

Tutkimus ChatGPT:llä luotujen käännösten laadusta

TURUN YLIOPISTO

Tietotekniikan laitos

Tietojenkäsittelytiede

LuK-tutkielma

Toukokuu 2024

Josefiina Heikkilä

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

LuK-tutkielma
Tietotekniikan laitos, Teknillinen tiedekunta
Turun yliopisto

Tutkinto-ohjelma: Tietojenkäsittelytiede

Tekijä: Josefiina Heikkilä

Otsikko: Tutkimus ChatGPT:llä tehtyjen käännösten laadusta

Sivumäärä: 34 sivua, 10 liitesivua

Päivämäärä: Toukokuu 2024

Tekoälyohjelmistot ovat kehittyneet paljon viime vuosien aikana. Myös niiden käyttö on yleistynyt niin yrityksissä kuin yksityisihmisillä. Suomen kielellä uusia malleja on tutkittu verrattain vähän, vaikka laajan käytön vuoksi on tärkeää ymmärtää, mihin tekoäly pystyy ja mitkä ovat sen heikkoudet. Tässä tutkielmassa tutkitaan, miten suosioon noussut ChatGPT onnistuu kääntämisessä englanti-suomi-kieliparilla.

ChatGPT hyödyntää GPT-3.5-kielimallia, joka on koulutettu vastaamaan käyttäjän antamaan kehotteeseen. GPT-kielimalli on arkkitehtuuriltaan transformer ja hyödyntää itsehuomiomekanismia. Tämä mahdollistaa esimerkiksi tekstissä kaukana toisistaan olevien osien välisten riippuvuuksien huomioimisen. Kääntämisessä itsehuomiomekanismi voi parantaa käännöksen laatua, koska konteksti huomioidaan.

Tutkimuksen käännösaineistoksi on kerätty 20 uutisartikkelia: 10 suomenkielistä ja 10 englanninkielistä. Uutisartikkelit syötetään ChatGPT:lle kerran ja kehotteessa pyydetään ohjelmistoa tuottamaan käännös. Tämän jälkeen tuotetut käännökset pisteytetään käyttäen MQM-pisteytysmenetelmää. MQM-menetelmässä virheet jaetaan tyypeittäin ja vakavuusasteittain. Jaottelun avulla niille lasketaan numeerinen virhearvo. Mitä suurempi tai vakavampi virhe, sitä suurempi on virheen virhearvo.

Tuotetuista käännöksistä käy ilmi, että ChatGPT tekee käännöksissään toistuvia virheitä. Yleisimpiä virhetyyppejä ovat suorat sanasta sanaan käännökset. Virheet johtavat epäidiomaattisiin ilmauksiin, jotka vaikeuttavat tekstin ymmärtämistä. Virheitä tapahtuu enemmän käännettäessä englannista suomeksi. Kummankaan kielen käännöksiä ei voisi käyttää ilman ihmisen tekemiä muokkauksia. Käännöksiä voisi kuitenkin hyödyntää alustavan käännösversion luomiseen tai saadakseen yleiskuvan tekstistä.

Asiasanat: ChatGPT, transformer-arkkitehtuuri, konekääntäminen, MQM

SISÄLLYSLUETTELO

1	JOHDANTO	1
2	KONEOPPIMINEN	3
2.1	KONEOPPIMISTYYLIT	3
2.2	NEUROVERKOT	4
3	LUONNOLLISEN KIELEN PROSESSOINTI	6
3.1	TRANSFORMER-ARKKITEHTUURI	7
3.2	KONEKÄÄNTÄMINEN	9
4	CHATGPT	10
4.1	GPT-3.5 ARKKITEHTUURI	10
4.2	GPT-3.5-KIELIMALLIN KOULUTTAMINEN	11
5	MQM-ARVOSTELUMENETELMÄ	12
5.1	KÄÄNNÖSTEN PISTEYTYS	12
5.2	VIRHETYYPIT	14
5.3	VIRHEIDEN VAKAVUUSTASOT JA KERTOIMET	15
6	KÄÄNNÖSTEN ARVIOINTI	16
6.1	TUTKIMUSMENETELMÄ	16
6.2	TUTKIMUSAINESTO	17
6.3	KÄÄNNÖSTEN PISTEYTYS	17
6.4	TUTKIMUSTULOSTEN ANALYYSI	24
7	YHTEENVETO	25
	LÄHTEET	26
	LIITE 1: TIEDEUUTISTEN PISTETAULUKOT	31
	LIITE 2: PAIKALLISUUTISTEN PISTETAULUKOT	36

1 JOHDANTO

Tekoäly ja sen hyödyntäminen päivittäisessä arjessa ovat trendanneet viime vuosien aikana. Yritykset panostavat aikaa ja rahaa tekoälyn kehittämiseen ja sen hyödyntämiseen toiminnassaan. (Matzelle, 2024) Samoin yksityisihmisten kiinnostus aihetta kohtaan on kasvanut merkittävästi. Muun muassa tekoälyn ja OpenAI:n kehittämän ChatGPT:n googlehaut kasvoivat merkittävästi. Tekoälyä googlataan nykypäivänä viisi kertaa enemmän kuin vielä kaksi vuotta sitten. (Google Trends¹)

Tässä työssä tutkitaan, miten sujuvasti tekoälypohjainen chatbotti ChatGPT pystyy kääntämään suomesta englanniksi ja toisin päin. Tutkimusaineisto koostuu uutisartikkeleista, jotka on kerätty Helsingin sanomien ja BCC News -uutistalojen nettisivuilta. Artikkelit voidaan jakaa kahteen aihealueeseen: tiede- ja paikallisuutisartikkelit. Molemmista aihealueista on kerätty viisi uutisartikkelia molemmilla kielillä. Uutisartikkelit on kopioitu nettisivuilta ja syötetty ChatGPT:lle minimaalisin muokkauksin. Tutkimuksen tarkoitus on löytää mahdollisia toistuvia virheitä ja käännöstopoja sekä eroavaisuuksia käännösuuntien välillä.

Tutkimustulokset auttavat ymmärtämään tekoälyn rajoitteita sekä sen tekemiä virheitä. Tulokset myös auttavat käyttäjää ymmärtämään, minkä tasoiseen käännöstyöhön ChatGPT kykenee ja minkä tasoisten tai tyylisten tekstien kääntämiseen ChatGPT:tä voi hyödyntää.

Koska ChatGPT:n käyttäjäkunta on hyvin laaja ja tekoälypohjaisten chattien käyttö yleistymässä, tutkimuksesta voivat hyötyä kaikki ne, jotka hyödyntävät tämän kaltaisia palveluita. Tutkimus auttaa myös arvioimaan, vastaako ChatGPT:n tuottaman käännöksen laatu käännöksen käyttötarkoituksen vaatimuksia.

ChatGPT ei ole erityisesti kääntämistä varten koulutettu. Tämän vuoksi voidaan olettaa, että käännöksistä löytyy virheitä ja virheellisiä kielenilmauksia. Koska ChatGPT:tä koulutettaessa englanninkielistä opetusmateriaalia on ollut enemmän

¹ Hakutermit: 'ChatGPT' ja 'Tekoäly'. Haettu 26.4.2024.

verrattuna suomenkieliseen tekstiaineistoon, myös tekstien kääntösuunnalla on merkitystä. Tutkimuskysymykset ovat:

1. Miten hyvin ChatGPT onnistuu kääntämisessä?
2. Tekeekö ChatGPT toistuvia tai samankaltaisia virheitä?
3. Onko käännösuunnalla merkitystä tai niiden välillä eroja?

Koska ChatGPT on uusi ohjelmisto ja tekoäly tieteenalana voimakkaasti kehittyvä, myös uusia tutkimuksia tarvitaan aiheesta. Suomen kielen asemaa tekoälysovellusten kehittämisessä on tutkittu verrattain vähän. Esimerkiksi Kousa (2023) on tutkinut ChatGPT:tä työkaluna suomenkielisten poliittisten tekstien tiedonhaussa. ChatGPT:n kykyä kääntää suomenkielisiä tekstejä englanniksi tai englanninkielisiä tekstejä suomen kielelle ei ole tutkittu.

Jotta voidaan ymmärtää ChatGPT:n tai muiden tekoälyohjelmistojen toimintaa ja suhtautua niiden tuottamiin tuloksiin kriittisesti, täytyy ymmärtää niiden taustalla olevia tekniikoita ja rajoitteita. Tutkielman alussa käynkin läpi koneoppimisen teoriaa ja neuroverkkojen rakennetta. Sen jälkeen käsittelen tarkemmin yhtä koneoppimisen haaraa: luonnollisen kielen prosessointia. Tässä yhteydessä esittelen transformer-rakenteen, jonka arkkitehtuuriin myös ChatGPT pohjautuu, ja kerron konekääntämisestä yhtenä tekoälyn sovellusalana. Yleisen koneoppimisen teorian jälkeen esittelen ChatGPT:n ja sen mallin rakenteen sekä mallin koulutusmenetelmät. Teorian jälkeen esittelen tutkimuksen menetelmät sekä tutkimusaineiston. Tutkittavista artikkeleista löytyneet virheet jaetaan virhetyyppeihin ja pisteytetään virheen vakavuuden mukaan. Tämän jälkeen käydään läpi tutkimustulokset ja pohditaan niiden syitä ja vaikutuksia.

