

Pronominien kääntyminen konekääntimillä ranskasta suomeen

Joanna Rajala

Pro gradu -tutkielma

Monikielisen käännösviestinnän tutkinto-ohjelma, ranska

Kieli- ja käännöstieteiden laitos

Humanistinen tiedekunta

Turun yliopisto

Huhtikuu 2024

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck –järjestelmällä

Pro gradu -tutkielma

Monikielisen käännösviestinnän tutkinto-ohjelma, ranska

Joanna Rajala

Pronominien kääntyminen konekääntimillä ranskasta suomeen

Sivumäärät: tutkielma 49 sivua, liite 7 sivua.

Tässä pro gradu -tutkielmassa tutkitaan pronominien kääntymistä konekääntimillä ranskasta suomeen. Tutkimuskohteet ovat seuraavat ranskan kielen pronominit: *il, elle, ils, elles, l', le, la, ja les*. Kyseiset pronominit valittiin tutkittaviksi sen takia, että useimmat niistä voivat ilmaista tarkoitteen sukupuolta, mukaan lukien kieliopillinen sukupuoli. Tämä on ranskan kielessä esiintyvä ilmiö, jota suomen kielessä ei ole. Suomen kielen pronominit eivät ilmaise sukupuolta, joten niiden kääntämisessä voidaan joutua käyttämään muita keinoja. Tämä ranskan ja suomen kielten välinen ero tekee aiheesta mielenkiintoisen.

Tässä tutkimuksessa tutkittiin kahta varsinaista konekääntintä (Google Translate ja DeepL Translator) sekä ChatGPT:tä, joka on tekoäly, mutta jota voidaan käyttää myös kääntimenä. Aineistona toimi osio ranskankielisestä Katherine Pancolin romaanista *La Valse lente des tortues* (2008). Myös teoksen suomennosta *Kilpikonnien hidas valssi* (Lotta Toivanen, 2012) käytettiin osana tutkimusta.

Tarkoituksena oli tutkia sitä, miten konekääntimet selviytyvät pronominien kääntämisestä. Erityisesti tutkittiin sitä, ovatko viittaukset oikein ja ovatko ne selviä. Tätä varten kehiteltiin oma arviointimalli. Eri konekääntimien tuloksia voidaan vertailla keskenään, jolloin saadaan selville, mikä niistä toimii parhaiten. Tutkimuksessa saatiin myös tietoa esimerkiksi siitä, mitä käännöksiä kääntimet käyttävät eniten.

Tulosten perusteella suurin osa jokaisen kääntimen viittauksista oli oikein. Tätä tärkeämpää on kuitenkin tarkastella sitä, ovatko viittaukset selviä. Selkeyden kohdalla esiintyi myös eroja eri kääntimien välillä. Eniten selviä viittauksia oli DeepL:llä (142 kohtaa 249:stä), toiseksi eniten ChatGPT:llä (94) ja vähiten Google Translatella (81). Täytyy kuitenkin pitää mielessä, että kyseessä on vain pieni aineisto, eikä näiden tulosten perusteella voida tehdä pitäviä johtopäätöksiä. Siispä olisikin kiinnostavaa tutkia aihetta lisää erilaisilla aineistoilla, kielipareilla ja kääntimillä.

Avainsanat: konekääntäminen, persoonapronominit, sukupuoli

Sisällysluettelo

1	Johdanto	5
2	Teoreettinen viitekehys	8
2.1	Arvioiva tutkimus	8
2.2	Konekääntimet	9
2.2.1	Sääntöpohjaiset konekääntimet	10
2.2.2	Tilastolliset konekääntimet	11
2.2.3	Neuroverkkokääntimet	11
2.2.4	Konteksti	12
2.3	Konekääntimet	13
2.3.1	Google Translate	13
2.3.2	DeepL Translator	13
2.3.3	ChatGPT	14
2.4	Ranskan pronominit	14
2.5	Suomen pronominit	15
2.5.1	Persoonapronominit	15
2.5.2	Demonstratiivipronominit	16
3	Aineisto ja menetelmät	18
3.1	Aineisto	18
3.2	Metodi	21
3.3	Arviointimalli	22
3.4	Konekääntimet	23
4	Analyysi ja tulokset	24
4.1	Ihmiskäännöksen kohtien luokittelu	24
4.2	Viittaukset oikein vai väärin	25
4.3	Viittaukset selviä vai epäselviä	26
4.4	Erot sanojen välillä	26
4.5	Erot luokkien välillä	28
4.6	Yleisimmät käännökset	30
4.7	Esimerkkejä kääntimien onnistuneista käännösratkaisuista	35

4.7.1	Ensimmäinen ryhmä: sukupuoli ilmaiseva substantiivi	36
4.7.2	Toinen ryhmä: nimen käyttäminen	38
4.7.3	Kolmas ryhmä: viittaus selvä muun lauseen perusteella	39
4.7.4	Viittaus selvä muun kontekstin perusteella	40
4.8	Esimerkkejä kääntimien epäonnistuneista viittauksista	41
4.8.1	Väärät viittaukset	42
4.8.2	Epäselvät viittaukset	44
4.8.3	Monitulkintainen pronomini	45
4.9	Huomautus kontekstista	46
5	Yhteenveto ja johtopäätökset	47
6	Lopuksi	49
	Lähteet	51
	Liite: Résumé en français	54

1 Johdanto

Tämän tutkimuksen tarkoituksena on selvittää, miten konekääntimet kääntävät kolmannen persoonan pronomineja ranskasta suomeen. Tutkimuskohteina ovat yksikön ja monikon kolmannen persoonan pronominit (*il, elle, ils, elles*) sekä seuraavat pronominit: *l', le, la, ja les*. Aihetta on mielenkiintoista tutkia, sillä konekääntimiä kehitetään ja käytetään aina vain enemmän. Mielestäni on siis tärkeää selvittää, miten konekääntimet selviävät lähtö- ja tulokielten välisistä eroista johtuvista mahdollisista ongelmakohdista. Konekääntimien tekemien virheiden tunnistaminen onkin ensimmäinen askel kääntimien kehittämiseen. Nykyisillä kääntimillä voi vielä olla ongelmia tässä tutkimuksessa tarkasteltavan ilmiön kanssa, mihin voi vaikuttaa osaltaan se, että kyseessä on harvinaisempi kielipari, jonka kehittämiseen ei käytetä yhtä paljon aikaa ja resursseja kuin enemmän käytettyjen kieliparien kohdalla.

Tutkimuksen aineistona on ranskankielinen kaunokirjallinen teos (Katherine Pancol: *La Valse lente des tortues*, 2008) ja sen suomennoksia. Kirjasta valittu osio on noin 60 sivun pituinen. Konekääntimiä ei tällä hetkellä juurikaan käytetä kaunokirjallisuuden kääntämisessä, mutta asia voi muuttua tulevaisuudessa. Tutkimuksessa ei ole kuitenkaan tarkoituksena keskittyä erityisesti kaunokirjallisuuden kääntämiseen konekääntimillä, vaan aineistoon valintaan vaikutti lähinnä se, että romaaneissa esiintyy tyypillisesti paljon pronomineja. Tulokset eivät siis rajoitu pelkästään kaunokirjallisuuteen vaan tätä ranskan ja suomen kielten pronominiin eroista johtuvaa ilmiötä esiintyy myös muissa tekstilajeissa.

Tarkasteltavia käännöksiä oli siis useita. Näistä yksi oli ihmisen tekemä käänös (Lotta Toivanen: *Kilpikonnien hidas valssi*, 2012), joka toimi eräänlaisena pohjana, johon muita käännöksiä voi verrata. Muut kolme käännöstä olivat koneiden tekemiä. Konekääntimiksi valittiin Google Translate (<https://translate.google.fi/?hl=fi>) ja DeepL Translator (<https://www.deepl.com/translator>). Käytössä oli DeepL:n ilmaisversio. Lisäksi teksti käännettiin ChatGPT:llä (<https://chat.openai.com/auth/login>), joka ei ole varsinainen konekäännin vaan tekoäly, mutta jota voidaan käyttää myös kääntämiseen. Tutkimuksessa käytettiin ChatGPT:n versiota 3.5.

Ranskan ja suomen kielten kolmannen persoonan pronominiin erona on se, että ranskassa on kaksi pronominia *il(s)* ja *elle(s)*. Ensimmäisellä viitataan miespuolisiin henkilöihin, ja jälkimmäisellä naispuolisiin henkilöihin. Henkilöiden lisäksi näillä pronomineilla voidaan

viitata myös vastaavasti maskuliinis- ja feminiinissukuisiin asioihin. Suomen kielessä tällaista erottelua ei ole, vaan kaikkiin viitataan samalla pronominilla hän tai he. Tämä ero aiheuttaa toisinaan ongelmia kääntämisessä. Ranskan kielessä voidaan erotella henkilö käyttämällä joko maskuliini- tai feminiinipronominia. Hän-pronominin käyttäminen käännöksessä voi aiheuttaa sekaannusta tapauksissa, joissa viittaus ei ole selvä. Tästä johtuen käännöksessä voidaan joutua käyttämään esimerkiksi henkilön nimeä tai jotakin muuta sanaa, joka selventää viittauksen. Sama ongelma voi ilmaantua myös muiden tutkittavien sanojen kohdalla. Samaan aihepiiriin liittyvää tutkimusta on tehty ainakin venäjä-suomi-kieliparissa (Kuzmina, 2016). Kyseisessä pro gradu -tutkimuksessa ei kuitenkaan tutkittu konekääntimiä vaan pelkästään ihmisen tekemää suomennosta venäjänkielisestä teoksesta. Kyseessä ei siis ole pelkästään ranska-suomi-kieliparissa esiintyvä ongelma, vaan monissa muissakin kielissä esiintyy samanlaisia pronomien eroista johtuvia vaikeuksia. Samanlaista tutkimusta voitaisiin tehdä myös muiden kielten osalta, ja esimerkiksi vertailla näiden tutkimusten tuloksia keskenään.

Tässä tutkimuksessa tarkoituksena on tehdä konekäännöksille analyysi, josta selviää, ovatko viittaukset oikein. Tähän käytän itse kehrittelemääni yksinkertaista arviointimallia, jossa tarkastellaan paitsi viittausten täsmällisyyttä myös niiden selkeyttä. Pääpaino onkin viittausten selkeydessä, sillä väärin kääntyneitä viittauksia esiintyi vain vähän, ja tulokset eri kääntimien välillä olivat pieniä. Selkeyden kohdalla eroja kuitenkin esiintyi enemmän, ja sitä tutkimalla voidaan paremmin arvioida konekääntimien käännösten onnistuneisuutta. Tarkasteltavat kohdat määräytyvät ihmisen tekemän käännöksen pohjalta. Aineistosta poimitaan kohdat, joissa ihmiskääntäjä on päätenyt käyttämään eri viittaustapaa kuin alkuperäisessä tekstissä. Näin voidaan tutkia vain aiheen kannalta kiinnostavia kohtia, joissa konekääntimillä voidaan olettaa olevan vaikeuksia. Aineistosta ei siis otettu mukaan sellaisia kohtia, joissa myös suomennoksessa on käytetty pronomina. Tutkimuksen aiheena olevia kohtia on tutkittavassa osiossa yhteensä 249 kappaletta.

Tarkasteltavat kohdat on jaoteltu kolmeen eri luokkaan. Ensimmäiseen luokkaan kuuluvat sellaiset kohdat, joissa pelkän pronominin käyttäminen suomennoksessa aiheuttaisi sekaannusta. Toisessa luokassa ovat sellaiset kohdat, joissa suomentaja on päätenyt käyttämään eri viittaustapaa, vaikka sekaannuksen vaaraa ei varsinaisesti ole. Tämän luokan tapauksissa voi olla kyse käännöksen eksplisiittistämisestä, eli siitä, että käännöksestä on tehty selkeämpi. Esimerkiksi nimen käyttäminen pronominin sijasta voi auttaa lukijaa pysymään selvillä siitä, kenestä on kyse. Kolmannessa luokassa puolestaan ovat sellaiset

kohdat, joissa pelkän pronominin käyttäminen ei olisi onnistunut, sillä kyseisessä kohdassa on jo käytetty samaa pronominia.

Tutkimuksen tarkoituksena on siis selvittää, ovatko konekääntimien tekemät ratkaisut oikein viittaussuhteiden kannalta. Tarkastelen myös sitä, mitkä ovat konekääntimien yleisimmin käyttämät käännökset. Lisäksi selvitän, onko eri sanojen ja luokkien välillä eroja. Oletan, että luokan kaksi viittaussuhteet onnistuvat konekääntimiltä, mutta ensimmäisen luokan kohdalla voi olla ongelmia. Toisen luokan kohdissa viittauksen kanssa ei pitäisi olla sekaannusta, joten sen kääntämisen pitäisi luonnistua konekääntimiltä hyvin. Kolmannen luokan kohdalla viittaukset ovat luultavasti oikein, mutta ne eivät ole välttämättä kovinkaan selviä. Tämän lisäksi vertailen myös konekääntimiä keskenään, ja selvitän, mikä niistä onnistuu parhaiten pronomien kääntämisessä.

Työn rakenne on seuraavanlainen. Ensin käsittelen luvussa 2 arvioivaa tutkimusta sekä konekääntimiä ja niiden toimintaa. Esittelen lyhyesti sääntöpohjaiset, tilastolliset ja neuroverkkokääntimet sekä tutkimuksessa käytetyt kääntimet: Google Translate, DeepL Translate ja ChatGPT. Samassa luvussa käydään myös läpi ranskan ja suomen persoonapronomineja. Sen jälkeen esittelen aineiston ja kuvaan metodin luvussa 3. Tässä luvussa käsitellään myös tutkimuksessa käytetty arviointimalli sekä kuvaillaan konekääntimiä. Lopuksi esittelen analyysin ja tulokset luvussa 4 tilastojen ja esimerkkien avulla. Yhteenveto ja johtopäätökset ovat luvussa 5. Luvussa 6 käydään vielä läpi tutkimusta kokonaisuutena.

2 Teoreettinen viitekehys

Teoriaosiossa tarkastelen ensin arvioivaa tutkimusta (2.1), konekääntimiä ja niiden toimintaa (luku 2.2) sekä tutkimuksessa käytettyjä ohjelmia (luku 2.3). Tämän jälkeen kuvaan tutkimuksen kannalta oleellisia ranskan ja suomen pronomineja luvuissa 2.4 ja 2.5.

2.1 Arvioiva tutkimus

Teknologia kehittyi tunnetusti nopeasti, eikä käännösteknologia ole poikkeus. Jotta teknologiaa voidaan kehittää, täytyy tietää, mitä ongelmia siinä on. Tätä varten tarvitaan tutkimusta, joka arvioi nykyisten teknologioiden ja niiden tuotteiden laatua. Kuten Doherty (2019: 340) aihetta käsittelevässä artikkelissaan mainitseekin, käännösteknologian kehitys kulkee käsi kädessä sen arvioinnin kanssa. Käännösteknologiaan kuuluvat sekä varsinaiset konekääntimet että tietokoneavusteisen kääntämisen (CAT: *Computer-assisted translation*) työkalut. Näiden molempien kehittämiseen tarvitaan tutkimusta. Tutkimusta voivat tehdä muun muassa tutkijat ja teknologioiden kehittäjät. Kaikkien päällimmäinen tarkoitus on kuitenkin arvioida sitä, miten kyseinen teknologia selviytyy tehtävästä, jota varten se on suunniteltu. (Doherty, 2019: 339).

Siispä konekäännintä arvioidessa keskitytään yleensä sen tuottamaan tekstiin. Konekääntimen tarkoitus on tuottaa automaattisesti käännös sille annetusta tekstistä. Toki konekääntimestä voidaan arvioida muitakin seikkoja, kuten sen tehokkuutta tai sen käytettävyyttä. Yleinen tutkimuksen kohde on kuitenkin kääntimen tuottama teksti, jota voidaan arvioida monipuolisesti eri kriteerien mukaisesti. Yleisiä ihmisen suorittamia käännöksestä tutkittavia asioita ovat muun muassa seuraavat: tarkkuus (*accuracy*), sujuvuus (*fluency*), luettavuus (*readability*) ja ymmärrettävyys (*comprehensibility*) (Doherty, 2019: 341–342).

Tässä tutkimuksessa keskitytään edellä esitellyistä kriteereistä tarkkuuteen, kun arvioidaan sitä, ovatko konekäännösten viittaukset oikein. Viittausten selkeyden arvioinnin puolestaan voidaan ajatella liittyvän sujuvuuteen. Selvät viittaukset mahdollistavat keskeytymättömän lukemisen, kun lukijan ei tarvitse pysähtyä miettimään sitä, kenestä tai mistä asiasta tekstissä milloinkin on kyse. Toisaalta tämä voi liittyä myös luettavuuteen.

Kaiken kaikkiaan arvioiva tutkimus sopii mielestäni erityisen hyvin konekäännösten tutkimiseen, sillä tarkoituksena on kehittää konekääntimistä entistä parempia. Arvioinnin avulla saadaan selville, millaisissa asioissa koneilla on hankaluuksia. Lisäksi eri

konekääntimiä voidaan vertailla keskenään, kun niiden tuottamia käännöksiä arvioidaan johdonmukaisesti. Arvioinnissa käytetään usein hyväksi jonkinlaista virheluokittelua, jonka avulla voidaan antaa pisteitä tai muuten laskea lukuja, joita on helppo vertailla keskenään. Tällaisen tutkimuksen ei kuitenkaan tarvitse olla pelkästään määrällinen, vaan siinä voidaan myös tarkastella laadullisesti aineistossa esiintyviä ilmiöitä. Esimerkiksi usein toistuvat ilmiöt ovat mielenkiintoisia, ja niitä voidaan tarkastella omina kohtinaan.

Omassa tutkimuksessani keskitytään konekääntimien arviointiin ja tarkoituksena on tutkia pronominien kääntymistä ranskasta ja suomeen. Arvioinnissa käytetään arviointimallia, jonka avulla saadaan laskettua lukuja, joita voidaan vertailla keskenään. Tutkimuksen päätteeksi on siis mahdollista määritellä, mikä kääntimistä toimi parhaiten tämän aineiston kohdalla. Tästä koostuu tutkimuksen määrällinen osuus. Tutkimuksen aineistosta voidaan kuitenkin myös tarkastella siinä usein toistuvia ilmiöitä. Tämän laadullisen puolen ansiosta voidaan lisäksi yrittää tehdä päätelmiä siitä, miten konekääntimet toimivat. Tutkimuksessa olisi siis mahdollista keskittyä myös tällaiseen ilmiötä selittävään suuntaukseen, mutta tämän tutkimuksen puitteissa sellainen ei ole mahdollista. Tämä kuitenkin osoittaa, että arvioiva tutkimussuuntaus avaa ovia myös muiden suuntausten käyttöön, eikä tarkoituksena ole etsiä virheitä, jotta voitaisiin kritisoida tutkimuksen kohteita. Tarkoituksena on saada tietoa siitä, mihin arvioiva tutkimus soveltuu hyvin. Tämän tiedon avulla voidaan myös kehittää teknologioista entistä parempia.

