on hyvin helppokäyttöinen ja saavutettava työväline maallikkokääntäjälle, joka pyrkii saavuttamaan nopeasti melko hyvää laatua.

Tutkimuskysymykseen 2. voidaan vastata taulukon 3 tietojen ja vastaajilta saamieni kommenttien perusteella, joiden mukaan jälkieditointi vie merkittävästi aikaa ja kognitiivista työpanosta, vaikka kyseessä onkin vain noin sivun pituinen teksti. Taulukon 3 mukaan sekä ChatGPT3.5, että Phrasen konekäännin ovat tehneet keskimäärin yhdestä kahteen virhettä per virke, joka ei sinänsä kuulosta paljolta, jos vaikka kyseessä olisi pienet kieliopilliset virheet. Mutta koska suurin osa virheistä kuuluu käsitteitä koskeviin virheisiin, voidaan ymmärtää jälkieditointiin kuluvan erityisen paljon aikaa, kun jälkieditoija joutuu miettimään esimerkiksi termistöä koskevia ongelmia. Vaikka tätä eroa työpanoksessa kieliopillisten ja käsitevirheiden käsittelyn välillä ei ole otettu huomioon lähdekirjallisuudessa, on selvää, että

esimerkiksi sijapäättöjen vaihtaminen oikeaan vie keskimäärin vähemmän jälkieditoijan aikaa kuin esimerkiksi termin oikeellisuuden tarkistaminen ja on kognitiivisesti vähemmän kuormittava tehtävä.

Vastatakseni ensimmäiseen tutkimuskysymykseen; vastauksissa saatiin aikaan luettavia ja ymmärrettäviä versioita kummastakin käännöksestä jälkieditoinnin seurauksena. Tässä tutkimuksessa havaittiin, että ne alueet, missä toinen näistä käännösteknologisista apuvälineistä menestyy, toinen kompastelee, ja päinvastoin. Erot kokonaistuloksissa näiden välineiden välillä (ks. taulukko 3) ovat kuitenkin hyvin pieniä ja päätöstä siitä, että kumpaa tällaisessa projektissa hyödyntäisi on mahdotonta tehdä näiden tilastojen perusteella ja on katsottava muita tekijöitä.

Kolmanteen ja viimeiseen tutkimuskysymykseen on siis vastattava tarkastelemalla kielenulkoisia tekijöitä ja metatyötä; vaikka ChatGPT3.5:n raakakäännös on tilastojen valossa hieman Phrasen konekäännöstä parempi, on se kokonaisuudessa selkeästi parempi vaihtoehto näistä kahdesta, kun otetaan huomioon, että se on saavutettava myös kansallisille lajiliitoille, joilla ei ole mahdollisuutta maksaa käännöstyöstä. ChatGPT3.5 pystyy myös hyödyntämään sille annettua termistöä ilman, että sen raakaversioon tarvitsee kajota, toisin kuin käännöstyökalussa luotu konekäännös. Selvitettäväksi jää, pystyykö generatiivisen tekoälyn käännös jälkieditoituna vastaamaan täysin ihmiskääntäjän manuaalisesti kääntämää tekstiä. Samoin tämän tutkimuksen rajauksen ulkopuolelle jää selvittää, onko käännetyillä ultimatien säännöillä lukijoita, ja onko siten urheilusääntöjen kääntäminen siten kannattavaa toimintaa.

Seuraavassa luvussa vertaan tutkimuksen löydöksiä alan aiempaan tutkimukseen, miten tämä tutkimus asettuu erikoisalojen kääntämisen tutkimukseen ja tarkastelen tutkimuksen tuloksia kriittisesti.

## 5 POHDINTA

Alasalmen pro gradu -tutkielmassaan tekemä urheilusääntöjen kääntämisen osittainen viitekehys pitää paikkaansa myös tämän tutkimuksen aiheen kohdalla. Vaikka tässä tutkimuksessa ei pyritty luomaan yhtä tarkkaa kuvausta urheilusääntöjen erikoisalakääntämisen konventioista, voidaan Alasalmen toteamien konventioiden pitkälti pätevän myös tämän tutkimuksen pienlajin sääntöjen kääntämisessä.

Teoreettisessa viitekehyksessä ja metodia käsittelevässä luvussa käsittelin jälkieditointia tutkimustapana konekääntimen laadun arvioinnissa, ja voin todeta sen toimineen hyvin tämän tutkimuksen tutkimuskysymysten vastauksia analysoidessa. Jälkieditointi toimi myös mielestäni hyvin tämän tutkimuksen toteuttamiseen, sillä vastaajien ei tässä tutkimuksessa tarvinnut osata kuvailla aineistoa tai sen virheitä, vaan pelkästään pyrkiä korjaamaan niitä, jolloin virheiden analysointi jäi vain yhden ihmisen harteille kuormittamatta jokaista vastaajaa. Uskon, että jälkieditoinnin käyttö aineistonkeruumenetelmänä sai myös osallistettua enemmän vastaajia, sillä tutkimukseen vastaaminen oli sen myötä helpompaa.

Toisin kuin tämän tutkimuksen teoreettisessa viitekehyksessä todettiin, generatiivinen tekoäly suoriutuu suhteellisen hyvin tämän erikoisalan käännöstehtävästä. Lähdekirjallisuudessa todetaan, että generatiivinen tekoäly tuskin kykenee erikoisalojen kääntämiseen sen yleisluontoisen treenidatan takia (Brown ym., 2020, s. 3; Siu, 2023, s. 10), mutta tässä tutkimuksessa tämä käsitys on todettu vääräksi. Kun generatiiviselle tekoälylle tarjotaan valmis termistö, mitä hyödyntää, se osaa kääntää hyvin ja sen tekemät virheet korostuvat etenkin kieliopillisten virheiden kategoriassa, jotka vievät vähemmän aikaa korjata kuin sisältöä tai termejä koskevat virheet. Tämä ristiriitaisuus teoreettisen viitekehysten kanssa osoittaa sen, että generatiivista tekoälyä ja sen hyödyntämistä kääntämisessä on tutkittu hyvin vähän ja yleistäviä johtopäätöksiä on tästä ilmiöstä mahdotonta tehdä toistaiseksi. Teoreettisen viitekehysten generatiivista tekoälyä käsittelevässä alaluvussa hyvin keskeisenä tutkimusaiheena ovat promptit, joita käytetään kääntämisen tehtävänannossa tekoälylle. En kuitenkaan koe hyötyneeni prompteista niin paljon tässä tutkimuksessa, sillä koin saavani hyvin yksinkertaisella promptaamisella hyvän

raakakäännöksen ulos. Samoin prompteja koskeva tieto on pitkälti rajautunut tutkimuskysymyksiä takia tutkimuksen ulkopuolelle, sillä en enää raakaversioiden saamisen jälkeen halunnut antaa tekoälylle lisäohjeita, jotta tutkimusasetelma säilyisi suunnitellusti raakakäännösten tutkimisessa. Lisäohjeiden promptaaminen voitaisiin tietyllä tavalla laskea jonkinasteiseksi jälkieditoinniksi, ja siksi olisi sotkenut tätä tutkimusasetelmaa.

Käsittelin teoreettisessa viitekehyksessä ja tutkimusmenetelmää kuvaavassa osiossa konekääntimen laadun arviointia ja siihen hyödynnettyä virheanalyysimenetelmää. Päädyin tässä tutkimuksessa käyttämään Koposen tarkempaa virheanalyysikategorisointia, joka osoittautui perusteelliseksi, sillä sen ansiosta tarkempi kuvaus tässä tekstilajissa tyypillisistä virheistä oli mahdollinen. Myös muuten vähäiset erot generatiivisen tekoälyn ja konekäännöksen välillä on helpompaa tarkastella ja löytää, kun virhekategorioiden tarkemmat.

Tarkat virhekategorioiden eivät kuitenkaan konekääntimen laadun tutkimuksessa ole välttämättömiä ja pitkälle voidaan päästä myös yksinkertaisella oikea–väärä- ja sisältö–kielioppi-asteikoilla, mutta oikean tarkkuuden löytäminen riippuu ennen kaikkea tutkimuskysymyksistä ja -näkökulmasta. Tässä tutkimuksessa ei pyritty sen suuremmin haastattelemaan vastaajia, mutta jos tutkimuksessa olisi esimerkiksi mitattu jälkieditointiin kulutettua aikaa tai käytetty katseenseurantamenetelmää, yksinkertaisempi virhekategorioiden olisi voinut olla paikallaan.

Tarkka tutkimusmenetelmän valinta on tämän kaltaisessa tutkimuksessa erityisen tärkeää, jotta voidaan välttyä editointiväsymykseltä. Tässä tutkimuksessa sitä ei onnistuttu täysin välttämään, ja siksi olisikin aiheellista pohtia tutkittavien tekstipätkien pituutta tai olisiko käännösten mahdollista olla eri tekstipätkistä. Syytä olisi myös pohtia, onko jokaisen vastaajan päästävä tekemään rinnakkaistekstien analyysiä, vai voisiko pelkästään toisen käännöksen tarkastelu riittää, jolloin vastaajat tulisi jakaa kahteen eri fokusryhmään. Mielestäni tämän tutkimuksen aiheeseen ei sovi yksikielinen jälkieditointi, eli esimerkiksi pelkästään yhden käännöksen tarkastelu ilman lähdetekstiä, jota Koponen ja Salmi (2015) ovat tutkimuksessaan hyödyntäneet, sillä tekstin aihe voi olla haastava jopa lajin asiantuntijalle. Mahdollisesti lakia ja urheilusääntöjä käsittelevät tekstit voitaisiin

pyrkii tulevassa tutkimuksessa rajaamaan tämän tutkimusmenetelmän ulkopuolelle niiden haastavan sisällön vuoksi.