2 KONEOPPIMINEN

Koneen kykyä käsitellä suuri määrä dataa, oppia siitä ja tuottaa haluttu tulos, kutsutaan koneoppimiseksi. Esimerkiksi kasvojentunnistusohjelmistoa kehitettäessä koneelle syötetään suuria määriä kasvojen kuvia, joista se oppii erilaisia piirteitä ja tunnistaa ihmisen kasvojen perusteella. Koneoppimisen kehitys on tekoälyn kehityksen keskiössä. (Morales & Escalante, 2022)

2.1 Koneoppimistyylit

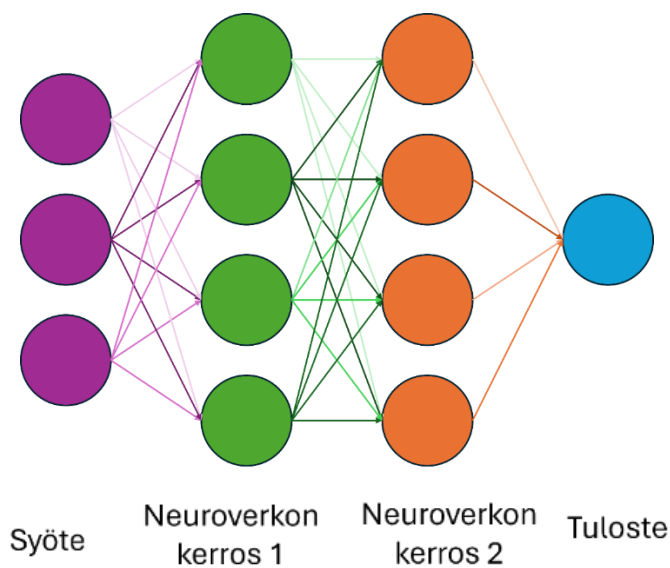
Jotta kone oppii halutun toiminnan, sitä pitää opettaa. Oppimistyylejä on useita, muun muassa ohjattu oppiminen (*supervised learning*), ohjaamaton oppiminen (*unsupervised learning*), vahvistusoppiminen (*reinforced learning*) (Janiesch ym., 2021 ja Morales & Escalante, 2022) sekä puoliohjattu oppiminen (*semi-supervised learning*) (Jurafsky & Martin, 2024).

Ohjatussa oppimisessa koneen saama syöte on ihmisen merkitsemää eli annotoimaa. Kone pyrkii oppimaan valmiiksi merkitystä datasta mallin, jolla haluttu tulos tuotetaan. Eli ohjatussa oppimisessa oikea tulos on koulutusdatassa mukana. Ohjaamattomassa oppimisessa koneelle syötetään suuri määrä dataa, josta kone pyrkii oppimaan piirteitä ilman ihmisen antamaa lisätietoa. (Janiesch ym., 2021) Puoliohjattu oppiminen on ohjatun ja ohjaamattoman oppimisen välimuoto. Myöhemmin kielen prosessoinnin yhteydessä (Kappale 3) esiteltävä suuren kielimallin koulutus on esimerkki puoliohjatusta oppimisesta. Vahvistusoppimisessa puolestaan tekoäly saa syötteen, yrittää tuottaa oikean tuloksen ja saa toimintansa jälkeen palkkion (positiivisen tai negatiivisen). Palaute voidaan siis antaa viiveellä. Toisin sanoen oppiminen perustuu virheistä oppimiseen. Malli yrittää oppia oikean toimintatavan, jolla saavuttaa halutun lopputuloksen. (Morales & Escalante, 2022)

2.2 Neuroverkot

Yksi koneoppimisessa käytetyistä rakenteista on neuroverkko (*artificial neural network*). Neuroverkon rakennetta voidaan ajatella ihmisaivojen kaltaisena. Se muodostuu joukosta yhteydessä olevia kerroksia (aivojen synapsit), jotka välittävät toisilleen painotettuja arvoja (signaaleja). Kerrosten määrä ja rakenne voivat vaihdella. (Janiesch ym., 2021 ja Jurafsky & Martin, 2024)

Kun neuroverkossa on vähintään kaksi kerrosta, puhutaan syväoppimisesta. Syväoppivat neuroverkot oppivat piirteitä (*automated feature learning*), joten ihmisen ei tarvitse määrittellä niitä syötteeseen etukäteen vähentäen ihmistä tarvitsevan työn määrää. Yleisimmät syväoppivat neuroverkkorakenteet ovat RNN-, CNN- sekä transformer-mallit. (Janiesch ym. 2021) Alla olevassa kuvassa (Kuva 1) on esitetty yksinkertaistettu eteenpäin kytketyn neuroverkon rakenne, jossa on kaksi kerrosta. Kuvassa näkyvät myös kerrosten sisäiset ja väliset yhteydet.



Kuva 1: Yksinkertaistettu kuva eteenpäin kytketyn neuroverkon rakenteesta.

Syvä neuroverkon rakenne sopii tilanteisiin, joissa käsiteltävä data on monimutkaista, kuten äänitteet tai videot. Kun syöte on kulkenut neuroverkon kerrosten

läpi, saadaan tuloste, joka voi olla esimerkiksi kuvan tunnistus (*image recognition*) tai arvaus lauseen seuraavasta sanasta (*next word prediction*). Neuroverkkojen kouluttaminen vaatii kuitenkin paljon dataa ja tehokkaan tietokoneen. (Jurafsky & Martin, 2024)

Tämän työn kannalta oleellisin neuroverkkoarkkitehtuuri on transformer-malli. Käsittelem sen seuraavassa kappaleessa kielen prosessoinnin kontekstissa.

3 LUONNOLLISEN KIELEN PROSESSOINTI

Luonnollisen kielen prosessointi (*natural language processing, NLP*) voidaan nähdä yhtenä koneoppimisen osa-alueena, jossa nykyään hyödynnetään syväoppivaa rakennetta. Luonnollisen kielen prosessointi mahdollistaa muun muassa koneiden kielen käytön, joka helpottaa ihmisten ja koneiden välistä kommunikointia. (Jain ym., 2018) Omar ym. (2022) kertoo, miten kielen prosessointia hyödyntävät ohjelmat ovat tällä hetkellä, ja jo pitkään olleet, suuri hitti. Päivittäin käytössämme ovat roskasähköpostin tunnistus, erilaiset konekääntäjät sekä muut älylaitteet kuten Alexa.

Kielen prosessoinnissa kone oppii luonnollisen kielen säännöt ja merkitykset, jolloin se voi analysoida ja tulkita kielidataa halutulla tavalla. (Jain ym., 2018 ja Omar ym., 2022) Toisin sanoen kone luo kielen todennäköisyyksiä edustavan mallin eli kielimallin. (Jurafsky & Martin, 2024) Nykypäivänä suuret kielimallit (*large language model*), kuten GPT-3.5, ovat lähes aina transformer-malleja. Suuret kielimallit luodaan suuresta kielidatasta kouluttamalla malli ensin puoliuhjatulla oppimisella ennustamaan tekstiaineistossa seuraavana tuleva sana. Koska dataa tarvitaan paljon, mallit koulutetaan usein pääosin netistä kerätyllä aineistolla. Aineisto on hyvin monimuotoista ja siihen voi kuulua esimerkiksi käännöstekstejä, kysymys-vastauspareja sekä eri tekstilajeja ja -tyylejä. Alun kouluttamisesta opittua mallia kutsutaan esikoulutetuksi kielimalliksi (*pretrained language model*). Alun kouluttamisen jälkeen mallia hienosäädetään (*fine-tuning*). (Jurafsky & Martin, 2024) Kielimalliarkkitehtuureja on useita, kuten tässä työssä esiteltävä transformer-malli.

Ennen kuin aineistoa voidaan käsitellä mallissa, täytyy se muuttaa muotoon, jota malli ymmärtää. Tätä vaihetta kutsutaan tokenisoinniksi (*tokenization*). Esimerkiksi tekstidata jaetaan sanoihin, tavuihin tai merkkeihin, jotka muutetaan numeeriseen muotoon. Sen jälkeen lisätään myös muita piirteitä, kuten merkitystä ja sijaintia lauseessa, esittäviä tokeneita, jotka mahdollistavat sanojen välisten semanttisten merkitysten säilyttämisen (Janiesch ym., 2021). Tämän jälkeen tokenit muutetaan vektoriarvoiksi (*embedding vectors*). Kun syöte on käsitelty, se voidaan syöttää malliin. (Vaswani ym., 2017)

3.1 Transformer-arkkitehtuuri

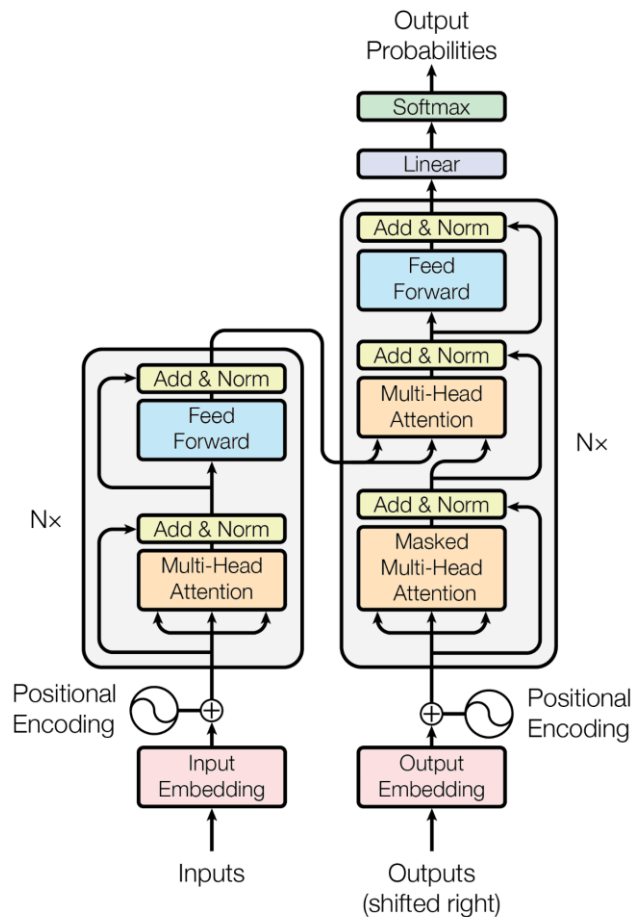
Transformer-mallit ovat arkkitehtuuriltaan eteenpäin kytkettyjä neuroverkkoja. Tämä tarkoittaa sitä, että sen jälkeen, kun syöte on käsitelty yhdessä kerroksessa, se siirtyy seuraavaan kerrokseen palaamatta takaisin aikaisemmille. (Jurafsky & Martin, 2024) Transformer-malli perustuu itsehuomiomekanismiin (*self-attention mechanism*), joka painottaa eri tavoin syötetyn datan eri osia. Transformer-malli soveltuu moniin tehtäviin, muun muassa kielen kääntämiseen, tekstin ymmärtämiseen, yhteenvedojen koostamiseen ja tekstin tuottamiseen (Vaswani ym., 2017).