2.2 Konekääntimet

Seuraavissa alaluvuissa esittelen konekääntimiä ja niiden kehitystä yleisesti. Käsittelen ensin sääntöpohjaiset kääntimet, sitten tilastolliset kääntimet ja lopuksi neuroverkkokääntimet, joilla on saatu aikaan aikaisempia lupaavampia tuloksia. Sääntöpohjaisten ja tilastollisten kääntimien esittely on lyhyt, koska tässä tutkimuksessa ei käytetä niihin pohjautuvia kääntimiä. Tässä yhteydessä on myös pieni katsaus kontekstiin, ja sen vaikutukseen konekääntämisessä. Tämän jälkeen esittelen alaluvussa 2.3 tässä tutkimuksessa käytetyt kääntimet, jotka ovat siis varsinaiset konekääntimet Google Translate ja DeepL Translator sekä tekoäly ChatGPT, jota voidaan käyttää myös kääntämiseen.

Konekääntimet ovat käännösteknologian apuvälineitä, joilla voidaan kääntää tekstiä automaattisesti. Niitä voidaan käyttää erilaisissa tarkoituksissa. Ammattimaisen käytön lisäksi niille voi olla käyttöä myös arkipäiväisissä tilanteissa. Konekääntimien tuottama teksti ei ole täydellistä, mutta sen avulla voi saada selville pääkohdat tekstistä, joka on kirjoitettu

sellaisella kielellä, jota ei itse osaa. Lisäksi konekääntimiä on vapaasti kaikkien saatavilla, ja ne ovatkin monille tuttuja. Esimerkiksi Google Translate on varmasti useimmille tuttu. Konekääntimiin voi törmätä myös muilla verkkosivuilla, joihin käännin on laitettu valmiiksi mukaan. Tällöin käyttäjän ei tarvitse edes avata käännintä erikseen, vaan tekstin voi kääntää suoraan sivulla. Aikaisemmin konekääntimien tuottama teksti oli lähinnä huvittavaa, mutta nykyään niiden taso on parantunut huomattavasti, ja konekääntimien käyttäminen voi olla hyödyllistä tietyissä tapauksissa.

Tällaisen arkikäytön lisäksi konekääntimiä voidaan hyödyntää myös ammattimaisessa kääntämisessä. Tällöin kääntimen antama teksti toimii pohjana, johon kääntäjä tekee korjauksia. Tällaista toimintaa kutsutaan yleisesti jälkieditoinniksi. Konekääntimien kehityksen myötä myös kääntäjien työnkuva on muuttunut ja tulee muuttumaan edelleen. Kääntäjän työ saattaa koostua yhä enemmän jälkieditoinnista kääntämisen sijaan. Sirkku Latomaan ja Mary Nurmisen artikkelissa (Latomaan & Nurminen, 2022) kerrotaankin, että vuoden 2018 eurooppalaisten kielipalvelualan yhdistysten tekemän kyselyn mukaan puolet vastaajista käytti konekäännöksiä työssään. Erityisesti neuroverkkokääntimien yleistyminen on parantanut kääntimien laatua, mistä johtuen niitä voi olla järkevää käyttää kääntämisen apuna. Konekääntimien tuottamaa tekstiä ei voi ainakaan vielä julkaista sellaisenaan, vaan kääntäjän tekemä jälkieditointi on välttämätöntä. Kuten Maarit Koponen (Koponen, 2017) huomauttaakin, teksti voi joissakin tapauksissa näyttää kieliopillisesti oikealta ja ihmisen kirjoittamalta, mutta asiasisältö saattaa olla väärin. Siispä kääntäjän tulee olla tarkkana, ja tarkastaa teksti huolellisesti. Konekääntimien käyttäminen on siis jo nyt yleistä käännoäsalalla, joten niiden kehittäminen edelleen on tärkeää. Mitä paremmin konekääntimet kääntävät, sitä vähemmän aikaa kääntäjällä menee konekäännösten tarkastamiseen ja korjaamiseen.

2.2.1 Sääntöpohjaiset konekääntimet

Kuten nimestäkin käy ilmi, sääntöpohjaiset konekääntimet (*rule-based machine translation*) perustuvat sääntöihin. Konekääntimet olivat enimmäkseen sääntöpohjaisia vielä 1980-luvun loppupuolella. Sääntöpohjaiset kääntimet käyttävät kääntämiseen sanakirjoja ja kielioppisääntöjä. Sääntöjen perustana toimivat muun muassa lähtö- ja tulokielien morfologia, syntaksi ja semantiikka. (Bai–Yu, 2023: 193). Vaikka näitä kääntimiä käytettiin pitkään, oli niissä omat ongelmansa. Ongelmat juontuivat erityisesti kielten sekä kääntämisen monimutkaisuudesta. Lisäksi kääntimien tekeminen oli työlästä ja kallista. (Bai–Yu, 2023: 199).

2.2.2 Tilastolliset konekääntimet

Tilastolliset konekääntimet (*statistical machine translation*) puolestaan toimivat hyvin erilaisella toimintaperiaatteella kuin sääntöpohjaiset konekääntimet. Ne perustuvat todennäköisyyksiin ja käyttävät apunaan rinnakkaistekstejä. Tilastollisia konekääntimiä yritettiin kehittää jo vuonna 1949, mutta silloin kehityksen esteenä oli tietokoneiden tehottomuus. Tietokoneiden kehityksen myötä myös tilastollisia konekääntimiä alettiin kehittää uudestaan, ja niitä on käytetty vuodesta 1991 lähtien. Toinen tärkeä ero sääntöpohjaisten ja tilastollisten konekääntimien välillä on se, että tilastolliset konekääntimet eivät ole riippuvaisia kieliparista. (Bai–Yu, 2023: 204). Tilastollisissa konekääntimissä on kuitenkin myös omat ongelmansa, mistä johtuen niiden käyttö on jäänyt vähäisemmäksi, kun paremmin toimivia konekääntimiä on saatu kehitettyä.

2.2.3 Neuroverkkokääntimet

Neuroverkkokääntimet perustuvat tekoälyyn ja syväoppimiseen (*deep learning*). Näiden kääntimien tekemiseen ei käytetä mitään eksplisiittistä käänösprosessin mallia, vaan niitä koulutetaan suurella määrällä dataa, joka koostuu käänöskorpuksista.

Neuroverkkokääntimissä käytetään apuna vektoreita (Liu–Li, 2023: 289–290). Neuroverkon toiminta perustuu siihen, että sille annetaan useita syötteitä, joiden perusteella se pyrkii muodostamaan oikean tuloksen (Koehn, 2020: 67).

Neuroverkkokääntimien nimi viittaa aivojen neuroneihin eli hermosoluihin. (Koehn, 2020: 30) Konekääntimissä käytetyt keinotekoiset neuronit muistuttavat näitä aivojen neuroneja, mutta niiden välillä on kuitenkin suuria eroja, jotka liittyvät esimerkiksi niiden keskinäiseen rakenteeseen ja oppimistapoihin. Keinotekoiset neuronit tarvitsevat oppimiseen valvottua kouluttamista. Näiden suurten erojen takia jotkut tutkijat eivät käytä kyseistä termiä, vaan suosivat termiä syväoppiminen. (Koehn, 2020: 31)

Kääntimien osaaminen kehittyi sen mukaan, millaisilla teksteillä niitä on koulutettu. Kouluttamiseen tarvitaan paljon tekstimateriaalia, mutta joissakin kielipareissa sitä ei ole tarpeeksi. Tällä hetkellä toinen kieliparin kielistä on yleensä englanti, mistä johtuen neuroverkkokääntimien tuottamat tekstit eivät ole välttämättä yhtä laadukkaita muissa kielipareissa. (Koponen, 2017). Eri kääntimien välillä voi siis olla koulutusmateriaalista johtuen suuria eroja, vaikka niiden toimintaperiaate olisikin sama. Tämän takia on tärkeää valita sopiva käännin tekstilajin ja tekstin aiheen perusteella.

Ensimmäiset neuroverkkomallit (*perceptrons*) olivat yksikerroksisia, ja niitä kehitettiin jo 1940- ja 1950-luvuilla. Ne eivät kuitenkaan toimineet niin hyvin kuin oli toivottu, ja neuroverkkojen kehittäminen keskeytyi joksikin aikaa, vaikka monikerroksisia neuroverkkoja kehitettiinkin. (Koehn, 2020: 31–32) Monikerroksisten neuroverkkojen kehitys kuitenkin edistyi, ja termiksi vaihdettiin syväoppiminen. Tutkijat saivat aikaan huomattavasti entistä parempia tuloksia useilla eri osa-alueilla. (Koehn, 2020: 33)

Konekääntäminen on yksi alue, joka on kiinnostanut tutkijoita jo vuosikymmeniä. Myös neuroverkkoja yritettiin hyödyntää konekääntämisen kehittämisessä jo 1980- ja 1990-luvuilla. Jotkin aiemmista malleista ovat hyvin samanlaisia nykyään käytettävien mallien kanssa, mutta siihen aikaan niiden kouluttamiseen ei ollut olemassa tarpeeksi tehokkaita tietokoneita. Tilastollisilla konekääntimillä saatiin kuitenkin aikaan parempia tuloksia, ja myöhemmin sitä alettiin yhdistää neuroverkkokääntimiin. Tilastolliset kääntimet kuitenkin jäivät pois, ja alettiin kehittää pelkästään neuroverkkoihin perustuvia kääntimiä. Neuroverkkokääntimistä saatiin kehitettyä nopeasti parempia kuin muista kääntimistä, ja nykyään keskitytään enimmäkseen niiden kehittämiseen (Koehn, 2020: 39). Tämä käy ilmi myös tämän tutkimuksen konekääntimistä, jotka kaikki hyödyntävät neuroverkkoja.

2.2.4 Konteksti

Pronominien kääntämisessä on haasteita myös muissa kuin tämän tutkimuksen aiheena olevissa tapauksissa. Eräässä tutkimuksessa (Vanmassenhove et al., 2018) tutkittiin neuroverkkokääntimiä, joiden koulutusvaiheessa oli merkitty puhujan sukupuoli, jotta konekääntäminen voisi kääntää morfologisesti oikein. Morfologisesti väärät muodot vaikuttavat käännöksen sujuvuuteen ja soveltuvuuteen (Vanmassenhove et al., 2018: 3003). Tulosten mukaan tämän mallin avulla voidaan saada parempia tuloksia, mutta ne eivät olleet aina johdonmukaisia. (Vanmassenhove et al., 2018: 3007). Tutkimuksen mukaan tekstin kirjoittajan sukupuolta ei välttämättä pystytä selvittämään pelkästään tekstistä. Ihmiskääntäjä pystyy käyttämään sukupuolen selvittämiseen kontekstia hyväkseen, mutta konekääntimet eivät yleensä pysty tähän. Konekääntimet käyttävät hyväkseen niiden kouluttamiseen käytettyä dataa, ja valitsevat tilastollisesti todennäköisimmän vaihtoehdon. Kontekstin puuttumiseen vaikuttaa erityisesti se, että konekääntimet kääntävät tekstiä virke kerrallaan. (Vanmassenhove et al., 2018: 3003).

Wong et al. (2020) osoittavat tutkimuksessaan, että tulevan kontekstin hyödyntämistä ei pidä väheksyä. Kontekstiriippuvaiset neuroverkkokonekääntimet (*context-dependent NMT*) ovat

tutkimusten mukaan parantaneet muun muassa pronominiinien kääntämistä. Kirjoittajat kuitenkin huomauttavat, että näissä ei ole otettu huomioon tulevien virkkeiden tuomaa kontekstia. (Wong et al., 2020: Introduction). Tutkimuksen tulokset osoittavat, että tulevan kontekstin huomioiminen tekee käännöksestä yhtä hyvän tai paikoitellen jopa paremman kuin edeltävän kontekstin käyttäminen. (Wong et al., 2020: Conclusions).

Kontekstin vaikutus pronominiinien kääntämiseen vaihtelee kuitenkin luultavasti paljon sen mukaan, mistä pronomineista on kyse. Esimerkiksi ensimmäisen persoonan pronominit pysyvät samana koko tekstissä, koska niillä viitataan tekstin kirjoittajaan tai kertojaan. Myös toisen persoonan pronomineissa voidaan olettaa esiintyvän tällaista jatkuvuutta. Sen sijaan kolmannen persoonan pronominiinien kohdalla viittauksen kohde voi vaihtua usein ja nopeasti tekstin sisällä. Tästä johtuen on mahdollista, että tuleva konteksti ei välttämättä paranna konekääntimien tuloksia näiden pronominiinien kohdalla. Viitattavaan asiaan tai henkilöön viitataan yleensä ensin nimellä tai jollakin sitä kuvaavalla sanalla, ennen kuin voidaan käyttää pronominia. Edeltävällä kontekstilla saattaisikin siis olla suurempi merkitys.

2.3 Konekääntimet

Seuraavissa alaluvuissa esitellään lyhyesti seuraavat konekääntimet: Google Translate, DeepL Translator ja ChatGPT.

2.3.1 Google Translate

Google Translate on Googlen kehittämä konekääntin (<https://translate.google.fi/?hl=fi>). Se julkaistiin jo vuonna 2006, ja aluksi se perustui tilastolliseen kääntämiseen. Kymmenen vuotta myöhemmin vuonna 2016 se aloitti siirtymävaiheen neuroverkkokääntimeksi, ja alkoi kääntää kokonaisia lauseita. (Sommerlad, 2021). Vuonna 2022 Google Translateen lisättiin 24 uutta kieltä, joten sillä pystyy nykyään kääntämään 133 eri kieltä (Caswell, 2022).

2.3.2 DeepL Translator

DeepL Translator (<https://www.deepl.com/translator>) on DeepL SE:n vuonna 2017 julkaisema konekääntin. Julkaisuvaiheessa se toimi seitsemän eri kielen välillä. Nämä kielet olivat englanti, saksa, ranska, espanja, italia, puola ja hollanti. Suomen kieli lisättiin valikoimaan vuonna 2021 usean muun eurooppalaisen kielen kanssa. (DeepL, 2023b).

2.3.3 ChatGPT

ChatGPT (<https://chat.openai.com/auth/login>) on OpenAI:n vuoden 2022 lopulla julkaisema ohjelma, joka perustuu tekoälyyn (Marr, 2023). Tekoälyn pyrkimyksenä on matkia ihmisen tapaa havaita, oppia, ratkoa ongelmia ja tehdä päätöksiä. Tekoälyä voidaan soveltaa lukuisilla eri osa-alueilla, ja niiden joukossa on myös kääntäminen. (Liu–Li, 2023: 280). ChatGPT:tä ei ole kehitetty varsinaisesti kääntämistä varten, mutta se on yksi asioista, joita se pystyy tekemään. ChatGPT on tekstipohjainen, ja se rakentuu keskustelumallisesti. Käyttäjä esittää kysymyksen, johon tekoäly vastaa. Tätä keskustelua voidaan jatkaa, jolloin ohjelman antamat vastaukset voivat muuttua tai tarkentua käyttäjän antamien tietojen mukaan. On tärkeää huomata, että ChatGPT:n tuottamat käännökset voivat olla erilaisia eri kerroilla, vaikka kysymys olisikin sama (Jiao et al., 2023: Limitations). Tämän tutkimuksen puitteissa ei kuitenkaan ole mahdollista tutkia useampaa käännöstä samalta kääntimeltä. Muilla tutkimuskerroilla voidaan myös saada hieman erilaisia tuloksia, jos tutkimus toistettaisiin.

2.4 Ranskan pronominit

Tutkimuksen kohteet voidaan jakaa kahteen eri ryhmään. Ensimmäisessä ryhmässä ovat persoonapronominit, jotka on kuvattu alla olevassa taulukossa 1. Persoonapronominien järjestelmä muistuttaa paljon suomen kielen järjestelmää, mutta eroja on kolmannen persoonan kohdalla, jossa pronomineilla on kaksi eri muotoa. Kolmannessa persoonassa on eri pronominit maskuliinille ja feminiinille sekä yksikössä (*il*, *elle*) että monikossa (*ils*, *elles*). Muut pronominit ovat yksikön ensimmäinen (*je*) ja toinen (*tu*), sekä monikon ensimmäinen (*nous*) ja toinen (*vous*) persoona. Näitä muita pronomineja ei kuitenkaan käsitellä ollenkaan tässä tutkimuksessa. Tässä kuvatut pronominit toimivat lauseessa subjektina. (Chevalier, 1974: 228).

Taulukko 1 Ranskan kielen persoonapronominit subjektina (Chevalier, 1974: 228)

Persoona	Yksikkö	Monikko
1.	je	nous
2.	tu	vous
3. maskuliini	il	ils
3. feminiini	elle	elles

Toinen tutkimuksessa tarkasteltava ryhmä koostuu seuraavista sanoista *l'*, *la*, *le* ja *les*. Ne on kuvattu alla olevassa taulukossa. Näitä pronomineja käytetään silloin, kun tarvitaan suoraa

objektia. Yksikössä sekä maskuliinilla että feminiinillä on omat muotonsa: *le* ja *la*. Vokaalin edellä näistä molemmista käytetään muotoa *l'*. Monikossa puolestaan erottelua sukupuolen mukaan ei ole, vaan molemmista käytetään muotoa *les*. (Chevalier, 1974: 228)

Taulukko 2 Ranskan kielen persoonapronominit suorana objektina (Chevalier, 1974: 228)

	Yksikkö	Monikko
Maskuliini	le/l'	les
Feminiini	la/l'	les

Persoonapronominien jakaminen maskuliiniin ja feminiiniin johtaa siihen, että niitä voidaan käyttää erottelemaan viittauksen kohdetta. Kuten edellä olevista taulukoista nähdään, tämä on mahdollista sekä pronominin ollessa subjektina että objektina. Siispä tutkimukseen on valittu mukaan juuri nämä pronominit.

Muiden persoonapronominien rinnalle on syntynyt uusi pronomini *iel* (monikossa *iels*). Tämä pronomini on kolmannen persoonan pronomini, jota voidaan käyttää viittamaan kehen tahansa sukupuolesta riippumatta (Dico en ligne Le Robert, n.d.) Ranskan kielessä on aikaisemminkin ollut vastaavia sanoja, mutta ne ovat jääneet pois käytöstä. Esimerkiksi pronominia *el* (joskus muodossa *al*) on käytetty muinaisranskassa (Marchello-Nizia, 1989: 174). *Iel*-sanalle on ehdotettu muitakin vaihtoehtoja, kuten *al*, *ol*, *ul*, *ille* tai *yol* (Elmiger, 2022: 2). Tämä pronomini on lähinnä suomen kielen sukupuolesta riippumatonta *hän*-pronominia. Sen käyttö ei kuitenkaan ole yleistä (Leblanc, 2022: 3), eikä se ole korvannut *il*- ja *elle*-pronomineja. Sitä voidaan käyttää viittaamaan non-binäärisiin henkilöihin tai sellaisissa tapauksissa, joissa viitattavan henkilön identiteetistä ei ole tietoa tai sillä ei ole merkitystä (Elmiger, 2022: 6). Ranskan kielen persoonapronomineissa on siis tapahtunut viime aikoina muutoksia, eikä ole selvää, miten niitä tullaan jatkossa käyttämään. Tätä kehitystä onkin mielenkiintoista seurata myös tulevaisuudessa.

2.5 Suomen pronominit

2.5.1 Persoonapronominit

Persoonapronominit ovat pronomineja, joilla voidaan viitata joko puhetilanteen osapuoliin tai puheenalaisiin henkilöihin (VISK § 716). Kolmannen persoonan pronominit kuuluvat jälkimmäiseen kategoriaan. Kaiken kaikkiaan suomen kielessä on kuusi persoonapronominia, jotka voidaan jakaa persoonan mukaan. Toinen jako tapahtuu yksikön ja monikon välillä.

