

HELSINGIN YLIOPISTO

Laajan kielimallin hyödyntäminen kolumnien kääntämisessä

Tapaustutkimus suomi–saksa-kieliparin tekoälykäännösten laadusta ja niiden
jälkieditoinnin kognitiivisesta työmäärästä

Kääntämisen ja tulkkauksen maisteriohjelma, saksa–suomi
Maisterintutkielma

Laatija:
Juuli Muona

Ohjaaja:
FT Marjut Alho

6.5.2025
Helsinki

Tiedekunta: Humanistinen tiedekunta

Koulutusohjelma: Kääntämisen ja tulkkauksen maisteriohjelma

Opintosuunta: Käännös- ja tulkkausviestintä, saksa–suomi

Tekijä: Juuli Muona

Työn nimi: Laajan kielimallin hyödyntäminen kolumnien kääntämisessä –
Tapaustutkimus suomi–saksa-kieliparin tekoälykäännösten laadusta ja niiden
jälkieditöinnin kognitiivisesta työmäärästä

Työn laji: Maisterintutkielma

Kuukausi ja vuosi: Toukokuu 2025

Sivumäärä: 62 sivua, 10 liitesivua

Avainsanat: konekääntäminen, tekoäly, ChatGPT, kolumni, tekstianalyysi,
virheanalyysi, jälkieditointi, kognitiivinen kuormitus, tapaustutkimus

Ohjaaja tai ohjaajat: Marjut Alho

Säilytyspaikka: Helda – Helsingin yliopiston avoin julkaisuarkisto

Muita tietoja:

Tiivistelmä: Maisterintutkielma käsittelee laajan kielimallin avulla käännettyjen kolumnien laatua ja niiden jälkieditöinnin kognitiivista työmäärää. Tutkielma on osa Aikakausmedian kansainvälistäminen -projektia, joka on toteutettu Helsingin yliopiston ja mediakonserni A-lehtien yhteistyönä. Projektin tarkoituksena on selvittää, miten generatiivista tekoälyä voidaan hyödyntää lifestyle-median kansainvälistämisessä. Tutkimuksessa pyritään selvittämään, kuinka hyvin ChatGPT 4.0 onnistuu kolumnien kääntämisessä, minkälaisia virheitä laaja kielimalli tekee käännöksissä, mikä on virheiden kognitiivinen työmäärä ja miten käännöksiä tulisi jälkieditöidä, jotta ne olisivat julkaisukelpoisia.

Tutkielman aineisto koostuu viidestä A-lehtien Meillä kotona -sivuston kolumnista, jotka on käännetty A-lehtien käyttöön koulutetulla ChatGPT 4.0 -tekoälytyökalulla suomesta saksaan. Kolumnien pituus vaihtelee 250 sanan ja 500 sanan välillä.

Aineiston laajuus on yhteensä 1 910 sanaa (käännökset 2 808 sanaa). Tutkimus on monimenetelmäinen. ChatGPT:llä käännetyt kolumnit on ensin jälkieditoitu. Tämän jälkeen jälkieditöinnissa tehdyt muutokset on luokiteltu Irina Temnikovan (2010) virheluokittelulla, jonka tarkoitus on selvittää yksittäisten virheiden kognitiivista työmäärää. Lisäksi tutkimuksessa on seurattu jälkieditöintiin kuluvaan aikaa.

Tutkimuksen teoriaosuudessa avataan kolumnien kääntämiseen, jälkieditointiin, käännosten laadun arvioimiseen ja tekoölyyn liittyvää tutkimusta.

Tutkimustuloksista käy ilmi, että kolumnit vaativat paljon muokkausta, jotta ne olisivat julkaisukelpoisia. Suurin osa aineistossa ilmenneistä virheistä liittyy idiomaattisuuteen ja yksittäisiin sanavalintoihin. Virheet johtuvat usein siitä, että tekoöly on kääntänyt tekstin hyvin suoraan tai se ei ole ottanut huomioon kulttuurista kontekstia. Suurin osa virheistä ei kuitenkaan ole kognitiivisesti kovin kuormittavia ja vain muutamat virheistä ovat kognitiivisesti erittäin kuormittavia. Aineistosta löytyy yhteensä 213 virhettä ts. muokkausta, jotka on luokiteltu Temnikovan virheluokkiin. Jälkieditointiin kulunut aika vaihtelee. Keskimäärin jälkieditointiin kului 84 minuuttia/teksti. Vaikka kolumneista löydetty virheet eivät ole kognitiivisesti kovin kuormittavia, voi jälkieditointiin kuluneen ajan perusteella päätellä, että jälkieditointiprosessi on kognitiivisesti kuormittava.

Sisällysluettelo

1	Johdanto	1
2	Teoreettinen viitekehys	4
2.1	Kolumnit, tekoäly ja kolumnien kääntäminen	4
2.1.1	Kolumnin määritelmä ja tekoälyn hyödyntäminen kolumnien kirjoittamisessa	4
2.1.2	Kolumnien kääntäminen	6
2.2	Generatiivinen tekoäly ja laajat kielimallit	8
2.3	Jälkieditoiminen ja jälkieditoinnin kognitiivisen työmäärän mittaaminen	12
2.4	Muita tapoja arvioida konekäännösten laatua	16
2.4.1	Konekäännösten automaattinen arviointi	17
2.4.2	Ihmisen tekemä arviointi: virheluokittelu	18
3	Tutkimusmenetelmät ja aineisto	22
3.1	Tutkimusmenetelmät	22
3.1.1	Lähestymistapa	22
3.1.2	Valitut tutkimusmenetelmät	23
3.1.3	Virheanalyysi	24
3.2	Analyysin vaiheet	26
3.3	Aineisto	29
4	Analyysi	32
4.1	Väärin tunnistetut idiomaattiset ilmaisut (luokka 3.5.)	34
4.2	Väärät sanat tai fraasit: tyyliin liittyvät virheet (luokka 3.3.)	37
4.3	Puuttuvat sanat, fraasit, lauseet, virkkeet tai selvennykset (luokka 1.)	40
4.4	Muut virheet	43
4.5	Tutkimustulosten yhteenveto	47
5	Johtopäätökset	52
	Lähteet	55
	Primäärilähteet	55
	Sekundäärilähteet	56
	Liitteet	63
	Liite 1. Jälkieditoidut kolumnit virheluokitteluineen	63

Liite 2. ChatGPT:lle annettu käännöskehote	71
Liite 3. ChatGPT:lle annettu jälkieditointikehote (käännös 2)	72
Deutschsprachige Kurzfassung	1
TAULUKKO 1. TEMNIKOVAN VIRHELUOKITTELU (2010)	24
TAULUKKO 2. TEMNIKOVAN VIRHELUOKITTELUN (2010) VIRHEET LUOKITELTUNA KIELEN ERI TASOILLE JA JÄRJESTETTYNÄ KOGNITIIVISEN TYÖMÄÄRÄN MUKAAN	25
TAULUKKO 3. KÄÄNNÖKSISTÄ LÖYTYNEET VIRHEET JAOTELTUNA TEMNIKOVAN (2010) VIRHELUOKKIIN	32
TAULUKKO 4. JÄLKIEDITOINTIIN KULUNUT AIKA	50
TAULUKKO 5. SANAMÄÄRÄT ALKUPERÄISTEKSTEISSÄ, CHATGPT:N KÄÄNNÖKSISSÄ JA JÄLKIEDITOIDUISSA KÄÄNNÖKSISSÄ	51
KAAVA 1. KÄÄNNÖKSISTÄ LÖYTYNEIDEN VIRHEIDEN KOGNITIIVINEN TYÖMÄÄRÄ TEMNIKOVAN (2010) LUOKITTELUN MUKAAN	48

1 Johdanto

Generatiivisen tekoälyn ja laajojen kielimallien käyttö on kasvanut viime vuosina räjähdysmäisesti. Vuoden 2022 lopulla OpenAI julkisti tekoälytyösovellus ChatGPT:n, joka toi kehittyneen tekoälytyökalun suuren yleisön saataville. Kahdessa kuukaudessa sovelluksella olikin jo 100 miljoonaa käyttäjää (Milmo 2023). Nykyisin tekoälyä hyödynnetään mitä moninaisempiin tehtäviin, ja erityisesti laajat kielimallit kuten ChatGPT ovat osoittaneet potentiaalinsa myös kääntämisessä. Sitä, miten hyvin generatiivinen tekoäly tai laajat kielimallit onnistuvat erilaisten tekstien kääntämisessä, on kuitenkin tutkittu vasta vähän. Tekoäly myös kehittyy ja sitä kehitetään jatkuvasti, joten uusien ohjelmistoversioiden ja käännösten laatuun vaikuttavien tekijöiden tutkiminen on tarpeellista.

Tutkielman tarkoituksena on selvittää, miten laaja kielimalli onnistuu kolumnien kääntämisessä suomesta saksaan ja mikä on käännösten jälkieditoinnin kognitiivinen työmäärä. Tutkimukseni on osa mediakonserni A-lehtien ja Helsingin yliopiston Aikakausmedian kansainvälistäminen -projektia. Projektin tarkoituksena on selvittää, miten generatiivista tekoälyä voidaan hyödyntää lifestyle-median kansainvälistämisessä. Projektissa on käännetty A-lehtien Meillä kotona -sivuston lifestyle-artikkeleita ChatGPT:llä suomesta saksaan ja englantiin. Sivusto sisältää eri tekstilajeja aina resepteistä henkilökuviin. (A-lehdet 2025.) Tutkimukseni keskittyy näistä yhteen, Meillä kotona -sivuston kolumneihin. Helsingin yliopiston projektukurssilla keväällä 2024 kävi ilmi, että ChatGPT:llä on hankaluuksia kääntää Meillä kotona -sivuston tekstejä suomesta saksaan. Valitsin tutkimuskohteeksi kolumnit, sillä arvelin, että niiden kääntäminen aiheuttaisi tekoälylle eniten haasteita tekstilajin ominaispiirteiden vuoksi.

Olen valinnut tutkimukseen viisi Meillä kotona -sivuston kolumnia, joiden pituus vaihtelee noin 250 sanasta 500 sanaan. Käännätin kolumnit tutkimusta varten A-lehtien omalla ChatGPT 4.0 -pohjaisella tekoälysovelluksella työkaluun valmiiksi syötettyä kehotetta hyödyntäen. Yksi käännöksistä on osana kurssityötä myös jälkieditoitu ChatGPT:llä ennen ihmisen jälkieditointia.

Jälkieditoin ChatGPT:n kääntämät tekstit ensin itse, jonka jälkeen saksan kielen ja kääntämisen lehtori Sanne Tschirpke muokkasi ja kommentoi editoimiani tekstejä. Kahden tutkielmani kolumnin jälkieditoimiseen osallistui lisäksi toinen saksan kielen ja kääntämisen opiskelija Helsingin yliopistosta, Anna Verronen.

Jälkieditointiprosessin aikana mittasimme myös editoimiseen kulunutta aikaa. Haluan kiittää Sannea ja Annaa heidän arvokkaasta avustaan kolumnien jälkieditoinnissa.

Jälkieditointiprosessin päätteeksi tein itsenäisesti virheanalyysin kolumneihin Irina Temnikovan (2010) virheluokittelua hyödyntäen. Virheluokittelun avulla analysoin paitsi käänöksissä esiintyneitä virheitä myös niiden jälkieditoinnin kognitiivista työmäärää. Keskustelin käännösten laadusta ja kognitiivisesta työmäärästä myös em. kahden muun saksan kielen asiantuntijan kanssa. Valitsemieni menetelmien avulla pyrin tutkielmassani vastaamaan seuraaviin tutkimuskysymyksiin:

1. Kuinka hyvin ChatGPT 4.0 onnistuu kolumnien kääntämisessä suomesta saksaan?
2. Minkälaisia virheitä laaja kielimalli tekee ko. käänöksissä?
3. Mikä on virheiden korjaamisen kognitiivinen työmäärä?
4. Miten käännöksiä tulisi jälkieditoida, jotta ne olisivat julkaisukelpoisia?

Hyödynnän siis tutkielmassani Irina Temnikovan konekäännösten virheluokittelua, jonka tarkoituksena on analysoida virheiden korjaamisen kognitiivista työkuormaa. Tekoälyllä käännettyjä tekstejä eikä tätä tekstilajia ole tietääkseni vielä juurikaan tutkittu tämän teorian valossa. Marjut Alho (tulossa) on osana samaa A-lehtien yhteistyöprojektia tutkinut Meillä kotona -sivuston eri tekstilajien käännöksiä ja hyödyntänyt Temnikovan virheluokittelua niiden analysoimiseen. Hänen menetelmänsä ja aineistonsa kuitenkin poikkeavat hieman tämän tutkimuksen menetelmästä. Kerron lisää alhon tutkimuksesta luvussa 2. Tutkielmani lopussa vertaan omia tutkimustuloksiani Alhon tuloksiin.

Hypoteesini on, että ChatGPT:llä on vaikeuksia kääntää kolumneja idiomattisesti ja virheettömästi. Odotan, että suurin osa virheistä liittyy nimenomaan idiomattisuuteen, mutta myös esimerkiksi sanajärjestykseen, välimerkkeihin ja kielioppiin. Odotan myös, että käännökset vaativat paljon jälkieditointia, jotta ne olisivat julkaisukelpoisia, ja että jälkieditoinnin kognitiivinen työmäärä on melko suuri.

On hyvä huomioida, että laajat kielimallit, niistä esimerkkinä tässä tutkimuksessa käytetty ChatGPT 4.0, eivät ole konekääntimiä. Niitä ei siis ole lähtökohtaisesti suunniteltu tai tarkoitettu nimenomaan kääntämiseen. Tästä näkökulmasta tekoälyllä tuotetut käännökset eivät siis ole konekäännöksiä. Suomen kieleen ei kuitenkaan ole vielä vakiintunut termiä, joka kuvaisi tekoälyn tuottamaa käännöstä. Siksi tässä tutkielmassa tekoälyllä tuotettuja käännöksiä käsitellään eräänlaisina konekäännöksinä. Keskustelen konekääntämisen ja tekoälyavusteisen kääntämisen eroista lisää luvussa 2.

Tutkimukseni tuottaa hyödyllistä tietoa paitsi A-lehdille, myös muille tekoälyavusteisesta kääntämisestä kiinnostuneille organisaatioille. Tutkimus on esimerkkitapaus siitä, miten ChatGPT onnistuu kolumnien kääntämisessä suomesta saksaan ja mitkä ovat laajan kielimallin heikkoudet ja vahvuudet kääntämisessä. Samalla se selventää sitä, minkälainen kyseisten käännösten jälkieditointiprosessi on ja onko tekoälyllä tuotettujen käännösten jälkieditointi ajallisesti ja siten myös taloudellisesti kannattavaa.

Tutkielmani rakenne on seuraava: avaan luvussa 2 tutkimukseni kannalta keskeistä teoriaa ja aiempaa tutkimusta muun muassa tekoälyavusteisesta kääntämisestä, jälkieditoinnista ja käännösten laadun arvioimisesta. Luvussa 3 esittelen tarkemmin tutkimusmenetelmiäni ja aineistoani. Luvussa 4 analysoin tutkimustuloksiani, esittelen esimerkkejä tyypillisimmistä virheistä aineistossani ja vastaan tutkimuskysymyksiin. Luvussa 5 kokoan havaintoni yhteen, reflektoin tutkimukseni onnistuneisuutta ja esitän ehdotuksia jatkotutkimukselle.

2 Teorettinen viitekehys

Luvun tarkoituksena on selventää tutkimukseni keskeisimpiä käsitteitä ja tutkimani ilmiön teoreettista taustaa. Kerron ensin kolumnista tekstilajina ja määrittelen, mitä kolumnilla tarkoitetaan tässä tutkimuksessa. Käyn kappaleessa lyhyesti läpi myös kolumnien kääntämistä.

Tekstilajin taustoittamisen jälkeen avaan lyhyesti käännöksiin hyödyntämäni laajan kielimallin toimintaa ja taustaa. Tutkimukseni tarkoituksena ei ole perehtyä syvällisesti tekoälyn toimintaan, joten keskityn kertomaan käytännönläheisesti tekoälyn mahdollisuuksista, erityisesti kääntämisen näkökulmasta. Myöskään lukijan ei tarvitse tuntea yksityiskohtaisesti tekoälyn toimintaa voidakseen seurata tutkimukseni kulkua. Koska tekoälykäännösten tutkimus on vielä verrattain uutta, sisältää luku 2.2 myös vertaisarvioimattomia versiota julkaisuista (*Preprint*). Julkaisujen tarkoitus on esitellä tutkimusten alustavia tuloksia tutkimusyhteisölle.

Tutkimukseni aineisto koostuu tekoälyllä käännettyjen tekstien jälkieditoituista versioista, joten avaan tässä luvussa jälkieditointia käsitteenä. Lisäksi kerron, miten jälkieditoinnin kognitiivista työmäärää on aiemmissa tutkimuksissa arvioitu. Lopuksi kartoitan, minkälaisin menetelmin konekäännösten laatua voidaan yleisesti arvioida. Kuten tutkielman ensimmäisessä luvussa mainitsin, laaja kielimalli ei ole varsinaisesti konekäännin esimerkiksi neuroverkkopohjaisten konekäänninten tapaan. Laajoilla kielimalleilla käännettyjä tekstejä voidaan kuitenkin lähestyä eräänlaisina konekäännöksinä. Siksi avaan teorettisessa viitekehyksessä konekäännösten tutkimusta ja arvioin esimerkiksi konekäännösten laadun tai työmäärän arvioimiseen suunnattujen mittareiden soveltuvuutta tekoälyllä tehtyihin käännöksiin.

2.1 Kolumnit, tekoäly ja kolumnien kääntäminen

2.1.1 Kolumnin määritelmä ja tekoälyn hyödyntäminen kolumnien kirjoittamisessa

Kolumni voidaan määritellä journalistisena tekstinä, joka on kirjoitustyyliiltään vapaa. Tyypillistä kolumnille on pohdiskeleva ja usein myös humoristinen tai ironinen lähestymistapa (Utismedian liitto n.d.). Toisaalta kolumni voi olla myös asiatyylinen (Mikkola ym. 2004, 196). Kolumnien pituus vaihtelee ja niiden

kirjoittaja voi olla joko julkaisevan median työntekijä tai median ulkopuolinen kolumnisti (Uutismedian liitto n.d.). Kolumnisti päättää itse kolumnin aiheen, ja tekstissä esitetyt mielipiteet ovat kirjoittajan omia (Mikkola ym. 2004, 193).

Vaikka kolumni voidaan määritellä julkaisukontekstin perusteella journalistiseksi tekstiksi, on kolumneissa myös kaunokirjallisia piirteitä. Kolumnien kirjoittamiseen kytkeytyy luovuus ja niiden ensisijainen tarkoitus on tarjota lukijalle kokemus (Hadley ym. 2022, 6). Kerron myöhemmin tässä luvussa tarkemmin kolumnien kielellisistä piirteistä ja siitä, mitä niiden kääntämisessä tulee ottaa huomioon.

Tutkimuksessani analysoituja kolumneja ei ole valittu tietyn ulkopuolisen määritelmän perusteella, vaan julkaisija on itse määritellyt ne kolumneiksi. Tekstit kuitenkin täyttävät edellä mainitut tekstilajin piirteet, sillä ne sisältävät muun muassa huumoria ja ironiaa. Lisäksi niiden tyyli on vapaa, ne pohjautuvat kirjoittajan henkilökohtaisiin kokemuksiin tai mielipiteisiin ja kirjoittaja tuo niissä persoonallisuuttaan esiin. Esittelen tutkimukseen valittuja kolumneja tarkemmin luvussa 3. Kolumnin ominaispiirteet on syytä pitää mielessä myös tutkimukseni käännöksiä arvioitaessa. Esimerkiksi tekstien julkaisukelpoisuus nimittäin liittyy julkaisevan tahon ohjeiden lisäksi myös siihen, miten hyvin ne täyttävät tekstilajin konventiot.

Keskeistä kolumnissa on, että se keskittyy johonkin tiettyyn ilmiöön tai tapahtumaan, josta kirjoittaja on kiinnostunut. Hyvä kolumnisti kiinnittää lukijan huomion ja saa hänet kiinnostumaan aiheesta. Tätä edesauttaa se, että kirjoittaja kirjoittaa omana itsenään ja tuo rohkeasti persoonaansa esiin tekstissä. (Kotilainen 2003, 68.)

Kolumni siis liitetään usein inhimillisiin kokemuksiin ja persoonalliseen tyyliin.

Kolumnien keskiössä on siis pitkään ollut nimenomaan ihmiskirjoittaja. Viime vuosina kuitenkin myös tekoälyä on alettu hyödyntämään kolumnien ja muiden journalististen tekstien luomiseen (ks. esim. Körner 2024; Tekniikka&Talous 2024). Esimerkiksi saksalainen päivälehti *Die Tageszeitung* on julkaissut vuodesta 2022 alkaen kuukausittain ChatGPT:llä kirjoitettuja kolumneja (taz 2025). Tekstit eivät kuitenkaan edes tekoälyn avulla synny sormia napsauttamalla. Kilgin ym. (2022) mukaan esimerkiksi sopivan kehotteen löytämiseen voi kulua aikaa. Tekstin täytyy nimittäin olla sopivan mittainen ja tiettyjen laatukriteerien mukainen. Kerron

seuraavassa alaluvussa (2.2) lisää tekoälyn toiminnasta ja siitä, minkälaisia haasteita saattaa liittyä kolumnien kääntämiseen tekoälyllä.

Journalististen tekstien luominen automaattisesti tietokoneella ei ole uusi ilmiö. Esimerkiksi automaattista tekstinsyöttöä on hyödynnetty journalismissa jo ennen generatiivisen tekoälyn käytön yleistymistä (Graefe ja Bohlken 2020, 50). Graefe ja Bohlken (2020) ovat tutkineet tietokoneella luotujen journalististen tekstien vastaanottoa. Määrällisen tutkimuksen tarkoituksena oli kartoittaa tuloksia vuosina 2017–2020 tehdyistä tutkimuksista, joissa tutkittiin journalististen tekstien vastaanottoa. Tutkimuksessa analysoitiin yhteensä 4 473 vastaajan vastausta ja tavoitteena oli tutkia lukijoiden mielikuvaa tekstien luotettavuudesta, laadusta ja luettavuudesta. Uutisartikkelit painoutuivat urheiluun, talouteen ja politiikkaan (Graefe ja Bohlken 2020, 52). Tutkimuksessa ei eritellä tai vertailla tekstin tuottamisessa hyödynnettyjä ohjelmia. Tuloksissa oman tutkimuksen kannalta kiinnostavaa on erityisesti se, että vastaajat arvioivat ihmisen tuottamat tekstit selvästi luettavammiksi kuin automaattisesti luodut tekstit. Tekstien luotettavuudessa ja laadussa taas ei ollut lukijoiden arvion mukaan merkittäviä eroja. Kiinnostavaa on myös, että ihmisen tuottamat tekstit arvioitiin paremmin, jos lukijalla oli tiedossa, että tekstin oli kirjoittanut ihminen (emt. 57–58). Vaikka tutkimuksessa käsitellyt tekstit eivät ole käännettyjä kolumneja, tarjoaa Graefen ja Boehlken tutkimus (2020) kiinnostavaa kokoavaa tietoa siitä, miten lukijat kokevat tietokoneen luomat tekstit verrattuna ihmisen kirjoittamiin teksteihin.

2.1.2 Kolumnien kääntäminen

Kolumneja ja niissä käytettyä kieltä on tutkittu erilaisista näkökulmista. Kolumnien kääntäminen ei kuitenkaan ole kovin laajasti tutkittu aihealue. Kolumneja, kuten muitakin journalistisia tekstejä, kuitenkin käännetään myös vieraille kielille (ks. esim. Kuusela ja Orell 2023). Käyn seuraavaksi läpi tarkemmin asioita, joita kolumnien kääntämisessä on hyvä ottaa huomioon. Tämän jälkeen kerron lyhyesti siitä, miten kolumnien kääntämistä on aiemmin tutkittu.

Yksi kolumnien merkittävin kielellinen ominaispiirre ovat idiomaattiset ilmaisut ja idiomit. Rune Ingo (1990, 245) määrittelee idiomit fraaseina tai ilmauksina, ”joiden merkitystä ei aina ole mahdollista päätellä yksittäisten ilmaukseen kuuluvien sanojen

perusteella.” Idiomi on siis eräänlainen merkityskokonaisuus, jonka merkitys syntyy kokonaisuudesta, ei sen yksittäisistä osista. Jokaisessa kielessä on omat idiominsa (osittain myös samoja), jotka tulee ottaa huomioon kääntämisessä. (emt.) Jos mahdollista, tulisi kohdekielestä aina löytää lähdekielen idiomia vastaava idiomi. Jos tällaista ei ole, täytyy idiomi selittää. Idiomi on teoriassa mahdollista kääntää myös sanasanaisesti, mutta tämä harvoin johtaa onnistuneeseen lopputulokseen. (Ingo 1990, 246–247.) Ingon mukaan (1990, 247) idiomeja käännettäessä on tärkeää ensin selvittää, mitä ilmauksella halutaan sanoa. Idiomien kääntämiseen sisältyy siis kaksi haastetta: itse idiomien tunnistaminen ja oikean käännösratkaisun löytäminen. Idiomien kääntämiseen liittyy siis aina mahdollisia kompastuskiviä, jopa ammattikäntäjälle (Ingo 1990, 246).

Idiomaattisilla ilmaisuilla viitataan tässä tutkimuksessa ilmauksiin, jotka ovat tietylle kielelle tyypillisiä. Ilmaisut ovat siis jollakin tavalla vakiintuneet äidinkielisten puhujien käyttöön huolimatta siitä, onko niiden merkitys pääteltävissä yksittäisistä sanoista vai ei. Tutkimusaineistossa voi esiintyä sekä idiomeja että muita idiomaattisia ilmaisuja.

Tehdessäni taustatutkimusta tätä työtä varten löysin suomalaisista korkeakouluista yksittäisiä kolumnien kääntämiseen liittyviä opinnäytetöitä. Esimerkiksi Kaisa Kujala (2011) on opinnäytetyössään kääntänyt suomalaiselle viittomakielelle kolumnin hyödyntäen kolmea eri lähestymistapaa kääntämiseen. Linné-yliopistossa Sofia Sundqvist (2011) taas on tutkinut kulttuurisidonnaisten ilmiöiden ja ilmaisujen kääntämistä amerikkalaisen kirjailijan Jeff Kliman blogin kolumneissa. Blogitekstit on siis julkaistu kolumneina amerikkalaisen *The Huffington Post* -uutissivustolle (nykyinen *HuffPost*). Sundqvistin opinnäytetyö on siinä mielessä oman tutkimukseni kannalta kiinnostava, että hän kiinnittää tutkimuksessa huomiota yleisten kulttuuristen näkökohtien (historiallisten tapahtumien, uskonnollisten juhlien, julkaisujen, kulttuuriin liittyvien stereotyyppien ja idiomien), kirosanojen ja henkilöviittausten kääntämiseen (Sundqvist 2011, 9–16). Omassa tutkimuksessani korostuvat myös kulttuurisidonnaiset ilmiöt ja ilmaisut ja niiden kääntämiseen liittyvät haasteet. Sundqvist on kääntänyt kolumnit siten, että ne julkaistaisiin ruotsinkieliselle yleisölle (emt. 6). Hän harkitsee erilaisia käännösratkaisuja ja lopputulemana on, että sanasanainen käänнос harvoin toimii kyseisissä teksteissä

(emt. 21). Oletettavasti myös omassa tutkimuksessani sanasanaiset käännökset kulttuurisidonnaisista ilmiöistä tai ilmaisuista eivät useinkaan toimi.

Taustatutkimukseni perusteella vaikuttaa, että nimenomaan tekoälyllä käännetyt kolumnit on kuitenkin aihealueena vielä melko tutkimaton. Seuraavaksi kerron lisää generatiivisesta tekoälystä ja havainnoista, joita tähän mennessä julkaistuissa tutkimuksissa on tehty sen hyödyntämisestä kääntämisessä.

2.2 Generatiivinen tekoäly ja laajat kielimallit

Sanastokeskuksen Tietotekniikan termitalkoot -projektissa generatiivinen tekoäly on määritelty tekoälynä, joka ”tuottaa sille annetun tekstisyötteen pohjalta uutta teksti-, kuva-, video- tai äänisisältöä” (Tietotekniikan termitalkoot 2024). Generatiivinen tekoäly siis viittaa tekoälyyn, joka tuottaa materiaalia erilaisissa muodoissa.

Tekoälystä keskusteltaessa esille nousevat monesti myös ns. laajat tai suuret kielimallit (*Large language models*, LLMs).¹ Laajat kielimallit toimivat samalla periaatteella kuin generatiivinen tekoäly ja perustuvat syväoppimiseen. Toisin kuin generatiivinen tekoäly, ne keskittyvät kuitenkin nimenomaan kielellisen sisällön tuottamiseen (Toner, 2023). Generatiivisen tekoälyn (esim. Microsoft Copilot) ja laajojen kielimallien (esim. ChatGPT) käsitteitä käytetään osittain myös toistensa synonyymeinä (emt.). Omassa tutkimuksessani voimme puhua käyttämästäni tekoälytyökalusta tarkemmin laajana kielimallina, sillä se on koulutettu ja tarkoitettu nimenomaan kielellisen sisällön tuottamiseen. On kuitenkin hyvä huomioida, että molempia käsitteitä käytetään joskus samaan tekoälyyn viitattaessa, ja niiden erot eivät aina ole täysin selviä.

