

Taloustekstien konekäännösten laadun arviointi saksasta suomeen käännettäessä

Esimerkkinä DeepL-käännin

Irina Dalfors

Pro gradu -tutkielma

Monikielinen käännösviestintä, saksan kieli

Kieli- ja käännöstieteiden laitos

Humanistinen tiedekunta

Turun yliopisto

Syyskuu 2024

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

Pro gradu -tutkielma

Monikielinen käännösviestintä, saksan kieli

Irina Dalfors

**Taloustekstien konekäännösten laadun arviointi saksasta suomeen käännettäessä
Esimerkkinä DeepL-käännin**

Sivumäärät: tutkielman sivumäärä 61, liitteiden sivumäärä 25

Nykyään neuroverkkokääntimien, kuten DeepL:n, käyttö on yleistynyt sekä maallikoiden että ammattikäntäjien keskuudessa. DeepL on noussut merkittäväksi toimijaksi konekääntämisen alalla ja ansainnut mainetta kyvystään tuottaa suhteellisen laadukkaampia ja sujuvampia käännöksiä, erityisesti yleisluontoisista teksteistä. DeepL-kääntimen käännöksiä on tarkasteltu useissa tutkimuksissa. Tulokset ovat osoittaneet, että vaikka DeepL-käännökset voivat olla kieliopillisesti oikeita, sujuvia ja usein jopa loogisia, niissä voi olla puutteita erikoisalan termeissä ja monimerkityksissä sanoissa.

DeepL-käännökset eivät näin ollen aina luotettavia, ja niiden käytössä on tärkeää tarkkailla laatua. Tässä tutkimuksessa keskityttiin DeepL-konekäännöksen laadun arviointiin erityisesti saksasta suomeen käännettäessä. Aineistona käytettiin talousaiheisia tekstikatkelmia, ja niiden laatua analysoitiin manuaalisesti moniulotteisen ja standardoidun MQM-käännöslaatumittarin avulla. MQM-mittarin virhekategorioiden analysoitavaksi valittiin ainoastaan tutkimuksen kannalta relevantit kategoriat: tarkkuus, sujuvuus ja terminologia. Näiden kategorioiden avulla tarkasteltiin virheitä tyypeittäin ja määriteltiin virheiden vakavuusasteet, ja lopulta laskettiin laatuasteen numeerisesti.

Tutkimuksen tarkoituksena oli selvittää, minkälaisia tarkkuuteen, sujuvuuteen ja terminologiaan liittyviä virheitä esiintyy konekäännöksessä ja minkälainen vaikutus niillä on lauseiden semanttisen merkityksen välittymiseen. Konekäännösten tulosta vertailtiin vielä ihmisen tekemän käännöksen kanssa, jotta saataisiin selville, minkälaisia eroja niiden välillä esiintyy. Tulokset osoittivat, että DeepL-käännös ei ollut virheetön: Eniten virheitä esiintyi tarkkuuden kategoriassa ja sujuvuuden kategoriassa, kun taas terminologiassa virheitä oli vähiten. DeepL-käännöksen virheet eivät olleet kriittisiä, ja ne eivät johtaneet koko käännöksen hylkäämiseen. Vakavuusasteiltaan virheet olivat joko merkittäviä tai vähäisiä, ja ne vaikeuttivat pääasiassa semanttisen merkityksen välittämistä. Aineiston analyysin perusteella päädyttiin tulokseen, että DeepL-käännin ei vielä ole ihmiskääntäjän kanssa samalla tasolla tietokirjojen tekstien kääntämisessä. Vaikka DeepL-käännökset tarjoavat hyvän lähtökohdan käännöstyölle, ne tarvitsevat edelleen jälkieditointia ennen julkaisua.

Avainsanat: konekääntäminen, laadunarviointi, virhekeskeinen arviointi, MQM

Sisällysluettelo

1	Johdanto	5
2	Konekääntäminen	8
2.1	Katsaus konekääntämisen historiaan ja menetelmiin	9
2.2	Konekääntämisen mahdollisuuksia ja haasteita	14
2.2.1	Konekääntämisen mahdollisuuksia	14
2.2.2	Konekääntämisen haasteita	15
2.3	DeepL-konekäännin	17
3	Kieli ja tekstilaji neuroverkkokääntämisen näkökulmasta	18
3.1	Suomen ja saksan kirjakielten erityispiirteitä	18
3.2	Tietokirja tekstilajina	21
4	Konekäännöksen laadunarviointi	23
4.1	Laadukas käännös ja käännöksen laatu	23
4.1.1	Funktionaalisuus eli toimivuus laadun kriteerinä	24
4.1.2	Ekvivalenssi eli vastaavuus laadun kriteerinä	25
4.2	Laadunarviointimenetelmät	26
4.2.1	Manuaalinen ja automaattinen lähestymistapa	26
4.2.2	Holistinen ja analyyttinen lähestymistapa	28
4.2.3	Yksiulotteinen ja moniulotteinen lähestymistapa	29
4.3	Moniulotteinen MQM-käännöslaatumittari	29
5	Tutkimusaineisto- ja menetelmät	32
5.1	Aineiston valinta ja kuvaus	32
5.2	Tutkimusmenetelmät	33
5.2.1	MQM-virhetytologia	33
5.2.2	Virhekatteerit tarkkuus, sujuvuus ja terminologia	34
5.2.3	Virheiden vakavuusasteet	36
5.2.4	Laatupistemäärän laskeminen	37
6	Tulokset ja analyysi	40
6.1	DeepL-konekäännöksen määrällinen analyysi	40
6.2	Ihmis- ja DeepL-käännöksen laadullinen analyysi	42
6.2.1	Tarkkuusvirheet	43

6.2.2	Sujuvuusvirheet	48
6.2.3	Terminologiavirheet	50
7	Yhteenveto ja pohdinta	53
	Lähteet	56
	Liitteet	62
	Liite 1. Deutsche Zusammenfassung	62
	Liite 2. DeepL-konekäännös	78
	Liite 3. DeepL-käännöksen virheiden dokumentointi	86

1 Johdanto

Tutkielmani aiheena on DeepL-konekäännöksen laadun arviointi sekä konekäännöksen ja ihmisen kääntämän tekstin vertailu. Nykyisin konekääntimet ovat yleisiä työkaluja käännösosalalla, mikä käy ilmi esimerkiksi ELIS (European Language Industry Survey) -kyselyn raportista. Vuoden 2022 raportin mukaan konekääntimiä käytti lähes 58 % käännösyrityksistä (n=264), ja lisäksi 20 % tutkimukseen osallistuneista käännösyrityksistä suunnitteli käyttävänsä niitä (ELIS Research 2022). Vaikka raportissa ei anneta tarkempaa tietoa siitä, kuinka usein ja minkä tyyppisiin töihin konekääntimiä yrityksissä hyödynnettiin, niiden käyttötavat ja käytön määrä todennäköisesti vaihtelevat kääntämisen eri osa-alueilla. Esimerkiksi teknisten ja hallinnollisten tekstien, kuten käyttöohjeiden tai EU-tekstien, kääntämisessä koneavusteiset menetelmät, mukaan lukien konekääntimet, ovat laajassa käytössä. Monissa muissa tekstilajeissa konekääntämisen käyttö on ollut vähäisempää johtuen tekstilajien erityispiirteistä (Koponen 2022, 270). Erityisesti sellaiset konekääntimet kuin Google Translate ja DeepL ovat käyttäjäystävällisiä ja laajasti tunnettuja internetissä ja sosiaalisessa mediassa. Tämä on tehnyt konekääntämisestä kaikkien saavutettavissa olevaa ja arkipäiväistänyt sen käyttöä (mt., 269).

Konekääntimiä on käytetty eri alojen tekstien kääntämisessä niin kaunokirjallisuudessa kuin erikoisalojen tekstien kääntämisessäkin. Esimerkiksi Belgian Gentin yliopistossa vuonna 2020 tehdyssä tutkimuksessa (Webster ym. 2020) tutkijat tarkastelivat muun muassa neuroverkkopohjaisten konekääntimien soveltuvuutta kaunokirjallisuuden kääntämiseen. Tutkimuksessa käytettiin Google Translatea ja DeepL-konekäännintä neljän englanninkielisen klassikkoromaanin kääntämiseen hollanniksi, ja näitä konekäännöksiä verrattiin ihmiskäännöksiin. Tutkimus osoitti, että konekäännökset sisälsivät enemmän virheitä kuin ihmiskäännökset, niiden sanasto oli suppeampi ja teksti oli vähemmän yhtenäistä. Lisäksi konekäännökset olivat taipuvaisempia seuraamaan lähdetekstin rakennetta, kun taas ihmiskäännökset erottuivat tyylillisesti konekäännöksistä. Tutkimuksen tuloksena oli, että konekäännökset eivät yltäneet ihmisten tekemien käännösten laatuun kirjallisuuden kääntämisessä. Myöskään talousartikkelien konekäännökset eivät ole virheettömiä. Esimerkiksi Kriston (2017) raportoi tutkimuksesta, jossa arvioitiin Google Translaten, SDL Free Translatorin ja Microsoft Bingin konekäännösten luotettavuutta verrattuna ihmiskäännökseen. Tutkimuksessa talousartikkeleita käännettiin romanian kielestä englanniksi. Näistä kolmesta neuroverkkokääntimestä parhaan tuloksen tarjosi Google

Translate, mutta kuitenkin Google-käännöksessä oli vielä paljon parannettavaa.

Haasteellisimmaksi konekääntimelle osoittautuivat muun muassa prepositioiden kääntäminen, oikean termin valitseminen kahden hyvin samankaltaisen merkityksen välillä ja lauseen subjektit. Lisäksi konekäännös sisälsi muutamia kieliopillisia virheitä, kuten miespuolisen subjektin ”he” lisääminen sanan ”pankki” eteen, mitä pidettiin vääränä mutta myös outona ratkaisuna, koska sana ”pankki” kuuluu romanian kielessä feminiinisukuun. Tutkimus antoi tukea ajatukselle, että laadukkaiden käännösten tuottamisessa kääntäjällä on yhä vahva rooli.

Konekääntämisen laatuun liittyy siis vielä ratkaisemattomia kysymyksiä. Viime vuosina konekäännösten sujuvuus on merkittävästi parantunut neuroverkkokääntämisen yleistymisen myötä (Latomaa–Nurminen 2022), mutta sujuvuus ei välttämättä takaa laatua, ja sen voisi jopa olettaa aiheuttavan vaikeuksia konekääntimen tekemien virheiden havaitsemisessa. Vaikka neuroverkkokonekäännökset alkavat kielellisen sujuvuuden osalta muistuttaa ihmisen tekemiä käännöksiä, ne saattavat edelleen sisältää selkeitä merkitysvirheitä. Tilanteissa, joissa informaation tarkkuus on elintärkeää ja virhe saattaisi aiheuttaa vakavaa vaaraa ja vahinkoa henkilöille, heidän terveydelleen tai omaisuudelleen, pelkkä konekäännös ei riitä (Koponen 2022, 277). Lisäksi neuroverkkokonekäännösten laatu vaihtelee huomattavasti eri kieliparien välillä. Laadun tasoon vaikuttaa ennen kaikkea se, kuinka paljon kääntimen kehittämiseen on käytettävissä kieliaineistoa. Koponen (mt., 278) korostaa, että parhaat tulokset saavutetaan kielissä, joissa aineistoa on runsaasti, kuten esimerkiksi englannissa ja espanjassa. Vuonna 2019 Sveitsin liittokanslerinviraston kielipalveluyksikön italian osaston aloitteesta toteutettiin tutkimus, jonka tavoitteena oli arvioida DeepL-konekääntimen soveltuvuutta liittovaltion hallinnon kielipalveluihin. Tutkimuksessa havaittiin muun muassa selviä laatueroja eri kielisuuntien välillä. Erityisesti italia-saksa-käännöksen laatu jäi jälkeen verrattuna ranska-saksa- tai englantia-saksa-käännöksiin (Bericht DeepL-Test 2019). Vaikka myöskään suomen kielen konekäännös ei ole vielä saavuttanut korkeinta tasoaan, tilanne on silti parempi verrattuna useisiin muihin vähemmistökieliin, joissa kieliaineisto on vieläkin rajallisempaa (Koponen 2022, 278).

Kuten edellä on käynyt ilmi, neuroverkkopohjaisilla konekääntimillä aikaansaadut käännökset eivät välttämättä ole sisällöltään luotettavia ja kielellisesti virheettömiä. Niiden käytössä on tärkeää tarkkailla laatua, erityisesti tietojen tarkkuutta vaativissa teksteissä ja jos käännös tapahtuu suomen kaltaiseen pienempään kieleen päin. Neuroverkkokäännösten laadun tutkiminen on tarpeellista, koska tämäntyyppinen tutkimus voi auttaa käännösohjelmien

kehittäjiä ymmärtämään, millä alueilla (syntaksi, morfologia jne.) järjestelmä kaipaa parannuksia, ja parantamaan siten neuroverkkokääntimien suorituskykyä.

Neuroverkkokääntimet ovat ammattikäytäjäiden työvälineitä jo nykyään ja varmasti myös tulevaisuudessa. Siksi tätä aihetta on myös kiinnostavaa tutkia.

Tämän tutkielman tavoitteena on selvittää, kuinka laadukkaita neuroverkkokonekäännökset ovat ja missä määrin ne voivat vastata ihmisten tekemien käännösten tasoa. Tällä tutkimuksella pyrin erityisesti arvioimaan DeepL-käännösohjelman soveltuvuutta taloudellisten tekstien kääntämiseen saksasta suomeen.

Tutkielman avulla pyrin vastaamaan seuraaviin kysymyksiin:

- 1) Minkälaisia tarkkuuteen, sujuvuuteen ja terminologiaan liittyviä virheitä esiintyy käännettäessä saksankielisiä talousalan tekstejä suomeksi DeepL-kääntimellä?
- 2) Missä määrin konekäännöksissä esiintyvät semanttisen merkityksen vääristymät ovat haitallisia tekstin ymmärtämisen tai käytön kannalta?
- 3) Missä määrin DeepL-kääntimen talousalan tekstistä tuottaman konekäännöksen laatu vastaa ihmisen tuottaman käännöksen laatua?

Tutkielmassa käytän aineistoina DeepL-kääntimen käännöstä ja suomenkielistä ihmisen tekemää referenssikäännöstä. Aineistoa kerään Ulrike Herrmannin tietokirjasta ”Der Sieg des Kapitals”, josta on julkaistu kääntäjä Mari Janatuisen tekemä suomennos ”Pääoman voitto”. Analysoitavan aineiston monipuolisuuden varmistamiseksi valitsen tietokirjasta talouden eri teemoja käsitteleviä osuuksia, ja analyysikehyksenä käytän MQM-virheluokittelujärjestelmää.

Luvussa 2 esittelen lyhyesti konekääntämisen historiaa, menetelmiä sekä konekäännösten käyttömahdollisuuksia. Samalla tarkastelen konekääntämisen tuomia etuja ja haasteita ja esittelen DeepL-konekääntimen. Luvussa 3 käsittelen saksan ja suomen kielen erityispiirteitä, jotka saattavat muodostaa haasteita konekääntimelle. Näitä piirteitä otan huomioon tutkielmani analyysissä. Luvussa 4 keskityn käännösten laatuun, määrittelen laadun käsitettä ja esittelen laadun arviointimenetelmiä. Luvussa 5 esittelen tutkimusaineiston ja tutkimusmenetelmät, kuten MQM-virheluokittelumallin. Luvussa 6 esittelen analyysin tulokset ja niiden pohjalta tehdyt johtopäätökset. Lopuksi luvussa 7 teen yhteenvedon tutkimustuloksista ja pohdin niiden merkitystä.

2 Konekääntäminen

Koska tutkimuksessani aion tutkia DeepL-konekääntimen tuotosta, pidän tärkeänä ensin määritellä, mitä ”konekääntäminen” tarkoittaa. Tässä vaiheessa haluan kiinnittää huomiota siihen, että ”konekääntäminen” ei ole sama asia kuin ”koneavusteinen kääntäminen”. Nämä termit voivat joskus sekoittua keskenään. ”Koneavusteinen kääntäminen” tunnetaan paremmin englanninkielisellä termillä ”CAT” (engl. computer-assisted translation) ja ”konekääntäminen” termillä ”MT” (engl. machine translation) (Somers 2012, 429–430).

Koneavusteisessa kääntämisessä pääasiallisena kääntäjänä toimii ihminen käyttäen käännösprosessissa erilaisia CAT-työkaluja, kuten termipankkeja, sanakirjoja, kieliopillisia tarkistuksia jne., jotka auttavat häntä tekemään tarkkoja ja johdonmukaisia käännöksiä (Somers 2012, 429–430). Sitä vastoin konekääntämisellä tarkoitetaan ”automaattista kääntämistä”, jossa konekääntimet kääntävät tekstin yhdestä kielestä toiseen automaattisesti eli ilman ihmisen välitöntä osallistumista. Konekääntämiselle on muitakin määritelmiä. Tarkoitukseni ei ole kuitenkaan käydä läpi niitä kaikkia, vaan keskittyä oman tutkimukseni kannalta olennaiseen. Kun lähdin tutkimaan määritelmien kirjoja, huomasin, että konekääntämiselle on esitetty sekä laajempia että suppeampia määritelmiä.

Esimerkiksi Tieteen termipankki (28.2.2024: Käännöstiede: konekääntäminen.) määrittelee konekääntämisen prosessiksi, jossa tietokoneohjelma muuntaa tekstiä yhdeltä kieleltä toiselle. Tämä määritelmä jättää tilaa tulkinnalle. Sitä vastoin Somersin (2012, 428) mukaan konekääntäminen on niin sanottua automaattista kääntämistä, jossa konekäännösohjelmat kääntävät tekstin automaattisesti yhdestä kielestä toiseen käyttäen erilaisia algoritmeja. Myös Kenny (2018, 428) määrittelee konekääntämisen automaattiseksi prosessiksi, mutta hän tarkentaa konekääntämisen määritelmää kielen suhteen. Kenny täydentää, että konekääntäminen on kirjoitetun kielen tekstin automaattista kääntämistä yhdestä ihmiskielestä toiseen. Tässä määritelmässä korostuu sana ”ihmiskieli”, joka herättää minussa kysymykset: Miten konekääntäminen voi tuottaa lopputuloksena toisen ihmiskielen? Onko koneen kyky tulkita ihmiskieltä riittävän kehittynyt, jotta lopputulos olisi aitoa ihmiskieltä? Vaikuttaa siltä, että tämä on mahdollista, sillä nykyään on olemassa monia erilaisia konekääntimiä, ja niitä käytetään aktiivisesti eri yhteyksissä usealla eri tavalla.

Konekääntimet ovatkin kehittyneet merkittävästi viime vuosina, ja niiden käyttö on jo yleistynyt nykypäivänä. Konekääntämisen tulevaisuudennäkymät vaikuttavat siis lupaavilta. Mutta miksi konekääntimiä on ylipäänsä lähdetty kehittämään? Tähän kysymykseen vastaaminen edellyttää mielestäni konekääntämisen kehityshistorian ja menetelmien tarkastelua, ja näitä aion sivuta hieman seuraavaksi.

2.1 Katsaus konekääntämisen historiaan ja menetelmiin

Konekääntämisen ajatus on kiehtonut tutkijoita kautta aikojen jo ennen kuin tietokoneita oli edes keksitty. Konekääntämisen kehittämisen taustalla on ollut useita erilaisia syitä, jotka ovat vaihdelleet aikakaudesta ja vallitsevasta tilanteesta riippuen. Alussa konekääntämisen järjestelmiä kehitettiin pääasiassa sotilaallisiin tarkoituksiin ja valtioiden välisiin viestintätarpeisiin (Wang 2022, 143). Joskus konekääntämisen kehittämiseen ei tarvittu edes mitään varsinaista syytä, vaan yksinkertaisesti kielitieteilijöiden ja insinöörien innostus ja näkemys konekääntämisen potentiaalista riittivät (Somers 2012, 429).

Somersin (2012, 429) mukaan konekääntämisen kehittämiseen liittyvän historian voidaan katsoa ulottuvan toisen maailmansodan jälkeiseen aikaan. Vuosina 1947–1949 merkittävä amerikkalainen matemaatikko Warren Weaver esitti ajatuksen, että samoja laskennallisia menetelmiä, joita hyödynnettiin toisen maailmansodan aikana koodinmurtotehtävissä, voitaisiin soveltaa myös käännoistyöhön. Hänen visionääriset ajatuksensa herättivät suurta kiinnostusta Yhdysvalloissa ja kansainvälisesti. Konekäännösteknologian potentiaali ja sen mahdolliset käyttökohteet koettiin merkittäviksi, mikä synnytti huomattavia taloudellisia investointeja alan tutkimukseen ja kehitystyöhön (mt.). 1950-luvulla ja 1960-luvun alkupuolella kehitettiin useita konekäänninjärjestelmiä. Ensimmäisten joukossa oli vuonna 1952 kehitetty Yhdysvaltain Georgetown-IBM-järjestelmä, joka pystyi kääntämään venäjäästä englanniksi (Chan 2023, 23). Konekääntämisen kehitys oli aktiivista, ja Yhdysvaltain lisäksi muut konekääntämisen edelläkävijämaat olivat Venäjä, Iso-Britannia, Japani, Kiina ja Tšekkoslovakia (mt., 24).

Aktiivinen vaihe ei kuitenkaan jatkunut pitkään. Erityisesti Yhdysvalloissa konekääntämisen tutkimus ja kehitys pysähtyivät vuosikymmeniksi. Tämä johtui siitä, että vuonna 1964 Yhdysvaltain hallitus päätti tarkastella konekääntämisen edistymistä ja perusti automaattisen kielenkäsittelyn neuvoa-antavan komitean (ALPAC). Tarkastuksen päätteeksi vuonna 1966

julkaistiin raportti, joka oli varsin kriittinen konekääntämistä kohtaan. Raportissa erityisesti painotettiin konekääntimien hitautta, epätarkkuutta ja kustannustehottomuutta verrattuna ihmiskääntämiseen (Somers 2012, 430). Raportin seurauksena konekääntämisen tutkimusrahoitukseen tehtiin merkittäviä leikkauksia. Somersin (mt.) mukaan nämä tulokset olivat ennustettavissa, koska varhaisen konekääntämisen kehitystä haittasivat teknologian alkeellisuus ja tutkijoiden, pääasiassa matemaatikkojen ja insinöörien, vähäinen kielitieteellinen tausta.

Vaikka ALPAC-raportin tulokset olivat kielteisiä ja konekäännöstutkimus väheni nopeasti, Kanadassa, Länsi-Euroopassa ja Japanissa tutkimus jatkui aktiivisesti 1970- ja 1980-luvuilla. Näiden tutkimusten pohjalta syntyi erilaisia konekäännösjärjestelmiä, kuten CULT-konekäännösjärjestelmä, TAUM-METEO-järjestelmä sääennusteiden kääntämistä varten sekä EUROTRA-järjestelmä, jolla voitiin kääntää kaikkia Euroopan talousyhteisön virallisia kieliä (Chan 2023, 24–25). Näiden 1970- ja 1980-lukujen tutkimuksissa painotettiin erityisesti kieliopillisia rakenteita, syntaksia ja sääntöpohjaisia lähestymistapoja (Somers 2012, 429).

Sääntöpohjaisessa menetelmässä (lyh. RBMT) käytettiin kaksikielisiä sanakirjoja ja manuaalisesti kirjoitettuja sääntöjä lähdekielisten tekstien kääntämiseksi kohdekielisiksi teksteiksi (Wang 2022, 143–153). Ohjelmistoja suunniteltiin niin, että kielitieteilijät ja kääntäjät voisivat kirjoittaa sääntöjä ilman syvällistä ymmärrystä tietokoneohjelmien toiminnasta (Somers 2012, 431). Vuonna 1978 SYSTRAN toi markkinoille kaupallisen käännösjärjestelmän, joka oli Wangin (2022, 143–153) mukaan tuolloin yksi tunnetuimmista esimerkeistä kaupallisesti onnistuneesta sääntöpohjaisesta järjestelmästä.

Vuosien varrella on kehitetty erilaisia sääntöpohjaisia menetelmiä, kuten *suora* (engl. direct MT) ja *epäsuora* (engl. undirect MT) konekääntäminen. Epäsuora lähestymistapa jaetaan vielä kahteen eri menetelmään: *siirtomenetelmään* (engl. transfer-based MT) ja *kieltenväliseen menetelmään* (engl. inter-lingua based MT).

Nitzken ja Hansen-Schirran (2021, 22) mukaan *suorassa konekääntämisessä* lähtökohtaisesti lähdetekstin sanat analysoidaan morfologisesti ja etsitään sen jälkeen sanakirjasta. Tämä tarkoittaa, että ihanteellisesti kaikki morfologiasäännöt on määritelty, joten sanakirjan tarvitsee sisältää vain sanojen vartalot. Seuraavaksi lähdekielen sanat korvataan kohdekielen sanoilla ja tehdään kaikki kohdekielen vaatimat morfologiset muutokset. *Siirtomenetelmässä*

puolestaan lähdetekstistä rakennetaan ensin syntaktinen esitys, joka on vapaa moniselitteisyydestä. Seuraavaksi tämä esitys luodaan kohdekielelle kieliopin avulla, joka sisältää kaksikieliset siirtosäännöt. *Kieltenvälisessä menetelmässä* taas peruseriaatteena on, että lähtöteksti käännetään ensin niin sanotuksi ”interlinguaksi”, joka edustaa merkitystä abstraktissa muodossa, ja sitten ”interlingua” käännetään kohdekieleksi.