Transformerit koostuvat enkoodereista (*encoder*) ja dekodereista (*decoder*). Malli voi sisältää molemmat tai vain toisen. Enkooderi-dekooderi-malli (*encoder-decoder*) sopii tehtäviin, joissa tuotetaan sisältöä, esimerkiksi kääntämiseen tai yhteenvedon luomiseen. Mallit, joissa on vain enkooderi, soveltuvat tehtäviin, joissa on tärkeää syötteen ymmärtäminen, kuten luokittelutehtävät. Dekooderi-mallit soveltuvat tuottamista vaativiin tehtäviin. (Cai ym., 2022)

Enkooderi ja dekooderi muodostuvat transformer-kerroksista. Enkooderissa transformer-kerrokset koostuvat itsehuomiokerroksesta (*multi-head attention*) ja eteenpäin kytketystä kerroksesta (*feedforward layer*). Molempia kerroksia seuraa aina normalisointikerros, joka nimensä mukaisesti normalisoi aikaisemmilta kerroksilta saadut vektoriarvot, jotta ne pysyvät rajatun skaalan sisällä. Enkooderin kerrokset käsittelevät koko syötteen kerralla. Syötteen osat etenevät ensimmäiseen kerrokseen ilman muiden osien vaikutusta. Ensimmäisessä itsehuomiokerroksessa vektoriarvoihin lisätään muiden osien painoarvot. Eli enkooderin läpi kulkiessaan syötteen vektoriarvoissa on mukana muiden syötteiden vaikutus. Koska enkooderissa tai dekooderissa syötteen mukana ei kulje paikkatietoa, paikkatieto mallinnetaan erikseen (*positional encoding*). (Vaswani, ym., 2017)

Dekooderin kerrosten rakenne on muuten enkooderin kaltainen, mutta siinä on lisäksi vielä yksi itsehuomiokerros, joka käsittelee myös enkooderin tulosteen. Dekooderi vastaa mallin tuottamasta tuloksesta olkoon se kokonainen käänös tai yksittäinen sana. Toisin kuin enkooderi, dekooderi näkee edeltävän dekodierikerroksen tuottaman tulosteen, esimerkiksi tuotetun sanan, sekä enkooderin tulosteen. (Vaswani, ym., 2017)

Kuvassa 2 on Vaswanin ym. (2017) luoma hahmotelma transformerin rakenteesta, jossa on kuvattuna syötteen kulku mallin läpi sekä eri kerrokset, joissa syötettä käsitellään.



Kuva 2: Transformer-mallin arkkitehtuuri. (Vaswani, ym. 2017)

Itsehuomiokerroksessa malli tarkastelee omaa syötettään. (Jurafsky & Martin, 2024) Se on funktio, joka laskee syötteen painotetut arvot. (Vaswani ym., 2017) Jokaisella syötteellä on kysely-, avain- ja arvovektori (Jurafsky & Martin, 2024). Kyselyvektori kuvastaa sen hetkistä haettavaa tietoa (syötettä). Avainvektorit ovat ns. hakuavaruus, joka määrittää, mitä tietoa huomioidaan. Viimeisenä arvovektori sisältää itse tiedon. Tuloste muodostetaan laskemalla kaikkien kolmen vektorin painotettu summa. Itsehuomiokerroksen laskenta on optimaalista ja sitä voidaan suorittaa samanaikaisesti. Itsehuomiokerroksen avulla saavutetaan se, että tekstin riippuvuuksia

voidaan oppia myös pitkien etäisyyksien päästä. (Vaswani ym., 2017) Eli esimerkiksi tekstissä toisistaan kaukana olevienkin osien väliset riippuvuudet voidaan säilyttää.

3.2 Konekääntäminen

Yksi kielenprosessoinnin tehtävistä ja tekoälyn sovelluskohteista on konekääntäminen. Kääntämistä vaikeuttaa kielten väliset erot kieliopissa, kuten sanajärjestys tai sanojen taivutus, sekä se, että sanoilla ja lauseilla voi olla useita eri merkityksiä. Kaikista kielistä ei myöskään löydy samoja rakenteita tai merkityksiä. (Wang ym., 2022)

Tyypillinen neuroverkkoja hyödyntävä konekääntäjä (*neural machine translation, NMT*) koostuu transformerin enkooderi- ja dekoderosista. Enkooderi muuttaa lähdekielen lauseen vektoriarvoiksi ja dekoderi tuottaa arvoista käännöksen. Transformerin itsehuomiokerroksen avulla käännoستهتävissä voidaan laskea lähtötekstin ja tuotettavan käännöksen suhde tai todennäköisyys. (Wang ym., 2022) Koska itsehuomiomekanismi mahdollistaa pitkienkin riippuvuussuhteiden huomioimisen (Vaswani ym., 2017), sen käyttäminen käännoستهتävissä parantaa koneella luotujen käännösten laatua (Wang ym., 2022).

4 CHATGPT

ChatGPT on OpenAI:n kehittämä chatbotti, joka julkaistiin 30. marraskuuta 2022. Ohjelmasta on tarjolla useita verkkoversioita: tässä tutkimuksessa käytettävä ilmaisversio, joka käyttää GPT-3.5-kielimallia sekä useita maksullisia versioita, jotka käyttävät uusinta kielimallia: GPT-4:ää. ChatGPT on erittäin suosittu työkalu, jonka käyttäjämäärä ohitti miljoonan vain viidessä päivässä (Buchholz, 2023). Sen vuoksi onkin hyvä ymmärtää mallin rajoitteet. OpenAI:n (2022a) ChatGPT:n esittelyartikkelissa mainitaan muun muassa seuraavat haasteet:

- ChatGPT saattaa antaa virheellisen tai järjettömän vastauksen. Vaikka, vastaus olisi virheellinen, se saattaa antaa käyttäjälle todentuntuisen vaikutuksen.
- Käyttäjän syötteen muotoilu voi vaikuttaa annettuun vastaukseen. Kysymystä, jolle malli ei aluksi osaa muodostaa vastausta, muokkaamalla ChatGPT saattaa antaa lopulta oikean vastauksen.
- Koulutusdatan puolueellisuus ja ylioptimoinnin ongelmat johtavat tiettyjen ilmausten toistuvaan käyttöön ja tarpeettoman monisanaisiin vastauksiin.
- Jos käyttäjän kehote on liian monitulkintainen, malli arvaa käyttäjän tarkoituksen ja vastaa arvauksen pohjalta. Sen sijaan tarkempia vastauksia saataisiin, jos malli kysyisi tarkentavia kysymyksiä.
- Mallille annettujen epäasiallisten kysymysten suodattamiseen hyödynnetään Moderation APIa. Joskus malli kuitenkin vastaa epäasiallisiin tai haitallisiin kehoitteisiin.

4.1 GPT-3.5 arkkitehtuuri

Toisin kuin kappaleessa 3.2 esitelty tyypillinen konekääntäjä, GPT-3.5 on pelkästä dekooderista koostuva transformer-malli. Käsiteltävä tekstidata on käyttäjän antama kehote. Kehote tokenisoidaan siten, että yksi tokeni edustaa noin neljää merkkiä. Tokeneille lasketaan vektoriarvot, jotka syötetään dekooderiin. (OpenAI) Ensimmäisellä kierroksella dekooderin syötteenä on käyttäjän kehote. Tämän jälkeen malli tuottaa

vastauksen sana kerrallaan, jolloin seuraavan kierroksen syöte on kehoite sekä edellisellä kierroksella tuotettu vastaus.