Tässä tutkimuksessa käsitellään kolmannen persoonan pronomineja sekä yksikössä että monikossa. Ne ovat anaforis-deiktisiä, mikä tarkoittaa sitä, että niiden tarkoite ei vaihdu puhujasta riippuen toisin kuin ensimmäisen ja toisen persoonan kohdalla (VISK § 716). Toisin kuin ranskan kielessä, suomessa ei käytetä eri pronomineja tarkoitteen sukupuolen mukaan. Taulukossa 3 on esitetty suomen kielen persoonapronominit.

Taulukko 3 Suomen kielen persoonapronominit (VISK § 716.)

Persoonana	Yksikkö	Monikko
1.	minä	me
2.	sinä	te
3.	hän	he

2.5.2 Demonstratiivipronominit

Aineistossa esiintyy persoonapronominien lisäksi myös joitakin demonstratiivipronomineja. Demonstratiivipronominit (taulukko 4) viittaavat tarkoitteisiin, jotka on jo mainittu puhetilanteessa tai tekstissä tai jotka havaitaan puhetilanteessa. Demonstratiivipronominien valintaan vaikuttaa huomiopiiri. (VISK § 720) Huomiopiiri tarkoittaa konkreettista tai abstraktia aluetta, joka kullekin keskustelun osapuolelle kuuluu (VISK Määritelmät). Pronominia *tämä* käytetään silloin, kun tarkoite kuuluu puhujan huomiopiiriin. Vastaavasti pronominia *se* käytetään, jos tarkoite kuuluu puhuteltavan huomiopiiriin. *Tuo* on käytössä silloin, kun tarkoite ei kuulu kummankaan huomiopiiriin. (VISK § 720).

Taulukko 4 Demonstratiivipronominit (VISK - § 720)

Yksikkö	Monikko
tämä	nämä
tuo	nuo
se	ne

Demonstratiivipronomineja voidaan käyttää itsenäisesti tai substantiivina määrittämässä. Pronominia *tämä* voidaan käyttää viittamaan henkilöön, joka on juuri mainittu tekstissä. (VISK § 720). Jos tekstissä on kaksi substantiivilauseketta, voi olla epäselvää kumpaan viitataan. Persoonapronominit *hän* tai *he* viittaavat yleensä etualaiseen lausekkeeseen. Etualaisuudella tarkoitetaan ”saman tai hallitsevan lauseen subjektin asemaa tai muuta teemapaikkaisuutta”. Tekstin merkitys voi selventää viittaussuhteen, mutta taka-alaiseen

tarkoitteeseen voidaan myös viitata pronomineilla *tämä* tai *nämä*. Lisäksi yleismerkityksistä substantiivia tai adjektiivia (esim. *tyttö*, *mies*) voidaan käyttää pronominin tavoin. (VISK § 1432).

3 Aineisto ja menetelmät

Tässä luvussa kuvaillaan tutkimuksen aineisto (3.1), metodi (3.2), arviointimalli (3.3) sekä konekääntimet (3.4).

3.1 Aineisto

Tutkimuksen aineisto koostuu ranskankielisestä romaanista ja sen suomennoksista.

Romaanista on valittu noin 60 sivun pituinen osio tarkasteluun. Suomennoksia on yhteensä neljä. Näistä yksi on ihmiskääntäjän tekemä, ja se on julkaistu painettuna kirjana. Kolme muuta ovat tätä tutkimusta varten koneilla muodostettuja. Konekääntimistä on valittu mukaan Google Translate ja DeepL Translator, ja teksti on käännetty myös ChatGPT-tekoälyllä.

Lähtöteoksena on Katherine Pancolin vuonna 2008 julkaistu teos *La Valse lente des tortues* (kustantanut Albin Michel, 1. painos). Romaani on trilogian toinen osa, ja se on jatkoa teokselle *Les yeux jaunes des crocodiles*, joka voitti vuonna 2006 Prix Maison de la Presse -palkinnon (*Prix : Des Maisons de la presse - Roman*, n.d.). Katherine Pancol syntyi Marokossa, mutta muutti Ranskaan viisivuotiaana. Hänen ensimmäinen romaaninsa *Moi d'abord* julkaistiin vuonna 1979, minkä jälkeen hän on kirjoittanut lukuisia muita romaaneja. (*Biographie de Katherine Pancol*, n.d.).

Tutkimuksen aineistona olevassa kirjassa seurataan Joséphine-nimisen naisen elämää keskittyen erityisesti hänen ihmissuhteisiinsa. Ihmissuhteisiin keskittyminen onkin yksi syy, minkä takia valitsin tutkimukseen juuri tämän kirjan. Tutkimuksen aiheena ovat pronominit, joten tutkimuksen kannalta sopiva teos on sellainen, jossa niitä käytetään paljon. Koska tutkimuksessa keskitytään kolmannen persoonan pronomineihin, oli luontevampaa valita kirja, joka on kirjoitettu yksikön kolmannessa persoonassa ensimmäisen persoonan sijasta. Muita oleellisia henkilöihahmoja ovat päähenkilön tyttäret Hortense ja Zoé sekä Joséphin sisaren aviomies Philippe.

Romaanin on suomentanut Lotta Toivanen, ja se ilmestyi vuonna 2012 nimellä *Kilpikonnién hidas valssi* (Bazar). Tutkimuksessa on käytetty kirjan ensimmäistä painosta. Tätä ihmisen tekemää käännoästä on käytetty tutkimuksessa eräänlaisena pohjana, jonka avulla on esimerkiksi rajattu tutkittavien kohtien valintaa sekä muodostettu luokittelu. Toivanen on suomentanut yli 60 teosta (*Lotta Toivanen - Koneen Säätio*, n.d.), ja on saanut työstään erilaisia palkintoja, joiden joukossa esimerkiksi Gummeruksen tunnustuspalkinto

”ansioituneesta työstä kaunokirjallisuuden suomentajana” (*Palkinnot kirjailija Markku Turuselle ja suomentaja Lotta Toivaselle*, 2011). Kyseessä on siis arvostettu suomentaja, jolla on paljon kokemusta kaunokirjallisuuden kääntämisestä. Siispä hänen tekemäänsä käännöstä voi mielestäni pitää tietynlaisena mallina hyvästä käännöksestä.

Muut aineistoon kuuluvat suomennokset ovat koneiden tekemiä. Ensimmäinen niistä on Googlen käännöspalvelun Google Translaten tekemä. Toinen konekääntimien tekemistä käännöksistä on DeepL Translatorin tekemä suomennos. Neljäs ja viimeinen suomennos ei ole varsinaisen konekääntimen tekemä, vaan se on muodostettu ChatGPT:n avulla. Viittaan jatkossa kaikkiin näihin kolmeen ohjelmaan konekääntiminä. Google Translaten käännös on tehty 28. marraskuuta 2023, DeepL Translatorin käännös 28–29. marraskuuta 2023 ja ChatGPT:n käännös 29. marraskuuta.

Tarkempi tutkimuskohta on romaanin toisen osion alusta. Romaanissa ei ole käytetty varsinaisia lukuja, vaan se on jaettu viiteen pidempään osioon. Tutkimuskohdan pituus on noin 60 sivua. Valitsin juuri tämän osion sen takia, että siinä mukaan tulee enemmän miespuolisia henkilöitä. Kirjan päähenkilö on nainen, ja ensimmäisessä osiossa puhutaan paljon hänestä ja hänen tyttäristään. Tästä johtuen tutkimusta varten tarvittavaa ilmiötä esiintyy enemmän vasta tässä toisessa osiossa. Jotta pronomineilla voidaan erotella mies- ja naispuolisia henkilöitä toisistaan, täytyy molempiin viitata samassa kohtauksessa. Tämän takia en ottanut tutkittavaksi osioksi kirjan alkua.

Sekä romaanin alkuteos että ihmisen tekemä käännös ovat painettuja kirjoja. Jotta niitä voidaan tutkia, täytyy teokset skannata ja muuttaa digitaaliseen muotoon OCR-ohjelmalla (*Optical character recognition*). Näin tekstiä voidaan käsitellä tietokoneella. Käytin tähän OneDriven skannaustoimintoa, jolla voi samalla tehdä OCR-tekstintunnistuksen. Sen jälkeen etsin aineistosta tutkimuksen kannalta oleellisia pronomineja.

Tutkimuksen aiheena ovat kolmannen persoonan pronominit, joiden tarkempi kuvaus on esitetty aiemmin teorialuvussa. Tutkimuskohteena ovat siis yksikön kolmannen persoonan pronominit *il* ja *elle* sekä monikon kolmannen persoonan pronominit *ils* ja *elles*. Tämän lisäksi tutkin muutamia muita sanoja, joilla voidaan viitata ihmisiin tai asioihin. Näitä ovat *l'* (esiintyy ennen vokaalia), *le* (maskuliini), *la* (feminiini) ja *les* (monikko). *L'* on vokaalia ennen esiintyvä lyhennetty muoto pronomineista *le* ja *la*. Se on kuitenkin tutkimuksessa mukana omana kohtanaan, sillä siitä ei voi suoraan tietää kummasta pronominista se on

lyhennetty. Koska kyseisessä tutkimuksessa on tärkeässä osassa viitattavien henkilöiden tai asioiden sukupuoli, on mielestäni perusteltua käsitellä sitä omana kohtanaan.

Suurin osa aineiston viittauksista kohdistuu ihmisiin, mutta siinä on myös joitakin tapauksia, joissa viittauksen kohteena on jokin eloton asia. En ottanut tutkimukseen mukana omistusmuotoja, vaikka niidenkin kohdalla voitaisiin tutkia samaa ilmiötä, sillä niissäkin käytetään maskuliini- ja feminiinimuotoja. En kuitenkaan käsittele niitä tässä tutkimuksessa, sillä asiaa on mielestäni selkeämpää tutkia vain nyt mukana olevien pronomien avulla. Lisäksi aineistosta olisi voinut tulla liian laaja, jos myös omistusmuodot olisivat olleet mukana.

Aineistoa on rajattu siten, että mukaan on otettu vain sellaiset kohdat, joissa suomentaja on päättänyt käyttämään jotakin muuta kuin alkuteoksessa käytettyä vastaavaa pronomia. Tällaisissa kohdissa henkilöön tai asiaan on viitattu yleisimmin nimellä tai jollakin muulla sanalla, josta ilmenee, kenestä puhutaan. Monet tällaisista kohdista ovat sellaisia, joissa pelkän pronomien käyttäminen tekisi viittauksesta epäselvän. Suomennoksessa on kuitenkin myös useita kohtia, joissa pronomienkin käyttäminen toimisi viittaussuhteiden osalta. Tällöin on käytetty esimerkiksi henkilön nimeä pronomien sijasta, vaikka sekaannuksen vaaraa ei olisikaan. Tämä voi johtua siitä, että suomentaja on joko tietoisesti tai tiedostamattaan tehnyt viittauksista selkeämpiä, jolloin käännöksestä on tullut eksplisiittisempi kuin alkuteoksesta. Eksplisiittistäminen tarkoittaa sitä, että käännös on eksplisiittisempi eli yksityiskohtaisempi kuin alkuperäinen (Behrens, 2022: 102). Se onkin yksi kääntämisen universaaleista (Kuusi, 2011: 14), ja sitä näyttäisi esiintyvän myös tässä aineistossa. Tämä rajaus mahdollistaa siis sen, että voidaan tutkia vain kaikkien mielenkiintoisimpia kohtia, joissa konekääntimillä voidaan olettaa olevan ongelmia. Niinpä tutkimuksen aineistosta on jäänyt pois kaikki sellaiset kohdat, joissa pronomini kääntyy myös suomeksi yksinkertaisesti pronominilla. Lisäksi aineistosta on otettu pois joitakin kohtia. Näihin kuuluvat kohdat, joissa suomennoksessa on käytetty sanaa 'tämä', kohdan lauserakenne on muuttunut paljon tai viittaus on vaihdettu ihmisestä asiaan.

Virkkeet, joissa esiintyy tutkimuksen kannalta relevantteja pronomineja, siirrettiin Microsoft Excel-ohjelmaan (Microsoft 2023), mikä teki analyysin tekemisestä helpompaa. Loput suomennokset muodostettiin konekääntimillä, ja niistä poimittiin samaan Excel-taulukkoon kyseisten kohtien käännökset. ChatGPT:n toimintatapa on erilainen kuin varsinaisilla konekääntimillä, sillä sen kohdalla täytyy muodostaa kysymys, joka käynnistää

kääntämisprosessin. Tämän kysymyksen muoto vaikuttaa tuotettuun käännökseen, joten sen muotoilemisessa täytyy olla huolellinen. (Jiao ym., 2023). Tässä tutkimuksessa käytetty suomenkielinen kysymys oli seuraavanlainen: ”Käännä ranskasta suomeksi: x”. X:n paikalle tuli kulloinkin käännettävä osio. Päädyin käyttämään neutraalia kysymystä, josta tulee ilmi myös tekstin lähtökieli ja toivottu tulokieli. Kuten aiemmin mainittiinkin, vaikuttaa kysymys käännökseen. Tämän tutkimuksen puitteissa ei ole mahdollista tutkia kysymyksen merkitystä.

3.2 Metodi

Tutkittavat kohdat on luokiteltu kolmeen luokkaan. Luokat määrittyvät ihmisen tekemän suomennoksen mukaan. Ensimmäisessä luokassa ovat sellaiset kohdat, joissa viittaussuhteiden kanssa voi olla ongelmia, jos käytetään vain pronominia. Tässä ryhmässä on 130 kohtaa. Toisessa luokassa on kohdat, joissa pronominin käyttäminen ei tuottaisi ongelmia lukijan ymmärtämiselle. Niitä oli aineistossa 109. Tämän luokan kohdissa on siis luultavasti kyse eksplisiittistamisesta. Kolmas ja viimeinen luokka puolestaan on pieni, ja siihen kuuluu vain 10 kohtaa. Nämä kohdat ovat sellaisia, joissa suomennoksessa on jo käytetty samaa pronominia, joten sen käyttäminen uudelleen ei ole mahdollista. Näissä tapauksissa konekääntimillä ei luultavasti ole ongelmia viittaussuhteen kanssa, mutta viittauksen selvyydessä voi ilmetä vaikeuksia.

Konekääntimien kehityksessä käytetään yleensä tekstejä, joita niiden käyttäjät ovat niillä käännettäneet. Kehittämistä varten tekstit siis tallentuvat jonnekin, mikä voi olla ongelmallista tekijänoikeudellisista syistä. Tämän takia koko tekstiä ei ole käännetty konekääntimillä, vaan teksti on jaettu pieniin palasiin. Konteksti on kuitenkin tärkeä asia tutkittavan ilmiön kannalta, sillä pronominit viittaavat taaksepäin. Tästä syystä jokaisen tutkittavan virkkeen lisäksi on otettu mukaan kaksi sitä edeltävää virkettä. Tutkimuksessa on siis käännetty 249 kolmen virkkeen mittaista osiota. Näin mukaan saadaan edes jonkin verran kontekstia, vaikka se ei aina olekaan riittävän paljon. Lisäksi kaikki nämä kolmen virkkeen kohdat on käännetty yksitellen, jotta koneet eivät käsitelisi kaikkia kohtia yhtenä pidempänä tekstinä. On kuitenkin mahdollista, että ainakin ChatGPT oppii edellisistä virkkeistä, ja käyttää tätä aiempaa kontekstia hyväkseen kääntämisessä. Kaikki virkkeet käännettiin samalla käyttäjätunnuksella sulkematta välillä keskustelua. On kuitenkin hieman epäselvää, käyttävätkö kaikki kääntimet kontekstia ollenkaan hyväkseen. On myös mahdollista, että ne kääntävät vain yhden virkkeen kerrallaan, jolloin aiempaa kontekstia ei otettaisi ollenkaan

huomioon. Tämä vaikuttaisi huomattavasti tuloksiin, sillä ilman kontekstia on vaikeaa kääntää viittaukset oikein.

3.3 Arviointimalli

Käännösten arvioimista varten on kehitelty erilaisia arviointimalleja ja monet niistä sopivat varmasti myös konekäännösten arvioimiseen. Yksi tällaisista työkaluista on DQF-MQM-virheluokittelu. Tämä luokittelu perustuu kahteen aikaisempaan luokitteluun (Lommel, 2018: 124). Kyseessä on laaja ja monipuolinen väline virheanalyysin tekemiseen. Siinä on monta luokkaa, mutta käyttäjät voivat valita ne, joita tarvitsevat omassa tutkimuksessaan (Lommel, 2018: 125).

Tämän tutkimuksen tutkimuskysymykset ovat kuitenkin hyvin tarkat, joten tarvittavia luokkia olisi vain muutama. Kyseisen virheluokittelun kohdat eivät myöskään oikein sovi tähän tutkimukseen. Tästä syystä päädyinkin kehittelemään oman suppeamman ja juuri tähän tutkimukseen sopivan arviointimallin. Virheisiin perustavan luokittelun sijaan päädyin käyttämään kriteeriperustaista arviointimallia. Tutkimuksen tarkoituksena on tutkia pronomien kääntymistä konekääntimillä, joten oleellista on arvioida sitä, säilyykö viittaus oikeana. Lisäksi arvioin myös sitä, onko viittaus selvä vai aiheuttaako se lukijalle hämmennystä. Näiden kahden arviointikriteerin lisäksi arvioin sitä, onko tarkasteltava virke kääntynyt niin hyvin, että siitä voidaan arvioida kahta edellistä kohtaa. Jos käännös on epäonnistunut, ei siitä voida määritellä viittausta ja sen onnistuneisuutta. Edellä esitelty arviointimalli on kuvattuna myös alla olevassa taulukossa.

Taulukko 5 Arviointimalli

1.	Onko käännös niin onnistunut, että siitä voi määritellä viittauksen?	Kyllä	Ei
2.	Onko viittaus oikein?	Kyllä	Ei
3.	Onko viittaus selvä?	Kyllä	Ei

Luokittelun kohdista erityisen mielenkiintoinen on kolmas kohta, jossa tarkastellaan viittauksen selkeyttä, sillä epäselvä viittaus tekee lukemisesta vaikeaa ja voi johtaa lukijaa harhaan. Tutkimus perustuu nimenomaan ranskan ja suomen kielten eroihin pronomineissa. Ranskassa voidaan paremmin käyttää pronomineja erottelemaan viittauksen kohdetta, mutta suomessa pelkän pronominin käyttäminen voi aiheuttaa epäselvyyttä. Tästä johtuen

viittauksen selkeyden arvioiminen on mielestäni yksi tämän tutkimuksen kaikista oleellisimmista asioista.

3.4 Konekääntimet

Tätä tutkimusta varten on valittu kolme ohjelmaa, joilla tehdään käännökset aineistona olevasta tekstistä. Näistä kaksi ovat perinteisiä konekääntimiä (Google Translate ja DeepL Translator), ja yksi on tekoäly (ChatGPT). Valitsin tutkimukseen mukaan myös ChatGPT:n, koska sitä voidaan käyttää myös kääntimenä. Lisäksi siitä on puhuttu viime aikoina paljon, ja se voi muuttaa sitä, miten työtä tehdään tulevaisuudessa. Tässä vaiheessa on toki mahdotonta arvioida sen vaikutuksia tulevaisuuden työhön ja opiskeluun, mutta näyttäisi siltä, että sen käyttö tulee vain lisääntymään. Siispä se on mielestäni hyvin ajankohtainen, ja on myös mahdollista, että tulevaisuudessa kääntämisen apuna voidaan käyttää tekoälyä. Se eroaa varsinaisista konekääntimistä erityisesti siinä, että sille voidaan antaa ohjeita siitä, millainen käännöksen tulisi olla. Näin pystyttäisiin tarkentamaan jo ennen kääntämistä sitä, millaista käännöstä ollaan tekemässä. Lisäksi ChatGPT:tä voidaan pyytää muokkaamaan jo käännettyä tekstiä. Sopivalla kysymyksellä voidaan saada vaivattomasti käännös, joka vastaa enemmän juuri kyseiseen tilanteeseen tarvittavaa käännöstä.