Sääntöpohjaisen menetelmän suurin haaste oli laajojen kaksikielisten sanakirjojen ja sääntökokoelmien manuaalinen luominen ja ylläpitäminen (Wang 2022, 143–153). Tästä syystä etsittiin uusia menetelmiä, joilla saavutettaisiin parempia tuloksia vähemmällä vaivalla. 1980-luvun lopulla konekääntämisen tutkijat tekivät merkittävän suunnanmuutoksen ja siirtyivät kokeilemaan tilastollista lähestymistapaa. He uskoivat, että tietokone voisi oppia tekemään käännöksiä analysoimalla tilastollisesti aiempia käännöksiä. Somersin (2012, 430) mukaan tilastollinen menetelmä perustui ajatukseen, että riittävällä laskentateholla ja aiemman käännösaineiston avulla voitaisiin ennustaa todennäköisimmät kohdesanat ja niiden järjestys lähtötekstin perusteella. Vuodesta 1989 lähtien korpuspohjaiset järjestelmät, jotka mahdollistivat suuren tarkkuuden vähemmällä vaivalla, ovat alkaneet saada suosiota konekääntämisessä (Sarkhel ja Tripathi 2010, 390).

Tilastollinen menetelmä (lyh. SMT) kuuluu korpuspohjaisten menetelmien joukkoon.

Forcadan (2010, 220) mukaan siinä ei käytetä mitään kielellistä tietoa, vaan tavoitteena on löytää kohdekielen vastaava lause, jonka todennäköisyys on suurin. Werthmannin ja Wittin (2014, 95) mukaan tilastollinen menetelmä koostuu seuraavista keskeisistä komponenteista:

- a) *Kohdekielinen malli*: Tämä malli perustuu kohdekieliseen yksikieliseen korpukseseen. Se vastaa kohdekielisten sanojen oikeasta muodostuksesta ja laskee kohdelauseen todennäköisyyden.
- b) *Rinnakkaiseen korpukseseen perustuva käännösmalli*: Tämä malli perustuu rinnakkaiseen korpukseseen, joka sisältää lähdetekstit ja niiden käännökset. Siinä lasketaan kohdelauseen ehdollinen todennäköisyys lähdelauseen perusteella.
- c) *Dekooderi*: Dekooderi suorittaa varsinaisen käännöstehtävän. Se etsii parhaan mahdollisen käännösparin maksimoimalla kahden edellä mainittujen mallien laskemat todennäköisyydet.

Werthmannin ja Wittin (2014, 95) mukaan tilastollinen menetelmä oli sääntöpohjaiseen menetelmään verrattuna edullisempi, koska se ei vaatinut aikaa vievää sääntökokonaisuuksien luomista. On kuitenkin huomioitava, että suurten korpusten luominen voi olla vaivalloista,

mikä oli tilastollisen menetelmän heikkous (mt., 96). Tilastollinen menetelmä saavutti pian vaiheen, jossa merkittävää parannusta oli vaikea enää saavuttaa. Siksi oli tärkeää etsiä muita lähestymistapoja konekääntämisen parantamiseksi.

Uudempi menetelmä löytyi *neuroverkkokääntämisestä*. Neuroverkkokääntäminen (lyh. NMT) alkoi saada merkittävää huomiota 2010-luvulla. Erityisen merkittävä edistysaskel tapahtui vuonna 2016, kun se tuli laajan käyttöön. Vaikka se kuuluukin korpuspohjaisten menetelmien joukkoon, sitä kutsutaan usein datapohjaiseksi menetelmäksi (Mohamed ym. 2021, 15919). Konekääntämisen alalla neuroverkkokääntäminen (NMT) on onnistunut voittamaan tilastollisen konekääntämisen (SMT), ja lyhyessä ajassa se on saavuttanut suuren suosion käännösalan tutkijoiden keskuudessa sen rakenteen yksinkertaisuuden ja lupaavien käännöstulosten takia (mt.).

NMT:n suosio johtuu todennäköisesti siitä, että se perustuu tekoälyyn ja tarjoaa merkittäviä parannuksia perinteisiin sääntöpohjaisiin ja tilastollisiin konekäännösmenetelmiin verrattuna. Schmalzin (2019, 196–197) mukaan neuroverkkokääntäminen pystyy tuottamaan sujuvampia ja ihmisen tekemän käännöksen laatua lähempänä olevia käännöksiä analysoimalla kokonaisia lauseita, ja se kykenee ottamaan huomioon monimutkaisempia asiayhteyksiä käännöksen tuottamisessa, mikä edistää käännösten luonnollisuutta. Tämä saavutetaan hyödyntämällä käännösalgoritmeja, jotka ovat itseoppivia järjestelmiä ja jotka eivät nojautu ennalta määriteltyihin kielisääntöihin.

NMT koostuu yhdestä suuresta hermoverkosta, joka sisältää enkooderin ja dekooderin. Enkooderi vastaanottaa lähdekielisen lauseen sanat, käsittelee ne ja muuntaa saadun tiedon kontekstivektoriksi, joka edustaa lauseen semanttista merkitystä. Tämän jälkeen dekooderi käyttää tätä vektoria kohdelauseen luomiseen sana sanalta. (Mohamed ym. 2021, 15919). Verkon sanotaan perustuvan ihmisaivojen neuronien malliin, ja sen avulla järjestelmä pystyy luomaan tärkeitä asiayhteyksiä sanojen ja fraasien välille. Cullenin (ei vuotta) mukaan tällaiset yhteydet ovat mahdollisia, koska järjestelmä on skannannut miljoonia esimerkkilauseita tietokannastaan yhteisten piirteiden löytämiseksi ja oppinut tunnistamaan kielisääntöjä niiden avulla. Tämän jälkeen kone käyttää näitä sääntöjä tilastollisten mallien laatimiseen, mikä auttaa sitä oppimaan, miten lauseet tulisi rakentaa.

Neuroverkkokääntimiä on tutkittu laajasti, ja tulokset ovat olleet sekä myönteisiä että kielteisiä. Matusovin (2019, 10–19) tutkimuksessa, jossa analysoitiin romaanien konekäännöksiä saksasta englanniksi ja englannista venäjäksi, todettiin käännösten laadun olevan riittävän korkea, jotta tarina voitiin ymmärtää. Käännöksiä oli jopa miellyttävä lukea. Noin 30 % konekäännetyistä lauseista oli laadultaan hyväksyttäviä, ja ne saattoivat vaatia vain kohdekielen oikolukua. Vakavia syntaktisia virheitä havaittiin hyvin vähän, mutta merkitysvirheitä moniselitteisten sanojen kohdalla oli edelleen paljon.

Eduistaan huolimatta neuroverkkokääntimillä on myös heikkoutensa. Yksi keskeisimmistä haasteista on suurten harjoitusaineistojen tarve. Koehn (2020, 296) toteaa, että pienemmät datamäärät johtavat usein heikompiin tuloksiin, kun taas suurempi määrä parantaa käännösten laatua. Jos harjoitusmateriaalia ei ole riittävästi, laatu on siis huono. Nitzke ja Hansen-Schirran (2021, 26) mukaan tämä on usein ongelmallista pienten kielten ja harvinaisten kieliyhdistelmien kohdalla, koska ne ovat usein aliedustettuja ja resursseiltaan köyhiä.

Koehnin (2020, 294) mukaan NMT-kääntimillä saattaa lisäksi olla haasteita eri erikoisalojen sanojen kanssa. Mikäli järjestelmää ei ole asianmukaisesti koulutettu tietylle erikoisalalle, sen tuottamat tulokset voivat olla heikkoja. Koehn (mt., 295) korostaa, että tämä on erityisen huolestuttavaa silloin, kun konekäännöstä käytetään tiedonhakuun. NMT-kääntimien tuotokset saattavat johtaa käyttäjää harhaan. Koehnin (mt., 295) mukaan tämä johtuu siitä, että kun NMT-käännin kohtaa erikoisalan ulkopuolista sisältöä, se yleensä yrittää kääntää sen. Tämän seurauksena käännös saattaa vaikuttaa sujuvalta, mutta se voi olla täysin irrallaan alkuperäisestä sisällöstä eli epätarkka. Sitä vastoin tilastollinen konekäännin jättää usein nämä erikoisalan ulkopuoliset osat kääntämättä vastaavissa tapauksissa.

NMT-kääntimet tuottavat siis sujuvia ja hyvin tarkkojakin käännöksiä, mutta ne nähdään usein myös ns. ”mustina laatikkoina” (Mohamed 2021, 15929). Mohamed (mt.) korostaa, että NMT-kääntimet ovat erittäin monimutkaisia, eikä niiden päätöksentekoprosesseja tunneta kovin hyvin, erityisesti mitä tulee sanojen valintaan dekodausvaiheessa. Tämä tekee virheiden korjaamisesta haastavaa. Jos järjestelmä oppii virheellisiä malleja, se voi toistaa virheitä jatkuvasti. Sveitsin liittokanslerinviraston kielipalveluyksikön italian osaston teettämässä neuroverkkopohjaisen DeepL-kääntimen suoritukseen liittyvässä tutkimuksessa (Bericht DeepL-Test 2019) kiteytettiin myös, että DeepL-käännin ei ajattele, eikä se opi jo tehdyistä korjauksista. Jotkut korjaukset oli tehtävä samaan tekstiin yhä uudelleen.

Mohamedin (2021, 15929) katsausartikkelin mukaan NMT-kääntimillä on haasteita monitulkintaisten ja harvinaisten sanojen sekä pitkien lauseiden kanssa. Esimerkiksi saksankielinen sana ”der Flügel” voi tarkoittaa sekä soitinta että linnun siipeä, ja oikean merkityksen valinta riippuu kääntimen kouluttamiseen käytetyn aineiston määrästä ja kohdekielellä saatavilla olevasta datasta. Koehnin (2020, 297) mukaan harvinaisten sanojen kääntäminen voi olla erityisen haastavaa voimakkaasti taivutetuissa kielissä, sillä monet taivutusmuodot voivat esiintyä harvoin. Mohamed (2021, 15929) korostaa, että NMT-järjestelmät toimivat hyvin lyhyitä lauseita sisältävissä teksteissä, mutta käännösten laatu heikkenee merkittävästi lauseen pituuden kasvaessa. Koehnin ja Knowlesin (2017, 6) mukaan yli 60 sanan lauseissa tilastolliset kääntämismenetelmät ovat NMT-kääntimiä parempia. Vaikuttaa siltä, että usein pitkiä virkkeitä sisältävien saksankielisten tekstien kääntäminen saattaa aiheuttaa haasteita NMT-kääntimille.

2.2 Konekääntämisen mahdollisuuksia ja haasteita

Nykyään maailmanlaajuinen yhteistyö ja kaupan lisääntyminen ovat kasvattaneet tarvetta kielten väliselle viestinnälle ja käännöksille monilla eri aloilla, kuten liike-elämässä ja kulttuurivaihdossa. Konekääntäminen tarjoaa tehokkaan tavan kääntää tekstiä eri kielille, ja sen avulla voidaan käsitellä suuria määriä tekstiä lyhyessä ajassa, mikä mahdollistaa monikielisen kommunikaation. Toisaalta konekääntämiseen liittyy myös haasteita. Seuraavaksi tuon lyhyesti esille, minkälaisiin tarkoituksiin konekääntimiä pystyy käyttämään ja minkälaisissa tilanteissa niiden käyttö ei ole suotavaa.

2.2.1 Konekääntämisen mahdollisuuksia

Konekääntimiä voidaan hyödyntää moniin eri tarkoituksiin, mutta niiden soveltuvuus riippuu suuresti käännösten laadullisista vaatimuksista. Konekääntämisen avulla voidaan esimerkiksi poistaa kielellisiä esteitä, edistää kielirajat ylittävää viestintää ja tarjota laajemmalle joukolle ihmisiä pääsy informaatioon (Schmalz 2019, 196–197). Esimerkiksi Google, Microsoft ja DeepL tarjoavat ilmaisia käännöspalveluita internetin käyttäjille. Vaikka tällaiset yleiskäyttöön tarkoitetut kääntimet pyrkivät pääasiassa välittämään alkuperäisen tekstin merkityksen, niiden avulla aikaansaadut käännökset saattavat sisältää virheitä. Koposen (2017) mukaan virheistä huolimatta nämä konekäännökset voivat auttaa lukijaa hahmottamaan ainakin tekstin keskeiset sisällöt. Koponen (2022, 205) korostaa, että

konekäännös voi olla hyödyllinen väline erityisesti silloin, kun ihmisen tekemää käännöstä ei ole saatavilla. Tuomisen (2023, 93) mukaan esimerkiksi Suomessa asuvat henkilöt, jotka eivät osaa suomea tai ruotsia, saattavat kokea konekäännöksen arvokkaaksi resurssiksi, kun he hakevat tietoa paikallisista uutisista ja ajankohtaisista tapahtumista. Konekääntimen tuottaman tekstityksen käytössä käyttäjät voivat hyväksyä jopa puutteellisen laadun, jos se mahdollistaa heille pääsyn haluamaansa tai tarvitsemaansa sisältöön (mt.).

Konekääntimiä voidaan hyödyntää myös ammattikäytäjäiden työssä esimerkiksi tuottamalla raakakäännös, jota sitten ihmiskääntäjä hioo julkaisukelpoiseksi käännökseksi. Tutkimuksissa on todettu, että tällainen konekäännöksen jälkieditointi voi nopeuttaa käännösprosessia säilyttäen samalla käännöksen laadun, edellyttäen, että käytetyt työkalut ovat sopivat ja alkuperäisteksti soveltuu konekäännettäväksi (Koponen 2017). Terribilen (2023) tutkimusartikkelissa esitellään laajamittainen tutkimus, jossa pyrittiin vertailemaan ihmiskääntämisen ja neuroverkkokääntämisen jälkieditoinnin käännös- ja tarkistusnopeuksia. Tutkimuksessa käytettiin data-analyysia, joka kattoi 879 kääntäjän kahden ja puolen vuoden ajalta 11 eri kieliparista kääntämiä tekstejä, yhteensä 90 miljoonaa sanaa. Tulokset osoittivat, että konekäännöksen jälkieditointi oli yleisesti ottaen nopeampaa kuin pelkkä ihmiskääntäminen, vaikka ihmisen käännösnopeus vaihtelikin suuresti. Käännösosalalla onkin esitetty rohkeita arvioita siitä, että jopa 75 % kääntäjien työtehtävistä saattaa olla konekäännösten jälkieditointia jo vuonna 2025 (Latomaa ja Nurminen 2022).

Konekääntäminen mahdollistaa myös verkkokeskustelujen seuraamisen eri kielillä, mikä voi auttaa muun muassa poliisia tunnistamaan mahdollisia rikoksia tai uhkia. Myös puolustusvoimat ja kansainväliset avustajajärjestöt kokevat konekääntimet hyödyllisiksi niiden tarjoamien mahdollisuuksien vuoksi, erityisesti kriisialueilla, joissa he kohtaavat usein vähemmän puhuttujen kielten edustajia (Schmalz 2019, 205). Konekääntimen merkitys korostuu myös kansainvälisten yritysten päivittäisessä toiminnassa erityisesti tilanteissa, joissa kaikki eivät välttämättä hallitse englantia kovin hyvin ja joissa esim. tuotantohäiriöistä tiedottaminen laajalle joukolla mahdollistaa nopean reagoinnin ongelmaan (Voutilainen 2021).

2.2.2 Konekääntämisen haasteita

Vaikka konekääntimet tarjoavat lukuisia mahdollisuuksia, on myös tilanteita, joissa niiden käyttö ei välttämättä ole suotavaa ja laatu ei ole odotetun mukaista. Esimerkiksi luovaa

kääntämistä vaativissa teksteissä, joissa esiintyy runsaasti tulkinnanvaraisia ilmaisuja ja kielikuvia, konekääntimien tarkkuus ja luotettavuus voivat olla kyseenalaisia (Backström 2020). Lisäksi erilaisten kohderyhmien ymmärtäminen voi olla haasteellista konekääntimille, mikä voi johtaa virheellisiin tai epätarkkoihin käännöksiin. Kone ei kykene ajattelemaan, kuka tekstiä tulee lukemaan ja millaista lisätietoa vastaanottaja saattaa tarvita (mt.).

Myöskään käyttöohjeiden, markkinointimateriaalien ja muiden julkisten asiakirjojen kääntämisessä kone ei pärjää, ja niiden kääntämisessä suositellaan edelleen ihmiskääntäjän käyttöä. Ihminen kykenee tuomaan tekstiin syvällisempiä merkityksiä ja tarvittavia yksityiskohtia, jotka ovat olennaisia esimerkiksi journalististen tai kaunokirjallisten tekstien tulkitsemisessä (Voutilainen 2021). Lisäksi silloin, kun käännettävä ohjeistus vaatii parempaa ymmärrettävyyttä tai on turvallisuuden kannalta kriittinen, on suositeltavaa jälkieditoida ja tarkistaa konekääntimen tuottama teksti ennen julkaisua (mt.). Puutteellinen laatu ei ole myöskään hyväksyttävää, kun konekäännöstä hyödynnetään kaupallisissa, poliittisissa tai oikeudellisissa ympäristöissä. Tällaisissa tilanteissa on ensisijaisen tärkeää, että käänнос on tarkka, koska yrityksen maine ja luotettavuus voivat olla uhattuina (Schmalz 2019, 196–197). Yleisillä kääntimillä tämän korkean laadun saavuttaminen voi olla vaikeaa, ja tarkat tulokset saattavat vaatia erikoistarkoitukseen kehitettyjä kääntimiä.

Konekääntimien, kuten DeepLin, käyttöön liittyy muun muassa tietoturva-aste: Niissä käännetty teksti saattaa tallentua kääntimen muistiin käyttäjän siitä tietämättä. Lisäksi monet ilmaisia käännospalveluita tarjoavat tahot voivat varata oikeuden käännettyihin materiaaleihin (Voutilainen 2021). Tämä tarkoittaa käytännössä sitä, että yleisesti saatavilla oleva konekäännin ei välttämättä ole turvallinen vaihtoehto tietoturvan näkökulmasta.

Suuren haasteen konekääntämiselle muodostaa myös yleisesti ottaen kielten monimutkaisuus, erityisesti leksikaalinen monitulkintaisuus ja syntaktinen epäselvyys (Somers 2012, 432–440). Kuten tiedämme, sanoilla voi olla eri merkityksiä kontekstista riippuen, ja tämä tekee kontekstiin sopivan sanamerkityksen valinnasta kääntimelle vaikeaa. On tärkeää huomioida, että haasteet, erityisesti kielelliset, voivat vaihdella merkittävästi sen mukaan, millaista käännosmenetelmää käytetään. Koska tutkimuksessani pyrin arvioimaan DeepL-konekäännöksen laatua, keskityn seuraavaksi tarkastelemaan nimenomaan DeepL-käännintä.

2.3 DeepL-konekäännin

DeepL-käännin kuuluu neuroverkkoja hyödyntäviin kääntimiin, joiden toimintaperiaatetta käsittelemme jo luvussa 2.1. Tässä luvussa keskitymme tarkastelemaan ainoastaan DeepL-kääntimen ominaisuuksia. DeepL-kääntimen tehokkuus perustuu laajaan ja monipuoliseen yleiskieliseen tietokantaan, jota se hyödyntää. Tämä tietokanta kattaa laajan kirjon eri teemoja ja aihealueita ilman erillistä koulutusta tietyille asiasisällöille tai tekstityypille (DeepL, ei vuotta). Näin ollen DeepL pystyy tuottamaan suhteellisen laadukkaita käännöksiä erityisesti yleisluonteisista teksteistä. Koska tutkimukseni aineisto koostuu talousaiheisen tietokirjan tekstikatkelmista, on mahdollista, että DeepL ei tunnista tai ota huomioon talouteen liittyvää terminologiaa tai tietokirjan kielellisiä vaatimuksia. Toisaalta, mikäli käsiteltävät talousasiat ovat yleistajuisia, DeepL saattaa kuitenkin selviytyä tehtävästä hyvin.

DeepL tukee tällä hetkellä hieman yli 30 kieltä ja kattaa lähes kaikki tunnetuimmat kielet, ja se tuottaa parempia käännöksiä silloin, kun kyseessä on eurooppalaiset kieliparit. Erityisen hyvin se soveltuu tietyntyyppisiin teksteihin, kuten käyttöohjeisiin, resepteihin ja yleisiin sopimusehtoihin, joissa käytetään usein samankaltaisia lauseita. Toisaalta luovien tekstin kohdalla DeepL saattaa johtaa käyttäjää harhaan ja vaatia enemmän vaivaa kuin käännös ilman sitä (Neidhardt 2020). DeepL-käännökset ovat yleensä sujuvia, kieliopillisesti oikeita ja usein jopa loogisia. Kuitenkin tekstin merkityksen välittymisessä saattaa olla puutteita ja virheitä. Virheiden havaitseminen saattaa vaatia useita tarkastuskertoja. (mt.)

Kokonaisuudessaan DeepL on tehokas ja hyödyllinen apuväline monenlaisten tekstien kääntämiseen, mutta sen käyttö vaatii tietoisuutta sen rajoituksista ja tarvittaessa huolellisuutta virheiden havaitsemiseksi.

3 Kieli ja tekstilaji neuroverkkokääntämisen näkökulmasta

Vaikka neuroverkkokääntimet pystyvät oppimaan monimutkaisia säännönmukaisuuksia ja kontekstuaalisia piirteitä kielestä suoraan suurten tekstikorpusten avulla, ne eivät vielä ole täydellisiä ja täysin luotettavia. Kuten aiemmista luvuista käy ilmi, neuroverkkokääntimet pystyvät tuottamaan tyydyttävän ja välillä erittäin hyvänkin tuloksen. Edellä todettiin, että esimerkiksi DeepL-konekäännökset ovat yleensä sujuvia, mutta tarkkuudessa ne eivät vielä ole luotettavia. Tarkkuusongelmiin on monia eri syitä.

Syy lienee luonnollisessa kielessä, jolla tarkoitetaan jonkin ihmisryhmän äidinkielenään käyttämää kieltä, joka on luonnollisen kehityksen tulos (Tieteen termipankki 30.4.2024: Kielitiede:luonnollinen kieli). Kielet syntyvät ja kehittyvät luonnollisesti satojen tuhansien vuosien ajan muovautuen erityisesti sanastonsa osalta soveliaiksi ilmaisemaan juuri sitä, mikä kulloisessakin kulttuuri- ja fyysisessä ympäristössä on ollut tarkoituksenmukaista (mt.). Tästä johtuu, että esimerkiksi saksankielisessä ympäristössä tietyn ilmiön nimeämiseksi keksitylle sanalle ei välttämättä löydy suoraa vastinetta suomenkielisessä kieliympäristössä, jos sellaista ilmiötä ei esiinnykään Suomessa. Juuri tällaisten tiettyyn kulttuuriin ja maanalueeseen sidottujen sanojen ja fraasien kääntäminen toisen kulttuurin kielelle saattaisi tuottaa ongelmia neuroverkkokääntimelle.

Kulttuuristen erojen lisäksi DeepL-käännöksen tarkkuuteen ja laatuun vaikuttavat kielipari ja niiden kielelliset piirteet sekä tekstilaji. Esimerkiksi saksa ja suomi kuuluvat eri kieliryhmiin, ja DeepL-käännösten laatu saattaa vaihdella, varsinkin kun suomi on kohdekielenä. Tähän ovat osittain syynä suomen kielen erityispiirteet. Koska tutkimuksessani tarkastelen DeepL-kääntimen tuotoksia saksasta suomeen käännettäessä, koen hyödylliseksi esitellä näiden molempien kielten erityispiirteitä, jotka saattavat tuottaa ongelmia konekääntimelle.

3.1 Suomen ja saksan kirjakielten erityispiirteitä

Suomen kieli kuuluu suomalais-ugrilaisen kielten ryhmään ja on yksi itämerensuomalaisista kielistä. Näissä kielissä ei ole kieliopillisia sukuja eikä artikkeleita, toisin kuin saksan kielessä. Saksa kuuluu indoeurooppalaiseen kielikuntaan ja germaanisiin kieliin, ja siinä käytetään sekä artikkeleita että kieliopillisia sukuja. Saksan kielessä substantiiveilla on kolme sukua: maskuliini, feminiini ja neutri. Kieliopilliset suvut ja artikkelit eivät kuitenkaan yleensä aiheuta ongelmia DeepL-kääntimelle. Esimerkiksi Zürichin ammattikorkeakoulun

toteuttamassa pilottitutkimuksessa analysoitiin 300 saksankielisen tieteellisen tiivistelmän ja artikkelin konekäännöksiä, jotka tuotettiin muun muassa ilmaisella DeepL-kääntimellä (Delorme Benites 2021). Yksi tutkimuksen havainnoista oli, että DeepL-kääntimen osalta kielioppi ja oikeinkirjoitus eivät yleensä tuottaneet ongelmia ja käännökset olivat näiden osalta virheettömiä. Kieliopillisten sukujen tulkitseminen ei siis yleensä aiheuta haasteita DeepL-kääntimelle, vaan hankaluuksina ovatkin enemmän suomen kielen rikas taivutusjärjestelmä, pitkät yhdyssanat, synteettisyys ja sanajärjestys.