Tutkimusta rajoitti ja samalla tutkittavan aihealueen rajauksen määrittä saavutettavuus. Tutkimukseen käytiin ja tutkimuksessa tutkittiin niitä kääntämisen työvälineitä, joihin minulla oli yliopiston kautta lisensoitu pääsy ja käyttöoikeus, tai työkalu oli ilmainen. Maksullisuuden takia esimerkiksi ChatGPT:n uusin versio on rajautunut tutkimuksen ulkopuolelle, vaikka se varmasti olisi koulutetumpi kuin tässä tutkimuksessa käytetty versio, eikä tutkimus siksi tarjoa täysin todenmukaista dataa sen kykeneväisyydestä tai laadusta tällaisessa tehtävässä. Toisaalta tämä tutkimus antaa todenmukaisen kuvan siitä, minkälaista käännöslaatu voi olla saavutettavissa esimerkiksi sellaisille pienurheilulajien kansallisille lajiliitoille, joilla ei ole budjettia käytettävissä sääntöjen kääntämiseen ja sen avulla edistää lajin saavutettavuutta uusille harrastajille.

Koska kyseessä on tapaustutkimus, jossa tutkitaan yhden lajin sääntöjen kääntämistä, ei tämän tutkimuksen perusteella voida tehdä yleistyksiä urheilsääntöjen kääntämisestä. Kuitenkin kun tätä tutkimusta ja sen tuloksia tarkastellaan yhdessä saman erikoisalan muiden tutkimusten, kuten Alasalmen ja Uyanikin tutkimusten kanssa, voidaan tehdä jonkinlaisia päätelmiä yleisesti urheilsääntöjen kääntämiseen pätevistä konventioista.

Tutkimusta ja sen tutkimustulosten analyysiä rajoitti myös vahvasti vertaisarvioimattomiin artikkeleihin nojaava analyysi. Aiemmin tässä tutkielmassa puhuttiin siitä, että generatiivisella tekoälyllä kääntämisestä löytyy hyvin vähän tutkimusta, joka johtuu osaltaan siitä, että tämänlainen ilmiö on käännösosalalla vielä hyvin uusi. Ensimmäiset tätä aihetta käsittelevät artikkelit on julkaistu vuonna 2023, eli vain vuosi ennen tämän tutkielman julkaisua. Nämä artikkelit myös tutkivat aiempaa versiota ChatGPT:stä, eivätkä niiden tulokset siksi ole täysin rinnastettavissa tämän tutkimuksen tuloksiin.

Samoin yksi tämän tutkielman teoreettisen viitekehyksen kulmakivistä, Alasalmen pro gradu -tutkielma on vertaisarvioimaton, eikä opinnäytetyönä varsinaisesti tieteellinen julkaisu ja siksi sitä tulee tarkastella kriittisesti. Alasalmen tutkimuksessa

on kuitenkin tarkasteltu hyvin läheisesti tämän tutkielman aihetta muistuttavaa ilmiötä, jota ei ole muissa tutkimuksissa käsitelty, ja on siksi relevantti tämän tutkielman tulosten kannalta. Myös generatiivista tekoälyä ja sillä kääntämistä tutkivissa vertaisarvioimattomissa lähteissä on käsitelty tämän tutkielman aiheen kannalta tärkeää ilmiötä, ja muuta vertaisarvioitua kirjallisuutta ei tämän ilmiön osalta ole tiettävästi julkaistu. Siksi olen päättänyt sisällyttää myös vertaisarvioimatonta kirjallisuutta tutkielmani teoreettiseen viitekehykseen, sillä ilman sitä, jäisi tutkittava käännöstieteen alue hyvin irralliseksi ja tämän tutkielman löydökset hypoteettisiksi.

Tutkimustulosten analyysin pohjalta heräsi ideoita myös muista materiaalista mahdollisesti tutkittavista asioista. Olisi mielenkiintoista tutkia termistön konkreettista käyttöä tarkemmin, vaikka sitä jo hieman tässä tutkielmassa sivuttiinkin tilastojen valossa. Tulevaisuudessa voisi siis tarkastella tulevatko tällaiset luodut ja kansallisilla lajiliitoilla hyväksytyt termistöt todellisuudessa käyttöön ja missä määrin ne omaksutaan osaksi pelaajien tilannekohtaista puhekieltä. Tämän ilmiön tutkimiseen kuitenkin vaadittaisiin valmis sääntöversio ja sen käyttöönotto lajissa ennen kuin ilmiötä voitaisiin tutkia, ja tätä ei ollut mahdollisuutta tämän tutkielman aikamääreiden puitteissa toteuttaa.

Olisi myös mielenkiintoista tutkia Uyanikin esille tuomaa sosiolingvististä näkökulmaa sääntöjen kääntämiseen, eli onko sääntöjen kääntäjän taustalla todella merkitystä sääntöjen ymmärrettävyyden tai jälkieditointityöpanoksen kannalta. Uyanik toteaa tutkimuksessaan (2017, s. 109, 111–112), että hyvän urheilutekstien kääntäjän tulisi osata käännösparin kielet hyvin, mutta myös olla tietoinen lajista ja sen säännöistä voidakseen kääntää kyseisen lajin säännöt ymmärrettävästi. Tämä on mielenkiintoinen hypoteesi tulevalle tutkimukselle, joka ei myöskään mahtunut tämän tutkielman sivumäärään. Jätin myös tämän näkökulman tutkimuksen ulkopuolelle, sillä halusin minimoida henkilötietojen käsittelyn tässä tutkielmassa ja tätä näkökulmaa varten olisi pitänyt tutkia vastaajien taustoja enemmän. Pienen otoksen takia tällainen tutkimusnäkökulma olisi myös ollut tutkimuseettisesti kyseenalainen.

Analyysivaiheessa tutkimusasetelma tässä tutkielmassa meni koetukselle, kun kävi ilmi, etteivät molemmat työkalut pysty hyödyntämään termistöä samalla tavalla ja tämä aiheuttaa perustavanlaatuisen ongelman aineiston analyysiin.

Raakakäännösten tasoero kuitenkin selitettiin ja siirryin pelkästä aineiston analyysistä kokonaisvaltaisempaan analyysiin. Tutkimustuloksia siirryttiin tämän jälkeen selittämään myös kielenulkoisten tekijöiden avulla ja pyrittiin realiteettien avulla selvittämään näistä työkaluista parempi vaihtoehto. Ratkaisun kolmanteen tutkimuskysymykseen tässä tutkimuksessa lopulta teki näkökulma, eli kokonaisvaltainen käännösprosessin tarkastelu.

Sen lisäksi, että jouduin vaihtamaan näkökulmaa vastatakseni tutkimuskysymyksiin, minun piti analyysivaiheessa vielä palata aineistoon ja luokittelemaan virheet uudelleen. Huomasin analyysivaiheessa, että olin luokitellut osan virheistä väärin ja jouduin siksi tehdä virheanalyysin aineiston pohjalta uudestaan. Tämän lisäksi luokittelin virheet vielä uudestaan ja loin uuden taulukon termistöä koskevien virheiden pohjalta, ja tämän ansiosta onnistuin vetämään analyysissä paremmin omia tuloksiani ja teoreettisen viitekehityksen tutkimusta yhteen.

Lukuisista rajoitteista ja hankaluuksista huolimatta, tässä tutkimuksessa on onnistuttu vastaamaan tutkimuskysymyksiin kattavasti ja kyseenalaistamaan hypoteesiä. Tutkimuskysymykset on onnistuneesti muotoiltu tämän tutkimuksen tutkimusalueen huomioon ottaen. Tämä tutkimus on tapaustutkimuksen luonteestaan huolimatta onnistuneesti tutkinut uudenlaista alaa käännöstieteessä ja pyrkinyt luomaan viitekehystä hyvin tarkasti rajattuun tutkimusalaan. Vaikka yleistyksiä ei tämän tutkimuksen tulosten perusteella voida tehdä, on niitä mahdollisuutta hyödyntää tulevaisuudessa, kun tätä tutkimusrakoa tarkastellaan enemmän.

Tässä tutkimuksessa on tutkittu hyvin uudenlaista, lähes tutkimatonta ilmiötä, joka on seurausta vähän tutkituista, isommista ilmiöistä, kuten generatiivisella tekoälyllä kääntämistä ja urheilusääntöjen, tai yleisemmin sääntöjen kääntämistä. Tämä tutkimus osoittaa sen, että näille erikoisaloille vaaditaan lisää tutkimusta ja samalla myös kannustaa erikoisalojen ja käännöstyökalujen tutkimiseen.