4.2 GPT-3.5-kielimallin kouluttaminen

Sekä GPT-3.5- että GPT-4 pohjautuvat GPT-malliin. Nimi tulee sanoista *Generative Pre-Trained Transformer*. Nimensä mukaan GPT on yleinen esikoulutettu kielimalli eli sitä ei ole jatkokoulutettu mihinkään tiettyyn tehtävään. GPT-malli on koulutettu pääosin netistä kerätyllä aineistolla (Jurafsky & Martin, 2024). GPT-mallista on päästy GPT-3.5-malliin jatkokouluttamalla se käyttäen sekä ohjattua oppimista että vahvistusoppimista ihmisen palautteen avulla. GPT-3.5-malli on koulutettu antamaan vastaus kehoitteeseen (*prompt*) tai toimimaan annettujen ohjeiden mukaisesti. (OpenAI, 2022b)

Kuten yllä mainittiin GPT-3.5 pohjautuu GPT-malliin. GPT-kielimallin jatkokouluttaminen ohjatulla oppimisella on koostunut ihmisen luomista esimerkkikeskusteluista. Keskusteluissa ihminen on esittänyt sekä kysyvän käyttäjän että vastaavan chatbotin roolia. Tällä malli on saatu tuottamaan halutunlaisia vastauksia. (Ouyang ym., 2022, OpenAI, 2022a ja OpenAI, 2022b) Ottamalla ihmisen mukaan kielimallin koulutukseen ChatGPT:n toiminnasta ja vastauksista on myös pystytty vähentämään valheellisia ja epäasiallisia vastauksia. (Ouyang ym., 2022)

Seuraavassa vaiheessa koulutusta on jatkettu vahvistusoppimisella. Tarkemmin ihmispalautteeseen pohjaavalla menetelmällä (*Reinforcement Learning from Human Feedback, RLHF*). (OpenAI, 2022a) Vahvistusoppimista varten kielimallille on luotu palkkiomalli (*reward model*). Palkkiomalli on koulutettu tuottamalla kielimallilla samasta kehoitteesta useita vastauksia, jotka ihminen on järjestänyt paremmuusjärjestykseen. Palkkiomallin avulla kielimallin tuottamia vastauksia voidaan järjestää ilman ihmistyötä. (Christiano ym., 2023 ja OpenAI, 2017) Ohjatussa oppimisessa tuotettujen vastausten arviointi on tehty ihmisvoimin, kun taas vahvistusoppimisessa se saadaan palkkiomallilta.

5 MQM-ARVOSTELUMENETELMÄ

MQM (*Multidimensional Quality Metrics*) on menetelmä, jonka avulla voidaan arvioida käännösten laatua. MQM tarjoaa yleisen viitekehyksen käännösten arvioinnille, ja sitä voidaan soveltaa tehtävän vaatimalla tasolla. MQM luotiin vuonna 2014 työkaluksi, jolla pystytään arvioimaan niin ihmisten kuin koneiden tekemiä käännöksiä ja yhtenäistämään niiden arviointia. Menetelmä luotiin osana EU:n rahoittamaa QTLaunchPad-projektia. (Lommel ym., 2014)

Menetelmä sopii hyvin myös tämän tutkimuksen aineiston arviointiin, koska se tarjoaa selkeät kriteerit, joilla käännöksiä arvioidaan. Tämä vähentää myös eroja tekstien välisissä arvioinneissa sekä henkilökohtaisten asenteiden vaikutusta (Lommel ym., 2014).

5.1 Käännösten pisteytys

Käännösten arviointi perustuu pisteytykseen. Menetelmässä virheet jaetaan virhetyypeittäin sekä vakavuustasoittain. Virheen vakavuustaso (*Severity Penalty Level*) määrittää kertoimen (*Severity Penalty Multiplier*), jonka mukaan jokaiselle virhetyypille lasketaan kokonaispistemäärä (*Error Type Penalty Total, ETPT*). Summaamalla jokaisen virhetyypin kokonaispistemäärät saadaan absoluuttinen virheiden pistemäärä (*Absolute Penalty Total, APT*). Kun absoluuttiset pisteet suhteutetaan arvioitavan tekstin sanamäärään (*Evaluation Word Count*) saadaan lopullinen pisteytys. Mitä matalampi lopullinen pistemäärä on, sitä parempi käännös on kyseessä. Käännöksen arvioinnille voitaisiin myös määrittää kynnsarvo (*Threshold Value*). (TheMQM²) Koska tässä tutkielmassa tarkoituksena on arvioida käännösten laatua, eikä hyväksyä käyttöön luotua käännöstä, kynnsarvoa ei tarvitse määrittää. Kuvassa 3 on esimerkki MQM-pistetaulukosta, josta käy ilmi pisteiden lasku.

² TheMQM. *MQM (Multidimensional Quality Metrics) - Concrete Example with Formulas*. Osoitteessa: <https://themqm.org/introduction-to-tqe/concrete-example-with-formulas/>.

MQM Scorecard: Top-Level Error Typology with 4 Severity Levels						
	Error Counts				Error Type Penalty Total	
Severity Penalty Levels:	Neutral	Minor	Major	Critical	ET Weights	ETPTs
Severity Penalty Multipliers:	0	1	5	25		
Error Types						
Terminology	2	3	1	0	1.0	8.0
Accuracy	4	5	2	0	1.0	15.0
Linguistic conventions	1	8	0	0	1.0	8.0
Style	5	3	0	0	1.0	3.0
Locale convention	1	2	0	0	1.0	2.0
Audience appropriateness	0	2	0	0	1.0	2.0
Design and markup	0	6	0	0	1.0	6.0
Custom	0	0	0	0	1.0	0.0
					Absolute Penalty Total (APT):	44.0
					Evaluation Word Count:	10184
					Penalty per 100 words:	0.43
					Overall score:	99.57
					Decision	PASS

Kuva 3: Esimerkki käännöksen pisteytyksestä MQM-metodilla. (TheMQM³)

Kuvan taulukossa virhetyypit on jaettu kahdeksaan: terminologia (*Terminology*), tarkkuus (*Accuracy*), kielelliset konvention (*Linguistic conventions*), tyyli (*style*), alueelliset konventiot (*Locale convention*), kohdeyleisölle sopivuus (*Audience appropriateness*), muotoilu (*Design and markup*) ja muut (*Custom*). Virheen vakavuustasot on jaettu neljään: neutraali, pieni, suuri ja kriittinen. Vakavuustasojen kertoimet ovat vastaavasti 0, 1, 5 ja 25. Esimerkissä on myös sarake virhetyypeille annettavista painokertoimista, jotka ovat tässä esimerkissä jokaisella tyypillä 1. Virhetyyppien määrästä on kertoimien mukaan laskettu kokonaispistemäärät, jotka summaamalla on saatu absoluuttiseksi virheiden pistemääräksi 44. Se on suhteutettu sanamäärään ja laskettu virhearvo per 100 sanaa. Tästä lasketaan lopullinen pistemäärä (*Overall score*) 99,57.

³ TheMQM. *MQM (Multidimensional Quality Metrics) - MQM Metrics and Scoring Models*. Osoitteessa: https://themqm.org/error-types-2/1_scorecards/.

5.2 Virhetyypit

Seuraavaksi määrittelen tarkemmin ne kuusi virhetyyppiä, joita tässä tutkimuksessa käytetään käännöksen arviointiin. Alkuperäisessä MQM-mallissa virhetyyppejä on määritelty yli 100 (Burchardt ym., 2015). Burchardt ym. (2015) ovat jakaneet mallin virhetyypit hierarkkisesti kategorioihin ja ulottuvuuksiin. Hierarkiaa edetessä ylhäältä alas kategorian virheet voidaan jakaa tarkempiin alakategorioihin. Näiden avulla on helpompi jakaa käännösten virheet oikeiden tyyppien alle. Kuitenkin tämän tutkimuksen laajuuden kannalta käännösten virheet luokitellaan suoraan seuraavien yläkategorioiden alle: terminologia, tarkkuus, tyyli, kielellinen sujuvuus (nyk. *Linguistic conventions*, aik. *Fluency*), kohdeyleisölle sopivuus (nyk. *Audience appropriateness*, aik. *Verity*) sekä alueelliset konventiot. Tämän lisäksi virheille on muut-kategoria, jonka alle luokitellaan kaikki sellaiset virheet, jotka eivät selkeästi kuulu edellä mainittuihin virhetyyppeihin. Burchardt ym. (2015) määrittelevät virhetyypit seuraavasti:

Terminologia. Terminologiavirheisiin kuuluu epä johdonmukainen termien käyttö. Esimerkiksi useat käännökset samalle termille, aihealueelle vieraiden termien käyttö tai väärän termin käyttö. Kategoriaan eivät kuulu yleiskielen käännösvirheet.

Tarkkuus. Tarkkuusvirheet käsittävät virheet tarkoituksen muutoksissa lähtö- ja kohdetekstin välillä. Tähän kategoriaan kuuluvat käänöksessä tapahtuvat lisäykset ja poistot, suorat ja kirjaimelliset käännökset sekä muut merkitystä muuttavat virheet.

Kielelliset konventiot. Kielellisten konventioiden arviointi keskittyy suoraan kohdetekstiin. Kategoriassa arvioidaan, onko teksti koherenttia ja kieliopillisesti oikeaa, seuraavatko välimerkit ja kirjoitusasu kohdekielen kielioppisääntöjä.

Tyyli. Käännöksen tyyliä arvioidaan tekstin muodollisuutta verrattuna lähtötekstiin sekä koheesiota läpi tekstin. Tyyli-kategoriaan luetaan myös epäidiomaattiset ilmaukset sekä tekstin rekisterinvaihdokset.

Alueelliset konventiot. Alueellisten konventioiden virheet rikkovat kohdekielellä käytössä olevia formaatteja. Esimerkiksi päivämäärien, osoitetietojen tai mittayksiköiden käännökset voivat olla oikeita, mutta ne eivät noudata kohdekielen muotoa, jolloin virhe luokitellaan alueellisten konventioiden virhekategoriaan.

Kohdeyleisölle sopivuus. Kohdeyleisölle sopimattomat ilmaukset voivat olla esimerkiksi kulttuuri- tai aluesidonnaiset ilmaukset. Tämän kaltaiset virheet muodostuvat kohdeyleisölle vieraiden tai virheellisten käännösten käytöstä.

Muut. Muut-kategoriaan kuuluvat vain ne virheet, joita ei voida sijoittaa edellä listattuihin kategorioihin. Virhe merkitään aina jonkin tarkemman kategorian alle, jos se vain on mahdollista. Muut-kategorian käytön pitäisi olla hyvin harvinaista.

5.3 Virheiden vakavuustasot ja kertoimet

Virheille on vielä määritelty vakavuustasot ja kertoimet. MQM-menetelmässä (TheMQM⁴) on annettu jaottelukriteerit, joiden mukaan käännöksen arvostelija voi jakaa virheet kriittisyyden perusteella. Seuraavaksi esittelen menetelmän mukaiset kriteerit.