Yhteistä erilaisille tekoälytyökaluille on, että ne pyrkivät jäljittelemään ihmiselle ominaista älyllistä toimintaa. Suurin osa kääntämisessä hyödynnettävistä tekoälyistä on joko neuroverkko pohjaisia konekääntimiä (esim. DeepL) tai laajoja kielimalleja. Molemmat hyödyntävät ns. syväoppimista, jossa tekoälyä koulutetaan syvien neuroverkkojen avulla. Keinotekoisten neuroverkkojen toiminta muistuttaa biologisia neuroverkkoja. (Siu 2023b, 2.)

¹ Käsitteitä ”laaja kielimalli” ja ”suuri kielimalli” käytetään synonyymeinä ((Tietotekniikan termitalkoot 2024b).

Generatiivisen tekoälyn ja laajojen kielimallien toimintaperiaatteet kuitenkin eroavat neuroverkkokonekääntimien toiminnasta. Siun (2023, 6) mukaan laajojen kielimallien oppiminen on epäsuoraa, kun taas neuroverkkokääntimet on usein koulutettu rinnakkaiskorpuksilla tiettyä tehtävää varten. Laajat kielimallit eivät siis välttämättä tarvitse suoria esimerkkejä tai niille riittää muutama esimerkki, jotta ne voivat tuottaa käännöksen (emt., 6). Luova toimintaperiaate on varmasti yksi syy generatiivisen tekoälyn suosioon. Neuroverkkokääntimet ovat kuitenkin selvästi kehittyneempiä kuin esimerkiksi edeltäjänsä tilastopohjaiset kääntimet (emt., 7; Hadley ym. 2022, 3). Neuroverkkokäänninten tapaan tilastopohjaiset kääntimet perustuvat korpuksiin. Tilastopohjaiset kääntimet kuitenkin pilkkovat lauseet osiin, kun taas neuroverkkokääntimet prosessoivat kokonaisia lauseita. Neuroverkkokääntimet myös analysoivat matemaattisesti sanojen välisiä suhteita, mikä usein johtaa sujuvampaan käännökseen. (Hadley ym. 2022, 5.)

Neuroverkkokääntimistä ja tilastopohjaisista kääntimistä poiketen generatiivinen tekoäly pystyy suorittamaan laajasti erilaisia toimintoja kääntämisen lisäksi. Sen avulla voi muun muassa jälkieditoida ja oikolukea tekstejä sekä ideoida sisältöä. Siihen, miten hyvin tekoäly onnistuu näissä tehtävissä, vaikuttaa muun muassa käytetty ohjelma, kieli ja kehote (ks. esim. Hendy ym. 2023; Alam ym. 2024; Peng ym. 2023). Lisäksi, koska generatiivinen tekoäly on usein koulutettu hyvin laajalla aineistolla eri lähteistä, se voi soveltua neuroverkkokääntimiä paremmin luovien tekstien kääntämiseen (Hadley ym. 2022, 7). Luovat tekstit nimittäin sisältävät hyvin paljon erilaisia viittauksia ja tyylejä, joita tiettyihin korpuksiin perustuvan konekääntimen voi olla vaikeaa hahmottaa (emt.).

Khoshafah (2023) on tutkinut ChatGPT:llä tehtyjä eri tekstilajien käännöksiä englanti-arabia-kieliparissa. Hänen mukaansa tekoäly onnistui melko tarkasti välittämään lähtötekstin viestin, mutta ihmisen tekemät verrokkikäännökset olivat tyyllisestään onnistuneempia. ChatGPT:n käännökset eivät sisältäneet samaa syvyyttä, luonnollisuutta ja vivahteita kuin ihmisen tuottamat käännökset. Haasteita aiheuttivat myös muun muassa idiomaattiset ilmaisut. (emt.) Toisaalta esimerkiksi Siu (2023a) on tutkimuksessaan havainnut, että ChatGPT on onnistunut tekemään sopivissa kohdissa luovia käännösratkaisuja. On huomioitava, että tutkimuksissa on

erilainen lähestymistapa kehoitteisiin. Khoshafahin kehote (2023, 4) pyrkii varmistamaan käännöksen tarkkuuden. Siu (2023a, 22) taas korostaa tutkimuksessaan kontekstin tarjoamisen merkitystä kehoitteessa. Jos ChatGPT:lle antaa kattavat tiedot kohderyhmästä ja kontekstista, voi lopputuloksena olla hyvinkin onnistunut käänнос (emt.). Käsittelen kehoitteiden merkitystä laajoilla kielimalleilla tuotettuihin käännöksiin luvun lopussa.

Kuten konekääntimetkään ei generatiivinen tekoäly siis aina onnistu käännöksissä halutulla tavalla. Generatiivisen tekoälyn suurimmat haasteet kääntämisessä liittyvät muun muassa ns. hallusinointiin sekä erikoissanaston ja kontekstin tunnistamiseen (Siu 2023a; Siu 2023b; Bang ym. 2023; Bubeck ym. 2023). Hallusinointi on yksi generatiivisen tekoälyn merkittävimmistä haittapuolista verrattuna neuroverkkokääntimiin ja saattaa pahimmillaan johtaa merkittäviin väärinkäsityksiin. Hallusinoinnilla tarkoitetaan, että tekoäly lisää fiktiivisiä yksityiskohtia tekstiin tai muuhun tuottamaansa sisältöön. (Siu 2023b, 12.) Hallusinointi on kuitenkin harvinaisempaa kääntämisessä kuin muunlaisessa luonnollisen kielen tuottamisessa generatiivisella tekoälyllä. Tämä johtuu todennäköisesti ainakin osittain siitä, että tehtävä on selkeämmin rajattu. (Guerreiro ym. 2023, 1502.)

Hallusinointia, kuten muitakin generatiivisen tekoälyn haittapuolia, esiintyy enemmän käännettäessä kielistä, joista ei löydy paljon tekoälyn koulutusmateriaalia (*Low resource languages*, ks. esim. Alam ym. 2024). Suomea puhuu saksaan verrattuna merkittävästi vähemmän ihmisiä (KOTUS 2024; Statista 2024). Erityisesti muihin suomalais-ugrilaisiin kieliin verrattuna kieliteknologiassa hyödynnettävää suomenkielistä materiaalia on kuitenkin olemassa paljon (Hämäläinen ja Alnajjar 2021). Vaikka suomi ja saksa eivät ole ns. vähäresurssisia kieliä, on mahdollista, että omassa tutkimusaineistossani esiintyy hallusinointia. Aiemman tutkimuksen valossa tämä ei kuitenkaan ole erityisen todennäköistä. Mahdollisen hallusinoinnin lisäksi generatiivinen tekoäly myös muotoilee käännöksiä vapaammin ja ns. omin sanoin kertoen verrattuna esimerkiksi neuroverkkokääntimiin (Hendy ym. 2023; Raunak ym. 2023). Tämä ei välttämättä sovellu kaikkiin tekstilajeihin, jos tavoitteena on alkutekstille uskollinen muotoilu (Hendy ym. 2023). Lisäksi Ingon (1990, 81) mukaan vapaa ”parafraastinen käänнос” ei ole yleensä paras käännösratkaisu, vaan

liian pitkä ja selittävä. Ingon mukaan tällaiset käännökset toimivat lähinnä erikoissanastoa sisältävissä teksteissä, joiden sisältöä on tarve selkeyttää (emt.). Kerron tutkielmani analyysiosiossa aineistossa mahdollisesti ilmenevästä hallusinoinnista ja vapaista käännöksistä.

Koska generatiivisen tekoälyn ja laajojen kielimallien toiminta perustuu ihmisen luomiin kehoitteisiin (*prompt*), on myös oikeanlaisen kehoitteen löytäminen keskeistä (Siu 2023b, 7). Mitä yksityiskohtaisempi ja tarkemmin muotoiltu kehote on, sitä todennäköisemmin käännös onnistuu. Kerron oman aineistoni luomisessa käytetystä kehoitteesta tarkemmin luvussa 3. Koska tutkimukseni pääasiallinen tarkoitus ei kuitenkaan ole arvioida kehoitteen onnistuneisuutta tai vertailla eri kehoitteita, en syvenny työssä käyttämäni kehoitteeseen tai kehoitteiden tutkimukseen. Tutkimustulosten arvioinnissa on kuitenkin hyvä huomioida, että kehote voi vaikuttaa merkittävästi laajan kielimallin tuottamaan käännökseen. Kehotteen lisäksi käännökseen vaikuttavat tekoälytyökalun asetukset, kuten lämpötila (*temperature*). Lämpötila-asetuksen kautta voidaan säädellä sitä, kuinka ennakoitava tai vaihtoehtoisesti sattumanvarainen ChatGPT:n tuottama sisältö on. Lämpötila vaihtelee asteikolla 0–2. Mitä korkeampi lämpötila on, sitä sattumanvaraisempi sisältö on. Jos tavoitellaan luovaa sisältöä, voi korkeampi lämpötila tuottaa todennäköisemmin tavoitellun lopputuloksen. (Davis ym. 2024.) Omassa tutkimuksessani käytetty lämpötila on 0,5 tai 1. Käännöksissä käytetyt lämpötila-asetukset on merkitty jälkieditoitujen käännösten yhteyteen (ks. liite 1). En tutkimuksessani arvioi lämpötilan vaikutusta käännökseen. On kuitenkin hyvä tiedostaa, että lämpötila voi vaikuttaa siihen, kuinka ennakoitava tai luova ChatGPT:n tuottama käännös on.

Seuraavassa alaluvussa kerron lisää siitä, miten konekäännettyjä tekstejä voidaan editoida julkaisukelpoiseksi. Lisäksi avaan lyhyesti aiempaa tutkimusta siitä, minkälaista kognitiivista työpanosta konekäännösten jälkieditointi vaatii.

2.3 Jälkieditoiminen ja jälkieditoinnin kognitiivisen työmäärän mittaaminen

Koska tutkimukseeni liittyy oleellisesti tekoälyllä käännettyjen tekstien jälkieditointi, kerron seuraavaksi lähdekirjallisuuden avulla, mitä jälkieditointi on. Lisäksi kerron, miten jälkieditointiin liittyvää työmäärää on arvioitu aiemmissä tutkimuksissa.

Jälkieditoinnilla tarkoitetaan konekäännettyjen tekstien muokkaamista niin, että ne ovat julkaisukelpoisia. Konekäännettyjen tekstien määrän kasvaessa myös tarve jälkieditoinnille kasvaa. Jälkieditointi poikkeaa prosessina tavallisesta ihmisen tekemästä kääntämisestä, sillä jälkieditoija muokkaa koneen tuottamaa tekstiä. Jälkieditoitavan tekstin laatu voi vaihdella paljon. Tämä saattaa tehdä jälkieditointiprosessista asiantuntijalle työlää ja jopa raskaamman tuntuisen verrattuna tavalliseen käännösprosessiin (Guerberof Arenas 2013; Alvarez-Vidal ym. 2020; Moorkens ym. 2018). Siksi jälkieditointiprosesseja ja niiden kannattavuutta on tärkeää tutkia.

Konekäännösten jälkieditointia on tutkittu eri näkökulmista. Jälkieditointia on analysoitu muun muassa siihen kuluvan ajan (ks. esim. Krings, 2001; Koponen ym. 2012), tehtyjen muokkausten (ks. esim. Marzal ja Vidal 1993; van Egdom ja Pluymaekers 2019) ja kognitiivisen työkuorman näkökulmasta (ks. esim. Krings 2001; Temnikova, 2010; Koponen ym. 2012; Lacruz ym. 2014; Lacruz 2017). Keskityn tässä luvussa avaamaan nimenomaan kognitiivisen työmäärän mittaamiseen liittyvää tutkimusta, sillä omat tutkimuskysymykseni liittyvät siihen.

Viittaan kognitiivisella työmäärällä tässä tutkimuksessa mahdolliseen kognitiiviseen kuormitukseen, jota kohdistuu jälkieditoijaan, kun hän muokkaa konekäännettyä tekstiä. Kognitiivinen kuorma liittyy siis siihen, miten kuormittavaa jälkieditointi on aivoille. Tietyt tekijät, esimerkiksi asioiden tekeminen yhtä aikaa, lisäävät kognitiivista kuormitusta (Työturvallisuuskeskus n.d.). Kerron luvussa 3 tarkemmin menetelmistä, joilla kognitiivista työkuormaa on arvioitu tässä tutkimuksessa.

Jälkieditoinnin kognitiivisen työmäärän arvioimiseen on käytetty pääasiassa kahta tapaa. Niin sanotussa TAP-metodissa (*think-aloud protocol*) jälkieditoija kommentoi ääneen tekemiänsä korjausvalintoja ja tutkija arvioi kommenttien pohjalta

jälkiedointiprosessin kognitiivista työmäärää (ks. esim. Krings 2001). CNA-menetelmässä (*choice-network analysis*) taas analysoidaan jälkiedoinnissa tehtyjen korjausten määrää (ks. esim. O'Brien, 2005). Kognitiivista työmäärää arvioidaan sen perusteella, kuinka monta kertaa jälkieditoija on muuttanut samaa sanaa (Temnikova 2010, 3487). Yksi haaste TAP-menetelmässä on, että sitä on vaikea vahvistaa ja toisintaa (Temnikova 2010, 3487). Menetelmää on kritisoitu myös sen hitaudesta ja siitä, että se muuttaa käytössä olevia kognitiivisia prosesseja (Koponen 2016, 11). CNA-metodissa taas kaikilla jälkieditoijilla ei välttämättä ole mahdollisuutta tehdä samoja muutoksia tekstiin. (Temnikova 2010, 3487.)

Irina Temnikova näki olemassa olevat arviointitavat puutteellisina ja ehdotti siksi näiden täydennykseksi uutta tapaa arvioida jälkieditoinnin kognitiivista työmäärää (Temnikova 2010). Temnikova loi virheluokittelun, jossa konekäännöksissä esiintyvät virheet jaotellaan morfologian, syntaksin ja sanaston tasolla sen perusteella, minkälaisia toimenpiteitä niiden korjaaminen vaatii. Virheluokat taas on arvoitettu sen mukaan, minkälainen kognitiivinen työkuorma virheen korjaamiseen liittyy (emt. 3487–3488). Temnikovan luokittelu perustuu psykolingvistiikkaa tutkineen Trevor Harleyn (2008) kognitiivisen lukemisen malliin, Alan Baddeleyn (Baddeley ja Hitch 1974) työmuistimalliin ja tekstissä ilmenevien virheiden havaitsemiseen liittyviin tutkimuksiin (Larigauderie, Gaonac'h ja Lacroix 1998). Temnikova pyrkii erottelemaan virheet mahdollisimman tarkasti niin, että niiden kognitiivisen työmäärän arvioiminen olisi mahdollista. Esimerkiksi sanajärjestystä koskevat virheet luokitellaan kahteen eri kategoriaan sen mukaan, johtavatko ne yksittäisten sanojen vai lausekkeiden siirtämiseen. Sanastovirheet taas luokitellaan eri kategorioihin riippuen siitä, vaihdetaanko sana tyyliin paremmin sopivaan synonyymiin vai täysin toisenlaiseen sanaan. Temnikovan luokittelussa eniten kognitiivista työtä vaativat syntaksiin liittyvät virheet ja vähiten morfologiaan liittyvät virheet. (Temnikova 2010, 3487–3488.) Esittelen Temnikovan virheluokkia tarkemmin luvussa 3.

Temnikova (2010) on luokittelun avulla tutkinut sitä, minkälainen on jälkieditoinnin kognitiivinen työmäärä, kun konekäännetään kontrolloitua kieltä (*controlled language*) verrattuna yleiskieleen. Tieteen termipankin mukaan kontrolloitu kieli on ”rakenteiltaan ja sanastoltaan rajattu, monitulkintaisuutta välttävä kielimuoto, jota

käytetään automaattisesti käännettävien tekstien tuottamiseen” (Tieteen termipankki 2025a). Temnikovan tutkimuksen kontrolloitu kieli perustuu psykolingvistisiin tutkimuksiin ihmisen käsityskyvystä stressitilanteissa (Temnikova 2010, 3485). Kieltä on siis muokattu niin, että se olisi mahdollisimman ymmärrettävää ja selkeää myös tilanteissa, joissa lukijaan kohdistuu stressiä. Omassa tutkimusaineistossani käytetty kieli on hyvin erilaista kuin Temnikovan tutkima kontrolloitu kieli. Siksi voidaan olettaa, että myös tutkimusten tulokset ovat erilaisia eikä niitä voi suoraan verrata toisiinsa. Toisaalta Temnikova on verrannut kontrolloitua kieltä yleiskieleen, joten yleiskieltä koskevien tutkimustulosten vertaaminen oman tutkimukseni tuloksiin voi olla kiinnostavaa. Alaluvun lopussa kerron lisää siitä, mitä Temnikovan virheluokittelun hyödyntämisessä on otettava huomioon tässä tutkimuksessa.

Temnikovan tutkimuksessa (2010) tutkimuskohteena oli kaksi viranomaisten kansalaisille laatimaa kirjallista ohjeistusta hätätilanteiden varalle: ohjeistus ydinonnettomuuden varalle ja ohjeistus terrori-iskujen varalle (esim. kemikaalionnettomuus ja bioterrorismi). Toinen teksteistä muokattiin ennen konekääntämistä kontrolloidun kielen normien mukaisesti (Temnikova 2010, 3486) ja toinen jätettiin alkuperäiseen muotoonsa. Tekstit oli käännetty aiempaa tutkimusta (Temnikova and Orasan 2009) varten tilastopohjaisella konekääntimellä (Google Translate) seitsemälle eurooppalaiselle kielelle. Aiemmassa tutkimuksessa tekstien jälkieditointia tutkittiin määrällisesti jälkieditointiin kuluvan ajan perusteella. Lisäksi Temnikova ja Orasan (2009) tutkivat sitä, kuinka monta merkkiä tekstissä piti muuttaa sitä korjattaessa (*editing distance*). Myöhempään tutkimukseen, jossa käsiteltiin jälkieditoinnin kognitiivista kuormittavuutta, valikoituivat tutkittavaksi bulgarian-, venäjän- ja espanjankieliset käännökset.

Temnikovan mukaan kontrolloidun kielen käyttäminen vähentää konekäännöksen jälkieditoinnin kognitiivista työmäärää (Temnikova 2010, 3490). Kontrolloidusta kielestä ei millään kielellä löytynyt kognitiivisesti kaikista kuormittavimpia virheitä. Vähemmän kuormittavia, esimerkiksi sanojen taivutukseen ja vääriin sanoihin liittyviä, virheitä löytyi taas paikoin enemmän kuin yleiskielisistä teksteistä. Yleiskielisistä teksteistä (poislukien venäjänkieliset tekstit) löytyi erilaisia virheitä eli myös kognitiivisesti kuormittavia syntaksiin ja semantiikkaan liittyviä virheitä. Tuloksissa näkyi kuitenkin myös vaihtelua eri kielten välillä. Koska esimerkiksi

venäjän kielessä on hyvin vapaa sanajärjestys, löytyi venäjän konekäännöksistä Temnikovan tutkimuksessa kognitiivisesti kuormittavia sanajärjestykseen liittyviä virheitä vähemmän (emt. 3489).

Temnikovan johtopäätöstä kontrolloidun kielen vaikutuksesta konekäännetyin tekstin kognitiiviseen kuormittavuuteen voisi kyseenalaistaa siitä, että virheitä on kuitenkin kaiken kaikkiaan hieman enemmän kontrolloidun kielen teksteissä verrattuna yleiskielisiin teksteihin. Jos kuitenkin katsotaan virheiden kognitiivista kuormittavuutta Temnikovan luokittelun avulla, ovat kontrolloidun kielen tekstien virheet vähemmän kognitiivisesti kuormittavia kuin yleiskielisten tekstien virheet.

Sen lisäksi, että Temnikovan aineisto poikkeaa oman tutkimukseni aineistosta, Temnikovan tutkimus ei perustu generatiivisella tekoälyllä tai laajoilla kielimalleilla kääntämiseen. On siis hyvä pohtia, miten Temnikovan luokittelu lopulta soveltuu tekoälyllä käännettyjen tekstien arviointiin. Esimerkiksi mahdollisia hallusinointiin liittyviä virheitä saattaa olla haastavaa arvioida Temnikovan luokittelun avulla, sillä luokittelua ei ole laadittu erityisesti tekoälypohjaisille käännöksille. Marjut Alho (tulossa) on osana Aikakausmedian kansainvälistäminen -tutkimusprojektia tutkinut laajalla kielimallilla käännettyjen A-lehtien artikkelien virheiden kognitiivista kuormittavuutta ja hyödyntänyt tutkimuksessaan nimenomaan Temnikovan luokittelua. Alhon mukaan ChatGPT:n tekemät virheet poikkeavat jonkun verran perinteisten konekääntimien tekemistä virheistä. Esimerkiksi konekääntimille tyypillisiä sanojen poisjättöön ja verbien monitulkintaisuuteen liittyviä virheitä esiintyi Alhon aineistossa suhteellisen vähän. Tutkimuksen mukaan suurin osa tekoälykäännösten virheistä ei ollut kognitiivisesti kovin kuormittavia. Alhon tutkimustuloksia on kiinnostavaa vertailla oman tutkimukseni tulosten kanssa, sillä tutkimusten aineisto on osittain sama ja niissä hyödynnetään samaa virheluokittelua. Tutkimusaineistoja on kuitenkin käsitelty hieman eri tavalla, sillä Alhon tutkimukseen ei liity tekstien jälkieditointia. Tutkimukseni siis täydentää Alhon havaintoja ChatGPT:llä tuotettujen käännösten virheiden kognitiivisesta kuormittavuudesta. Vertailen tuloksia tarkemmin luvussa 4.

Temnikovan (2010) tutkimustuloksia ei siis voi suoraan verrata oman tutkimukseni tuloksiin, sillä tutkimusten aineistot on käännetty kahdella hyvin erilaisella

työkalulla. Tekniikka on myös yleisesti kehittynyt valtavasti Temnikovan tutkimuksen jälkeen. Tutkimukseni yhtenä tavoitteena onkin selvittää, miten Temnikovan luokittelu soveltuu nykyajan tekoälyllä käännettyihin teksteihin. Arvioin Temnikovan luokittelun lopullista soveltuvuutta omaan tutkimukseeni luvussa 5.

Temnikovan luokittelun lisäksi aion tutkimuksessani myös mitata jälkieditointiin kulunutta aikaa. Jälkieditointiin kuluneen ajan arvioiminen antaa lisätietoa jälkieditoinnin kognitiivisesta kuormittavuudesta, vaikka tutkimukseni pääpaino onkin kognitiivisen työmäärän mittaamisessa nimenomaan Temnikovan luokittelun avulla. Muun muassa Koponen ym. (2012) ovat esittäneet, että jälkieditointiin kuluvan ajan mittaaminen on yksi varteenotettava tapa arvioida jälkieditoinnin kognitiivista kuormittavuutta. Menetelmän vahvuutena voi nähdä sen, ettei se ota kantaa yksittäisten virheiden kognitiiviseen kuormittavuuteen, vaan tarjoaa selkeää tietoa siitä, kuinka kauan jälkieditointiin on kulunut aikaa. Myös Koponen ym. hyödynsivät tutkimuksessaan Temnikovan (2010) luokittelua kognitiivisen kuormittavuuden arvioimiseen. He kuitenkin tutkivat jälkieditointiin kulunutta aikaa virkkeiden tasolla, kun taas oma tutkimukseni keskittyy kokonaisten kolumnien jälkieditointiin kuluvaan aikaan. Koposen ym. (2012) havaintojen mukaan vähiten aikaa kului Temnikovan luokittelussa kognitiivisesti vähemmän kuormittavien pienten kuten taivutukseen ja sanavalintoihin liittyvien virheiden korjaamiseen. Eniten aikaa kului kognitiivisesti kuormittavampien täysin väärin käännettyjen sanojen tai kääntämättä jääneiden kohtien korjaamiseen sekä sanajärjestykseen ja idiomaattisiin ilmauksiin liittyvien virheiden muokkaamiseen. Koposen ym. (2012) valitsema menetelmä, jota myös itse hyödynnän tutkimuksessani hieman eri näkökulmasta, antaa lisätietoa siitä, minkälaiset virheet ovat kognitiivisesti kuormittavimpia. Vertailen Koposen ym. (2012) tutkimustuloksia oman tutkimukseni tuloksiin luvussa 4. Seuraavaksi avaan tarkemmin sitä, miten konekäännettyä tekstejä voidaan arvioida muuten kuin kognitiivisen kuormittavuuden näkökulmasta.

2.4 Muita tapoja arvioida konekäännösten laatua

Tutkimukseni päätavoitteena on arvioida valitsemieni tekoälyllä käännettyjen tekstien jälkieditoinnin kognitiivista työmäärää ja sitä kautta niiden laatua. Tässä luvussa kerron erilaisista menetelmistä, joiden avulla konekäännettyjen tekstien

laatua voidaan arvioida. Luvun tarkoituksena on avata laadunarviointiin tarkoitettujen menetelmien kirjoa, haasteita, etuja ja lopulta sitä, miksi päädyin valitsemaan juuri Temnikovan virheluokittelun tutkimukseni menetelmien tueksi.

Laatua voidaan arvioida erilaisilla mittareilla arvioiden esimerkiksi käännöksen sujuvuutta, tarkkuutta tai tarkoituksenmukaisuutta (Koponen 2010, 1).

Arviointimenetelmät taas voidaan jaotella automaattisen arvioinnin menetelmiin ja ihmisen tekemän arvioinnin menetelmiin. Käyn ensin läpi esimerkkejä automaattisen arvioinnin menetelmistä ja niiden vahvuuksista ja heikkouksista. Sen jälkeen kerron tarkemmin ihmisen tekemästä laadunarvioinnista.

2.4.1 Konekäännösten automaattinen arviointi

Yksi tunnetuimmista automaattisista konekäännösten arviointimenetelmistä on BLEU-metodi (Papineni ym. 2002). BLEU (*Bilingual evaluation understudy*) perustuu siihen, että konekäännöstä verrataan ihmisen tekemiin käännöksiin ja sen onnistuneisuus arvioidaan sen perusteella, kuinka yhteneväinen se on ns. verrokkitekstien kanssa. Yhteneväisyys lasketaan matemaattisesti käännöksissä esiintyviä sanoja vertailemalla ja käännös pisteytetään. Oletuksena on, että mitä korkeampi pistemäärä on, sitä suurempi yhteneväisyys tekstien välillä on ja sitä paremmin käännös on onnistunut. (Papineni ym. 2002, 311–312.) BLEU-metodia on kritisoitu siitä, ettei testin korkea tulos aina korreloi käännöksen laadun kanssa (Callison-Burch ym. 2006). BLEU-testissä usea eri versio samasta käännöksestä saattaa nimittäin saada saman korkean arvosanan, mutta ihminen arvioi käännökset erilaatuisiksi. BLEU siis sallii käännökseen paljon enemmän variaatiota kuin normaalisti olisi suotavaa ja saattaa arvioida erikoisetkin sanayhdistelmät onnistuneiksi. (emt., 251.) Vaikka BLEU on pitkään ollut suosittu konekäännösten arvioinnin väline, on sen rinnalle tullut tekoälykehityksen myötä muita, kehittyneempiä arviointitapoja (ks. esim. Kocmi ym. 2024).

Konekäännösten arvioinnin menetelmiä on BLEU lisäksi lukuisia muita. Yksi esimerkki automaattisen arvioinnin menetelmästä, joka muistuttaa BLEU-metodia, on HTER-metodi (Snover ym. 2006). HTER-metodissa (*human-targeted translation error rate*) lasketaan niiden muokkausten määrää, joita konekäännös

vaatii vastatakseen ihmisen tekemää käännöstä. Muita samankaltaisia automaattisia käännösten arviointimetoodeja ovat muun muassa nk. NIST-metodi (Doddington, 2002) ja METEOR (Banerjee ja Lavie 2005). Tekoäly on kuitenkin mullistanut konekääntämisen lisäksi myös konekäännösten arvioinnin. Freitag ym. (2024) ovat tutkineet konekäännösten arvioinnin mittareiden soveltuvuutta nimenomaan laajoilla kielimalleilla käännettyihin teksteihin. Heidän mukaansa syväoppimismalleja hyödyntävät arviointitavat soveltuvat näiden tekstien laadunarviointiin paremmin kuin käännösten päällekkäisyyttä arvioivat menetelmät, kuten BLEU (Freitag ym. 2024, 64).

Konekäännösten onnistuneisuutta voi arvioida myös automaattisilla menetelmillä, joihin ei liity ihmisen tekemää verrokkikäännöstä. Yksi tällainen tutkimusmetodi on nk. CE-menetelmä (*confidence estimate*), jossa arvioidaan matemaattisesti todennäköisyyttä sille, että käännös onnistuu (ks. esim. Blatz ym. 2004 ja Specia ym. 2009). CE-menetelmässä käännöksiä ei siis lähtökohtaisesti verrata ihmisen tekemään verrokkikäännökseen.