Käännöksen muokkaaminen kysymyksillä on mahdotonta silloin, kun käytetään perinteisiä konekääntimiä, kuten tässä tutkimuksessa käytettyjä Google Translatea ja DeepL Translatoria. Tällaisia kääntimiä käytettäessä ei yleensä ole mahdollista antaa minkäänlaisia tarkennuksia, vaan käyttäjältä kysytään vain lähtö- ja tulokielet sekä lähtöteksti. DeepL Translatorin Pro-versiossa on kuitenkin mahdollisuus valita kahdesta vaihtoehdosta: *formal* ja *informal*. Tämä valinta vaikuttaa erityisesti pronomineihin, ja kyseinen toiminto on tarjolla yhdeksälle eri kielelle. (DeepL, 2023a). Tämä on kuitenkin vain yksi asia, johon käyttäjä voi vaikuttaa, kun taas ChatGPT:n kohdalla valittavat täsmennykset ovat paljon monipuolisempia. Ei edes ole olemassa tarkkaa tietoa siitä, minkälaisia täsmennyksiä sillä pystytään tekemään, sillä käyttäjä voi pyytää melkein mitä vain. Ei tietenkään ole takeita siitä, että tekoäly suorittaa kaiken, mitä pyydetään, ja laadustakaan ei voida olla varmoja.

4 Analyysi ja tulokset

Tässä luvussa käsitellään tutkimuksessa saatuja tuloksia. Aineiston avulla voidaan selvittää monia eri asioita ja laskea tilastoja. Kyseinen luku onkin jaettu alalukuihin. Alaluvussa 4.1 on kuvattu ihmisen tekemän käännöksen luokittelu, joka siis toimii pohjana joillekin muille analyysin osille. Alaluvussa 4.2 tarkastellaan sitä, ovatko konekääntimien viittaukset oikein vai väärin, ja alaluvussa 4.3 tutkitaan viittausten selkeyttä. Sen jälkeen tarkastellaan eroja sanojen välillä (4.4) ja luokkien välillä (4.5). Tämän jälkeen käydään läpi sekä ihmisen tekemän käännöksen että konekääntimien käännösten yleisimpiä käännöksiä alaluvussa 4.6. Sen jälkeen siirrytään tarkastelemaan joitakin esimerkkejä kääntimien tekemistä käännöksistä. Onnistuneita käännösratkaisuja esitellään luvussa 4.7 ja sen neljässä alaluvussa. Epäonnistuneita viittauksia puolestaan käsitellään luvussa 4.8, joka on jaettu kolmeen alalukuun. Alaluvussa 4.9 esitellään vielä lyhyt huomautus konekääntimien kontekstin käytöstä.

4.1 Ihmiskäännöksen kohtien luokittelu

Taulukko 6 Luokittelu ihmiskäännöksestä

Ryhmä	Määrä
1	130
2	109
3	10
Yhteensä	249

Yllä olevassa taulukossa näkyy, miten ihmiskäännöksessä tehdyt valinnat pronomiinien kääntämisestä jakautuvat ryhmiin. Luokittelun ensimmäisessä ryhmässä ovat sellaiset kohdat, joissa pelkän pronominin käyttäminen suomennoksessa voisi aiheuttaa sekaannusta siitä, kenestä tai mistä on kyse. Toisessa ryhmässä puolestaan ovat ne pronominit, joiden kohdalla vaaraa tällaisesta sekaannuksesta ei ole. Kolmannessa ryhmässä ovat sellaiset kohdat, joissa jonkin muun kuin pronominin käyttäminen on tarpeellista, sillä muuten yhdessä lauseessa olisi kaksi kertaa pronomini 'hän'. Kuten taulukosta nähdään, suurin osa kohdista kuuluu joko ryhmään 1 (130 kohtaa) tai 2 (109 kohtaa). Kolmannessa ryhmässä on vain kymmenen kohtaa. Tätä luokittelua hyödynnetään analyysin myöhemmässä vaiheessa, jossa tarkastellaan konekääntimien käännöksiä onnistuneisuutta tämän luokittelun ryhmien avulla. Kohtien luokittelu ryhmiin perustuu vain omaan arviooni, ja muut voisivat saada erilaisia tuloksia. Tämä luokittelu ei siis ole ainoa mahdollinen, mutta kaikkien kohtien luokittelussa on

käytetty samoja periaatteita. Tämä koskee myös muita tässä tutkimuksessa tehtyjä luokitteluja.

4.2 Viittaukset oikein vai väärin

Taulukossa 7 on kuvattu jokaisen kääntimen kohdalla erikseen se, ovatko viittaukset säilyneet oikein vai eivät. Lisäksi joukossa on joitakin kohtia, joissa koneen antama käänнос on ollut epäonnistunut, eikä siitä ole pystytty määrittelemään viittauksen onnistuneisuutta. Tällaiset kohdat on luokiteltu kohtaan 'mahdoton analysoida'. Jokaiselle kohdalle on myös laskettu prosenttiosuus, joka kuvaa sitä, kuinka monta prosenttia kyseinen luku on koko tutkittavan osion kohdista.

Taulukko 7 Viittaukset oikein vai väärin

	Google Translate	%	DeepL	%	ChatGPT	%
Viittaus oikein	244	98,0	244	98,4	245	98,4
Viittaus väärin	5	2,0	4	0,4	1	0,4
Mahdoton analysoida	0	0,0	1	1,2	3	1,2
Yhteensä	249	100,0	249	100,0	249	100,0

Kuten taulukosta käy ilmi, suurin osa tutkittavista kohdista on kääntynyt onnistuneesti, eli siis niin, että niistä voidaan määritellä viittauksen onnistuneisuus. Google Translaten käänносistä kaikki ovat tältä osin onnistuneita. DeepL:llä on yksi ja ChatGPT:llä kolme kohtaa, joissa viittausta ei voida tarkastella, eli analyysin tekeminen on mahdotonta. Tämän lisäksi suurin osa viittauksista on myös kääntynyt oikein. Seuraavassa alaluvussa (4.3) tarkastellaan erikseen sitä, ovatko onnistuneet viittaukset selviä vai eivät. Google Translaten käänносistä oikein oli 244 ja väärin 5 kohtaa. DeepL:n käänносistä puolestaan oikein oli 244 kohtaa ja väärin 4. ChatGPT:llä vastaavat luvut ovat 245 ja 1. Näiden tulosten perusteella voidaan siis sanoa, että kääntimet kääntävät vain harvoin niin, että viittaukset olisivat varsinaisesti väärin. Kriteerit viittauksen onnistumiseen olivat sellaiset, että riitti, että viittaus ei ole väärin. Tästä johtuen kohdat, joissa viittaus on oikein, eivät välttämättä ole selviä, ja näiden kahden arvioinnin tulosten välillä voi olla suuriakin eroja.

4.3 Viittaukset selviä vai epäselviä

Seuraavasta taulukosta nähdään, miten hyvin kääntimet ovat suoriutuneet, kun tarkastelun kohteena on viittauksen selkeys. Joissakin tapauksissa viittaus on ollut väärin tai koko virkkeen käänös on epäonnistunut. Nämä tapaukset on merkitty yhteiseen luokkaan 'mahdoton analysoida' taulukon kolmannelle riville. Oleellisimmat kohdat tässä taulukossa ovat kuitenkin ensimmäisellä ja toisella rivillä.

Taulukko 8 Viittausten selkeys

	Google Translate	%	DeepL	%	ChatGPT	%
Viittaus selvä	81	32,5	142	57,0	94	37,8
Viittaus epäselvä	163	65,5	102	41,0	151	60,6
Mahdoton analysoida	5	2,0	5	2,0	4	1,6
Yhteensä	249	100,0	249	100,0	249	100,0

Ensimmäisessä sarakkeessa on Google Translaten tulokset. Sen käännösten viittauksista selviä oli 81 (32,5 % kaikista sillä käännettyistä kohdista) ja epäselviä 163 (65,5 %). DeepL puolestaan sai viittaukset oikein huomattavasti useammin: 142 selvää kohtaa (57,0 %) ja 102 epäselvää (41,0 %). ChatGPT:llä vastaavat luvut ovat melko samanlaiset Google Translaten kanssa, mutta kuitenkin hieman paremmat: selviä 94 (37,8 %) ja epäselviä 151 (60,6 %).

4.4 Erot sanojen välillä

Alla olevaan taulukkoon on koottu kaikkien kääntimien tulokset yhteen. Siinä on eritelty jokainen tutkittava sana omaksi kohdaksi, ja siitä nähdään, kuinka monen viittaus on oikein. Niin kuin aiemmissakin taulukoissa, myös tässä on oma sarakkeensa niille kohdille, jotka eivät kääntyneet tarpeeksi hyvin. Tällaisia kohtia, joita ei voitu analysoida, oli yhteensä neljä. Kaikista viittauksista väärin oli vain kymmenen (1,3 %), ja oikein oli 733 (98,1 %).

Taulukko 9 Viittaukset oikein vai väärin – kaikki kääntimet yhdessä

	Väärin	%	Oikein	%	Mahdoton analysoida	%	Yhteensä	
Elle	6	1,5	390	98,5	0	0,0	396	100,0 %
Il	1	0,4	243	98,8	2	0,8	246	100,0 %
Elles	0	0,0	12	100,0	0	0,0	12	100,0 %
Ils	0	0,0	30	100,0	0	0,0	30	100,0 %

	Väärin	%	Oikein	%	Mahdoton analysoida	%	Yhteensä	
L'	3	16,7	13	72,2	2	11,1	18	100,0 %
La	0	0,0	18	100,0	0	0,0	18	100,0 %
Le	0	0,0	21	100,0	0	0,0	21	100,0 %
Les	0	0,0	6	100,0	0	0,0	6	100,0 %
Yhteensä	10	1,3	733	98,1	4	0,5	747	100,0 %

Samasta taulukosta voidaan myös saada selville, oliko eri sanojen välillä eroja. Kyseessä on kuitenkin niin pieni määrä kohtia, että näiden tulosten perusteella on vaikea tehdä pitäviä johtopäätöksiä. Tämän aineiston kohdalla näyttäisi kuitenkin siltä, että eniten vääriä viittauksia oli sanan *l'* kohdalla (16,7 %). Muita vääriä viittauksia oli vain pronominiin *elle* (98,5 %) ja *il* (98,8 %) kohdalla, muut viittaukset kääntyivät sataprosenttisesti oikein.

Taulukossa 10 puolestaan on myös kaikki kääntimet yhdessä, ja siihen on eritelty viittausten selkeys sanojen mukaan. Kaikkien kääntimien viittauksista 318 oli selviä (42,6 %), ja epäselviä 415 (55,6 %). Epäselviä viittauksia oli siis enemmän, mutta ero ei ole huomattavan suuri.

Taulukko 10 Viittausten selkeys – kaikki kääntimet yhdessä

	Selvä	%	Epäselvä	%	Mahdoton analysoida	%	Yhteensä	
Elle	157	39,6	233	58,8	6	1,5	396	100,0 %
Il	94	38,2	149	60,6	3	1,2	246	100,0 %
Elles	12	100,0	0	0,0	0	0,0	12	100,0 %
Ils	21	70,0	9	30,0	0	0,0	30	100,0 %
L'	7	38,9	6	33,3	5	27,8	18	100,0 %
La	10	55,6	8	44,4	0	0,0	18	100,0 %
Le	14	66,7	7	33,3	0	0,0	21	100,0 %
Les	3	50,0	3	50,0	0	0,0	6	100,0 %
Yhteensä	318	42,6	415	55,6	14	1,9	747	

Kaikki viittaukset olivat selviä pronominiin *elles* kohdalla. Tämän jälkeen selkeimpiä olivat sanojen *ils* (70,0 %), *le* (66,7 %), *la* (55,6 %) ja *les* (50,0 %) viittaukset. Täytyy kuitenkin huomioida, että tutkittavien sanojen määrä aineistossa vaihtelee suuresti. Joitakin sanoja esiintyi aineistossa vain muutamia, joten niiden perusteella ei voida tehdä johtopäätöksiä.

4.5 Erot luokkien välillä

Seuraavaksi tarkastelen viittausten selkeyttä aiemmin esitellyn luokittelun perusteella. Luokittelussa tutkittavat kohdat on jaettu kolmeen ryhmään suomentajan tekemän käännöksen perusteella (alaluku 4.1). Tätä tarkastellaan jokaisen kääntimen kohdalla erikseen, ja lisäksi kaikkien kääntimien tuloksia yhdessä.

Ensimmäisenä on Google Translaten tulokset, jotka näkyvät alla olevassa taulukossa 11. Ryhmään yksi kuuluvat siis sellaiset kohdat, joissa pelkän pronominin käyttäminen voisi tehdä viittauksesta epäselvän. Tähän ryhmään kuuluvista kohdista viittaus on selvä vain seitsemässä kohdassa (5,4 % kaikista tämän ryhmän kohdista), kun taas epäselviä on 122 (93,8 %). Toisessa ryhmässä on kohdat, joissa pronominin käyttäminen ei todennäköisesti aiheuttaisi sekaannusta. Tämän ryhmän kohdista selviä on 74 (67,9 %) ja epäselviä 31 (28,4 %). Vaikka tämänkin ryhmän viittauksista melkein kolmasosa on epäselviä, on tulos kuitenkin huomattavasti parempi kuin ensimmäisessä ryhmässä. Ryhmän kolme muodostavat kohdat, joissa on jo käytetty pronominia, joten sen käyttäminen uudelleen ei ole mahdollista. Näiden kohtien viittauksista yksikään ei ollut selvä, vaan kaikki kymmenen olivat epäselviä. Yhteensä selviä viittauksista oli 81 eli 32,5 prosenttia.

Taulukko 11 Google Translate – viittausten selkeys luokittelun mukaan

Ryhmä	Selvä	%	Epäselvä	%	Mahdoton analysoida	%	Yhteensä
1	7	5,4	122	93,8	1	0,8	130
2	74	67,9	31	28,4	4	3,7	109
3	0	0,0	10	100,0	0	0,0	10
Yhteensä	81	32,5	163	65,5	5	2,0	249

Taulukossa 12 on puolestaan vastaavat luvut DeepL:n osalta. DeepL:n kääntämistä viittauksista selviä oli huomattavasti enemmän kuin Google Translatella: 142 kappaletta eli 57,0 prosenttia kaikista kohdista. Myös DeepL:n kohdalla voidaan kuitenkin todeta, että enemmän selviä viittauksia oli ryhmässä kaksi, 82 kappaletta (75,2 %). Tämä konekäännin onnistui kuitenkin usein tekemään myös ryhmän yksi viittauksista selviä: 55 kohtaa (42,3 %). Myös ryhmän kolme kohdista puolet eli viisi kohtaa on selviä, kun Google Translatella kaikki olivat epäselviä.

Taulukko 12 DeepL – viittausten selkeys luokittelun mukaan

Ryhmä	Selvä	%	Epäselvä	%	Mahdoton analysoida	%	Yhteensä
1	55	42,3	72	55,4	3	2,3	130
2	82	75,2	25	22,9	2	1,8	109
3	5	50,0	5	50,0	0	0,0	10
Yhteensä	142	57,0	102	41,0	5	2,0	249

Taulukosta 13 nähdään kolmannen ja viimeisen konekääntimen eli ChatGPT:n tulokset. Sen tulokset ovat lähellä Google Translaten tuloksia, mutta ovat hieman paremmat. Ensimmäisen ryhmän viittauksista selviä oli 16 kohtaa (12,3 %) ja epäselviä 112 (86,2 %). Myös ChatGPT:n kohdalla toisen ryhmän viittaukset olivat selvempiä kuin ensimmäisen ryhmän: 78 selvää (71,6 %) ja 29 epäselvää (26,6 %). Kuten Google Translate, ei ChatGPT:n kolmannen ryhmän viittauksista yksikään ollut selvä.

Taulukko 13 ChatGPT – viittausten selkeys luokittelun mukaan

Ryhmä	Selvä	%	Epäselvä	%	Mahdoton analysoida	%	Yhteensä
1	16	12,3	112	86,2	2	1,5	130
2	78	71,6	29	26,6	2	1,8	109
3	0	0,0	10	100,0	0	0,0	10
Yhteensä	94	37,8	151	60,6	4	1,6	249

Taulukkoon 14 on koottu kaikkien vielä kääntimien tulokset yhteensä. Tämän taulukon avulla voidaan vertailla kaikkien konekääntimien tuloksia yhdessä. Taulukosta nähdään käännosten onnistuminen ryhmien mukaan jaoteltuna. Ensimmäisen ryhmän viittauksista selviä oli 78 (20,0 %) ja epäselviä 306 (78,5 %). Toisessa ryhmässä vastaavat luvut ovat 234 (71,6 %) ja 85 (26,0 %). Kolmannessa ryhmässä puolestaan luvut ovat 5 (16,7 %) ja 25 (83,3 %).

Taulukko 14 Kaikki kääntimet – viittausten selkeys luokittelun mukaan

Ryhmä	Selvä	%	Epäselvä	%	Mahdoton analysoida	%	Yhteensä
1	78	20,0	306	78,5	6	1,5	390
2	234	71,6	85	26,0	8	2,4	327
3	5	16,7	25	83,3	0	0,0	30
Yhteensä	317	42,4	416	55,7	14	1,9	747

Tämän perusteella voidaan todeta, että parhaiten kääntyivät ryhmän 2 viittaukset, kun taas sekä ryhmän 1 että ryhmän 3 kohdalla oli eniten vaikeuksia. Tämä selittyy sillä, että ryhmään kaksi kuuluivat sellaiset kohdat, joissa jonkin muun ilmaisun käyttäminen pronominin sijasta

ei ole välttämätöntä. Pronominin vaihtaminen johonkin muuhun selittynee siis muilla seikoilla kuin selkeyden kautta välttämättömillä muutoksilla. Ryhmissä 1 ja 3 puolestaan olisi pitänyt käyttää jotakin muuta ratkaisua pronominin käyttämisen sijasta, mutta konekääntimillä oli yleisesti ottaen tämän kanssa vaikeuksia. Seuraavissa kohdissa tarkastelenkin sitä, mitkä olivat suomentajan sekä konekääntimien käyttämät yleisimmät käännökset (4.6) sekä joitakin esimerkkejä konekääntimien tekemistä käännösratkaisuista (4.7).

4.6 Yleisimmät käännökset

Tässä aluvussa esittelen käännökset, joita suomentaja ja konekääntimet ovat käyttäneet. Käännökset on jaoteltu sanojen mukaan, ja ne käydään läpi seuraavassa järjestyksessä: *elle, il, elles, ils, l', la, le* ja *les*. Käännökset puolestaan ovat aakkosjärjestyksessä.