Neutraali. Neutraaleja ovat virheet, joissa teksti olisi parempi toisella tavalla käännettynä, mutta virheen vakavuus ei vaadi virhepistettä. Tästä syystä neutraalien virheiden kerroin on 0, eikä siitä koidu sanktioita käännöksen arvioinnissa.

Pieni. Virheellä on pieni vaikutus käännöksen sisällön oikeellisuuteen, mutta se ei muuta käännöksen luotettavuutta tai käytettävyyttä. Pienen virheen kerroin on 1, koska sillä ei ole suurta merkitystä käännöksen kannalta.

Suuri. Virhe on selvästi huomattavissa ja se vaikuttaa käännöksen ymmärrettävyyteen, luotettavuuteen ja käytettävyyteen. Kerroin on tällä vakavuustasolla 5. Suuret virheet vaikuttavat jo selvästi käännöksen laatuun.

Kriittinen. Virhe on niin suuri, että käännöstä ei voida käyttää, koska se voi aiheuttaa mm. fyysistä, rahallista tai maineellista haittaa. Kriittisten virheiden kerroin on suuri, 25, koska tämän tasoiset virheet tekevät käännöksestä käyttökelvottoman.

⁴ TheMQM. *MQM (Multidimensional Quality Metrics) - Values and Scores*. Osoitteessa: https://themqm.org/error-types-2/1_scorecards/values-and-scores/.

6 KÄÄNNÖSTEN ARVIOINTI

Tässä kappaleessa käydään läpi käännösten arviointiin käytetyt menetelmät, tutkimustulokset sekä niiden merkitys. Tutkimuksen käännösartikkelit löytyvät lähdeluettelosta ja artikkelikohtaiset pistetaulukot liitteistä (Liitteet 1–2).

6.1 Tutkimusmenetelmä

Tutkimukseen käytettävät artikkelit on kerätty suoraan uutissivustoilta. Koska ChatGPT:lle ei voida syöttää artikkeleissa käytettyjä kuvia, kuvatekstit on siirretty hakasulkeiden sisään erottamaan ne muusta leipätekstistä. Ennen varsinaisten tutkimusartikkelien syöttämistä vertailtiin erilaisten kehoitteiden ja eri tarkkuudella annettujen ohjeiden vaikutusta käännöksiin. Eri kehoitteiden välillä ei vaikuttanut olevan suuria eroja, joten kehoitteena käytetään ”*Translate the text below to English.*” tai ”*Translate the text below to Finnish.*” kehoitteen yksinkertaisuuden vuoksi. Artikkelit syötetään ChatGPT:lle yksitellen.

Kun ChatGPT on tuottanut käännöksen, se kopioidaan sellaisenaan tekstitiedostoon. Käännökset pisteytetään käyttäen edellisessä kappaleessa esiteltyä MQM-menetelmää.

Huomioitavaa tutkimuksessa on, että jokainen artikkeli syötetään ChatGPT:lle vain kerran, minkä jälkeen käännöksestä ei anneta palautetta, eikä käännökseen pyydetä parannuksia. Kaikki artikkelit syötetään saman keskustelun aikana. Kehotteessa ei anneta teksteistä mitään lisätietoa, esimerkiksi kohdeyleisöä, tekstilajia tai muotoilutietoa. Vaikka pisteytysmenetelmässä sekä virhetyyppien että vakavuustasojen kriteerit ovat tarkasti määriteltyjä, arvostelijan subjektiivinen mielipide vaikuttaa lopputuloksiin. Mielipiteiden vaikutusta tasapainottaa se, että kaikki artikkelit ovat saman henkilön arvostelema.

6.2 Tutkimusaineisto

Tutkimusaineisto on kerätty BBC Newsin ja Helsingin sanomien nettisivuilta. Tutkittaviksi artikkeleiksi valittiin 20 aineiston keräyshetkellä tuoreita uutista. Aineistoon kuuluu viisi tiedeaihealueen artikkeleita molemmilla kielillä sekä viisi paikallisuutisartikkelia molemmilla kielillä.

Tutkimukseen valittiin nämä kaksi aihealuetta, jotta saadaan testattua käännösten laatua laaja-alaisesti. Paikallisuutisissa voidaan tutkia kotoutetun sanaston ja paikkasidonnaisten käännösten laatua. Tiedeteksteissä taas käsitellään enemmän maailmanlaajuisia uutisia, jotka eivät vaikuta vain paikallisesti. Niissä nousevat esiin myös oikean termistön käyttö.

Aineistoa valittaessa otettiin huomioon myös se, että GPT-3.5:n koulutus saatiin päätökseen vuoden 2022 alussa (OpenAI, 2022a). Artikkelit valittiin tuoreimmista uutisista, jotta voidaan olla varmoja, että ne eivät ole olleet osana mallin koulutusdataa.

6.3 Käännösten pisteytys

Arvioin ensin käännösten yhteispisteitä käännössuuntakohtaisesti. Sen jälkeen verrataan vielä, minkälaisia eroja voidaan löytää verratessa käännöksiä aihealueittain.

Suomesta englanniksi käännettyissä uutisartikkeleissa on yhteensä 172 virhettä. Eniten käännöksistä löytyy pieniä kielellisiä konventioita rikkovia virheitä (52 kpl), toiseksi eniten pieniä tarkkuusvirheitä (38 kpl) ja kolmanneksi neutraaleja kielellisten konventioiden virheitä (24 kpl). Vakavuustasoltaan pieniä virheitä löytyy eniten (112 kpl). Vain yhdestä suomesta englanniksi käännetystä artikkelista löytyy kriittinen virhe, joka on tyylivirhe. Utinen, jossa virhe on, on lapsille tarkoitettu artikkeli, jossa vastataan lasten esittämiin kysymyksiin. Lähtötekstissä asiat ovat yksinkertaisesti selitetyt, mutta käännöksessä terminologia on haastavaa, eikä teksti ole ymmärrettävää lapselle. Taulukossa 1 on näkyvissä kaikkien suomesta englantiin käännettyjen artikkelien arvioinnit yhteenlaskettuna.

Taulukko 1: Yhteispisteitys suomesta englanniksi käännetuille artikkeleille.

Yhteispisteitys suomesta englanniksi käännetuille artikkeleille						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin virheet:	Virhetyypin pisteet:
Vakavuustason kerroin:	0	1	5	25		
Virhetyypit:						
Terminologia	1	16	7	0	24	51
Tarkkuus	8	38	14	0	60	108
Kielelliset konventiot	24	52	4	0	80	72
Tyyli	0	1	0	1	2	26
Alueelliset konvention	1	1	0	0	2	1
Kohdeyleisölle sopivuus	0	4	0	0	4	4
Muut	0	0	0	0	0	0
Vakavuustason virheet:	34	112	25	1		

Kaikki virheet yhteensä:	172
Absoluuttinen virhepistemäärä:	262
Aineiston sanamäärä:	5986
Virheet suhteutettuna sanamäärään:	0,044

Kun virhemäärä suhteutetaan aineiston kokonaissanamäärään, saadaan virhearvoksi 0,044.

Englannista suomeksi käännettyissä artikkeleissa virheitä on 187 kpl. Tässä käännoissuunnassa eniten löytyy pieniä tarkkuusvirheitä (41 kpl), sitten pieniä kielellisten konventioiden virheitä (36 kpl) ja kolmantena suuria tarkkuusvirheitä (30 kpl). Suomenkielisistä käännoksistä löytyy 2 kriittistä virhettä, jotka muuttavat tekstin merkitystä niin paljon, että käännoksen käyttäminen aiheuttaisi mainehaittaa. Virhe on artikkelin otsikossa, ”*Adviser warns London a 'no-go zone for Jews every weekend'*” (Liite 2), joka on käännetty ”Neuvonantaja varoittaa, että Lontoo on ”juutalaisten viikonloppuina kielletty alue””. Tämä käännosvirhe muuttaa käännoartikkelin otsikon

merkityksen alkuperäistä päinvastaiseksi. Sama virhe toistuu myöhemmin uutisen leipätekstissä. Suomeksi käännettyistä artikkeleista löytyy myös yksi muut-kategorian virhe. Käännöksessä on ohitettu kuvateksti, mutta sen muotoiluun käytetyt hakasulkeet on otettu seuraavan kappaleen alkuun. Taulukossa 2 on englannista suomeksi käännettyjen artikkelien yhteispisteitys.

Taulukko 2: Yhteispisteitys englannista suomeksi käännettyille artikkeleille.

Yhteispisteitys englannista suomeksi käännettyille artikkeleille						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin virheet:	Virhetyypin pisteet:
Vakavuustason kerroin:	0	1	5	25		
Virhetyyppi:						
Terminologia	7	17	3	0	27	32
Tarkkuus	13	41	30	2	86	241
Kielelliset konventiot	10	36	10	0	56	86
Tyyli	6	6	1	0	13	11
Alueelliset konvention	0	0	0	0	0	0
Kohdeyleisölle sopivuus	0	4	0	0	4	4
Muut	1	0	0	0	1	0
Vakavuustason virheet:	37	104	44	2		

Kaikki virheet yhteensä:	187
Absoluuttinen virhepistemäärä:	374
Aineiston sanamäärä:	5068
Virheet suhteutettuna sanamäärään:	0,074

Englannista suomeksi käännettyjen artikkelien virhearvo on 0,074, mikä on huomattavasti suurempi kuin suomesta englanniksi käännettyillä (0,044). Kuitenkin virheiden kappalemäärä on lähellä toisiaan molemmissa käännösuunnissa.