Automaattisen arvioinnin etu on sen vaivattomuus, nopeus ja kustannustehokkuus verrattuna ihmisen tekemään arviointiin (Papineni ym. 2002, 311; Snover ym. 2006, 223). Koska mahdollisia käännösratkaisuja on lukuisia, ei automaattisen arvioinnin tuloksia kuitenkaan voi pitää täysin tarkkoina (Freitag ym. 2021, 1460; Vilar ym. 2006, 697). Automaattisen arvioinnin menetelmät eivät myöskään sovellu oman tutkimukseni aineistoon ja resursseihini parhaalla tavalla. Avaan tutkimusmenetelmiäni ja menetelmien valintaa tarkemmin luvussa 3. Seuraavassa alaluvussa esittelen lyhyesti ihmisen tekemän arvioinnin menetelmiä ja avaan erityisesti omassa tutkimuksessani hyödyntämäni arviointitapaa, virheluokittelua.

2.4.2 Ihmisen tekemä arviointi: virheluokittelu

Ihmisen tekemään konekäännösten arviointiin on hyödynnetty erilaisia menetelmiä. Osaan menetelmistä liittyy jälkieditointia, kuten omassa tutkimuksessani, kun taas toisissa käännöstä arvioidaan sellaisenaan esimerkiksi arviointiasteikkoa hyödyntäen (ks. esim. Bentivogli ym. 2018). Käännöstä saatetaan siis ensin muokata ja sitten

arvioida esimerkiksi muokkausten määrää tai luonnetta. Toisaalta tekstiä voidaan arvioida suoraan esimerkiksi ihmisen tekemään verrokkikäännökseen verraten.

Keskityn tässä alaluvussa virheanalyysiin, sillä hyödynnän tätä arviointimenetelmää omassa tutkimuksessani. Esittelin omassa tutkimuksessani hyödynnettävää Temnikovan virheluokittelua luvussa 2.3, joten avaan seuraavaksi toista luokittelua, jota harkitsin tutkimukseeni. Kerron alaluvussa lyhyesti virheanalyysin vahvuuksista ja heikkouksista, jonka jälkeen luvussa 3 kerron tarkemmin tutkimukseni menetelmistä ja siitä, miksi ne soveltuvat mielestäni parhaiten omaan aineistooni. Erilaisia laadunarviointiin tarkoitettuja virheluokitteluja on myös muita (ks. esim. The MQM Council 2025; Vilar ym. 2006), mutta harkitsin omaan tutkimukseeni pääasiassa Koposen (2010) virheluokittelua. Esittelen seuraavaksi lyhyesti Koposen virheluokittelun ja tärkeimmät tutkimustulokset.

Suomessa konekääntämistä tutkinut Maarit Koponen on laatinut pilottitutkimuksessaan (2010) konekäännöksissä ilmenevien virheiden analysoimista varten luokittelun, jossa käsitteisiin (*concept*) ja viittaussuhteisiin (*relation*) liittyvät virheet jaotellaan eri kategorioihin. Virheiden luokittelu keskittyy kielen semanttiseen tasoon eli luokittelu ei ota erikseen kantaa esimerkiksi kielioppivirheisiin. Koponen konekänsi tutkimusta varten pätkän kolmen eri tekstilajin tekstistä englannista suomeksi: National Geographicin artikkelista, tietokoneen virustorjuntaohjelmiston käyttöohjeesta ja Euroopan komission vihreästä kirjasta. Jokainen teksti käännettiin kahdella eri konekääntimellä: sääntöpohjaisella ja tilastopohjaisella kääntimellä. Lisäksi konekäännetyjä tekstejä verrattiin olemassa oleviin ihmisen tekemiin käännöksiin. (Koponen 2010, 2–6.) Tutkimus valottaa sitä, miten sääntö- ja tilastopohjainen konekäännin soveltuvat eri tekstilajien kääntämiseen. Samalla tutkimus kertoo virheiden määrästä ja sitä kautta jälkieditoinnin työmäärästä.

Koposen löysi sääntö- ja tilastopohjaisen konekääntimen väliltä eroja.

Tilastopohjainen käännin teki kokonaisuudessaan enemmän virheitä ja käännosten laadussa oli enemmän vaihtelua kuin sääntöpohjaisen konekääntimen käännoksissä. Tilastopohjainen konekäännin teki selvästi eniten sekä käsitevirheitä että viittaussuhteisiin liittyviä virheitä lehtiartikkelin käännoksessä. Sääntöpohjaisen

konekääntimen käsitevirheet taas jakautuivat tekstilajien välillä tasaisemmin kuin tilastopohjaisen kääntimen. Viittaussuhteisiin liittyviä virheitä löytyi kuitenkin eniten lehtiartikkelin käännöksestä. (Koponen 2010, 7–9.) Koposen tutkimustulokset tukevatkin hypoteesiani, jonka mukaan konekääntimet tekevät asiateksteihin verrattuna enemmän virheitä lehtiartikkeleissa, joiden yksi tarkoitus on viihdyttää lukijaa. Toki on huomioitava, että Koponen tutki nimenomaan sääntö- ja tilastopohjaisia kääntimiä eikä generatiivista tekoälyä tai laajoja kielimalleja.

Koposen tutkimuksessa myös ihmisen tekemistä käännöksistä löytyi virheitä, mutta ne poikkeavat konekääntimien tekemistä virheistä. Konekääntimien tyypillisimmät virheet liittyvät vääriin käännöksiin (*mistranslation*) kun taas ihmisen tekemät virheet lisäyksiin ja poistoihin. (Koponen 2010, 7–9.) Koponen korostaa, että tulosten perusteella ei voi tehdä johtopäätöstä siitä, mikä käännöksistä onnistui parhaiten välittämään lähtötekstin semanttiset merkitykset, sillä virheluokittelussa ei jaotella virheitä niiden vakavuuden perusteella (emt., 9). Vertaan Koposen tutkimustuloksia omiin tuloksiini luvussa 4.

Koposen luokittelu tarjoaa kiinnostavaa tietoa tilasto- ja sääntöpohjaisten konekäänninten käännöksissä esiintyvistä virheistä. Luokittelun ulkopuolelle jää kuitenkin virheitä, jotka liittyvät muihin kuin kielen semanttisiin piirteisiin. En siksi hyödynnä lopullisessa tutkimuksessani Koposen virheluokittelua, vaan analysoin konekäännöksiä luvussa 2.3 esittelemäni Temnikovan virheluokittelun avulla (2010). Valitsemani luokittelun tarkoitus on pääasiassa arvioida jälkieditoinnin kognitiivista työmäärää. Samalla se kuitenkin kertoo myös virheiden määrästä ja luonteesta. Siksi sovellan sitä tutkimuksessani myös käännösten laadun arvioimiseen.

Freitag ym. (2021, 1460) nostavat esiin ihmisen tekemän konekäännösten arvioinnin kompastuskiviä. Arviointia ei voi pitää aukottomana ja täydellisenä, sillä se perustuu arvioijan subjektiiviseen arvioon käännöksen onnistuneisuudesta. Konekäännösten arvioinnissa on huomioitava, ettei tuloksia kannata tiivistää yhteen lukuun tai luokitukseen, vaan tuloksia on tulkittava kokonaisuutena. (emt.) Aiemmissä tutkimuksissa on myös havaittu, että ihmisen tekemän arvioinnin tulokset saattavat poiketa toisistaan. Tulokset eivät siis tältä osin ole täysin luotettavia. (Graham ym. 2013, 33.) Toki on myös huomioitava, että mahdollisia käännösratkaisuja on aina

useita ja subjektiivisuudesta ei pääse täysin eroon, kun kyse on ihmisistä. Ihmisen tekemä arviointi voi kuitenkin paljastaa käännosten laadusta asioita, joita ei automaattisella arvioinnilla olisi mahdollista selvittää.

Freitag ym. (2021, 1460) mukaan virheluokittelu on yksi parhaista tavoista arvioida konekäännösten onnistuneisuutta. Virheluokittelun avulla käännöksen laatua on mahdollista arvioida perusteellisesti, kun virheet luokitellaan ja arvotetaan yksitellen (emt.). Myös omaan tutkimukseeni virheanalyysi ja valitsemani virheluokittelu tarjoaa rungon, jonka avulla käännösten laatua on mahdollista arvioida mahdollisimman tarkasti ja luotettavasti.

3 Tutkimusmenetelmät ja aineisto

Tässä luvussa esittelen tutkimukseni lähestymistavan, menetelmät ja aineiston. Kerron myös, miten olen päätenyt valitsemaani aineistoon ja menetelmään ja miten aineistoa on kerätty ja analysoitu.

3.1 Tutkimusmenetelmät

3.1.1 Lähestymistapa

Tutkimukseni lähestymistapa on tapaustutkimus. Tapaustutkimus-käsitteen tarkka määrittelyminen on haastavaa muun muassa siksi, että lähestymistapaa hyödynnetään eri tieteenaloilla. Lisäksi tapaustutkimuksissa käytettävät menetelmät vaihtelevat. (Eriksson ja Koistinen 2014, 4; Bamberg, Jokinen ja Laine 2007, 9.)

Yhteistä tapaustutkimuksille on se, että tutkimuskohteena on yksi tai muutama rajattu tapaus, joiden perusteelliseen kuvaamiseen tutkimus keskittyy (Bamberg, Jokinen ja Laine 2007, 9–10). Tapaustutkimuksen tavoitteena on siis ymmärtää paremmin tutkittavaa tapausta ja sen kontekstia (emt. 10). Tapaus voi olla esimerkiksi jokin henkilö, ilmiö tai organisaatio. Aineisto on pääasiassa laadullista, mutta se voi olla myös määrällistä. (Eriksson ja Koistinen 2014, 4–5.)

Nojaan tutkimuksessani Päivi Erikssonin ja Katri Koistisen määritelmään, jonka mukaan tapaustutkimuksessa “tarkastellaan yhtä tai useampaa tapausta, joiden määrittely, analysointi ja ratkaisu on tapaustutkimuksen keskeisin tavoite” (Eriksson ja Koistinen 2014, 4). Olen määritellyt oman tutkimukseni tapaukseksi yksittäisen ilmiön eli konekääntämisen tekoälyllä, tarkemmin sanottuna ChatGPT:llä. Tutkimuskohteeni taas ovat aikakauslehtiä kustantavan A-lehden Oy:n meilläkotona.fi-sivuston kolumnit, jotka on käännetty ChatGPT:llä suomeksi saksaksi. Tutkimuskohde on se osa tapauksesta, josta olen erityisen kiinnostunut ja jonka analysoimiseen haluan tutkimuksessani paneutua (Bamberg, Jokinen ja Laine 2007, 10).

Koska kyseessä on tapaustutkimus, eivät sen tulokset ole sellaisinaan yleistettävissä kaikkeen tekoälyavusteiseen kääntämiseen (Eskola, Saarela ja Vilkkä 2018, 191).

Tapaustutkimusta onkin kritisoitu lähestymistapana muun muassa edustavuuden puutteesta (Eriksson ja Koistinen 2014, 4). Tutkimustulokset kuitenkin tarjoavat esimerkin siitä, miten tekoäly onnistuu kolumnien kääntämisessä suomesta saksaan, minkälaisia virheitä käännoksissä voi ilmetä ja mikä on jälkieditointiprosessin kognitiivinen työmäärä. On siis mahdollista ja toivottavaa, että tutkimukseni tuloksista on hyötyä laajemmallekin yleisölle. Erikssonin ja Koistisen (2014, 4) mukaan tapaustutkimus voi tuottaa ”rikasta, monipuolista ja yhteen kietoutuvia rakenne- ja toimijasuhteita analysoivaa tietoa”. Kerron seuraavaksi tarkemmin valitsemistani tutkimusmenetelmistä.

3.1.2 Valitut tutkimusmenetelmät

Tutkielmani menetelmien keskiössä on tekstianalyysi, tarkemmin sanottuna virheanalyysi, jonka olen tehnyt ChatGPT:llä suomesta saksaan käännettyihin kolumneihin. Kolumnit on ensin käännetty tekoälyllä, jonka jälkeen ne on jälkieditoitu. Jälkieditointiin on osallistunut lisäksi kaksi muuta saksan kielen asiantuntijaa. Asiantuntijoilla tarkoitetaan tässä tutkimuksessa saksan kielen ja kääntämisen maisteriopiskelijaa sekä saksan kielen ja kääntämisen yliopistonlehtoria, jonka äidinkieli on saksa.

Ottamalla jälkieditointiprosessiin mukaan muita kielen asiantuntijoita olen pyrkinyt varmistamaan mahdollisimman objektiivisen ja perusteellisen lopputuloksen. Luvussa 3.2 kerron tarkemmin työskentelyprosessista ja kahden muun asiantuntijan panoksesta tutkimukseeni. Jälkieditoinnin jälkeen olen tehnyt virheiden lopullisen jaottelun virheluokkiin itsenäisesti. Virheiden tyypittely tarjoaa mahdollisuuden aineiston konkreettiseen ja mahdollisimman objektiiviseen kuvailuun.

Käytän tutkimuksessani sekä laadullista että määrällistä menetelmää. Pääasiallinen menetelmäni on laadullinen, sillä jaottelen käännoksissä ilmenevät virheet virhekategoriioihin ja analysoin käännosten laatua ja jälkieditoinnin kognitiivista työmäärää. Laadullisen analyysin lisäksi hyödynnän määrällistä menetelmää vertailllessani kategoriioihin jaettujen virheiden määrää toisiinsa. Vertailun tarkoituksena on selventää sitä, minkälaisia virheitä teksteissä esiintyy eniten. Lisäksi olen mitannut jälkieditointiin kuluvaan aikaa. Ajan mittaaminen auttaa

hahmottamaan jälkieditoinnin vaatimaa työpanosta ja kognitiivista kuormitusta. Kerron analyysin vaiheista tarkemmin luvussa 3.2.

Valitsemieni menetelmieni tarkoituksena on valottaa sitä, minkälaista työpanosta tekoälyllä käännettyjen kolumnien jälkieditointi vaatii, ja lopulta pohtia, onko tekoälyn hyödyntäminen tämän tekstilajin kääntämisessä kannattavaa.

3.1.3 Virheanalyysi

Harkitsin tutkimukseeni erilaisia tapoja arvioida tekoälyllä tehtyjä käännöksiä. Päädyin kuitenkin ensisijaisesti laadulliseen menetelmään, sillä ainoastaan määrällisen tutkimusmenetelmän käyttäminen ei olisi ollut mielekästä tutkimukseni resurssien, aineiston laajuuden ja tutkimuskysymykseni huomioon ottaen.

Määrällisten automaattisen arvioinnin menetelmien tavoitteena on usein valottaa suurempien tekstimassojen luonnetta. Lisäksi automaattisen arvioinnin tulokset eivät aina kerro luotettavasti käännösten laadusta (ks. luku 2.4.1). Määrällistä menetelmää tutkimuksessani edustaa kuitenkin käännöksissä ilmenneiden virheiden vertaileminen ja jälkieditointiin kuluvan ajan seuraaminen.

Päädyin siis valitsemaan tutkimusmenetelmäni ihmisen tekemien arviointitapojen joukosta. Valitsin lopulta virheanalyysiin kappaleessa 2.3 esittelemäni Irina Temnikovan virheluokittelun (2010, 3487–3488):

Taulukko 1. Temnikovan virheluokittelu (2010)

Error	Correction
(1.) Missing word	Error correction requires adding the missing word
(2.1.) Word order error	Error correction requires moving single words
(2.2.) Word order error	Error correction requires moving whole phrases

(3.1.) An incorrect word	Error correction requires replacing with a completely different lexical item
(3.2.) Correct word with an incorrect ending (e.g. number or case)	Error correction requires replacing with a different ending
(3.3.) An incorrect word	Error correction requires a different style synonym
(3.4.) Extra word	Error correction requires deleting the extra word
(3.5.) Error due to incorrectly recognised idiomatic expressions	Error correction requires replacing with the correct translation of the idiomatic expression
(4.1.) Missing punctuation sign	Error correction requires adding the missing punctuation sign(s)
(4.2.) Incorrect punctuation sign	Error correction requires replacing with the correct punctuation sign(s)

Päädyin Temnikovan luokitteluun, sillä sen avulla voin tutkia sekä virheiden tyyppiä että niiden jälkieditöinnin kognitiivista kuormittavuutta. Alla Temnikovan luokittelu kognitiivisen työkuorman määrittämiseksi:

Taulukko 2. Temnikovan virheluokittelun (2010) virheet luokiteltuna kielen eri tasoille ja järjestettynä kognitiivisen työmäärän mukaan

Listan ensimmäisen virheen kognitiivinen työmäärä on pienin ja viimeisen suurin.

Morphol. level	1. Correct word, incorrect form (CInF)
Lexical level	2. Incorrect style synonym (Styl)
	3. Incorrect word (InW)
	4. Extra word (ExW)
	5. Missing word (MissW)
	6. Idiomatic expression (Idiom)
Syntactic level	7. Wrong punctuation (InP)
	8. Missing punctuation (MissP)
	9. Word order at word level (WoW)
	10. Word order at phrase level (WoPh)

Koska Temnikova on hyödyntänyt virheluokittelua tutkimuksessaan kolmeen hyvin erilaiseen kieleen, luokittelun pitäisi soveltua hyvin myös suomi–saksa käännöksiin. On kuitenkin hyvä huomioida, ettei luokittelu välttämättä ole täysin tyhjentävä tai sovellu yhtä hyvin kaikkiin tekstilajeihin. Temnikova kehottaa soveltamaan virheluokittelua kunkin kielen ominaispiirteiden mukaisesti (2010, 3489). Saksa-suomi-kieliparin käännöksissä tulee esimerkiksi ottaa huomioon syntaksiin liittyvät eroavaisuudet. Kommentoin aineistossa ilmenneitä käännöshaasteita saksan ja suomen välillä luvussa 4. Uskon, että Temnikovan luokittelu tarjoaa mahdollisuuden työmäärän kattavaan ja yksityiskohtaiseen arviointiin, kun sitä sovelletaan tutkittavan kielen ja tekstilajin ominaispiirteiden mukaan. Lisäksi voin virheluokittelun avulla arvioida yleisemmin käännösten laatua.

Yksi arviointimenetelmä, jota harkitsin tutkimukseeni ja ensin sovelsin aineistoni analysoimiseen, on kappaleessa 2.3 mainitsemani Maarit Koposen virheluokittelu (2010). Tämänkaltainen joustavuus analyttisen kehyksen määrittelyssä on tyypillistä tapaustutkimukselle (Bamberg, Jokinen ja Laine 2007, 20). Koposen luokittelu paljasti mielenkiintoisia huomioita käännöksistä, mutta sen soveltaminen juuri kolumnien käännöksiin ei tuntunut täysin mielekkäältä. Tämä johtuu muun muassa siitä, että Koposen virheluokittelu keskittyy semantiikkaan, kun taas Temnikovan luokittelu tarjosi mahdollisuuden analysoida kolumneissa esiintyneitä käännösvirheitä kielen eri tasoilla.

Seuraavaksi kerron tarkemmin analyysin vaiheista.

3.2 Analyysin vaiheet

Tutkimukseni alkoi syksyllä 2024 osana Helsingin yliopiston Aikakausmedian kansainvälistäminen -projektiin liittyvää kurssityötä. Projektissa käännettiin aikakausmedian artikkeleita tekoälyn avulla suomesta saksaan ja englantiin (Helsingin yliopisto 2024). Jälkieditoin kurssilla toisen saksan kielen ja kääntämisen opiskelijan kanssa yhteensä viisi suomesta saksaan konekäännettyä A-lehtien meilläkotona.fi-sivuston artikkeleita. Kaksi teksteistä oli kolumneja, joita päädyin hyödyntämään tutkimuksessani.

Kurssilla jälkieditointityötämme arvioi ja kommentoi saksan kielen ja kääntämisen lehtori, joka puhuu saksaa äidinkielenään. Täydentääkseni aineistoani lisäsin siihen kolme kolumnia, joita olin analysoinut samaan projektiin liittyvällä aiemmalla kurssilla keväällä 2024. Tuolloin teksteihin tehtiin kuitenkin vain virheanalyysi ilman varsinaista jälkieditointia.

Tutkimuksessani analysoidut käännösversiot ovat uusia eli en päätenyt tutkielmassani analysoimaan versioita, joihin olin jo aiemmalla kurssilla tehnyt virheanalyysin. Käännätin tekstit tekoälyllä uudestaan, jotta jälkieditoitavat käännökset olisi käännetty ajallisesti mahdollisimman lähellä toisiaan. Tällä pyrin varmistamaan niiden vertailukelpoisuuden. Kaikki tutkimuksessani käytetyt tekstit on siis käännetty tekoälyllä loka- ja marraskuussa 2024.

Kolumnit oli ladattu valmiiksi A-lehtien ChatGPT 4.0 -pohjaiseen käännöstyökaluun. Työkalusta löytyi myös valmis kaikissa käännöksissä käytettävä julkaisukontekstia kuvaileva kehote, jonka avulla käänsin tutkimukseni tekstit (ks. liite 2). Kehote on muotoiltu englanniksi, ja se on alun perin tarkoitettu englanninkielisen käännöksen tuottamiseen. Jotta kehote sopisi saksankielisen käännöksen tuottamiseen, olen muokannut sanan ”English” tilalle sanan ”German”. Jälkieditointi ja virheluokittelu tehtiin Microsoft Wordissa. Käännettävän materiaalin käsittely ei siis vaatinut esimerkiksi aineiston kopiointia erilliseen tekoälytyökaluun.

Kurssilla jälkieditoituja kahta kolumnia oli kommentoimassa yliopistonlehtorin lisäksi saksan kielen ja kääntämisen opiskelija, joka teki omat jälkieditoituidut versionsa teksteistä. Vertailimme tekstejämme toisiinsa ja saimme lopulliseen yhteiseen palautukseemme palautetta yliopistonlehtorilta. Kolmen muun tekstin jälkieditoinnissa apunani oli ainoastaan yliopistonlehtori, joka kommentoi ja muokkasi jälkieditoimiani kolumneja. Lehtori huomioi muokkauksissaan vain suomenkielisen alkuperäistekstin ja jälkieditoitun käännöksen. Hän ei siis verrannut jälkieditoitua versiota alkuperäiseen ChatGPT:n käännökseen.

Jälkieditoinnin jälkeen keskustelimme jälkieditointiprosessista ja kävimme läpi teksteissä esiintyneitä virheitä. Keskustelut avaavat tarkemmin jälkieditointiratkaisujen taustoja ja sitä, miltä jälkieditoinnin työmäärä käytännössä tuntui. Kerron tärkeimmistä havainnoistamme luvussa 4. Yksi teksteistä (Sanna Stellanin kolumni *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa*) on

osana kurssitehtävää ennen ihmisten jälkieditointia jälkieditoitu tekoälytyökalulla. Kuten itse käännökset, myös tekoälyn jälkieditointi tehtiin A-lehtien itse muotoilemaa kehotetta käyttäen, joka löytyi valmiiksi tekoälytyökalusta (ks. liite 3). Kerron analyysissä, miten ylimääräinen työvaihe vaikutti tämän kolumnin kohdalla lopullisen käännökseen.

Jälkieditointiin ei annettu yksityiskohtaisia ohjeita. Kurssilla tavoitteena oli sujuva, julkaisukelpoinen teksti. Saimme A-lehdiltä lisätietoa brändistä sekä ohjeet muun muassa tekstien sävyn ja tyylin muokkaamiseen. Tavoiteltu sävy käy ilmi myös ChatGPT:lle annetusta kehotteesta (liite 2). Brändin keskiössä on muun muassa inspiroivuus, kannustaminen, ilo, kestävä elämäntapa ja lämminhenkisyys. Lisäksi saimme taustatietoa tekstien kohderyhmistä. Näiden tietojen perusteella pyrimme arvioimaan, milloin teksti on julkaisukelpoinen. Seurasimme prosessin aikana myös jälkieditointiin kuluvaan aikaan, johon sisältyi muun muassa tiedonhaku. Tähän ei kuitenkaan käytetty erillistä ajanseurantaan tarkoitettua ohjelmaa, joten kyse on aika-arvioista. Kommentoin jälkieditointiin kulunutta aikaa tarkemmin luvussa 4.

Kahden muun asiantuntijan kommenttien ja yhteisten keskustelujen pohjalta tein viimeiset muutokset jälkieditoituihin teksteihin. Tämän jälkeen luokittelin virheet Temnikovan (2010) luokittelun mukaisesti. Virheitä olisi voinut erotella ja luokitella eri tavoin. Päädyin erottelemaan virheet siten, että laskin kunkin lähtötekstin puutteen tai virheen yhdeksi virheeksi, riippumatta siitä johtiko se yhteen vai useampaan muutokseen. Joissain tapauksissa yksi virke saattaa sisältää useita virheitä. Se saattaa olla esimerkiksi kokonaisuudessaan epäidiomaattinen ja siitä voi samalla puuttua jokin lähtötekstin merkitys. Tällaisissa tapauksissa olen erottanut virheet omikseen. Jos virke on kokonaisuudessaan ollut epäidiomaattinen, mutta siitä on vaikea erotella yksittäisiä virheitä, on se luokiteltu yhdeksi virheeksi. Myös kokonaiset fraasit, jotka kuuluvat lähtötekstiin mutta puuttuivat käännöksestä eikä niiden poisjätö ole perusteltua, on luokiteltu yhdeksi virheeksi. Verbin tai substantiivin taivutukseen liittyviä korjauksia ei ole luokiteltu omiksi virheikseen, jos taivutusta on pitänyt muuttaa siksi, että virheellinen sana on vaihdettu tyyliin paremmin sopivaan synonyymiin ja tämä on vaikuttanut kongruenssiin. Kuvaan edellä mainittuja tapauksia tarkemmin analyysissä konkreettisten esimerkkien avulla.

Osa virheistä sopisi useampaan kategoriaan. Olen analyysissäni luokitellut kunkin virheen yhteen luokkaan sen mukaan, mikä luokka kuvaa niitä parhaiten ja pyrkinyt luokittelemaan kaikki samantyylliset virheet johdonmukaisesti. Kuvatessani kunkin kolumnin analyysiä ja havaintojani kerron tarkemmin virheiden luokittelusta ja niistä virheistä, joiden luokittelu aiheutti eniten haasteita.

Virheiden erotteluun liittyvien haasteiden takia virheiden määrän tarkka analysoiminen ei ole kovin mielekäästä. Myös Temnikova (2010, 3489) pohtii virheiden määriä vertaillen sitä, kuinka muun muassa kielten väliset erot ja jälkieditoijille annetut ohjeet vaikuttavat virheiden määriin ja luonteeseen. Määrät kuitenkin antavat osviittaa siitä, kuinka paljon eri virhetyyppejä kolumneissa esiintyi. Toisin kuin Temnikova keskityn tutkimuksessani vain yhden kielen analysoimiseen. Koska kolumnien virheet on siis eroteltu ja luokiteltu samalla tavalla, voidaan virheiden määriä eri kolumneissa verrata keskenään ja siten saada tietoa siitä, mitkä virheet ovat kaikista yleisimpiä kyseisissä kolumneissa. Jos kieliä olisi useampia, kuten Temnikovalla, olisi virheiden määrien vertailu haastavampaa.

Temnikovan tutkimuksessa virheiden luonne ja määrät siis vaihtelivat kielen mukaan. Sekä venäjän, bulgarian että espanjan käännöksissä virheet liittyivät kuitenkin pääasiassa taivutukseen, tyyliin ja väärin sanavalintoihin (Temnikova 2010, 3489). Omassa tutkimuksessani virheet painottuvat oletettavasti enemmän idiomaattisuuteen. Temnikova ei ole artikkelissaan esitellyt esimerkkejä eri kategorioiden virheistä. On siis mahdollista, että Temnikova on sisällyttänyt *idiomatic expression* -luokkaan ainoastaan idiomit. Olen kuitenkin valinnut omassa tutkimuksessani sisällyttää luokkaan sekä väärin käännetty idiomit että muut idiomaattiset fraasit ja ilmaukset. Vertaan luvussa 4 Temnikovan tuloksia oman tutkimukseni tuloksiin.

3.3 Aineisto

Tutkimukseni aineisto koostuu A-lehtien meilläkotona.fi-sivuston vapaasti luettavista artikkeleista, jotka on konekäännetty suomesta saksaan yrityksen omalla ChatGPT 4.0 -käännöstyökalulla. Meilläkotona.fi on suomalaisen vuonna 1933 perustetun A-lehden Oy:n asumisen medioiden sivusto. Sivusto sisältää erilaisia sisustukseen,

remontointiin, puutarhaan, ruokaan liittyviä tekstejä ja tee se itse -ohjeita. Sivustolta löytyy siis erilaisia tekstilajeja.