Suomen kieli kuuluu niin kutsuttuihin synteettisiin kieliin, joille on tyypillistä, että kieliopilliset suhteet ilmaistaan sanoja taivuttamalla ja liittämällä niihin päätteitä (Vilokkinen 2017, 83). Tämä suomen kielen rikas taivutusjärjestelmä tekee suomen kielestä erikoisen ja haasteellisen konekääntimelle. Esimerkiksi nomineilla voi olla jopa 2 000 ja verbeillä yli 12 000 taivutusmuotoa. Erilaisten muotojen määrä johtuu suomen agglutinatiivisesta luonteesta: sanaan voidaan liimata suuri joukko taivutuspäätteitä ja muita affikseja, esimerkiksi *halu+tu+imm+i+lla+mme+ko* (Koskeniemi ym. 2020, 11). Taivutuspäätteen tulkinta syntaksin tasolla saattaa olla haastavaa konekääntimelle, esimerkiksi: *Poliisi tarkkaili miestä mäellä* ja *Poliisi tarkkaili miestä kiikarilla*. (Koskeniemi ym. 2020, 23). Sitä vastoin saksan kielessä taivutusjärjestelmä on suppeampi kuin suomen kielessä. Haasteita konekääntimelle voivat muodostaa saksan kielessä pikemminkin eriävät yhdysverbit ja niiden etuliitteet, jotka ovat tyypillisiä saksan kielelle. Suomen kielessä sellaisia verbejä ei ole. Esimerkiksi jos saksankielisessä pitkässä virkkeessä käytetään eriävää yhdysverbiä ja sijoitetaan sen etuliite kauas verbistä, konekääntäminen voi hankaloitua entisestään. Esimerkiksi verbi *vorstellen* voi esiintyä seuraavanlaisissa lauseissa: *Er stellte sich, nachdem er mir ein Getränk angeboten hatte und wir ins Gespräch gekommen waren, vor* (Burchardt ym. 2020, 12). Verbin etuliitteet yksinään tuskin tuottavat vaikeuksia DeepL-kääntimelle. Haasteita syntyy vasta silloin, kun niitä käytetään pitkissä virkkeissä, jolloin riski etuliitteen virheellisestä tulkinnasta kasvaa.

Vilokkinen (2017, 83) korostaa, että synteettiselle kielelle on tyypillistä myös pyrkimys tiiviyyteen. Hänen mukaansa suomessa se ilmenee muun muassa lauseenvastikkeina ja muina rakenteina, joilla korvataan sivulauseita, kuten partisiippiattribuuttirakenteet, agenttipartisiippi ja lauseenvastikemainen komitatiivi. Vilokkisen (mt., 84) mukaan tiiviyyteen liittyy myös se, että suomi karsastaa paranteeseja eli virkkeen sisään upotettuja irrallisia huomioita. Tämä suomen kielen pyrkimys tiiviyyteen saattaa olla haasteellista DeepL-kääntimelle. Jos konekäännin ei onnistu tiivistämään tekstiä suomen kielen tyylille ominaiseen tapaan,

käännöksestä tulee kömpelö, mikä voi vaikuttaa lukukokemukseen. Sitä vastoin saksan kielelle parenteesit ovat hyvin tyypillisiä. Saksan kielen monimutkaisuus näkyy myös sen taipumuksessa käyttää pitkiä sisäkkäisiä lauseita, jotka voivat olla haasteellisia konekääntimille.

Taivutuspäätteiden ja tiiviiden lisäksi suomessa on taipumus muodostaa pitkiä yhdyssanoja, mikä voi aiheuttaa haasteita neuroverkkokääntimelle. Yhdyssanat koostuvat usein yksittäisistä sanoista, jotka on ymmärrettävä kokonaisuutena, jotta ne voidaan kääntää oikein. On myös tärkeää huomata, että pitkiä yhdyssanoja kirjoitetaan yhdeksi sanaksi usein ilman välilyöntejä tai yhdysmerkkejä, mikä vaikeuttaa tilannetta entisestään. Lisäksi yhdyssanoista voidaan edelleen muodostaa uusia yhdistelmiä (Koskenniemi ym. 2020, 11). Teoriassa yhdyssanan pituudelle ei ole rajoja. Konekääntimet saattavat kuitenkin kohdata vaikeuksia näiden sanojen pilkkomisessa ymmärrettäviin osiin, sillä ne eivät aina tunnista sanojen rajoja. Yhdyssanat voivat johtaa virheellisiin käännöksiin, esimerkiksi jos yhdyssanaa ei ole sanakirjassa (Delorme Benites 2021). Myös saksan kielessä on tapana muodostaa pitkiä yhdyssanoja. Saksan kielessä on mahdollista luoda äärettömän pitkiä sanoja, sillä sananmuodostusmahdollisuuden ansiosta useita sanoja ja liitepäätteitä voidaan helposti yhdistää (Burchardt ym. 2020, 12). Konekääntimet eivät aina kykene tulkitsemaan pitkien sanojen merkitystä oikein kontekstin perusteella.

Seuraava suomen kielelle tyypillinen ominaispiirre on sanajärjestys. Suomen perussanajärjestys on sama kuin useimmissa muissakin kielissä: ensin tulee subjekti, sitten predikaatti ja sitten objekti. Suomen kieli voi kuitenkin käyttää joustavaa sanajärjestystä, mikä voi tehdä konekääntämisestä haastavaa, kun on ymmärrettävä, mikä osa lauseesta viittaa mihinkin (Koskenniemi ym. 2020, 11). Suomen sanajärjestys vaihtelee sen mukaan, mikä on lauseen informaatorakenne (mt.). Informaatorakenne näkyy suomen kielessä siten, että ensin ilmoitetaan tuttu tieto ja kerrotaan siitä sitten jotakin uutta (Vilokkinen 2017, 85). Asian painotuksen tulisi osua kohdalleen, jotta lauseen ymmärtäminen helpottuisi. Suomessa lauseen loppu tapaa saada enemmän painoa, ja sinne täytyy siis sijoittaa painavaa asiaa eikä mitään sivuseikkaa (mt., 85). Tämä saattaa olla hankalaa DeepL-kääntimelle, jos se pyrkii säilyttämään alkuperäisen kielen sanajärjestyksen, jolloin olennaiset asiat eivät välttämättä välity oikein. Saksan kieli eroaa tässä suhteessa suomen kielestä, koska saksassa sanajärjestys on melko kiinteä. Saksan kielessä uusi ja vanha asia esitetään usein epämääräisen tai määräisen artikkelin avulla. Saksankielinen lause tai virke noudattaa usein tarkoin määriteltyä

sanajärjestystä: Päälauseen sanajärjestys on usein: *subjekti + predikaatti + objekti*, mutta sivulauseessa kaikki verbit kuuluvat lauseen loppuun. Suomessa sanajärjestys pysyy yleensä samana riippumatta siitä, onko kyseessä pää- vai sivulause. Tällaiset erot voivat aiheuttaa vaikeuksia konekääntimelle.

Kun olen tarkastellut molempien kielten DeepL-konekääntämisen kannalta hankalia piirteitä, esittelen seuraavaksi lyhyesti tietokirjan ominaispiirteitä. Kuten jo aiemmissa luvuissa totesin, neuroverkkokääntimet, kuten myös DeepL, pärjäävät muun muassa käyttöohjeiden kääntämisessä erittäin hyvin. Sitä vastoin tietokirjan kääntämisessä DeepL-kääntimellä saattaa ilmetä ongelmia johtuen esimerkiksi tietokirjan sisällöllisistä piirteistä ja käyttötarkoituksesta.

3.2 Tietokirja tekstilajina

Tietokirjallisuus on tietopainotteista ja tarkoitettu yksityiskäyttöön (Schikowski 2011, 66). Tietokirjat määritellään yleensä niiden sisällön ja käyttötarkoituksen mukaan. Sisältö on tietoa ja informaatioaiheista, eikä se ole toimintapainotteista. Koska tietokirjat voivat kuitenkin käsitellä kaikkia yhteiskunnallisia aiheita ja reagoida nopeasti yhteiskunnan uusiin ongelmiin, tietokirjan määritelmä on aina joko liian laaja tai liian suppea sisällön suhteen. Tietokirjat voivat käsitellä esim. taloutta, politiikkaa ja ajan historiaa, ja niiden kohderyhmänä ovat yleensä maallikot, jotka haluavat syventyä tiettyyn aiheeseen. (mt.) Lukijat voivat olla kaikenlaisia, kouluja käyneitä tai käymättömiä ihmisiä, joten tietokirjan on oltava ymmärrettävissä ilman syvällisiä pohjatietoja.

Schikowskin (2011, 67) mukaan tietokirjallisuus määritellään useammin sen tehtävän mukaan. Tietokirjojen tehtävänä on esittää monimutkaisia asiayhteyksiä selkeällä kielellä, mikä saavutetaan muun muassa erikoisalojen termejä selittämällä ja ilmaisemalla yksityiskohdat lyhyesti ja ytimekkäästi. Tämä tekee tietokirjoista houkuttelevia erityisesti niille, joilla ei ole syvällistä tietoa aiheesta. Rajalan (2013) mukaan on tärkeää, että tietokirjassa esiintyvä tieto pitää paikkansa, mutta samalla tietokirjan tyyli on tarinankerrontaa muistuttavaa ja subjektiivisuus on sallittua.

Vilokkinen (2017, 16) korostaa, että tietokirjan käännökseen laadun määrittelevät käännökseen luontevuus ja sujuvuus, jotka syntyvät suomenkielisistä rakenteista ja ilmaisuista sekä monipuolisesta sanastosta. Oleellista on myös, että tekstin punainen lanka säilyy ja että käännos on sekä sanastollisesti että rakenteellisesti monipuolinen eikä orjallisesti toista

lähdekielen ilmaisutapoja. Hyvä käännös on helppolukuinen, tarkka mutta ei jäykkä, ja se pyrkii välittämään alkuperäisen tekstin ajatuksia eikä vain sanoja.

Tietokirjoilta odotettavien ominaisuuksien takia niiden suomentaminen ei ole yleensä yksinkertaista. Niiden suomentamisessa täytyy huomioida tietokirjalta toivotut rakenteet, kielelliset ominaisuudet ja oletetun lukijakunnan odotukset. Tutkimuksessani en kykene arvioimaan kaikkia näitä ominaisuuksia kattavasti ja objektiivisesti, joten laadunarvioinnissa keskityn pääasiassa tarkastelemaan tietokirjan kielellisiä piirteitä.

4 Konekäännöksen laadunarviointi

Käännösten laadunarviointi on olennainen osa käännösprosessia, sillä sen avulla voidaan seurata konekääntimien käännöslaadun kehittymistä. Konekäännösten arviointi ei ole kuitenkaan yksinkertaista, eikä siihen ole olemassa yhtä oikeaa menetelmää. Arviointia hankaloittaa esimerkiksi se, että samalle tekstikohdalle löytyy yleensä useampia toimivia käännösvaihtoehtoja (Burchardt ym. 2020, 203). Eri käännösvaihtoehtojen valinta riippuu monista tekijöistä, kuten kääntäjistä, kohderyhmästä, käytettävissä olevasta ajasta ja kieliparista (Popović 2018, 130). Käännöksen laatu ei ole myöskään muuttumaton käsite, vaan se määräytyy pitkälti käännöstehtävän vaatimuksen mukaan (Burchardt ym. 2020, 203). Käännös voi olla tarkka ja laadukas, mutta silti hyödytön, jos se ei noudata tiettyjä vaadittuja kriteereitä. Hyväksyttävien käännösten kirjo on siis laaja. Konekäännösten arvioinnin monimutkaisuutta on tutkittu laajasti, ja siitä on esitetty useita suositeltuja käytäntöjä ja menetelmiä, joihin palaan myöhemmissä luvuissa.

4.1 Laadukas käännös ja käännöksen laatu

Ennen kuin syvennyn konekäännöksen laadunarviointimenetelmien kuvaamiseen, on olennaista määritellä, mitä *laadukkaalla käännöksellä* ja *käännöksen laadulla* tarkoitetaan. Käännöksen laadulle ei ole vakiintunutta määritelmää, joka kattaisi koko ihmis- ja konekääntämisen kirjon ja toimisi luotettavana mittarina laadun arvioinnissa. Melbyn (2012, 5) mukaan käännöksen laatu riippuu muun muassa siitä, miten kääntäminen määritellään, mitä käännökseltä odotetaan ja mitä kohdetta arvioidaan. Konttinen ja Veivo (2008, 6) korostavat, että arvioinnin kohteet käännösosalalla voivat vaihdella ja kohdistua eri tasoille, kuten 1) organisaation tai yksilön käännös- ja käännättämistoimintaan, 2) kääntäjän suoritukseen, 3) käännösprosessiin tai 4) itse käännökseen.

Koska valitsin tutkimukseni arviointikohteeksi itse käännöksen, arvioin tässä työssä laatua nimenomaan käännöksen näkökulmasta ja tarkastelen ainoastaan arviointimenetelmiä ja lähestymistapoja, jotka ovat merkittäviä nimenomaan itse käännöksen arvioinnissa.

Tarkastelemani menetelmät ovat sovellettavissa sekä ihmisen tekemän että konekäännösten arvioinnissa.

4.1.1 Funktionaalisuus eli toimivuus laadun kriteerinä

Käännöksen laatu saattaa olla sidoksissa siihen, miten tehokkaasti käännös palvelee tiettyä käyttötarkoitusta, vastaa asiakkaan odotuksia tai noudattaa lähdetekstin sisällöllisiä ja muodollisia piirteitä. Tiittula (2014, 3) korostaa, että käännöksellä on aina tietty käyttötarkoitus eli skopos, ja onnistunut käännös palvelee juuri tätä tarkoitusta. Hänen mukaansa esimerkiksi artikkeli syöpähoidosta vaatii erilaista käännöstapaa lääketieteelliseen julkaisuun verrattuna yleistajuisempaan terveystieteeseen, ja käännöksen on sopeuduttava kohdeyleisön lääketieteelliseen tietämykseen. Käännöksen laatu riippuu siis kontekstista, jossa sitä käytetään, ja käännöksen ei tarvitse olla kaikin puolin täydellinen täyttääkseen tehtävänsä.

Skoposteorian ovat kehittäneet alun perin Katharina Reiss ja Hans J. Vermeer, ja se perustuu toimintateoriaan (Nord 2010, 121). Siinä käännöstä pidetään tavoitteellisena toimintana: Käännöksen tarkoituksena on toimia välittäjänä eri kulttuuriyhteisöjen jäsenten välillä, ja arvioitaessa käännöksen laatua on ensisijaisen tärkeää huomioida oletettu tai ilmoitettu käyttötarkoitus. Skoposteoriassa käännöstä pidetään onnistuneena ja laadukkaana, kun se täyttää tarkoituksensa ja tehtävänsä. Siinä käännöksen arvioinnin mittapuuna ei siis ole lähdeteksti, vaan kohdetekstin toimivuus tai sopivuus (Nord 2010, 122). Myös Melbyn (2012, 13) mukaan joskus riittää, että laadukas käännös täyttää ainoastaan kaikki neuvotellut vaatimukset eli spesifikaatiot ottaen huomioon loppukäyttäjän tarpeet. Jos esimerkiksi loppukäyttäjän tarve on saada sujuva käännös mahdollisimman nopeasti, niin loppukäyttäjä voi hyväksyä sisällöllisesti epätarkankin käännöksen. Tässä tilanteessa on kuitenkin muistettava, missä yhteydessä käännöstä tullaan käyttämään. Tietyissä konteksteissa, kuten terveydenhuoltomateriaalien kääntämisessä, nopeus ei aina voi korvata tarkkuuden puutetta.

Tutkielmassani aineistona käytin talousaiheista tietokirjaa, ja sen käännöksen laadun arvioimisessa yhdeksi kriteeriksi valitsin nimenomaan funktionaalisuuden näkökulman. Tietokirjan odotettu käyttötarkoitus on palvella lukijaa tarjoamalla selkeää ja olennaista tietoa. Mikäli talousaiheet on ilmaistu tarkasti ja sujuvasti, teksti on vaivatonta lukea, ja se täyttää tarkoituksensa palvella lukijaa. Vaikka funktionaalisessa näkökulmassa ei ole tarkemmin määritelty tämän teorian soveltuvuutta konekäännösten arvioinnissa, väitän sen soveltuvan myös konekäännösten laadun arviointiin, sillä myös konekäännösten käyttäjinä on tietty ihmisryhmä.

4.1.2 Ekvivalenssi eli vastaavuus laadun kriteerinä

Funktionaalisuuden lisäksi valitsin toiseksi tietokirjan käännösten laatua määritteleväksi kriteeriksi ekvivalenssin eli vastaavuuden. Tiittulan (2014, 7) mukaan lähdetekstin ja kohdetekstin välillä oletetaan vallitsevan jonkinlainen vastaavuus. Myös Yinhua (2011, 169–170) korostaa, että käännöksen vastaavuus on välttämätöntä. Hän perustelee sen sillä, että kääntäminen on pohjimmiltaan viestinnällistä toimintaa ja yksi viestinnän perusvaatimuksista on varmistaa, että viesti siirtyy alkuperäisestä lähteestä tarkoituksenmukaisesti ja ymmärrettävästi vastaanottajalle. Tämä periaate pätee myös kääntämiseen: kääntäjän on pyrittävä mahdollisimman tarkasti toistamaan alkuperäistekstin merkitys kohdetekstissä, jotta lukija ymmärtäisi lähteen viestin asianmukaisesti (mt., 170). Ilman vastaavuutta kääntäminen ei täyty viestinnän tavoitetta, ja ilman tietynasteista tai tietyiltä osin ilmenevää vastaavuutta käännettyä tekstiä ei voida pitää onnistuneena käännöksenä alkuperäisestä tekstistä.

Yinhua (2011, 171) kuitenkin painottaa, että vastaavuuden saavuttaminen ja alkuperäisen tekstin koko sanoman siirtäminen kohdetekstiin koko tekstin osalta on yksinkertaisesti mahdoton tehtävä. Hän ehdottaa, että vastaavuus tulisi ymmärtää vain eräänlaisena samankaltaisuutena tai lähentymisenä. Tämä tarkoittaa, että lähde- ja kohdetekstin välinen vastaavuus voidaan todeta eri tasoilla ja eri näkökohdissa. Tiittulan (2014, 7) mukaan tekstit voivat vastata toisiaan esimerkiksi sisällöllisesti, tyylillisesti, muodollisesti tai niillä voi olla vastaanottajaan sama vaikutus. Yleensä kaikkia vastaavuuden tasoja ei voi saavuttaa, vaan on mietittävä, mikä kulloisessakin tilanteessa on tärkeintä. Esimerkiksi jos tekstillä halutaan saada sama vaikutus aikaiseksi, se ei aina muodoltaan voi kovinkaan tarkasti noudattaa lähdetekstiä. (mt., 7)

Tutkimuksessani olen kiinnostunut pääasiassa sisällöllisen merkityksen välittymisestä, tarkemmin sanottuna semanttisesti ekvivalentista käännöksestä. Tässä yhteydessä semanttinen tarkoittaa siis sisältöä tai merkitystä, ja ekvivalentti viittaa samankaltaisuuteen tai vastaavuuteen. Semanttisesti ekvivalentissa käännöksestä ei keskitytä sanojen, lauseiden ja kielen kääntämiseen, vaan lähdetekstin ymmärtämiseen ja käännöksen tuottamiseen kohdekielelle ominaisten menetelmien avulla (Roslöf ja Veitonen 2006, 113). Toisin sanoen semanttisesti ekvivalentti käännos välittää alkuperäisen tekstin sanoman, tarkoituksen ja merkityksen kohdekielellä, vaikka sanat tai ilmaisut voisivatkin olla erilaisia.

Koska tietokirjan tarkoitus on välittää uutta tietoa, sanojen ja ajatusten merkitys on käännettävä tarkasti kohdekielelle. Semanttisesti riittävällä ekvivalenssilla varmistetaan, että alkuperäisen tekstin sisältö säilyy mahdollisimman tarkkana, mikä tekee tietokirjasta sekä houkuttelevamman että helpommin luettavan. Tästä syystä ekvivalenssi on mielestäni keskeistä tietokirjojen kääntämisessä.

4.2 Laadunarviointimenetelmät

Käännöksen laadun määrittelyyn vaikuttavat siis monet tekijät, eikä sille ole vakiintunutta määritelmää. Melbyn (2012, 7) mukaan on kuitenkin tärkeää saavuttaa yhteisymmärrys laadun käsitteestä ennen arvioinnin aloittamista, sillä ilman sitä arviointia ei voi tehdä.

Käännöksen laatua voidaan arvioida monin eri tavoin, ja arviointimenetelmät vaihtelevat sen mukaan, millaiset laatuvaatimukset eli spesifikaatiot käännökselle asetetaan, ja näiden vaatimusten tulisi vastata loppukäyttäjien odotuksia ja tarpeita (Melby 2012, 19).

Laatuvaatimusten lisäksi on olennaista määrittää, mihin tarkalleen arviointi kohdistetaan. Kuten jo aiemmin todettiin, arvioinnin fokus voi olla käännöksessä, käännösprosessissa tai koko käännösprojektissa. Koska oman tutkimukseni arviointikohde on kääntämisen tuote, niin seuraavaksi tarkastelen nimenomaan sellaisia yleisiä lähestymistapoja, joita voidaan käyttää, kun laadun arviointikohteena on itse käännös. Tutkielmassani en tule käyttämään kaikkia arviointimenetelmiä, sillä en tarvitse niitä kaikkia omille käännöksille asetettujen laatuvaatimusten arvioinnissa. Esittelen nämä yleisimmät menetelmät vain perustellakseni oman lopullisen valintani valitun laadunarviointimenetelmän osalta.

4.2.1 Manuaalinen ja automaattinen lähestymistapa

Yksi tapa arvioida ja mitata konekäännöksen laatua on ihmisen suorittama arviointi eli manuaalinen arviointi. Kääntämisen ja kielialan asiantuntijat saattavat arvioida konekäännöksen laatua kahdesta näkökulmasta. Manuaalisessa arvioinnissa tämä prosessi liittyy yleensä käännöksen tarkkuuden ja sujuvuuden arviointiin (Rivera-Trigueros 2022, 593–619). Tarkkuus tarkoittaa semanttisen laadun arviointia eli sitä, onko käännöksessä välitetty tieto oikea vai ei. Tämä edellyttää vertailua joko vertailukäännöksiin (yksikielinen arviointi) tai alkuperäistekstiin (kaksikielinen arviointi). Sujuvuus puolestaan kohdistuu syntaktiseen laatuun; tässä tapauksessa vertailu alkuperäistekstiin on tarpeetonta, ja arviointi suoritetaan yksikielisesti (mt.). Sujuvuuden näkökulma keskittyy kohdetekstin ja kohdekielen

normien noudattamiseen, kuten kieliopillisiin ja selkeyteen liittyviin piirteisiin (Maučec & Donaj 2019, 149–151).

Myös Popović (2018, 130) tuo esille tarkkuuden ja sujuvuuden kriteerit ehdottaessaan manuaalisia tapoja arvioida konekäännöksiä. Hän esittää kaksi arviointitapaa. Ensimmäisessä menetelmässä lähtö- ja kohdekieliä ymmärtävä ihmisarvioija antaa käännökselle pisteitä asteikolla 1–5 erilaisten ennalta määriteltyjen kriteerien perusteella. Näitä kriteereitä voivat olla esimerkiksi tarkkuus ja sujuvuus, ja näiden kahden pohjalta määritellään lopulta käännöksen kokonaislaatu. Toisessa menetelmässä eri konekäännöksiä vertaillaan keskenään ilman erillistä pisteytystä, ja ne asetetaan järjestykseen parhaasta huonoimpaan.

Maučec ja Donaj (2019, 149–151) painottavat, että ihmisen suorittama arviointi on aikaa vievää, työlästä, kallista ja luonteeltaan subjektiivista. Heidän mukaansa subjektiivisuusongelmaa on yritetty lieventää yleisesti pyytämällä useampia ammattikäytäjiä arvioimaan samat käännökset. Näiden mainittujen haasteiden vuoksi on kehitetty automaattisia arviointimenetelmiä, joiden avulla voidaan arvioida konekäännösten laatua ilman ihmisen osallistumista.

Escriven (2019, 36) mukaan automaattiset mittarit ovat tehokkain ratkaisu konekäännösten arvioinnissa, koska ne ovat objektiivisia, nopeita ja edullisia verrattuna ihmisen tekemään arviointiin. Automaattisessa arvioinnissa konekäännöksen laatua verrataan referenssikäännökseen, joka on arvioitu hyvälaatuiseksi. Koska ihmiskäännöksissä on merkittävää vaihtelua, on olennaista, että jokaiseen arvioitavaan konekäännökseen on saatavilla useita ihmisen tekemiä käännöksiä (Maučec ja Donaj 2019, 147–151).

Konekäännösten arvioimiseen on kehitetty lukuisia automaattisia mittareita, kuten TER, METEOR, BLEU ja NIST. BLEU on näistä suosituin, koska sen katsotaan tuottavan erittäin tarkkoja tuloksia, jotka korreloivat vahvasti ihmisten tekemien arviointien kanssa (Shterionov yms. 2018, 217).

BLEU-mittarin avulla arviointi toteutetaan vertaamalla konekäännöstä referenssikäännökseen. Yksi tai useampi konekäännin tuottaa käännöksen, ja algoritmi analysoi erot referenssikäännöksen ja eri konekäännösten välillä. Tämän jälkeen saadaan BLEU-pistemäärä, joka esitetään lukuna 0:n ja 1:n välillä tai prosentteina 0–100 %. BLEU-mittarissa 100 % kuvastaa täydellistä käännöstä, mikä tarkoittaa identtisyyttä referenssikäännöksen

kanssa, kun taas 0 % edustaa täysin huonoa käännöstä, jolla ei ole mitään yhtäläisyyksiä referenssikäännöksen kanssa (Melby 2012, 20). BLEU-pistemäärä ei anna arviota siitä, onko käännös hyvä vai huono, se ainoastaan mittaa samankaltaisuutta konekäännöksen ja referenssikäännöksen välillä. On mahdollista, että konekäännös on suoriutunut hyvin, mutta saanut alhaisen pistemäärän, jos se käytti erilaisia ilmaisuja kuin vertailukäännös. Jos saadut pisteet on laskettu huonon referenssitekstin perusteella, tämä ei myöskään anna tietoa konekäännöksen laadusta, vaan se kertoo ainoastaan siitä, että konekäännös vastaa heikkoa referenssikäännöstä.