## 6 JOHTOPÄÄTÖKSET

Tässä maisterintutkielmassa tutkittiin täysin uudenlaista tutkimusrakoa erikoisalakääntämisen tutkimusalalla. Tämän tutkimuksen tavoitteena oli selvittää, onko jokseenkin tutkimattoman erikoisalan, urheilusääntöjen, kääntäminen mahdollista käännösteknologisia työkaluja hyödyntäen. Tutkimuksessa pyrittiin myös vastaamaan kolmeen tutkimuskysymykseen, joissa pohdittiin käännösteknologisilla työkaluilla käännettyjen tekstien laatua, niihin kuluva jälkieditointipanosta ja näiden käännösteknologisten työkalujen keskeisimpiä, käännösten laatuun merkittävästi vaikuttavia eroavaisuuksia.

Tutkimuskysymyksiin haettiin vastauksia ensin esittelemällä tutkimuksen teoreettista viitekehystä, siirtyen tutummista ja yleisemmistä käännöstieteen tutkimuksen alueista kohti tuntemattomampia ja kapeampia tutkimusaloja. Ensimmäisenä esiteltiin lyhyesti konekääntämisen historiaa, erilaisia konekääntimiä ja neuroverkkokääntimen toimintamallia. Sitten esiteltiin konekäännöksen jälkieditoinnin tutkimusta, jossa keskeisessä osassa on Maarit Koposen tekemä tutkimus. Näiden jälkeen siirryttiin tuntemattomampiin tutkimusaloihin; ensin generatiivisella tekoälyllä kääntämistä käsitteleviin artikkeleihin ja lopuksi urheilun erikoisalakääntämiseen, josta löytyi aiempaa tutkimusta vain kahden tutkimuksen verran.

Teoreettisen viitekehysten tarkastelun jälkeen siirryttiin tutkimusmenetelmän kuvaukseen, jossa erilaisten tutkimusmenetelmien pohjalta muokkautui tähän tutkimukseen sovellettava tutkimusmenetelmä. Keskeisessä osassa tutkimusmenetelmässä toimi Koposen virheanalyysimenetelmä ja tämä valikoitui tutkimusmenetelmäksi tutkimuskysymysten perusteella. Virheanalyysin nähtiin analysoivan suhteellisen suurtakin aineistoa tehokkaasti ja loi mahdollisuuden kerätä aineiston asiantuntijoilta.

Aineistonkeruussa sovellettiin myös muutamia eri menetelmiä parhaan mahdollisen aineiston saavuttamiseksi pienellä otoksella. Vastaajia haettiin tutkimukseen sosiaalisen median kanavissa ja vastaajat ilmoittautuivat tutkimukseen internet-lomakkeella. Lomakkeen yhteydessä vastaajat saivat luettavakseen tutkimuksen

tietosuojailmoituksen ja heiltä kerättiin ilmoittautumisen yhteydessä sähköpostiosoitteet tutkimuksen toteuttamista ja yhteydenpitoa varten. Vastaajien henkilötiedot eivät ole osa tutkimuksen aineistoa.

Tutkimusmenetelmän ja aineiston kuvauksen jälkeen siirryttiin analyysiin, jossa ensimmäisenä esiteltiin virheanalyysin yleiset tulokset. Taulukosta 3 kävi ilmi, kuinka paljon virheitä kumpikin käännösteknologinen työkalu oli tehnyt, ja mitä nämä määrät olivat suhteessa virkkeiden määrään. Taulukosta 3 löytyi myös alkuperäistekstin perustietoja, kuten sanojen ja virkkeiden lukumäärä.

Analyysi eteni teoreettisen viitekehyksen tapaan yleisemmistä löydöksistä spesifimpiin, mitä pidemmälle analyysissä edettiin. Yleislaatuisten tietojen jälkeen esiteltiin konsepteja koskevat, eli tekstin sisältöä koskevat virheet, joita oli toiseen kategoriaan nähden merkittävästi enemmän. Käsitteitä koskevien virheiden jälkeen esiteltiin termistöä koskevat virheet, jotka olivat kaikki osa sisältöä koskevista virheistä, ja siksi tämä taulukko esiteltiin välittömästi käsitevirheiden jälkeen. Termistöä koskevat virheet olivat erityisen tärkeitä teoreettisen viitekehyksen tutkimusten vertailun ja analyysin kannalta, ja tässä tutkimuksessa löytyikin yhtäläisyyksiä termistön tarkastelun osalta. Näiden tarkastelun perusteella tutkimuksen hypoteesi siitä, että Phrasen konekäännin tekee vähemmän virheitä kuin generatiivinen tekoäly, osoittautui vääräksi.

Sisältöä koskevien virheiden tarkastelun jälkeen siirryttiin tarkastelemaan kielioppia koskevia virheitä. Kielioppia koskevia virheitä oli taulukoiden mukaan lukumäärältään merkittävästi vähemmän kuin sisältöä koskevia virheitä. Merkittävä löydös tässä kohtaa oli se, että Phrasen konekäännin oli tehnyt vähemmän virheitä tässä kategoriassa kuin generatiivinen tekoäly, eli hypoteesi tälle tutkimukselle ei ollut täysin väärässä.

Konekääntimen ja generatiivisen tekoälyn merkittävimmistä eroista sekä virheanalyysissä, että todellisuudessa, on käytettävyyys. Generatiivinen tekoäly pystyy suoraan, yhdellä promptilla hyödyntämään termistöä ja lähtötekstiä käännoksessään, kun taas konekäännöksen raakaversiota tulisi jälkieditoida, jotta termistö tulisi käyttöön käännoksessä, elleivät termit ole konekääntimen treenidatan kautta sille jo

tuttuja. Toisaalta konekäännöksen raakaversiota pääsee helposti muokkaamaan suoraan käännöstyökalussa, kun taas generatiivisen tekoälyn käännöstä tulee jälkieditoida tekstinmuokkaustyökalussa, jos ja kun sitä ei saa uudelleenpromptaamalla hiottua huippuunsa ja jos ei halua nähdä vaivaa siirtääkseen generatiivisen tekoälyn käännöksen käännöstyökaluun, jossa sitä voisi jälkieditoida.

Toinen merkittävistä eroista on saavutettavuus. Tämän tutkimuksen motiivina on lajinedistämistyö ja lajin saavutettavuuden lisääminen, ja tällöin myös käännöstyökalujen saavutettavuus on tärkeää. Vaikka Phrasen konekääntimen hyödyntäminen tässä projektissa oli minulle ilmaista, tarvitaan sen käyttöön kuitenkin lisenssi, joka minulla oli yliopiston alaisuudessa olemisen ansiosta. Tällaiset lisenssit eivät kuitenkaan kaikille lajiliitoille ole mahdollisia ja siksi generatiivinen tekoäly palvelee ilmaisena työvälineenä paremmin tarkoitusta.

Johtopäätöksissä todettiin siis, että generatiivinen tekoäly on tämän tapaustutkimuksen kannalta parempi työväline toteuttamaan käännös. Tutkimuksessa tullaan myös siihen johtopäätökseen, että urheilusääntöjen kääntäminen käännösteknologisilla työkaluilla on mahdollista, mutta toistaiseksi jälkieditointityö on suuri ja käännösteknologisten työkalujen hyöty verrattuna ihmiskääntäjän alusta loppuun tekemään käännökseen jää epäselväksi. Tämän tekstilajin kääntäminen käännösteknologisten apuvälineiden avulla on siis sellainen ilmiö, jota voisi myös tulevaisuudessa tutkia.

Koska kyseessä on tapaustutkimus, jossa käännetään yhden urheilulajin sääntöjä, yhdessä kieliparissa ja tietyillä käännösteknologisilla työkaluilla, ei voida tehdä yleistyksiä urheilusääntöjen kääntämisestä, tai niiden kääntämisestä käännösteknologisilla työkaluilla. Yleistysten tekemistä rajoittaa myös tutkittavien alueiden, urheilun erikoisalakääntämisen tutkimuksen ja generatiivisella tekoälyllä kääntämisen tutkimuksen tutkimattomuus ja vertaisarvioimattomuus.

Vaikka yleistyksiä ei voida tämän tutkimuksen perusteella tehdä, on tässä tutkimuksessa tehty tietoa syventäviä löydöksiä ja niiden peilaus teoreettiseen

viitekehukseen auttaa saamaan paremman ymmärryksen näistä kääntämisen tutkimuksen pienaloista.

Tulevaisuuden tutkimuksessa olisi hyvä ensinnäkin tutkia laajemmin näitä pienaloja, generatiivisella tekoälyllä kääntämistä ja urheilun erikoisalakääntämistä, kuten myös erilaisten urheilulajien kääntämistä. Mielenkiintoista olisi myös tutkia, kuinka luotu termistö tulee käyttöön urheiltavan lajin puhekielessä tai kuinka erilaiset asiantuntijaryhmät kääntävät tai jälkieditoivat urheilusääntöjä. Olisi erittäin mielenkiintoista myös tutkia Uyanikin hypoteesiä urheilusääntöjen kääntämiseen liittyen, määrittääkö kääntäjän tausta todella käännöksen laadun.

Tämä tutkimus nostaa odotuksia tämän, ja muiden erikoisalojen kääntämisen tutkimukselle ja kannustaa tutkimaan erikoisalojen kääntämistä tarkemmin. Tämä tutkimus havainnollistaa myös sen, kuinka nopeasti käänösala muuttuu ja tutkittavat alueet käänöstieteessä lisääntyvät. Tämä tutkimus osoittaa myös kääntämisen tutkimuksen tärkeyden, ja kuinka olennainen osa kääntäminen on saavutettavuutta.