Olen valinnut keskittyä tutkimuksessani yhteen sivustolta löytyvään tekstilajiin; kolumneihin. Valitsin nimenomaan tämän tekstilajin siksi, että kolumnit vaativat usein tarkkaa lukutaitoa (ks. luku 2.1). Siksi oletan, että tekoälyn voi olla vaikeaa hahmottaa kolumnien kieltä ja niiden välittämiä implisiittisiä merkityksiä. Analysoin yhteensä viittä kolumnia, joiden pituus vaihtelee noin 200 sanasta 500 sanaan:

1. *”Keitetäänpä kahvit” on Antti Holman mielestä suomen kielen täysin lause – kolumnisti saa suodatinkahvilla yhteyden synnyinmaahansa*
(Antti Holma 4.8.2021) 467 sanaa, käänös 710 sanaa
2. *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa*
(Sanna Stellan, 14.6.2023) 474 sanaa, käänös 682 sanaa
3. *Miltä meillä haisee? Lapseni kaverin kommentti pysäytti*
(Anu Karttunen 13.9.2023) 239 sanaa, käänös 369 sanaa
4. *Kun töissä iski ruuhka, äiti tuli apuun – ja näytti heti, miksi hän on korvaamaton* (Sanna Stellan 13.12.2023) 468 sanaa, käänös 670 sanaa
5. *Osallistuin viiden vaatteen vuosi -haasteeseen, ja näin siinä kävi*
(Tytti Kontula 12.1.2024) 262 sanaa, käänös 377 sanaa

Kolumneista kolme on hieman alle 500 sanan pituisia ja kaksi hieman yli 200 sanan pituisia. Kolumnien sanamäärässä on huomioitu otsikko ja ingressi. Suomenkielisen materiaalin laajuus on siis yhteensä 1 910 sanaa (käännökset 2 808). Vertaan myöhemmin tutkielmassani editoimattomien käännösten sanamäärää lopullisten, jälkieditoitujen tekstien sanamäärään.

Antti Holman kolumni (4.8.2021) kertoo suomalaisesta kahvikulttuurista ulkosuomalaisen silmin. Sanna Stellanin kolumni (14.6.2023) keskittyy suomalaisen juhlakulttuurin kuvailemiseen. Anu Karttunen taas käsittelee kolumnissaan (13.9.2023) kotien ominaishajua ja erilaisiin tuoksuihin liittyviä muistoja. Sanna Stellan kertoo toisessa kolumnissaan (13.12.2023) äidistään ja siitä, kuinka on saanut tältä apua haastavassa tilanteessa. Tytti Kontulan kolumni (12.1.2024) taas kertoo kirjoittajan kokemuksista Viiden vaatteen vuosi -somehaasteesta.

Kolumnit 3–5 on käännetty 8.11.2024 peräkkäin samassa istunnossa. Kaksi ensimmäistä kolumnia on käännetty erillisissä istunnoissa noin kuukautta aiemmin lokakuussa 2024. Tarkemmat ajankohdat löytyvät liitteestä 1. On mahdollista, että istunnon aikana annetut tiedot ovat vaikuttaneet samassa istunnossa seuraavaksi käännettyihin teksteihin. Vaikutusta on kuitenkin hyvin vaikeaa arvioida aineistostani eikä se ole tutkimukseni tarkoitus, joten en keskity tutkielmassani tämän analysoimiseen.

Analysoitavat kolumnit on valittu niiden tekstien joukosta, joita A-lehdet ovat ladanneet käännöstyökaluun. Aineiston valinnalle ei ole muita kriteereitä kuin tekstilaji. Niitä ei siis ole valikoitu esimerkiksi kirjoittajan tai julkaisuajankohdan perusteella. Näillä ominaisuuksilla ei ole tutkielmani kannalta väliä, sillä keskityn konekäännösten laadun ja jälkieditoinnin työmäärän arvioimiseen tietyn tekstilajin sisällä.

Valitsin tutkimukseeni viisi tekstiä, sillä tämä määrä tarjosi minulle mahdollisuuden tutkimukseni puitteissa analysoida tapausta mahdollisimman kattavasti mutta kuitenkin myös tarpeeksi syvällisesti. Jos analysoisin tutkielmassani suurempaa tekstimäärää, saattaisi analyysi jäädä liian pintapuoliseksi. Toisaalta esimerkiksi vain yhden käännöksen analysoiminen ei olisi antanut minulle mahdollisuutta paneutua tapaukseen tarpeeksi kattavasti.

4 Analyysi

Kerron seuraavaksi tutkimustuloksistani. Avaan ensin tutkimustuloksia yleisellä tasolla ja kerron myös analyysiin vaikuttaneista tekijöistä ja mahdollisista haasteista. Tämän jälkeen esittelen alaluvuissa 4.1–4.3 esimerkkejä kolmesta suurimmasta virheluokasta. Alaluvussa 4.4 avaun lyhyesti muita kolumneissa esiintyneitä virheitä. Lopuksi kokoan analyysini havainnot yhteen ja vastaan aiemmin esittämiini tutkimuskysymyksiin.

Kolumneista löytyi kaiken kaikkiaan 213 virhettä. Virheitä löytyi miltei kaikista Temnikovan virheluokista, mutta ne painottuivat selvästi kolmeen luokkaan: väärin tunnistettuihin tai tunnistamattomiin idiomaattisiin ilmaisuihin (3.5. Error due to incorrectly recognised idiomatic expression), vääräntyyliin sanoihin (3.3. An incorrect word) ja puuttuviin sanoihin tai ilmaisuihin (1. Missing word). Näihin luokkiin kuuluvia virheitä löytyi kaikista viidestä kolumnista. Muiden virheluokkien virheitä taas ei välttämättä löytynyt jokaisesta tekstistä. Välimerkkeihin liittyviä virheitä (luokat 4.1. ja 4.2.) ei löytynyt kolumneista ollenkaan. Luokan 2.2. virheitä (Word order error), jotka liittyvät kokonaisten fraasien siirtämiseen, löytyi kaikista kolumneista yhteensä vain neljä.

Taulukko 3. Käännöksistä löytyneet virheet jaoteltuna Temnikovan (2010) virheluokkiin

	1. Holma (4.8.2021)	2. Stellan (14.6.2023)	3. Karttunen (13.9.2023)	4. Stellan (13.12.2023)	5. Kontula (12.1.2024)	Yht.
(1.) Missing word	3	21	2	9	3	38
(2.1.) Word order error	1	2	0	5	0	8
(2.2.) Word order error	0	2	0	0	0	2
(3.1.) An incorrect word	0	1	0	0	1	2
(3.2.) Correct word	7	2	0	3	0	12

with an incorrect ending (e.g. number or case)						
(3.3.) An incorrect word	9	23	8	20	8	68
(3.4.) Extra word	0	7	0	0	1	8
(3.5.) Error due to incorrectly recognised idiomatic expressions	13	28	5	17	12	75
(4.1.) Missing punctuation sign	0	0	0	0	0	0
(4.2.) Incorrect punctuation sign	0	0	0	0	0	0
	Yht. 33	Yht. 86	Yht. 15	Yht. 54	Yht. 25	Yht. 213

Eniten virheitä löytyi Sanna Stellanin kolumnin *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa* (2023a) käännöksestä. Kolumni on myös teksteistä pisin (alkuperäisteksti 474 sanaa). Se ei kuitenkaan ole merkittävästi pidempi verrattuna muihin tutkimuksen kolumneihin; Antti Holman kolumnin (2021) pituus on 467 sanaa ja Sanna Stellanin toisen kolumnin (2023) pituus 468 sanaa.

Käyn seuraavaksi tuloksia läpi tarkemmin. Esittelen löytöjäni kolmesta suurimmasta virheluokasta alaluvuissa 4.1–4.3. Alaluvussa 4.4 esittelen muita kolumneista löytyneitä virheitä. Jälkieditoidut tekstit virheluokitteluineen löytyvät tutkielman liitteistä. Tässä luvussa käsitelen esimerkkejä teksteissä esiintyneistä virheistä. Olen

lihavoitunut esimerkkien virheet. Jos virhe koskee koko esimerkkiä, sitä ei ole lihavoitu.

Yksi kolumneista (Stellan 2023a) on kääntämisen jälkeen myös jälkieditoitu ChatGPT:llä ennen ihmisen tekemää jälkieditointia. Siksi tämän kolumnin esimerkeissä on esitelty myös ChatGPT:n jälkieditoima versio. Tutkimuksen keskiössä ei kuitenkaan ole ChatGPT:n tekemän jälkieditoinnin arvioiminen. Analyysissä ei siis keskitytä vertaamaan tekoälyn jälkieditoimaa versiota tekoälyllä tehtyyn käännökseen ja ihmisen jälkieditoimaan tekstiin. ChatGPT myös teki jälkieditoinnissa hyvin vähän muutoksia Stellanin (2023a) tekstiin (yht. 6 kpl).

4.1 Väärin tunnistetut idiomaattiset ilmaisut (luokka 3.5.)

Kolumneista löytyi kaiken kaikkiaan 75 luokkaan 3.5. jaoteltavaa virhettä. Olen laskenut tähän luokkaan virheet, jotka liittyivät siihen, ettei saksankielinen käänнос jostain syystä ollut idiomaattinen ja korjaus vaati kokonaisen fraasin, lauseen tai virkkeen muuttamista.

Luokan 3.5. virheitä löytyi kaiken kaikkiaan eniten (28 kpl) Sanna Stellanin kolumnin *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa* (2023a) käännöksestä. Vähiten tämän luokan virheitä (5 kpl) oli Anu Karttusen kolumnin *Miltä meillä haisee? Lapseni kaverin kommentti pysäytti* (2023) käännöksestä. Muissa kolumneissa 3.5.-luokan virheitä oli keskenään suurin piirtein saman verran (12–17 kpl). Karttusen teksti on kolumneista lyhyin ja Stellanin pisin, joten tältä osin virheiden määrä korreloi tekstin pituuden kanssa.

Alaotsikon nimestä huolimatta on huomioitava, että tähän luokkaan jaotellut virheet eivät ole kolumnien ainoita idiomaattisuuden liittyviä virheitä. Idiomaattisuuden liittyviä virheitä on jaoteltu myös muun muassa kahteen muuhun suurimpaan virheluokkaan, jotka esittelen seuraavissa alaluvuissa. Virheet on kuitenkin jaoteltu eri luokkiin sen perusteella, minkälaisia muokkauksia niiden korjaaminen vaatii. Lähtötekstissä on saattanut olla esimerkiksi idiomi tai muu idiomaattinen ilmaisu, jota tekoäly ei ole tunnistanut ja on kääntänyt sen väärin tai epäidiomaattisesti. Koska kolumnien tyyli on usein hyvin persoonallinen, on mahdollista, että kaikki

suomenkielisen tekstin ”idiomaattiset ilmaisut” eivät ole laajassa käytössä, vaan osittain kirjoittajan itse keksimiä. Olen kuitenkin laskenut nämäkin ilmaisut idiomaattisiksi, sillä käänösprosessissa pitäisi tunnistaa, että ilmaus on lähtökielelle tyypillinen eikä sitä todennäköisesti voi kääntää sellaisenaan suoraan kohdekielelle. Esimerkki tällaisesta ilmauksesta löytyy Sanna Stellanin kolumnin (2023a) käännöksestä:

Alkuperäisteksti: *LAPSET SURISIVAT ONNESTA heti, kun mummu tuli.*

ChatGPT:n käänös: *DIE KINDER SUMMSTEN VOR FREUDE, sobald Oma ankam.*

Jälkieditoitu käänös: *SOBALD OMA KAM, WAREN DIE KINDER VOR GLÜCK GANZ AUS DEM HÄUSCHEN.*

Onnesta suriseminen ei ole suomeksi niin vakiintunut ilmaus, että voitaisiin puhua idiomista. Kyseessä on myös ilmaus, jonka merkitys ei ole kirjaimellinen vaan kuvainnollinen. Siksi kääntäjän, tai tässä tapauksessa kääntimen, pitäisi tunnistaa, ettei ilmaus välttämättä toimi kohdekielessä suoraan käännettynä. Tässä tapauksessa suora käänös ei toimi ja jälkieditoidussa versiossa tilalle on muokattu idiomaattisempi vastine. Myös sanajärjestystä on muutettu (virheluokka 2.2.) virkkeen sujuvoittamiseksi.

Toinen vastaavanlainen esimerkki löytyy Anu Karttusen kolumnista (2023). Sana *hajukommentti* ei ole vakiintunut ilmaus suomen kielessä, mutta käänöksessä sille pitäisi löytää idiomaattinen vastine:

Alkuperäisteksti: [...] hankin **lapsen hajukommentin** jälkeen tuoksudiffuuserin ja päätin [...]

ChatGPT:n käänös: [...] kaufte ich nach dem **Geruchskommentar meines Kindes** einen Duftdiffusor und beschloss [...]

Jälkieditoitu käänös: [...] kaufte ich nach dem **Kommentar** einen Duftdiffusor und beschloss [...]

Geruchskommentar on suora käänös sanasta *hajukommentti*. Jälkieditoinnissa tilalle on pyritty löytämään neutraalimpi ja idiomaattisempi käänösratkaisu. Kyseessä on näennäisesti pieni muutos ja virheen voisi periaatteessa luokitella

luokkaan 3.1. tai 3.3. Muutos ei kuitenkaan ole jäänyt vain yksittäisen sanan tasolle, vaan jälkieditoidusta tekstistä on myös jätetty pois määrite *meines Kindes*. Laajempi muutos kokonaisuuteen on johtanut siihen, että olen luokitellut virheen luokkaan 3.5. Väärin käännetyn määritteen *meines Kindes* (*minun lapseni*, vrt. *lapsen*) olisi voinut muokata lähtötekstiä vastaavaksi vastineella *des Kindes* (*lapsen*). Idiomaattisimmalta ratkaisulta tuntui kuitenkin jättää määrite kokonaan pois.

Edellä mainitut esimerkit edustavat eräänlaisia poikkeuksia. Tyypillinen esimerkki luokkaan 3.5. jaotellusta virheestä löytyy Antti Holman kolumnista (2021):

Alkuperäisteksti: *Sydämeni löi miten sattuu, ja mieheni pakotti minut tutkimuksiin.*

ChatGPT:n käänнос: *Mein Herz schlug unregelmäßig, und mein Mann zwang mich zu Untersuchungen.*

Jälkieditoitu käänнос: *Mein Herz schlug unregelmäßig, und mein Mann zwang mich dazu, mich untersuchen zu lassen.*

Rakenne *sich untersuchen lassen* on tässä tapauksessa idiomaattisempi käännosratkaisu kuin *zu Untersuchungen zwingen*. ChatGPT:n käänнос on hyvin suora ja sanasanainen, ja se on pitänyt muuttaa jälkieditoinnissa. Vaikka virkkeestä löytyy virhe, voi samalla kiinnittää huomiota siihen, että tekoäly on kääntänyt virkkeen alkuosan onnistuneesti. Käänнос lauseelle *sydämeni löi miten sattuu* (*Mein Herz schlug unregelmäßig*) ei ole sanasanainen, vaan nimenomaan idiomaattinen ja virkkeeseen sopiva.

Sanna Stellanin kolumni (2023a) käsittelee suomalaisia juhlaperinteitä, joten se sisältää paljon ilmaisuja, jotka viittaavat kulttuurisidonnaisiin ilmiöihin. Esimerkiksi juhlakauden konsepti on kulttuurisidonnainen, eikä aukene sellaisenaan saksalaiselle kohdeyleisölle. ChatGPT:n tuottama sanasanainen käänнос on jälkieditoinnissa muotoiltu uudelleen:

Alkuperäisteksti: *Juhlasesonki on täällä taas.*

ChatGPT:n käänнос: *Die Feier-Saison ist wieder da.*

ChatGPT:n jälkieditoima käännös: *Die Feier-Saison ist wieder da.* (ei muutoksia alkuperäiseen käännökseen)

Jälkieditoitu käännös: *Die Sommersaison der Familienfeiern hat wieder begonnen.*

Kyseinen virhe on myös esimerkki siitä, kuinka joitakin virheitä olisi ollut mahdollista jaotella eri kategorioihin. Esimerkiksi lauseen subjektin *Die Feier-Saison* voisi laskea yhdeksi virheeksi (luokka olisi tällöin 3.1.) ja loppuosan *ist wieder da* toiseksi (luokka 3.5.). Lisäksi virkkeeseen on jälkieditoinnissa lisätty tietoa (*Sommersaison der Familienfeiern*), jolloin mukaan voisi laskea myös luokan 1. virheen. Edellä mainitun kaltaisissa tapauksissa, joissa koko virke on virheellinen ja vaatii korjausta, eikä virheiden erottaminen toisistaan ole yksiselitteistä, olen kuitenkin käsitellyt virkettä yhtenä virheenä. Yleensä tällaiset tapaukset asettuvat juuri luokkaan 3.5.

Suurin osa luokan 3.5. virheistä liittyy siis nimenomaan siihen, että idiomaattinen ilmaus on käännetty sanasanaisesti suomesta saksaan. Virheet eivät ole siinä mielessä kriittisiä, että ne olisivat esimerkiksi loukkaavia kohdekielellä. Alkuperäistekstin merkitys jää kuitenkin monissa tapauksissa välittämättä ja tekstistä välittyy käännösmäinen vaikutelma.

4.2 Väärät sanat tai fraasit: tyyliin liittyvät virheet (luokka 3.3.)

Luokan 3.3. virheitä löytyi aineistosta yhteensä 68 kappaletta. Tähän luokkaan kuuluvat virheet, joiden korjaaminen on vaatinut sanan tai lausekkeen vaihtamista toiseen lähes samanmerkityksiseen rakenteeseen. Luokkaan ei siis kuulu virheitä, jotka koostuvat kokonaisista lauseista tai virkkeistä. Syy muutokselle on ollut epäidiomaattisuus, epätarkkuus tai epäsopiva tyyli.

Eniten luokan 3.3. käännösvirheitä löytyi Sanna Stellanin kolumneista. Kolumnin *Kun töissä iski ruuhka, äiti tuli apuun – ja näytti heti, miksi hän on korvaamaton* (2023b) käännöksessä oli yhteensä 20 luokan 3.3. virhettä ja kolumnin *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa* (2023a) käännöksessä yhteensä 19 virhettä. Vähiten virheitä, 6 kappaletta, löytyi Holman kolumnin (2021) käännöksestä.

Anu Karttusen kolumni (2023) käsittelee kotien ominaisuutta. Teksti ei siis tässä tapauksessa käsittele suomalaiselle kulttuurille ominaista ilmiötä. Kolumni sisältää kuitenkin joitain sanoja, joiden kääntäminen saattaa aiheuttaa hankaluuksia, jos ei ole tietoinen niiden konnotaatioista ja käyttötavoista lähde tai kohdekulttuurissa. Esimerkki tästä on suomen kielen sana *koti*, jota käytetään myös Karttusen kolumnissa useaan otteeseen:

Alkuperäisteksti: Oman **kodin** ominaisuus voi tuntea vain hetkisen [...]

ChatGPT:n käännös: Den charakteristischen Duft des eigenen **Zuhauses** kann man nur für einen kurzen Moment wahrnehmen [...]

Jälkieditoitu käännös: Den charakteristischen Duft der eigenen **Wohnung** kann man nur für einen kurzen Moment wahrnehmen [...]

Suomeksi *koti* on hyvin monikäyttöinen sana. Sillä voi viitata sekä konkreettiseen tilaan että abstraktiin käsitteeseen. Saksan *Zuhause*-sana muistuttaa merkitykseltään suomen *kotia*, mutta sitä käytetään harkitummin kuin suomeksi. Siksi Karttusen kolumnissa *Zuhause* on useammassa kohdassa vaihdettu toiseen, tekstiin paremmin sopivaan lähes samanmerkityksiseen sanaan. Virhe on luokiteltu kategoriaan 3.3.

Myös Antti Holman kolumni (2021) sisälsi muutamia kiinnostavia luokkaan 3.3. kuuluvia virheitä. Seuraavassa esimerkissä onomatopoeettinen eli kahvinkeitin äänestä muistuttava *rupluttaa* on ChatGPT:n käsissä kääntynyt sanaan *plätschern* (liplattaa). *Blubbern* (kuplia, porista) sopii kontekstiin paremmin:

Alkuperäisteksti: Sielussani **rupluttaa** noin vuonna 1998 automarketista ostettu valkoinen kahvinkeitin [...]

ChatGPT:n käännös: In meiner Seele **plätschert** eine weiße Kaffeemaschine [...]

Jälkieditoitu käännös: In meiner Seele **blubbert** eine weiße Kaffeemaschine [...]

Sanna Stellanin kolumnissa (2023b) taas sanan *ura* kääntäminen on aiheuttanut tekoälylle haasteita:

Alkuperäisteksti: Minulla ei olisi **tätä monipuolista uraa** eikä neljää lasta [...]

ChatGPT:n käännös: *Ich hätte nicht **diese vielfältige Karriere** und vier Kinder*
[...]

Jälkieditoitu käännös: *Ich hätte nicht **diesen vielseitigen Beruf** und vier Kinder*
[...]

Kohta on alkuperäiskielelläkin monitulkintainen. Jälkieditoinnissa tulkitsimme, että sanalla ura viitataan tässä tapauksessa enemmänkin kirjoittajan ammatin monipuolisuuteen kuin itse urapolkuun. Siksi *Beruf* (ammatti) tuntui luontevammalta käännösratkaisulta.

Osa luokan 3.3. virheistä on selkeämpiä. Esimerkiksi Sanna Stellanin toisessa kolumnissa (2023a) tekoäly on ensin kääntänyt sanan historia melko onnistuneesti, mutta jälkieditoi kuitenkin tilalle toisen vastineen:

Alkuperäisteksti: *Juhlissa otetuista kuvista muodostuu **historiaa**.*

ChatGPT:n käännös: *Die bei den Feiern aufgenommenen Fotos werden zu **Geschichte**.*

ChatGPT:n jälkieditoima käännös: *Die bei den Feiern aufgenommenen Fotos werden zu **Erinnerungen**.*

Jälkieditoitu käännös: *Die Fotos, die auf Familienfeiern gemacht werden, werden zur **Familiengeschichte**.*

Lopullinen jälkieditoitu versio ei vastaa täysin kumpaakaan ChatGPT:n tarjoamaa vaihtoehtoa. On kuitenkin kiinnostavaa, että jälkieditointi ei ole parantanut käännöstä, vaan päinvastoin ehkä heikentänyt sen laatua. Virkkeeseen on tehty myös muita muutoksia, ja se sisältää myös luokan 3.5. virheen, koska ChatGPT:n kääntämässä ja jälkieditoimassa versiossa virkkeen alku ei ollut idiomaattinen.

Virheluokka 3.3. sisältää siis erilaisia virheitä, joista osa liittyy epäidiomaattisuuteen ja osa väärin käännettyihin sanoihin. Tämä virheluokka sekoittuu eniten luokan 3.5. virheiden kanssa. Huomasin nimittäin virheanalyysia tehdessäni, että osa virheistä olisi sopinut sekä luokkaan 3.3. että 3.5. Näissä tapauksissa luokittelun on ratkaissut se, minkälaista muokkausta virheen korjaaminen vaatii. Jos muutokset jäävät

yksittäisten sanojen tasolle, on virhe luokiteltu luokkaan 3.3. Laajemmat fraasien ja lauseiden tason muutokset on luokiteltu luokkaan 3.5.

4.3 Puuttuvat sanat, fraasit, lauseet, virkkeet tai selvennykset (luokka 1.)

Kolmanneksi suurimpaan virheluokkaan eli luokkaan 1. jaottelin yhteensä 38 virhettä. Virheluokka on sikäli kiinnostava, että päätin jaotella siihen kahdenlaisia virheitä, joita lähestyin en niinkään jonkin lähtötekstin materiaalin puuttumisen vaan lisäyksen tarpeen näkökulmasta. Virheluokka sisältää siis virheitä, jotka liittyvät siihen, että lähtötekstissä ollut sana, fraasi, lause tai virke puuttuu käännöksestä. Olen kuitenkin luokitellut virheluokkaan myös sellaiset tapaukset, joissa käännökseen on pitänyt lisätä jotakin, mitä ei ole ollut lähtötekstissä. Lisäykset ovat yleensä jonkinlaisia selvennyksiä kulttuurisidonnaisiin ilmiöihin. Ne voivat olla myös partikkeleita, joiden lisääminen kolumniin on lisännyt tekstin idiomaattisuutta.

Eniten (21 kpl) tämän virheluokan virheitä löytyi Sanna Stellanin kolumnin (2023a) käännöksestä. Muissa kolumneissa kyseisen luokan virheitä oli selvästi vähemmän ja keskenään lähes saman verran. Kolumnista löytyy esimerkkejä lisäyksistä, joissa ei ole kyse siitä, että tekoäly olisi jättänyt jonkin alkutekstin osan kääntämättä. Lisäysten tarkoitus on huomioida käännöksen kohdeyleisö ja tehdä tekstistä heille mahdollisimman sopiva. Seuraavassa esimerkissä erisnimen yhteyteen on lisätty tietoa siitä, mikä Kotivinkki on:

Alkuperäisteksti: [...] *Kotivinkin kolumnisti* [...]

ChatGPT:n käännös: [...] *Kolumnistin bei Kotivinkki.*

ChatGPT:n jälkieditoima käännös: [...] *Kolumnistin bei Kotivinkki.* (ei muutoksia alkuperäiseen käännökseen)

Jälkieditoitu käännös: [...] *Kolumnistin **bei der Zeitschrift** Kotivinkki.*

Erityisesti tässä tapauksessa, kun erisnimi on suomalainen, on kohdeyleisölle hyvä selventää, mistä on kyse. Toisaalta myös näkökulman muuttuminen voi johtaa siihen, että tekstiin tulee lisätä elementtejä. Samassa Sanna Stellanin kolumnissa on myös tästä esimerkki:

Alkuperäisteksti: *Kansana olemme kykeneväisiä [...]*

ChatGPT:n käänös: *Als Nation sind wir fähig [...]*

ChatGPT:n jälkieditoima käänös: *Als Nation sind wir fähig [...]* (ei muutoksia alkuperäiseen käänökseen)

Jälkieditoitu käänös: [...] *Als **finnische** Nation sind wir fähig, auf eine Art und Weise zu feiern [...]*

Stellan kirjoittaa suomalaisista muille suomalaisille, joten puhe ”meistä” on luontevaa. Saksankielisessä tekstissä viittaussuhteet kuitenkin hämärtyvät. Tekstissä on kyseistä virkettä ennen keskusteltu nimenomaan suomalaisen juhkakulttuurin ominaispiirteistä. Yhtäkkäinen puhe ”meistä” voi hämärtää suhteita sen välillä, kenelle puhutaan ja mistä. Siksi lisäys *finnische* (suomalaisina) on paikallaan.

Kolumnien käänöksistä löytyi vain yksittäisiä kohtia, joissa osa lähtötekstistä oli jäänyt ilman syytä kokonaan käänöksestä pois. Sanna Stellanin kolumnin (2023a) käänöksestä löytyy kuitenkin myös tästä esimerkki:

Alkuperäisteksti: [...] että se näyttää ja **tuntuu** rangaistukselta.

ChatGPT:n käänös: [...] die wie eine Strafe wirkt.

ChatGPT:n jälkieditoima käänös: [...] die fast wie eine Strafe wirkt.

Jälkieditoitu käänös: [...] die nach außen wie eine Bestrafung wirkt und **die sich auch so anfühlt.**

ChatGPT on jättänyt käänöksestä pois sen, että asia ei vain näytä vaan myös tuntuu rangaistukselta.

Seuraava esimerkki Tytti Kontulan kolumnista (2024) sisältää kaksi lisäystä, joiden puuttuminen ChatGPT:n käänöksestä on laskettu virheeksi:

Alkuperäisteksti: *Se perustui saksalaisen Hot or Cool -instituutin raporttiin [...]*

ChatGPT:n käänös: *Diese basierte auf einem Bericht des deutschen Hot or Cool-Instituts [...]*

Jälkieditoitu käänös: *Diese Social-Media-Challenge in **Finnland** basierte sich auf einem Bericht ders deutschen **Denkfabrik** Hot or Cool Institute [...]*

Virkkeessä on lisätty maininta siitä, että se (haaste) on nimenomaan suomalainen. Lisäksi, vaikka Hot or Cool Institute onkin saksalainen toimija, on selvennykseksi hyvä lisätä maininta siitä, mitä instituutti tekee (*Denkfabrik/ajatushautomo*). Esimerkistä käy ilmi, että virkkeeseen on lisätty muitakin osia, sillä persoonapronominin *se* tilalle on vaihdettu *Diese Social-Media-Challenge*. Tämä johtuu siitä, että edeltävää virkettä on muokattu niin, että se on vaikuttanut sitä seuraavan virkkeen sisältöön.

Joihinkin kolumneihin on lisätty myös partikkeleita, joiden tarkoitus on toimia vastineena suomen liitepartikkeleille ja tehdä käännöksestä idiomaattisempi. Näiden puuttuminen on merkitty virheeksi. Esimerkki tällaisesta tapauksesta löytyy Sanna Stellanin kolumnista (2023a):

Alkuperäisteksti: *Eihän* siinä olisi mitään kerrottavaa [...]

ChatGPT:n käännös: *Es gäbe nichts zu erzählen [...]*

ChatGPT:n jälkieditoima teksti: *Es gäbe nichts zu erzählen [...]* (ei muutoksia alkuperäiseen käännökseen)

Jälkieditoitu teksti: ***Und*** es gäbe ***ja*** nichts zu erzählen [...]

Eihän sanan tarkempi merkitys ja sävy välittyi siis käännökseen partikkelin *ja* kautta. Virkkeen alkuun on lisätty myös konjunktio *und (ja)*, jonka tarkoitus on lisätä tekstin sidosteisuutta ja idiomaattisuutta. Katharina Reissin mukaan (1971, 21) partikkeleilla on hyvin tärkeä rooli tekstin idiomaattisuuden varmistamisessa, ja ne kuuluvat oleellisesti sujuvaan tekstiin.

Kuten kahteen muuhun suurimpaan virhekkategoriaan, olen jaotellut myös virheluokkaan 1. erilaisia virheitä. Virheet kuuluvat kuitenkin joitain luokkien 3.3. ja 3.5. virheitä selkeämmin nimenomaan tähän luokkaan, sillä niistä suurin osa ei sopisi mihinkään muuhun Temnikovan virheluokkaan.