Myös uutisartikkelien aihealueiden pisteissä on eroja. Suurin suhteellinen virhearvo on englannista suomeksi käännettyillä tiedeuutisilla. Englannista suomeksi käännettyillä paikallisuutisilla ja suomesta englanniksi käännettyillä tiedeuutisilla suhteutetut virhearvot ovat samat: 0,066. Alla olevista taulukoista 3–6 käy ilmi artikkelien pisteet aihealueittain. Taulukoiden sarakkeissa on muun muassa sana- sekä virhemäärät, joista voidaan huomata, että virheiden määrä korreloi tekstin pituuden kanssa. Sama ei kuitenkaan päde sanamäärään ja suhteutettuihin virheisiin. Tämä johtuu virheiden vakavuuksista ja niiden laskentakertoimista.

Ensimmäisenä alla olevien taulukoiden 3 ja 4 avulla voidaan verrata paikallisuutisten pisteitä.

Taulukko 3: Englannista suomeksi käännettyjen paikallisuutisten yhteispisteitys.

Paikallisuutiset englannista suomeksi				
Paikallisuutiset englannista suomeksi:	Sanamäärä:	Virheitä yhteensä:	Absoluuttinen virhepistemäärä:	Suhteutetut virheet:
Nottingham attacks: Sons upset at not being told of vigil	868	23	36	0,041
£40k orphanage donor feels 'cheated' by charity	719	28	17	0,024
Patients facing uphill struggle to see GPs - report	297	14	12	0,040
Adviser warns London a 'no-go zone for Jews every weekend'	225	13	72	0,320
Theresa May to stand down as MP at next election	74	2	6	0,081
Yhteensä:	2183	80	143	
Virheet suhteutettuna:	0,066			

Taulukko 4: Suomesta englanniksi käännettyjen paikallisuutisten yhteispisteitys.

Paikallisuutiset suomesta englanniksi				
Paikallisuutiset suomesta englanniksi:	Sanamäärä	Virheitä yhteensä	Absoluuttinen virhepistemäärä	Suhteutetut virheet
Arvotalo Helsingin Punavuorella rakensi tuhansia euroja tuottavan pien-	1114	32	31	0,028
Eturivin poliitikot siirretään sivuun historiallisissa päätöksissä	1075	28	37	0,034
Jäteauton kuljettaja: Rotat ovat monilla Helsingin seudun alueilla erityisen paha	783	10	16	0,020
Keravalle jumppakepit hankkinut toimialajohtaja siirtyy toisiin tehtäviin	308	11	15	0,049
Polisi varoittaa: Liikkeellä on taas uusi huijaus	231	6	10	0,043
Yhteensä:	3511	87	109	
Virheet suhteutettuna:	0,031			

Koska paikallisuutisissa teemat ovat enemmän paikkasidonnaisia, korostuvat näissä artikkeleissa erityisesti sellaisten sanojen ja lauseiden käännökset, joille ei löydy kohdekieleltä vastinetta. Tämä johtaa usein väärin termien käyttöön, josta taas seuraa merkitysten muutoksia tai epäselvyyksiä merkityksessä. Samanlaisia virheitä aiheuttavat myös sanontojen ja idiotismien käännökset.

Esimerkiksi Pertun (2024) Punavuorella olevaa arvotaloa käsittelevän artikkelin käännöksessä ”arvotalo” on käännetty kirjaimellisesti ”*valuable building*”, vaikka suomenkielinen termi ei välttämättä viittaa suoraan talon rahalliseen arvoon. Samassa artikkelissa puhutaan myös taloyhtiöstä, jolle ei ole vastinetta englannin kielessä. ”Taloyhtiö” onkin käännetty suoralla käännöksellä ”*housing company*”, josta lukija voi saada väärän käsityksen taloyhtiön merkityksestä. Taas Trigglen (2024) englannista suomeen käännettyssä artikkelissa lyhenne ”*NHS*” on jäänyt kokonaan käännättämättä. NHS on lyhenne sanoista *National Health System*, eli se tarkoittaa julkista terveydenhuoltoa.

Toisena verrataan tiedeuutisten pisteitä. Seuraavan sivun taulukoissa 5 ja 6 ovat tiedeuutisten koostetut pisteet.

Taulukko 5: Englannista suomeksi käännettyjen tiedeuutisten yhteispisteytys.

Tiedeuutiset englannista suomeksi				
Tiedeuutiset englannista suomeksi:	Sanamäärä:	Virheitä yhteensä:	Absoluuttinen virhepistemäärä:	Suhteutetut virheet:
Robot ships: Huge remote controlled vessels are setting sail	901	28	51	0,057
More climate records fall in world's warmest February	622	26	77	0,124
Mass die-offs among farmed salmon on the rise around the world	533	13	17	0,032
Michelle Donelan told to pay damages to academic over Hamas claim	443	26	58	0,131
Why fat Labradors can blame their genes	386	14	28	0,073
Yhteensä:	2885	107	231	
Virheet suhteutettuna:	0,080			

Taulukko 6: Suomesta englanniksi käännettyjen tiedeuutisten yhteispisteytys.

Tiedeuutiset suomesta englanniksi				
Tiedeuutiset suomesta englanniksi:	Sanamäärä:	Virheitä yhteensä:	Absoluuttinen virhepistemäärä:	Suhteutetut virheet:
Hybridilentokone voi nousta jopa jalkapallokenttää pienemmältä aukiolta	964	24	34	0,035
Salaatista ja tomaatista luvataan entistä terveellisempiä, jos	615	12	22	0,036
Korallit haalistuvat Isolla valliriutalla	347	10	18	0,052
Miksi Suomessa pitää opiskella suomen kieltä, vaikka osaisi hyvin suomea?	308	38	84	0,273
Egyptistä löytyi valtava faarao Ramses Suuren patsas	241	7	5	0,021
Yhteensä:	2475	91	163	
Virheet suhteutettuna:	0,066			

Tiedeuutisten välinen piste-ero on pienempi kuin paikallisuutisilla. Tiedeuutisissa aiheet eivät sijoitu lehden kotimaahan, vaan niissä käsitellään myös globaaleja aiheita, joten on todennäköisempää, että termille löytyy vastine kohdekielellä. Artikkeleissa käsitelläänkin osittain samoja aihealueita, kuten ilmastonmuutosta sekä teknologian kehitystä. Tiedeuutisissa onkin vähemmän terminologiatyyppin virheitä verrattuna paikallisuutisiin. Näissä lukemista vaikeuttavat suorat käännökset, kuten ”korallien lämpöstressi Ison valliriutan etelä- ja keskialueella on ollut korkeinta ja riutan pohjoisosissa toiseksi korkeinta tilastohistorian aikana” on käännetty kömpelösti suoraan ”*coral heat stress in the southern and central areas of the Great Barrier Reef has been the highest, and second highest in the northern parts of the reef in recorded history*”.

6.4 Tutkimustulosten analyysi

Edellisessä kappaleessa esitettyjen pisteiden pohjalta voidaan arvioida ChatGPT:n kykyä kääntää. Suomen kielelle käännettyissä artikkeleissa suhteutettu virhearvo on korkeampi. Tähän voi vaikuttaa englanninkielisen aineiston suurempi edustus mallin koulutusvaiheessa. Myös artikkelien sanamäärillä on vaikutusta tuloksiin, koska käännosten lopulliset pisteet suhteutetaan sanamäärään. Virhemäärät eroavat käännoissuuntien välillä, mutta kummassakaan tapauksessa käännoksiä ei voisi käyttää uutisartikkeleina. Käännoksissä on paljon suoria sanasta sanaan käännoksiä, joista koituu epäidiomaattisia ja kieliopin vastaisia ilmauksia. Vaikka ilmaukset eivät suoraan olisi virheellisiä, ne tekevät teksteistä vaikealukuisia ja kankeita.

Kummassakaan käännoissuunnassa ei myöskään ole muutettu kohdekielen tekstilajille tai -tyylille ominaisia piirteitä. Esimerkiksi englannista suomeen käännettyissä teksteissä käytetään ajatusviivoja ja sinä-passiivia, jotka tulevat suoraan englannin kielen käytännöistä. Englanniksi käännettyissä artikkeleissa taas muun muassa otsikoissa käytettiin artikkeleita, joita ei yleensä uutisartikkeleissa ole.

Vaikka ChatGPT:n tuottamia käännoksiä ei voi käyttää suoraan ilman ihmisen tarkistusta tai muutoksia, ChatGPT:tä voi käyttää aputyökaluna esimerkiksi käännoksen ensiversion luonnissa. ChatGPT:n käännoksestä saa hyvän pohjan, jota voi jatkokehittää. Tämä vähentää käännoksen alussa tarvittavan työn määrää. Toisaalta ChatGPT:tä voi käyttää yksittäisten termien tai lauseiden kääntämisessä. Käännookset ovat myös tarpeeksi tarkkoja antamaan yleiskuvan tekstistä. Tämä auttaa esimerkiksi ymmärtämään tekstiä, joka on kirjoitettu kielellä, jota lukija ei osaa. Sanatarkalla tasolla artikkeleihin ei kuitenkaan voisi luottaa. Joka tapauksessa ChatGPT:llä tuotettuihin käännoksiin kannattaa suhtautua kriittisesti ja muistaa ohjelman tunnetut heikkoudet.

7 YHTEENVETO

Tämän työn tarkoitus on ollut arvioida ChatGPT:llä tuotettujen käännösten laatua ja muodostaa käsitys siitä, miten käyttökelpoisia tekoälyn tuottamat käännökset ovat. Ensimmäinen tutkimuskysymys oli: Miten hyvin ChatGPT onnistuu kääntämisessä? Kuten käännösten arviointia käsittelevistä kappaleista (kappaleet 6.3 ja 6.4) käy ilmi, ChatGPT:n käännöksissä on paljon virheitä. Useimmat virheet ovat vakavuustasoltaan pieniä, jolloin yksittäinen virhe ei haittaisi tekstin käytettävyyttä. Useimmissa artikkeleissa virheitä on kuitenkin niin monta, että ne vaikuttavat tekstin ymmärrettävyyteen.