Taulukosta 15 voidaan nähdä kaikki käännökset, joita pronominilla *elle* esiintyi.

Ensimmäisessä sarakkeessa on käännös, ja muihin on laskettu, kuinka monta kertaa kyseinen käännös esiintyi käännöksissä. Myös prosentit on laskettu vertailua varten.

Taulukko 15 *Elle* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
Berthier	1	0,8		0,0		0,0		0,0
emäntä	1	0,8		0,0		0,0		0,0
Henriette	1	0,8		0,0		0,0		0,0
Hortense	6	4,5		0,0		0,0		0,0
hän		0,0	129	97,7	109	82,6	127	96,2
Iris		0,0		0,0	1	0,8		0,0
Jéphine		0,0		0,0	1	0,8		0,0
Jo		0,0		0,0	1	0,8		0,0
Josefine		0,0		0,0		0,0	1	0,8
Joséphine	47	35,6		0,0	1	0,8		0,0
Josiane	4	3,0		0,0		0,0		0,0
mies		0,0		0,0	1	0,8		0,0
Mylène	32	24,2		0,0		0,0		0,0
nainen	3	2,3		0,0	13	9,8		0,0
poliisi	1	0,8		0,0		0,0		0,0
rouva	4	3,0		0,0		0,0		0,0
se		0,0	3	2,3	1	0,8		0,0
se luuska	1	0,8		0,0		0,0		0,0

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
se nainen	1	0,8		0,0		0,0		0,0
Shirley	4	3,0		0,0		0,0		0,0
sisar	1	0,8		0,0		0,0		0,0
Suzanne	1	0,8		0,0		0,0		0,0
tuo muija	1	0,8		0,0		0,0		0,0
tyttö	3	2,3		0,0		0,0		0,0
tytär		0,0		0,0		0,0	1	0,8
tämä		0,0		0,0		0,0	1	0,8
vaimo	3	2,3		0,0		0,0		0,0
Zoé	9	6,8		0,0	2	1,5		0,0
äiti	8	6,1		0,0	2	1,5	2	1,5
Yhteensä	132	100,0	132	100,0	132	100,0	132	100,0

Kuten yllä olevasta taulukosta nähdään, on suomentajan käännösratkaisut huomattavasti monipuolisempia kuin konekääntimien ratkaisut. Suomentajan käännöksistä yleisimmät ovat *Joséphine* (47 kertaa eli 35,6 %) ja *Mylène* (32 eli 24,2 %). Molemmista henkilöahmoista kerrottiin paljon tutkittavassa osiossa, mistä syystä niitä esiintyy paljon. Nimien käyttäminen onkin yksi paljon käytetyistä käännösratkaisuista. Konekääntimien käyttämiä käännöksiä on huomattavasti paljon vähemmän. Kaikkien kääntimien käyttämistä käännöksistä yleisin on *hän*, Google Translate käyttää sitä 129 kertaa (97,7 %), DeepL 109 kertaa (82,6 %) ja ChatGPT 127 kertaa (96,2 %). DeepL käyttää eniten muitakin vaihtoehtoja, joista yleisin on *nainen* (13 kertaa eli 9,8 %). Muita konekääntimien käyttämiä ratkaisuja ovat *se*, *Zoé*, *Iris*, *äiti*, *tytär*, *tämä*, *mies* sekä *Joséphine* erilaisissa muodoissa (*Joséphine*, *Jéphine*, *Jo* ja *Josefine*). Näitä kaikkia on kuitenkin käytetty vain muutamia kertoja, eivätkä kaikkien niiden viittaukset ole oikein.

Taulukossa 16 on esitettyä samat tiedot pronominille *il*. Joissakin kohdissa viittaus on kadonnut käännöksen yhteydessä, mitä merkitään taulukossa viivalla (-).

Taulukko 16 // – käännökset

Käännös	Suomentaja	%	Google Translate	%	Deep L	%	ChatGP T	%
Antoine	3	3,7		0,0		0,0		0,0
Gary	1	1,2		0,0		0,0		0,0
herra	1	1,2		0,0		0,0		0,0

Käännös	Suomentaja	%	Google Translate	%	Deep L	%	ChatGPT	%
hollantilainen	1	1,2		0,0		0,0		0,0
hän		0,0	82	100,0	62	75,6	77	93,9
isä	4	4,9		0,0		0,0		0,0
Josephine		0,0		0,0	1	1,2		0,0
Juniori	3	3,7		0,0		0,0		0,0
lapsi	2	2,4		0,0		0,0		0,0
Marcel	10	12,2		0,0		0,0		0,0
mies	15	18,3		0,0	17	20,7		0,0
naapuri	1	1,2		0,0		0,0		0,0
Philippe	23	28,0		0,0		0,0		0,0
poika	5	6,1		0,0	1	1,2	1	1,2
se	0	0,0		0,0		0,0	1	1,2
se jätkä	1	1,2		0,0		0,0		0,0
se tyyppi	1	1,2		0,0		0,0		0,0
se vauva	1	1,2		0,0		0,0		0,0
tyyppi	1	1,2		0,0		0,0		0,0
tämä		0,0		0,0	1	1,2	1	1,2
uusi johtaja	1	1,2		0,0		0,0		0,0
Vittorio	3	3,7		0,0		0,0		0,0
Wei	5	6,1		0,0		0,0		0,0
-		0,0		0,0		0,0	2	2,4
Yhteensä	82	100,0	82	100,0	82	100,0	82	100,0

Kuten pronominin *elle* kohdalla, myös tässä suomentajan eniten käyttämä käännös on nimi: Philippe 23 kertaa eli 28 prosenttia. Muita usein käytettyjä käännöksiä ovat mies (15 kertaa, 18,3 %) ja Marcel (10 kertaa, 12,2 %). Konekääntimistä Google Translate on käyttänyt pelkästään pronominia 'hän'. Se on myös muiden kääntimien yleisin käännös: DeepL 62 kertaa (75,6 %) ja ChatGPT 77 kertaa (93,9 %). Tämänkin pronominin kohdalla DeepL on käyttänyt usein sanaa *mies*, joka vastaa edellisessä kohdassa olevaa käännöstä *nainen*. Näitä oli 17 eli 20,7 prosenttia DeepL:n käännöksistä. Muita konekääntimien vähän käytettyjä käännöksiä ovat *poika*, *se*, *tämä* ja *Josephine*.

Taulukko 17 *Elles* – käännökset

Käännös	Suomentaja	%	Google Translate	%	Deep L	%	ChatGPT	%
he		0	4	100	4	100	4	100

Käännös	Suomentaja	%	Google Translate	%	Deep L	%	ChatGP T	%
naiset	2	50		0		0		0
tytöt	1	25		0		0		0
ystävättäret	1	25		0		0		0
Yhteensä	4	1	4	100	4	100	4	100

Taulukossa 17 nähdään pronominin *elles* käännökset. Suomentajan ratkaisut ovat *naiset* (2 kertaa), *tytöt* (1) ja *ystävättäret* (1). Kaikki kääntimet ovat käyttäneet vain yhtä käänösratkaisua eli kaikissa neljässä kohdassa on käytetty pronominia *he*.

Taulukko 18 *ils* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
asukkaat	1	10		0		0		0
he		0	10	100	10	100	10	100
kaikki	1	10		0		0		0
kiinalaiset	1	10		0		0		0
muut	2	20		0		0		0
naapurit	2	20		0		0		0
Van den Brockit	2	20		0		0		0
äiti ja poika	1	10		0		0		0
Yhteensä	10	100	10	100	10	100	10	100

Myös pronominin *ils* kohdalla konekääntimet ovat käyttäneet vain yhtä käännöstä eli pronominia *he* kaikissa kymmenessä kohdassa. Suomentajan käyttämät käännökset ovat tässäkin kohdassa paljon vaihtelevampia. Suomentajan käännökset ovat *asukkaat*, *kaikki*, *kiinalaiset*, *muuta*, *naapurit*, *Van den Brockit* sekä *äiti ja poika*. Jokaista näistä on käytetty yksi tai kaksi kertaa, kuten taulukosta 18 nähdään.

Taulukko 19 *L'* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
Berthier	1	16,7		0,0		0,0		0,0
hän		0,0	4	66,7	5	83,3	4	66,7
Joséphine	1	16,7		0,0		0,0		0,0
miehenne	1	16,7		0,0		0,0		0,0
mies	1	16,7		0,0		0,0		0,0

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
poika	1	16,7		0,0		0,0		0,0
se		0,0	2	33,3		0,0	1	16,7
tyttö	1	16,7		0,0		0,0		0,0
-		0,0		0,0	1	16,7	1	16,7
Yhteensä	6	100,0	6	100,0	6	100,0	6	100,0

Taulukossa 19 on puolestaan sanan *l'* käännökset. Konekääntimien käännöksistä selvästi yleisin on *hän*. Toinen vaihtoehto on *se*, ja joissakin tapauksissa viittaus on kadonnut kokonaan käännöksestä. Tässäkin kohdassa nähdään, että suomentajan ratkaisut ovat huomattavasti vaihtelevampia: *Berthier, Joséphine, miehenne, mies, poika* ja *tyttö*. Kaikkia näitä on käytetty kerran.

Taulukko 20 *La* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
asia	1	16,7		0,0		0,0		0,0
Berthier	1	16,7		0,0		0,0		0,0
Hortense	1	16,7		0,0		0,0		0,0
hän		0,0	4	66,7	4	66,7	3	50,0
Joséphine	1	16,7		0,0		0,0		0,0
käsi	1	16,7		0,0		0,0		0,0
nainen		0,0		0,0	1	16,7		0,0
se		0,0	2	33,3	1	16,7	1	16,7
tilanne		0,0		0,0		0,0	1	16,7
tytär		0,0		0,0		0,0	1	16,7
äiti	1	16,7		0,0		0,0		0,0
Yhteensä	6	100,0	6	100,0	6	100,0	6	100,0

Sanan *la* käännökset taulukossa 20 ovat samankaltaisia edellisen kohdan kanssa, mutta tässä konekääntimien käännökset ovat vaihtelevampia. Pronomineja *hän* ja *se* käytetään edelleen, mutta niiden rinnalla esiintyy DeepL:n käyttämä *nainen* sekä ChatGPT:n *tilanne* ja *tytär*. Tässäkin suomentajan käännösratkaisut ovat kuitenkin vielä monipuolisempia: *asia, Berthier, Hortense, Joséphine, käsi* ja *äiti*.

Taulukko 21 *Le* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
hän		0,0	7	100,0	5	71,4	7	100,0
mies	4	57,1		0,0	1	14,3		0,0
Philippe	1	14,3		0,0		0,0		0,0
poika	1	14,3		0,0		0,0		0,0
tämä		0,0		0,0	1	14,3		0,0
tämä mies	1	14,3		0,0		0,0		0,0
Yhteensä	7	100,0	7	100,0	7	100,0	7	100,0

Myös taulukossa 21 nähtävät sanan *le* käännökset noudattavat edellä tutuksi tullutta kaavaa. Suurin osa käännoksistä on pronomini *hän*, mutta DeepL on käyttänyt lisäksi kerran molempia seuraavista sanoista: *mies* ja *tämä*. Myös suomentaja on käyttänyt usein sanaa *mies* (4 kertaa, 57,1 %), ja muita käännoiksi ovat *Philippe*, *poika* ja *tämä mies*.

Taulukko 22 *Les* – käännökset

Käännös	Suomentaja	%	Google Translate	%	DeepL	%	ChatGPT	%
he		0	2	100	2	100	2	100
muut	1	50		0		0		0
tytöt	1	50		0		0		0
Yhteensä	2	100	2	100	2	100	2	100

Viimeisenä taulukosta 22 nähdään käännökset sanalle *les*. Tätä esiintyi aineistossa vain kaksi kertaa, ja konekääntimien käännökset ovat molemmissa kohdissa *he*. Suomentaja puolestaan on käyttänyt sanoja *muut* ja *tytöt*.

4.7 Esimerkkejä kääntimien onnistuneista käännostratkaisuista

Seuraavaksi esittelen joitakin esimerkkejä konekääntimien tekemistä käännoksistä, joissa viittaus on onnistunut ja selvä. Onnistuneet käännökset voidaan jakaa ryhmiin alla olevan taulukon 23 mukaisesti. Ensimmäisessä ryhmässä ovat sellaiset kohdat, joissa on käytetty pronominin sijasta sellaisia sanoja, jotka ilmaisevat kohteen sukupuolta. Tällaisia sanoja ovat esimerkiksi *mies*, *nainen*, *äiti*, *poika* ja *tytär*. Toisessa ryhmässä ovat kohdat, joissa on käytetty henkilöhahmon nimeä. Kolmannen ryhmän kohdissa viittauksesta tulee selvä muun virkkeen perusteella. Neljännessä ryhmässä ovat kohdat, joissa konekääntimien on käyttänyt hyväkseen kontekstia jostakin muualta kuin kyseisestä kohdasta. Saman kohdan eri virkkeen

kontekstia hyödyntäviä kohtia on aineistossa useita, mutta aiemman kohdan kontekstia on käytetty vain harvoin.

Taulukko 23 Onnistuneiden viittausten luokittelu

1	sukupuolta ilmaiseva substantiivi
2	nimen käyttäminen
3	viittaus on selvä muun lauseen perusteella
4	viittaus on selvä muun kontekstin perusteella

Esimerkit on numeroitu, ja ensimmäisenä esitetään ranskankieliset alkuperäiset virkkeet. Tätä merkitään lyhenteellä LT (lähtöteksti). Esimerkeissä annetaan tutkittavan virkkeen lisäksi kaksi edeltävää virkettä kontekstin vuoksi, ja myös sen takia, että konekääntimille annettiin samanlainen kolmen virkkeen pituinen kokonaisuus. Toisena annetaan vertailun vuoksi suomentajan tekemä käännös samasta osuudesta. Tätä merkitään lyhenteellä S (suomentaja). Kolmantena annetaan konekääntimen käännös. Näitä merkitään lyhenteillä GT (Google Translate), DL (DeepL) ja CG (ChatGPT). Tarkasteltava viittaus on lihavoitu kaikissa kohdissa. Myös esimerkin numero aineistossa on merkitty.

4.7.1 Ensimmäinen ryhmä: sukupuolta ilmaiseva substantiivi

Seuraavissa esimerkeissä on välttytty viittauksen epäselvyydeltä käyttämällä pronominin *hän* sijasta sanoja, joista selviää viitattavan kohteen sukupuoli.

(1)

LT: Monsieur et madame Van den Brock étaient venus lui rendre visite après avoir appris la mort de madame Berthier. Ils avaient sonné à sa porte, plus solennels que des candélabres. **Elle**, fantaisiste, ronde, primesautière, lui, sérieux et maigre. (10, s. 127)

S: Herra ja rouva Van den Brock olivat käyneet hänen luonaan kuultuaan rouva Berthierin kuolemasta. He olivat soittaneet ovikelloa jäykempinä kuin kynttilänjalat. **Rouva** värikkäissä vaatteissa, pulleana, impulsiivisena, herra totisena ja laihana. (s. 138)

DL: Van den Brockin pariskunta oli tullut käymään hänen luonaan saatuaan tietää rouva Berthierin kuolemasta. He soittivat hänen ovikelloaan, juhlallisemmin kuin kynttilänjalat. **Nainen** oli oikukas, pyöreä ja leikkisä, mies oli vakava ja hoikka.

Esimerkistä 1 nähdään, että pronomini *elle* on suomentajan käännöksessä *rouva*. DeepL puolestaan on käyttänyt sanaa *nainen*, joka tekee viittauksesta selvän.

(2) (3)

LT: Ils avaient sonné à sa porte, plus solennels que des candélabres. Elle, fantaisiste, ronde, primesautière, lui, sérieux et maigre. **Elle** avait les yeux qui roulaient dans tous les sens et tentait obstinément de les ramener à un point fixe ; **il** fronçait les sourcils et agitait de longs doigts de moine apothicaire comme de gigantesques paires de ciseaux. (11 ja 12, s. 127)

S: He olivat soittaneet ovikelloa jäykempinä kuin kynttilänjalat. Rouva värikäissä vaatteissa, pulleana, impulsiivisena, herra totisena ja laihana. **Rouvalla** oli silmät, jotka pyörivät levottomasti mutta jotka hän yritti kohdistaa yhteen pisteeseen, **herra** kurtisteli otsaansa ja heilutteli pitkiä lääkintämunkin sormiaan kuin valtaisia saksia. (s. 138)

DL: He soittivat hänen ovikelloaan, juhllisemmin kuin kynttilänjalat. Nainen oli oikukas, pyöreä ja leikkisä, mies oli vakava ja laiha. **Naisen** silmät pyörivät joka suuntaan ja yrittivät itsepäisesti saada ne takaisin kiinteään pisteeseen; **mies** nyrpisti otsaansa ja heilutti pitkiä apteekkarin sormiaan kuin jättimäisiä saksia.

Yllä olevissa esimerkeissä 2 ja 3 on kyse samasta ilmiöstä. Ranskan pronominit *il* ja *elle* ovat suomentajan käännöksessä *rouva* ja *herra*. DeepL on käyttänyt samantapaista periaatetta, ja käyttänyt käännöksiä *nainen* ja *mies*.

(4) (5)

LT: J'ai bien fait de n'en parler à personne, je serais devenue une curiosité, on viendrait me lancer des cacahuètes sur le paillason.

Au premier étage de son immeuble vivaient un fils et sa mère, les Pinarelli. **Il** devait avoir cinquante ans, **elle** quatre-vingts. (25 ja 26, s. 129)

S: Onneksi en kertonut asiasta kenellekään, minusta olisi tullut outo otus, jonka kynnysmatolle olisi tultu heittelemään maapähkinöitä.

Talon toisessa kerroksessa asuivat äiti ja poika Pinarelli. **Poika** oli viisissäkymmenissä, **äiti** kahdeksankymmentävuotias. (s. 139–140)

DL: Onneksi en kertonut kenellekään, minusta olisi tullut kuriositeetti, ja ihmiset olisivat tulleet ja heittäneet pähkinöitä ovimattoni päälle.

Hänen talonsa ensimmäisessä kerroksessa asui poika ja hänen äitinsä, Pinarellin perhe. **Poika** oli varmaan viisikymppinen, **äiti** kahdeksankymppinen.

CG: Oli hyvä, etten puhunut siitä kenellekään, olisin tullut uteliaisuuden kohteeksi, ihmiset tulisivat heittämään minulle maapähkinöitä matolle.

Hänen talonsa ensimmäisessä kerroksessa asui äiti ja poika, Pinarellit. **Poika** oli noin viisikymmentä vuotta vanha, ja **äiti** kahdeksankymmentä.

Esimerkeissä 4 ja 5 on kyse äidistä ja pojasta, joihin viitataan ranskaksi pronomineilla *il* ja *elle*. Suomentaja on käyttänyt pronominin sijasta sanoja *äiti* ja *poika*. Konekääntimistä sekä DeepL että ChatGPT ovat käyttäneet tätä samaa ratkaisua, joka tekee viittauksista yksiselitteiset. Tämä kohta toimii samalla myös esimerkkinä sellaisesta kohdasta, jossa edellisten virkkeiden kontekstia on käytetty hyväksi, jotta viittaus saadaan oikein. Tällaisia kohtia on aineistossa useita.