4.4 Muut virheet

Kolumneista löytyi myös viiden muun virheluokan virheitä: yksittäisiä virheitä sanajärjestyksessä (2.1.), laajempia virheitä sanajärjestyksessä (2.2.), vääriä sanoja, jotka on pitänyt korvata täysin toisenlaisella vastineella (3.1.), väärin taivutettuja sanoja (3.2.) ja ylimääräisiä sanoja (3.4.). Näihin luokkiin jaoteltavia virheitä ei löytynyt kaikista viidestä kolumnista eli niissä ei ilmennyt samanlaista toistuvuutta kuin kolmen suurimman virheluokan virheissä. Annan seuraavaksi esimerkin jokaisesta virheluokasta.

Luokkaan 2.1. kuuluvia virheitä löytyi kolumneista yhteensä seitsemän kappaletta. Olen jaotellut luokkaan sanajärjestykseen liittyvät virheet, jotka ovat vaatineet yksittäisten sanojen siirtämistä. Esimerkki luokan 2.1. virheestä löytyy Sanna Stellanin kolumnista (2023b):

Alkuperäisteksti: *Tuon aikakauden sitkeät muorit painavat vaihteettomilla pyörillä kirkonkylälle monta kilometriä **päivässä** säässä kuin säässä.*

ChatGPT:n käännös: *Diese zähen Frauen jener Zeit fahren mit ihren ganglosen Fahrrädern viele Kilometer **täglich** ins Dorf, bei jedem Wetter.*

Jälkieditoitu käännös: *Die zähen Frauen ihrer Generation fahren mit ihren ganglosen Fahrrädern **täglich** viele Kilometer ins Dorf, bei jedem Wetter.*

ChatGPT on sijoittanut käännöksessä sanan *täglich* samalle paikalle kuin millä se on alkuperäistekstissä (*monta kilometriä päivässä/viele Kilometer täglich*).

Idiomaattisempi muotoilu saksaksi on kuitenkin *täglich viele Kilometer*. Kyseessä on tyyppinen esimerkki tämän virheluokan virheestä. Olen jaotellut luokkaan myös sellaiset virheet, jotka ovat vaatineet useamman sanan siirtämistä. Edelleen kyse on kuitenkin yksittäisistä sanoista tai yhtenäisistä lausekkeista tai lauseista. Arvioin, että vaikka kyse on useammasta sanasta, ei kognitiivinen kuormitus sinänsä kasva, koska lause tai lauseke liikkuu yhtenäisenä kokonaisuutena. Antti Holman kolumnissa sivulauseen siirtäminen on liitetty luokan 2.1. virheeseen:

Alkuperäisteksti: [...] *joka ryystäisi tassilta, **jos kehtaisi**.*

ChatGPT:n käänös: [...] *die aus einer Untertasse schlürfen würde, **wenn sie sich trauen würde.***

Jälkieditoitu käänös: [...] *die, **wenn sie sich trauen würde,** den Kaffee von einer Untertasse schlürfte.*

ChatGPT:n käänökseen tehty muutos liittyy idiomaattisuuteen ja tyyliin. Jälkieditoidussa käänöksessä painottuu sivulauseen ehto, kehtaaminen.

Luokan 2.2. virheet liittyvät niin ikään sanajärjestykseen. Nämä virheet kuitenkin liittyvät siihen, että useampi sana, lauseke tai lause on vaihtanut paikkaa keskenään. Luokkaan 2.2. kuuluvia virheitä löytyi aineistosta kaksi, mutta ne ovat sama virhe. Virke siis toistuu kaksi kertaa kolumnissa. Virhe löytyy Sanna Stellanin kolumnin (2023b) käänöksestä:

Alkuperäisteksti: *Kenellekään ei oikein tule puhuttuakaan.*

ChatGPT:n käänös: *Mit niemandem spricht man so richtig.*

ChatGPT:n jälkieditoima käänös: *Mit niemandem spricht man so richtig.* (ei muutoksia alkuperäiseen käänökseen)

Jälkieditoitu käänös: *Man spricht auch mit niemandem so richtig.*

ChatGPT on kääntänyt sanajärjestyksen niin suoraan kuin kieliopillisesti mahdollista. Lopullisessa, ihmisen jälkieditoimassa, versiossa sanajärjestys on kuitenkin muokattu idiomaattisemmaksi ja sujuvammaksi.

Luokkien 3.1. ja 3.3. virheiden eli sanavalintoihin liittyvien virheiden erotteluun liittyy haasteita. Luokittelin luokkaan 3.1. virheet, joissa sana on jälkieditoinnissa jouduttu vaihtamaan toiseen, selvästi erimerkityksiseen sanaan tai fraasiin. Luokan 3.3. virheet taas liittyvät siihen, että sana on korvattu kontekstiin sopivammalla synonyymillä eli yleensä yhdellä sanalla. Huomasin analyysia tehdessäni, että luokka 3.1. saattaa sekoittua myös luokkaan 3.5. Luokan 3.1. virheet kuitenkin vaativat vähemmän korjauksia kuin luokan 3.5. virheet. Esimerkiksi Sanna Stellanin kolumnista (2023b) löytyy tapaus, joka voitaisiin periaatteessa jaotella kumpaan luokkaan (3.1. tai 3.3.) tahansa:

Alkuperäisteksti: **Tuon aikakauden** sitkeät muorit painavat vaihteettomilla pyörillä [...]

ChatGPT:n käännös: *Diese zähen Frauen **jener Zeit** fahren mit ihren ganglosen Fahrrädern [...]*

Jälkieditoitu käännös: *Die zähen Frauen **ihrer Generation** fahren mit ihren ganglosen Fahrrädern [...]*

ChatGPT:n käännösratkaisu *Jener Zeit* on hyvin suora käännös alkuperäistekstistä (*tuon aikakauden*). Idiomaattisempi ratkaisu tässä tapauksessa on *ihrer Generation* (*hänen sukupolvensa*). Vaikka *jener Zeit* ja *ihrer Generation* eivät ole täysin toistensa synonyymejä, ovat ne mielestäni merkitykseltään niin lähellä toisiaan, että ne voidaan luokitella luokkaan 3.3. Temnikovan mukaan luokkaan 3.1. kuuluvat virheelliset sanat, joiden korjaaminen vaatii niiden korvaamista täysin erilaisella leksikaalisella yksiköllä. Temnikovan artikkeli ei kuitenkaan tarkenna, mitä leksikaalisella yksiköllä tutkimuksessa tarkoitetaan. Tieteen termipankin (2025b) mukaan leksikaalinen yksikkö on ”mentaalisen leksikon varastoima sanatason aines”. Leksikaalisia yksiköitä ovat sanojen lisäksi muun muassa idiomit. Jokaisen sanan voi siis nähdä omana leksikaalisena yksikkönään, vaikka kyse olisikin synonyymeistä. Siksi ero luokkaan 3.1., joka käsittää virheelliset sanat, joiden tilalle vaihdetaan sanan synonyymi, on joissain tapauksissa hieman epäselvä. Löysin aineistosta kuitenkin kaksi tapausta, jossa virheellinen sana oli korvattu niin erilaisella merkityksellä, että virhe sopii selkeästi luokkaan 3.1. Toinen virheistä löytyi Sanna Stellanin kolumnin (2023a) käännöksestä:

Alkuperäisteksti: **Tänäkin juhlakautena** tuhannet suomalaiset ajavat [...]

ChatGPT:n käännös: *Auch in dieser **Feiersaison** fahren tausende Finnen [...]*

ChatGPT:n jälkieditoima käännös: *Auch in dieser **Feiersaison** fahren tausende Finnen [...]*

Jälkieditoitu käännös: *Auch in **diesem Sommer** fahren wieder tausende Finnen [...]*

ChatGPT on jälleen tarjonnut hyvin suoraa käännöstä sanalle *juhlakausi*. Suomalainen juhlakauden konsepti ei kuitenkaan todennäköisesti avaudu saksankieliselle yleisölle. Myös Suomessa *juhlakausi* voidaan ymmärtää monella

tavalla. Koska teksti on julkaistu kesäkuussa ja siinä käsitellään erityisesti kesän juhlia, on luonteva ratkaisu selvyuden vuoksi *in diesem Sommer (tänä kesänä)*.

Kaksi viimeistä kolumneissa esiintynyttä virheluokkaa, 3.2. ja 3.4., olivat hieman helpompia luokitella. Luokkaan 3.2. olen luokitellut taivutukseen liittyviä virheitä. Virhe voi koskea aikamuotoa taikka sanan lukua tai sukua. Suurin osa luokan 3.2. virheistä liittyy jollakin tavalla tyyliin tai idiomaattisuuteen, ei niinkään kieliopillisuuteen. Jälkieditoinnissa tavoiteltiin toimeksiantajan toiveiden mukaan myös mahdollisimman sukupuolineutraalia kieltä, joten osa luokan 3.2. virheistä liittyy tähän. Esimerkki tällaisesta tapauksesta löytyy Antti Holman kolumnista (2021):

Alkuperäisteksti: *Tupakoitsija väittää, ettei tarvitse nikotiinia [...]*

ChatGPT:n käänнос: ***Ein Raucher** behauptet, **er** brauche kein Nikotin [...]*

Jälkieditoitu käänнос: ***Rauchende** behaupten, **sie** brauchen kein Nikotin [...]*

ChatGPT on kääntänyt tupakoitsijan maskuliinimuotoon *ein Raucher/er (tupakoitsija/hän)*. Tekstissä ei kuitenkaan viitata tiettyyn, miespuoliseen tupakoitsijaan, vaan puhutaan yleisesti tupakoitsijoista. Siksi jälkieditoinnissa tilalle on valittu sukupuolineutraali *Rauchende (tupakoitsijat)*, joihin viitataan pronomiinilla *sie*.

Luokan 3.4. virheet liittyvät siihen, että ChatGPT on lisännyt käännökseen sanan, fraasin, lauseen tai virkkeen, jota ei ole ollut lähtötekstissä. Luokkaan on jaoteltu myös ChatGPT:n tekstiin lisäämät turhat etuliitteet ja artikkelit. Monessa tapauksessa virheet eivät merkittävästi muuta tekstin merkitystä. Voisi siis jopa argumentoida, että ne voisi jättää tekstiin. Olen kuitenkin luokitellut kaikki ChatGPT:n tekemät lisäykset virheiksi, jos ne eivät vastaa lähtötekstiä ja tuo lisäarvoa tekstiin. Esimerkki tällaisesta tapauksesta löytyy Sanna Stellanin kolumnista (2023a):

Alkuperäisteksti: [...] *että se näyttää ja tuntuu rangaistukselta.*

ChatGPT:n käänнос: [...] *die wie eine Strafe wirkt.*

ChatGPT:n jälkieditoima käänнос: [...] *die **fast** wie eine Strafe wirkt.*

Jälkieditoitu käännös: [...] *die nach außen wie eine Bestrafung wirkt und die sich auch so anfühlt.*

ChatGPT on käännöksen tuottamisen jälkeen omassa jälkieditoinnissaan lisännyt lauseeseen sanan *fast* (melkein). Lisäys on turha, vaikka ei sinänsä myöskään esimerkiksi tee lauseesta kieliopillisesti virheellistä. Koska tässä tapauksessa sana ei tuo lisäarvoa tekstiin, luokitellaan se virheeksi.

Edellä olen esitellyt esimerkkejä virheluokista, joita tutkimusaineistossani esiintyi. Seuraavaksi teen yhteenvedon tuloksista ja vastaan tutkimuksen alussa esittämiini tutkimuskysymyksiin.

4.5 Tutkimustulosten yhteenveto

Tutkimuksen tarkoituksena oli vastata seuraaviin tutkimuskysymyksiin:

1. Kuinka hyvin ChatGPT 4.0 onnistuu kolumnien kääntämisessä Suomesta saksaan?
2. Minkälaisia virheitä laaja kielimalli tekee ko. käännöksissä?
3. Mikä on virheiden korjaamisen kognitiivinen työmäärä?
4. Miten käännöksiä tulisi jälkieditoida, jotta ne olisivat julkaisukelpoisia?

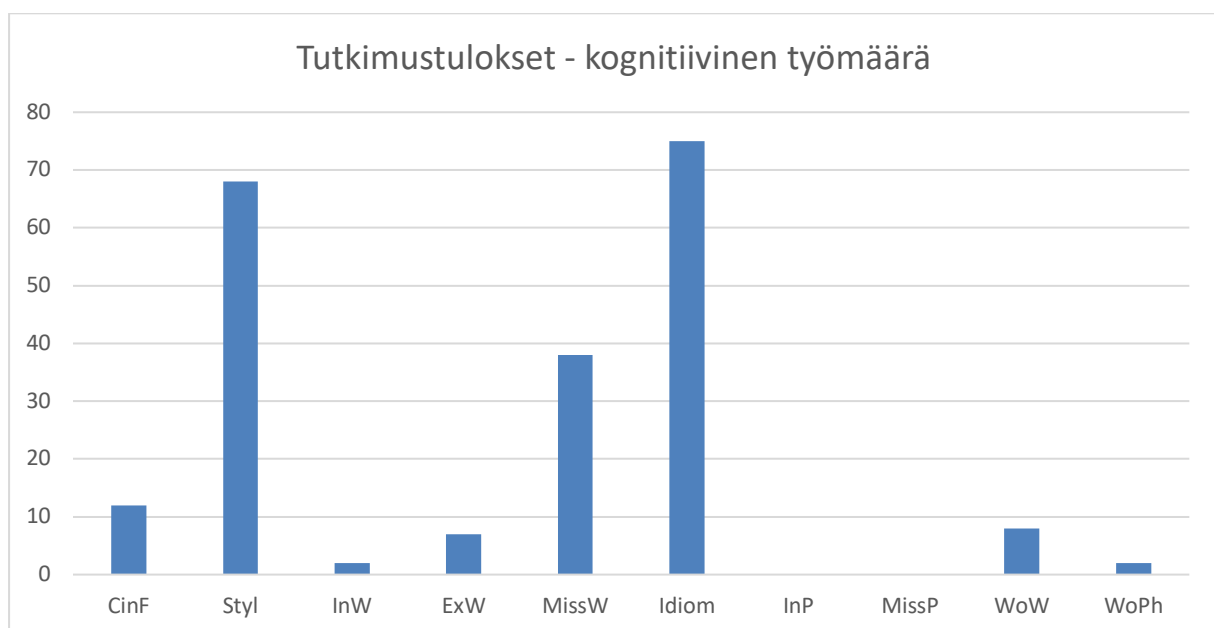
Tutkimuskysymyksiin 2 ja 4 olen vastannut luvussa 4. Suurin osa ChatGPT:n käännösvirheistä liittyy tässä tutkimuksessa idiomaattisuuteen ja tyyliin. Käännökset ovat usein hyvin suoria ja sanasanaisia, mikä johtaa siihen, että ne saattavat olla epäidiomaattisia ja merkitykseltään virheellisiä. ChatGPT ei tehnyt välimerkkeihin liittyviä virheitä, eikä juurikaan kielioppivirheitä. Mainitsin luvussa 2, että aiempien tutkimusten perusteella ChatGPT merkittävimmät virheet liittyvät hallusinointiin sekä erikoissanaston ja kontekstin tunnistamiseen. Kuten ennakoin, omassa aineistossani ei ilmennyt hallusinointia, lukuun ottamatta muutamia ChatGPT:n tekstiin lisäämiä sanoja. Kolumneissa ei ollut varsinaista erikoisalanastoa, mutta tekstit sisälsivät paljon kulttuurisidonnaisia ja kirjoittajien itse keksimiä käsitteitä. Näiden kääntämisessä ja samalla kontekstin tunnistamisessa ChatGPT ei tässäkään tutkimuksessa usein onnistunut.

Jälkiedointiprosessin kautta olen tehnyt ehdotelman siitä, miten käännöksiä tulisi jälkiedoitaa, jotta ne olisivat julkaisukelpoisia. ChatGPT:n käännökset eivät olleet sellaisinaan valmiita julkaistavaksi, vaan vaativat muun muassa tekstin pituudesta ja aihepiiristä riippuen paljon tai erittäin paljon editointia. Vastaus tutkimukseni ensimmäiseen tutkimuskysymykseen on, että ChatGPT onnistuu kieliopin ja oikeinkirjoituksen osalta hyvin kolumnien kääntämisessä, mutta epäonnistuu usein idiomattisen, tavoitellun tyylin mukaisen ja sujuvan käännöksen tuottamisessa (vrt. Khoshafah 2023).

Tutkimustulosten perusteella voidaan todeta, että Temnikovan luokittelun mukaan käännöksissä esiintyvät virheet eivät ole kognitiivisesti kovin kuormittavia. (ks. kaavio 1). Eniten kolumnien käännöksissä esiintyi idiomattisiin ilmauksiin liittyviä virheitä, jotka ovat kognitiiviselta kuormittavuudeltaan asteikon keskivaiheilla. Kolumneissa esiintyi hyvin vähän virheitä, jotka ovat kognitiivisesti erittäin kuormittavia. Tutkimustulokset ovat tältä osin linjassa luvussa 2 esittelemieni Alhon tutkimuksen (tulossa) havaintojen kanssa. Kuten mainitsin luvussa 2, myös Temnikovan (2010) tutkimuksessa konekäännin teki kognitiivisesti kaikista kuormittavimpia virheitä vain vähän.

Kaava 1. Käännöksistä löytyneiden virheiden kognitiivinen työmäärä Temnikovan (2010) luokittelun mukaan

Virheiden kognitiivinen kuormittavuus kasvaa vasemmalta oikealle. Vähiten kuormittavin virhe on taivutukseen liittyvä virhe (CinF) ja eniten kuormittava sanajärjestysvirhe, joka vaatii kokonaisten lausekkeiden siirtämistä (WoPh).



Virheluokittelun lisäksi olen keskustellut käännösten laadusta ja jälkieditoinnin kuormittavuudesta jälkieditoinnissa auttaneen saksan kielen ja kääntämisen lehtorin kanssa. Hänen kokemuksensa oli, että jälkieditointikurssin aikana jälkieditoidut kolumnit eli Sanna Stellanin kolumni (2023a) ja Antti Holman kolumni (2021) tuntuivat raskaimmilta editoida ja jälkieditoinnin työmäärä oli liian suuri. On hyvä huomioida, että lehtori on editoinut minun ja toisen saksan kielen opiskelijan editoimaa tekstiä. Hänen kokemukseensa jälkieditoinnin raskaudesta on siis vaikuttanut se, minkälaisia jälkieditointiratkaisuja tekstiin on tehty ennen hänen editointiaan. Koska kurssin aikana jälkieditoidut tekstit olivat ensimmäisiä tutkimusta varten jälkieditoituja tekstejä, on mahdollista, että tekstejä on jälkieditoitu vähemmän kuin muita tutkimuksen tekstejä. Koen, että jälkieditointikurssin jälkeen tekoälyllä kääntämiäni tekstejä rohkeammin, sillä olin saanut tästä palautetta lehtorilta.

Itse koen, että jälkieditoinnin raskaus korreloi pitkälti virheiden määrän kanssa. Raskaimpia jälkieditoitavia olivat tekstit, joissa oli eniten virheitä. Usein näissä teksteissä oli myös paljon kognitiivisesti kuormittavia luokan 3.5. virheitä. On kuitenkin vaikeaa subjektiivisesti arvioida tunnetta jälkieditoinnin kuormittavuudesta, sillä myös virheluokittelun tekeminen on ollut kognitiivisesti työlästä. Virheluokittelujen päätteeksi olikin vaikea enää erottaa, mikä teksteistä on tuntunut raskaimmalta jälkieditoida. Siksi on tärkeää arvioida kognitiivista työmäärää ensisijaisesti Temnikovan (2010) luokittelun avulla. Näin olen pyrkinyt varmistamaan, että voin vastata tutkimuskysymyksiin mahdollisimman objektiivisesti.

Jälkieditointiprosessissa mitattiin myös jälkieditointiin kulunutta aikaa. Kuluneen ajan analysoimisessa on huomioitava, että jälkieditointiin ei ollut annettu aikarajaa, vaan tekstejä jälkieditoitiin niin kauan, että ne koettiin julkaisukelpoisiksi. Yliopistonlehtorin (S) jälkieditointiin käyttämä aika pitää sisällään myös käännöksen ja virheiden kommentoimisen. On siis oletettavaa, että jälkieditointiin olisi kulunut hieman vähemmän aikaa, jos hän ei olisi myös kommentoinut virheiden laatua. Kokonaisuutena voidaan kuitenkin todeta, että jälkieditointiin kului merkittävän paljon aikaa. Keskimäärin yhden kolumnin jälkieditointiin kului noin 84 minuuttia.

Taulukko 4. Jälkieditointiin kulunut aika

Toinen saksan kielen ja kääntämisen opiskelija (A) osallistui vain kahden kolumnin jälkieditointiin, joten muiden kolumnien jälkieditointiin kulunut aika on merkattu vain tutkielman kirjoittajan (J) ja saksan kielen ja kääntämisen lehtorin (S) sarakkeisiin.

	A	J	S
1. Holma (4.8.2021)	40 min	60 min	120 min
2. Stellan (14.6.2023)	75 min	105 min	90 min
3. Karttunen (13.9.2023)	-	40 min	45 min
4. Stellan (13.12.2023)	-	75 min	210 min
5. Kontula (12.1.2024)	-	90 min	60 min

Koposen ym. (2012) tutkimuksessa eniten aikaa kului juuri Temnikovan luokittelussa kognitiivisesti kuormittavampien virheiden korjaamiseen. Omassa tutkimuksessani yksittäisten virheiden korjaamiseen kulunutta aikaa ei tutkittu. Jälkieditointiin kulunutta aikaa voi kuitenkin yrittää arvioida suhteessa virheiden luonteeseen kussakin kolumnissa. Esimerkiksi Sanna Stellanin kolumnin (13.12.2023) jälkieditoiminen vei keskimäärin eniten aikaa (J 75 min ja S 210 min). Kolumni ei kuitenkaan sisältänyt eniten virheitä (54 kpl) eikä se ollut teksteistä pisin, vaikkakin pisinten joukossa. Voidaan siis olettaa, että virheiden laatu on vaikuttanut kognitiiviseen työkuormaan ja siten myös jälkieditointiin kuluneeseen aikaan. Kolumnissa ei ole kognitiivisesti kuormittavia sanajärjestykseen liittyviä virheitä, mutta siitä löytyy paljon kognitiivisesti melko kuormittavia luokkien 3.5. ja 3.3 virheitä. Kaiken kaikkiaan näin pienen aineiston perusteella on kuitenkin vaikeaa arvioida korrelaatiota jälkieditointiin kuluneen ajan ja virheiden luonteen välillä.

Mitä tulee käännösten laatuun, saksan kielen ja kääntämisen lehtori koki äidinkielenä saksan kielen puhujana erityisesti Sanna Stellanin kolumnin (2023a) ja Antti Holman (2021) kolumnien käännösten laadun niin heikoksi, että tekstit kannattaisi kääntää kokonaan uudelleen jälkieditoinnin sijaan. Myös muiden käännösten laatu koettiin melko heikoksi. Jälkieditoinninkaan jälkeen tekstit eivät lopulta tuntuneet täysin onnistuneilta muun muassa siksi, ettei ChatGPT välttämättä onnistunut välittämään lähtötekstin humoristista sävyä käännökseen.

Vaikuttaa siis siltä, että tekstilajille tyypillisten piirteiden kääntäminen tuottaa laajalle kielimallille haasteita. Vaikka tekoäly teki myös onnistuneita

käännösratkaisuja, se ei monissa kohdin ollut tarpeeksi luova tai osannut ottaa huomioon kulttuurista kontekstia. Tutkimustulokset ovat siis tältä osin linjassa muun muassa Khoshafahin (2023) alustavien havaintojen kanssa. Koposen (2010) tutkimustuloksiin verrattuna voidaan myös todeta, että ChatGPT:n haasteet kääntämisessä ovat osittain samanlaisia kuin perinteisillä konekääntimillä. Koposen tutkimuksessa sekä tilastopohjainen- että neuroverkkopohjainen konekäännin tekivät eniten virheitä lehtiartikkelin käännöksessä verrattuna virustorjuntaohjelman käyttöohjeen ja Euroopan komission vihreän kirjan käännökseen (ks. luku 2.4.2).

Editoinnin määrä ja virheiden laatu heijastuu myös jälkieditoidun tekstin lopulliseen sanamäärään. Kaikkien käännösten sanamäärä kasvoi jälkieditoinnissa (taulukko 4). Tämä selittyy varmasti juuri sillä, että teksteihin on pitänyt lisätä selvennyksiä kulttuurisidonnaisista asioista ja ilmiöistä. Suorat sanasanaiset käännökset ovat harvoin toimineet, vaan niiden tilalle on kaivattu erilaista vastinetta, joka on saattanut muodostua useammasta sanasta.

Taulukko 5. Sanamäärät alkuperäisteksteissä, ChatGPT:n käännöksissä ja jälkieditoiduissa käännöksissä

	Sanamäärä (suomi)	Sanamäärä (saksa, ChatGPT)	Sanamäärä (saksa, ChatGPT:n jälkieditointi)	Sanamäärä (saksa, lopullinen jälkieditoitu)
1. Holma (4.8.2021)	467	710	-	740
2. Stellan (14.6.2023)	474	682	683	738
3. Karttunen (13.9.2023)	239	369	-	374
4. Stellan (13.12.2023)	468	670	-	734
5. Kontula (12.1.2024)	262	377	-	430

Käyn seuraavaksi läpi tutkimuksen haasteita ja onnistumisia kokonaisuutena. Lisäksi esitän ehdotuksia mahdolliselle jatkotutkimukselle.

5 Johtopäätökset

Tutkimuksen tarkoituksena on ollut tutkia A-lehtien Meillä kotona -sivuston kolumneista tehtyjen tekoälykäännösten laatua ja jälkieditoinnin kognitiivista työmäärää. Tutkimuksessa on hyödynnetty erilaisia menetelmiä: virheanalyysiä, jälkieditointia ja jälkieditointiin kuluvan ajan mittaamista sekä keskustelua kognitiivisesta työmäärästä ja käännösten laadusta äidinkielen saksan kielen puhujan kanssa. Tutkimuksessa ilmeni, että suurin osa kolumneissa esiintyneistä virheistä ei ole kognitiivisesti kovin kuormittavia. Jälkieditoijien kokemus kognitiivisesta kuormittavuudesta on kuitenkin hieman erilainen. He kokivat jälkieditoinnin kuormittavaksi ja olisivat suosineet tekstien kääntämistä itse kokonaan uudelleen.

Viidestä käännetystä kolumnista (aineiston laajuus 2 808 sanaa) löytyi yhteensä 213 jälkieditoinnissa korjattua virhettä. Puhe ”virheistä” saattaa olla tässä tapauksessa hieman harhaanjohtavaa, sillä kaikki korjatut tapaukset eivät vaikuta samalla tavalla negatiivisesti käännöksen laatuun. Kuten kääntämisessä käännösratkaisuja, on myös jälkieditoinnissa mahdollisia muokkausmahdollisuuksia useita. Ei siis ole yhtä oikeaa tapaa jälkieditoida tekstiä. Jokainen jälkieditoija tekee omat ratkaisunsa, eikä tutkimukseni siksi ole täysin toisinnettavissa.

Koska saksan kielen ja kääntämisen lehtori ei verrannut tekemiäni jälkieditointiratkaisuja alkuperäiseen tekoälykäännökseen, on mahdollista, että olen esimerkiksi tehnyt tekstiin tarpeettomia muokkauksia. Keskeistä oli, että lopullinen jälkieditoitu versio vastaa merkitykseltään mahdollisimman tarkasti lähtötekstiä sekä on sujuva ja saksankieliselle yleisölle kohdennettu. En myöskään kiinnittänyt jälkieditoinnissa erityistä huomiota tekstin ulkoasuun, joka on myös osa julkaisukelpoisuutta. Tutkimukseni tarkoitus oli analysoida nimenomaan itse käännöstä, eikä niinkään tekoälyn kykyä ottaa huomioon ulkoasuun liittyviä seikkoja. Tulevissa tutkimuksissa myös ulkoasu olisi hyvä ottaa huomioon.

Jälkieditoinnin subjektiivisuuden lisäksi on hyvä huomioida myös virheiden luokittelun subjektiivisuus. Virheitä olisi voinut luokitella monin eri tavoin, jopa samaa virheluokittelua käyttämällä. Koen, ettei tämän tutkimuksen tulosten kannalta

sillä ole kuitenkaan suurta merkitystä, mihin virheluokkiin yksittäiset rajatapaukset asettuvat. Selvästi suurimmat virheluokat olivat joka tapauksessa 3.5., 3.3. ja 1. Temnikovan luokittelu soveltui sinänsä hyvin myös tekoälyllä käännettyjen kolumnien analysoimiseen. Tulevissa tutkimuksissa voisi vielä tarkastella, voisiko Temnikovan luokittelua tarkentaa tai täydentää niin, että se antaisi vielä yksityiskohtaisempaa tietoa virheiden tyypistä ja kognitiivisesta työmäärästä. Mielestäni on oleellista pohtia virheiden kokonaismäärää ja sen vaikutusta kognitiiviseen kuormitukseen. Vaikka virheet eivät olisi luokittelun mukaan kuormittavia, niin jälkieditointiprosessista voi tulla kuormittava, jos virheitä on paljon. Jälkieditointiin kulunut aika antoikin tutkimuksessani viitteitä tästä.