Toinen tutkimuskysymys oli: Tekeekö ChatGPT toistuvia tai samankaltaisia virheitä? Käännösten pisteytyksistä käy ilmi, että kielellisten konventioiden, terminologian ja tarkkuuden virheet ovat yleisiä kaikissa artikkeleissa sekä molemmissa käännössuunnissa. Virheiden tarkempi jaottelu ja analyysi voisi paljastaa lisää virheiden laadusta.

Kolmanneksi kysyttiin, onko käännössuunnalla merkitystä tai niiden välillä eroja. Tutkimuksessa englannista suomeksi käännettyissä uutisartikkeleissa virheitä on enemmän. Tämä ei kuitenkaan tarkoita, että englanninkieliset käännökset olisivat hyviä, vaan molemmissa käännössuunnissa tuotetut käännökset ovat lähtötekstin tasoon verrattuna käyttökelvottomia ilman lisämuokkauksia sekä ihmisen läpikäyntiä.

Toisaalta, kuten edellisessä kappaleessa mainittiin, ChatGPT:n luomille käännöksille on omat käyttökohteensa, vaikka ne eivät virheettömiä olekaan. ChatGPT soveltuu esimerkiksi informaalin tekstin kääntämiseen, jossa virheet eivät ole niin haitallisia. Esimerkiksi kaukaiselle Amerikan-serkulle lähetettävän kirjeen käännös voitaisiin tehdä ChatGPT:llä. ChatGPT:n eduksi käännöstyökaluna on myös se, että se on saatavilla tavallisille ihmisille, ja sen käyttö verrattain helppoa.

Tämä tutkimus antaa hyvin yleisen tason käsityksen ChatGPT:llä luotujen käännösten laadusta. Tekoäly on alana ripeästi kehittyvä ja parempia malleja on jo kehitetty. Aiheesta tarvitaankin lisää tutkimuksia niin konekääntämisen kuin tekoälypohjaisten chatbottien osalta.

LÄHTEET

- Buchholz, K. (2023). *Threads Shoots Past One Million User Mark at Lightning Speed*. Statista. Osoitteessa: <https://www.statista.com/chart/29174/time-to-one-million-users/>.
- Burchardt, A., Görög, Lommel, Melby, Uszkoreit. (2015). *Multidimensional Quality Metrics Issue Types*. German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad. Osoitteessa: <https://web.archive.org/web/20211216145215/https://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>.
- Cai, P. X., Fan, Y. C., & Leu, F. Y. (2022). *Compare Encoder-Decoder, Encoder-Only, and Decoder-Only Architectures for Text Generation on Low-Resource Datasets*. BWCCA 2021, LNNS 346, 216–225.
- Christiano, P., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2023). *Deep reinforcement learning from human preferences*. (arXiv:1706.03741). arXiv. Osoitteessa: <http://arxiv.org/abs/1706.03741>.
- Google Trends. Osoitteessa: <https://trends.google.com/trends?geo=CH&hl=fi>. Viitattu: 26.4.2024
- Jain, A., Kulkarni, G., Shah, V. (2018). *Natural Language Processing*. International Journal of Computer Sciences and Engineering, 6(1), 161-167. Osoitteessa: <https://doi.org/10.26438/ijcse/v6i1.161167>.
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). *Machine learning and deep learning*. Electronic Markets, 31(3), 685–695. Osoitteessa: <https://doi.org/10.1007/s12525-021-00475-2>.
- Jurafsky, D., Martin. (2024). *Speech and Language Processing - An Introduction to Natural Language processing, Computational Linguistics, and Speech Recognition*. 3rd ed. draft. Osoitteessa: https://web.stanford.edu/~jurafsky/slp3/ed3bookfeb3_2024.pdf.
- Kousa, I. (2023). *Introducing ChatGPT to a researcher's toolkit: An empirical comparison between rule-based and large language model approach in the*

context of qualitative content analysis of political texts in Finnish. Proceedings of the Joint 3rd NLP4DH and 8th IWCLUL. (s. 102-113).

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). *Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics*. *Tradumàtica Technologies de La Traducció*, 12, 455–463. Osoitteessa: <https://doi.org/10.5565/rev/tradumatica.77>.

Matzelle, E. (29.2.2024). *Top Artificial Intelligence Statistics and Facts for 2024*. Comptia Community. Osoitteessa: <https://connect.comptia.org/blog/artificial-intelligence-statistics-facts#:~:text=Similarweb%20reports%20the%20global%20AI,rate%20of%2019%25%20from%202023>.

Morales, E., Escalante, H. s. 111-129. (2022). *Chapter 6 - A brief introduction to supervised, unsupervised, and reinforcement learning*. *Biosignal Processing and Classification Using Computational Learning and Intelligence*, Academic Press. Osoitteessa: <https://doi.org/10.1016/B978-0-12-820125-1.00017-8>.

Omar, M., Choi, S., Nyang, D., & Mohaisen, D. (2022). *Robust Natural Language Processing: Recent Advances, Challenges, and Future Directions*. *IEEE Access*, 10, 86038–86056. <https://doi.org/10.1109/ACCESS.2022.3197769>.

OpenAI. (2017) (13.6.2017). *Learning from human preferences*. Osoitteessa: <https://openai.com/research/learning-from-human-preferences>.

OpenAI. (2022a) (30.11.2022). *Introducing ChatGPT*. Osoitteessa: <https://openai.com/blog/chatgpt>.

OpenAI. (2022b) (27.1.2022). *Aligning language models to follow instructions*. Osoitteessa: <https://openai.com/research/instruction-following>.

OpenAI. *Documentation* – *Introduction*. Osoitteessa: <https://platform.openai.com/docs/introduction>.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022).

Training language models to follow instructions with human feedback (arXiv:2203.02155). arXiv. Osoitteessa: <http://arxiv.org/abs/2203.02155>.

TheMQM. *MQM (Multidimensional Quality Metrics) - What Is MQM?*. Osoitteessa: <https://themqm.org/>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. Osoitteessa: <http://arxiv.org/abs/1706.03762>.

Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). *Progress in Machine Translation*. *Engineering*, 18, 143–153. Osoitteessa: <https://doi.org/10.1016/j.eng.2021.03.023>.

TUTKIMUSAINEISTON LÄHTEET

Tiedeuutisartikkelit

Amos, J., Morelle, R. & Francis, A. (7.3.2024). *Robot ships: Huge remote controlled vessels are setting sail*. BBC News. Osoitteessa: <https://www.bbc.com/news/science-environment-68486462>. Viitattu: 8.3.2024.

Briggs, H. (7.3.2024). *Why fat Labradors can blame their genes*. BBC News. Osoitteessa: <https://www.bbc.com/news/science-environment-68492504>. Viitattu: 8.3.2024.

Francis, S. (6.3.2024). *Michelle Donelan told to pay damages to academic over Hamas claim*. BBC News. Osoitteessa: <https://www.bbc.com/news/uk-politics-68486206>. Viitattu: 8.3.2024.

Mäkelä, A. (8.3.2024). *Korallit haalistuvat Isolla valli-riutalla*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/tiede/art-2000010279778.html>. Viitattu: 8.3.2024.

McGrath, M. (8.3.2024). *Mass die-offs among farmed salmon on the rise around the world*. BBC News. Osoitteessa: <https://www.bbc.com/news/science-environment-68502267>. Viitattu: 8.3.2024.

- Poynting, M. (7.3.2024). *More climate records fall in world's warmest February*. BBC News. Osoitteessa: <https://www.bbc.com/news/science-environment-68428348>. Viitattu: 8.3.2024.
- Puttonen, M. (3.3.2024). *Salaatista ja tomaatista luvataan entistä terveellisempiä, jos geenisakset sallitaan EU:ssa*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/tiede/art-2000010254350.html>. Viitattu: 8.3.2024.
- Rajatorpan koulun Maailman ihmeet -valinnaisaineen oppilaat, Vantaa. (8.3.2024). *Miksi Suomessa pitää opiskella suomen kieltä, vaikka osaisi hyvin suomea?*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/tiede/art-2000010213627.html>. Viitattu: 8.3.2024.
- Rantanen, K. (7.3.2024). *Hybridilentokone voi nousta jopa jalkapallokenttää pienemmältä aukiolta*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/tiede/art-2000010260333.html>. Viitattu: 8.3.2024.
- Soininen, V. (2024a) (4.3.2024). *Egyptistä löytyi valtava faarao Ramses Suuren patsas*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/tiede/art-2000010271622.html>. Viitattu: 8.3.2024.

Paikallisuutisartikkelit

- Baker, G. & Seddon, P. (8.3.2024). *Theresa May to stand down as MP at next election*. BBC News. Osoitteessa: <https://www.bbc.com/news/uk-politics-68509239>. Viitattu: 8.3.2024.
- Bartram, A. (8.3.2024). *Nottingham attacks: Sons upset at not being told of vigil*. BBC News. Osoitteessa: <https://www.bbc.com/news/uk-england-nottinghamshire-68481707>. Viitattu: 8.3.2024.
- Kalin, O. (7.3.2024). *Jäteauton kuljettaja: Rotat ovat monilla Helsingin seudun alueilla erityisen paha ongelma*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/kaupunki/art-2000010278366.html>. Viitattu: 8.3.2024.