4.7.2 Toinen ryhmä: nimen käyttäminen

Toinen jonkin verran käytetty käänösratkaisu oli pronominin korvaaminen henkilön nimellä. Tämä on konekääntimille vaikeampi ratkaisu, sillä niiltä puuttuu usein kontekstia, eikä ole selvää, kenestä milloinkin puhutaan. Tällaisiakin ratkaisuja kuitenkin esiintyi aineistossa, kuten alla olevista esimerkeistä käy ilmi.

(6)

LT: —C'est bon de te revoir, Jo. Ça fait si longtemps...

Elle leva la tête vers lui et rougit. (94, s.145)

S: ”Mukava nähdä sinua, Jo. Siitä on niin kauan...”

Joséphine kohotti päätään ja punastui. (s. 158)

DL: -Mukava nähdä sinua taas, Jo. Siitä on niin kauan...

Jo katsoi miestä ja punastui.

Esimerkistä 6 nähdään, että DeepL on korvannut pronominin kyseessä olevan hahmon nimellä. Suomenoksessa on käytetty virallista nimeä *Joséphine*, mutta konekääntin on ottanut käyttöön samassa kohtauksessa käytetyn lempinimen *Jo*. Kuten edellisessäkin esimerkissä, myös tässä saadaan edellisistä virkkeistä kontekstia viittaukseen.

(7)

LT: Iris les recevait avec condescendance comme si elle daignait lui pardonner d'être assis à la place de son père, Joséphine hésitait à lui sauter au cou devant les mines réprobatrices de sa mère et de sa sœur. Ce soir, Marcel Grobz doit fêter son premier Noël avec Josiane et son fils. **Elle** irait le voir bientôt. (53, s. 135)

S: Iris otti ne vastaan alentuvasti, ikään kuin hän olisi suvainnut suoda miehelle anteeksi sen, että tämä oli istunut hänen isänsä paikalle, Joséphine epäröi, mutta kapsahti sitten miehen kaulaan äidin ja sisaren seurattessa paheksuen vierestä. Tänä iltana Marcel Grobz viettää varmaan ensimmäistä jouluaan Josianen ja poikansa kanssa. **Joséphine** kävisi hänen luonaan pian. (s. 146–147)

DL: Iris otti heidät vastaan alentuvasti, ikään kuin hän antaisi anteeksi, että hän istui isänsä tuolissa, ja Joséphine epäröi hypätä hänen niskaansa äitinsä ja sisarensa moittivien ilmeiden edessä. Tänä iltana Marcel Grobzin oli määrä viettää ensimmäistä jouluaan Josianen ja hänen poikansa kanssa. **Joséphine** menisi pian tapaamaan häntä.

Toinen esimerkki nimen käyttämisestä nähdään esimerkissä 7, joka on myös DeepL:n tuotos. Tässä pronomini *elle* on käännetty sekä suomentajan versiossa että DeepL:n käännöksessä nimellä *Joséphine*. Toisin kuin esimerkissä 5, on tässä esimerkissä enemmän vaihtoehtoja pronominin *elle* mahdolliseksi käännökseksi. Kyseisessä kohdassa viitataan moneen muuhunkin naispuoliseen henkilöön: *Iris*, *äiti*, *sisar* ja *Josiane*. DeepL on kuitenkin saanut viittaukset oikein, vaikka Joséphine ei olekaan ainoa kohtauksessa mainittu henkilö.

4.7.3 Kolmas ryhmä: viittaus selvä muun lauseen perusteella

Tähän ryhmään kuuluu sellaisia kohtia, joissa käännöksenä on usein ollut pronomini. Sellaisenaan viittaus olisi voinut olla epäselvä, mutta muun lauseen ansiosta myös pelkän pronominin käyttäminen on ollut mahdollista.

(8)

LT: —Ainsi je pénètre dans votre sanctuaire ! C'est un grand honneur...

Il sembla sur le point d'ajouter quelque chose, mais se tut. **Elle** sortit une bouteille de champagne du Frigidaire et la lui tendit pour qu'il l'ouvrît. (59, s. 136)

S: ”Tunkeudun nyt siis pyhäkköönne! Tämä on suuri kunnia...”

Hän oli aikeissa lisätä jotakin mutta vaikenä. **Joséphine** otti jääkaapista samppanjapullon ja ojensi sen miehelle avattavaksi. (s.147)

DL: -Selvä, että astun sisään pyhäkköönne! Tämä on suuri kunnia...

Hän näytti olevan sanomassa jotain, mutta vaikenä. **Hän** otti jääkaapista samppanjapullon ja ojensi sen miehelle avattavaksi.

Esimerkissä 8 ranskankielisessä virkkeessä on selvää, kenestä puhutaan, sillä käytetty persoonapronomini on feminiini. Suomentajaksi selkeyttä on lisätty viittaamalla naispuoliseen henkilöön hänen nimellään ja kohtauksessa esiintyvään miespuoliseen henkilöön sanalla *mies*. DeepL:n käännöksessä käytetään pronominia *hän*, joka ei itsessään kerro, kenestä on kyse. Virkkeessä käytetty sana *mies* kuitenkin selventää sen, että *hän* viittaa kohtauksen toiseen osapuoleen.

(9)

LT: Joséphine et Philippe étaient assis sur le canapé, près du feu. Il s'empara de la main de Jo dans son dos. **Elle** devint cramoisie et le supplia du regard de lâcher sa main. (140, s. 154)

S: Joséphine ja Philippe istuivat sohvalla takkatulen ääressä. Philippe tarttui Joséphinen käteen tämän selän takana. **Joséphine** lehahti tulipunaiseksi ja pyysi häntä päästämään irti. (s. 168)

CG: Joséphine ja Philippe istuivat sohvalla takan läheisyydessä. Hän tarttui Jo'n käteen selkänsä takana. **Hän** punastui ja pyysi silmillään Philippea päästämään irti hänen kädestään.

Esimerkissä 9 on suomennoksessa käytetty henkilön nimeä selventämään viittausta.

ChatGPT:n käännöksessä on käytetty pronominia *hän*, mutta kohtauksessa mukana olevaan mieheen on viitattu hänen nimellään (*Philippe*). Tämän ansiosta viittaus on selvä. Tässä esimerkissä kiinnostavaa on myös se, että konekäännöksessä on käytetty henkilön nimeä, eikä jotakin geneeristä sanaa kuten *mies*. Hahmon nimi on mainittu ranskankielisessä versiossa kaksi virkettä aiemmin, mutta käännin on onnistunut käyttämään sitä käännöksessä oikein.

4.7.4 Viittaus selvä muun kontekstin perusteella

Seuraavissa esimerkeissä on kyse sellaisista kohdista, joissa konekäännin on hyödyntänyt viittauksessa aiempaa kontekstia. Konteksti ei tule saman kohdan muista virkkeistä, vaan jostakin aiemmasta käännöksestä. Kaikki kohdat käännettiin kolmen virkkeen osioissa, mutta tässä konteksti tulee jostakin muualta, ja onkin hieman epäselvää, miten tämä on tapahtunut. Molemmat tämän kohdan esimerkeistä ovat ChatGPT:itä.

(10)

LT: C'est plus facile de faire naître ce mirage-là que de donner le conseil, la présence, l'assistance de l'âme que je ne sais pas offrir, empêtrée dans mes maladresses. Nous payons toutes les deux ma négligence, ma chérie, ma beauté, mon amour que j'aime à la folie. Elle **la** retint un instant dans ses bras et lui répéta à l'oreille ces derniers mots : —Ma chérie, ma beauté, mon amour que j'aime à la folie. (86, s.144)

S: On helpompaa luoda sellainen kangastus kuin antaa ohjeita, läsnäoloa ja sielunopastusta, johon en kykene, koska olen kömpelyyteni vanki. Saamme molemmat maksaa minun laiminlyönneistäni, kultaseni, kaunokaisen, rakas tyttöni, jota niin kovasti rakastan. Joséphine rutisti **Hortensea** hetken ja kuiskasi tämän korvaan sanat: ”Kultaseni, kaunokaiseni, rakas tyttöni, jota niin kovasti rakastan. (s. 156–157)

CG: On helpompaa synnyttää tuo harhakuva kuin antaa neuvoja, olla läsnä ja tarjota sielun apua, jota en osaa tarjota, sotkeutuneena kömpelöihin yrityksiini. Maksamme molemmat huolimattomuudestani, kultaseni, kauneuteni, rakkaani,

jota rakastan mielettömästi. Hän piti **tytärtään** hetken sylissään ja toisti nämä viimeiset sanat tämän korvaan: —Kultaseni, kauneuteni, rakkaani, jota rakastan mielettömästi.

Ensimmäisessä esimerkissä 10 on kyse kohdasta, jossa Joséphine haluaa tytärtään Hortensea. Ranskankielisessä tekstissä on käytetty pronomineja *elle* ja *la*. Suomentaja on korvannut molemmat pronominit henkilöiden nimillä. ChatGPT puolestaan käyttää Joséphinesta pronominia *hän*, mutta Hortensesta sanaa *tytär*. Tämä kohta on mielenkiintoinen sen takia, että kyseissä kohdassa ei tule ilmi se, että Hortense on Joséphin tytär. On siis mahdollista, että konekäännin on saanut viittauksen oikein käyttämällä kontekstia jostakin aiemmasta kohdasta. ChatGPT ei kuitenkaan ole varsinainen konekäännin, vaan tekoäly, joka oppii jatkuvasti.

(11)

LT: —Oui.

—C'est bien.

Il murmurait dans son cou, **elle** sentait les mots s'imprimer sur sa peau. (108, s. 147)

S: ”Niin.”

”Hyvä.”

Philippe kuiskaili hänen kaulalleen, ja **Joséphine** tunsi sanojen painuvan iholleen. (s. 160)

GT: —Kyllä.

—Hyvä niin.

Hän kuiskasi hänen kaulaansa, **Josefine** tunsi sanojen painuvan hänen iholleen.

Esimerkissä 11 on vastaavanlainen tilanne, jossa ranskaksi käytetään pronominia *elle*. Sekä suomentaja että ChatGPT ovat käyttäneet pronominin sijasta henkilön nimeä. ChatGPT on päätenyt myös suomentamaan nimen, mutta viittaus on sinällään oikein. Oleellista on kuitenkin se, että nimeä ei käytetä kyseissä kohdassa, mutta konekäännöksessä sitä on silti käytetty oikein.

4.8 Esimerkkejä kääntimien epäonnistuneista viittauksista

Seuraavaksi esittelen joitakin konekääntimien esimerkkejä, joissa viittaus on joko väärin tai epäselvä. Ensimmäiseksi käsittelen joitakin tapauksia, joissa viittaus on täysin väärä. Sen

jälkeen vuorossa ovat sellaiset kohdat, joissa viittaus itsessään on oikein, mutta se ei ole täysin selvä. Lopuksi esittelen vielä yhden kohdan, jossa alkutekstin pronomini on monitulkintainen.

4.8.1 Väärät viittaukset

Kuten taulukosta 7 alaluvussa 4.2 nähdään, jokainen konekäännin tuotti ainakin yhden sellaisen kohdan, jossa viittaus on väärin. Esittelenkin seuraavaksi yhden epäonnistuneen käännöksen jokaiselta konekääntimeltä.

(12)

LT: Il chercha du bout de la langue un morceau de pruneau, suçà un peu d'armagnac. Il souriait béatement dans la pénombre. La prochaine fois que je vais à New York, je l'emmène avec moi. (152, s. 156)

S: Hän tunsi kielellään luumunhitusen, maistoi armanjakin. Hän hymyili autuaana hämärässä. Ensi kerralla kun menen New Yorkiin, otan **Joséphinen** mukaani. (s. 170)

GT: Hän etsi kielensä kärjellä luumua, imi vähän armanjakkia. Hän hymyili autuaasti pimeässä. Kun seuraavan kerran menen New Yorkiin, otan **sen** mukaani.

Esimerkissä 12 on kohta, jossa Google Translaten käännöksessä on väärä viittaus. Suomennoksessa on käytetty viitattavan hahmon nimeä, mutta käännin on käyttänyt pronominia *se*. On epäselvää, mihin konekääntimen pronomini viittaa, mutta ainakaan se ei viittaa oikeaan asiaan. Tässä kohdassa kontekstin puute voi johtaa viittauksen epäonnistumiseen. DeepL ja ChatGPT ovat kuitenkin saaneet tässä kohdassa viittauksen oikein käyttämällä pronominia *hän*.

(13)

LT: A-t-il revu Zoé sans que je le sache ? Est-ce lui qui a eu l'idée de cette mise en scène grotesque ou elle, toute seule ? **Elle** restait immobile, face à Antoine en carton-pâte, essayant de comprendre. (159, s. 157)

S: Onko hän tavannut Zoén minun tietämättäni? Onko Antoine keksinyt tämän naurettavan esityksen vai Zoé yksin? **Joséphine** istui liikkumatta vastapäätä pahvista Antoinea ja yritti ymmärtää. (s. 171)

DL: Tapasiko hän Zoén taas minun tietämättäni? Keksikö hän tämän groteskin lavastuksen idean, vai tekikö Zoé sen itse? **Zoé** pysyi liikkumatta ja katsoi paperimassasta tehtyyn Antoineen ja yritti ymmärtää.

Esimerkissä 13 puolestaan esimerkki sellaisesta kohdasta, jossa DeepL:n viittaus ei ole oikein. Ranskankielisessä virkkeessä on käytetty pronominia *elle*, jonka suomentaja on kääntänyt kyseessä olevan henkilöahmon nimellä: Joséphine. Myös DeepL on käyttänyt henkilön nimeä, mutta se on käyttänyt väärää nimeä eli Zoéta. Kyseisessä kohdassa on mainittu nimeltä vain Zoé, mistä voi johtua se, että käännin on päätenyt käyttämään sitä. Sekä Google Translate että ChatGPT ovat käyttäneet tässä kohdassa pronominia *hän*, joka on sinänsä oikein, mutta voi jättää epäselväksi sen, kenestä on puhe.

(14)

LT: Elle sentait une écoute bienveillante de la part de Josiane qui la contemplait de son regard chaud et attentif en lissant son jabot blanc.

—Vous croyez que j'ai des visions ?

—Non... mais madame Suzanne l'a vu dans la gueule d'un crocodile et elle se trompe rarement. (243, s. 181–182)

S: Hän tunsi, että Josiane kuunteli hyväntahtoisesti, katsellen häntä suopeasti ja huolekkaasti, sivellen paitansa valkoista röyhelöä.

”Luuletteko, että näen näkyjä?”

”En... mutta Suzanne näki **miehenne** krokotiilin kidassa ja hän erehtyy harvoin. (s. 199)

CG: Hän tunsi Josianen myötätuntoisen kuuntelun, kun tämä katseli häntä lämpimillä ja tarkkaavaisilla silmillään valkoisen jabotin laskiessaan.

—Luuletteko, että minulla on näkyjä?

—Ei... mutta rouva Suzanne näki **sen** krokotiilin kidassa, eikä hän yleensä erehdy.

Yllä on esimerkki ChatGPT:n väärin kääntämästä viittauksesta. Suomennoksessa pronomini *l'* on käännetty oikein viittamaan päähenkilö Joséphin mieheen, mutta kääntimen versiossa käännös on *se*. *Se* viittaa yleensä elottomiin asioihin tai sitä käytetään puhekielisenä versiona pronominista *hän*. Tässä kohdassa on kuitenkin kyse ihmisestä, eikä puhekielen käyttäminen sovi kyseiseen kohtaukseen. Kyseessä on siis väärä viittaus. ChatGPT ei kuitenkaan ollut ainoa käännin, jolle tämä kohta tuotti vaikeuksia, sillä myös Google Translate käänsi viittauksen sanalla *se*. DeepL kuitenkin käytti pronominia *hän*, jolloin viittauksesta tuli oikea.

4.8.2 Epäselvät viittaukset

Kuten kohdan 4.3 taulukosta 8 kävi ilmi, oli aineistossa paljon kohtia, joissa viittaus oli epäselvä. Edellä esiteltiin jokaiselta kääntimeltä yksi väärä viittaus, ja teen samoin myös tässä kohdassa epäselvien viittausten kanssa.

(15)

LT: Marcel la prévint qu'il quitterait la table si elle refusait son cadeau. Il se considérerait offensé. **Elle** insista, **il** renchérit, **elle** s'obstina, **il** tint bon, **elle** s'entêta, **il** ne voulut pas en démordre. (219, 220, 221, 222, 223, 224, s. 172–173)

S: Marcel uhkasi lähteä pöydästä, jos Shirley kieltäytyisi ottamasta lahjaa vastaan. Hän loukkaantuisi. **Shirley** kinasi, **Marcel** sanoi vastaan, **Shirley** intti, **Marcel** piti pintansa, **Shirley** jankkasi, **Marcel** ei halunnut antaa periksi. (s. 189)

GT: Marcel varoitti häntä, että hän poistuisi pöydästä, jos hän kieltäytyisi hänen lahjastaan. Hän luulisi olevansa loukkaantunut. **Hän** vaati, **hän** meni pidemmälle, **hän** jatkoi, **hän** piti kiinni, **hän** jatkoi, **hän** ei perääntynyt.

Esimerkissä 15 on virke, jossa on kuusi tutkimuksessa tarkastelua pronominia: *elle* ja *il*, jotka molemmat toistuvat kolme kertaa. Suomennoksessa sekaannukselta on välttytty käyttämällä henkilöiden nimiä. Google Translaten käännöksessä puolestaan on kaikissa käytetty pronominia *hän*, mistä aiheutuu se, että viittaukset eivät ole selviä. DeepL on ratkaissut ongelman käyttämällä sanoja *nainen* ja *mies*. ChatGPT on myös käyttänyt pronominia *hän* kaikissa paitsi viimeisessä viittauksessa, jossa viittaus on jäänyt kokonaan pois. Kyseessä ei siis ollut kääntimille kovinkaan helppo kohta, sillä vain yksi kääntimistä sai viittauksista selvät.

(16)

LT: En entendant ces chiffres macabres, Joséphine avait serré ses mains entre ses cuisses pour réprimer le tremblement qui la secouait.

—La semelle de la chaussure devait être drôlement épaisse, énonça le capitaine comme si elle essayait de se convaincre. **Il** a frappé à l'emplacement du cœur. (4, s. 124)

S: Kuullessaan nuo makaaberit luvut Joséphine oli painanut kätensä reisien väliin tukahduttaakseen häntä vavisuttavan tärinän.

”Sen teidän kenkäne pohjan on täytynyt olla todella paksu”, totesi ylikonstaapeli ikään kuin olisi vakuutellut itseään. ”**Mies** puukotti Berthieriä sydämeen.” (s. 134)

DL: Kun Joséphine kuuli nämä makaaberit hahmot, hän puristi kätensä reisiensä väliin tukahduttaakseen vapinaa, joka ravisteli häntä.

-Kengänpohjan on täytynyt olla todella paksu", kapteeni sanoi kuin yrittäen vakuuttaa itseään. **Hän** löi häntä sydämeen.

DeepL:n esimerkissä on suomennoksessa käytetty pronominin *il* vastineena sanaa *mies*. Käännin on kuitenkin käyttänyt pronominia *hän*, joka olisi voinut kontekstin puolesta ehkä ollakin riittävän selvä viittaus. Samassa lauseessa on kuitenkin toisen kerran sama pronomini, mikä tekee kohdan viittauksista epäselviä. Muutkin kääntimet ovat käyttäneet käännöksissään samaa pronominia *hän*, mutta ne ovat jättäneet pois viittauksen toiseen henkilöön, kuten myös alkuperäisessä lauseessa on tehty.