Tulevissa tutkimuksissa olisi hyödyllistä tutkia myös kehoitteiden ja lämpötilan vaikutusta käännökseen. ChatGPT ei tässä aineistossa onnistunut kovin hyvin ottamaan kohdeyleisöä huomioon. Olisikin mielenkiintoista ja hyödyllistä selvittää, minkälaisilla kehoitteilla ja millä lämpötilalla ChatGPT:n puutteet voitaisiin ratkaista ja virheiden määrää vähentää. Yksi tutkimuksessani ilmennyt haaste ChatGPT:lle oli persoonallisen tyylin ja muun muassa huumorin kääntäminen. Olisikin hyödyllistä tutkia, mitä persoonallisen tyylin kääntäminen A-lehtien kolumneissa vaatii tai mistä kolumnien tyyli rakentuu. Tämä voisi tarjota vastauksia muun muassa siihen, minkälaisilla kehoitteilla saa aikaan mahdollisimman onnistuneen käännöksen. Myöhemmin olisi kiinnostavaa tutkia myös sitä, miten saksankielinen yleisö vastaanottaa tekoälyllä käännetyt kolumnit. Tämä on kuitenkin ajankohtaista vasta silloin, kun (jälkieditoidun) käännöksen laatu on niin hyvä, että se täyttää toimeksiantajan odotukset.

Tekoäly voi toimia hyödyllisenä apuvälineenä kääntäjälle. Tähän mennessä tutkimukset eivät ole kuitenkaan vielä osoittaneet vakuuttavia tuloksia siitä, että tekoäly voisi kääntää tekstejä ilman kielten ja kääntämisen asiantuntijoita. Käännösprosessi, kuten jälkieditointi ja oikolukukin, vaatii syvällistä ymmärrystä kielestä ja kulttuurista. Tekoälyn hyödyllisyys kääntämisessä riippuu myös paljon muun muassa tekstilajista, mikä käy ilmi myös tämän tutkimuksen tuloksista.

Jos tekoälyä halutaan tavalla tai toisella hyödyntää kääntämisessä, on tekoälytyökalujen kehittämisessä tärkeää hyödyntää kielten ammattilaisten

osaamista. Tutkimuksen aikana huolenaiheeksi nousivat myös laatuodotukset (tekoälyn tuottamille) teksteille. Tulevissa tutkimuksissa onkin mielestäni tärkeää kiinnittää erityistä huomiota siihen, minkälainen on onnistunut ja julkaisukelpoinen teksti ja minkälaista laatua tekoälyn generoimilta teksteiltä odotetaan.

Lähteet

Primäärilähteet

Holma, A. (2021) ”Keitetäänpä kahvit” on Antti Holman mielestä suomen kielen täysin lause – kolumnisti saa suodatinkahvilla yhteyden synnyinmaahansa. meillakotona.fi-sivuston kolumni 4.8.2021.
<https://www.meillakotona.fi/artikkelit/antti-holman-kolumni-kahvista> (viitattu 25.4.2025).

Karttunen, A. (2023) Miltä meillä haisee? Lapseni kaverin kommentti pysäytti. meillakotona.fi-sivuston kolumni 13.9.2023.
<https://www.meillakotona.fi/artikkelit/anu-karttusen-kolumni-kodin-tuoksusta> (viitattu 25.4.2025).

Kontula, T. (2024) Osallistuin viiden vaateen vuosi -haasteeseen, ja näin siinä kävi. meillakotona.fi-sivuston kolumni 12.1.2024.
<https://www.meillakotona.fi/artikkelit/tytti-kontulan-kolumni-viiden-vaateen-vuosi-haasteesta> (viitattu 25.4.2025).

Stellan, S. (2023a) Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa. meillakotona.fi-sivuston kolumni 14.6.2023.
<https://www.meillakotona.fi/artikkelit/sanna-stellanin-kolumni-juhlimisesta> (viitattu 25.4.2025).

Stellan, S. (2023b) Kun töissä iski ruuhka, äiti tuli apuun – ja näytti heti, miksi on korvaamaton. meillakotona.fi-sivuston kolumni 13.12.2023.
<https://www.meillakotona.fi/artikkelit/kun-toissa-iski-ruuhka-aiti-tuli-apuun> (viitattu 25.4.2025).

Sekundäärilähteet

Alam, F. *ym.* (2024) ‘LLMs for Low Resource Languages in Multilingual, Multimodal and Dialectal Settings’, in M. Mesgar and S. Loáiciga (eds) *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts*. St. Julian’s, Malta: Association for Computational Linguistics, s. 27–33. <https://aclanthology.org/2024.eacl-tutorials.5/> (viitattu 24.3.2025).

A-lehdet (2025) *A-lehdet vie suomalaisen elämäntavan maailmalle – kansainvälinen Kotona-sivusto avaa uuden aikakauden lifestyle-mediassa*. <https://www.a-lehdet.fi/artikkelit/a-lehdet-vie-suomalaisen-elamantavan-maailmalle-kansainvalinen-kotona-sivusto-avaa-uuden-aikakauden-lifestyle-mediassa> (viitattu 14.4.2025).

Alho, M. (tulossa) ‘Open AI and translation quality in translations from Finnish into German – a case study’.

Alvarez-Vidal, S., Oliver, A. ja Badia, T. (2020) ‘Post-editing for Professional Translators: Cheer or Fear?’, *Tradumàtica: tecnologies de la traducció*, s. 49–69. <https://doi.org/10.5565/rev/tradumatica.275>.

Baddeley, A.D. ja Hitch, G. (1974) ‘Working Memory’, G.H. Bower (ed.). Academic Press (Psychology of Learning and Motivation), s. 47–89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1).

Bamberg, J., Jokinen, P. ja Laine, M. (2007) *Tapaustutkimuksen taito / toimittaneet Markus Laine, Jarkko Bamberg & Pekka Jokinen*. Helsinki: Gaudeamus.

Banerjee, S. ja Lavie, A. (2005) ‘METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments’.

Bang, Y. *ym.* (2023) ‘A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity’, in J.C. Park et al. (eds) *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. Nusa Dua, Bali: Association for Computational Linguistics, s. 675–718. <https://doi.org/10.18653/v1/2023.ijcnlp-main.45>.

Bentivogli, L. *ym.* (2018) ‘Machine Translation Human Evaluation: an investigation of evaluation based on Post-Editing and its relation with Direct Assessment’, M. Turchi, J. Niehues, and M. Federico (eds) *Proceedings of the 15th International Conference on Spoken Language Translation*. Brussels: International Conference on Spoken Language Translation, s. 62–69. <https://aclanthology.org/2018.iwslt-1.9/> (viitattu 6.3.2025).

- Blatz, J. *ym.* (2004) 'Confidence Estimation for Machine Translation', *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*. COLING 2004, Geneva, Switzerland: COLING, s. 315–321. <https://aclanthology.org/C04-1046> (viitattu 11.10.2024).
- Bubeck, S. *ym.* (2023) *Sparks of Artificial General Intelligence: Early experiments with GPT-4*. <https://doi.org/10.48550/arXiv.2303.12712>
- Callison-Burch, C., Osborne, M. ja Koehn, P. (2006) 'Re-evaluating the Role of Bleu in Machine Translation Research', D. McCarthy and S. Wintner (eds) *11th Conference of the European Chapter of the Association for Computational Linguistics*. *EACL 2006*, Trento, Italy: Association for Computational Linguistics, s. 249–256. <https://aclanthology.org/E06-1032/> (viitattu 16.1.2025).
- Davis, J. *ym.* (2024) 'The Temperature Feature of ChatGPT: Modifying Creativity for Clinical Research', *JMIR Hum Factors*, 11, p. e53559. Available at: <https://doi.org/10.2196/53559>.
- Doddington, G. (2002) 'Automatic evaluation of machine translation quality using n-gram co-occurrence statistics', *Proceedings of the second international conference on Human Language Technology Research - the second international conference*, San Diego, California: Association for Computational Linguistics, s. 138–145. <https://doi.org/10.3115/1289189.1289273>.
- van Egdom, G.-W. ja Pluymaekers, M. (2019) 'Why go the extra mile? How different degrees of post-editing affect perceptions of texts, senders and products among end users', *The Journal of Specialised Translation*, 2019(31), s. 158–176.
- Eriksson, P. ja Koistinen, K. (2014) *Monenlainen tapaustutkimus / Päivi Eriksson, Katri Koistinen*. Helsinki: Kuluttajatutkimuskeskus (Kuluttajatutkimuskeskuksen tutkimuksia ja selvityksiä, 11/2014).
- Eskola, J., Saarela, M. ja Vilkkä, H. (2018) 'Riittääkö yksi? Tapaustutkimus kuvaajana ja selittäjänä.', in *Ikkunoita tutkimusmetodeihin. 1, Metodien valinta ja aineistonkeruu: virikkeitä aloittelevalle tutkijalle / Raine Valli (toim.) ; kirjoittajat: Eila Aarnos [ja 17 muuta]*. 5., uudistettu painos. Jyväskylä: PS-kustannus.
- Freitag, M. *ym.* (2021) 'Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation', *Transactions of the Association for Computational Linguistics*. B. Roark and A. Nenkova (eds), 9, s. 1460–1474. Available at: https://doi.org/10.1162/tacl_a_00437.
- Freitag, M. *ym.* (2024) 'Are LLMs Breaking MT Metrics? Results of the WMT24 Metrics Shared Task', B. Haddow et al. (eds) *Proceedings of the Ninth Conference on Machine Translation*. *WMT 2024*, Miami, Florida, USA: Association for Computational Linguistics, s. 47–81. <https://doi.org/10.18653/v1/2024.wmt-1.2>.

- Graefe, A. ja Bohlken, N. (2020) 'Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News', *ResearchGate* [Preprint]. <https://doi.org/10.17645/mac.v8i3.3019>.
- Graham, Y. *ym.* (2013) 'Continuous Measurement Scales in Human Evaluation of Machine Translation', A. Pareja-Lora, M. Liakata, and S. Dipper (eds) *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse. LAW 2013*, Sofia, Bulgaria: Association for Computational Linguistics, s. 33–41. <https://aclanthology.org/W13-2305/> (viitattu 6.3.2025).
- Guerberof Arenas, A. (2013) 'What do professional translators think about post-editing', *Jostrans The Journal of Specialised Translation*, 19, s. 75–95.
- Guerreiro, N.M. *ym.* (2023) 'Hallucinations in Large Multilingual Translation Models', *Transactions of the Association for Computational Linguistics*, 11, s. 1500–1517. https://doi.org/10.1162/tacl_a_00615.
- Hadley, J.L. *ym.* (eds) (2022) *Using Technologies for Creative-Text Translation*. New York: Routledge. <https://doi.org/10.4324/9781003094159>.
- Harley, T.A. (2008) *The psychology of language: from data to theory / Trevor A. Harley*. 3rd ed. Hove: Psychology Press.
- Hendy, A. *ym.* (2023) 'How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation'. arXiv. <https://doi.org/10.48550/arXiv.2302.09210>.
- Helsingin yliopisto (no date) *Aikakausedian kansainvälistäminen*. <https://researchportal.helsinki.fi/en/projects/aikakausedian-kansainv%C3%A4list%C3%A4minen> (viitattu 26.3.2025).
- Hämäläinen, M. ja Alnajjar, K. (2021) 'The Current State of Finnish NLP', F.A. Pirinen *ym.* (eds) *Proceedings of the Seventh International Workshop on Computational Linguistics of Uralic Languages. IWCLUL 2021*, Syktyvkar, Russia (Online): Association for Computational Linguistics, s. 65–72. <https://aclanthology.org/2021.iwclul-1.8/> (viitattu 24.3.2025).
- Ingo, R. (1990) *Lähtökielestä kohdekieleen : johdatusta käänntieteeseen / Rune Ingo., Lähtökielestä kohdekieleen : johdatusta käänntieteeseen*. Helsinki: Werner Söderström.
- Khoshafah, F. (2023) *ChatGPT for Arabic-English Translation: Evaluating the Accuracy*. [Preprint] <https://doi.org/10.21203/rs.3.rs-2814154/v1>.

Kilg, M. *ym.* (2022) 'Kolumne einer künstlichen Intelligenz: Wie schreibt eine Robo-Autor*in?', *Die Tageszeitung: taz*, 26.11.2022. <https://taz.de/Kolumne-einer-kuenstlichen-Intelligenz/!5898282/> (21.1.2025).

Kocmi, T. *ym.* (2024) 'Navigating the Metrics Maze: Reconciling Score Magnitudes and Accuracies', L.-W. Ku, A. Martins, ja V. Srikumar (eds) *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. ACL 2024, Bangkok, Thailand: Association for Computational Linguistics, s. 1999–2014. <https://doi.org/10.18653/v1/2024.acl-long.110>.

Koponen, M. (2010) 'Assessing Machine Translation Quality with Error Analysis', *MikaEL: Kääntämisen ja tulkkauksen tutkimuksen symposiumin verkkojulkaisu. Kääntämisen ja tulkkauksen tutkimuksen symposiumi*, Suomen kääntäjien ja tulkkien liitto.

Koponen, M. (2016) 'Machine Translation Post-editing and Effort : Empirical Studies on the Post-editing Process'. <http://hdl.handle.net/10138/160256> (viitattu 11.12.2024).

Koponen, M., Aziz, W., Ramos, L. ja Specia, L. (2012) 'Post-editing time as a measure of cognitive effort', S. O'Brien, M. Simard, and L. Specia (eds) *Workshop on Post-Editing Technology and Practice. AMTA 2012*, San Diego, California, USA: Association for Machine Translation in the Americas. <https://aclanthology.org/2012.amta-wptp.2/> (viitattu 5.3.2025).

Kotilainen, Lauri. (2003) *Parempi lehtijuttu / Lauri Kotilainen., Parempi lehtijuttu*. Helsinki: Inforviestintä.

Krings, H.P. (2001) *Repairing Texts: Empirical Investigations of Machine Translation Post-editing Processes*. Kent State University Press.

Kujala, K. (2011) *YKSI KOLUMNI, KOLME KÄÄNNÖSTÄ: Käännösten kolme erilaista lähestymistapaa*. Diakonia-ammattikorkeakoulu.

Kuusela, A. ja Orell, J. (2023) *Financial Times | Turun Sanomat jatkaa brittiläisen laatulehden juttujen julkaisua: 'Entistä enemmän juttuja asioista, jotka jäävät helposti suomalaisessa uutisvirrassa syrjään', ts.fi*. <https://www.ts.fi/uutiset/5866887> (viitattu 18.2.2025).

Körner, T. (2024) *Generalisiertes Vertrauen in automatisierten Journalismus : Bedeutung und Einflussfaktoren auf das Vertrauen deutscher Leser*innen*. [S.l.]: Springer VS. <https://research.ebsco.com/linkprocessor/plink?id=8d89872e-f856-3429-835b-da817b499771> (viitattu 17.1.2025).

- Lacruz, I. (2017) 'Cognitive Effort in Translation, Editing, and Post-editing', *The Handbook of Translation and Cognition*. John Wiley & Sons, Ltd, s. 386–401. <https://doi.org/10.1002/9781119241485.ch21>.
- Lacruz, I., Denkowski, M. ja Lavie, A. (2014) 'Cognitive demand and cognitive effort in post-editing', S. O'Brien, M. Simard, ja L. Specia (eds) *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas. AMTA 2014*, Vancouver, Canada: Association for Machine Translation in the Americas, s. 73–84. <https://aclanthology.org/2014.amta-wptp.6/> (viitattu 26.3.2025).
- Larigauderie, P., Gaonac'h, D. ja Lacroix, N. (1998) 'Working memory and error detection in texts: What are the roles of the central executive and the phonological loop?', *Applied Cognitive Psychology*, 12(5), s. 505–527. [https://doi.org/10.1002/\(SICI\)1099-0720\(199810\)12:5<505::AID-ACP536>3.0.CO;2-D](https://doi.org/10.1002/(SICI)1099-0720(199810)12:5<505::AID-ACP536>3.0.CO;2-D).
- Marzal, A. ja Vidal, E. (1993) 'Computation of normalized edit distance and applications', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9), s. 926–932. <https://doi.org/10.1109/34.232078>.
- Milmo, D. (2023) 'ChatGPT reaches 100 million users two months after launch', *The Guardian*, 2 February. <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app> (viitattu 25.4.2025).
- Mikkola, A.-M. *ym.* (2004) *Äidinkieli ja kirjallisuus. Käsikirja*. Porvoo: WSOY.
- Moorkens, J. *ym.* (2018) 'Translators' perceptions of literary post-editing using statistical and neural machine translation', *Translation Spaces*, 7(2), s. 240–262. <https://doi.org/10.1075/ts.18014.moo>.
- O'Brien, S. (2005) 'Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability', *Machine Translation*, 19(1), s. 37–58.
- Orasan, C. (2009) 'Post-editing experiments with MT for a controlled language | Request PDF', in *ResearchGate*. https://www.researchgate.net/publication/270878440_Post-editing_experiments_with_MT_for_a_controlled_language (viitattu 11.12.2024).
- Papineni, K. *ym.* (2002) 'BLEU: a Method for Automatic Evaluation of Machine Translation'. <https://doi.org/10.3115/1073083.1073135>.
- Peng, K. *ym.* (2023) 'Towards Making the Most of ChatGPT for Machine Translation', in H. Bouamor, J. Pino, and K. Bali (eds) *Findings of the Association for Computational Linguistics: EMNLP 2023. Findings 2023*, Singapore: Association for

Computational Linguistics, s. 5622–5633.

<https://doi.org/10.18653/v1/2023.findings-emnlp.373>.

Raunak, V. *ym.* (2023) ‘Do GPTs Produce Less Literal Translations?’, A. Rogers, J. Boyd-Graber, and N. Okazaki (eds) *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. ACL 2023, Toronto, Canada: Association for Computational Linguistics, s. 1041–1050.
<https://doi.org/10.18653/v1/2023.acl-short.90>.

Reiss, K. (1971) *Möglichkeiten und Grenzen der Übersetzungskritik : Kategorien und Kriterien für eine sachgerechte Beurteilung von Übersetzungen / Katharina Reiss*. München: Hueber.

Siu, S.C. (2023a) ‘ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation’, *SSRN Electronic Journal* [Preprint] s. 1–36. <https://doi.org/10.2139/ssrn.4448091>.

Siu, S.C. (2023b) ‘Revolutionizing Translation with AI: Unravelling Neural Machine Translation and Generative Pre-Trained Large Language Models’, [Preprint] s. 1–19. <https://dx.doi.org/10.2139/ssrn.4499768>.

Snover, M. *ym.* (2006) ‘A Study of Translation Edit Rate with Targeted Human Annotation’, *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*. AMTA 2006, Cambridge, Massachusetts, USA: Association for Machine Translation in the Americas, s. 223–231. <https://aclanthology.org/2006.amta-papers.25> (viitattu 11.10.2024).

Specia, L. *ym.* (2009) ‘Estimating the Sentence-Level Quality of Machine Translation Systems’, L. Màrquez and H. Somers (eds) *Proceedings of the 13th Annual Conference of the European Association for Machine Translation*. EAMT 2009, Barcelona, Spain: European Association for Machine Translation.
<https://aclanthology.org/2009.eamt-1.5> (viitattu 11.10.2024).

Sundqvist, S. (2011) *Dealing with Cultural Issues in Translating Blog Columns by Jeff Klima*. Linné-yliopisto.

taz (2025) ‘kolumne intelligenzbestie’. <https://taz.de/Kolumne-Intelligenzbestie/!t5898202/> (viitattu 17.1.2025).

Tekniikka&Talous (2024) *Googlen Gemini-tekoälyn kirjoittama kolumni: Yksi laite muutti elämäni*, *Tekniikka&Talous*. <https://www.tekniikkatalous.fi/uutiset/googlen-gemini-tekoalyn-kirjoittama-kolumni-yksi-laite-muutti-elamani/5ebc1bc5-723c-452d-96fe-bb2feb44eddb> (viitattu 17.1.2025).

Temnikova, I. (2010) ‘Cognitive Evaluation Approach for a Controlled Language PostEditing Experiment’, in *Proceedings of the Seventh International Conference on*

Language Resources and Evaluation (LREC'10). Valletta, Malta: European Language Resources Association (ELRA).

Temnikova, I. ja Orasan, C. (2009) 'Postediting Experiments with MT for a Controlled Language. Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages (ISMTCL).', *Postediting Experiments with MT for a Controlled Language. Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages (ISMTCL)*., 2009.

The MQM Council (2025) 'The MQM Error Typology – MQM (Multidimensional Quality Metrics)'. <https://themqm.org/error-types-2/typology/> (viitattu 6.3.2025).

Tieteen termipankki (2025a) *Nimitys: kontrolloitu kieli*. https://tieteentermipankki.fi/wiki/Nimitys:kontrolloitu_kieli (viitattu 5.3.2025).

Tieteen termipankki (2025b). *Kielitiede:leksikaalinen yksikkö (psykolingvistiikka)*. [https://tieteentermipankki.fi/wiki/Kielitiede:leksikaalinen_yksikk%C3%B6_\(psykolingvistiikka\)](https://tieteentermipankki.fi/wiki/Kielitiede:leksikaalinen_yksikk%C3%B6_(psykolingvistiikka)) (viitattu 5.5.2025).

Tietotekniikan termitalkoot (2024a) *generatiivinen tekoäly; sisältää tuottava tekoäly, sanastokeskus.fi*. https://sanastokeskus.fi/tsk/fi/termitalkoot/hakemistot-267.html?page=get_id&id=ID822&vocabulary_code=TSKTT (viitattu 18.2.2025).

Tietotekniikan termitalkoot (2024b) *suuri kielimalli; laaja kielimalli, sanastokeskus.fi*. https://sanastokeskus.fi/tsk/fi/termitalkoot/hakemistot-267.html?page=get_id&id=ID824&vocabulary_code=TSKTT (viitattu 1.3.2025).

Toner, H. (2023) 'What Are Generative AI, Large Language Models, and Foundation Models?', *Center for Security and Emerging Technology*, 12.5.2023, <https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/> (viitattu 18.2.2025).

Uutismedian liitto (no date) *Kolumni, Juttutyypit*. <https://juttutyypit.fi/juttutyypit/kolumni/> (viitattu 18.10.2024).

Vilar, D. *ym.* (2006) 'Error Analysis of Statistical Machine Translation Output', N. Calzolari et al. (eds) *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy: European Language Resources Association (ELRA). <https://aclanthology.org/L06-1244/>.

Liitteet

Liite 1. Jälkieditoidut kolumnit virheluokitteluiheen

1. "Keitetäänpä kahvit" on Antti Holman mielestä suomen kielen täysin lause – kolumnisti saa suodatinkahvilla yhteyden synnyinmaahansa (Antti Holma 4.8.2021)

Istunnon tiedot

Aika: 2.10.2024 klo 6.20

seed (ei tiedossa), temperature (ei tiedossa)

<p>"Lassen Sie uns Kaffee kochen" ist für Antti Holma der perfekte finnische Satz – der Kolumnist fühlt sich durch Filterkaffee mit seiner Heimat verbunden"</p>	<p>Kommentoinut [JM1]: 3.2. correct word with an incorrect ending</p>
<p>"Es war war einfach, sich an die Gewohnheiten eines französischen Mannes in der amerikanischen Kultur zu gewöhnen, aber das finnische Kaffeetrinken werdegebe ich nicht aufgeben gebe ich nicht auf, entschied der Kolumnist von Kotivinkki, Antti Holma.";-</p>	
<p>„Mein Herz schlug unregelmäßig, und mein Mann zwang mich dazu, mich untersuchen zu lassen. zu Untersuchungen-Ich hatte die Herzrhythmusstörungen schon vor Jahren bemerkt, wollte aber nichts sagen. Denn wenn es keine Diagnose gibt, gibt es auch keine Krankheit – so ist das Verhältnis eines finnischen Mannes zum Gesundheitswesen.</p>	<p>Kommentoinut [JM2]: 3.5. incorrectly recognized idiomatic expression</p>
<p>„Lassen Sie zunächst den Kaffee weg“, sagte der Arzt. „Klar“, antwortete ich und fuhr nach Hause, um Kaffee zu trinken.</p>	
<p>Koffeinfreien Kaffee trinke ich nicht, weil das kein Kaffee, sondern Pupswasser-eine eklige Plörre ist.</p>	<p>Kommentoinut [JM3]: 3.3. incorrect word</p>
<p>Die meisten Abhängigen sagen, dass Abhängigkeit nur eine Gewohnheit ist. Eine rRauchender PersonRaucher behauptet, licher brauchen kein Nikotin, sondern eine Pause. Jemand, der abends drei Gläser-Glas Wein trinkt, sagt, er könne auch ohne, aber es sei zur Gewohnheit geworden, ein paar Gläser-Glas zu trinken. Wenn es wirklich nur eine Gewohnheit wäre, könnte der Raucher an einem Lutscher nuckeln und der Weingenießer Saft trinken, aber das tun sie nicht.</p>	<p>Kommentoinut [JM4]: 3.2. correct word with an incorrect ending Kommentoinut [JM5]: 3.2. correct word with an incorrect ending Kommentoinut [JM6]: 3.2. correct word with an incorrect ending</p>
<p>Auch ich trinke keinen koffeinfreien Kaffee, weil das kein Kaffee, sondern Pupswasser-eklige Plörre ist. Dennoch lassen sich Abhängigkeiten nicht nur durch die Wirkstoffe erklären. Abhängigkeiten sind ein komplexes Gemisch aus psychischen und physischen Phänomenen, das die sich von Situation zu Situation und von Mensch zu Mensch unterschiedlich unterscheidet sind. Ihnen gemeinsam ist nur, dass man normalerweise nicht einmal an die Abhängigkeit denkt, bis jemand oder etwas einem befiehlt, aufzuhören.</p>	<p>Kommentoinut [JM7]: 3.2. correct word with an incorrect ending Kommentoinut [JM8]: 3.3. incorrect word Kommentoinut [JM9]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Die Anweisung des Arztes, mit dem Kaffeetrinken aufzuhören, fühlte sich wie ein unangemessener Angriff Hassrede an. Alkohol und Zigaretten habe ich schon vor Jahren aufgegeben, aber Kaffee ist doch kein Rauschmittel! Es ist das Elixier meines Lebens, das Herzstück meiner finnischen Identität der Kern meines Vaterlandes. In meiner Seele plätschert-blubbert eine weiße Kaffeemaschine, die ich etwa 1998 in einem Supermarkt gekauft wurde habe, deren Heizplatte fleckig geworden ist und deren Scharniere und Innereien braun verfärbt sind, weil die Maschine den Kaffeesatz dorthin gespuckt hat, wo kein Mikrofasertuch hinkommt.</p>	<p>Kommentoinut [JM10]: 3.3. incorrect word Kommentoinut [JM11]: 3.5. incorrectly recognized idiomatic expression Kommentoinut [JM12]: 3.3. incorrect word Kommentoinut [JM13]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Wenn die Gäste nach einem Besuch nach Hause kommen, kochen sie noch Abendkaffee, der immer besser ist als der Kaffee bei Besuchen auf dem Besuch.</p>	<p>Kommentoinut [JM14]: 1. missing word Kommentoinut [JM15]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Im Finnischen gibt es keinen perfekteren Satz als „Lassen Sie uns Kaffee kochen“. Er ist immer angebracht, sowohl in Momenten der Freude als auch der Trauer, in Krisen und bei Feierlichkeiten gleichermaßen nützlich.</p>	<p>Kommentoinut [JM16]: 3.2. correct word with an incorrect ending Kommentoinut [JM17]: 3.5. incorrectly recognized idiomatic expression</p>

Während meiner Zeit hier in den USA vermisse ich am meisten das knackige die kurze und knackige traffic Art der finnischen Kultur des Kaffeetrinkens. Wenn Gäste kommen, wird der Kaffeefesttag bereits vor dem Austausch von Neuigkeiten Aufmerksamkeit geschenkt darüber informiert, ob der Kaffee schon fertig ist. Man steht setzt sich an dem neben dem Tisch steht und entschuldigt sich für das die schlechte Angebot Auswahl, das die genauso gut aus Notfallkekse, die nach Gesangbuch schmecken, wie aus sieben Sorten Sandwiches und Törtchen bestehen kann. Man verhandelt über den Milch- und Sahnebedarf, weil man nicht wusste, ob normale Milch reicht oder ob Hafermilch besser ist. Man erlaubt sich, am Tisch zu sitzen wird zur Tafel Tisch gebeten, und sitzt dort, bis der Kaffee getrunken ist und der Arm am Wachtuch mit Mohnblumenmuster klebt. Wenn man aufsteht, gibt die Haut ein schmatzendes Geräusch von sich. Die Gäste bedanken sich, schauen auf die Uhr und sagen, dass sie gehen müssen.

Wenn die Gäste nach Hause kommen, kochen sie noch Abendkaffee, der immer besser ist als der Kaffee bei Besuchen auf dem besuch.

Wenn ich über Sodbrennen klage, fühle ich mich mit meiner Heimat verbunden.