## Lähteet

### Aineisto

ChatGPT3.5 (2024) '18.1. Merkkkausrikkeet'

Phrase (2024) '18.1. Rikkomusten merkitseminen:'

The World Flying Disc Federation. (2021) 'Marking Infractions', *Rules of Ultimate 2021-2024*. Colorado Springs: The World Flying Disc Federation, 12. Saatavilla osoitteessa: <https://rules.wfdf.sport> (Avattu 15.3.2024)

### Kirjallisuus

Alasalmi, T. (2014) *Problems of Translating the Laws of Rugby Union from English into Finnish*. Pro gradu -tutkielma. Jyväskylä: Jyväskylän yliopisto. Saatavilla osoitteessa: <http://urn.fi/URN:NBN:fi:juu-201411193295> (Avattu 19.4.2024)

Alvarez-Vidal, S., Badia, T. ja Oliver, A. (2020) 'Post-editing for Professional Translators: Cheer or Fear?', *Revista Tradumàtica*, numero 18. Saatavilla osoitteessa: [https://www.researchgate.net/publication/348571533\\_Post-editing\\_for\\_Professional\\_Translators\\_Cheer\\_or\\_Fear](https://www.researchgate.net/publication/348571533_Post-editing_for_Professional_Translators_Cheer_or_Fear) (Avattu 25.3.2024)

Brown, T., ym. (2020) 'Language Models are Few-Shot Learners', arXiv. Saatavilla osoitteessa: <https://doi.org/10.48550/arXiv.2005.14165> (Avattu 14.4.2024)

Carl, M. ym. (2011) 'The Process of Post-Editing: a Pilot Study', *Copenhagen studies in language*. Frederiksberg: Department of International Language Studies and Computational Linguistics Copenhagen Business School, s. 131–142. Saatavilla osoitteessa: <https://aclanthology.org/www.mt-archive.info/NLPSC-2011-Carl-1.pdf> (Avattu 4.3.2024)

Daems, J. ym. (2017) 'Identifying the Machine Translation Error Types with the Greatest Impact on Post-editing Effort' *Frontiers in Psychology*, vol. 8. Saatavilla osoitteessa: [10.3389/fpsyg.2017.01282](https://doi.org/10.3389/fpsyg.2017.01282) (Avattu 4.3.2024)

Ehrensberger-Dow, M. (2019) 'Ergonomics and the translation process', *ResearchGate*. Saatavilla osoitteessa: [https://www.researchgate.net/publication/331105807\\_Ergonomics\\_and\\_the\\_translation\\_process](https://www.researchgate.net/publication/331105807_Ergonomics_and_the_translation_process) (Avattu 25.3.2024)

Koehn, P. (2020a) 'Uses of Machine Translation', *Neural Machine Translation*, 1<sup>st</sup> Edition. Cambridge: Cambridge University Press, s. 19–28. Saatavilla

- osoitteessa: <https://doi-org.libproxy.helsinki.fi/10.1017/9781108608480>  
(Avattu 27.2.2024)
- Koehn, P. (2020b) 'History', *Neural Machine Translation*, 1<sup>st</sup> Edition. Cambridge: Cambridge University Press, s. 29–40. Saatavilla osoitteessa: <https://doi-org.libproxy.helsinki.fi/10.1017/9781108608480> (Avattu 27.2.2024)
- Koehn, P. (2020c) 'Neural Networks', *Neural Machine Translation*, 1<sup>st</sup> Edition. Cambridge: Cambridge University Press, s. 67–87. Saatavilla osoitteessa: <https://doi-org.libproxy.helsinki.fi/10.1017/9781108608480> (Avattu 27.2.2024)
- Koponen, M. (2010) 'Assessing Machine Translation Quality with Error Analysis', *MikaEL: Kääntämisen ja tulkauksen tutkimuksen symposiumin verkkojulkaisu*, vol. 4. Suomen kääntäjien ja tulkkien liitto. Saatavilla osoitteessa: [https://sktl-fi.directo.fi/@Bin/40701/Koponen\\_MikaEL2010.pdf](https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf) (Avattu 13.11.2023)
- Koponen, M. (2016) 'Is Machine Translation Post-editing Worth the Effort? A Survey of Research into Post-editing and Effort', *The Journal of Specialised Translation*, issue no 25, s. 131–148. Saatavilla osoitteessa: [https://www.researchgate.net/publication/299345698\\_Is\\_Machine\\_Translation\\_Post-editing\\_Worth\\_the\\_Effort\\_A\\_Survey\\_of\\_Research\\_into\\_Post-editing\\_and\\_Effort](https://www.researchgate.net/publication/299345698_Is_Machine_Translation_Post-editing_Worth_the_Effort_A_Survey_of_Research_into_Post-editing_and_Effort) (Avattu 4.3.2024)
- Koponen, M., Salmi, L. (2015) 'On the correctness of machine translation: A machine translation post-editing task', *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing, s. 170–199. Saatavilla osoitteessa: [https://www.academia.edu/70037232/On\\_the\\_correctness\\_of\\_machine\\_translation\\_A\\_machine\\_translation\\_post\\_editing\\_task?hb-sb-sw=48354818](https://www.academia.edu/70037232/On_the_correctness_of_machine_translation_A_machine_translation_post_editing_task?hb-sb-sw=48354818) (Avattu 29.2.2024)
- Koponen, M. ym. (2012) 'Post-editing time as a measure of cognitive effort', *Proceedings of the AMTA 2012 Workshop on Post-editing Technology and Practice*, s. 11–20. Saatavilla osoitteessa: [https://www.semanticscholar.org/paper/Post-editing-time-as-a-measure-of-cognitive-effort-Koponen-Aziz/717b0220273e8b4916d9e15573570cc3116b4627?utm\\_source=direct\\_link](https://www.semanticscholar.org/paper/Post-editing-time-as-a-measure-of-cognitive-effort-Koponen-Aziz/717b0220273e8b4916d9e15573570cc3116b4627?utm_source=direct_link) (Avattu 29.2.2024)

- Maučec, M. S., Donaj, G. (2019) 'Machine Translation and the Evaluation of Its Quality', *Recent Trends in Computational Intelligence*. IntechOpen. Saatavilla osoitteessa: [10.5772/intechopen.89063](https://doi.org/10.5772/intechopen.89063) (Avattu 6.3.2024)
- Peng, K., ym. (2023) 'Towards Making the Most of ChatGPT for Machine Translation' *Arxiv*, vol. 4. Saatavilla osoitteessa: <https://doi.org/10.48550/arXiv.2303.13780> (Avattu 23.11.2023)
- Siu, S. C. (2023) 'Revolutionizing Translation with AI: Unravelling Neural Machine Translation and Generative Pre-trained Large Language Models.' Saatavilla osoitteessa: <https://ssrn.com/abstract=4499768> (Avattu 14.4.2024)
- Salmi, L., ja Koponen, M. (2020). 'Valintaverkkoanalyysi konekäännöksen jälkieditoinnin tarkastelun apuna', *MikaEL Kääntämisen ja tulkkauksen tutkimuksen symposiumin verkkojulkaisu*, Vol. 13. Suomen kääntäjien ja tulkkien liitto, s. 58–74. Saatavilla osoitteessa: <http://hdl.handle.net/10138/320216> (Avattu 27.2.2024)
- Song, H. (2023) 'SelfSeg: A Self-supervised Sub-word Segmentation Method for Neural Machine Translation', *ACM Transactions on Asian and Low-Resource Language Information Processing*, Vol. 22. Association for Computing Machinery (ACM), s. 1–24. Saatavilla osoitteessa: <http://dx.doi.org/10.1145/3610611> (Avattu 6.3.2024)
- SLKL. (2023) *Suomen Liitokiekkoliiton (SLKL) Toimintakertomus 2022*. Saatavilla osoitteessa: <https://ultimate.fi/wordpress/ultimate/wp-content/uploads/2023/05/SLKL-Toimintakertomus-2022-paivitetty-2.pdf> (Avattu 31.3.2024)
- Toner, H. (2023) 'What Are Generative AI, Large Language Models, and Foundation Models?', *Center for Security and Emerging Technology*. Saatavilla osoitteessa: <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/> (Avattu 19.4.2024)
- Uyanik, G. B. (2017) 'Translation and Interpreting in Sports Contexts', *Translating and Interpreting Specific Fields: Current Practices in Turkey*. Saatavilla osoitteessa: [https://www.academia.edu/50762400/Translation\\_and\\_Interpreting\\_in\\_Sports\\_Contexts](https://www.academia.edu/50762400/Translation_and_Interpreting_in_Sports_Contexts) (Avattu 14.4.2024)
- Valli, R. (2018) 'Aineistonkeruu kyselylomakkeella', *Ikkunoita tutkimusmetodeihin 1*, 5. painos. Jyväskylä: PS-kustannus, s. 92–116.