(17)

LT: Il devait avoir cinquante ans, elle quatre-vingts. Il était grand, mince, les cheveux teints en noir. **Il** ressemblait, en plus âgé, à Anthony Perkins dans Psychose. (28, s.129)

S: Poika oli viisissäkymmenissä, äiti kahdeksankymmentävuotias. Poika oli pitkä ja hoikka, tukka mustaksi värjätty. **Poika** oli kuin *Psykon* Anthony Perkinsin vanhentunut versio. (s. 140)

CG: Hän oli noin viisikymmentä vuotta vanha, ja äiti kahdeksankymmentä. Hän oli pitkä, hoikka, ja hänellä oli mustaksi värjättyt hiukset. **Hän** muistutti, vanhempana, Anthony Perkinsia elokuvassa *Psyko*.

Esimerkissä 17 nähdään, miten suomennoksessa on käytetty sanaa *poika*, kun taas ChatGPT:n käännöksessä on käytetty pronominia *hän*. Edellä on puhuttu äidistä ja pojasta, joten pelkän pronominin käyttäminen ei tee viittauksesta selvää. Kääntimen versiossa pronomini *hän* voisi viitata myös äitiin. Myös Google Translate ja ChatGPT ovat käyttäneet samaa käännösratkaisua, ja niidenkin käännöksissä viittaus on siis epäselvä.

4.8.3 Monitulkintainen pronomini

Seuraavassa esimerkissä on kohta, jossa lähtötekstin pronomini voidaan nähdä monitulkintaisena. Tämä näkyy myös konekääntimien käännöksistä, joissa on tehty eri tulkintoja.

(18)

LT: Joséphine sourit. Elle allait rapporter un détail cocasse. Le capitaine croirait qu'elle le faisait exprès ou qu'elle ne **la** prenait pas au sérieux. (6, s. 124)

S: Joséphine hymyili. Hän aikoi mainita naurettavan yksityiskohdan. Ylikonstaapeli ajattelsi, että hän teki sen tahallaan tai että hän ei suhtautunut **asiaan** vakavasti. (s. 134)

GT: Josephine hymyilee. Hän aikoi kertoa hauskan yksityiskohdan. Kapteeni luuli, että hän teki sen tahallaan tai ettei hän ottanut **sitä** vakavasti.

DL: Joséphine hymyili. Hän aikoi kertoa jotain hauskaa. Kapteeni luulisi, että hän teki sen tahallaan tai että hän ei ottaisi **häntä** vakavasti.

CG: Joséphine hymyili. Hän oli aikeissa kertoa hassusta yksityiskohdasta. Kapteeni saattaisi luulla, että hän teki sen tarkoituksella tai ettei hän ottanut **tilannetta** vakavasti.

Alkutekstin pronomini *la* on käännetty ihmisen tekemässä käänöksessä sanalla *asia*. Google Translaten käyttämä pronomini *se* on hieman epäselvä, eikä ole selvää, mihin se viittaa. ChatGPT on päätenyt samantapaiseen ratkaisuun kuin suomentaja, ja käyttänyt sanaa *tilanne*. DeepL puolestaan on kääntänyt kohdan pronominilla *hän*, jolloin viittauksen kohteena on kohtauksessa oleva ylikonstaapeli. Kohta on siis käännetty niin, että alkutekstin pronominilla viitataan joko kohtauksen ylikonstaapeliin tai yleisemmin asiaan tai tilanteeseen.

4.9 Huomautus kontekstista

Alaluvussa 2.2.4 tarkasteltiin kontekstin vaikutusta konekääntimien tuloksiin. Siinä mainittiin konekääntimien toimivan niin, että ne kääntävät vain yhden virkkeen kerrallaan. Oman tutkimukseni tulosten perusteella vaikuttaa kuitenkin siltä, että tämä ei pidä enää paikkaansa. Joissakin kohdissa näkyy selvästi, että konekäännin on käyttänyt edeltävien virkkeiden kontekstia tai jopa edeltävien kohtien kontekstia. Tekstiä käännettiin kolmen virkkeen osioissa, mutta joissakin käänöksissä on hyödynnetty tietoa, jota kyseisessä kohdassa ei ole. Näyttäisi siis siltä, että ainakin jotkin nykyisistä konekääntimistä ovat oppineet hyödyntämään kontekstia, eivätkä ne käännä vain yhtä virkettä kerrallaan. Aineistossani on lukuisia kohtia, joissa on käytetty saman kohdan kontekstia hyväksi. Näitä näkyy myös analyysiluvussa annetuissa esimerkeissä. Edellisten kohtien kontekstia käyttäviä kohtia ei sen sijaan ole yhtä paljon, mutta niistäkin annettiin esimerkkejä analyysiluvun alaluvussa 4.7.4.

5 Yhteenveto ja johtopäätökset

Analyysissä tarkasteltiin yhteensä 249 viittausta, jotka oli luokiteltu suomentajan tekemän käännöksen perusteella luokkiin. Tämä jako on esitetty taulukossa 6, ja siitä käy ilmi, että suurin osa kohdista kuuluu ryhmään 1 (130 kohtaa) tai ryhmään 2 (109 kohtaa). Lisäksi ryhmässä kolme on 10 kohtaa. Ensimmäisessä ryhmässä ovat kohdat, joissa voi olettaa esiintyvän sekaannusta, jos käytettäisiin pelkkää pronominia. Ryhmän 2 kohdissa tällaista ei oletettavasti esiinny. Ryhmä 3 muodostuu kohdista, joissa pronominin käyttäminen johtaisi tilanteeseen, jossa samassa lauseessa olisi kaksi hän-pronominia.

Ensimmäisenä analyysissä tarkasteltiin sitä, ovatko konekäännösten tekemät viittaukset säilyneet oikein. Tässä kohdassa tulokset olivat oikein hyviä, sillä suurin osa viittauksista oli oikein. Eri konekääntimien välillä ei ollut juuri lainkaan eroja. Google Translatella ja DeepL:llä viittauksista oikein oli 244 kohtaa, ja ChatGPT:llä 245. Siispä vääriä viittauksia oli kolmen kääntimen välillä yhteensä vain 10 kappaletta. Lisäksi oli muutama kohta, joissa käännös oli sen verran huonosti onnistunut, että viittauksen onnistuneisuutta ei pystytty määrittelemään. Nämä kohdat luokiteltiin ryhmään 'mahdoton analysoida'.

Viittausten onnistuneisuutta tärkeämpää oli kuitenkin tutkia sitä, ovatko viittaukset selviä. Tässä nähtiinkin jo eroja eri kääntimien välillä, sillä DeepL:n tulokset olivat huomattavasti parempia kuin kahden muun kääntimen tulokset. DeepL:n viittauksista selviä oli 142 kohtaa, kun taas Google Translatella selviä oli 81 ja ChatGPT:llä 94 kappaletta. Tämän jaottelun mukaan voitaisiin siis todeta, että pronominiinien kääntämisestä parhaiten selvisi DeepL, toiseksi parhaiten ChatGPT ja huonoimman tuloksen sai Google Translate. Tässä täytyy kuitenkin muistaa, että aineisto on pieni ja että tulokset koskevat pelkästään tätä tutkimusta. Näitä tuloksia ei siis voi yleistää, eikä pelkästään tämän tutkimuksen perusteella voida jakaa konekääntimiä paremmuusjärjestykseen.

Analyysissä on lisäksi tarkasteltu tuloksia eri näkökulmista. Yksi näistä koskee sitä, onko tutkittavien sanojen välillä eroja. Koska suurin osa viittauksista oli oikein, ei vääriä viittauksia ole paljon. Siispä ei voida sanoa, onko eri sanojen välillä oikeasti eroja, sillä tällaisia kohtia on hyvin vähän. Sama koskee viittausten selkeyttä, joten siitäkään on vaikea tehdä mitään johtopäätöksiä.

Tuloksia on tarkasteltu myös aluksi esitellyn suomennoksen mukaan tehdyn luokittelun mukaan. Hypoteesina oli se, että ryhmän 2 viittaukset onnistuvat hyvin, mutta erityisesti

ryhmän 1 kohdissa voisi olla ongelmia. Tulosten perusteella näyttää siltä, että tämä pitää paikkansa. Google Translaten käännöksessä ryhmässä 2 selviä oli 74 kappaletta eli 67,9 prosenttia. Ryhmässä 1 näin oli vain 7 kohdassa, eli 5,4 prosenttia koko ryhmän kohdista. Ryhmässä kolme kaikki 10 kohtaa olivat epäselviä. DeepL:n käännöksessä ryhmän 2 käännöksistä selviä oli 82 eli 75,2 prosenttia. DeepL onnistui kuitenkin tekemään selviä viittauksia myös ryhmässä 1: 55 selvää eli 42,3 prosenttia. Lisäksi ryhmän kolme kohdista puolet olivat selviä. ChatGPT:n tulokset muistuttavat Google Translaten tuloksia: ryhmässä 2 oli 78 selvää kohtaa (71,6 %), ryhmässä 1 puolestaan 16 selvää (12,3 %) ja ryhmän 3 kohdista yksikään ei ollut selvä. Yleisesti ottaen voidaan siis sanoa, että hypoteesi vahvistui, ja ryhmän 2 viittaukset olivat konekääntimille helpoimpia.

Analyysin tuloksissa on myös käyty läpi sitä, mitkä olivat suomentajan ja kääntimien yleisimmin käyttämät käännökset. Tarkasteltavat kohdat oli valittu suomennoksen perusteella, ja kriteerinä oli nimenomaan se, että kyseissä kohdassa on käytetty jotakin muuta kuin pronominia. Siispä suomentajan tekemässä käännöksessä ei esiinny pronomineja, vaan pelkästään muita sanoja, joilla voidaan viitata johonkin asiaan tai henkilöön. Useasti kyseessä oli erisnimi tai jokin substantiivi, joka jotenkin kuvaa viittauksen kohdetta. Konekääntimillä nimiä tai muita substantiiveja kuitenkin käytettiin vain harvoin. Hyvin yleisiä konekääntimien käyttämiä käännöksiä ovatkin *hän* ja *se*, joita esiintyy huomattavasti eniten kaikilla konekääntimillä. Erityisesti DeepL on kuitenkin käyttänyt jonkin verran myös muita käännöksiä, joista yleisiä ovat *nainen* ja *mies*. Mielenkiintoista on se, että Google Translate ja ChatGPT eivät käyttäneet näitä sanoja kertaakaan, vaikka niillä voisi helposti korvata pronominit *elle* ja *il*. Muita käytettyjä käännöksiä ovat erityisesti nimet.

Näiden lukujen lisäksi analyysissä on käsitelty onnistuneita sekä epäonnistuneita esimerkkejä aineistosta. Onnistuneet viittaukset voitiin luokitella neljään kohtaan: sukupuolta ilmaiseva substantiivi, nimen käyttäminen, kohdat, joissa viittaus oli selvä muun lauseen perusteella ja kohdat, joissa viittaus oli selvä muun kontekstin perusteella. Väärien ja epäselvien viittausten kohdalla vastaavaa ryhmittelyä ei tehty, mutta niistä on esitelty joitakin esimerkkejä, jotka tuovat esiin sitä, millaisia ongelmia konekääntimillä oli.

6 Lopuksi

Tässä tutkimuksessa tarkasteltiin sitä, miten hyvin eri konekääntimet onnistuvat kääntämään pronomineja ranskasta suomeen. Käsittelyssä olivat enimmäkseen sellaiset pronominit, jotka ilmaisevat viittauksen kohteen sukupuolta. Tutkimus ei rajoitu pelkästään henkilöihin, vaan tutkimuksessa oli mukana myös asioita, sillä niillä on ranskan kielessä kieliopillinen sukupuoli. Tämän johdosta asioihin voidaan viitata toisinaan pelkällä pronominilla, mutta suomen kielessä vastaava ei ole mahdollista, koska suomen kielen pronomineilla ei voida erotella sukupuolta. Tämä ranskan ja suomen kielten välinen ero tekee pronominiin kääntämisen tutkimuksesta mielenkiintoista. Kääntämisessä joudutaan usein käyttämään muita keinoja, jotta viittaus saataisiin oikein, ja että se olisi selvä.

Kyseisen ilmiön tutkimista varten valittiin aineistoksi osio ranskankielisestä romaanista. Myös sen suomennosta on hyödynnetty tutkimuksessa. Esimerkiksi tutkittavat kohdat valikoituivat suomennoksen perusteella, sillä mukaan otettiin vain sellaiset kohdat, joissa kääntäjä oli käyttänyt käännöksessä jotakin muuta kuin pronominia. Oletuksena oli, että tällaiset kohdat ovat jossakin mielessä ongelmallisempia kuin suoraan pronominilla käännetty kohdat. Koska tutkimuksessa tutkitaan konekääntimiä, oli mielekkäämpää tutkia sellaisia kohtia, joissa ongelmia voisi olettaa esiintyvän kuin kohtia, jotka kääntyvät lähtökohtaisesti hyvin. On tietysti mahdollista, että konekääntimillä esiintyy ongelmia myös yksinkertaisemmissa kohdissa, ja tätä olisi voinut tutkia osana tätä tutkimusta.

Aineiston valintaan puolestaan vaikutti erityisesti se, että romaaneissa esiintyy tyypillisesti enemmän pronomineja kuin asiateksteissä. Lisäksi aineiston saatavuus oli asia, joka tuli ottaa huomioon valintaa tehtäessä. Aineisto itsessään on kuitenkin vain pieni otos, eikä näin pienen tutkimuksen perusteella ole mahdollista tehdä pitäviä johtopäätöksiä. Tämän tutkimuksen aineisto koostuu loppujen lopuksi vain yhden kirjailijan ja yhden suomentajan teksteistä, joten useampien eri tekstien tutkiminen voisikin olla mielenkiintoista, jotta saataisiin kartoitettua myös tapauksia, joita ei tässä aineistossa esiintynyt. Myös muiden konekääntimien ottaminen mukaan olisi mielenkiintoista. Esimerkiksi sääntöpohjaisten tai tilastollisten kääntimien vertaaminen neuroverkkokääntimiin voisi olla jatkotutkimuksen aiheena.

Tutkimuksessa käytetty arviointimalli on itse kehitetty, mikä tarkoittaa erityisesti sitä, että tämän tutkimuksen tuloksia ei voida helposti verrata samaa aihetta käsittelevien tutkimusten tuloksiin. Tästä syystä tutkimus voi jäädä tietyssä mielessä irralliseksi. Arviointimallin

subjektiivisuus on myös harmillista. Viittausten selkeyden arvioiminen on jossakin määrin henkilöstä riippuvaa, ja tässä tutkimuksessa arviointi perustuu pelkästään yhden tutkijan arvioon. Olisikin parempi, että saman tekstin arvioisi monta eri tutkijaa, jotta voitaisiin välttyä tällaiselta subjektiivisuudelta. Vaihtoehtoisesti arviointimallia voisi kehittää, ja tehdä tarkemmat kriteerit sille, mikä lasketaan selväksi viittaukseksi ja mikä ei. Olen arviointia tehdessäni käyttänyt samoja kriteerejä jokaisen tekstin kohdalla, joten tuloksia voi vertailla keskenään. Sen sijaan jonkun muun tekemän tutkimusten tulosten vertaileminen tämän tutkimuksen tuloksiin voi olla vaikeaa. Lisäksi arviointimallissa voisi olla isompi asteikko. Tällä hetkellä viittauksesta voidaan sanoa vain se, onko se selvä vai ei. Aina on kuitenkin olemassa tapauksia, jotka ovat näiden kahden rajalla. Asteikkoa voisi siis suurentaa, ja lisätä siihen sellaisia kategorioita, kuten 'melko selvä' tai 'hyvin selvä'. Näin arviointiin saataisiin enemmän vivahteita, ja tarkempaa tietoa. Tällaisten kategorioiden kriteerien määrittäminen olisi kuitenkin tehtävä huolellisesti.

Tässä tutkimuksessa käsitelty aihe on sellainen, jota voitaisiin tulevaisuudessa tutkia vielä enemmänkin. Ensinnäkin pronomien kääntämisen vaikeus ei ole kyseessä vain ranskasta suomeen käännettäessä. Samaa ilmiötä sukupuolta ilmaisevista pronomineista esiintyy monissa muissakin kielissä, joten olisi mielenkiintoista tehdä vastaavanlaista tutkimusta tällaisten kielten kohdalla. Tuloksia voisi vertailla keskenään, ja selvittää näin, onko eri kieliparien välillä eroja. Lisäksi olisi mielenkiintoista toistaa tutkimus erilaisella aineistolla.

Tässä tutkimuksessa käännettiin vain lyhyitä kolmen virkkeen pituisia osioita, mutta tutkimuksen voisi tehdä myös pidemmällä teksteillä. Tällöin konekääntimet saisivat käyttöönsä enemmän kontekstia, ja tulokset voisivat olla parempia. Ainakin osa konekääntimistä käyttää hyväkseen aiempaa tekstiä, joten kontekstin vaikutusta tuloksiin olisi hyvä tutkia. Myös tulevien virkkeiden tuoman kontekstin vaikutusta olisi mielenkiintoista tutkia, ja vertailla tuloksia. Tässä tutkimuksessa huomioitiin vain edeltävä konteksti, mutta olisi mielenkiintoista toistaa koe ottamalla huomioon esimerkiksi kaksi tulevaa virkettä tai sekä edeltävää että tulevaa kontekstia. Näitä tuloksia vertailemalla voitaisiin saada mielenkiintoista tietoa kontekstin vaikutuksesta pronomien kääntämiseen neuroverkkokääntimillä.

Viimeisenä täytyy ottaa huomioon se, että konekääntimiä kehitetään koko ajan, ja ne toimivat aina vain paremmin. Saman tutkimuksen toistaminen suhteellisen lyhyenkin ajan päästä voi antaa aivan erilaisia tuloksia.

Lähteet

Aineistolähteet

Pancol, Katherine 2008. *La valse lente des tortues*. Pariisi, Albin Michel.

Pancol, Katherine 2012. *Kilpikonnien hidas valssi*. Suom. Lotta Toivanen. Helsinki, Bazar.

Tutkimuskirjallisuus

Bai, Xiaojiang – Yu, Shinwen 2023. Rule-Based Machine Translation. *Routledge encyclopedia of translation technology*, s. 193–207. Toim. Sin-wai Chan. Toinen painos. Routledge, Abingdon.