Es war einfach, sich an die Gewohnheiten meines französischen Mannes in der amerikanischen Gesellschaft zu gewöhnen, aber in Sachen Kaffee verhandle ich nicht mehr. Ermüdet von der Presskanne, die einen schwarzen Schlamm am Boden der Tasse hinterlässt, und der Kapselmaschine, die Kaffee mit Abwassergeschmack produziert zubereitet, fuhr bin ich zum Supermarkt gefahren und habe kaufte eine anständige Kaffeemaschine gekauft, deren Thermoskanne so starken Mokka liefert, dass der Gaumen funktbrennt. Wenn ich über Sodbrennen klage, fühle ich mich mit meiner Heimat verbunden.

Trotz allem versprach ich, das Kaffeetrinken aufzugeben, wenn bei mir eine Herzkrankheit diagnostiziert wird. Dann sagte der Arzt, dass meine Herzstolperer normal seien. Freudig rief ich die guten Nachrichten nach in Finnland an und erzählte einem guten Freund die guten Nachrichten einem guten Freund zu einem Freund, der ebenfalls wegen Herzrhythmusstörungen beim Arzt gewesen war. Wir lachten über das Älterwerden wie alte Hühner, und ich teilte ihm mit, dass ich endlich nach Finnland zu Besuch kommen würde. „Lass uns Kaffee trinken gehen“, sagte der Freund.

Antti Holma ist eine als Kosmopolit getarnter finnische Hauswirtschaftsmeister Großbäuerin, die errie, wenn sie sich trauen würde, -den Kaffee aus von einer Untertasse schlürfen würde, wenn er sie sich trauen würde.

Kommentoinut [JM18]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM19]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM20]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM21]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM22]: 3.3. incorrect word

Kommentoinut [JM23]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM24]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM25]: 1. missing word

Kommentoinut [JM26]: 3.3. incorrect word

Kommentoinut [JM27]: 3.2. correct word with an incorrect ending

Kommentoinut [JM28]: 3.2. correct word with an incorrect ending

Kommentoinut [JM29]: 3.3. incorrect word

Kommentoinut [JM30]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM31]: 3.3. incorrect word

Kommentoinut [JM32]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM33]: 1. Missing word

Kommentoinut [JM34]: 3.2. correct word with an incorrect ending

Kommentoinut [JM35]: 3.3. incorrect word

Kommentoinut [JM36]: 2.1. word order error

2. Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa (Sanna Stellan 14.6.2023)

Istunnon tiedot

Aika: 7.10.2024 klo 17.30

seed 1, temperature 0.5

„**Kann-Darf** man schon nach Hause gehen, fragt sich der Finne auf Familienfeiern!“ In der finnischen Feiertradition gibt es eine Besonderheit: Es ist völlig normal, 600 Kilometer an einem Tag hin und zurück zu fahren, um still in einer Ecke **bei jemandes Feier zu sitzen auf einer Party herum zu sitzen**, schreibt Sanna Stellan, Kolumnistin bei **der Zeitschrift** Kotivinkki.“ **Die Sommersaison der Familienfeiern Die Festsier-Saison ist wieder dahat wieder begonnen. Es wird Häuser werden** geputzt, gewischt und renoviert, damit man sich traut, Gäste **nach Hause** einzuladen. Das Angebot wird sorgfältig geplant, Kühlschränke **werden** geleert, Geschirrssets gemietet. **Jede Feiern sind einzigartig gibt es nur einmal und, sie müssen muss** gelingen. **Die Fotos, die auf Familienfeiern gemacht werden, beiden Feiern aufgenommenen Fotos** werden zur **Erinnerungen Familiengeschichte**.

In der finnischen Feiertradition gibt es eine Besonderheit: Es ist völlig normal, 600 Kilometer an einem Tag hin und zurück zu fahren, um still in einer Ecke **bei jemandes Feier auf einer Party herum zu sitzen**. Als **finnische** Nation sind wir fähig, auf eine **Art und Weise** zu feiern, die **nach außen fest** wie eine **Strafe Bestrafung wirkt und die sich auch so anfühlt**.

Man spricht auch mit niemandem so richtig. Mit niemandem spricht man so richtig. Und das sollte man auch deshalb nicht tun, damit man kein falsches Interesse Vortauschen muss. es lohnt sich auch nicht, damit es nicht zur Heuchelei wird., weit man nicht vortauschen will Spaß zu haben.

Auch in diese **mr Festiersaison Sommer** fahren **wieder** tausende Finnen mit einer **Designats Geschenk verats verpackten Aalto Vase von Alvar Aalto als Geschenk** zu den Feiern entfernter Verwandter, wo sie ein Stück **der in den nordischen Ländern traditionellen Sandwichkuchen Sandwichtorte essen** und eine Tasse Kaffee **zu sich nehmen trinken genießen** und darauf warten, dass die Zeit vergeht, damit sie sich trauen, **wieder** nach Hause zu **gehen fahren**.

Mit niemandem spricht man so richtig. Man spricht auch mit niemandem so richtig. Und das sollte man auch deshalb nicht tun, damit man kein falsches Interesse Vortauschen muss. Man spricht mit niemandem so richtig. Und es lohnt sich auch nicht, damit es nicht zur Heuchelei wird weit man nicht vortauschen will Spaß zu haben. Eine kleine Nachricht über den Studienerfolg des eigenen Kindes, die **fast nebenbei** beim Einatmen erzählt wird, geht auch **am n den** Zuhörern vorbei, weil genau in dem Moment Onkel Reijo auf dem Sofa **die seine** Sitzposition wechselt und das Quietschen des Leders die vielsagende Stille der Feier durchbricht.

Zum Glück gibt es immer diesen einen Entertainer. Die gleichen Geschichten hört man auf jeder Feier, aber alle sind dankbar für die Erzählungen, die **im mit dem eigenen** schallenden Lachen des Erzählers enden.

Es ist für uns natürlich, in **der Gruppe Gemeinschaft stit zu sein zu schweigen** und nicht auf Gesprächseröffnungen einzugehen. **Das sei uns gegönnt. Das muss man uns zugestehen.**

M Daacht war nichts kein Problem, der Pfarrer machte weiter, und das Paar heiratete. Die Ehe läuft immer noch gut.

Bei der Hochzeit **eines Freundes meiner Freunde** war es anders. **Einer der EheDer pPartner ist**

Kommentoinut [JM1]: 3.3. an incorrect word
 Kommentoinut [JM2]: 3.5. incorrectly recognized (... [11])
 Kommentoinut [JM3]: 1. missing word
 Kommentoinut [MA4]: 3.5. incorrectly recognized (... [2])
 Kommentoinut [MA5]: 3.3 incorrect word
 Kommentoinut [MA6]: 1. missing word
 Kommentoinut [MA7]: 1. missing word
 Kommentoinut [MA8]: 3.5. incorrectly recognized (... [3])
 Kommentoinut [MA9]: 3.5. incorrectly recognized (... [4])
 Kommentoinut [MA10]: 3.3. incorrect word
 Kommentoinut [MA11]: 3.5. incorrectly recognized (... [5])
 Kommentoinut [MA12]: 1. missing word
 Kommentoinut [MA13]: 3.5. incorrectly recognized (... [6])
 Kommentoinut [MA14]: 3.5. incorrectly recognized (... [7])
 Kommentoinut [MA15]: 3.4. extra word
 Kommentoinut [MA16]: 3.3. incorrect word
 Kommentoinut [MA17]: 1. missing word
 Kommentoinut [MA18]: 1. missing word
 Kommentoinut [MA19]: 2.2. word order error
 Kommentoinut [MA20]: 3.5. incorrectly recognized (... [8])
 Kommentoinut [MA21]: 3.1. incorrect word
 Kommentoinut [MA22]: 1. missing word
 Kommentoinut [JM23]: 3.4. extra word
 Kommentoinut [JM24]: 1. missing word
 Kommentoinut [MA25]: 3.5. incorrectly recognized (... [9])
 Kommentoinut [MA26]: 1. missing word
 Kommentoinut [MA27]: 3.3. incorrect word
 Kommentoinut [MA28]: 3.3. incorrect word
 Kommentoinut [MA29]: 1. Missing word
 Kommentoinut [MA30]: 3.3. incorrect word
 Kommentoinut [MA31]: 2.2. word order error
 Kommentoinut [MA32]: 1. missing word
 Kommentoinut [MA33]: 3.5. incorrectly recognize (... [10])
 Kommentoinut [MA34]: 1. Missing word
 Kommentoinut [MA35]: 3.2. Correct word with inc (... [11])
 Kommentoinut [MA36]: 3.3. incorrect word
 Kommentoinut [MA37]: 3.5. incorrectly recognize (... [12])
 Kommentoinut [MA38]: 3.4. extra word
 Kommentoinut [MA39]: 3.3. Incorrect word
 Kommentoinut [MA40]: 3.3. incorrect word
 Kommentoinut [MA41]: 3.5. incorrectly recognize (... [13])
 Kommentoinut [MA42]: 3.5. incorrectly recognize (... [14])
 Kommentoinut [MA43]: 3.2. correct word with an (... [15])
 Kommentoinut [MA44]: 3.3. incorrect word

<p>Italiener kommt aus Italien und die Feier verlief als Geschehen entsprechend. Die Gäste aus Sizilien waren alle in Schwarz gekleidet. Bei Auf einer Hochzeit im finnischen Sommer war das natürlich sehr unüblich, aber es verhinderte den Erfolg der Feier nicht war trotzdem ein Erfolg.</p>	<p>Kommentoinut [MA45]: 3.3. incorrect word</p> <p>Kommentoinut [MA46]: 3.3. incorrect word</p> <p>Kommentoinut [MA47]: 1. missing word</p> <p>Kommentoinut [MA48]: 3.3. incorrect word</p> <p>Kommentoinut [MA49]: 3.3. incorrect word</p>
<p>Die Gäste blieben auch nicht still in der Kirche sitzen. Amüsant war es, als ein Herr zum Altar ging und den Pfarrer so weit zur Seite schob, damit dass er ein gutes Foto vom Brautpaar direkt vom Altar aus machen konnte. Da war nichts Macht nichts, Kein Problem der Pfarrer machte weiter, und das Paar heiratete. Die Ehe läuft immer noch gut.</p>	<p>Kommentoinut [MA50]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentoinut [MA51]: 3.4. extra word</p> <p>Kommentoinut [MA52]: 3.3. incorrect word</p> <p>Kommentoinut [MA53]: 3.5. incorrectly recognized ... [11]</p>
<p>Als die Gäste ankamen, war noch vieles noch nicht fertig im Gange. Der erste Gast musste die Sahne schlagen, der zweite Kaffee kochen.</p>	<p>Kommentoinut [MA54]: 3.4. Extra word</p> <p>Kommentoinut [MA55]: 3.5. incorrectly recognized ... [2]</p>
<p>Es wird oft steif auf Feiern wirken oft verkrampft, Feiern werden oft steif, wenn man versucht, alles richtig perfekt zu machen. Zu viel Förmlichkeit und die Angst vor Fehlern lassen erstarren die Stimmung erstarren.</p>	<p>Kommentoinut [MA56]: 3.4. extra word</p> <p>Kommentoinut [MA57]: 3.5. incorrectly recognized ... [3]</p> <p>Kommentoinut [MA58]: 3.3. incorrect word</p>
<p>Ein Freund erzählte: A, dass er als er Jüngerer jünger war, hatte er das Tablett der Gastgeberin mit den Begrüßungsgetränken der Gastgeberin ergriffgenommen und fragte: „Wohin soll ich das bringen?“ Das ist lustig, aber gleichzeitig spüre ich ein leichtes Ziehen der Fremdscham unter meinem Brustbein, aber es zwick auch ein wenig vor Mitgefühl in der bekannten Stelle unter dem Brustbein. Das hätte ich sein können. Aber daraus wurde nun eine gute Geschichte. Und es gäbe ja nichts zu erzählen, wenn der Freund einfach das Begrüßungsgetränk genommen und ein natürliches lockeres Gespräch begonnen hätte.</p>	<p>Kommentoinut [MA59]: 3.3. incorrect word</p> <p>Kommentoinut [MA60]: 1. Missing word</p> <p>Kommentoinut [MA61]: 2.1. word order error</p> <p>Kommentoinut [MA62]: 3.5. incorrectly recognized ... [4]</p> <p>Kommentoinut [MA63]: 3.5. incorrectly recognized ... [5]</p> <p>Kommentoinut [MA64]: 2.1. word order error</p> <p>Kommentoinut [MA65]: 3.3. incorrect word</p> <p>Kommentoinut [MA66]: 3.5. incorrectly recognized ... [6]</p>
<p>Zu viel Mühe Wenn man sich zu viel Mühe gibt, kann es auch schiefgehen. Die Namensgebungsfeier meines zweiten Kindes war völlig übertrieben. In meinem Hormonrausch Das Mein Hormonchaos hat mich dazu gebracht, machte ich Dutzende Meter Girlanden aus rustikalen Stoffen zu machen, die ich auf Flohmärkten zusammen gesammelt hatte. Ich kaufte alte, verzierte Kaffeetassen mit Spitzenrand für 50 Personen und schleppte sie ohne Auto, mit dem Bus nach Hause. Ich backte habe alles selbst gebacken und das alles machte ich mit einem neugeborenen Neugeborenen Baby im Tragetuch.</p>	<p>Kommentoinut [MA67]: 1. Missing word</p> <p>Kommentoinut [MA68]: 1. Missing word</p> <p>Kommentoinut [MA69]: 1. Missing word</p> <p>Kommentoinut [MA70]: 3.3. incorrect word</p> <p>Kommentoinut [JM71]: 1. Missing word</p> <p>Kommentoinut [JM72]: 3.5. incorrectly recognized ... [7]</p>
<p>Als die Gäste ankamen, war noch vieles im Gange noch nicht fertig. Der erste Gast musste die Sahne schlagen, der zweite Kaffee kochen.</p>	<p>Kommentoinut [JM73]: 3.5. incorrectly recognized ... [8]</p> <p>Kommentoinut [JM74]: 3.3. incorrect word</p> <p>Kommentoinut [JM75]: 3.5. incorrectly recognized ... [9]</p>
<p>Was kannst du lernen wir daraus lernen? Komm nicht pünktlich zur geplanten angegebenen Zeit zu meiner Feier, sonst musst kann es sein, dass du arbeiten musst. Ich erinnere mich kaum an die Feier, aber den Fotos nach zu urteilen, hatten die Menschen Spaß. Das war wohl das Wichtigste.</p>	<p>Kommentoinut [JM76]: 1. Missing word</p> <p>Kommentoinut [JM77]: 3.5. incorrectly recognized ... [10]</p> <p>Kommentoinut [JM78]: 3.3. incorrect word</p> <p>Kommentoinut [JM79]: 3.5. incorrectly recognized ... [11]</p> <p>Kommentoinut [JM80]: 3.4. extra word</p>
<p>Sanna Stellan findet es immer schwierig, zu einer Feier zu gehen, aber noch schwieriger, sie zu verlassen wieder loszugehen, weil es dann doch so unglaublich schön war ist.</p>	<p>Kommentoinut [JM81]: 3.5. incorrectly recognized ... [12]</p> <p>Kommentoinut [JM82]: 3.5. incorrectly recognized ... [13]</p> <p>Kommentoinut [JM83]: 1. Missing word</p> <p>Kommentoinut [JM84]: 3.5. incorrectly recognized ... [14]</p> <p>Kommentoinut [JM85]: 1. Missing word</p> <p>Kommentoinut [JM86]: 3.3. incorrect word</p>

3. Miltä meillä haisee? Lapseni kaverin kommentti pysäytti (Anu Karttunen 13.9.2023)

Istunnon tiedot

Aika: 8.11.2024 klo 12.25

seed 1, temperature 0.5

"Wie riecht es bei uns? Der Kommentar eines Freundes meines Kindes ließ mich innehalten", "Den charakteristischen **Duft** ~~Duft~~ ~~des~~ ~~eigenen~~ ~~Zuhauses~~ ~~Wohnung~~ kann man nur für einen kurzen Moment wahrnehmen, schreibt Anu Karttunen, Redaktionsleiterin des Magazins Unelmien Talo&Koti.", "Ich hörte einen Satz aus dem Flur, der mich aufhorchen ließ. „Ich hatte vergessen, dass es bei euch so riecht“, sagte der Freund meines jüngsten Kindes, als er hereinkam. Wie, schlecht? Riecht es bei uns, ohne dass wir es selbst merken? Etwa **nach** Schimmel?"

„Nicht schlecht, sondern nach euch“, fuhr der Freund fort, während er seine Schuhe auszog, und ich atmete erleichtert auf. Dann begannen die **Freunde** ~~beiden~~, die Gerüche der **Häuser** ~~Wohnungen~~ ihrer gemeinsamen **Bekanntem** ~~Freunde~~ zu vergleichen. Auch ich erinnere mich noch an die Duftnoten aus den **Häusern** ~~Wohnungen~~ meiner Kindheitsfreunde. Oder daran, wie es bei Oma oder Verwandten roch.

WENN MAN nach langer Zeit nach Hause kommt, kann man für einen kurzen Moment den eigenen Hausgeruch wahrnehmen. Nach ein paar Sekunden ist man schon zu sehr daran gewöhnt, um ihn noch zu riechen. Für mich riecht unsere Veranda immer nach Holz, das in der Sonne erwärmt wurde. Es ist der Duft eines ewigen Sommers. Jemand anderes würde den Geruch vielleicht als verrottenden Holzstapel beschreiben.

Für mich riecht unsere Veranda immer nach Holz, das in der Sonne erwärmt wurde. Es ist der Duft eines ewigen Sommers.

Obwohl ich den charakteristischen Duft unseres Hauses mag, kaufte ich nach dem **Geruchskommentar** ~~Kommentar~~ ~~meines Kindes~~ einen Duftdiffusor und beschloss, dass es bei uns künftig nur noch betörend herrlich riechen soll. Ich verliebte **dem Geruch von** Zitronengras und Pfefferminze ~~und~~ ~~goss~~ ~~stellte~~ die aromatischen Ölmischungen **mit so großer** ~~Begeisterung~~ ~~her~~, dass die **Familienmitglieder** ~~übrige Familie~~ anfang **sich** ~~zu~~ ~~beklagen~~. Ich beschloss; mich zu beruhigen und den Diffusor nur bei besonderen Anlässen **duften zu lassen** ~~Duft zu verbreiten~~.

ICH BEKAM EINMAL einen unschlagbaren Geruchstipp von einem Experten. Wenn du **vermutest** ~~denkst~~, dass der Geruch deines Hauses **etwas** ~~is~~ ~~Besorgniserregendes~~ ~~ist~~ ~~hat~~, solltest du einen Stoff aus dem Haus nehmen, ihn fest in einen Plastikbeutel verschließen und den Beutel beispielsweise in der Umkleidekabine eines Schwimmbades oder in einer anderen ziemlich neutralen Geruchsumgebung öffnen. Wenn ein Schimmelgeruch oder muffiger Gestank aus dem Beutel strömt, solltest du etwas dagegen tun – etwas anderes als einen Duftdiffusor zu kaufen.

Wenn nicht, kannst du tief einatmen und seufzen: Ah, der **Duft** ~~vondes~~ ~~eigenen~~ ~~Zuhauses~~!"

Kommentoinut [JM1]: 3.3. incorrect word

Kommentoinut [JM2]: 1. missing word

Kommentoinut [JM3]: 3.3. incorrect word

Kommentoinut [JM4]: 3.3. incorrect word

Kommentoinut [JM5]: 3.3. incorrect word

Kommentoinut [JM6]: 3.3. incorrect word

Kommentoinut [JM7]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM8]: 1. missing word

Kommentoinut [JM9]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM10]: 3.3. incorrect word

Kommentoinut [JM11]: 3.3. incorrect word

Kommentoinut [JM12]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM13]: 3.3. incorrect word

Kommentoinut [JM14]: 3.5. incorrectly recognized idiomatic expression

Kommentoinut [JM15]: 3.5. incorrectly recognized idiomatic expression

4. Kun töissä iski ruuhka, äiti tuli apuun – ja näytti heti, miksi hän on korvaamaton (Sanna Stellan 13.12.2023)

Istunnon tiedot

Aika: 8.11.2024 klo 12.24

seed 1, temperature 0.5

"Als die Arbeit stressig wurde ich überfordert mit der Arbeit überfordert Arbeit war, kam meine Mutter zur Hilfe - und zeigte sofort, warum sie unersetzlich ist", "Meine siebzigjährige Mutter kam zu Besuch und brachte das ganze Haus in nur drei Tagen auf Vordermann. In den diesen Omas steckt eine Kraft, die so ich fürchtet scheint mir → eine verschwindende Ressourcen aussterbender Schatz ist zu sein, schreibt Sanna Stellan.", "Ohne meine Mutter wäre ich nicht die, die ich bin. Das ist natürlich biologisch offensichtlich, aber auch im Hinblick auf die Bewältigung meines Alltags. Ich hätte nicht diesen vielfältige vielseitigen Karriere/Beruf und vier Kinder, wenn meine Mutter nicht an meiner Seite wäre. Auch jetzt noch, mit wo sie 75 Jahre alt ist n.

Als es in der auf der Arbeit hektisch wurde, kam meine Mutter zur Hilfe. Mit dem Zug, aus der Region Nordsatakunta im Nord-Satakunta Südwesten Finnlands, mit einem Rollkoffer voller im Wald gesammelter Beeren, selbstgemachter Marmeladen und Säfte. In den Säften hatte sie die Sommerernte konserviert – von Äpfeln, Johannisbeeren, Aronia und Rhabarber in verschiedenen Mischungen. konserviert. Einige waren mit Ingwer, Zimt und Minze gewürzt. Alle mit viel Liebe gemacht.

Als Handtasche benutzt meine Mutter immer noch nur die alte Tasche in altes Taschen, obwohl der Weihnachtsmann schon eine neue gebracht hat. Die spart sie für besondere Anlässe. Ihren Mantel hängt sie immer ordentlich auf einen Bügel, niemals lässt sie ihn auf einem Stuhl liegen. Das Material würde leiden. Um alles Sie kümmert Sie sich um alles gut mit Sorgfalt, Besonders Vor allem um die Kinder.

„Innnerhalb von drei Tagen räumte verlagerte räumte meine Mutter die Fahrräder in den Fahrradkeller, lagerte in den Keller die Pelargonien in den Winterschutz Keller für den Winter ein, paarte diesortierte Socken paare und ordnete brachte Ordnung in die deckellosen Dosen und Gläser.“

Sobald Oma kam, SOBALD OMA KAM, WAREN waren d DIE KINDER SUMMSTEN VOR FREUDE WAREN VOR GLÜCK GANZ AUS DEM HÄUSCHEN, sobald Oma kam. Ein es reservierte einen Schlafplatz an ihrer Seite, ein anderer übernachtete ging doch nicht bei einem Freund Freunden übernachten. Meine Position als Familienm Mittelpunkt der Familie wurde unbedeutend, und das störte mich nicht. Den Alltag mit einem anderen Erwachsenen zu teilen, ist wunderbar. Mit jemandem, der meine Kinder mindestens genauso liebt wie ich.

Innnerhalb von drei Tagen räumte hat meine Mutter die Märbacka-Pelargonien, die im Sommer zartrosa blühen, von den Terrassen in den Winterschutz Keller verlagert. Sie fattete bewältigte den Wäscheberg, paarte sortierte die Socken paare, ordnete die deckellosen Dosen, die sich im Trockenschrank Geschirrabtropfschrank über der Spüle angesammelt hattenen, und deckelosen Dosen und dasie Geschirr, dasie per Hand gespült werden muss und danach im Schrank vergessen wird, hort steckengebliebenen handwaschbaren Sachen. Sie bereitete leckeres Essen zu, reinigte den Kühlschrank, zupfte die trockenen Blätter von den Pflanzen. All das natürlich mit meiner vierjährigen Jüngsten, die überall mit dabei war.

Der Höhepunkt war, dass meine Mutter die Fahrräder der Familie in den Keller des Hauses brachte. Man muss durch eine kleine schmale Tür, danach gibt es über einen unverhältnismäßig viel zu steile Trepp hohen Abstieg. Wie diese Frau, die altersbedingt vom

Kommentoinut [JM1]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM2]: 3.3. incorrect word
 Kommentoinut [JM3]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM4]: 3.3. incorrect word
 Kommentoinut [JM5]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM6]: 3.3. incorrect word
 Kommentoinut [JM7]: 1. missing word
 Kommentoinut [JM8]: 2.1. word order error
 Kommentoinut [JM9]: 3.3. incorrect word
 Kommentoinut [JM10]: 3.3. incorrect word
 Kommentoinut [JM11]: 2.1. word order error
 Kommentoinut [JM12]: 3.3. incorrect word
 Kommentoinut [JM13]: 3.3. incorrect word
 Kommentoinut [JM14]: 3.3. incorrect word
 Kommentoinut [JM15]: 3.3. incorrect word
 Kommentoinut [JM16]: 1. missing word
 Kommentoinut [JM17]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM18]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM19]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM20]: 1. missing word
 Kommentoinut [JM21]: 2.1. word order error
 Kommentoinut [JM22]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM23]: 3.2. correct word with incorrect ending
 Kommentoinut [JM24]: 3.2. correct word with incorrect ending
 Kommentoinut [JM25]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM26]: 3.2. correct word with incorrect ending
 Kommentoinut [JM27]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM28]: 3.3. incorrect word
 Kommentoinut [JM29]: 1. missing word
 Kommentoinut [JM30]: 3.5. incorrectly recognized idiomatic expression
 Kommentoinut [JM31]: 3.3. incorrect word
 Kommentoinut [JM32]: 3.5. incorrectly recognized (... [1])
 Kommentoinut [JM33]: 3.5. incorrectly recognized (... [2])
 Kommentoinut [JM34]: 3.5. incorrectly recognized (... [3])

"Als die Arbeit stressig wurde ich überfordert mit der Arbeit überfordert Arbeit war, kam meine Mutter zur Hilfe - und zeigte sofort, warum sie unersetzlich ist", "Meine siebzigjährige Mutter kam zu Besuch und brachte das ganze Haus in nur drei Tagen auf Vordermann. In **den diesen** Omas steckt eine Kraft, die **so ich fürchtet scheint mir - eine verschwindende Ressourcen aussterbender Schatz ist zu sein**, schreibt Sanna Stellan.", "Ohne meine Mutter wäre ich nicht die, die ich bin. Das ist natürlich biologisch offensichtlich, aber auch im Hinblick auf die Bewältigung meines Alltags. Ich hätte nicht **diesen vielfältige vielseitigen Karriere Beruf** und vier Kinder, wenn meine Mutter nicht an meiner Seite wäre. Auch jetzt noch, **mit wo sie 75 Jahre alt ist**."

Als es **in der auf der** Arbeit hektisch wurde, kam meine Mutter zur Hilfe. **Mit dem Zug, aus der Region Nordsatakunta im Nord-Satakunta Südwesten Finnlands**, mit einem Rollkoffer voller im Wald gesammelter Beeren, selbstgemachter Marmeladen und Säfte. In den Säften hatte sie die Sommerernte **konserviert - von Äpfeln**, Johannisbeeren, Aronia und Rhabarber in verschiedenen Mischungen. **konserviert**. Einige waren mit Ingwer, Zimt und Minze gewürzt. Alle mit viel Liebe gemacht.

Als Handtasche benutzt meine Mutter immer **noch nur die alte Tasche in alten Taschen**, obwohl der Weihnachtsmann schon eine neue gebracht hat. Die spart sie für besondere Anlässe. Ihren Mantel hängt sie immer ordentlich auf einen Bügel, niemals lässt sie ihn auf einem Stuhl liegen. Das Material würde leiden. **Um alles Sie kümmert Sie sich um alles gut mit Sorgfalt, Besonders Vor allem** um die Kinder.

„**Innnerhalb von** drei Tagen **räumte vertagerte räumte** meine Mutter die Fahrräder **in den Fahrradkeller**, lagerte **in den Keller** die Pelargonien in den **Winterschutz Keller für den Winter** ein, **paarte diesortierte** Socken **paare** und **ordnete brachte Ordnung in** die deckellosen Dosen **und Gläser**.“

Sobald Oma kam, SOBALD OMA KAM, WAREN waren dDIE KINDER SUMMSTEN VOR FREUDE WAREN VOR GLÜCK GANZ AUS DEM HÄUSCHEN, sobald Oma kam. Ein **es** reservierte einen Schlafplatz an ihrer Seite, ein anderer **übernachtete ging** doch nicht bei **einem Freund Freunden** übernachten. Meine Position als **Familienm Mittelpunkt der Familie** wurde unbedeutend, und das störte mich nicht. Den Alltag mit einem anderen Erwachsenen zu teilen, ist wunderbar. Mit jemandem, der meine Kinder mindestens genauso liebt wie ich.

Innnerhalb von drei Tagen **räumte hat** meine Mutter die Märbacka-Pelargonien, **die im Sommer zartrosa blühen**, von den Terrassen in den **Winterschutz Keller verlagert**. Sie **fattete bewältigte** den Wäscheberg, **paarte sortierte** die Socken **paare**, ordnete die **deckellosen Dosen, die sich im Trockenschrank Geschirrabtropfschrank über der Spüle** angesammelt **hattenen, und deckellosen Dosen und dasie Geschirr, dasie per Hand gespült werden muss und danach im Schrank vergessen wird, hort steckengebliebenen handwaschbaren Sachen**. Sie bereitete leckeres Essen zu, reinigte den Kühlschrank, zupfte die trockenen Blätter von den Pflanzen. All das natürlich mit meiner vierjährigen Jüngsten, die überall mit dabei war.