BLEU-mittarin suurin haaste on oletus, että jokainen teksti voidaan kääntää hyvin vain yhdellä tavalla. Lähtöteksteille voi kuitenkin olla useita sopivia käännösvaihtoehtoja. Koska tutkimusaineistoni on tietokirjallisuus, jossa saattaa usein olla monia oikeita käännösvaihtoehtoja, BLEU-mittari ei sovellu minun tutkimukseni tarpeisiin. Lisäksi automaattiset mittarit eivät tunnista virheiden vakavuutta (Silva ym. 2024, 1–2), mikä myös tekee niistä riittämättömiä käännösvirheiden analysoinnissa.

4.2.2 Holistinen ja analyttinen lähestymistapa

Manuaalisessa arvioinnissa käännöksen laatua voidaan arvioida joko holistisesti eli kokonaisvaltaisesti tai analyttisesti. Nämä lähestymistavat mahdollistavat arvioinnin tarkastelemalla arvioitavaa tekstiä tai käännöstä joko kokonaisuutena tai segmentteittäin.

Melbyn (2012, 19) mukaan *holistisessa arvioinnissa* keskitytään kolmeen keskeiseen osatekijään: lähdetekstiin, kohdetekstiin ja laatuvaatimukseen eli spesifikaatioihin. Arvioija arvioi käännöstä kokonaisuutena suhteessa laatuvaatimukseen eli spesifikaatioihin käyttäen asteikkoa, jossa vaihtoehdot ovat hyvä, huono tai keskinkertainen. Snow (2015, 15) korostaa, että kokonaisvaltainen lähestymistapa tarjoaa yleensä nopean arvion siitä, kuinka hyvin teksti on käännetty, mutta ei välttämättä anna riittävän yksityiskohtaista tietoa käännöstyön kohdennettuun parantamiseen.

Analyttinen arviointi puolestaan keskittyy yksityiskohtaisiin tekstin osiin, kuten sanoihin ja lauseisiin, ja paljastaa konkreettisesti, missä kohdassa tekstiä käännösvirheet sijaitsevat (Snow 2015, 15). Analyttinen arviointi perustuu samoihin kolmeen keskeiseen elementtiin eli lähdetekstiin, kohdetekstiin ja spesifikaatioihin, ja siinä käytetään huomattavasti yksityiskohtaisempaa mittaria (Melby 2012, 19). Tämä mittari sisältää yleensä valittuja ja

painotettuja virheluokkia, ja virheet voidaan merkitä aina sanatasolle saakka. Virheet myös luokitellaan niiden vakavuusasteen mukaan. Yhdistelmä virheiden määrästä ja niiden vakavuudesta tuottaa numeron, joka kuvastaa, kuinka kaukana käänнос on laadukkaasta käännoksestä eli siitä, että käännos täyttää sille asetetut laatuvaatimukset. (mt.)

Virhekeskeinen arviointi tuo esille konkreettisia ongelmakohtia käännoksessa. Se auttaa keskittymään virheiden korjaamiseen ja parantamaan käännoksen yksittäisiä puutteita.

Vaikka analyttinen arviointi auttaa tunnistamaan käännoksen virheelliset kohdat tarkasti, sopivan mittarin valinta voi olla haastavaa.

4.2.3 Yksiulotteinen ja moniulotteinen lähestymistapa

Virhekeskeisessä arvioinnissa käännostekstejä voidaan tarkastella yhden tai useiden erilaisten virhetyyppien suhteen. Siinä virheet jaetaan virhetyyppihin, jotka muodostavat ulottuvuuksia. Arvioinnissa voidaan joutua ottamaan huomioon tai jättämään pois kokonaisia ulottuvuuksia käännoksen tarpeiden perusteella (Snow 2015, 16). Käännosten arvioijat voivat keskittyä yhteen virhetyyppiin, jolloin puhutaan yksiulotteisesta lähestymistavasta, tai he voivat keskittyä useisiin eri virhetyyppihin, jolloin kyseessä on moniulotteinen lähestymistapa.

On olemassa erilaisia virheluokittelumalleja, mutta tutkielmassani käytin ainoastaan MQM-virheluokittelumittaria sen moniulotteisuuden ja joustavuuden takia. Valitsin MQM-laatumittarin myös sen takia, että se mahdollistaa virhekeskeisen arvioinnin, ja sen avulla analysoin DeepL-käännoksen laatua manuaalisesti. Seuraavaksi esittelen MQM-mittarin ja sen keskeiset piirteet.

4.3 Moniulotteinen MQM-käännoslaatumittari

Käännosten laadun arviointiin on kehitetty erilaisia menetelmiä, mutta nämä menetelmät eivät olleet standardoituja ja yleisesti hyväksytyjä. Tämä tarkoittaa sitä, että subjektiivisuus ja epä johdonmukaisuus ovat leimanneet käännosten laadunarviointiprosessia (Lommel 2018, 111–112). 1990-luvulla on alettu kehittämään menetelmiä, jotka olisivat objektiivisia ja yhdenmukaisia käännosten laadun arvioinnissa. Ensimmäisten joukossa olivat muun muassa 1900-luvun lopulla julkaistut SAE J2450 TQM ja LISA QA-malli. Nämä mallit perustuivat virheiden laskemiseen ja luokitteluun, ja niiden taustalla oli halu luoda nopea ja tehokas tapa valvoa käännosten laatua. Valitettavasti näissä virheluokittelumaleissa oli muutamia

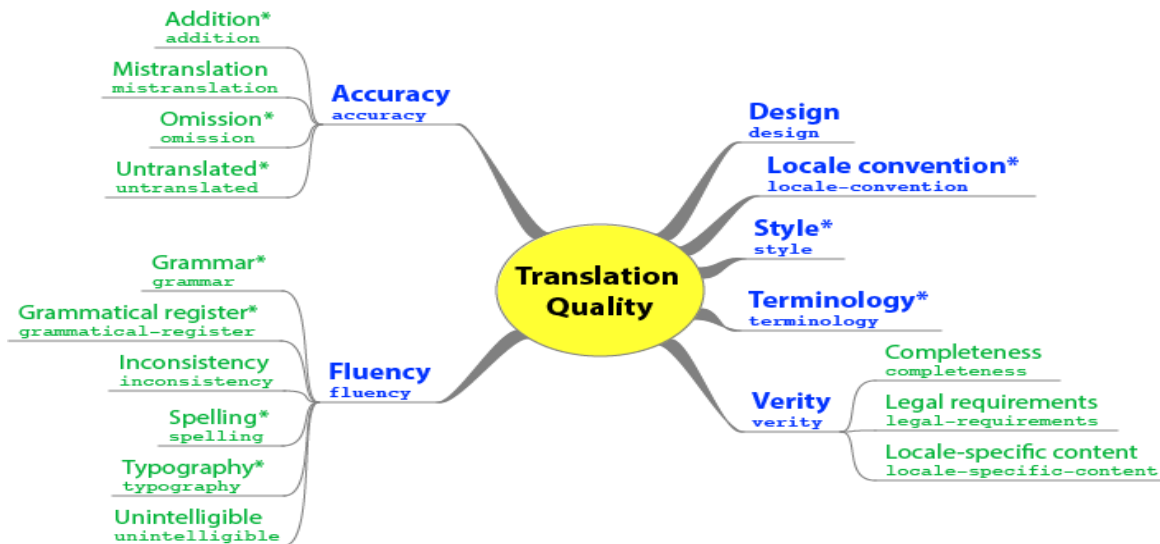
ongelmia: Niissä käytettiin joskus eri nimiä samoille virheille, ne luokittelivat virheet eri tavoin, ne eivät kattaneet kaikkia mahdollisia virheitä ja ne eivät sisältäneet uusia virheluokituksia, joita tarvittiin esimerkiksi konekäännösten arvioimiseen (Lommel 2018, 112–113). Kaikki tämä vaikeutti käännöksen laadun arviointia ja tietojen vertailua. Kun käännösmäärät kasvoivat ja tarve saada käyttökelpoista tietoa lisääntyi, näihin edellä mainittuihin ongelmiin kiinnitettiin entistä enemmän huomiota, ja vuonna 2015 Euroopan unionin rahoittamassa QT21-hankkeessa luotiin moniulotteinen laatumittari MQM. Tämä mittari oli hyvin kattava, ja sen kehittämisessä otettiin huomioon aiemmat mittarit.

MQM on käännöslaadun arviointikehys, jonka avulla arvioijat voivat mukauttaa omia mittareitaan laadunarviointiin. Tämä arviointimalli on alun perin kehitetty konekäännösten laadun arviointia varten, mutta sitä käytetään nykyään myös ihmisten tekemien käännösten arviointiin (Hagemann 2022, 53). MQM:n tavoitteena on mahdollistaa käyttäjille tasapuolinen ja puolueeton vertailu sekä kone- että ihmiskäännösten välillä ja samalla välttää monien nykyisten luokittelumenetelmien subjektiivisuutta keskittymällä tiettyihin virhetyyppeihin. MQM:n avulla voidaan luokitella käännöksiä virhetytologian kriteerien mukaisesti, ja virheiden määrien ja niiden vakavuusasteen pohjalta arvioida, kuinka laadukkaita käännökset ovat. Virheiden luokittelun ja pisteytyksen tuloksena syntyy niin sanottu laatumittari, joka auttaa määrittämään, onko käännös onnistunut ja sopiiko se tarkoitukseensa (Lommel 2018, 121–122). MQM-mittarin avulla käännösten laatu voidaan siis määrittää numeerisesti.

MQM-virhekategorioiden avulla tunnustetaan mahdolliset ongelmat kohdetekstissä, jotka voivat johtua virheistä tai tarkoituksellisista muutoksista. Esimerkiksi, jos kohdetekstin sisältö eroaa lähdetekstistä, kyseessä voi olla tarkkuusvirhe, mutta poikkeama voi myös olla tahallinen. MQM-mittari mahdollistaa näiden kohdetekstin ja lähdetekstin vertailun niin, että poikkeamat tunnustetaan. Tämä auttaa arvioimaan, kuinka tarkasti alkuperäisen tekstin merkitys on säilynyt käännöksessä. MQM-mallia ei voida käyttää täysimääräisesti kaikissa käännöstoissa, koska sen käyttäminen on työlästä. Snowin (2015, 17) mukaan tarkasteltavaksi otettavien virheluokkien määrä määräytyy käännöksen laatuvaatimusten pohjalta. Hagemann (2022, 63) korostaa, ettei ole olemassa yleispätevää sääntöä siitä, mitkä MQM-mallin osat alueet on otettava huomioon käännösten laadun arvioinnissa. Ne määräytyvät aina tilanteen ja arvioitavien asioiden perusteella.

MQM-mittari sisältää yhteensä seitsemän korkean tason ulottuvuutta eli yläkategoriaa: tarkkuus, sujuvuus, terminologia, tyyli, paikallinen konventio, muotoilu ja asianmukaisuus.

Jokainen näistä ulottuvuuksista sisältää tarkempia virheen alatyyppejä eli alakategorioita, jotka kattavat erilaisia laatuun liittyviä ongelmia. Esimerkiksi yläkategoriaan *tarkkuus* kuuluvat sellaiset alakategoriat, kuten lisäys, käännösvirhe ja poisto. On tärkeää myös huomioda, että MQM-laatumittarista on saatavilla kaksi versiota: laajempi MQM Full ja suppeampi MQM Core. MQM Full tarjoaa yksityiskohtaisemman virhetyyppien valikoiman, jota voidaan käyttää silloin, kun käänöksiltä vaaditaan suurempaa tarkkuutta. MQM Core puolestaan rajoittuu suppeampaan valikoimaan virhekategorioiden. Tutkielmassani käytin MQM-mittarin suppeampaa versiota, sillä sen tarjoamat virhetyypit olivat riittäviä tutkimuskysymyksiini vastaamiseksi. Valitsemani yläkategoriat ja alakategoriat esittelen tarkemmin *Tutkimusmenetelmät* -osiossa. MQM Coren rakenne voidaan esittää graafisesti seuraavasti:



Lähde: <https://sites.miis.edu/runyul/2018/03/04/translation-quality-assessment-mqm-multidimensional-quality-metrics/> / [Haettu 20.12.2023.]

5 Tutkimusaineisto- ja menetelmät

5.1 Aineiston valinta ja kuvaus

Tutkielmassa aineistona käytin DeepL-neuroverkkokääntimen saksasta suomeksi käännettyjä talousaiheita käsitteleviä tekstikatkelmia, jotka kuvaavat kansantaloutta yleistajuisesti. Käännettävän aineiston poimin Ulrike Herrmannin teoksesta *Der Sieg des Kapitals*, joka on 288-sivuinen talousalan tietokirja. Kirjassa Herrmann kuvaa Saksan talousjärjestelmän historiallista kehitystä ja tarjoaa havainnollistavia esimerkkejä kapitalismista ja markkinataloudesta. Hänen mukaansa kapitalismi koskee kaikkia, koska se on kokonaisvaltainen järjestelmä, joka läpäisee kaikki elämänalueet. Teoksessa Ulrike Herrmann selittää kapitalismin jatkuvaa alttiutta kriiseille ja avaa modernin kapitalismin erityispiirteitä. Lisäksi teoksessa käsitellään rahoitusmarkkinoiden, globalisaation ja kriisien historian käännekohtia ja annetaan vinkkejä, miten historian oppeja voisi hyödyntää nykyisten talousongelmien ratkaisemisessa. Kirja on rakennettu siten, että ensin kuvataan kapitalismin syntyä, nousua ja kehitystä, sen jälkeen kapitalismin kokemia kriisejä ja lopuksi pohditaan kapitalismin mahdollista romahdusta. Kirjan aiheen takia siinä esiintyy paljon talousaiheisia termejä.

Referenssikäännöksenä käytin kääntäjä Mari Janatuisen tekemää suomennosta *Pääoman voitto*, joka on julkaistu suomeksi vuonna 2015 ja jonka on kustannut Into Kustannus Oy. Referenssikäännöstä käytin vertailemaan konekäännöstä ihmiskäännökseen saadakseen selville, missä määrin konekäännös eroaa ihmiskäännöksestä.

Koska laajuudeltaan lähes 300-sivuisen kirjan huolellinen analysointi ei mahtunut pro gradu -työlle asetettuihin mittoihin, analysoin tekstistä vain otteita. Sisällöllisen monipuolisuuden ja aineiston johdonmukaisen etenemisen varmistamiseksi valitsin kirjan luvuista 6,9 ja 11 kaksi ensimmäistä sivua, joissa käsitellään talouden eri ilmiöitä, kuten kapitalismia, rahaa, velkaa, korkoja, finanssikriisejä ja inflaatiota. Lähtötekstin otteissa esiintyy lisäksi sellaisia saksan kielen ominaispiirteitä, joiden oletin olevan hankalia DeepL-kääntimelle ja joita käsitelinkin jo luvussa 3.1.

5.2 Tutkimusmenetelmät

Aineistoni tutkimisessa käytin sekä määrällistä että laadullista menetelmää. Määrällisen tutkimusmenetelmän osuus rajoittui virheanalyysin tulosten tilastointiin, kun taas laadullisessa osiossa keskityin virheiden tunnistamiseen, luokitteluun ja niiden vakavuusasteen määrittämiseen käyttäen apuna MQM-laatumittaria. Käyttämässäni virhekeskeisessä lähestymistavassa merkitsin kaikki havaitsemani virheet, laskin niiden määrän ja luokittelin ne vastaaviin virheluokkiin. Näiden pohjalta laskin lopuksi kokonaisprosenttipistemäärän, jota kutsutaan tutkielmassani (raaka)laatumittamääräksi. Tämä pistemäärä auttoi sitten määrittämään, onko käännös hyväksyttävä vai ei, kun pistemäärää verrattiin ennalta määriteltyyn hyväksymisrajaan.

Tässä vaiheessa on syytä mainita, että analyysini tekemisen jälkeen julkaistiin vuonna 2024 uusi kansainvälinen ISO 5060-standardi, mutta se ei enää ehtinyt mukaan tutkimukseeni. MQM ja ISO 5060 ovat pitkälti samankaltaiset. ISO 5060 tarjoaa kattavat ohjeet käännösten arvioinnille, myös konekäännöksille, ja korostaa virhekeskeistä analyysiä käyttäen vastaavia virhekatteorioita ja -tyyppjä kuin MQM.

5.2.1 MQM-virhetypologia

Havaittujen virheiden luokittelua varten tutkielmassani käytin suppeampaa MQM Core -typologiaa ja tarkastelin vain muutamia pääluokkia. Kuten aiemmin todettiin, käännöksen laadunarvioinnissa ei välttämättä tarvitse käyttää kaikkia MQM-mittarin ulottuvuuksia ja alaluokkia, vaan ne valitaan käännösmääritysten perusteella.

Tutkimuksessani keskityin arvioimaan ainoastaan sujuvuutta, tarkkuutta ja terminologiaa, jotka ovat olennaisia osa-alueita valitsemalleni talousaiheiselle tietokirjalle. Valintaan vaikuttivat merkittävästi tietokirjan sisältö, asiasanasto, tietokirjan tarkoitus ja oletettu kohdeyleisö. Koska arvioitavana oli talouteen liittyvä tietokirja, painotin erityisesti oikeiden ja yhdenmukaisten taloudellisten termien käyttöä käännöksessä. Tietokirjat sisältävät usein lukuisia faktoja, joten käännöksen on välitettävä näiden tarkat merkitykset. Lisäksi otin huomioon, että tietokirjan lukijakunta voi olla monipuolinen, joten helppolukuisuus ja ymmärrettävyys ovat keskeisiä. Alaluokista valitsin tarkisteltaviksi ne, jotka liittyvät olennaisesti sujuvuuteen, tarkkuuteen ja terminologiaan. Esittelen seuraavaksi nämä pääkategoriat ja niiden alakategorioista ne, jotka olen valinnut tutkimukseeni tarkasteltaviksi.

5.2.2 Virhekategorioiden tarkkuus, sujuvuus ja terminologia

Tarkkuus -yläkatgoria arvioi tekstin sisällön tarkkuutta, erityisesti lähde- ja kohdetekstin vastaavuutta toisiinsa. Tähän kategoriaan kuuluvat tarkkuusvirheet. Ne ovat virheitä, jotka syntyvät, kun kohdeteksti ei vastaa tarkasti alkuperäisen lähdetekstin sisältöä.

Tarkkuusvirheet ovat merkittäviä, koska ne voivat vaikuttaa muun muassa tekstissä esitettyjen käsitteiden tulkintaan ja ymmärtämiseen. Tarkkuusvirhe voi syntyä, kun käännöksessä ei säilytetä esimerkiksi alkuperäisen ilmaisun merkitystä. Tämä yläkatgoria jakautuu seitsemään alakategoriaan, joista tutkimukseeni valitsin seuraavat:

- 1) **Lisäys:** Lisäysvirhe käännöksessä tapahtuu, kun käännös sisältää ylimääräisiä sanoja tai ilmauksia, joita ei ole alkuperäisessä tekstissä. Tällainen virhe voi muuttaa alkuperäisen tekstin merkitystä, heikentää käännöksen tarkkuutta ja aiheuttaa virheellisiä tulkintoja.
- 2) **Virheellinen käännös:** Kohdetekstin sisältö ei vastaa tarkasti lähdetekstin sisältöä. Tämä voi johtua virheellisestä sanavalinnasta, käsitteen virheellisestä tulkinnasta tai yleisesti ottaen siitä, että käännös ei välitä alkuperäistä merkitystä oikein. Virheelliset käännökset voivat vääristää tekstin merkitystä.
- 3) **Poisto:** Poistovirhe käännöksissä viittaa tilanteeseen, jossa käännöksestä puuttuu sisältö, joka on olemassa lähdekielellä. Tämä voi tapahtua esimerkiksi, kun tietty sana, ilmaisu tai lause jätetään vahingossa huomioimatta tai virheellisesti poistetaan. Poistovirheet voivat muuttaa kohdetekstin merkitystä, aiheuttaa väärinymmärryksiä tai antaa virheellisen kuvan alkuperäisen viestin tarkoituksesta.
- 4) **Ei käännetty/Kääntämätön:** Virheet, jotka johtuvat siitä, että käännettäväksi tarkoitettu tekstisegmentti jää käntämättä kohdesisällössä. Jos tietty osa tekstistä jätetään käntämättä, se voi johtaa puutteelliseen tai virheelliseen käännökseen, joka ei välitä alkuperäisen tekstin tarkkaa merkitystä. Käännös voi olla epätäydellinen.

Sujuvuus -yläkatgoriaan kuuluvat virheet liittyvät tekstin kielelliseen muotoon ja voivat sisältää kieliopillisia sekä oikeellisuuteen liittyviä ongelmia. Tässä kategoriassa on viisi alakategoriaa, joista tutkimukseeni valikoituivat seuraavat:

- 1) **Kielioppi:** Virhe, joka ilmenee, kun käännöksessä oleva lause, fraasi tai muu rikkoo kohdekielen kielioppisääntöjä. Kielioppivirheet voivat vaikuttaa merkittävästi tekstiin. Ne voivat vaikeuttaa ymmärtämistä, heikentää tekstin ammattimaisuutta ja jopa johtaa virheellisiin tulkintoihin. Esimerkiksi virheellinen verbin muoto, sanajärjestys tai muu

kielioppivirhe voi häiritä tekstin sujuvuutta ja vaikeuttaa lukemista. Tämä on erityisen haitallista talousteksteissä, joissa pienetkin tarkkuusvirheet voivat vaikuttaa merkityksen selkeyteen ja tulkintaan ja näin ollen vähentää käännöksen luotettavuutta.

- 2) **Välimerkit:** Tällä alakategoriolla tarkoitetaan tilannetta, jossa välimerkkejä käytetään epäasianmukaisesti, virheellisesti, tyylille tai paikalliselle alueelle sopimattomasti. Virheellinen käyttö voi tarkoittaa välimerkkien puuttumista, liiallista käyttöä tai niiden väärää sijoittelua lauseisiin. Väärinkäytetyt tai puuttuvat välimerkit voivat johtaa epäselvyyksiin ja vaikeuttaa tekstin ymmärtämistä.
- 3) **Oikeinkirjoitus:** Tämä alakategoria käsittää oikeinkirjoitusvirheitä, joilla tarkoitetaan virheitä sanojen kirjoittamisessa tai niiden oikeinkirjoitusmuodon käytössä. Oikeinkirjoitusvirheet voivat vaikuttaa tekstin luettavuuteen.
- 4) **Epäselvä teksti:** Epäselvällä tekstillä tarkoitetaan tilannetta, jossa käännös ei ole selkeää tai ymmärrettävää lukijalle. Tämä voi johtua muun muassa epätarkasta sanavalinnasta, puutteellisesta lauserakenteesta, vaikeaselkoisesta ilmaisusta tai virheistä käännöksessä.

Terminologia -yläkategoriaan kuuluvat terminologiavirheet. Terminologiavirhe käännöksissä viittaa tilanteeseen, jossa käytetty termi ei vastaa erikoisalan tai organisaation normatiivisia terminologisia standardeja, termi käännetään jollain muulla kuin odotetulla tai muuten määritetyllä termillä tai termejä ei ole valittu oikein asiayhteyden mukaisesti. Tämä kategoria sisältää kolme alakategoriaa, joista valitsin tarkasteluun seuraavat:

- 1) **Virheellinen termi:** Virheellisellä termillä tarkoitetaan tilannetta, jossa käännetty termi ei vastaa alkuperäistä termiä tai voi aiheuttaa väärinkäsityksiä tai virheellistä tulkintaa.
- 2) **Termien epäjohdonmukainen käyttö:** Tämä alakategoria kuvaa tilannetta, jossa samalle käsitteelle käytetään useita eri termejä, vaikka johdonmukaisuus on toivottavaa. Termien epäjohdonmukainen käyttö voi sekoittaa käännöksen käyttäjää ja täten vaikeuttaa tekstin ymmärtämistä.

Vaikka MQM-laatumittari mahdollistaa virheiden tarkemman määrittelyn ja korjaamisen, pelkkä virheiden luokittelu ei kuitenkaan riitä laadun arvioinnissa. Yksittäinen virhe ei vielä kerro, onko käännös laadukas. Snow (2015, 20–21) korostaa, että vielä yksi olennainen tekijä on virheiden painotukset, joita voidaan käyttää virheiden merkityksen määrittämiseen sen mukaan, mikä on niiden merkitys tietyllä käännökselle. Jos kaikki virhetyypit ovat yhtä

tärkeitä, niin ne saavat saman painotuskertoimen, esim. 1. Jos joillekin virhetyypeille halutaan antaa suurempi merkitys, niin voidaan käyttää suurempaa painotuskerrointa, esim. 2. Koska tarkasteltavana aineistona oli talousaiheisen tietokirjan tekstikatkelmia, kaikkien valikoitujen virhekategorioiden eli tarkkuuden, sujuvuuden ja terminologian ja niiden virhetyyppien oletin olevan yhtä tärkeitä ottaen huomioon tietokirjan ominaisuudet ja käyttötarkoitus. Näin ollen kaikki virhetyypit saivat saman painotuskertoimen eli 1.

5.2.3 Virheiden vakavuusasteet

MQM-mittarissa käännöksen laadun arvioiminen ei kuitenkaan rajoitu pelkästään virhetyyppien määrän tietämiseen ja painotuksen määrittämiseen, vaan arvioijien on myös tiedettävä, kuinka vakavia virhetyypit ovat käsiteltävän aineiston kannalta (Lommel 2018, 120). Näin ollen arvioinnin kannalta on tärkeä määrittää myös virheiden vakavuusaste. Mitä vakavampi virhe on kyseessä, sitä todennäköisemmin se vaikuttaa käännöksen käytettävyyteen (mt., 120). Vakavuusaste on indikaattori virheen merkittävyydestä. Päätettäessä virheen vakavuusasteesta arvioijien tulisi ottaa huomioon käännöslaatumääritykset eli spesifikaatiot ja pohtia, millaiset seuraukset virheestä voi aiheutua loppukäyttäjille, mikäli se olisi jäänyt huomaamatta, ja miten tämä vaikuttaisi tekstin käytettävyyteen sen kohderyhmän ja tarkoituksen kannalta (MQM, ei vuotta). Koska yksi tutkimuskysymyksistäni on selvittää semanttisten vääristymien vaikutusta käännosten ymmärrettävyyteen ja käyttökelpoisuuteen, tarkastelin tutkimuksessa havaittuja virheitä niiden vakavuuden näkökulmasta.