- Valli, R., Perkkilä, P. (2018) 'Sähköinen kyselylomake ja sosiaalinen media aineistonkeruussa', *Ikkunoita tutkimusmetodeihin 1*, 5. painos. Jyväskylä: PS-kustannus, s. 117–128.
- Vehmas-Lehto, I. (2010) 'Termit kääntäjän näkökulmasta', *Käännösteoria, ammattikielet ja monikielisyys*, N:o 37. Vaasa: VAKKI:n julkaisut, s. 361–372.
- Vilkka, H. (2018) 'Havainnot ja havainnointimenetelmät tutkimuksessa', *Ikkunoita tutkimusmetodeihin 1*, 5. painos. Jyväskylä: PS-kustannus, s. 156–171.
- Vilkka, H., Saarela, M. ja Eskola, J. (2018) 'Riittääkö yksi? Tapaustutkimus kuvaajana ja selittäjänä', *Ikkunoita tutkimusmetodeihin 1*, 5. painos. Jyväskylä: PS-kustannus, s. 190–201.
- WFDF. (ei julkaisuvuotta) 'Ultimate', *World Flying Disc Federation*. Saatavilla osoitteessa: <https://wfdf.sport/disciplines/ultimate/> (Avattu 20.4.2024)

## Liitteet

### Liite 1. Lähtöteksti: “Marking infractions”

18.1. Marking Infractions:

18.1.1. Marking infractions include the following:

18.1.1.1. “Fast Count” – the marker:

18.1.1.1.1. starts or continues the stall count illegally,

18.1.1.1.2. does not start or restart the stall count with “Stalling”,

18.1.1.1.3. counts in less than one second intervals,

18.1.1.1.4. does not correctly reduce or reset the stall count when required, or

18.1.1.1.5. does not start the stall count from the correct number.

18.1.1.2. “Straddle” – a line between a defensive player’s feet comes within one disc diameter of the thrower’s pivot point.

18.1.1.3. “Disc Space” – any part of a defensive player is less than one disc diameter away from the torso of the thrower. However, if this situation is caused solely by movement of the thrower, it is not an infraction.

18.1.1.4. “Wrapping” – a line between a defensive player’s hands or arms comes within one disc diameter of the thrower’s torso, or any part of the defensive player’s body is above the thrower’s pivot point. However, if this situation is caused solely by movement of the thrower, it is not an infraction.

18.1.1.5. “Double Team” – a defensive player other than the marker is within three (3) metres of the thrower’s pivot point without also guarding another offensive player. However, merely running across this area is not a double team.

18.1.1.6. “Vision” – a defensive player uses any part of their body to intentionally obstruct the thrower’s vision.

18.1.2. A marking infraction may be contested by the defence, in which case play stops.

18.1.2.1. If a pass has been completed, a contested or retracted marking infraction must be treated as a violation by the offence, and the disc must be returned to the thrower.

18.1.3. After all marking infractions listed in 18.1.1 that are not contested, the marker must resume the stall count with the number last fully uttered before the call, minus one (1).

18.1.4. The marker may not resume counting until any illegal positioning has been corrected. To do otherwise is a subsequent marking infraction.

18.1.5. Instead of calling a marking infraction, the thrower may call a marking violation and stop play if;

18.1.5.1. the stall count is not corrected,

18.1.5.2. there is no stall count,

18.1.5.3. there is an egregious marking infraction, or

18.1.5.4. there is a pattern of repeated marking infractions.

18.1.6. If a marking infraction, or a marking violation, is called and the thrower also attempts a pass before, during or after the call, the call has no consequences (unless 18.1.2.1 applies) and if the pass is incomplete, then the turnover stands.

## Liite 2. Phrasen konekäännös

18.1. Rikkomusten merkitseminen:

18.1.1. Merkintärikkomuksia ovat muun muassa seuraavat:

18.1.1.1. "Nopea laskenta" - merkki:

18.1.1.1.1. aloittaa tai jatkaa sakkolaskentaa laittomasti,

18.1.1.1.1.2. ei käynnistä tai käynnistä uudelleen sakkauslaskentaa "Sakkaus",

18.1.1.1.1.3. laskee alle sekunnin välein,

18.1.1.1.1.4. ei vähennä tai nolaa sakkolaskentaa oikein tarvittaessa tai 18.1.1.1.1.5. ei aloita sakkolaskentaa oikeasta numerosta.

18.1.1.2. "Straddle" - puolustavan pelaajan jalkojen välinen linja tulee yhden kiekon halkaisijan päähän heittäjän kääntöpisteestä.

18.1.1.3. "Kiekkotila" - mikä tahansa puolustavan pelaajan osa on alle yhden kiekon halkaisijan päässä heittäjän vartalosta. Jos tilanne kuitenkin johtuu ainoastaan heittäjän liikkeestä, se ei ole rikkomus.

18.1.1.4. "Kietoutuminen" - puolustavan pelaajan käsien tai käsivarsien välinen linja tulee yhden kiekon halkaisijan päähän heittäjän vartalosta tai mikä tahansa puolustavan pelaajan vartalon osa on heittäjän kääntymispisteen yläpuolella. Jos tilanne kuitenkin johtuu ainoastaan heittäjän liikkeestä, se ei ole rikkomus.

18.1.1.5. "Kaksoisjoukkue" - muu puolustava pelaaja kuin merkkajaaja on kolmen (3) metrin sisällä heittäjän kääntöpisteestä ilman, että hän vartioi myös toista hyökkäävää pelaajaa. Pelkkä tämän alueen yli juokseminen ei kuitenkaan ole kaksoistiimi.

18.1.1.6. "Näkeminen" - puolustava pelaaja käyttää mitä tahansa ruumiinosaansa estääkseen tahallaan heittäjän näkyvyyden.

18.1.2. Puolustus voi kiistää merkintärikkomuksen, jolloin peli keskeytyy.

18.1.2.1. Jos syöttö on suoritettu loppuun, kiistetty tai peruutettu merkintärikkomus on käsiteltävä rikkomuksena, ja kiekko on palautettava heittäjälle.

18.1.3. Kaikkien 18.1.1 kohdassa lueteltujen merkintärikkomusten jälkeen, joita ei ole kiistetty, merkitsijän on jatkettava karsintalaskentaa sillä numerolla, joka oli viimeksi kokonaan lausuttu ennen kutsua, vähennettynä yhdellä (1).

18.1.4. Merkkiainetta ei saa laskea uudelleen ennen kuin laitton sijainti on korjattu.

Muussa tapauksessa kyseessä on myöhempi merkintärikkomus.

18.1.5. Sen sijaan, että heittäjä kutsuisi merkintärikkomusta, hän voi kutsua merkintärikkomusta ja keskeyttää pelin, jos;

18.1.5.1. Sakkauslukemaa ei ole korjattu,

18.1.5.2. Sakkolaskentaa ei ole,

18.1.5.3. kyseessä on törkeä merkintärikkomus, tai

18.1.5.4. merkintärikkomukset ovat toistuvia.

18.1.6. Jos merkintärikkomus tai merkintärikkomus tuomitaan ja heittäjä yrittää syöttää ennen merkintärikkomusta, sen aikana tai sen jälkeen, merkintärikkomuksella ei ole seurauksia (paitsi jos 18.1.2.1 kohtaa sovelletaan), ja jos syöttö on epätäydellinen, vaihto pysyy voimassa.

### Liite 3. ChatGPT3.5:n käännös

18.1. Merkkkausrikkeet:

18.1.1. Merkkkausrikkomuksiin kuuluvat seuraavat:

18.1.1.1. "Nopea Laskenta" – markkeri:

18.1.1.1.1. aloittaa tai jatkaa pysäytyslaskentaa laittomasti,

18.1.1.1.2. ei aloita tai käynnistä pysäytyslaskentaa "Lasken" kanssa,

18.1.1.1.3. laskee alle yhden sekunnin välein,

18.1.1.1.4. ei vähennä tai nolaa pysäytyslaskentaa oikein tarvittaessa, tai

18.1.1.1.5. ei aloita pysäytyslaskentaa oikeasta numerosta.

18.1.1.2. "Haarukointi" – puolustavan pelaajan jalkojen väliin tulee yksi kiekon halkaisijan etäisyydellä heittäjän tukipisteestä.

18.1.1.3. "Kiekontila" – mikä tahansa puolustavan pelaajan osa on alle yhden kiekon halkaisijan päässä heittäjän ylävartalosta. Kuitenkin, jos tilanne aiheutuu pelkästään heittäjän liikkeestä, se ei ole rike.

18.1.1.4. "Halailu" – linja puolustavan pelaajan käsien tai käsivarsien välillä tulee yhden kiekon halkaisijan etäisyydelle heittäjän ylävartalosta, tai mikä tahansa puolustavan pelaajan ruumiinosasta on heittäjän tukipisteen yläpuolella. Kuitenkin, jos tilanne aiheutuu pelkästään heittäjän liikkeestä, se ei ole rike.

18.1.1.5. "Liikaa Puolustajia" – puolustava pelaaja muu kuin markkeri on kolmen (3) metrin etäisyydellä heittäjän tukipisteestä ilman, että samalla vartioi toista hyökkäävää pelaajaa. Kuitenkin, pelkästään tämän alueen ylittäminen ei ole liikaa puolustajia.

18.1.1.6. "Näkö" – puolustava pelaaja käyttää tahallaan kehoaan estääkseen heittäjän näkemystä.