Der Höhepunkt war, dass meine Mutter die Fahrräder der Familie in den Keller des Hauses brachte. Man muss durch eine kleine schmale Tür, danach **gibt es über einen unverhältnismäßig viel zu steile Treppchen hohen Abstieg**. Wie diese Frau, die **altersbedingt vom**

Kommentoinut [JM1]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM2]: 3.3. incorrect word
Kommentoinut [JM3]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM4]: 3.3. incorrect word
Kommentoinut [JM5]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM6]: 3.3. incorrect word
Kommentoinut [JM7]: 1. missing word
Kommentoinut [JM8]: 2.1. word order error
Kommentoinut [JM9]: 3.3. incorrect word
Kommentoinut [JM10]: 3.3. incorrect word
Kommentoinut [JM11]: 2.1. word order error
Kommentoinut [JM12]: 3.3. incorrect word
Kommentoinut [JM13]: 3.3. incorrect word
Kommentoinut [JM14]: 3.3. incorrect word
Kommentoinut [JM15]: 3.3. incorrect word
Kommentoinut [JM16]: 1. missing word
Kommentoinut [JM17]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM18]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM19]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM20]: 1. missing word
Kommentoinut [JM21]: 2.1. word order error
Kommentoinut [JM22]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM23]: 3.2. correct word with incorrect ending
Kommentoinut [JM24]: 3.2. correct word with incorrect ending
Kommentoinut [JM25]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM26]: 3.2. correct word with incorrect ending
Kommentoinut [JM27]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM28]: 3.3. incorrect word
Kommentoinut [JM29]: 1. missing word
Kommentoinut [JM30]: 3.5. incorrectly recognized idiomatic expression
Kommentoinut [JM31]: 3.3. incorrect word
Kommentoinut [JM32]: 3.5. incorrectly recognized ... [1]
Kommentoinut [JM33]: 3.5. incorrectly recognized ... [2]
Kommentoinut [JM34]: 3.5. incorrectly recognized ... [3]

5. Osallistuin viiden vaatteen vuosi -haasteeseen, ja näin siinä kävi (Tytti Kontula 12.1.2024)

Istunnon tiedot

Aika: 8.11.2024 klo 12.21

seed 1, temperature 1

<p>"Ich nahm an der Fünf-Kleidungsstücke-pro-Jahr-Challenge teil habe versucht, nur fünf neue Kleidungsstücke pro Jahr zu kaufen – und so ist es gelaufen", "Als ich am 2. Juni einen neuen Oberenteil für eine Feier kaufte, wusste ich, dass es mein letztes neues Kleidungsstück für das Jahr sein würde, schreibt Tytti Kontula, die in der Redaktion des finnischen Medienkonzerns A-Lehdet arbeitet. Chefredakteurin Tytti Kontula: "Das Experiment zwang mich, zwischen Wünschen und echten Bedürfnissen zu unterscheiden.", "Anfang letzten Jahres beteiligte habe ich die Herausforderung Challenge angenommen, nur fünf neue Kleidungsstücke pro Jahr zu kaufen, mich an der Fünf-Kleidungsstücke-pro-Jahr-Social-Media-Challenge. Diese Social-Media-Challenge in Finnland basierte sich auf einem Bericht des deutschen Denkfabrik Hot or Cool-Instituts Institute, dem zufolge der effektivste Weg zur Reduzierung des CO₂-Fußabdrucks von Kleidung darin besteht, weniger davon zu kaufen. Fünf neue Kleidungsstücke wären in Ordnung.</p>	<p>Kommentointu [JM1]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM2]: 3.3. an incorrect word</p> <p>Kommentointu [JM3]: 1. missing word</p> <p>Kommentointu [JM4]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM5]: 1. missing word</p> <p>Kommentointu [JM6]: 1. missing word</p> <p>Kommentointu [JM7]: 3.3. incorrect word</p>
<p>Ich war habetrefte schon seit langem eine überlegte Käuferin neuer Kleidungbewusste Entscheidungen beim Kleidungskauf getroffen. Dennoch war ich überrascht, als ich die neuen Anschaffungen des Jahres 2022 zählte: Es waren 12. Das bedeutet ein Kleidungsstück pro Monat, obwohl ich das Gefühl hatte, kaum etwas zu kaufen. Deshalb erschien mir die Fünf-Kleidungsstücke-Challenge als ein interessantes Experiment.</p>	<p>Kommentointu [JM8]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Ich war begeistert von einem gestreiften Shirt und klickte es innerhalb von Sekunden befand es sich schon in dem Warenkorb. Mein Kontingent war bereits bei 3/5 Es war schon das dritte neue Kleidungsstück des Jahres, ein schreckliches Tempo! Ich werde es nie schaffen!</p>	<p>Kommentointu [JM9]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM10]: 3.5. incorrectly recognized idiomatic expression</p>
<p>WIE VERLIEF DAS JAHR? Ich kaufte sofort im Januar Outdoor-Schuhe. Die alten hielten kein waren nicht mehr wasser dicht mehr, also waren sie ein Muss musste ich neue kaufen. Im Januar legte ich mir auch eine Jeans zu. Ich ging auf eine Geschäftsreise und hatte das Gefühl, keine ordentlichen Hosen zu haben. Meine dritte Anschaffung machte ich im März. Ich war begeistert von einem Shirt mit grünen und weißen -weiß gestreiften Shirt Streifen und klickte es innerhalb von Sekunden befand es sich schon im den Online-Warenkorb des Onlineshops. Mein Kontingent war bereits bei 3/5, ein schreckliches Tempo! Es war schon das dritte neue Kleidungsstück des Jahres. Ich werde es nie schaffen!</p>	<p>Kommentointu [JM11]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM12]: 3.3. incorrect word</p> <p>Kommentointu [JM13]: 3.3. incorrect word</p> <p>Kommentointu [JM14]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM15]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Als ich am 2. Juni kaufte ich einen neuen Oberenteil ein Top für eine Feier. Ich wollte dieses unter einem Blazer tragen, den ich auf dem Flohmarkt gefunden hatte, auf dem Flohmarkt gefundenen Blazer kaufte, Ich wusste ich, dass es mein letztes neues Kleidungsstück des Jahres sein würde. Ich musste für die Möglichkeit für das fünfte 5/5-Kleidungsstück zu kaufen musste ich eine Option für den Rest des Jahres lassen noch offenhalten, falls ich wirklich etwas benötigen sollte.</p>	<p>Kommentointu [JM16]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM17]: 3.5. incorrectly recognized idiomatic expression</p>
<p>Die Hälfte des Jahres kaufte ich also überhaupt keine neue Kleidung Kleider. Als meine Handschuhe kaputtgingen, benutzte ich ein anderes Paar, das ich im Schrank gefunden habe meine Ersatzhandschuhe – die waren auch gut genug. Feiern Ich ging zu Ich feiern ging ich in meinen alten Outfits. Wenn ich etwas „unbedingt haben musste“, versuchte ich es auf dem Flohmarkt zu finden. Ich finde, die Kleider-Challenge hat ihren Zweck erfüllt, denn währenddessen in der Zeit der Challenge musste ich wirklich über meine Bedürfnisse und Wünsche beim Konsum nachdenken und den Unterschied zwischen ihnen beiden erkennen.</p>	<p>Kommentointu [JM18]: 3.1. incorrect word</p> <p>Kommentointu [JM19]: 3.5. incorrectly recognized idiomatic expression</p> <p>Kommentointu [JM20]: 3.4. extra word</p> <p>Kommentointu [JM21]: 3.3. incorrect word</p> <p>Kommentointu [JM22]: 3.3. incorrect word</p> <p>Kommentointu [JM23]: 3.3. incorrect word</p> <p>Kommentointu [JM24]: 3.3. incorrect word</p>
<p>Auch mein klimabezogener Weltschmerz das weltweite Klima-Leiden angst wurde dadurch etwas gelindert. Bewertung des Experiments: 6/5!"</p>	<p>Kommentointu [JM25]: 3.5. incorrectly recognized idiomatic expression</p>

Liite 2. ChatGPT:lle annettu käännöskehote

You are an expert translator specializing in Finnish to German translations.

You have a deep understanding of both Finnish and international German-speaking cultures, idioms, and linguistic nuances.

Your task is to translate content for www.meillakotona.fi, a Finnish home and lifestyle website, ensuring that the translations are culturally appropriate and resonate with a global German-speaking audience while maintaining the original Finnish essence.

Key points about the brand:

- Finland's leading online service for home enthusiasts (decorators, renovators, builders, cooks, bakers, DIY and craft enthusiasts, and gardeners)

Brand values and tone:

- Inspiring, encouraging, and joyful
- Reliable, knowledgeable, and contemporary
- Inclusive, warm, and respectful of diverse lifestyles
- Promotes sustainable living
- Direct and conversational, addressing readers as equals

Translate the given Finnish text into standard German, following these guidelines:

- Accurately convey the meaning and spirit of the original Finnish content
- Adapt cultural references and idioms for an international German-speaking audience
- Maintain the brand's friendly and inspiring tone
- Use standard German spelling, vocabulary, and expressions

Liite 3. ChatGPT:lle annettu jälkieditointikehote (käännös 2)

As an expert German language editor, refine and polish the given text, adhering to these guidelines:

1. Precision: Ensure the text accurately reflects the original text's meaning and intent.
2. Linguistic excellence: Apply advanced German grammar, syntax, and style principles.
3. Tone refinement: Maintain and enhance the friendly, inspiring, and conversational tone.
4. Cultural nuance: Fine-tune cultural references for a global German-speaking audience.
5. Stylistic enhancement: Elevate the prose for clarity, fluency, and impact.
6. Lexical consistency: Ensure uniform terminology and phrasing throughout.
7. Structural integrity: Preserve and optimize formatting elements.
8. Rhetorical finesse: Improve argumentation, coherence, and logical flow.
9. Factual accuracy: Verify and correct any factual assertions as needed.
10. Idiomatic mastery: Refine idiomatic expressions for natural, native-like language.
11. Readability optimization: Adjust sentence structure and vocabulary for optimal comprehension.

Deutschsprachige Kurzfassung

KURZFASSUNG

Universität Helsinki

Humanistische Fakultät

Masterprogramm Übersetzen und Dolmetschen

Juuli Muona

Masterarbeit 62 Seiten, Anhang 10 Seiten

Mai 2025

Einsatz eines großen Sprachmodells zur Übersetzung von Kolumnen

Fallstudie zur Qualität von KI-Übersetzungen und kognitiven Aufwand bei ihrer
Nachbearbeitung im Sprachpaar Finnisch–Deutsch

1. Einleitung

Diese Arbeit untersucht die Qualität von ChatGPT 4.0-Übersetzungen finnischer Kolumnen ins Deutsche sowie den damit verbundenen kognitiven Arbeitsaufwand des Nachbearbeitungsprozesses. ChatGPT und andere große Sprachmodelle haben in den letzten Jahren ihre Nützlichkeit in vielen Bereichen, u.a. im Übersetzen, bewiesen. Die vorliegende Arbeit trägt zur bislang wenig erforschten Frage zur Qualität und Leistungsfähigkeit von ChatGPT 4.0 bei der Übersetzung von Kolumnen bei. Da das Übersetzen mithilfe großer Sprachmodelle aber immer mehr Aufmerksamkeit weckt, ist es wichtig, das Thema zu forschen.

Diese Arbeit ist ein Teil des Projekts „Aikakausmedian kansainvälistäminen“ (Internationalisierung von Zeitschriften), welches in Zusammenarbeit zwischen der Universität Helsinki und dem finnischen Medienkonzern A-lehdet durchgeführt wird. Das Ziel des Projekts ist es herauszufinden, inwiefern der Gebrauch künstlicher Intelligenz bei der Internationalisierung der Lifestyle-Medien von A-lehdet von Hilfe ist. (A-lehdet 2025.) Die Inhalte stellen verschiedene Textsorten dar. Der Fokus dieser Arbeit ist auf Kolumnen der Webseite „Meillä kotona“ (meillakotona.fi), weil diese Textsorte vermutlich am meisten Schwierigkeiten für die KI beim Übersetzen

bereitet. Das Forschungsmaterial besteht aus fünf Kolonnen von vier verschiedenen Autoren. Die Länge der Texte variiert zwischen 250 und 500 Wörtern.

Die Forschungsfragen sind folgende:

1. Inwieweit kann ChatGPT 4.0 qualitativ hochwertige Übersetzungen finnischer Kolonnen der Webseite *Meillä kotona* ins Deutsche übersetzen?
2. Welche Art Fehler kommen in den Übersetzungen vor?
3. Wie hoch ist der mit den generierten Übersetzungen verbundene kognitive Arbeitsaufwand des Nachbearbeitungsprozesses?
4. In welchem Umfang müssen die Texte nachbearbeitet werden, damit sie veröffentlicht werden könnten?

Um diese Fragen zu beantworten, werden verschiedene Forschungsmethoden benutzt, u.a. Fehleranalyse und Messung des Zeitverbrauchs des Nachbearbeitungsprozesses (vgl. Kapitel 3). Die Hypothese ist, dass ChatGPT 4.0 die Kolonnen nicht erfolgreich übersetzt und dass idiomatische Ausdrücke und kulturspezifische Begriffe besonders schwierig zu übersetzen sind. Dazu werden u.a. Fehler in Wortfolge, Interpunktion und Deklination erwartet. Eine weitere Hypothese ist, dass der kognitive Arbeitsaufwand des Nachbearbeitungsprozesses hoch ist.

Der Aufbau der Arbeit ist folgender: Als erstes werden die wichtigsten Begriffe und Theorien erläutert. Danach werden die Forschungsmethoden und das Forschungsmaterial vorgestellt. Als nächstes folgt die Analyse des Materials, in der auch Beispiele aus dem Material vorgestellt werden. Zum Schluss werden u.a. Vorschläge für weitere Forschung präsentiert.

2. Theoretischer Hintergrund

In diesem Kapitel werden zuerst die Begriffe *Kolonne* und *großes Sprachmodell* erklärt. Danach wird frühere Forschung zu KI-Übersetzungen, Fehleranalyse, Post-Editing und kognitive Arbeitsaufwand vorgestellt.

Eine Kolumne kann als ein journalistischer Text bezeichnet werden. Im Gegensatz zu vielen anderen journalistischen Texten, ähneln Kolumnen oft literarischen Texten, da sie die eigenen Erfahrungen und Persönlichkeit der Autoren widerspiegeln.

Kolumnen enthalten oft Humor oder Ironie, die auch im Material dieser Arbeit zu finden sind. (Uutismedian liitto n.d.; Kotilainen 2003, 68.) Kolumnen verfolgen oft kein bestimmtes Ziel, sondern sie stellen ein subjektives Erlebnis dar. Deswegen steht auch die Kreativität im Mittelpunkt des Schreibprozesses. (Hadley et al. 2022, 6.)

Obwohl Kolumnen manchmal übersetzt werden (siehe z. B. Kuusela und Orell 2023) gibt es nicht viel Forschung, die sich mit dem Thema auseinandersetzt. Es gibt aber bestimmte Eigenschaften in Kolumnen, die schwierig zu übersetzen sein können, wie zum Beispiel idiomatische Ausdrücke. Das Übersetzen idiomatischer Ausdrücke oder Idiome kann sogar für professionelle Übersetzer schwierig sein (Ingo 1990, 246). Es ist also anzunehmen, dass die KI auch damit Schwierigkeiten hat.

Große Sprachmodelle (*Large language model* oder LLM) sind KI-Modelle, die dazu geeignet sind, unterschiedliche sprachliche Inhalte zu empfangen und generieren. Der Begriff *generative KI* wird manchmal als Synonym für ein großes Sprachmodell verwendet. (Toner 2023.) Das Arbeitsprinzip ist das gleiche: beide bestehen aus mehrschichtigen neuronalen Netzen. Generative KI beschränkt sich aber nicht nur auf sprachliche Inhalte, sondern generiert auch andere Art Inhalte, wie zum Beispiel Bilder, Videos und Audio (Tietotekniikan termitalkoot 2024). Die in dieser Arbeit verwendete KI (ChatGPT 4.0) wird als großes Sprachmodell bezeichnet, weil es speziell dazu geeignet ist, sprachliche Inhalte zu produzieren und verarbeiten.

In den vergangenen Jahren gab es Beispiele dafür, dass KI beim Schreiben journalistischer Texte – einschließlich Kolumnen – verwendet wurde (siehe z. B. Taz 2025; Körner 2024; Tekniikka&Talous 2024). Es ist anzunehmen, dass KI auch Kolumnen übersetzt. Es scheint aber, dass es kaum Forschung darüber gibt, inwiefern ChatGPT oder andere große Sprachmodelle es schaffen, Kolumnen qualitativ hochwertig zu übersetzen. Bisherige Studien haben sich u.a. darauf konzentriert, wie sich KI von früheren Technologien wie statistische oder neuronale maschinelle Übersetzung unterscheidet (siehe z. B. Hadley 2022). Außerdem ist untersucht worden, wie KI unterschiedliche Sprachen und Textsorten übersetzen kann (siehe z. B. Khoshafah 2023) sowie wie unterschiedliche

Eingabeaufforderungen (*Prompt*) die Qualität der Texte beeinflusst, (siehe z. B. Peng et al. 2023). Es gibt aber immer noch viel zu forschen, weil die Technologie sich schnell weiterentwickelt.

Um zu beurteilen, mit welchem Erfolg ChatGPT 4.0 die Texte übersetzt, sollte man sich mit früheren Studien zur Bewertung der Qualität der maschinellen Übersetzungen auseinandersetzen. Es gibt verschiedene Methoden, die Qualität einer maschinellen Übersetzung zu bewerten, von denen einige automatisch sind und andere von Menschen verwirklicht werden. Ein bekanntes Beispiel einer automatisierten Bewertungsmethode ist die BLEU-Metrik (Papineni et al. 2002). Dabei werden maschinelle Übersetzungen mit menschlichen Referenzübersetzungen verglichen. Es wird statistisch ausgewertet, inwiefern die Texte miteinander übereinstimmen. Je höher der BLEU-Wert, desto mehr stimmen die Texte überein und besser ist die maschinelle Übersetzung. (Papineni et al. 2002, 311–312.) Andere Beispiele für ähnliche automatisierte Bewertungsmethoden sind HTER-Methode (Snover et al. 2006), NIST-Methode (Doddington 2002) und METEOR (Banerjee und Lavie 2005).

Ein Vorteil der automatischen Methoden ist, dass große Mengen von Texten sehr schnell bewertet werden können. Jedoch ist der Gebrauch automatisierter Bewertungsmethoden auch mit Einschränkungen verbunden. Weil es mehrere gute Übersetzungsmöglichkeiten für einen Text gibt, ist es problematisch, die Qualität einer Übersetzung mit bestimmten Referenztexten zu beurteilen. Außerdem ist beispielsweise BLEU für seine Ungenauigkeit kritisiert worden (Callison-Burch et al. 2006). Dazu kann man sagen, dass es durch die technologischen Vorschritte heutzutage bessere Alternativen für automatische Bewertungen, als z. B. BLEU gibt (Kocmi et al. 2024).

Es gibt aber auch andere Möglichkeiten, die Qualität der Übersetzungen zu beurteilen, bei den keine automatische Methoden benutzt werden. Post-Editing, also die Nachbearbeitung maschineller Übersetzungen, ist eine Methode, die bei der Bewertung der Qualität der ursprünglichen maschinellen Übersetzung verwendet werden kann. Außerdem kann man mit Hilfe einer Fehlerklassifizierung identifizieren, welche Fehler es in der Übersetzung gibt und mit welchem Erfolg der

Text übersetzt ist. Diese beiden Methoden – Post-Editing und Fehlerklassifizierung – wurden in dieser Arbeit eingesetzt.

Den Schwerpunkt dieser Studie bildet die Fehlerklassifizierung von Irina Temnikova (2010), die Fehler in maschinellen Übersetzungen in zehn Kategorien einteilt, u.a. Deklinationsfehler, Fehler bei der Wortfolge und Fehler bei idiomatischen Ausdrücken. Jede Kategorie ist mit einem bestimmten kognitiven Aufwand verbunden. Fehler auf der syntaktischen Ebene beispielsweise sind mit dem höchsten kognitiven Arbeitsaufwand verbunden. Dagegen sind Fehler auf der morphologischen Ebene am wenigsten belastend. (Temnikova 2010, 3487–3488.) Obwohl auch andere Fehlerklassifikationssysteme existieren, z. B. die von Koponen (2010), The MQM Council (2025) und Vilar et al. (2006), die bei der Bewertung der Übersetzungen verwendet werden könnten, bietet Temnikovas Ansatz interessante Information mit Hinblick auf den kognitiven Arbeitsaufwand des Nachbearbeitungsprozesses.

3. Methoden und Material

In der Arbeit werden sowohl qualitative und als auch quantitative Methoden verwendet. Als erstes wurden die Kolumnen aus dem Deutschen ins Finnische mithilfe des KI-Modells (ChatGPT 4.0) übersetzt. Im KI-Tool von A-lehdet gab es bereits die finnische Ausgangstexte, deren Übersetzungen in dieser Arbeit analysiert werden. Außerdem gab es Eingabeaufforderungen zum Übersetzen und die Nachbearbeitung der Texte. Alle Kolumnen wurden mithilfe der Eingabeaufforderung zum Übersetzen mit ChatGPT übersetzt. Dazu wurde einer der Texte vor der menschlichen Nachbearbeitung als ein Teil einer Kursaufgabe durch ChatGPT 4.0 nachbearbeitet. Danach wurden die Kolumnen von Experten der deutschen Sprache und Übersetzung nachbearbeitet (Verfasserin der Arbeit, ein anderer Studierender sowie eine Dozentin in deutscher Sprache und Übersetzung, deren Muttersprache Deutsch ist). Die für den Nachbearbeitungsprozess benötigte Zeit, wurde systematisch erfasst. In der Folge des Post-Editings habe ich die vorgenommenen Änderungen in Temnikovas (2010) Fehlerkategorien eingeordnet und den damit verbundenen kognitiven Arbeitsaufwand bewertet. Die Mengen der

Fehler in den Texten wurden miteinander verglichen, um einen Überblick über die Ergebnisse zu gewinnen.

Das Material besteht aus fünf Kolumnen auf der Meillä kotona -Webseite:

1. *”Keitetäänpä kahvit” on Antti Holman mielestä suomen kielen täysin lause – kolumnisti saa suodatinkahvilla yhteyden synnyinmaahansa*

(Antti Holma, 4.8.2021) Originaltext 467 Wörter, Übersetzung 710 Wörter

2. *Joko kehtaa lähteä kotiin, pohtii suomalainen sukujuhlissa*

(Sanna Stellan, 14.6.2023) Originaltext 474 Wörter, Übersetzung 682 Wörter

3. *Miltä meillä haisee? Lapseni kaverin kommentti pysäytti*

(Anu Karttunen, 13.9.2023) Originaltext 239 Wörter, Übersetzung 369 Wörter

4. *Kun töissä iski ruuhka, äiti tuli apuun – ja näytti heti, miksi hän on korvaamaton* (Sanna Stellan, 13.12.2023) Originaltext 468 Wörter,

Übersetzung 670 Wörter

5. *Osallistuin viiden vaatteen vuosi -haasteeseen, ja näin siinä kävi*

(Tytti Kontula, 12.1.2024) Originaltext 262 Wörter, Übersetzung 377 Wörter

Für die Auswahl der Texte gibt es keine anderen Kriterien als die Textsorte. Fünf Kolumnen, insgesamt 1 910 Wörter (Übersetzungen 2 808 Wörter), bereiten genug Material, um einen Überblick über den kognitiven Arbeitsaufwand des Post-Editings zu bekommen. Weniger Material hätte die Analyse beschränkt, während ein größerer Umfang an Material eine tiefgründige Analyse im Rahmen dieser Arbeit ausgeschlossen hätte.

4. Analyse

In den Kolumnen gab es insgesamt 212 Änderungen, die in die Fehlerkategorien von Temnikova (2010) eingeordnet wurden. Mit 75 Fällen traten die meisten Fehler in der Kategorie 3.5 – Fehler aufgrund einer fehlerhaften Erkennung idiomatischer Ausdrücke – auf. Das heißt, dass die KI entweder einen idiomatischen Ausdruck nicht oder fehlerhaft erkannt hat. Mehrmals hat ChatGPT 4.0 den Ausdruck wortwörtlich übersetzt. Die zweitgrößte Kategorie mit 68 Fehler war 3.3. (An

incorrect word). In dieser Kategorie wurden Wörter, deren Gebrauch stilistisch oder kontextual nicht passend war, eingeordnet. Ein Beispiel ist das Wort *Zuhause*, das in der Übersetzung der Kolumne von Anu Karttunen falsch verwendet wurde. Im Finnisch wird das Wort *koti* teilweise auf einer anderen Weise verwenden als im Deutschen das Wort *Zuhause*.

In der Analyse wird über „Fehler“ gesprochen. Es soll aber berücksichtigt werden, dass nicht alle Änderungen bzw. Fehler gleichermaßen schwerwiegend sind. Als Fehler wurden zum Beispiel auch solche Fälle klassifiziert, bei denen die Übersetzung eine Erklärung benötigte, obwohl es diese im Originaltext nicht gab (Temnikovas Kategorie 1. *Missing word*). Diese hängen meistens mit kulturspezifischen Phänomenen oder Ausdrücken zusammen. Kategorie 1. war die drittgrößte Fehlerkategorie mit 38 Fehlern. Bei der Interpunktion wurden in den ChatGPT-Übersetzungen keine Fehler festgestellt. Dazu gab es kaum Grammatikfehler in den Übersetzungen.

Aus der Analyse lässt sich schließen, dass ChatGPT 4.0 Schwierigkeiten hatte die Zielgruppe zu berücksichtigen. Wenn die KI die Zielgruppe berücksichtigt hätte, würden die übersetzten Kolumnen weniger wortwörtliche Übersetzungen enthalten. Außerdem gäbe es Erklärungen zu kulturspezifischen Phänomenen und Ausdrücke in den Übersetzungen.

Einer der Kolumnen (Stellan 14.6.2023) wurde nach dem Übersetzen von ChatGPT 4.0 nachbearbeitet. Die Nachbearbeitung war aber nicht sehr erfolgreich. ChatGPT hat insgesamt sechs Änderungen am Text vorgenommen. Die Änderungen haben den Text nicht wirklich verbessert, sondern in einigen Fällen sogar Fehler hinzugefügt. Das Ziel dieser Arbeit ist es, die Übersetzungen und nicht die Nachbearbeitungsfähigkeit von ChatGPT 4.0 zu analysieren. Deswegen geht die Analyse nicht tiefgründig darauf ein, auf welche Weise ChatGPT 4.0 den Text nachbearbeitet hat.

Die Fehleranalyse lässt darauf hindeuten, dass der mit der Korrektur verbundene kognitive Aufwand überwiegend niedrig ist; die Mehrheit der Fehler fällt in die Kategorien geringer bzw. moderater kognitiver Belastung. Fehler mit sehr hoher kognitiver Belastung konnten nicht identifiziert werden. Wenn man aber die für die Nachbearbeitung benötigte Zeit anschaut, zeigt es das Gegenteil. Die

Nachbearbeitung hat im Durchschnitt 84 Minuten pro Kolumne gedauert. Dazu haben die Experten, die die Texte nachbearbeitet haben, geäußert, dass der Arbeitsaufwand besonders bei bestimmten Kolumnen zu hoch ist. Es muss aber auch beachtet werden, dass keine eindeutigen Anweisungen für den Nachbearbeitungsprozess vorlagen (z. B. wie viel Zeit soll gebraucht werden).

Letztendlich heißt dies, dass die Übersetzungen von ChatGPT 4.0 umfassende Nachbearbeitung benötigen, mit besonderem Augenmerk auf stilistische und idiomatische Ausdrücke. Nach Temnikovas (2010) Fehlerklassifizierung sind die Fehler kognitiv nicht sehr belastend, obwohl der Nachbearbeitungsprozess viel Zeit gekostet hat.

5. Schlussfolgerungen

Es soll berücksichtigt werden, dass man die Fehler in dieser Arbeit auf verschiedene Weise hätte klassifizieren können, sogar innerhalb der gleichen Fehlerklassifizierung. Obwohl einzelne Fehler in unterschiedliche Kategorien hätten eingeordnet werden können, sind die größten Fehlerkategorien trotzdem eindeutig. Diese Arbeit bietet also einen Überblick über die Fehler in den von ChatGPT 4.0 übersetzten Kolumnen.

Einer der größten Probleme in den übersetzten Kolumnen war, dass ChatGPT viele Stellen wortwörtlich übersetzt hat. Darüber hinaus hatte ChatGPT Schwierigkeiten, die Zielgruppe zu berücksichtigen und kulturspezifische Phänomene zu erkennen. Zukünftige Studien könnten sich darauf konzentrieren, wie Eingabeaufforderungen verfeinert werden können, so dass die Übersetzung idiomatischer und zielgruppengerechter ist. Dazu könnte man die Temperatureinstellung berücksichtigen, da diese Einstellung beeinflusst, wie vorhersehbar oder zufällig die Übersetzung wird (Davis et al. 2024).

Weil der persönliche Schreibstil, einschließlich Humor, in Kolumnen besonders wichtig ist, könnte auch das erforscht werden, was diesen Stil ausmacht und wie ChatGPT so gesteuert werden kann, dass es eine qualitativ hochwertige Übersetzung generiert.