MQM-virhetypologiassa virheet luokitellaan niiden vakavuusasteen mukaan neljään tasoon (kriittinen, merkittävä, vähäinen ja neutraali), ja kukin vakavuusaste pisteytetään. MQM määrittelee virheille eri vakavuuspisteitä, jotka eivät ole universaaleja ja voivat vaihdella 1–100 käyttötarkoituksesta ja käännöksen kontekstista riippuen. Vakavuuspisteitä voidaan myös mukauttaa omiin tarpeisiin, ja suosituksena on, että neutraalien, vähäisten, merkittävien ja kriittisten virheiden arvojen välillä olisi eksponentiaalinen ero, esimerkiksi 0–1–5–25 (MQM, ei vuotta). Tämä eksponentiaalinen suhdeasteikko kuvastaa virheiden riskin ja vaikutusten lisääntymistä. Vastaavaa pisteytystä käytin myös tutkielmassani, jossa vähäiset virheet saivat 1 vakavuuspisteen, merkittävät 5 vakavuuspistettä ja kriittiset 25 vakavuuspistettä.

Kriittiset virheet ovat sellaisia käännösvirheitä, jotka yksinään tekevät käännöksestä hyödyttömän tai käyttökelvottoman. Tämä määräytyy kohderyhmän ja määritellyn tarkoituksen perusteella. Kriittiset virheet voivat aiheuttaa vahinkoa ihmisille, laitteille tai organisaation maineelle, mikäli niitä ei korjata ennen kuin käänнос otetaan käyttöön (MQM, ei vuosilukua). Käännöksen laatua arvioitaessa yksikin kriittinen virhe voi johtaa siihen, että teksti saa automaattisesti arvosanan ”hylätty” (mt.).

Merkittävät virheet ovat virheitä, jotka vaikuttavat sisällön käytettävyyteen tai ymmärrettävyyteen, mutta eivät tee siitä käyttökelvotonta. Esimerkiksi väärin kirjoitettu sana saattaa vaatia lukijalta lisäponnistuksia lauseen merkityksen ymmärtämiseksi. Jos esimerkiksi tarkoitettu merkitys ei ole niin selvä, että lukija voisi luotettavasti korjata virheen, mutta virhe ei tee sisällöstä käyttökelvotonta, virhe tulisi luokitella merkittäväksi virheeksi (MQM, ei vuosilukua.) Merkittävät käännösvirheet tekevät tekstin merkityksen epäselväksi käyttäjälle. Merkittävä käännösvirhe ei kuitenkaan vahingoita käyttäjää merkittävästi, koska tämä ymmärtää, että tekstissä on virhe.

Vähäiset virheet puolestaan eivät vaikuta käytettävyyteen. Ne edustavat virheitä, jotka eivät vaikuta sisällön käytettävyyteen tai ymmärrettävyyteen. Jos käännökseen käyttäjä pystyy korjaamaan virheen luotettavasti eikä se vaikuta sisällön käytettävyyteen, virhe olisi luokiteltava vähäiseksi (MQM, ei vuotta). Monissa tapauksissa käyttäjä jättää tällaiset virheet tietoisesti huomiotta. Vähäiset virheet eivät aiheuta ongelmia merkityksen välittämisessä.

Mitättömiä tai neutraaleja virheitä ei varsinaisesti pidetä virheinä. Ne ovat virheitä, jotka lähtökohtaisesti sivuutetaan ennen kuin tekstin arviointi alkaa, koska ne eivät vaaranna kohdetekstin ymmärrettävyyttä. Neutraaleilla virheillä voidaan tarkoittaa esimerkiksi toistuvia virheitä, kuten saman sanan jatkuvaa kirjoitusvirhettä, joka ei kuitenkaan vaikuta tekstin kokonaisuuteen. Neutraalit virheet jätetään huomioimatta arvioinnissa (MQM, ei vuotta). Omassa tutkielmassani jätin kaikki neutraaleiksi luokitellut virheet huomiotta laskussa ja käännökseen laadun arvioinnissa.

5.2.4 Laatupistemäärän laskeminen

Kun virheet on tunnistettu, laskettu ja luokiteltu, seuraavaksi lasketaan laatupistemäärä ja määritetään laatu. MQM ehdottaa kahta laatupistemäärän laskentamenetelmää: *ei-kalibroitu* ja *kalibroitu*. Tutkimuksessani yhdistelin nämä laskentamenetelmät.

Ei-kalibroidun menetelmän avulla saatu laatu pistemäärä edustaa arvioitavan käännöksen raakatulosta, ja sitä kutsutaan myös raakapisteytysmalliksi. Ei-kalibroidussa menetelmässä (raaka)laatu pistemäärä saadaan laskemalla tekstin virheellinen osuus ja vähentämällä tämä arvo 1:stä tai 100 %:sta, jolloin saadaan arvo, joka edustaa virheetöntä osaa arvioidusta tekstistä. Laatu pistemäärä ilmaisee siis sen osuuden arvioidusta kohdesisällöstä, joka on virheetön. (MQM, ei vuotta.) Arvot 1 ja 100 % ilmaisevat tässä oletettua maksimilaatu pistemäärää. Maksimilaatu pistemäärää ei lasketa mistään erityisistä perusteista, vaan se on mielivaltaisen arvo, joka yksinkertaisesti valitaan alan parhaiden käytäntöjen mukaan ja jonka tarkoituksena on tehdä saadun laatu pistemäärän ymmärtämisestä helppoa muuntamalla lopuksi saatu laatu pistemäärä esimerkiksi prosentuaaliseksi arvoksi. Tutkimuksessani enimmäislaatu pistemäärän prosentuaaliseksi arvoksi valitsin myös 100 %.

Sekä ei-kalibroidussa että kalibroidussa menetelmässä arvioitaville käännöksille määritellään hyväksymiskynnys ennen arviointia. Hyväksymiskynnys on pistemäärä tai prosentuaalinen arvo, joka määrittelee hyväksyty/hylätty -rajan. Yleinen käytäntö ei-kalibroidussa menetelmässä on asettaa hyväksymiskynnykseksi esimerkiksi 0,99 tai 99 % käytetyn asteikon mukaan. Tässä tapauksessa 99 % ilmaisee sen osuuden arvioidusta käännöksestä, joka on virheetön, ja ”virheellisen osuuden” väli on 1 %. Näin ollen mikä tahansa laatu pistemäärä välillä 100–99 % luokitellaan ”hyväksytyksi”. 99 %:n hyväksymisraja käytännössä tarkoittaa, että 1000 sanan mittaisessa tekstissä saisi olla korkeintaan 10 rangaistuspistettä edellyttäen, että kaikkien virhetyyppien painotuskerroin on 1. Tämä tarkoittaa sitä, että 1000 sanan mittaisessa tekstissä 10 rangaistuspistettä saadaan jo, jos tekstissä esiintyy 10 vähäistä virhettä tai 2 merkittävää virhettä. Jos arvioitavan tekstin rangaistuspisteiden kokonaismäärä on yli 10 rangaistuspistettä/1000 sanaa, niin käännös ei ole hyväksyttävä. Tämä ei mielestäni tuntunut realistiselta vaatimukselta minun tutkimukseni kannalta, joten tutkielmassani käytin myös kalibroituja menetelmiä, jonka mukaan hyväksymisrajaa voidaan mukauttaa vastaamaan käännöksen käyttäjän vaatimuksia. Arvioitaville käännöksille asetin näin ollen 85 %:n hyväksymisrajan, mikä tuntuu realistisemmalta vaatimukselta. Tämä 85 %:n raja tarkoittaa käytännössä sitä, että 1000 sanan mittaisessa tekstissä saa olla yhteensä 150 rangaistuspistettä. Tutkimukseni konekäännöksessä, joka sisältää 2501 sanaa, saa vastaavasti olla enintään 375 rangaistuspistettä, jota käännöstä voidaan pitää vielä hyväksyttävänä.

Hyväksymisrajan lisäksi asetin käännöksille toisen hyväksymisehdon: kriittisen virheiden puuttumisen. Tämä on siksi, että kriittiset virheet tekevät käännöksestä käytännössä

käyttökelvottoman ja johtavat automaattisesti käännöksen hylkäämiseen, vaikka se olisi muuten laadukas (MQM, ei vuotta).

Laatupistemäärän laskemisessa noudatin MQM:n ei-kalibroidulle menetelmälle ehdottamaa kolmea vaihetta:

1) Ensimmäisessä vaiheessa laskin rangaistuspisteiden määrän valituille virhetyypeille – tarkkuus-, sujuvuus- ja terminologiavirheille – huomioiden niiden vakavuuspisteet (1,5 ja 25) ja painotuskertoimet (1). Lopuksi rangaistuspisteet laskin yhteen, jolloin sain rangaistuspisteiden kokonaismäärän.

Rangaistuspisteiden kokonaismäärä = [(vähäinen virhe x vakavuuspiste) + (merkittävä virhe x vakavuuspiste) + (kriittinen virhe x vakavuuspiste) x painotuskerroin]

2) Toisessa vaiheessa laskin sanakohtaisen rangaistuspistemäärän jakamalla saatu rangaistuspisteiden kokonaismäärä arvioitavan käännöksen sanamäärällä.

Sanakohtainen rangaistuspistemäärä = rangaistuspisteiden kokonaismäärä/arvioitavan käännöksen sanamäärä

3) Kolmannessa vaiheessa laskin laatupistemäärän prosentuaalisesti ilmaistuna.

Laatupistemäärä = 100 % - (sanakohtainen rangaistuspistemäärä x 100 %)

Oletetaan esimerkiksi, että käänöksessä on 1000 sanaa ja tarkastaja löytää 10 vähäistä virhettä, 2 merkittävää virhettä ja 0 kriittistä virhettä. Lisäksi kaikkien virhetyyppien painotuskerroin on 1. Lisäksi vähäiset virheet saavat vakavuuspisteen 1, merkittävät 5 ja kriittiset 25. Sijoitetaan nämä tiedot ylempänä esitettyihin kaavoihin:

Rangaistuspisteiden kokonaismäärä = [(10 vähäistä virhettä x 1) + (2 merkittävää virhettä x 5) + (0 kriittistä virhettä x 25) x 1] = 20 rangaistuspistettä

Sanakohtainen rangaistuspistemäärä = 20 rangaistuspistettä / 1000 sanaa = 0,02

(Raaka)laatupistemäärä = 100 % - (0,02 x 100 %) = **98 %**

Saadun laatupistemäärän pohjalta voi lopuksi määrittää, onko käänнос hyväksyttävä.

6 Tulokset ja analyysi

Kuten aiemmin mainitsin, analysoin DeepL-konekäännöstä käyttäen MQM-laatumittaria ja vertailin sitä ihmisen tekemään käännökseen. MQM-mittarin avulla pystyin luokittelemaan havaitsemiani virheitä sujuvuus-, tarkkuus- ja terminologiavirheisiin sekä arvioimaan niiden vakavuusasteet. Kun kaikki virheet oli luokiteltu, laskin laatupistemäärän hyödyntämällä virhetyyppien vakavuuspisteet ja painotuskertoimet. Tämän laatupistemäärän hyödynsin arvioidakseni, oliko konekäännös hyväksyttävä vai ei. Seuraavaksi esittelen tutkimukseni määrälliset ja laadulliset tulokset.

6.1 DeepL-konekäännöksen määrällinen analyysi

Käännöksessä esiintyi yhteensä 61 virhettä, joista 41 virhettä liittyi tarkkuuteen, 13 virhettä sujuvuuteen ja loput 7 terminologian käyttöön. Näistä 61 virheestä kriittisiä virheitä oli 0 kpl, merkittäviä 22 kpl ja vähäisiä 39 kpl. Tulokset esitetään taulukossa 1.

Taulukko 1. Virheiden määrä vakavuusasteen mukaan

Virheiden vakavuusaste	Virheiden kappalemäärä
-kriittinen virhe	0
-merkittävä virhe	22
-vähäinen virhe	39
Virheiden määrä yhteensä:	61

Tarkasteluun valikoiduista virheluokista kaikista eniten virheitä esiintyi tarkkuuden kategoriassa, seuraavaksi eniten sujuvuuden kategoriassa ja vähiten terminologian luokassa. Tarkemmat tulokset esitetään taulukossa 2.

Taulukko 2. Virheiden kappalemäärä virhetyypeittäin

Virheiden yläkategoriat ja niiden virhetyypit	Virheiden kappalemäärä
Tarkkuus	41
-virheellinen käännös	37
-lisäys	1
-poisto	1

-ei käännetty/kääntämätön	2
Sujuvuus	13
-kielioppi	13
-välimerkit	0
-oikeinkirjoitus	0
-epäselvä teksti	0
Terminologia:	7
-virheellinen termi	6
-termien epäjohdonmukainen käyttö	1

Virheiden määrittelyn ja analyysin jälkeen laskin lopuksi laatu pistemäärän hyödyntämällä MQM:n ehdottamaa kolmea vaihetta, jota esittelin *Tutkimusmenetelmät* -osiossa. Tässä osiossa esittelen ainoastaan laskutoimitusten pohjalta saadut tulokset.

Laskutoimituksissa käytin seuraavia tietoja: DeepL-käännöksessä oli yhteensä 2501 sanaa. Tarkkuusvirheitä löytyi 41 kappaletta, joista 18 oli merkittäviä ja 23 vähäisiä; kriittisiä virheitä ei ollut. Sujuvuusvirheitä oli 13 kappaletta, joista 2 oli merkittäviä ja 11 vähäisiä; kriittisiä virheitä ei ollut. Terminologiavirheitä oli 7 kappaletta, joista 2 oli merkittävää ja 5 vähäistä. Kaikkien virhetyyppien painotuskertoimena käytin arvoa 1. Vähäiset virheet saivat 1 vakavuuspisteen, merkittävät 5 vakavuuspistettä ja kriittiset 25 vakavuuspistettä.

Näiden tietojen pohjalta suoritettujen laskutoimitusten tulokset olivat seuraavanlaisia:

Vaihe 1:

Tarkkuusvirheiden rangaistuspistemäärä = 113 rangaistuspistettä.
Sujuvuusvirheiden rangaistuspistemäärä = 21 rangaistuspistettä.
Terminologiavirheiden rangaistuspistemäärä = 15 rangaistuspistettä.
Rangaistuspisteiden kokonaismäärä = 113 + 21 + 15 = 149

Vaihe 2:

Sanakohtainen rangaistuspistemäärä = 149 (rangaistuspisteiden kokonaismäärä) / 2 501
(arvioitavan tekstin kokonaissanamäärä) = 0,0596

Prosentteina ilmaistuna se on: $0,0596 \times 100 \% = 5,96 \%$

Vaihe 3:

(Raaka)laatu pistemäärä = 100 % - 5,96 % = 94,04 %.

Pyöristetynä (raaka) laatu pistemäärä on 94 %. Tämä on se pistemäärä, joka ilmaisee sen osuuden arvioidusta käännöksestä, jossa ei todettu virheitä. Ennen analyysiä käännökselle asetin hyväksymiskynnykseksi 85 %. Tulosta 94 % voidaan näin ollen pitää hyväksyttävänä tuloksena. Tulosta arvioitaessa täytyy kuitenkin huomioida, että arvioitavina kategorioina olivat ainoastaan tarkkuus, sujuvuus ja terminologia.

Koska yhtenä tutkimuskysymyksenä oli selvittää, minkälaisia tarkkuuteen, sujuvuuteen ja terminologiaan liittyviä virheitä esiintyy käännöksessä, niin siirryn seuraavaksi tarkastelemaan niitä laadullisesti. En esitele kaikkia virheitä, vaan keskityn ennemminkin sellaisiin virheisiin, joilla oli lauseen semanttista merkitystä vääristävä vaikutus osittain tai kokonaan. Tämä on siksi, että toisena tutkimuskysymyksenä oli selvittää, miten käännöksessä esiintyvät semanttisen merkityksen vääristymät vaikuttavat tekstin ymmärtämiseen ja käyttöön.

6.2 Ihmis- ja DeepL-käännöksen laadullinen analyysi

Tutkimuksessani kolmantena kysymyksenä oli selvittää, missä määrin DeepL-konekääntimen talousalan tekstistä tuottamat käännökset vastaavat ihmisen tekemien käännösten laatua. Tässä osiossa esittelen siksi rinnakkain konekäännöksen ja ihmisen tuottaman käännöksen ja vertailen niitä keskenään. Laadullisessa osiossa esittelen ainoastaan vähäisiä ja merkittäviä virheitä, ja neutraalit virheet eivät ole joukossa mukana. Sivuutin neutraalit virheet käännöksen laadun arvioinnissa siksi, että niiden esiintymisellä ei ollut vaikutusta tekstin sisällön ymmärtämiseen ja ne eivät myöskään vääristäneet lauseiden semanttista merkitystä. Sitä vastoin vähäisten ja merkittävien virheiden esiintymisellä oli tekstin lauseiden merkitystä osittain tai kokonaan vääristävä vaikutus. Kriittisiä virheitä tekstissä ei ilmennyt. Suurin osa virheistä oli vähäisiä tai merkittäviä. Erityisesti merkittävät virheet tekivät joistakin lauseista vaikeiksi ymmärtää. Merkittäviin virheisiin luokittelin sanojen virheelliset käännökset ja virheelliset termit sekä virheelliset sanajärjestykset, joilla oli lauseen merkitystä vääristävä vaikutus tai jotka eivät välittäneet semanttista merkitystä ollenkaan. Vähäisiin virheisiin luokittelin termien epä johdonmukaisen käytön, joka vaikutti pääasiassa lukukokemukseen. Vähäisiksi luokittelin myös tapauksia, joissa lauseessa painotettiin epäolennaista asiaa.

6.2.1 Tarkkuusvirheet

Konekäännöksessä esiintyi eniten tarkkuusvirheitä, jotka ilmenivät lähes aina virheellisinä käännoksinä. Virheelliset käännot liittyivät usein alkuperäisen sanan merkityksen virheelliseen tulkintaan. Näin syntyi tilanne, jossa kohdelause ei vastannut tarkasti alkuperäisen lähdelauseen sisältöä. Suuri osa tarkkuusvirheistä oli merkittäviä, koska ne vaikuttivat häiritsevästi muun muassa lauseiden merkitysten ymmärtämiseen. Ihmisen tekemästä käännoksestä en löytänyt tarkkuusvirheitä. Esimerkkejä konekääntimen tuottamista virheellisistä käännoksista on nähtävissä taulukossa 3.

Taulukko 3. Esimerkit virheellisistä käännoksista

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Marktwirtschaft, das klingt <u>kuschelig</u> .	Markkinatalous, se kuulostaa <u>hellyttävältä</u> .	Markkinatalous kuulostaa <u>mukavalta</u> .
(2) ..., dass Marktwirtschaft und Kapitalismus so <u>anhaltend</u> verwechselt werden.	..., että markkinatalous ja kapitalismi sekoitetaan niin <u>sitkeästi</u>, ja markkinatalouden ja kapitalismin <u>jatkuva</u> sekoittaminen toisiinsa on jo johtanut...
(3) Viele Menschen <u>ahnen</u> instinktiv, dass der Kapitalismus nicht das Gleiche ist wie eine Marktwirtschaft.	Monet ihmiset <u>epäilevät</u> vaistomaisesti, että kapitalismi ei ole sama asia kuin markkinatalous.	Monet <u>aavistavat</u> vaistomaisesti, ettei kapitalismi ole sama asia kuin markkinatalous.
(4) Und wie kann man es <u>retten</u> ?	Ja miten sitä voi <u>säästää</u> ?	Miten sen voi <u>pelastaa</u> ?
(5) Schuld klingt zudem nach <u>Schuldturn</u> , nach <u>drohender Knechtschaft</u> und Unfreiheit.	Velka kuulostaa myös <u>syllisyyden tornilta</u> , <u>orjuuden uhalta</u> ja vapauden puutteelta.	Kun sitä käytetään tarkoittamaan taloudellista velkaa, se luo mielikuvia <u>taakasta</u> , <u>uhkaavasta orjuudesta</u> ja vapauden menetyksestä.
(6) Die »kleine Morgen- und <u>Abendtoilette</u> « kostet zum Beispiel 12,91 Euro, während die »Hilfe bei der Nahrungsmittelaufnahme einer Zwischenmahlzeit« mit nur 4,78 Euro berechnet wird.	Esimerkiksi "pieni aamu- ja <u>iltavessa</u> " maksaa 12,91 euroa, kun taas "apu välipalan syömisessä" maksaa vain 4,78 euroa.	"Pieni aamu- tai iltapesu" on esimerkiksi 12,91 euroa, kun taas "apu välipalan nauttimisessa" maksaa 4,78 euroa.
(7) Aber <u>sie sehen einen Unterschied</u> zwischen Märkten und Marktwirtschaft.	Mutta <u>he tunnustavat eron</u> markkinoiden ja markkinatalouden välillä.	<u>Heidän mielestään</u> markkinoilla ja markkinataloudella <u>on</u> kuitenkin <u>ero</u> .
(8) Diese bemerkenswerte Kontinuität bei den führenden Aktiengesellschaften zeigt, dass <u>gegen ihre schiere Größe niemand mehr ankommt</u> .	Tämä johtavien osakeyhtiöiden huomattava jatkuvuus osoittaa, <u>ettei kukaan voi kilpailla niiden pelkän koon kanssa</u> .	Johtavien osakeyhtiöiden merkittävä jatkuvuus osoittaa, että <u>niiden sulan koon vuoksi kukaan ei enää pärjää niille</u> .

Esimerkissä 1 konekäännin on valinnut kontekstiin nähden epäsoivan käännöksen eli tulkinut sanan ”*kuschelig*” virheellisesti. Ilmaisun ”*markkinatalous kuulostaa hellyttävältä*” on ongelmallinen, koska se on lukijalle epäselvä ja jättää tulkinnanvaraa. Vaikka lukija saattaisi ymmärtää, että kyse on jostakin positiiviselta kuulostavasta, tällainen epäsoiva sananvalinta hidastaa lukemista. Luokittelin virheen vähäiseksi ja annoin sille vakavuuspisteen 1. Ihmiskäännös ”*markkinatalous kuulostaa mukavalta*” olisi tässä tapauksessa neutraalimpi ja suomalaiselle lukijalle selkeämpi.

Esimerkissä 2 nähdään jälleen, että DeepL on valinnut saksankieliselle sanalle ”*anhaltend*” epäsoivan vastineen, mikä on johtanut tarkan merkityksen katoamiseen. Suomenkielinen lause ”...*, että markkinatalous ja kapitalismi sekoitetaan niin sitkeästi*” viittaa kyllä siihen, että markkinatalous ja kapitalismi sekoitetaan keskenään, mutta sana ”*sitkeästi*” ei tarkalleen kerro, mitä sillä tässä yhteydessä tarkoitetaan. Lauseen merkitys ei siis välity täysin, vaan osittain hämärtyy. Luokittelin virheen vähäiseksi ja annoin sille vakavuuspisteen 1. Ihmiskääntäjän ehdottama ratkaisu, eli käännös ”*jatkuvasti*”, välittää alkuperäisen sanoman täsmällisemmin.

Esimerkissä 3 konekäännin on kääntänyt saksankielisen sanan ”*ahnen*” virheellisesti suomeksi sanalla ”*epäilevät*”, mikä kadottaa osittain lauseen alkuperäisen sanoman. Luokittelin tämän virheen merkittäväksi ja annoin sille vakavuuspisteen 5. Ihmiskääntäjän vaihtoehto ”*aavistavat*” on mielestäni oikea ja välittää sanan alkuperäisen merkityksen tarkasti.

Esimerkissä 4 on vastaava tilanne, jossa kone on kääntänyt sanan ”*retten*” suomeksi ”*säästää*”, mikä on täysin virheellinen tässä kontekstissa. Tämän virheellisen sanavalinnan takia lauseen merkitys ei välity lainkaan. Luokittelin virheen merkittäväksi, ja se sai vakavuuspisteen 5. Vaikka lausetasolla tämä virhe voisi olla kriittinenkin, koko käännöstekstiin suhteutettuna se ei tee koko tekstistä käyttökelpotonta. Ihmiskääntäjän vaihtoehto ”*pelastaa*” välittää alkuperäisen sanoman virheettömästi.

Esimerkissä 5 koneen ehdottamat käännökset ”*syyllisyyden torni*” ja ”*orjuuden uhka*” saksankielisille sanoille ”*Schuldturm*” ja ”*drohende Knechtschaft*” ovat ongelmallisia, sillä ne eivät välitä sanojen alkuperäistä merkitystä oikein. Luokittelin nämä virheet merkittäviksi ja annoin niille vakavuuspisteet 5. Käännös ”*syyllisyyden torni*” havainnollistaa DeepL-

kääntimen kohtaamia haasteita kaksimerkityksisten sanojen kääntämisessä. Saksan kielessä sana ”*Schuld*” tarkoittaa sekä ”*syylisyyttä*” että ”*velkaa*”. DeepL on valinnut näistä merkityksistä väärän, vaikka oikea valinta olisi ollut ”*velka*”. Sana ”*Schuldturm*” viittasi 1800-luvulle saakka erityisvankilaan, jossa pidettiin henkilöitä, jotka eivät olleet täyttäneet velanmaksuvelvoitteitaan. Oikeampi käännös ”*Schuldturm*” -sanalle voisi olla ”*velkatorni*” tai ”*velkavankila*”. Nämä vaihtoehdot saattavat kuitenkin vaatia lisäselvitystä, sillä ”*velkatorni*” ei välttämättä ole tuttu sana kaikille lukijoille, ei ainakaan itselleni. Sen sijaan ihmiskääntäjä on valinnut selkeämmän lähestymistavan ja käyttänyt sanaa ”*taakka*”, josta lukija voi helposti päätellä, että kyse on velkataakasta.