- Khatun Dewan, Y. (8.3.2024). *£40k orphanage donor feels 'cheated' by charity*. BBC News. Osoitteessa: <https://www.bbc.com/news/uk-england-birmingham-68493379>. Viitattu: 8.3.2024.
- Malmberg, R. (7.3.2024). *Poliisi varoittaa: Liikkeellä on taas uusi huijaus*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/kaupunki/art-2000010278199.html>. Viitattu: 8.3.2024.
- Nissinen, E. (7.3.2024). *Eturivin poliitikot siirretään sivuun historiallisissa päätöksissä*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/kaupunki/art-2000010274846.html>. Viitattu: 8.3.2024.
- Perttu, J. (7.3.2024). *Arvotalo Helsingin Puna-vuorella rakensi tuhansia euroja tuottavan pien-voimalan*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/kaupunki/art-2000010128323.html>. Viitattu: 8.3.2024.
- Soininen, V. (2024b) (7.3.2024). *Keravalle jumppakepit hankkinut toimialajohtaja siirtyy toisiin tehtäviin*. Helsingin sanomat. Osoitteessa: <https://www.hs.fi/kaupunki/art-2000010279349.html>. Viitattu: 8.3.2024.
- Triggle, N. (8.3.2024). *Patients facing uphill struggle to see GPs – report*. BBC News. Osoitteessa: <https://www.bbc.com/news/health-68506063>. Viitattu: 8.3.2024.
- Whannel, K. (8.3.2024). *Adviser warns London a 'no-go zone for Jews every weekend'*. BBC News. Osoitteessa: <https://www.bbc.com/news/uk-politics-68508351>. Viitattu: 8.3.2024.

LIITE 1: TIEDEUUTISTEN PISTETAULUKOT

Robot ships: Huge remote controlled vessels are setting sail (Amos ym., 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		2	1		3	7
Tarkkuus	2	7	4		13	27
Kielelliset konventiot	1	3	2		6	13
Tyyli	1	3			4	3
Alueelliset konvention Kohdeyleisölle sopivuus		1			0	0
Muut	1				1	0
	Virhemäärä yhteensä:	28			APT:	51
					Aineiston sanamäärä:	901
					Virheet suhteutettuna sanamäärään:	0,057

Why fat Labradors can blame their genes (Briggs, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	1				1	0
Tarkkuus		2	2		4	12
Kielelliset konventiot		6	2		8	16
Tyyli	1				1	0
Alueelliset konvention Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	14			APT:	28
					Aineiston sanamäärä:	386
					Virheet suhteutettuna sanamäärään:	0,073

Michelle Donelan told to pay damages to academic over Hamas claim (Francis, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	3	1	1		5	6
Tarkkuus	1	7	6		14	37
Kielelliset konventiot		3	2		5	13
Tyyli		1			1	1
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus		1			1	1
Muut					0	0
	Virhemäärä yhteensä:	26				
					APT:	58
					Aineiston sanamäärä:	443
					Virheet suhteutettuna sanamäärään:	0,131

Korallit haalistuvat Isolla valli-riutalla (Mäkelä, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		1			1	1
Tarkkuus		3	1		4	8
Kielelliset konventiot		4	1		5	9
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	10				
					APT:	18
					Aineiston sanamäärä:	347
					Virheet suhteutettuna sanamäärään:	0,052

Mass die-offs among farmed salmon on the rise around the world (McGrath, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	1	1			2	1
Tarkkuus	2	3	2		7	13
Kielelliset konventiot	1	3			4	3
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	13				
					APT:	17
					Aineiston sanamäärä:	533
					Virheet suhteutettuna sanamäärään:	0,032

More climate records fall in world's warmest February (Poynting, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		2	1		3	7
Tarkkuus	1	4	7		12	39
Kielelliset konventiot		6	4		10	26
Tyyli			1		1	5
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	26				
					APT:	77
					Aineiston sanamäärä:	622
					Virheet suhteutettuna sanamäärään:	0,124

Salaatista ja tomaatista luvataan entistä terveellisempiä, jos geenisakset sallitaan EU:ssa (Puttonen, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		2	1		3	7
Tarkkuus		3	2		5	13
Kielelliset konventiot	2	2			4	2
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	12				APT: 22
						Aineiston sanamäärä: 615
						Virheet suhteutettuna sanamäärään: 0,036

Miksi Suomessa pitää opiskella suomen kieltä, vaikka osaisi hyvin suomea? (Rajatorpan koulun Maailman ihmeet -valinnaisaineen oppilaat, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia			1		1	5
Tarkkuus		5	5		10	30
Kielelliset konventiot	6	17	1		24	22
Tyyli				1	1	25
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus		2			2	2
Muut					0	0
	Virhemäärä yhteensä:	38				APT: 84
						Aineiston sanamäärä: 308
						Virheet suhteutettuna sanamäärään: 0,273

Hybridilentokone voi nousta jopa jalkapallokenttää pienemmältä aukiolta (Rantanen, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia					0	0
Tarkkuus	2	7	4		13	27
Kielelliset konventiot	3	6			9	6
Tyyli		1			1	1
Alueelliset konvention	1				1	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	24				APT: 34
						Aineiston sanamäärä: 964
						Virheet suhteutettuna sanamäärään: 0,035

Egyptistä löytyi valtava faarao Ramses Suuren patsas (Soininen, 2024a)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		1			1	1
Tarkkuus		1			1	1
Kielelliset konventiot	2	3			5	3
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	7				APT: 5
						Aineiston sanamäärä: 241
						Virheet suhteutettuna sanamäärään: 0,021

LIITE 2: PAIKALLISUUTISTEN PISTETAULUKOT

Theresa May to stand down as MP at next election (Baker & Sheddon, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia					0	0
Tarkkuus		1	1		2	6
Kielelliset konventiot					0	0
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	2				APT: 6
						Aineiston sanamäärä: 74
						Virheet suhteutettuna sanamäärään: 0,081

Nottingham attacks: Sons upset at not being told of vigil (Bartram, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	1	5			6	5
Tarkkuus	1	5	4		10	25
Kielelliset konventiot	1	5			6	5
Tyyli		1			1	1
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	23				APT: 36
						Aineiston sanamäärä: 868
						Virheet suhteutettuna sanamäärään: 0,041

Jäteauton kuljettaja: Rotat ovat monilla Helsingin seudun alueilla erityisen paha ongelma (Kalin, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		1	1		2	6
Tarkkuus		2	1		3	7
Kielelliset konventiot	2	3			5	3
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	10				
					APT:	16
					Aineiston sanamäärä:	783
					Virheet suhteutettuna sanamäärään:	0,020

£40k orphanage donor feels 'cheated' by charity (Khatun Dewan, 2024)

MQM-pisteytys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		3			3	3
Tarkkuus	4	7			11	7
Kielelliset konventiot	3	6			9	6
Tyyli	4	1			5	1
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	28				
					APT:	17
					Aineiston sanamäärä:	719
					Virheet suhteutettuna sanamäärään:	0,024

Poliisi varoittaa: Liikkeellä on taas uusi huijaus (Malmberg, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		1			1	1
Tarkkuus		1	1		2	6
Kielelliset konventiot		3			3	3
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	6				
					APT:	10
					Aineiston sanamäärä:	231
					Virheet suhteutettuna sanamäärään:	0,043

Eturivin poliitikot siirretään sivuun historiallisissa päätöksissä (Nissinen, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		3	1		4	8
Tarkkuus	3	8	1		12	13
Kielelliset konventiot	4	6	2		12	16
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus					0	0
Muut					0	0
	Virhemäärä yhteensä:	28				
					APT:	37
					Aineiston sanamäärä:	1075
					Virheet suhteutettuna sanamäärään:	0,034

Arvotalo Helsingin Puna-vuorella rakensi tuhansia euroja tuottavan pien-voimalan (Perttu, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	1	5	2		8	15
Tarkkuus	3	7			10	7
Kielelliset konventiot	5	7			12	7
Tyyli					0	0
Alueelliset konvention		1			1	1
Kohdeyleisölle sopivuus		1			1	1
Muut					0	0
	Virhemäärä yhteensä:	32				
					APT:	31
					Aineiston sanamäärä:	1114
					Virheet suhteutettuna sanamäärään:	0,028

Keravalle jumppakepit hankkinut toimialajohtaja siirtyy toisiin tehtäviin (Soininen, 2024b)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		3	1		4	8
Tarkkuus		2			2	2
Kielelliset konventiot		4			4	4
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus		1			1	1
Muut					0	0
	Virhemäärä yhteensä:	11				
					APT:	15
					Aineiston sanamäärä:	308
					Virheet suhteutettuna sanamäärään:	0,049

Patients facing uphill struggle to see GPs – report (Triggle, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia	1	2			3	2
Tarkkuus	2	1	1		4	6
Kielelliset konventiot	3	3			6	3
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus		1			1	1
Muut					0	0
	Virhemäärä yhteensä:	14				
					APT:	12
					Aineiston sanamäärä:	297
					Virheet suhteutettuna sanamäärään:	0,040

Adviser warns London a 'no-go zone for Jews every weekend' (Whannel, 2024)

MQM-pisteitys						
Vakavuustaso:	Neutraali	Pieni	Suuri	Kriittinen	Virhetyypin määrä	ETPT
Vakavuustason kertoimet:	0	1	5	25		
Virhetyypit:						
Terminologia		1			1	1
Tarkkuus		4	3	2	9	69
Kielelliset konventiot	1	1			2	1
Tyyli					0	0
Alueelliset konvention					0	0
Kohdeyleisölle sopivuus		1			1	1
Muut					0	0
	Virhemäärä yhteensä:	13				
					APT:	72
					Aineiston sanamäärä:	225
					Virheet suhteutettuna sanamäärään:	0,320