18.1.2. Merkkkausrikkeeseen voidaan tehdä vastaväite puolustuksen toimesta, jolloin peli pysähtyy.

18.1.2.1. Jos syöttö on suoritettu, kiistetyin tai peruutetun merkkkausrikkeen on oltava rikkomus hyökkäyksen toimesta, ja kiekko on palautettava heittäjälle.

18.1.3. Kaikkien 18.1.1 kohdissa lueteltujen merkkkausrikkeiden jälkeen, jotka eivät ole kiistettyjä, markkerin on jatkettava pysäytyslaskentaa viimeksi täysin lausutulla numerolla ennen huutoa, vähennettynä yhdellä (1).

18.1.4. Markkeri ei saa jatkaa laskentaa ennen kuin kaikki laittomat asennot on korjattu. Toimiminen toisin on seuraava merkkkausrikke.

18.1.5. Sen sijaan, että kutsuisi merkkaustrikkomusta, heittäjä voi kutsua merkkaustrikkomusta ja pysäyttää pelin, jos;

18.1.5.1. pysäytyslaskentaa ei korjata,

18.1.5.2. pysäytyslaskentaa ei ole,

18.1.5.3. on vakava merkkaustrikkomus, tai

18.1.5.4. on toistuvia merkkaustrikkomuksia.

18.1.6. Jos merkkaustrikkeestä tai merkkaustrikkomuksesta huudetaan, ja heittäjä yrittää myös syöttöä ennen, sen aikana tai sen jälkeen, huudolla ei ole seurauksia (ellei 18.1.2.1 koske), ja jos syöttö on epätäydellinen, silloin menetys pysyy voimassa.

**Liite 4. Termistö**

|                     |                                     |
|---------------------|-------------------------------------|
| act of throwing     | heittäminen                         |
| affect the play     | vaikutusmerkitys vaikuttaa peliin   |
| attacking end zone  | hyökkävä maali-alue                 |
| best perspective    | paras näkökulma                     |
| breach              | sääntöriike                         |
| call                | huuto                               |
| catch               | kiinniotto                          |
| central zone        | keski-alue                          |
| check               | pelin käynnistys                    |
| defence             | puolustus                           |
| defending end zone  | puolustettava maali-alue            |
| defensive player    | puolustaja                          |
| end zone            | maali-alue                          |
| foul                | virhe                               |
| goal                | maali                               |
| goal line           | maaliviiva                          |
| ground              | pelialueen ympäristö                |
| ground contact      | maakosketus                         |
| guarding            | puolustaminen                       |
| infraction          | riike markkerivirhe                 |
| injury              | loukkaantuminen                     |
| interception        | katko                               |
| legitimate position | oikeutettu sijoittautuminen         |
| line                | viiva                               |
| making a play on    |                                     |
| the disc            | kiekontavoittelu tavoitella kiekkoa |
| marker              | markkeri                            |
| minor contact       | vähäinen kontakti                   |
| non-player          | muu kuin pelaaja                    |
| offence             | hyökkäys                            |
| offensive player    | hyökkääjä                           |
| out-of-bounds       | pelikentän ulkopuolella ulkona      |

|                        |  |
|------------------------|--|
| perimeter lines        | rajaviivat   |
| pivot                  | tukipiste  |
| pivot location         | tukipisteen sijainti                                   |
| play                   | pele pelata  |
| player                 | pelaaja  |
| playing field          | pelikenttä   |
| point                  | piste  |
| possession of the disc | hallussapito   |
| pre-stall              | ennakkolaskenta  |
| pull                   | aloitusheitto heittää aloitusheitto                    |
| receiver               | hakija   |
| score                  | maalinteko tehdä maali                                 |
| stall count            | pysäytyslaskenta                                       |
| stoppage of play       | seis-tilanne   |
| technical stoppage     | tekninen keskeytys                                     |
| throw                  | heitto heittää   |
| thrower                | heittäjä   |
| throwing motion        | heittoliike  |
| turnover               | hyökkäysvuoron vaihtuminen                             |
| violation              | sääntörikkomus<br>kiekon pysähtymispaikka minne kiekko |
| where the disc stops   | pysähtyy   |
| marking                | markkerointi   |
| straddle               | haarukointi  |
| wrapping               | halailu  |
| double team            | liikaa puolustajia                                     |
| vision                 | näkö   |
| disc space             | kiekontila   |
| stalling               | lasken   |

**Liite 5. Tieteellisen tutkimuksen tietosuojailmoitus**

**TIETEELLISEN TUTKIMUKSEN  
TIETOSUOJASELOSTE/-ILMOITUS  
EU:n yleinen tietosuoja-asetus  
13–14 artiklat  
Laatimispäivä: 8.2.2024**

**Tietoa henkilötietojen käsittelystä *Konekääntimen ja generatiivisen tekoälyn vaatima jälkieditointityö urheilusääntöjen kääntämisessä* -tutkimuksessa**

*Konekääntimen ja generatiivisen tekoälyn vaatima jälkieditointityö urheilusääntöjen kääntämisessä*-tutkimuksessa käytetään aineistona henkilötietoja. Tämän selosteen tarkoitus on antaa tietoa käsiteltävistä henkilötiedoista sekä siitä, mistä henkilötiedot ovat peräisin ja miten niitä käytetään tutkimuksessa. Selosteen lopussa kerrotaan tarkemmin, mitä oikeuksia rekisteröidyillä on.

Tutkimukseen osallistuminen ja henkilötietojen antaminen on vapaaehtoista. Sinuun ei kohdistu mitään negatiivista seuraamusta, jos et osallistu tutkimukseen tai jos keskeytät osallistumisesi tutkimukseen.

**Tutkimuksen rekisterinpitäjä**

Helsingin yliopisto

Osoite: PL 3 (Fabianinkatu 33), 00014 Helsingin yliopisto

Vilma Heinonen

vilma.heinonen@helsinki.fi

+358407219875

**Yhteyshenkilö ja vastuullinen tutkija**

Yhteyshenkilö tutkimusta koskevissa asioissa: Vilma Heinonen

Tiedekunta/osasto/yksikkö: Humanistinen tiedekunta

Puhelinnumero: +358407219875

Sähköpostiosoite: vilma.heinonen@helsinki.fi

Vastuullinen tutkija:

Nimi: Marjut Alho

Sähköpostiosoite: marjut.alho@helsinki.fi

**Tietosuojavastaavan yhteystiedot**

Helsingin yliopiston tietosuojavastaavaan saa yhteyden sähköpostiosoitteesta tietosuoja@helsinki.fi.

## **Kuvaus tutkimushankkeesta ja henkilötietojen käsittelyn tarkoitus**

*Henkilötietoja käsitellään osana maisterin tutkielmaa varten Helsingin yliopistolla. Vilma Heinonen vastaa aineiston käytännön hallinnasta kuten käsiteltävien tietojen minimoinnista, täsmällisyydestä sekä säilytysaikojen määräytymisestä ja asianmukaisten järjestelmien/sovellusten käyttämisestä. Käsittelyyn liittyvät riskit arvioidaan yhteistyössä ohjaajansa kanssa. Yliopisto ohjaa ja tukee henkilötietojen käsittelyssä suunnittelusta lopullisen työn julkaisuun ja tarjoaa tarvittavan infrastruktuurin sekä välineet tutkimuksen toteuttamiseksi.*

## **Mitä tietoja tutkimusaineisto sisältää**

*Vastaajilta kerätään sähköpostiosoitteet tutkimukseen liittyvää viestintää varten. Sähköpostiosoitteet eivät ole osa tutkimukseen kerättyä aineistoa.*

## **Mistä lähteistä henkilötietoja kerätään**

*Henkilötiedot kerätään vastaajilta. Vastaajia haetaan sosiaalisen median kanavissa.*

## **Arkaluonteiset henkilötiedot**

Tutkimuksessa ei käsitellä tietosuojasetuksen 9 artiklan mukaisia erityisiä henkilötietoryhmiä (eli arkaluonteisia henkilötietoja).

## **Henkilötietojen käsittelyn oikeusperuste**

Henkilötietoja käsitellään seuraavalla yleisen tietosuoja-asetuksen 6 artiklan 1 kohdan mukaisella perusteella:

yleistä etua koskeva tehtävä:

tieteellinen tai historiallinen tutkimus tai tilastointi (tietosuojalain 4 §:n 3 kohta)

tutkimusaineistojen ja kulttuuriperintöaineistojen arkistointi  
(tietosuojalain 4 §:n 4 kohta)

tutkittavan suostumus

rekisterinpitäjän lakisääteisen velvoitteen noudattaminen

rekisterinpitäjän tai kolmannen osapuolen oikeutettujen etujen toteuttaminen  
mikä oikeutettu etu on kyseessä:

Jos henkilötietojen käsittely perustuu tutkittavan suostumukseen, tutkittavalla on koska tahansa oikeus peruuttaa suostumuksensa. Suostumuksen peruminen ei vaikuta ennen peruuttamista suoritettujen käsittelyjen lainmukaisuuteen.

## **Tietojen vastaanottajat**

*Tietosuojailmoituksen laatimishetkellä tiedossa olevat vastaanottajat:  
Vilma Heinonen, vastaajien sähköpostiosoitteet tutkimukseen liittyvää viestintää varten.*

## **Tietojen siirto Euroopan talousalueen ulkopuolelle**

Tietoja ei siirretä Euroopan talousalueen ulkopuolelle, niitä käsitellään ainoastaan EU:n alueella.