Esimerkkien 6 ja 7 konekäännökset ovat myös ongelmallisia. Esimerkissä 6 koneen tuottama ilmaisu ”*pieni aamu- ja iltavessa*” on epäselvä ja jättää tulkintavaraa lukijalle. Tarkempia käännösvaihtoehtoja voisivat olla esimerkiksi ”*aamu- ja iltatoimi*” tai ihmiskääntäjän ehdottama ”*aamu- ja iltapesu*”. Esimerkissä 7 kone ei ymmärtänyt lähdekielen lauseen kontekstia oikein ja kääntänyt saksankielisen sanan ”*sehen*” virheellisesti, mikä vääristää lauseen alkuperäisen merkityksen. Sen sijaan ihmiskäännöksessä alkuperäinen sanoma välittyy oikein. Pidin esimerkeissä 6 ja 7 esiintyneitä virheitä merkittävinä ja annoin niille vakavuuspisteen 5.

Esimerkki 8 on jälleen esimerkki virheellisestä käännöksestä, jossa kone ei ole pystynyt tulkitsemaan oikein, kuka kilpailee kenen kanssa. Tämä virheellinen tulkinta kadottaa alkuperäisen tekstin sanoman, minkä vuoksi luokittelin virheen merkittäväksi ja annoin sille vakavuuspisteen 5. Vaikka lauseen merkitys on mahdollisesti ymmärrettävissä, lukija joutuu näkemään sen eteen ylimääräistä vaivaa. Tämä virhe on merkittävä, muttei koko käännöstekstiin suhteutettuna ole kriittinen. Ihmiskäännös sen sijaan välittää alkuperäisen lauseen sanoman oikein.

Virheellisten käännösten lisäksi konekäännöksessä esiintyi muita tarkkuusvirheitä, kuten poisto (kts. taulukko 5), lisäys (kts. taulukko 6) ja kääntämätön sana (kts. taulukko 4).

Taulukko 4. Kääntämätön -virhetyyppi

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Er beeindruckte seine griechischen Nachbarn mit derart opulenten Geschenken an die Orakel von Delphi, Milet und Ephesus, das sich der Ausdruck »reich wie Krösus« bis heute erhalten hat.	Hän teki vaikutuksen kreikkalaisiin naapureihinsa antamalla niin runsaita lahjoja Delfoin, Miletoksen ja Efesoksen oraakeleille, että ilmaisu " <u>rikas kuin Krösus</u> " on säilynyt tähän päivään asti.	Hän teki kreikkalaisiin naapureihinsa vaikutuksen antamalla Delfoin, Miletoksen ja Efesoksen oraakkeleille sen verran yllättäviä lahjoja, että ilmaisu " <u>rikas kuin Kroisos</u> " on säilynyt näihin päiviin asti.
(2) Im Mittelalter gab es zwar zahllose Märkte, wie Straßenschilder namens »Kornmarkt«, »Gänsemarkt«, »Tuchmarkt«, »Fischmarkt«, »Heumarkt« oder »Holzmarkt« bezeugen.	Keskiajalla oli lukemattomia toreja, kuten " <u>Kornmarkt</u> ", " <u>Gänsemarkt</u> ", " <u>Tuchmarkt</u> ", " <u>Fischmarkt</u> ", " <u>Heumarkt</u> " tai " <u>Holzmarkt</u> " -nimiset katukyltit osoittavat.	Keskiajalla oli lukuisia markkinapaikkoja, kuten esimerkiksi monista saksankielisistä katukylteistä voi edelleen havaita: <u>Kornmarkt – viljatori</u> , <u>Gänsemarkt – hanhitori</u> , <u>Tuchmarkt – kangastori</u> , <u>Fischmarkt – kalatori</u> , <u>Heumarkt – heinätori</u> tai <u>Holzmarkt – puutori</u> .

Esimerkissä 1 kone on jättänyt sanan "*Krösus*" kääntämättä. Tämä kääntämättä jättäminen johtaa tässä tapauksessa puutteelliseen ja epätäydelliseen käännökseen, joka ei välitä sanonnan "*rikas kuin Kroisos*" tarkkaa merkitystä. Suomalaiselle lukijalle ei välttämättä avaudu, mitä sana "*Krösus*" tarkoittaa. Toisaalta voisi olettaa, että lukija mahdollisesti pystyisi itsekkin päättämään/johtamaan sanonnalle "*reich wie Krösus*" oikean suomalaisen vastineen "*rikas kuin Kroisos*" helposti, koska sanontaa edeltävässä tekstissä puhuttiinkin jo aiemmin Kroisos-nimisestä kuninkaasta. Sanontaa edeltävä teksti on seuraava:

"Lydiaalaiset olivat indoeurooppalainen kansa, joka asui Vähä-Aasiassa, ja heidän kuuluisin kuninkaansa oli nimeltään Kroisos. Hän teki vaikutuksen kreikkalaisiin naapureihinsa antamalla niin runsaita lahjoja Delfoin, Miletoksen ja Efesoksen oraakeleille, että ilmaisu "rikas kuin Krösus" on säilynyt tähän päivään asti."

Tästä kuitenkin huomaa, että sanonnan kääntäminen DeepL-kääntimelle ei aina ole helppoa. Tämä puutteellinen käännös ei kuitenkaan tee koko lauseesta eikä tekstistä käyttökelpotonta. Luokittelin tämän virheen vähäiseksi ja annoin vakavuuspisteen 1. Ihmiskääntäjä on suomentanut sanonnan oikein.

Toinen esimerkki kääntämättä jättämisestä oli merkittävämpi. Konekäännöksessä nähdään, että torien nimiä ei ole käännetty. Tästä seuraa, että lauseen tarkka merkitys ei välity kokonaan. Lukijalle selviää ainoastaan, että toreja oli kaikenlaisia. Lukija ei saa tarkkaa kuvaa

eli tietoa siitä, minkä tyyppisiä torit olivat. Pidin tätä virhettä merkittävänä ja annoin sille vakavuuspisteen 5. Toisaalta kyseistä virhettä voisi pitää vähäisenäkin, sillä olennainen tieto lukemattomien torien olemassaolosta kuitenkin välittyy, vaikkakin yksityiskohdat eli torien nimet jäävät epämääräisiksi lukijalle. Koko käännostekstiin suhteutettuna näitä yksityiskohtia voisi pitää epäolennaisina. Ihmiskääntäjä on käännoksessään säilyttänyt torien saksankieliset nimet ja kääntänyt ne, mikä voi olla toimiva ratkaisu lukijan kannalta.

Konekäännöksessä esiintyneet poisto- ja lisäysvirheet eivät olleet vaikutukseltaan merkittäviä, koska ne eivät muuttaneet kohdetekstin merkitystä eivätkä aiheuttaneet väärinymmärryksiä. Ne kuitenkin vaikuttivat tekstin luettavuuteen. Esimerkkejä on taulukossa 5 ja 6.

Taulukko 5. Poistovirhe

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Großkonzerne kommen <u>darin</u> kaum vor.	(1) Suuryritykset eivät esiinny juuri ollenkaan.	Suuryrityksiä ei <u>joukossa</u> juuri ole.

Konekäännöksessä nähdään, että kone on poistanut sanan ”*darin*”. Luokittelin tämän virheen vähäiseksi ja annoin sille vakavuuspisteen 1, koska tämän sanan poisto käännoksestä ei kuitenkaan vääristä lauseen merkitystä ja lukija voi korjata puuttuvan tiedon helposti. Ihmiskäännös välittää mielestäni lauseen merkityksen täsmällisesti.

Taulukko 6. Lisäysvirhe

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) In seinem Buch <i>Wohlstand für alle</i> lässt er fast keine Seite <u>verstreichen</u> , ohne die sozialen Segnungen des Wettbewerbs <u>zu preisen</u> .	(1) Kirjassaan Wohlstand für alle (Hyvinvointia kaikille) <u>hän ei jätä sivua lukematta ylistämättä</u> kilpailun sosiaalisia siunauksia.	Kirjassaan <i>Wohlstand für alle</i> [Saksan ihme, suom. Jorma Aaltonen, Gummerus 1959] <u>hän muistaa ylistää</u> kilpailun sosiaalista siunauksellisuutta <u>melkein joka sivulla</u> .

Konekäännöksessä ilmaisu ”*er lässt keine Seite verstreichen, ohne ... zu preisen*” on käännetty suomeksi muotoon ”*hän ei jätä sivua lukematta ylistämättä*”. Sana ”*lukematta*” on tässä ylimääräinen lisäys, joka häiritsee lauseen tarkan merkityksen välittymistä ja vaikeuttaa olennaisen ymmärtämistä. Lukijalle voi olla epäselvää, ylistetäänkö sivua vai sosiaalisen kilpailun siunauksellisuutta. Tämä lause on mielestäni vaikeaselkoinen ja osittain vääristöä alkuperäistä merkitystä. Siksi luokittelin virheen merkittäväksi ja annoin siitä

vakavuuspisteen 5. Esimerkiksi käännös ”*hän ei jätä melkein yhtään sivua ylistämättä kilpailun sosiaalisia siunauksia...*” tai ihmiskääntäjän ehdottama ”*hän muistaa ylistää kilpailun sosiaalista siunauksellisuutta melkein joka sivulla*” välittäisi alkuperäisen merkityksen tarkemmin ja lukijaystävällisemmin kuin konekäännös.

6.2.2 Sujuvuusvirheet

Tarkkuusvirheiden jälkeen konekäännöksessä esiintyi eniten sujuvuusvirheitä, joista suurin osa liittyi virheelliseen sanajärjestykseen. Konekäännöksessä ei kuitenkaan ilmennyt lainkaan vähäisiä tai merkittäviä oikeinkirjoitus- tai välimerkkivirheitä, eikä teksti ollut epäselvää. Esittelen muutamia esimerkkejä sanajärjestyksen virheistä, sekä vähäisistä että merkittävistä, havainnollistaakseni, millainen vaikutus virheellisellä sanajärjestyksellä voi olla lauseen ymmärrettävyyteen. Esimerkkejä esitellään taulukossa 7.

Sanajärjestysvirheet voivat vaikeuttaa ymmärtämistä, häiritä tekstin sujuvuutta ja vaikeuttaa lukemista. Tämä on erityisen haitallista talousteksteissä, joissa pienetkin virheet voivat heikentää merkityksen selkeyttä ja tulkintaa.

Taulukko 7. Sujuvuusvirhe

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Doch <u>nicht nur</u> die Tilgung von Krediten war oft schwierig.	(1) Lainojen takaisinmaksu ei kuitenkaan ollut usein <u>pelkästään</u> vaikeaa.	Velkojen takaisinmaksu ei usein ollut ainoa ongelma.
(2) »Du sollst von deinem Bruder nicht Zinsen nehmen, weder für Geld noch für Speise noch für alles, wofür man Zinsen nehmen kann. ...«	(2) ”Älä ota korkoa veljeltäsi, älä rahasta, älä ruoasta äläkä mistään sellaisesta, mistä voi ottaa korkoa. ...”	”Älä peri toiselta israelilaiselta korkoa rahasta, ruoasta äläkä mistään muustakaan, mitä yleensä korkoa vastaan lainataan...” (23:20–21.)
(3) Schon hier zeigt sich, dass die Bibel Kredite nicht missbilligt, sondern ausdrücklich fordert, <u>dass die Reichen den Armen Geld oder Getreide leihen sollen</u> , wenn Notlagen drohen.	(3) Jo tämä osoittaa, että Raamattu ei paheksu lainoja, vaan nimenomaan vaatii rikkaita <u>lainaamaan rahaa tai viljaa köyhille</u> , kun hätä uhkaa.	Jo tästä huomaa, ettei Raamattu suinkaan tuomitse luottoja vaan nimenomaan vaatii rikkaita <u>lainaamaan köyhille rahaa tai viljaa</u> , kun nämä ovat vaarassa joutua hädänalaisiksi.
(4) <u>Doch</u> wer dessen Pressemitteilungen liest, gewinnt den Eindruck, dass meist eher unbedeutende Unternehmen attackiert werden.	(5) Jos <u>kuitenkin</u> lukee sen lehdistötiedotteita, saa sen vaikutelman, että se hyökkää lähinnä melko merkityksettömiä yrityksiä vastaan.	Sen lehdistötiedotteita lukiessa syntyy <u>kuitenkin</u> vaikutelma, että valvonta kohdistui ennemmin

		merkityksettömiin yrityksiin.
(5) Um den Wettbewerb zu schützen, hat Erhard 1958 sogar eigens das Bundeskartellamt gegründet.	(6) Vuonna 1958 Erhard perusti jopa liittovaltion kartelliviraston nimenomaan kilpailun suojelemiseksi.	Kilpailua suojelemaan Erhard perusti jopa erillisen Saksan liittotasavallan kartelliviraston vuonna 1958.

Esimerkissä 1 nähdään, että konekäännöksessä sana ”*pelkästään*” on sijoitettu väärin. Tämä virheellinen sanajärjestys on ongelmallinen, koska se vaikeuttaa lauseen ymmärtämistä ja lukija joutuu näkemään ylimääräistä vaivaa sen tarkan sisällön ymmärtämisen eteen. Luokittelin tämän virheen merkittäväksi ja annoin sille vakavuuspisteen 5, koska alkuperäisen lauseen merkitys jää epäselväksi ja sanoma ei välity. Koko käännöstekstiin suhteutettuna virhe ei kuitenkaan ole kriittinen. Korjaus, jossa sana ”*pelkästään*” sijoitetaan esimerkiksi lauseen eteen, selkeyttäisi lausetta: ”*Pelkästään lainojen takaisinmaksu ei kuitenkaan ollut usein vaikeaa.*”. Tämäkään lause ei kuitenkaan kuulosta täysin sujuvalta. Sen sijaan ihmiskäännös välittää alkuperäisen merkityksen tarkasti.

Esimerkissä 2 konekäännin on valinnut lauseen alkuosalle virheellisen sanajärjestyksen, mikä tekee lauseesta epäselvän. Lukija joutuu miettimään, mikä on lauseen ydinajatus. Tämä lähdekielen sitaatti on sellaisenaan saksankielisessä Raamatussa, mutta kone ei osannut löytää sille suoraa vastinetta suomenkielisestä Raamatusta, vaan on kääntänyt sen. Jos konekäännöksen sanajärjestyksestä muutetaan vähän, viesti selkenee: ”Älä ota veljeltäsi korkoa rahasta, ruoasta äläkä mistään sellaisesta, mistä voi ottaa korkoa.” Luokittelin tämän virheen merkittäväksi ja annoin sille vakavuuspisteen 5. Ihmiskääntäjä sen sijaan on käyttänyt suoraan suomenkielisen Raamatun käännöstä, mikä välittää alkuperäisen merkityksen täsmällisesti.

Esimerkissä 3 konekäännöksessä *että*-lauseessa sana ”*köyhille*” on sijoitettu väärin, mikä johtaa lauseen painotuksen vääristymiseen. Tämä virheellinen sanajärjestys ei tee lausetta käyttökelvottomaksi, vaikka painottaakin vähemmän tärkeää asiaa. Koska virhe ei haittaa lauseen kokonaismerkityksen välittymistä, se on vähäinen ja ansaitsee vakavuuspisteen 1. Sijoittamalla sanan ”*köyhille*” sanojen ”*rahaa tai viljaa*” eteen lauseesta tulee sujuvampi, mikä on tehty esimerkiksi ihmiskäännöksessä.

Esimerkissä 4 kone on sijoittanut sanan ”*kuitenkin*” väärään kohtaan eikä osannut liittää sitä oikein päälauseeseen kuuluvaksi. Vaikka virheellinen sanajärjestys ei estä lauseen

merkityksen ymmärtämistä, se muuttaa alkuperäisen lauseen painotusta. Virhe on vakavuusasteeltaan vähäinen, joten annoin sille vakavuuspisteen 1. Ihmiskääntäjä on sen sijaan tulkinnut yhteyden oikein.

Esimerkissä 5 konekäännöksen perusmerkitys on suunnilleen sama kuin lähdelauseessa, mutta painotus on erilainen. Lähdelauseessa painotetaan kartellinviraston perustamista, kun taas konekäännöksessä painotetaan kartellinviraston perustamiseen johtanutta syytä. Virheen luokittelin vähäiseksi, koska konekäännös ei kuitenkaan vääristä sanoman alkuperäistä merkitystä. Ihmiskäännöksessä painottuu kartellinviraston perustaminen, mikä välittää lähdekielen lauseen painotuksen oikein.

6.2.3 Terminologiavirheet

Konekäännöksessä esiintyi vähiten virheitä terminologiakategoriassa. Nämä virheet liittyivät lähes aina virheellisten termien valintaan. Käännetty termit eivät usein vastanneet alkuperäistä, minkä vuoksi lauseen merkitys ei välittynyt tarkasti tai jäi kokonaan epäselväksi. Esimerkkejä esitellään taulukossa 8.

Taulukko 8. Termien virheellinen käyttö

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Doch obwohl Babylonier und Assyrer noch keine Münzen kannten, wussten sie schon ganz genau, wie man <u>Zinseszins</u> berechnet.	(1) Vaikka babylonialaiset ja assyrialaiset eivät vielä tunteneet kolikoita, he tiesivät jo tarkalleen, miten <u>korkokorko</u> lasketaan.	Vaikka babylonialaiset ja assyrialaiset eivät vielä tunteneet kolikoita, he tiesivät jo tarkalleen, miten laskea <u>koronkorkoa</u> .
(2) Für Christen galt generell, dass sie keine <u>Zinsen nehmen</u> durften, weil alle Menschen als Brüder zu betrachten waren.	(2) Kristityt eivät yleensä saaneet <u>ottaa korkoa</u> , koska kaikkia ihmisiä oli pidettävä veljinä.	Kristityillä päti yleinen <u>koronperintäkielto</u> , koska kaikkia ihmisiä tuli kohdella veljinä.
(3) Manche waren dazu gedacht, den <u>Fernhandel</u> etwa mit Anatolien zu finanzieren.	(3) Joidenkin tarkoituksena oli rahoittaa <u>pitkän matkan kauppaa</u> esimerkiksi Anatolian kanssa.	Joillakin oli rahoitettu <u>kauppaa</u> esimerkiksi Anatolian kanssa.

Esimerkissä 1 kone on kääntänyt saksankielisen termin ”*Zinseszins*” suomeksi ”*korkokorko*”, mikä on virheellinen ja epätarkka käännös. Lähdetekstiä tuntematon lukija tuskin ymmärtää, mitä ”*korkokorko*” tarkoittaa. Lukija saattaisi olettaa, että kyse on pelkästään koron laskemisesta, vaikka todellisuudessa lasketaankin koronkorkoa, mikä on eri asia kuin pelkkä korko. Luokittelin tämän virheen merkittäväksi ja annoin sille vakavuuspisteen 5, koska

”korkokorko” ei välitä termin alkuperäistä merkitystä täsmällisesti. Ihmiskääntäjä sen sijaan on osannut valita termille oikean vastineen.

Esimerkissä 2 kone on kääntänyt ilmaisun ”Zinsen nehmen” suomeksi ”ottaa korkoa”. Vaikka tämä käänнос välittää kaiken olennaisen eikä vääristä lauseen merkitystä, se on silti ongelmallinen, koska se on epäsopiva asiayhteyteen. Talousalan kontekstissa sopivampi ilmaisu olisi ”periä korkoa”. Tämä virhe on juuri sellainen vähäinen virhe, jonka talouskirjallisuuden lukemiseen tottunut lukija pystyisi helposti korjaamaan tai päättelemään. Luokittelin virheen vähäiseksi, ja se sai vakavuuspisteen 1. Ihmiskääntäjä on valinnut termille oikean vastineen.

Esimerkissä 3 kone on kääntänyt termin ”Fernhandel” suomeksi ”pitkän matkan kauppa”, ja ihmiskääntäjä on valinnut käännökseksi sanan ”kauppa”. Luokittelin nämä käännosvaihtoehdot vähäisiksi virheiksi, ja ne saivat vakavuuspisteen 1. Vaikka nämä käännot eivät vääristä alkuperäisen termin merkitystä, ne eivät mielestäni ole täysin tarkkoja. Itse pitäisin parempana ja tarkempaa käännoksenä vaihtoehtoa ”kaukokauppa”.

Tapaus, jossa termiä käytettiin epä johdonmukaisesti, esitetään taulukossa 8.

Taulukko 8. Termien epä johdonmukainen käyttö

Lähdekieli (saksa)	Konekäännös	Ihmiskäännös
(1) Im Mittelalter gab es zwar zahllose <u>Märkte</u> , wie Straßenschilder namens »Kornmarkt«, »Gänsemarkt«, »Tuchmarkt«, »Fischmarkt«, »Heumarkt« oder »Holzmarkt« bezeugen. Aber diese <u>Märkte</u> machten einen eher kleinen Teil der Wirtschaft aus, während der Rest der Produktion feudal geprägt war.	Keskiajalla oli lukemattomia <u>toreja</u> , kuten "Kornmarkt", "Gänsemarkt", "Tuchmarkt", "Fischmarkt", "Heumarkt" tai "Holzmarkt" -nimiset katukyltit osoittavat. Nämä <u>markkinat</u> muodostivat kuitenkin melko pienen osan taloudesta, kun taas muulle tuotannolle oli ominaista feodalismi.	Keskiajalla oli lukuisia <u>markkinapaikkoja</u> , kuten esimerkiksi monista saksankielisistä katukylteistä voi edelleen havaita: Kornmarkt – viljatori, Gänsemarkt – hanhitori, Tuchmarkt – kangastori, Fischmarkt – kalatori, Heumarkt – heinätori tai Holzmarkt – puutori. Nämä <u>torit ja markkinapaikat</u> olivat vain pieni osa taloutta, muu tuotanto oli luonteeltaan feodaalista.

Esimerkissä 1 kone on kääntänyt sanan ”Märkte” suomeksi kahdella eri tavalla: ”toreja” ja ”markkinat”. Tämä on ongelmallista, koska sanoilla ”tori” ja ”markkinat” on eri merkitykset. Tämän takia käännoksen merkitys ei välity tarkasti. Saksassa sana ”Märkte” voi tarkoittaa

joko a) markkinoita tai b) kaupporeja/markkinapaikkoja. DeepL-käännöksessä näitä kahta sanaa on käytetty synonyymeinä, mikä on virheellistä. Oikea käännös tässä olisi joko *"kauppatorit/torit"* tai *"markkinapaikat"*.

7 Yhteenveto ja pohdinta

DeepL-käännöksen määrällisen ja laadullisen analyysin pohjalta voidaan nyt vastata tutkimuskysymyksiini, jotka olivat seuraavat:

- 1) Minkälaisia tarkkuuteen, sujuvuuteen ja terminologiaan liittyviä virheitä esiintyy käännettäessä saksankielisiä talousalan tekstejä suomeksi DeepL-kääntimellä?
- 2) Missä määrin konekäännöksissä esiintyvät semanttisen merkityksen vääristymät ovat haitallisia tekstin ymmärtämisen tai käytön kannalta?
- 3) Missä määrin DeepL-konekääntimen talousalan tekstistä tuottaman konekäännöksen laatu vastaa ihmisen tuottaman käännöksen laatua?

Sekä määrällinen että laadullinen analyysi osoittaa, että DeepL-käännös ei ole täysin virheetön. On myös syytä mainita, että DeepL-aineisto on tuotettu elokuussa 2023. Vaikka virheitä ilmeni, niiden vakavuus vaihteli, eivätkä ne tehneet käännöksestä suoraan käyttökeltvotonta. DeepL ei aina onnistunut tulkitsemaan kontekstia oikein, mikä heikensi käännöksen tarkkuutta. Lisäksi käännösohjelma kohtasi ajoittain vaikeuksia termien valinnassa ja monimerkityksisten sanojen tulkinnassa. Konekäännin ei aina osannut erottaa olennaista epäolennaisesta ja pyrki usein kääntämään lauseet sanatarkasti, mikä näkyi esimerkiksi lauseessa virheellisenä sanajärjestyksenä ja alkuperäisen merkityksen hämärtyminenä.

Tarkkuuskategoriassa havaitut virheet liittyivät pääasiassa virheellisiin käännöksiin, jotka johtivat lauseiden merkityksen epätarkkaan välittymiseen ja toisinaan semanttisiin vääristymiin. Nämä vääristymät heikensivät käännöslauseen ymmärrettävyyttä ja hankaloittivat sen tulkintaa, mutta eivät kuitenkaan tehneet koko arvioitua tekstiä käyttökeltvottomaksi.

Sujuvuuskategoriassa esiintyneet virheet koskivat pääasiassa virheellisiä sanajärjestyksiä. Konekäännin pyrki noudattamaan lähdetekstin sanajärjestyttä, mikä johti toisinaan vaikeaselkoisiin ilmaisuihin suomen kielessä. Tällaiset virheet toivat epätarkkuutta ja epäselvyyttä lauseiden merkitykseen, mikä hankaloitti alkuperäisen viestin ymmärtämistä, lisäksi viestin tulkintaan vaadittavaa aikaa ja hidasti lukemista.