Behrens, Bergljot 2022. Translated Text. *The Cambridge Handbook of Translation*, s. 96–116. Toim. Kirsten Malmkjær. Cambridge, Cambridge University Press. Saatavissa: <https://doi.org/10.1017/9781108616119.006>

Biographie de Katherine Pancol. [Verkkójulkaisu.] Viitattu 29.9.2023. Saatavissa: <https://www.katherine-pancol.com/biographie>

Caswell, Isaac. 11.5.2022. *Google Translate learns 24 new languages*. [Verkkójulkaisu.] Viitattu 16.1.2024. Saatavissa: <https://blog.google/products/translate/24-new-languages/>

Chevalier, Jean-Claude 1974. *Grammaire Larousse du français contemporain*. Paris, Larousse

DeepL 2023a. About the formal/informal feature. [Verkkójulkaisu.] DeepL Help Center. Viitattu 16.1.2024. Saatavissa: <https://support.deepl.com/hc/en-us/articles/4406432463762-About-the-formal-informal-feature>

DeepL 2023b. Press Information. [Verkkójulkaisu.] Viitattu 3.10.2023. Saatavissa: <https://www.deepl.com/press.html>

Dico en ligne Le Robert. Iel. Viitattu 9. 10.2023. Saatavissa: <https://dictionnaire.lerobert.com/definition/iel>

Doherty, Steven 2019. Translation technology evaluation research (*The Routledge Handbook of Translation and Technology*, s. 339–353). Toim. Minago O’Hagan. United Kingdom, Routledge

Elmiger, Daniel 2022. Quel est mon/ton/son pronom ? Invariabilité, autodétermination et le pronom *iel*. *GLAD!. Revue sur le langage, le genre, les sexualités* 12, 192–199. Viitattu 7.11.2023. Saatavissa: <https://doi.org/10.4000/glad.4215>

- Jiao, Wenxiang – Wang, Wenxuan – Jen-tse, Huang – Wang, Xing – Tu, Zhaopeng 2023. Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine. *arXiv.Org*. Viitattu 24.10.2013. Saatavissa: <https://doi.org/10.48550/arxiv.2301.08745>
- Koehn, Philipp 2020. *Neural machine translation*. Cambridge, Cambridge University Press
- Koponen, Maarit 2017. Neuroverkkokääntäminen – vastaus kieltenvälisiin ongelmiin? [Verkkójulkaisu.] Viitattu 6.10.2023. Saatavissa: <https://www.utu.fi/fi/ajankohtaista/uutinen/neuroverkkokaantaminen-vastaus-kieltenvälisiin-ongelmiin>
- Kuusi, Päivi 2011. *Miksi näkökulma muuttuu käänöksessä? Eksplisiittistämisen ja normaalistamisen selitysvoima ja seuraukset*. Väitöskirja. Tampere University Press. Viitattu 3.10.2023. Saatavissa: <https://trepo.tuni.fi/handle/10024/66777>
- Kuzmina, Ekaterina 2016. *Kolmannen persoonan henkilöviittaukset Boris Akuninin romaanin Turetski gambit suomennoksessa* [pro gradu -tutkielma, Helsingin yliopisto]. <http://urn.fi/URN:NBN:fi:hulib-201605111654>
- Latomaa, Sirkku – Nurminen, Mary 2022. Työnkuvan muutos: Kääntäjästä jälkieditoijaksi. *Kieli, koulutus ja yhteiskunta* 13(3). Viitattu 6.10.2023. Saatavissa: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2022/tyonkuvan-muutos-kaantajasta-jalkieditoijaksi>
- Leblanc, Isabelle 2022. Devenir non-binaire en français contemporain. *Langage et société* 2:173, 223–226. Viitattu 9.10.2023. Saatavissa: <https://doi.org/10.3917/ls.179.0218>
- Liu, Xueting – Li, Chengze 2023. Artificial Intelligence and Translation. *Routledge encyclopedia of translation technology*, s. 280–302. Toim. Sin-wai Chan. Toinen painos. Routledge, Abingdon.
- Lommel, Arle 2018. Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. *Translation Quality Assessment*, s. 109–127). Toim. Joss Moorkens, Sheila Castilho, Federico Gaspari, Stephen Doherty. Springer International Publishing, Cham. Saatavissa: https://doi.org/10.1007/978-3-319-91241-7_6
- Lotta Toivanen—Koneen Säätö. [Verkkójulkaisu.] Viitattu 3.10.2023. Saatavissa: <https://koneensaatio.fi/saaren-kartanon-residenssi/residenssitaiteilijat-ja-tutkijat/lotta-toivanen/>
- Marchello-Nizia, Christiane 1989. Le neutre et l’impersonnel. *LINX*, 21(1), 173–179. Viitattu 9.10.2023. Saatavissa: <https://doi.org/10.3406/linx.1989.1139>
- Marr, Bernard. 19.5.2023. *A Short History Of ChatGPT: How We Got To Where We Are Today*. [Verkkójulkaisu.] Viitattu 23.1.2024. Saatavissa:

<https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/>

Palkinnot kirjailija Markku Turuselle ja suomentaja Lotta Toivaselle. 29.1.2011

[Verkkojulkaisu.] Viitattu 3.10.2023. Saatavissa:

<https://news.cision.com/fi/gummerus/r/palkinnot-kirjailija-markku-turuselle-ja-suomentaja-lotta-toivaselle,c540751/>

Prix: Des Maisons de la presse - Roman. [Verkkojulkaisu.] Viitattu 3.10.2023. Saatavissa:

<https://www.babelio.com/prix/6/des-Maisons-de-la-presse>

Sommerlad, Joe. 24.3.2021. *Google Translate: How does the multilingual interpreter actually work?* [Verkkojulkaisu.] Viitattu 16.1.2024. Saatavissa:

<https://www.independent.co.uk/tech/how-does-google-translate-work-b1821775.html>

Vanmassenhove, Eva – Hardmeier, Christian – Way, Andy 2018. Getting Gender Right in Neural Machine Translation. Konferenssijulkaisu (*Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, s. 3003–3008). Brussels, Belgium. Association for Computational Linguistics. [Preprint]. Viitattu: 19.1.2024. Saatavissa: 10.18653/v1/D18-1334

VISK = Auli Hakulinen, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen ja Irja Alho 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio, viitattu 9.10.2023. Saatavissa: <http://scripta.kotus.fi/visk> URN:ISBN:978-952-5446-35-7

Wong, KayYen – Sameen, Maruf – Haffari, Gholamreza 2020. Contextual Neural Machine Translation Improves Translation of Cataphoric Pronouns. [Preprint]. Viitattu: 19.1.2024. Saatavissa: <https://doi.org/10.48550/arXiv.2004.09894>

Liite: Résumé en français

Introduction

L'objectif de cette étude est de comparer trois moteurs de traduction automatique. Les langues concernées dans cette étude sont le français et le finnois. L'étude a pour objet d'étudier les pronoms français suivants : *elle, il, elles, ils, l', le, la* et *les*. Ces pronoms ont été choisis parce que la plupart d'entre eux peuvent indiquer le genre de leur référent. Ce qui rend cette étude intéressante est le fait que dans le finnois le genre n'est pas marqué dans les pronoms personnels. En français, il est possible de faire référence à quelqu'un ou à quelque chose dans certains cas en utilisant simplement un pronom, ce qui ne serait pas possible en finnois à cause des différences entre ces deux langues.

Dans cette étude, il est question d'étudier si les moteurs de traduction automatique sont capables de traduire les pronoms dans une façon correcte et claire. L'étude permet également d'étudier quelles sont les traductions utilisées par les moteurs de traduction automatique. Selon les traductions des moteurs, il est également possible de voir quelles stratégies sont utilisées. Dans cette étude, nous avons pu identifier quatre stratégies différentes qui sont présentées plus en détail dans la section consacrée aux résultats.

Cadre théorique

Les moteurs de traduction automatique sont utilisés de plus en plus. En fait, il est important d'évaluer les moteurs, car pour pouvoir les développer, il faut savoir ce dont ils sont capables et notamment ce qui pose problème. Quand on sait où sont les problèmes, il est possible de commencer à développer les moteurs de traduction automatique pour résoudre ces problèmes. Comme Doherty (2019 : 340) mentionne, il est impossible de séparer l'évaluation du développement.

Les moteurs de traduction automatique sont utilisés pour traduire un texte d'une langue à une autre. Il existe différents types de moteurs de traduction automatique qui peuvent être classifiés selon leur façon de fonctionnement. Les trois types qui sont probablement les plus courants sont les suivants : la traduction à base de règles, la traduction statistique et la traduction automatique neuronale. Les trois moteurs étudiés dans cette étude sont tous les trois des moteurs de traduction automatique du troisième type : la traduction automatique neuronale. La qualité des traductions réalisées par les moteurs qui utilisent de la traduction

automatique neuronale est souvent encore mieux que les résultats des autres types des moteurs de traduction automatique.

Les moteurs de traduction automatique étudiés dans cette étude sont Google Translate, DeepL Translator et ChatGPT. Les versions gratuites ont été utilisées. Il faut noter que ChatGPT n'est pas un moteur de traduction automatique mais un agent conversationnel qui est basé sur intelligence artificielle. Il est quand même possible de l'utiliser pour faire des traductions et c'est pour cette raison qu'il fait partie de cette étude. Il n'est pas clair comment les différents moteurs de traduction automatique fonctionnent, mais il est possible que la mode de fonctionnement de ChatGPT en particulier est différente des moteurs de traduction automatique traditionnels. Il est donc intéressant de voir comment il fonctionne.

Pour traduire les pronoms, le contexte est important. Souvent, les pronoms font référence à un mot qui a été utilisé antérieurement. C'est notamment le cas avec les pronoms étudiés dans cette étude. Il est pour cette raison qu'en plus de la phrase qui contient le pronom, les deux phrases précédentes ont également été traduites par les moteurs de traduction automatique. Il n'est quand même pas clair comment les moteurs fonctionnent et s'ils utilisent le contexte ou pas. Mais si le contexte précédent a un effet sur la traduction, il est notamment le contexte précédent qui est important dans le cadre de cette étude.

Plus de contexte pourrait donner comme résultat une traduction plus correcte et plus claire, mais dans cette étude ce n'est pas possible, puisque nous avons utilisé des moteurs de traduction automatique gratuits. En utilisant des moteurs gratuits il est possible que les phrases traduites soient utilisées pour développer les moteurs de traduction automatique. Si cela est le cas, il peut y avoir des problèmes avec le droit d'auteur. Pour ces raisons, le contexte a été limité à deux phrases précédentes ce qui offre un peu de contexte. Comme on peut voir plus tard dans les résultats, le contexte précédent a eu un effet sur les traductions.

Méthodologie

Pour évaluer les phrases, un modèle d'évaluation a été créé. Il existe des modèles d'évaluation pour les traductions et des modèles qui sont spécifiquement conçus pour les traductions des moteurs de traduction automatique, mais il était mieux de créer un modèle qui convient à cette étude. Au lieu de se concentrer sur les fautes et les classer, nous avons décidé d'utiliser un modèle où on détermine si les traductions atteignent un certain niveau spécifié. Ce modèle se consiste en trois questions. Est-ce que la phrase a été traduit suffisamment bien pour qu'on

puisse analyser la traduction du pronom ? Est-ce que la référence est correcte ? Et finalement peut-être la question la plus importante : Est-ce que la référence est claire ?

Les questions étudiées dans cette étude sont très spécifiques et c'est pour cette raison qu'un nouveau modèle d'évaluation a été créé. Il existe quand même des problèmes quand on utilise un nouveau modèle. Il est impossible ou très difficile de comparer les résultats de cette étude avec d'autres études sur le même sujet. Malgré cela, nous avons décidé de l'utiliser parce qu'il permet d'étudier les questions pertinentes de cette étude.

Pour étudier les pronoms, un roman en français a été choisi. Il s'agit de roman qui s'appelle *La Valse lente des tortues* (2008) par Katherine Pancol. La traduction en finnois de ce roman (*Kilpikonnién hidás valssi*, Lotta Toivanen, 2012) fait également parti de cette étude. La traduction fonctionne comme une base. Les pronoms étudiés ont été choisis selon cette traduction car nous n'étudions que les pronoms qui ont été traduits par quelque chose d'autre qu'un pronom en finnois.

Au total, 249 pronoms ont été étudiés dans cette étude. Il est donc possible d'étudier seulement les phrases les plus intéressantes qui posent plus probablement des problèmes pour les moteurs de traduction automatique. Si la traductrice a utilisé quelque chose d'autre que le pronom, il est possible qu'un pronom ait rendu la phrase difficile à comprendre. L'utilisation d'un pronom en finnois n'est pas toujours possible même s'il est utilisé en français. Cette différence est liée au genre et elle a été décrite plus haut. Il existe quand même aussi d'autres raisons pour utiliser quelque chose d'autre qu'un pronom dans la traduction. Parfois les traducteurs font des traductions qui sont plus explicites que les textes originaux. Ce phénomène est appelé « explicitation ». (Behrens, 2022 : 102). Il est un des universaux de la traduction (Kuusi, 2011 : 14).

Les pronoms ont été divisés en trois groupes. Le premier groupe se compose des phrases où il est possible que l'utilisation d'un pronom ait pour résultat une traduction difficile à comprendre. Le deuxième groupe en revanche se compose des phrases où l'utilisation d'un pronom dans la traduction ne pose pas de problèmes pour la compréhension. Le troisième groupe est un peu particulier car il se compose des phrases où un pronom a déjà été utilisé. Il est donc impossible d'utiliser le même pronom une autre fois dans la même phrase. L'hypothèse concernant ces groupes est la suivante : les références des phrases dans le premier groupe posent probablement plus de problèmes que celles dans le deuxième groupe.

Résultats

Comme déjà mentionné plus haut, 249 phrases ont été examinées. Ces phrases ont été divisées en trois groupes, comme on peut voir dans le tableau 1 ci-dessous. La plupart des phrases font partie du premier groupe (130). Dans le deuxième groupe, il y a 109 phrases et seulement 10 phrases dans le troisième groupe. Le nombre des pronoms est similaire dans les deux premiers groupes, mais le troisième groupe ne contient pas beaucoup de pronoms. Il est donc difficile de tirer des conclusions concernant ce troisième groupe.

Tableau 1. Classification selon la traduction officielle

Groupe	Nombre dans le texte
1	130
2	109
3	10
Total	249

La première question dans le modèle d'évaluation était la suivante : est-ce que la phrase a été traduit suffisamment bien pour qu'on puisse analyser la traduction du pronom. Les résultats pour cette première question étaient très réussis : seulement 4 phrases n'étaient pas suffisamment bien traduites. Quant à la deuxième question, il s'agissait d'analyser si la référence était correcte. Pour cette question, les résultats étaient également bons : au total, 10 phrases avaient une référence qui n'était pas correcte. Ces résultats montrent que la plupart des pronoms étaient traduits d'une façon correcte. Mais la question la plus importante est si ces références étaient claires. Cette question est traitée dans le tableau 2 ci-dessous.

Tableau 2. Est-ce que les références sont claires ?

	Google Translate	%	DeepL	%	ChatGPT	%
La référence est claire	81	32,5	142	57,0	94	37,8
La référence n'est pas claire	163	65,5	102	41,0	151	60,6
L'analyse est impossible	5	2,0	5	2,0	4	1,6
Total	249	100,0	249	100,0	249	100,0

Comme on peut voir dans le tableau 2 ci-dessus, il existe plusieurs phrases où la référence n'est pas claire. Il existe également des différences entre les moteurs de traduction automatique. 142 phrases (57,0 %) ont une référence claire quand il s'agit les traductions

faites par DeepL. Quant à Google Translate, 81 pronoms (32,5 %) ont été traduits d'une façon claire. Les résultats de ChatGPT sont similaires à Google Translate : 94 références claires (37,8 %). Selon ces résultats, DeepL a le mieux réussi de traduire les phrases d'une façon claire. Il faut quand même noter que tous les moteurs de traduction automatique ont produit un grand nombre de phrases où la référence n'était pas claire. Cela montre qu'il existe toujours de difficultés avec la traduction des pronoms du français en finnois. Mais, selon ces résultats on peut dire que le moteur de traduction automatique qui a le mieux réussi avec ces phrases est DeepL. La différence est assez claire quand on examine les pourcentages donnés au tableau 2.

Dans cette étude, nous avons également étudié les différents pronoms individuellement. Le nombre de pronoms étant très limité, il est impossible de tirer de conclusions des différences entre les pronoms. Pour étudier ces questions, il faudrait étudier un corpus considérablement plus vaste.

En ce qui concerne les résultats selon les différents groupes, il semble que l'hypothèse présentée plus haut a été confirmée. Dans le tableau ci-dessous, on peut voir que les références dans le deuxième groupe étaient considérablement plus souvent claires que les références dans les deux autres groupes. Dans ce tableau, les résultats de tous les moteurs de traduction automatique ont été combinés. Cela permet de comparer les résultats de tous les moteurs selon les groupes. Le nombre des pronoms est donc 747. Pour pouvoir comparer ces chiffres, les pourcentages sont également donnés dans le tableau.

Tableau 3 Est-ce que les références sont claires ? (Les résultats de tous les moteurs de traduction automatique)

	Claire	%	Pas claire	%	L'analyse est impossible	%	Total
1	78	20,0	306	78,5	6	1,5	390
2	234	71,6	85	26,0	8	2,4	327
3	5	16,7	25	83,3	0	0,0	30
Total	317	42,4	416	55,7	14	1,9	747

Au total, 234 références du deuxième groupe étaient claires, ce qui signifie que 71,6 pourcents étaient claires. En revanche, dans le premier groupe seulement 78 références étaient claires, soit 20,0 pourcents. Les résultats du troisième groupe étaient encore moins bien réussis : seulement 5 références claires (16,7 %).

Les traductions réussies des moteurs de traduction automatique peuvent être divisées en quatre groupes. Il est possible de dire que ces groupes utilisent des stratégies, même si les moteurs de traduction automatique ne fonctionnent pas de la même manière que les traducteurs humains et il est impossible de dire qu'ils utilisent des stratégies d'une façon consciente. Mais si on essaie de trouver des tendances dans les traductions, il est possible de faire une classification. Le tableau 4 ci-dessous représente cette classification.

Tableau 4 Classification des traductions réussies

1	nom qui indique le genre
2	nom propre
3	la référence est claire grâce aux autres éléments de la phrase
4	la référence est claire grâce au contexte

Le premier groupe consiste en des noms qui indiquent le genre. Ce sont des mots comme *mies* (homme), *nainen* (femme), *äiti* (mère) ou *poika* (fils). Dans le deuxième groupe sont les phrases où un nom propre a été utilisé. Il s'agit donc du nom de personnage dans le texte. Quant au troisième groupe, il contient des phrases où la référence est claire grâce aux autres éléments de la phrase. Le quatrième groupe est similaire au troisième groupe, mais dans ces phrases, la référence est claire grâce au contexte autour de la phrase. Pour les traductions qui n'étaient pas réussies, une classification n'a pas été faite.

En ce qui concerne les traductions les plus souvent utilisées par les moteurs de traduction automatique, il est clair que des pronoms y apparaissent souvent. Les pronoms utilisés étaient *hän* (il/elle, utilisé pour les personnes) et *se* (il/elle, utilisé pour les objets inanimés). En plus, DeepL a parfois utilisé les mots suivants : *nainen* (femme) et *mies* (homme). Les noms propres étaient également utilisés. Mais il est clair que les traductions utilisées par les moteurs de traduction automatique étaient remarquablement moins variées que les traductions de la traductrice.

Conclusion

Pour conclure, nous allons proposer quelques sujets intéressants pour les recherches futures, car il en existe plusieurs. Tout d'abord, le problème avec la traduction des pronoms n'est pas une particularité qui concerne seulement les traductions du français en finnois. Il y a d'autres langues où un phénomène similaire existe, et il serait intéressant d'étudier si les moteurs de

traduction réussissent mieux dans d'autres paires de langue. Il est également possible que la même étude donnerait des résultats différents après un certain temps, car le développement de la technologie est rapide. Il est donc possible que certains moteurs de traduction automatique réussissent à traduire les pronoms dans une façon plus claire. En plus, il serait intéressant d'étudier l'influence du contexte sur les traductions. Dans cette étude, seulement les deux phrases antérieures ont été utilisées pour le contexte. Il serait donc intéressant de traduire des passages plus longs et voir quel effet le contexte a pour la traduction.