## **Automatisoitu päätöksenteko**

Tutkimuksessa ei tehdä automaattisia päätöksiä, joilla on merkittävä vaikutus tutkittaviin.

## Henkilötietojen suojaus

Tutkimusaineistoon sisältyviä henkilötietoja käsitellään ja säilytetään suojattuna niin, että ainoastaan niitä tarvitsevat henkilöt pääsevät tarkastelemaan tietoja.

Tietojärjestelmissä käsiteltäviä tietoja suojataan seuraavilla tavoilla:

- käyttäjätunnus ja salasana
- käytön rekisteröinti/lokitus
- kulunvalvonta
- salaus/kryptaus
- kaksivaiheinen tunnistautuminen
- muu, mikä:

Suorien tunnistetietojen käsittely:

- Rekisterinpitäjä kerää henkilötiedot ilman suoria tunnistetietoja
- Suorat tunnistetiedot poistetaan analysointivaiheessa ja säilytetään erillään analysoitavasta tutkimusaineistosta
- Aineisto analysoidaan suorien tunnistetiedoin, koska (peruste suorien tunnistetietojen säilyttämiselle):

## Henkilötietojen käsittelyn kesto tässä tutkimuksessa

*Henkilötietojen käsittelyn kesto on 8.-28.2.2024. Henkilötiedot poistetaan analysoitavasta aineistosta vastausajan päättymisen jälkeen.*

## Henkilötietojen käsittely tutkimuksen päättymisen jälkeen

- Tutkimusaineisto hävitetään
- Tutkimusaineisto säilytetään tämän tutkimuksen tulosten luotettavuuden arvioimista varten:
  - ilman suoria tunnistetietoja  tunnistetiedoin

Tutkimusaineisto säilytetään myöhempää, yhteensopivaa tieteellistä tutkimusta varten tietosuoja-asetuksen vaatimusten mukaisesti:

ilman suoria tunnistetietoja  tunnistetiedoin

Tutkimusaineiston säilytys perustuu tietosuoja-asetuksen 5 artiklan 1 kohdan b ja e alakohtiin.

Tutkimusaineiston uudesta tutkimuskäytöstä lähetetään rekisteröidyille uusi tietosuojailmoitus, paitsi jos rekisterinpitäjä ei enää pysty tunnistamaan rekisteröityjä tutkimusaineistosta.

Uudesta tutkimuksesta ei voida lähettää ilmoitusta rekisteröidylle myöskään silloin, jos tietojen toimittaminen olisi mahdotonta tai kohtuuttoman vaivalloista tai jos se estäisi tai vaikeuttaisi suuresti tutkimustarkoitusten saavuttamista (tietosuoja-asetuksen 14 artiklan 5 kohdan b alakohta).

Missä aineistoa säilytetään ja kuinka kauan: Vilma Heinosen arkisto, 8.2.-31.4.2024.

## **Mitä oikeuksia rekisteröidyllä on ja oikeuksista poikkeaminen**

Yhteyshenkilö tutkittavan oikeuksiin liittyvissä asioissa on tämän ilmoituksen kohdassa 1 mainittu yhteyshenkilö.

### **Rekisteröidyn oikeudet**

Tietosuoja-asetuksen mukaan rekisteröidyllä on oikeus:

- saada pääsy omiin tietoihin
- oikaista tietojaan
- poistaa tiedot ja tulla unohdetuksi
- rajoittaa omien tietojensa käsittelyä
- siirtää tiedot rekisterinpitäjältä toiselle
- vastustaa tietojen käsittelyä
- olla joutumatta automaattisen päätöksenteon kohteeksi.

Rekisteröity ei kuitenkaan voi käyttää kaikkia oikeuksia kaikissa tilanteissa.

Tilanteeseen vaikuttaa esimerkiksi se, millä perusteella henkilötietoja käsitellään.

Tarkempaa tietoa rekisteröidyn oikeuksista eri tilanteissa löytyy tietosuojavaltuutetun verkkosivuilta: <https://tietosuoja.fi/rekisteroidyn-oikeudet-eri-tilanteissa>

### **Oikeuksista poikkeaminen**

Tietosuoja-asetus ja Suomen tietosuojalaki mahdollistavat tietyistä rekisteröidyn oikeuksista poikkeamisen silloin, kun henkilötietoja käsitellään tieteellisessä tutkimuksessa ja oikeuksien toteuttaminen estäisi tai vaikeuttaisi suuresti käsittelyn tarkoitusten saavuttamista.

Tarvetta poiketa rekisteröidyn oikeuksista arvioidaan aina tapauskohtaisesti.

### **Valitusoikeus**

Sinulla on oikeus tehdä valitus tietosuojavaltuutetun toimistoon, mikäli katsot, että henkilötietojesi käsittelyssä on rikottu voimassa olevaa tietosuojalainsäädäntöä.

Yhteystiedot:

Tietosuojavaltuutetun toimisto

Käyntiosoite: Ratapihantie 9, 6. krs, 00520 Helsinki

Postiosoite: PL 800, 00521 Helsinki

Vaihde: 029 56 66700

Faksi: 029 56 66735

Sähköposti: tietosuoja(at)om.fi

## Abstract

**Faculty:** Faculty of Arts

**Degree programme:** Master's Programme in Translation and Interpreting

**Study track:** Translation technology, English into Finnish

**Author:** Vilma Heinonen

**Title:** Post-editing Effort when Translating Sports Rules with Machine Translation System and Generative AI – A Case Study of Translating the Rules of Ultimate Frisbee

**Level:** Master's thesis

**Month and year:** April 2024

**Number of pages:** 64 pages, attachments 15 pages

**Keywords:** machine translation, translation of sports rules, translating with AI, ChatGPT, case study

**Supervisor or supervisors:** Marjut Alho

**Where deposited:** Helda – open repository of University of Helsinki

### Additional information:

**Abstract:** This master's thesis examines the post-editing effort required when translating, and specifically machine translating, sports rules using generative AI. This thesis focused on translating the rules of *ultimate frisbee* using these translation tools. This master's thesis was inspired by the ongoing evolution in the translation industry, where neural network machine translation systems and generative AI are being employed increasingly as translation tools, and their application being increasingly expanded for various areas of special field translation.

Although played around the world, ultimate, formerly known as ultimate frisbee attracts a relatively small following when compared with very popular sports such as football or ice hockey., The Finnish Flying Disc Association currently has some 450 members. In ultimate frisbee, *Spirit of the Game*, or 'SOTG' is an essential concept. According to WFDF, "Spirit of the Game is the mindful behaviour practiced by players worldwide prior to, during and after a game. It encompasses attitudes and skills such as good knowledge and application of the rules, fair-mindedness, safe play, and spatial awareness, clear and calm communication, and positive and respectful attitude towards teammates, fans, and opponents, in a mutual effort to protect the basic joy of play." The game is played without any referees and the refereeing is left to the players. This requires that each player has sound knowledge and a good understanding of the rules of the game.

Personally, I am also a passionate ultimate frisbee player and wish to make the sport more accessible to Finns, because the rules of ultimate frisbee are not currently available in Finnish. The rules of the game are updated every four years, but they are published only in English, Spanish, French and in a number of other languages, but not Finnish. This is the reason why I studied the post-editing effort required when translating the rules of ultimate frisbee with a machine translation system and with generative AI. I also drew a conclusion regarding the cost and time feasibility of translating and post-editing the rules.

The impetus for this master's thesis was motivated by the lack of earlier research in the special field of translation of sports, as well as a lack of earlier research focusing on translating with the help of generative AI. Generative AI is such a new phenomenon, that it has not yet secured a market of its own in translation, however, it has been used for translation. The first publications of the use of Generative AI in translation date from 2023 but since then there has been a new version of ChatGPT open to the public and has not yet been scrutinized in research.

ChatGPT is the generative AI created by OpenAI and is used in this thesis along with Memsources' Phrase machine translation system.

In my master's thesis, I intended to answer the following three research questions:

1. What is the quality of the machine translation and the translation created with generative AI, when translating sports rules with a term bank?
2. How much post-editing effort is required when translating sports rules with generative AI and a machine translation system in order to present a readable and usable translation?
3. What are the key differences between the machine translation system and generative AI that influence the translation quality?

My hypothesis is that the machine translation system will make fewer errors in translation than generative AI, because the MT has been perfected for this purpose only, and its quality is particularly evinced in repetitive document texts, such as the rules of a sport.

This master's thesis is a case study; it focuses on the translation of rules in one sport and is not a comparative study of translating different sports, or even different versions of the rules of the same sport. This is a case study also insofar as it examines the translation of sports rules from a novel perspective: translating this type of texts using neural machine translation and generative AI.

This thesis studies a field of translation that has not yet been established as others are. This being the case, the theoretical framework for this study has been derived by combining those in different fields of study, in order to study a very specific phenomenon in the field of translation. Despite the paucity of case-specific studies, there are several publications from each field of study of this thesis, and the theoretical background section of my study has been divided accordingly.