Terminologiakategoriassa DeepL käänsi suurimman osan termeistä oikein, mikä oli positiivinen yllätys, sillä odotin löytäväni enemmän virheitä. Kuitenkin kaksimerkityksisten

termien kohdalla konekäännin kohtasi ongelmia. Näissä tapauksissa konekäännin ei aina onnistunut valitsemaan talouden kielen asiayhteyteen sopivaa termiä. Tämä havainto osoittaa, että konekääntimen kyky ymmärtää ja käyttää oikeita termejä voi olla rajallinen, erityisesti silloin, kun kyseessä ovat kaksi- tai monimerkityksiset sanat. Virheellisesti käännetty termit haittasivat lauseen ymmärtämistä ja käyttöä silloin, kun termi oli lauseen avainasia.

Vertailtuani DeepL-käännöstä ihmiskäännökseen huomasin selviä eroja. Ihmiskääntäjä tuotti sujuvampaa ja luonnollisempaa tekstiä, joka sopii paremmin kohdekielen kielellisiin normeihin. Konekäännin taas noudatti toisinaan lähdetekstin sanajärjestyttä liian tarkasti, mikä johti epäluonteviin lauseisiin. Tarkkuudessa ihmiskääntäjän käänös välitti lauseiden merkityksen oikein, myös vivahteet huomioiden, kun taas DeepL teki virheitä erityisesti kaksi- ja monimerkityksisten sanojen kohdalla. Ihmiskääntäjä osasi valita oikeat termit talouden kielen asiayhteyden perusteella, mutta DeepL käänsi joitakin termejä epätarkasti. Kokonaisuudessaan ihmiskäännös oli tarkempi ja paremmin kohdekieleen mukautettu, kun taas DeepL-käännös jäi epätarkemmaksi. Voidaankin todeta, että DeepL-käännös vaatii vielä jälkieditointia, jotta sen laatu ja ymmärrettävyys vastaisivat ihmiskäännöksen laatua.

Mitä tulee MQM-laatumittariin, se osoittautui hyödylliseksi työkaluksi virheiden luokittelussa ja niiden tyyppien määrittämisessä. Virhekategorioiden soveltaminen eli käänöksessä esiintyvien virheiden luokittelu oli mielestäni helppoa ja selkeää, mutta virheiden vakavuusasteiden määrittely tuotti vaikeuksia. Vaikka MQM-laatumittari tarjoaakin selkeät määritelmät vakavuusasteille, kriittisten ja merkittävien virheiden luokittelu ei ollutkaan niin yksinkertaista kuin olin odottanut – erityisesti virheellisten käännösten ja sanajärjestyksen osalta. Olin toisinaan epävarma, tulisiko virheellinen käänös luokitella kriittiseksi vai merkittäväksi virheeksi. Kriittinen virhehän merkitsisi koko käännöksen hylkäämistä. Virheellisten käännösten vakavuusasteen määrittelyssä oli otettava huomioon niiden vaikutus koko tekstiin ja lukijaan. Esimerkiksi lukijalle, joka ei ole aiemmin lukenut talousaiheista tietokirjaa, virheellinen tieto saattaisi näyttäytyä uutena ja saada hänet omaksumaan väärän tiedon. Tämän pohtiminen – millainen vaikutus virheellisellä tiedolla voi olla lukijalle ja millaisia seurauksia siitä voi aiheutua – oli aikaa vievää ja työlästä.

Vaikka MQM-mittari tarjoaa objektiivisen arviointikehyksen, arvioijan subjektiivisuus saattaa heikentää tulosten yleistettävyyttä ja luotettavuutta. Arvioijan aiemmat tiedot arvioitavasta aineistosta ja kokemus vaikuttavat siihen, miten hän luokittelee virheitä – sama virhe voi

hänen näkemyksestään riippuen olla joko vähäinen tai merkittävä. Jos arvioija tuntee arvioitavan tekstin aiheen hyvin, tekstin ymmärtäminen ja virheiden korjaaminen on helpompaa, ja siksi hän saattaa pitää virheitä vähäisinä. Vastaavasti kokemattomampi arvioija, jolla on lisäksi vähemmän tietoa aiheesta, voi pitää virheitä vakavimpana, koska tekstin ymmärtäminen on haastavampaa. Sama pätee myös lukijaan: Jos käännöksen lukija tuntee aihepiirin hyvin, käännöksessä esiintyvät virheet eivät välttämättä häiritse tekstin ymmärrettävyyttä, koska lukija pystyy korjaamaan virheet helposti mielessään. Jos taas lukija tutustuu talousaiheiseen kirjaan ensimmäistä kertaa, käännös voi vaikuttaa vaikeaselkoiselta.

Omaa arviointia tehdessäni luokittelin joitakin sujuvuusvirheitä vähäisiksi, vaikka toinen arvioija olisi saattanut pitää ne merkittävinä. Tämä johtui siitä, että olin lukenut lähdetekstin jo ennen sujuvuusarviointia ja minulla oli myös yleistietoa talousasioista. Tämä saattoi vaikuttaa objektiivisuuteeni, ja näin ollen monia sujuvuusvirheitä luokittelin vähäisiksi, koska koin ne helpoiksi korjata. Sujuvuusarvioinnin objektiivisuutta voisi parantaa, jos arvioijana toimisi henkilö, joka ei osaa saksaa mutta tuntee talousaiheet.

Tutkimustulokseni olivat suuntaa antavia, ja aihetta olisi hyvä tutkia edelleen.

Jatkotutkimuksen mahdollisuuksia voisi olla useita, erityisesti DeepL-kääntimen kehittämisen näkökulmasta. Esimerkiksi, voitaisiin tutkia, kuinka DeepL-käännösten laatu vaihtelee eri kieliparien välillä. Voisiko käännös olla tarkempi englannista suomeen kuin saksasta suomeen? Tämä tutkimus voisi paljastaa, mitkä kieliparit ovat haasteellisempia ja missä konekäännöksiä tulisi kehittää. Toinen jatkotutkimuksen aihe voisi olla käyttäjien, kuten ammattikäytäjäiden tai lukijoiden, käännöskokemusten tutkiminen. Voisi selvittää, miten he kokevat DeepL-käännösten sujuvuuden ja sisällön ymmärrettävyyden. Tämä voisi tarjota arvokasta tietoa siitä, missä tilanteissa konekäännös on riittävän hyvä ja milloin tarvitaan ihmiskääntäjää. Kolmas mahdollinen tutkimusaihe voisi liittyä jälkieditointiin. Voitaisiin tutkia, kuinka paljon DeepL-käännökset vaativat jälkieditointia verrattuna muihin konekääntimiin tai ihmiskääntäjien muokkaamiin käännöksiin. Tämä auttaisi arvioimaan, kuinka paljon työpanosta tarvitaan, jotta DeepL-käännöksiä voidaan hyödyntää tehokkaasti ammattikäytäjäissä ja käännöstöiden laskuttamisessa. Lisäksi voitaisiin tutkia tekoälyn parannusmahdollisuuksia, jotta DeepL oppisi käsittelemään monimerkityksisiä sanoja ja tunnistamaan erikoisalan kontekstin paremmin.

Lähteet

Aineslähteet

Herrmann, Ulrike 2013. *Der Sieg des Kapitals – Wie der Reichtum in die Welt kam: Die Geschichte von Wachstum, Geld und Krisen*. Westend Verlag GmbH, Frankfurt.

Herrmann, Ulrike 2015. *Pääoman voitto: Kasvun, rahan ja kriisien historia*. Suom. Mari Janatuinen. Into Kustannus Oy, Helsinki.

Tutkimuskirjallisuus

Backström, Heidi 2020. *Ihminen vastaan käänöskone*. [Verkojulkaisu.] Saatavissa: <https://www.goethe.de/ins/fi/fi/kul/mag/21750881.html?forceDesktop=1> [viitattu 19.12.2023].

Bericht DeepL-Test 2019. *Auswertung der Testergebnisse und Empfehlungen der Arbeitsgruppe „Maschinelle Übersetzung“ z.H. der KOSD*. [Verkojulkaisu.] Schweizerische Eidgenossenschaft. Saatavissa: <https://www.newsd.admin.ch/newsd/message/attachments/59735.pdf> [viitattu 19.11.2023].

Burchardt, Aljoscha – Egg, Markus – Eichler, Kathrin – Krenn, Brigitte – Kreutel, Jörn – Leßmöllmann, Annette – Rehm, Georg – Stede, Manfred – Uszkoreit, Hans – Volk, Martin 2020. Teoksessa: Georg Rehm ja Hans Uszkoreit (toim.) *Die deutsche Sprache im digitalen Zeitalter*. META-NET White Paper Series, 11–23. Saatavissa: <http://www.meta-net.eu/whitepapers/e-book/german.pdf> [viitattu 12.02.2024].

Chan, Sin-wai 2023. The development of translation technology. – *Routledge Encyclopedia of Translation Technology*, 2, 3–39. Routledge, London. Saatavissa: <https://doi-org.ezproxy.utu.fi/10.4324/9781003168348>

Cullen, Alana. *Understanding Neural Machine Translation. How artificial intelligence “works” in literary translation*. [Verkojulkaisu.] Saatavissa: <https://www.goethe.de/ins/gb/en/kul/lue/ail/21967556.html> [viitattu 19.02.2024].

Delorme Benites, Alice 2021. *Machinelle Übersetzung für akademische Texte*. [Verkojulkaisu.] ZHAW Zürcher Hochschule für Angewandte Wissenschaften. Saatavissa: <https://doi.org/10.21256/zhaw-22301> [viitattu 19.01.2024]

- ELIS Report 2022. *European Language Industry Survey 2022: Trends, expectations, and concerns of the European Language Industry*. [Verkkójulkaisu.] Saatavissa: <https://elis-survey.org/repository/> [viitattu 1.11.2023].
- Forcada, Mikel L. 2010. Machine Translation Today. Teoksessa: Yves Gambier ja Luc Van Doorslaer (toim.) *Handbook of Translation Studies*, 1, 215–222. Saatavissa: https://www.researchgate.net/profile/Barbara-Tannuri/publication/326381184_Handbook_of_Translation_Studies_-_Vol_1/links/5b4912aba6fdccadaec7cbe4/Handbook-of-Translation-Studies-Vol-1.pdf [viitattu 11.12.2023].
- Hagemann, Susanne 2022. *Recherche im Translationsprozess: Ein Lehrbuch- und Studienbuch*, 129, 53–63. Frank&Timme, Berlin. Saatavissa: <https://library.oapen.org/handle/20.500.12657/53673>
- Kenny, Dorothy 2018. Machine Translation. Teoksessa: Rawling, J.Piers ja Wilson, Philip (toim.) *The Routledge Handbook of Translation and Philosophy*, 1, 428–445. Routledge, London. Saatavissa: <https://doi-org.ezproxy.utu.fi/10.4324/9781315678481>
- Koehn, Philipp – Knowles, Rebecca 2017. *Six Challenges for Neural Machine Translation*. [Verkkójulkaisu.] Saatavissa: <https://doi.org/10.48550/arxiv.1706.03872> [viitattu 20.11.2023].
- Koehn, Philipp 2020. Current Challenges. – *Neural Machine Translation*, 293–309. Cambridge University Press, Cambridge. Saatavissa: <https://doi-org.ezproxy.utu.fi/10.1017/9781108608480>
- Konttinen, Kalle – Veivo, Outi 2008. *Kääntämisen arviointi työelämässä ja koulutuksessa*. *MikaEL*, 2, 1–16. [Verkkójulkaisu.] Saatavissa: https://www.sktl.fi/@Bin/41023/Konttinen_Veivo.pdf [viitattu 10.12.2023].
- Koponen, Maarit 2017. *Korvaako kone kääntäjän?* [Verkkójulkaisu.] Saatavissa: <https://blogit.utu.fi/utu/2017/04/28/korvaako-kone-kaantajan/> [viitattu 20.12.2023].
- Koponen, Maarit 2022. Teknologiavälitteinen kääntäminen: Kääntäjä ja kone käännosten tuotantotaloudessa. – *Tiede ja edistys*, 47 (4), 277–278. Saatavissa: <https://doi.org/10.51809/te.120121> [viitattu 23.11.2023].
- Koskenniemi, Kimmo – Lindén, Krister – Carlson, Lauri – Vainio, Martti – Arppe, Antti – Mietta, Lennes – Westerlund, Hanna – Hyvärinen, Mirka – Bartis, Imre – Nuolijärvi, Pirkko – Piehl, Aino 2020. Teoksessa: Georg Rehm ja Hans Uskoreit (toim.) *Suomi*

- digitaalisella aikakaudella*. META-NET White Paper Series, 11–23. Saatavissa: <http://www.meta-net.eu/whitepapers/e-book/finnish.pdf> [viitattu 12.02.2024]
- Kriston, Andrea 2017. Machine translation in translating business texts: Myth versus Reality. – *Professional Communication and Translation Studies*, 10, 126–136. Saatavissa: https://dspace.upt.ro/jspui/bitstream/123456789/5263/1/BUPT_ART_Kriston_f.pdf [viitattu 10.10.2023].
- Latomaa, Sirkku – Nurminen, Mary 2022. Työnkuvan muutos: kääntäjästä jälkieditoijaksi. – *Kieli, koulutus ja yhteiskunta*, 13(3). Saatavissa: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2022/tyonkuvan-muutos-kaantajasta-jalkieditoijaksi> [viitattu 10.11.2023].
- Lommel, Arle 2018. Metrics for Translation Quality Assessment: A case for Standardising Error Typologies. Teoksessa: Joss Moorkens, Sheila Castilho, Federico Gaspari ja Stephen Doherty (toim.) *Translation Quality Assessment: From Principles to Practice*, – *Translation Quality Assessment: From Principles to Practice*, 109–127. Springer Cham. Saatavissa: <https://doi.org/10.1007/978-3-319-91241-7>
- Matusov, Evgeny 2019. The Challenges of Using Neural Machine Translation for Literature. – *Proceedings of the Qualities of Literary Machine Translation*, 10–19. European Association for Machine Translation, Ireland. Saatavissa: <https://aclanthology.org/W19-7302.pdf> [viitattu 12.11.2023].
- Maučec, Mirjam Sepesy – Donaj, Gregor 2019. Machine Translation and the Evaluation of Its Quality. Teoksessa: Ali Sadollah ja Tilendra Shishir Sinha (toim.) *Recent Trends in Computational Intelligence*, 149–151. IntechOpen, London. Saatavissa: https://mts.intechopen.com/storage/books/8465/authors_book/authors_book.pdf [viitattu 12.11.2023].
- Melby, K. Alan 2012. Human and machine translation quality: Definable? Achievable? Desirable? 1–29. [Verkkójulkaisu.] Saatavissa: <https://www.ttt.org/wp-content/uploads/2019/01/2018-12-17-LACUS2012-melby-formatted-v9h.pdf> [viitattu 10.12.2023].
- Mohamed, Shereen A. – Ashraf, A. Elsayed – Hassan Y.F. – Mohamed A. Abdou 2021. Neural machine translation: Past, present, and future. – *Neural Computing & Applications*, 33 (23), 15919–15931. Saatavissa: <https://www.proquest.com/scholarly-journals/neural-machine-translation-past-present-future/docview/2592776230/se-2> [viitattu 10.12.2023].

- MQM (Multidimensional Quality Metrics). [Online-tietokanta.] Saatavissa: <https://themqm.org/> [viitattu 9.11.2023].
- Neidhardt, Miriam 2020. *Buchübersetzungen mit DeepL?!* [Verkkajulkaisu.] Saatavissa: <https://www.miriam-neidhardt.de/2020/09/03/buchuebersetzungen-mit-deepl/> [viitattu 12.01.2024].
- Nitzke, Jean – Hansen-Schirra, Silvia 2021. A short guide to post-editing. – *Translation and Multilingual Natural Language Processing*, 16, 22–26. Language Service Press, Berlin. Saatavissa: <https://doi.org/10.5007/2175-7968.2023.e85818> [viitattu 3.2.2024].
- Popović, Maja 2018. Error Classification and Analysis for Machine Translation Quality Assessment. Teoksessa: Joss Moorkens, Sheila Castilho, Federico Gaspari ja Stephen Doherty (toim.) *Translation Quality Assessment: From Principles to Practice*, 129–158. Toim. Cham Springer. Saatavissa: <https://doi.org/10.1007/978-3-319-91241-7>
- Rajala, Aino 2013. Mitä eroa on tiedekirjoilla ja tietokirjoilla? – *Suomalaisen Kirjallisuuden Seura (SKS)*. Helsinki. Saatavissa: <https://www.finlit.fi/ajankohtaista/blogi/mita-eroa-on-tiedekirjoilla-ja-tietokirjoilla/> [viitattu 14.12.2023].
- Rivera-Trigueros, Irene 2022. Machine translation systems and quality assessment: a systematic review. – *Lang Resources & Evaluation*, 56, 593–619. Saatavissa: <https://doi.org/10.1007/s10579-021-09537-5> [viitattu 13.1.2024].
- Roslöf, Raija – Veitonen, Ulla 2006. Suomalaisen viittomakielen kääntäminen ja käännösteoriat. Teoksessa: Niina Hytönen ja Terhi Rissanen (toim.) *Käden käänteessä: Viittomakielen kääntämisen ja tulkkauksen teoriaa sekä käytäntöä*, 113. Finn Lectura, Helsinki.
- Sarkhel, Juran Krishna – Tripathi, Sneha 2010. Approaches to machine translation. – *Annals of Library and Information Studies*, 57, 388–393. [Verkkajulkaisu.] Saatavissa: file:///C:/Users/OMISTAJA/Downloads/Approaches_to_machine_translation.pdf [viitattu 13.2.2024].
- Schikowski, Michael 2011. *Sachbuch, Ratgeber, Fachbuch -Unterschiede und Überschneidungen*, 63–80. [Verkkajulkaisu.] Saatavissa: <https://www.immer-schoen-sachlich.de/wp-content/uploads/Schikowski-2011-Sachbuch-Ratgeber-Fachbuch-Unterschiede-und-Ueberschneidungen.pdf> [viitattu 23.1.2024].
- Schmalz, Antonia 2019. Maschinelle Übersetzung. Teoksessa: Volker Wittpahl (toim.) *Künstliche Intelligenz*, 194–211. Springer Vieweg, Berlin, Heidelberg. Saatavissa: https://doi.org/10.1007/978-3-662-58042-4_12 [viitattu 13.11.2023].

- Shterionov, Dimitar – Superbo, Riccardo –Nagle, Pat – Casanellas, Laura – O’ Dowd, Tony – Way, Andy 2018. Human versus automatic quality evaluation of NMT and PBSMT. – *Machine Translation* 32(3), 217–235. Springer Netherlands. Saatavissa: <https://doi.org/10.1007/s10590-018-9220-z> [viitattu 14.11.2023].
- Silva, Beatriz – Buchicchio, Marianna – van Stigt, Daan – Stewart, Craig – Moniz, Helena 2024. Data-driven Asian Adapted MQM Typology and Automation in Translation Quality Workflow. – *The Journal of Specialised Translation*, 41, 1–29. Saatavissa: <https://www.jostrans.org/article/view/4713/4240> [viitattu 10.2.2024].
- Snow, Tyler A. 2015. *Establishing the Viability of the Multidimensional Quality Metrics Framework*, 15–21. [Pro gradu -tutkielma.] Brigham Young University, Provo. Saatavissa: <https://scholarsarchive.byu.edu/etd/5593/> [viitattu 4.03.2024].
- Somers, Harold 2012. Machine Translation: History, Development, and Limitations. Teoksessa: Kirsten Malmkjær ja Kevin Windle (toim.) *The Oxford Handbook of Translation Studies*, 427– 440. Oxford Academic. Saatavissa: <https://doi-org.ezproxy.utu.fi/10.1093/oxfordhb/9780199239306.013.0029>
- Terribile, Silvia 2023. Is post-editing really faster than human translation? –*Translation Spaces*. [Verkkójulkaisu.] Saatavissa: <https://arxiv.org/ftp/arxiv/papers/2312/2312.12660.pdf> [viitattu 10.01.2024].
- Tieteen termipankki 2024. Kielitiede: luonnollinen kieli. [Online-tietokanta.] Saatavissa: https://tieteentermipankki.fi/wiki/Kielitiede:luonnollinen_kieli [viitattu 30.4.2024].
- Tieteen termipankki 2024. Käännöstiede: konekääntäminen. [Online-tietokanta.] Saatavissa: <https://tieteentermipankki.fi/wiki/Käännöstiede:konekääntäminen> [viitattu 28.2.2024].
- Tiittula, Liisa 2014. *Johdatusta kääntämiseen ja käännöstieteeseen*, 4–7. [Verkkójulkaisu.] Saatavissa: <http://docplayer.fi/21520733-Johdatusta-kaantamiseenja-kaannostieteeseen.html> [viitattu 05.01.2024].
- Tuominen, Tiina – Koponen, Maarit – Vitikainen, Kaisa – Sulubacak, Umut – Tiedemann, Jörg 2023. Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution. – *The Journal of Specialised Translation* 39, 77–98. Saatavissa: https://jostrans.soap2.ch/issue39/art_tuominen.php [viitattu 20.11.2023].
- Vilokkinen, Natasha 2017. *Tiedontuojat – Opas tietokirjan suomentajalle*, 16–85. Vastapaino, Tampere.
- Voutilainen, Antti 2021. *Mihin konekäännintä kannattaa käyttää?* [Verkkójulkaisu.] Saatavissa: <https://blog.acolad.com/fi/mihin-konekaanninta-kannattaa-kayttaa> [viitattu 1.02.2024].

- Wang, Haifeng – Wu, Hua – He, Zhongjun – Huang, Liang – Church, Kenneth Ward 2022. Progress in Machine Translation. – *Engineering* 18, 143–153. Elsevier. Saatavissa: <https://doi.org/10.1016/j.eng.2021.03.023> [viitattu 3.01.2024].
- Way, Andy 2018. Quality Expectations of Machine Translation. Teoksessa: Joss Moorkens, Sheila Castilho, Federico Gaspari ja Stephen Doherty (toim.) *Translation Quality Assessment: From Principles to Practice*, 164–165. Springer International Publishing AG. Saatavissa: <https://doi.org/10.1007/978-3-319-91241-7>
- Webster, Rebecca – Fonteyne, Margot – Tezcan, Arda – Macken, Lieve – Daems, Joke 2020. Gutenberg Goes Neural: Comparing Features of Dutch Human Translations with Raw Neural Machine Translation Outputs in a Corpus of English Literary Classics. – *Informatics* 7 (3), 32. Gent. Saatavissa: <https://doi.org/10.3390/informatics7030032> [viitattu 10.11.2023].
- Werthmann, Antonina – Witt, Andreas 2014. Maschinelle Übersetzung: Gegenwart und Perspektiven. Teoksessa: Gerhard Stickel (toim.) *Translation and interpretation in Europe: Contributions to the Annual Conference 2013 of EFNIL in Vilnius*, 79–103. Lang, Frankfurt am Main. Saatavissa: <https://core.ac.uk/download/pdf/83654027.pdf> [viitattu 10.02.2024].

Liitteet

Liite 1. Deutsche Zusammenfassung

Bewertung der Qualität maschineller Übersetzungen von Wirtschaftstexten beim Übersetzen vom Deutschen ins Finnische

Der Einsatz maschineller Übersetzungssysteme wird heute aufgrund ihrer Vorteile immer üblicher. Sie bieten eine effiziente Möglichkeit, Texte in verschiedene Sprachen zu übersetzen, und tragen dazu bei, Sprachbarrieren zu überwinden und die mehrsprachige Kommunikation zu fördern. Maschinelle Übersetzungen können als Hilfsmittel eingesetzt werden, wenn eine menschliche Übersetzung nicht zur Verfügung steht, und helfen, den Hauptinhalt eines Textes zu verstehen. Maschinenübersetzer werden auch von professionellen Übersetzern eingesetzt, z. B., um eine Rohübersetzung anzufertigen. Sie werden auch eingesetzt, um Online-Chats in verschiedenen Sprachen zu überwachen, was z. B. bei kriminalpolizeilichen Ermittlungen nützlich sein kann. Maschinenübersetzer werden auch von den Streitkräften und internationalen Hilfsorganisationen als nützlich angesehen, da sie eine gute Möglichkeit zur Verständigung bieten, insbesondere in Krisengebieten, wo sie häufig auf Menschen treffen, die seltene oder wenig verbreitete Sprachen sprechen.

Geschichte der maschinellen Übersetzung

Die ersten maschinellen Übersetzer wurden auf der Grundlage der regelbasierten Methode (RBMT) entwickelt, bei der zweisprachige Wörterbücher und manuell geschriebene Regeln verwendet wurden. Die manuelle Erstellung und Pflege großer zweisprachiger Wörterbücher und Regelsammlungen war jedoch sehr zeitaufwändig, so dass nach neuen Methoden gesucht wurde, die mit weniger Aufwand bessere Ergebnisse erzielen konnten. Ende der 1980er Jahre erfuhr der Bereich der maschinellen Übersetzung mit der Erprobung der statistischen Methode (kurz SMT), die zu den korpusbasierten Methoden gehört, einen großen Richtungswechsel. Dabei werden keine sprachlichen Informationen verwendet, sondern das Ziel besteht darin, mit höchster Wahrscheinlichkeit den entsprechenden Satz in der Zielsprache zu finden. Die Methode basiert auf der Idee, dass es mit ausreichender Rechenleistung und Übersetzungsdaten möglich ist, aus dem Ausgangstext die wahrscheinlichsten Zielwörter/Wörter der Zielsprache und deren Reihenfolge vorherzusagen.

Im Vergleich zur regelbasierten Methode war die statistische Methode kostengünstiger. Die Erstellung großer Korpora war jedoch aufwändig. Die statistische Methode erreichte bald ein Stadium, in dem eine signifikante Verbesserung nur schwer zu erreichen war, was zur Suche nach neuen Ansätzen führte.