For the first part of my theoretical framework, I have considered approaches for different kinds of machine translation systems, such as statistical machine translation and neural machine translation systems. In keeping with the purpose of this, I broadly review the operating principle of the neural machine translation system. The history of machine translation is also briefly discussed in this section. Next, machine translation and post-editing are considered: why the research into post-editing effort is so crucial in analyzing the quality of machine translation, and why post-editing is such a good measure of machine translation quality. Post-editing effort is widely discussed in Maarit Koponen's research, and the manner in which she approaches her research is highly relevant for this thesis, which is why it includes several of her studies: they represent different perspectives on a very specific field of study and thus all are essential for the purpose of this thesis.

In the third component of my theoretical framework addresses generative AI, and its operating principle is also briefly discussed. The theoretical background on generative AI mostly discusses how generative AI is currently used in translation and how it is prompted to carry out a successful translation. A prompt is a cue given to the generative AI to assign a task. Even though generative AI and a neural machine translation system operate based on the same principle, only generative AI is able to utilize information given within the prompt in addition to utilizing its training data given to it prior to the task at hand. As the studies in the theoretical background section concerning generative AI are not peer reviewed, they should be viewed with a critical eye. Nevertheless, they are scientific publications, published on a reliable platform, and were thus included in my thesis. It is important to note that currently there are no peer reviewed studies of generative AI in translation.

The last component of my theoretical framework addresses the field of translation of sports. There is a paucity of research in this special field: I was able to identify only two relevant studies, which investigate this field of translation: Uyanik (2017) has conducted a study on sports texts to establish a framework for the genre of sports text, and the conventions of translations in this genre; the second was an MA thesis by Alasalmi (2014), which examined translation problems faced when translating the rules of rugby from English into Finnish, i.e. the subject of this case study is quite similar to mine.

The method adopted in my thesis relies heavily on Koponen's (2010) research on machine translation quality, and the error classification used in that study. I considered error analysis classifications employed by Koponen (2010), as well as Koponen and Salmi (2015). I decided upon Koponen's (2010) classification, because its more specific classification would provide more data for the analysis in this thesis and thus furnish more information concerning the translating of sports rules via machine

translation system and generative AI. All possible data from my study could be useful for future research in this field, and could provide the opportunity to conduct a comparative analysis based on the results extracted in this research.

Translation students and experts on the sport of ultimate frisbee were asked to review and revise the translations, after which I applied Koponen's (2010) error analysis framework. Data collection was conducted online. Participants were recruited through social media channels, and the documents were posted for revision on a cloud-based word processor, that was accessible for all and accessed via email. Based on the revisions made by the participants of the study, I classified the errors and exported them to tables, in order to allow a quantitative analysis alongside a qualitative analysis.

The participants were required to provide informed consent to use their emails for research purposes, but their emails are not part of this study's research material. The participants' emails were used for communication related to the research and to be able to share the document via a cloud-based word processor. A copy of the Data Protection Notice for Scientific Research appears as an appendix to this thesis.

Some methods were adapted to optimize data collection, due to the time and scope limitations of this research project.

The task assigned to my research participants was a complicated one due to its unconventional data collection method. In several other post-editing effort studies, such as Koponen and Salmi (2015) and Maučec and Donaj (2019), amongst others, a simple questionnaire was used for data collection; participants usually have to either assess the translation on a specific scale, or assess it as right or wrong, based on linguistic and content correctness. However, in this study I targeted a deeper understanding of the errors made by the machine translation system and generative AI, thus I employed a specific error classification. This, more specific error classification is more difficult, for which reason the participants were not troubled to apply it, rather I, as researcher, decided to complete the error classification myself, based on the corrections the participants had made in the translations. The participants highlighted every correction they would have in the document, and corrected every error as they best could. Based on these corrections, and comparing them to the errors, I was able to classify the errors. There were a few instances in which errors were not corrected, and in those cases I simply classified them as "mistranslated concepts" (Koponen 2010).

Results of the error analysis are presented in several different tables in the analysis section, which highlight the results. The results of the error analysis were also analyzed by comparing them with the theoretical framework of this thesis, and based on those comparisons, some conclusions can be drawn from the data.

Table 3 illustrates that both the machine translation system and the generative AI made from one to two errors per each sentence, which is a significant number of errors for a short, one-paged translation. However, even though both tools made a significant number of errors, it was somewhat unexpected that they made approximately the same number of mistakes. This indicates that my hypothesis at the outset of this thesis was wrong. When scrutinized, the error rates reveal that generative AI has made fewer errors than the machine translation, thus my hypothesis is proved wrong.

Based on Table 3, both generative AI and the machine translation are not able to adapt the sentence structure from the source language into the target language, given the only number in Table 3, which has remained unchanged between the original text and both translations, is the number of sentences in all texts. Some of the translation errors could be caused by this inability to adapt the sentence structure between languages, even though the generative AI and the machine translation system were able to produce somewhat readable translations without making the necessary adaptations to the sentence structure. This is adaptation is something a human translator would have found a more efficient way of translating.

Upon closer inspection of the results, the hypothesis at the outset of this thesis is not proved completely wrong, since the errors are divided by their characteristics into conceptual errors and into relation errors. Conceptual errors are those errors, which alter the meaning of the sentence drastically and can be considered errors on content, whereas relation errors are considered as linguistic errors and

errors concerning the language of the translation. Findings in Table 6 align with the findings Table 3, which means that generative AI made fewer conceptual errors than the machine translation system. However, the findings in Table 8 contradict earlier findings; the machine translation system made fewer linguistic errors than generative AI, and thus my hypothesis is not completely wrong. This finding is unexpected, not only because of the earlier findings in the analysis, but also because generative AI is considered a very reliable proofreader and language checker. In this study's research material, generative AI's translation has been corrected for simple typos, which seems a bit uncharacteristic for a machine.

Alasalmi's (2014) study on translating sports rules handles the errors concerning terminology in the translations, for which reason this was included in the data analysis section of this thesis. As evinced in Table 7, generative AI made fewer errors in this domain than the machine translation system. This can be explained by the generative AI's ability to use the terminology given to it without tampering with the rough translation, whereas the machine translation system was not able to use the term bank without being post-edited, thus it lacks in the terminology. This can also account for the findings in Tables 3 and 6, to some degree. Errors concerning terminology are inherently classified as conceptual errors, because according to Vehmas-Lehto, terms are the key to a text of a specialty field, because the extralinguistic information is accumulated in them (2010, p. 362).

*Editing fatigue* is a relevant phenomenon, which was studied in the theoretical framework by Koponen and Salmi (2015, p. 132). There were a number of different manifestations of editing fatigue in this thesis: inconsistencies within the post-editing process, not all appearances of an error term being noticed; inconsistency of correction throughout the translations; or not completing the post-editing of the document. Some of the participants reported that the process of post-editing of two, one-paged translations exceeded an hour in duration, in contrast to the referential timeframe, which was twenty minutes. This indicates that editing fatigue should be expected, especially from participants, who usually do not carry out these kinds of tasks.

Table 3 illustrates that generative AI made altogether slightly fewer errors than the machine translation system. However, once the extralinguistic factors and accessibility were taken into consideration, generative AI emerged as the preferable alternative, when translating texts for smaller businesses or sport federations. Smaller businesses or sports federations may not be able to afford to hire a professional translator to carry out the translation, or even invest in a license to use a machine translation system. This then is the reason why translating with generative AI is a viable option, even though the task still requires time and effort to be invested.

This study demonstrated that both generative AI and the machine translation system lacked in the ability to translate the text well. Table 3 indicates that there were on average 1.7 (ChatGPT3.5) and 1.9 (Phrase machine translation system) errors per sentence, thus it is evident that both translations required a considerable amount of post-editing effort, regardless of whether the errors were linguistic or conceptual. Key differences between generative AI and the machine translation system that affected the quality of translation were extralinguistic factors, such as usability and accessibility.

There are various limitations to this study. Since this was a case study focusing on the rules of one sport, no generalizations on translating sports text or translating with generative AI can be drawn. Previous research on both translating sports texts and translating with generative AI has been limited thus far. The results of this study have been limited due to the small number of participants in this study, even though two different participant groups, translation students and experts of ultimate, have both been involved in this project to allow for broader perspectives on the translations.

The research questions were successfully answered and chosen for the scope of this thesis, despite its various limitations. However, in future research, some measures should be taken to avoid editing fatigue. Shortening the length of text passages that are post-edited by the participants is one such measure. One could also divide the participants into focus groups that post-edit different translations, if comparative analysis is desired.

Based on the analysis section of my thesis, a more specific error categorization is justified for a comparative study such as this one; the more specific categorization differences between generative AI and machine translation allow for easier distinction.

In future research, it would be interesting to study the differences in the quality of translations influenced by sociolinguistic factors, such as the differences between professional translators and experts of the sport. This research perspective was discussed in Uyanik's (2017) article on translating sports texts, where Uyanik stated that the translator of sports text would ideally be a professional translator with a deep understanding of the sport (2017, p. 109, 111–112). Although an interesting research perspective, this is, however, outside the scope of this study, due to the length of this thesis, and because handling any personal participant information was kept to a minimum.

Terminology of sports may also provide an intriguing topic of research. Research into the use of established terms in the spoken language of a sport would be one such novel approach. This is a research perspective I was able to touch upon in the analysis of my thesis, but the analysis was very limited due to the scope of this project.