Eine neuere Methode wurde in der neuronalen Netzwerkübersetzung (NMT) gefunden, die seit 2016 weit verbreitet ist. Obwohl sie zu den korpusbasierten Methoden gehört, wird sie oft als datenbasierte Methode bezeichnet. NMT verwendet Übersetzungsalgorithmen, die selbstlernende Systeme sind und sich nicht auf vordefinierte Sprachregeln stützen. Ein neuronales Netzwerkübersetzer (NMT) besteht aus einem großen neuronalen Netzwerk, das einen Encoder und einen Decoder umfasst. Der Encoder empfängt die Wörter des Ausgangssatzes, verarbeitet sie und wandelt die Informationen in einen semantischen Kontextvektor um, den der Decoder verwendet, um den Zielsatz Wort für Wort zu erzeugen. Das Netzwerk soll auf einem Modell der Neuronen im menschlichen Gehirn beruhen.

Die Vorteile und Herausforderungen der NMT

Die neuronale Netzwerkübersetzung (NMT) hat die statistische maschinelle Übersetzung (SMT) erfolgreich ersetzt und ist dank ihrer Einfachheit und vielversprechenden Übersetzungsergebnisse sehr beliebt. NMT-Übersetzungen sind oft flüssig, grammatikalisch korrekt und verständlich. Nichtsdestotrotz kann ihre Flüssigkeit die Genauigkeit beeinträchtigen. Während sich NMTs hervorragend für die Übersetzung von Texten wie Bedienungsanleitungen eignen, sind sie für Belletristik oder kreative Texte weniger zuverlässig. Auch wenn der übersetzte Text flüssig zu sein scheint, kann es Probleme bei der Vermittlung der ursprünglichen Bedeutung geben.

NMTs haben auch Probleme, wenn sie in eine Zielsprache mit weniger Daten und einem komplexen Flexionssystem übersetzen müssen oder wenn sich Ausgangs- und Zielsprache grammatikalisch stark unterscheiden und zu verschiedenen Sprachgruppen gehören. Auch die Übersetzung von Fachgebieten kann Probleme verursachen. So sind die NMT-Übersetzungen zwar oft flüssig, können aber dennoch Probleme mit der Genauigkeit haben. Aus diesem Grund ist es wichtig, die Qualität von maschinellen Übersetzern zu untersuchen.

NMT-Übersetzungen sind nicht immer zuverlässig, was den Inhalt oder die sprachliche Genauigkeit betrifft, insbesondere wenn sie in eine kleinere Sprache wie Finnisch übersetzt

werden. Daher ist es wichtig, die Qualität von NMT-Übersetzungen zu überwachen, insbesondere bei Texten, bei denen es auf Genauigkeit ankommt. Untersuchungen zu NMT-Übersetzungen können den Entwicklern von Übersetzungssoftware helfen zu verstehen, wo das System verbessert werden muss, und so die Leistung von NMT-Übersetzern zu verbessern.

Ziele der Untersuchung und Forschungsfragen

Ziel dieser Arbeit war es, die Qualität von NMT-Übersetzungen zu bewerten und festzustellen, inwieweit sie die Qualität menschlicher Übersetzungen erreichen können. Sie untersuchte die Eignung des DeepL-Übersetzers für die Übersetzung von Wirtschaftstexten aus dem Deutschen ins Finnische.

Ziel der Untersuchung war es, die folgenden Fragen zu beantworten:

- 1) Welche Art von Fehlern in Bezug auf Genauigkeit, Flüssigkeit und Terminologie treten bei der Übersetzung deutscher Wirtschaftstexte ins Finnische mit dem Übersetzer DeepL auf?
- 2) Inwieweit beeinträchtigen semantische Bedeutungsverzerrungen in maschinellen Übersetzungen das Verständnis oder die Verwendung des Textes?
- 3) Inwieweit entspricht die Qualität der maschinellen Übersetzung eines Wirtschaftstextes, die mit DeepL erstellt wurde, der Qualität einer menschlichen Übersetzung?

In der Arbeit wurden Daten aus einer Übersetzung des Web-Übersetzers DeepL und einer finnischen Humanübersetzung verwendet. Bei den analysierten Daten handelte es sich um Textauszüge zu Wirtschaftsthemen, die die Volkswirtschaft in allgemeiner Form beschreiben, und zwar Auszüge aus dem Sachbuch *Der Sieg des Kapitals* von Ulrike Herrmann sowie dieselben Passagen des Buches, die jedoch mit dem Übersetzer DeepL übersetzt wurden. Da die Analyse des fast 300 Seiten umfassenden Buches nicht in den Rahmen der Arbeit passte, habe ich exemplarisch einige Auszüge aus den Texten analysiert. Um die inhaltliche Vielfalt zu gewährleisten, habe ich die ersten beiden Seiten der Kapitel 6, 9 und 11 des Buches ausgewählt, die sich mit verschiedenen wirtschaftlichen Phänomenen wie Kapitalismus, Geld, Schulden, Zinsen, Finanzkrisen und Inflation befassen. Als Referenzübersetzung diente die

Arbeit der Übersetzerin Mari Janatuinen von dem Buch *Pääoman voitto*. Die Referenzübersetzung wurde zum Vergleich der maschinellen Übersetzung mit der menschlichen Übersetzung herangezogen, um festzustellen, inwieweit sich die maschinelle Übersetzung von der menschlichen Übersetzung unterscheidet.

Die Qualität der maschinellen Übersetzungen

Die Bewertung der Qualität maschineller Übersetzungen ist nicht einfach, und es gibt keine einzig richtige Methode. Es gibt auch keine Standarddefinition von Übersetzungsqualität. Die Qualität hängt u. a. davon ab, wie die Übersetzung verstanden wird, was von der Übersetzung erwartet wird und was das Ziel ist, das bewertet wird. Die Qualität einer Übersetzung ist kein festes Konzept, sondern wird durch die Anforderungen der Übersetzungsaufgabe bestimmt. Eine Übersetzung kann genau und qualitativ hochwertig sein, aber dennoch nutzlos, wenn sie bestimmte Kriterien nicht erfüllt. Auch wenn es keine eindeutige Definition von Übersetzungsqualität gibt, ist es dennoch wichtig, sich vor einer Bewertung auf ein gemeinsames Verständnis des Qualitätsbegriffs zu einigen, denn ohne diese Einigung kann keine Bewertung durchgeführt werden. Die Bewertungsmethoden variieren je nach den Qualitätsanforderungen, die an die Übersetzung gestellt werden. Diese Anforderungen sollten die Erwartungen und Bedürfnisse der Endbenutzer widerspiegeln. Es ist auch wichtig zu bestimmen, was bewertet wird: die Übersetzung selbst, der Übersetzungsprozess oder das Übersetzungsprojekt als Ganzes.

Der Gegenstand meiner Untersuchung war das Produkt der Übersetzung, die Übersetzung selbst. Da ich ein Sachbuch über Wirtschaft als Ausgangsmaterial verwendet habe, wurde die Qualität der Übersetzung anhand ihrer Funktionalität, d. h. des Zwecks des Textes, bewertet. Ein Sachbuch sollte den Leser mit klaren und präzisen Informationen versorgen, und die Übersetzung sollte die wirtschaftlichen Themen genau und flüssig wiedergeben. Da es sich bei der Übersetzung um eine Form der Kommunikation handelt, ist es auch wichtig, dass die Botschaft aus dem Ausgangstext in den Zieltext auf verständliche Weise übertragen wird. Dies erfordert ein gewisses Maß an Äquivalenz, das sich in Inhalt, Stil oder Form ausdrücken lässt. Nicht alle Äquivalenzgrade werden immer erreicht, aber man muss abwägen, was in einer bestimmten Situation am wichtigsten ist. In meiner Untersuchung habe ich mich vor allem auf die semantische Äquivalenz konzentriert, d. h. auf die Vermittlung von Bedeutung

der Sachinhalte. In einem Informationsbuch ist dies wichtig, weil der Leser neue Informationen erhalten soll.

Qualitätsbewertungsmethoden für die maschinelle Übersetzung

Wenn der Gegenstand der Qualitätsbewertung die Übersetzung als Produkt ist, besteht eine Möglichkeit, die Qualität einer maschinellen Übersetzung manuell, d.h. menschlich, zu bewerten. Bei einer manuellen Bewertung bezieht sich die Übersetzungsqualität in der Regel auf die Genauigkeit und die Flüssigkeit der Übersetzung. Die Genauigkeit bezieht sich auf die Bewertung der semantischen Qualität, d. h. ob die in der Übersetzung vermittelten Informationen korrekt sind oder nicht. Die Flüssigkeit/Geläufigkeit hingegen bezieht sich auf die syntaktische Qualität. Der Aspekt der Geläufigkeit/Flüssigkeit konzentriert sich auf die Einhaltung der Normen des Zieltextes und der Zielsprache, wie z. B. grammatikalische Merkmale und Klarheit.

Die manuelle Bewertung ist zeitaufwändig, mühsam, teuer und subjektiv. Diese Herausforderungen haben zur Entwicklung automatisierter Bewertungswerkzeuge wie TER, METEOR, BLEU und NIST geführt, die angeblich schneller, billiger und objektiver sind. BLEU ist das beliebteste Verfahren, da es hochpräzise Ergebnisse liefert, die stark mit menschlichen Urteilen korrelieren. BLEU vergleicht die maschinelle Übersetzung mit der Referenzübersetzung und erzeugt eine BLEU-Bewertung/Score. Die BLEU-Bewertung gibt keine Auskunft darüber, ob die Übersetzung gut oder schlecht ist, sondern misst nur die Ähnlichkeit zwischen der maschinellen Übersetzung und der Referenzübersetzung. Es ist möglich, dass eine maschinelle Übersetzung gut abschneidet, aber eine niedrige Punktzahl erreicht, wenn sie andere Begriffe als die Referenzübersetzung verwendet.

Die größte Herausforderung bei der BLEU-Metrik ist die Annahme, dass jeder Text nur auf eine Weise gut übersetzt werden kann. Für Ausgangstexte kann es jedoch mehrere geeignete Übersetzungsmöglichkeiten geben. Da es sich bei meinen Forschungsdaten um Sachtexte handelt, für die es oft viele richtige Übersetzungsmöglichkeiten gibt, ist BLEU für meine Forschungszwecke nicht geeignet. Darüber hinaus erkennen die automatischen Messungen nicht die Schwere der Fehler, was sie für die Analyse von Übersetzungsfehlern ebenfalls ungeeignet macht. In meiner Studie wurde eine standardisierte MQM-Qualitätsmetrik verwendet, die eine manuelle Bewertung von Fehlern ermöglicht.

Bei einer manuellen Bewertung kann die Qualität der Übersetzung entweder ganzheitlich oder analytisch beurteilt werden. Eine ganzheitliche Bewertung gibt in der Regel einen Überblick über die Qualität der Übersetzung, liefert aber möglicherweise nicht genügend Detailinformationen für eine gezielte Verbesserung der Übersetzung. Die analytische Bewertung hingegen konzentriert sich auf detaillierte Teile des Textes, wie z. B. Wörter und Sätze, und zeigt genau auf, wo im Text Fehler zu finden sind.

Obwohl die analytische Bewertung genau ist, kann die Wahl der richtigen Metrik eine Herausforderung darstellen. In meiner Untersuchung wurde die MQM-Fehlerklassifikationsmetrik wegen ihrer Vielseitigkeit und Flexibilität verwendet. Die MQM-Metrik ermöglicht eine fehlerzentrierte Bewertung, und ich habe sie zur manuellen Analyse der Qualität der Übersetzung DeepL verwendet.

Die MQM-Qualitätsmetrik

Die MQM-Qualitätsmetrik ist sehr umfassend und wurde unter Berücksichtigung früherer Indikatoren wie SAE J2450 TQM und dem LISA-QA-Modell entwickelt. Der MQM ist ein Rahmen für die Bewertung der Übersetzungsqualität, der es den Prüfern ermöglicht, ihre eigenen Metriken für die Qualitätsbewertung anzupassen. Dieses Bewertungsmodell wurde ursprünglich entwickelt, um die Qualität von maschinellen Übersetzungen zu bewerten, wird aber inzwischen auch für die Bewertung menschlicher Übersetzungen verwendet. Mit dem MQM können Benutzer faire und unvoreingenommene Vergleiche zwischen maschinellen und menschlichen Übersetzungen anstellen und dabei die Subjektivität vieler bestehender Bewertungsmethoden vermeiden, indem sie sich auf bestimmte Fehlertypen konzentrieren. Das MQM umfasst in der Regel ausgewählte und gewichtete Fehlerkategorien, und Fehler können bis hinunter zur Wortebene gekennzeichnet werden, was eine fehlerzentrierte Bewertung ermöglicht. Die fehlerzentrierte Bewertung hebt spezifische Problembereiche in der Übersetzung hervor und hilft, einzelne Mängel in der Übersetzung zu verbessern. Bei der fehlerzentrierten Bewertung können Übersetzungstexte im Hinblick auf eine oder mehrere verschiedene Fehlerarten untersucht werden. Sie unterteilt Fehler in Fehlertypen, die Dimensionen bilden.

MQM-Fehlerkategorien/-typen helfen dabei, potenzielle Probleme im Zieltext zu erkennen, die durch Fehler oder absichtliche Änderungen verursacht werden können. Die MQM-Metrik

kann nicht in vollem Umfang für alle Übersetzungen verwendet werden, da sie umständlich in der Anwendung ist. Bei der Bewertung der Übersetzungsqualität müssen nicht unbedingt alle Dimensionen und Unterkategorien des MQM verwendet werden, sondern sie werden je nach den Anforderungen der Übersetzung ausgewählt.

Es gibt zwei Versionen der MQM-Qualitätsmetrik: das umfangreichere MQM Full und das übersichtlichere MQM Core. MQM Full bietet eine detailliertere Auswahl an Fehlertypen, die verwendet werden kann, wenn eine größere Genauigkeit bei Übersetzungen erforderlich ist. MQM Core hingegen ist auf eine engere Auswahl von Fehlerkategorien beschränkt. Für meine Studie habe ich die engere Version der MQM-Metrik verwendet, da sie, obwohl die Auswahl kleiner ist, doch genügend Fehlerkategorien zur Beantwortung meiner Forschungsfragen bietet.

Ober -und Subkategorien der MQM-Fehlertypologie

Das MQM umfasst insgesamt sieben übergeordnete Dimensionen bzw. Oberkategorien: Genauigkeit, Geläufigkeit/Flüssigkeit, Terminologie, Stil, lokale Konventionen, Design und Relevanz. Jede dieser Dimensionen enthält spezifischere Fehler-Subtypen oder Unterkategorien, die eine Reihe von qualitätsbezogenen Problemen abdecken. In meiner Analyse habe ich mich nur auf den flüssigen Ausdruck, die Genauigkeit und die Terminologie konzentriert, da diese Bereiche für das von mir ausgewählte Sachbuch über Wirtschaft von wesentlicher Bedeutung sind. Der Inhalt des Sachbuchs, das Glossar, der Zweck des Sachbuchs und das erwartete Zielpublikum waren wichtige Faktoren bei meiner Auswahl. Da es sich bei dem zu forschenden Sachbuch um ein Wirtschaftsbuch handelte, legte ich besonderen Wert auf die Verwendung korrekter und einheitlicher Wirtschaftsbegriffe in der Übersetzung. Sachbücher enthalten oft zahlreiche Fakten, so dass die Übersetzung deren genaue Bedeutung vermitteln muss. Außerdem habe ich berücksichtigt, dass die Leserschaft eines Sachbuchs sehr unterschiedlich sein kann, so dass Lesbarkeit und Verständlichkeit von entscheidender Bedeutung sind. Von den Unterkategorien habe ich diejenigen für die Forschung ausgewählt, die im Wesentlichen mit Sprachfluss, Genauigkeit und Terminologie zu tun haben.

Die Oberkategorie Genauigkeit enthält 7 Subkategorien, von denen jedoch nur 4 für die Studie ausgewählt wurden: Hinzufügung, falsche Übersetzung, Auslassung und Unübersetzt.

Die Oberkategorie Geläufigkeit/Flüssigkeit enthält 5 Subkategorien, aber nur 4 wurden für die Studie ausgewählt: Grammatik, Zeichensetzung, Rechtschreibung und mehrdeutiger Text. Die Oberkategorie Terminologie enthält 3 Subkategorien, von denen nur 2 für die Analyse ausgewählt wurden: falsche Terminologie und inkonsistente Verwendung von Begriffen. Obwohl die MQM-Qualitätsmetrik eine genauere Identifizierung und Korrektur von Fehlern ermöglicht, ist die bloße Klassifizierung von Fehlern für die Qualitätsbewertung nicht ausreichend. Ein einzelner Fehler sagt nichts darüber aus, ob eine Übersetzung von hoher Qualität ist. Die MQM-Metrik legt daher auch Wert auf die Gewichtung der Fehler, mit der die Bedeutung der Fehler entsprechend ihrer Auswirkung auf eine bestimmte Übersetzung bestimmt werden kann. Wenn alle Fehlertypen gleich wichtig sind, erhalten sie den gleichen Gewichtungsfaktor 1. Soll einigen Fehlertypen eine größere Bedeutung beigemessen werden, kann ein höherer Gewichtungsfaktor 2, verwendet werden. Da es sich bei den betrachteten Daten um Textauszüge aus einem wirtschaftsbezogenen Sachbuch handelte, wurde davon ausgegangen, dass alle ausgewählten Fehlerkategorien, d. h. Genauigkeit, Geläufigkeit/Flüssigkeit und Terminologie, und ihre jeweiligen Fehlertypen angesichts der Merkmale und des Zwecks des Sachbuches von gleicher Bedeutung waren. Daher erhielten alle Fehlertypen den gleichen Gewichtungsfaktor, d.h. 1.

Bei der MQM-Metrik beschränkt sich die Bewertung der Übersetzungsqualität nicht darauf, die Anzahl und Gewichtung der Fehlertypen zu kennen, sondern auch zu wissen, wie schwerwiegend und wichtig die Fehlertypen für das zu bearbeitende Material sind. Daher ist es auch von Bedeutung, den Schweregrad der Fehler zu bestimmen. Je schwerwiegender der Fehler ist, desto wahrscheinlicher ist es, dass er die Verwendbarkeit der Übersetzung beeinträchtigt. Der Grad der Schwere ist ein Indikator für die Bedeutung des Fehlers. Da eine meiner Forschungsfragen darin besteht, die Auswirkungen semantischer Verzerrungen auf die Verständlichkeit und Verwendbarkeit von Übersetzungen zu ermitteln, habe ich die in der Studie festgestellten Fehler auf ihren Schweregrad hin untersucht.

Fehlerklassifizierung nach der Fehlerschwere

In der MQM-Fehlertypologie werden Fehler je nach Schweregrad in vier Stufen (kritisch, signifikant, geringfügig und neutral) eingeteilt, und jede Schwerstufe wird bewertet. Die MQM weist den Fehlern verschiedene Schweregrade zu, die nicht allgemeingültig sind und je nach Zweck und Kontext der Übersetzung von 1 bis 100 reichen können. Die Schweregrade

können auch an individuelle Bedürfnisse angepasst werden, und es wird empfohlen, dass es eine exponentielle Differenz zwischen den Werten für neutrale, geringfügige, signifikante und kritische Fehler geben sollte, z. B. 0-1-5-25. Diese exponentielle Verhältnisskala spiegelt das zunehmende Risiko und die Auswirkungen von Fehlern wider. In meiner Forschung habe ich ein ähnliches Bewertungssystem verwendet, bei dem geringfügige Fehler mit 1 Schweregradpunkt, signifikante Fehler mit 5 Schweregradpunkten und kritische Fehler mit 25 Schweregradpunkten bewertet wurden.

In der MQM-Typologie sind kritische Fehler solche Übersetzungsfehler, die allein eine Übersetzung unbrauchbar machen. Dies wird durch das Zielpublikum und den definierten Zweck bestimmt. Bei der Bewertung der Qualität einer Übersetzung kann ein einziger kritischer Fehler automatisch zu einem „fail“ führen.

Signifikante Fehler hingegen sind solche, die die Verwendbarkeit oder Verständlichkeit des Inhalts beeinträchtigen, ihn aber nicht unbrauchbar machen. Erhebliche Übersetzungsfehler machen die Bedeutung des Textes für den Benutzer unklar.

Geringfügige Fehler beeinträchtigen weder die Verwendbarkeit noch Verständlichkeit des Inhalts. Wenn der Benutzer der Übersetzung den Fehler zuverlässig korrigieren kann, und dieser Fehler die Verwendbarkeit des Inhalts nicht beeinträchtigt, sollte er als geringfügig eingestuft werden. In vielen Fällen werden solche Fehler vom Benutzer absichtlich ignoriert. Geringfügige Fehler verursachen keine Probleme bei der Vermittlung von Bedeutung.

Irrelevante oder neutrale Fehler werden nicht wirklich als Fehler angesehen. Es handelt sich um Fehler, die im Allgemeinen ignoriert werden, bevor die Bewertung des Textes beginnt, da sie die Verständlichkeit des Zieltextes nicht beeinträchtigen. Neutrale Fehler können z. B. Wiederholungsfehler sein, wie z. B. die wiederholte falsche Schreibweise desselben Wortes, die den Gesamtsinn des Textes nicht beeinträchtigt. Neutrale Fehler werden bei der Bewertung außer Acht gelassen. In meiner eigenen Arbeit habe ich alle als neutral eingestuften Fehler bei der Berechnung und bei der Bewertung der Übersetzungsqualität ausgeschlossen, weil ihr Vorhandensein keine Auswirkungen auf das Textverständnis hat, und sie die semantische Bedeutung der Sätze nicht verfälschen. Das Vorhandensein geringfügiger und signifikanter Fehler hatte hingegen zur Folge, dass die Bedeutung der Sätze im Text teilweise oder vollständig verfälscht wurde.

Methoden zur Berechnung der Qualitätskennzahl

Nachdem die Fehler identifiziert, berechnet und klassifiziert wurden, habe ich im nächsten Schritt die Berechnung der Qualitätskennzahl und die Bestimmung der Qualität durchgeführt. Das MQM schlägt zwei Methoden zur Berechnung der Qualitätskennzahl vor: nicht kalibriert und kalibriert. In meiner Studie habe ich diese Berechnungsmethoden kombiniert. Die mit der nicht kalibrierten Methode ermittelte Qualitätsbewertung stellt die Rohbewertung der zu bewertenden Übersetzung dar und wird auch als Rohbewertungsmodell bezeichnet.

Bei der nicht kalibrierten Methode wird die (rohe) Qualitätsbewertung durch Berechnung des prozentualen Anteils des Textes mit Fehlern und Subtraktion dieses Wertes von 1 oder 100 % ermittelt, um einen Wert zu erhalten, der den fehlerfreien Teil des bewerteten Textes darstellt. Die Qualitätsbewertung drückt also den Anteil des geschätzten Zielinhalts aus, der fehlerfrei ist. Die Werte 1 und 100% stellen die hier angenommene maximale Qualitätskennzahl dar. Die maximale Qualitätskennzahl wird nicht auf einer bestimmten Grundlage berechnet, sondern ist ein willkürlicher Wert, der einfach in Übereinstimmung mit der besten Praxis in diesem Bereich gewählt wurde, und soll das Verständnis der erhaltenen Qualitätskennzahl erleichtern, indem die endgültige Qualitätskennzahl beispielsweise in einen Prozentwert umgerechnet wird. In meiner Untersuchung habe ich ebenfalls 100 % als Prozentsatz der maximalen Qualitätsbewertung gewählt.

Sowohl für die nicht kalibrierte als auch für die kalibrierte Methode wird vor der Auswertung eine Akzeptanzschwelle für die zu bewertenden Übersetzungen festgelegt. Die Akzeptanzschwelle ist ein Punktwert oder ein Prozentsatz, der die Grenze zwischen bestanden und nicht bestanden definiert. Bei der unkalibrierten Methode ist es üblich, eine Akzeptanzschwelle von z.B. 0,99 oder 99% festzulegen, je nach der verwendeten Skala. In diesem Fall geben 99 % den Anteil der geschätzten Übersetzung an, der fehlerfrei ist, und das Intervall der „Fehlerquote“ beträgt 1 %. Somit wird jede Qualitätsbewertung zwischen 100 und 99% als „akzeptiert“ eingestuft. In der Praxis bedeutet die Akzeptanzschwelle von 99%, dass ein Text von 1000 Wörtern maximal 10 Strafpunkte erhalten sollte, vorausgesetzt, der Gewichtungsfaktor für alle Fehlertypen ist 1. Das bedeutet, dass ein Text von 1000 Wörtern 10 Strafpunkte erhalten sollte, wenn er 10 kleine Fehler oder 2 große Fehler enthält. Wenn die Gesamtzahl der Strafpunkte in dem zu bewertenden Text 10 Strafpunkte/1000 Wörter

überschreitet, ist die Übersetzung nicht akzeptabel. Dies schien mir keine realistische Anforderung für meine Forschung zu sein, so dass in meiner Arbeit auch eine kalibrierte Methode verwendet wurde, um die Akzeptanzschwelle an die Bedürfnisse des Übersetzungsbenutzers anzupassen. Daher wurde eine Akzeptanzschwelle von 85 % für die zu bewertenden Übersetzungen festgelegt, was eine realistischere Anforderung zu sein schien. In der Praxis bedeutet diese 85% Grenze, dass ein Text von 1000 Wörtern insgesamt 150 Strafpunkte erhalten kann. Die maschinelle Übersetzung meiner Studie, die 2.501 Wörter enthielt, durfte maximal 375 Strafpunkte erhalten, was noch als akzeptabel angesehen werden konnte.

Neben der Akzeptanzschwelle habe ich ein weiteres Akzeptanzkriterium für Übersetzungen festgelegt: die Abwesenheit von kritischen Fehlern. Denn kritische Fehler machen die Übersetzung praktisch unbrauchbar und führen automatisch zu ihrer Ablehnung, auch wenn die Übersetzung ansonsten von hoher Qualität ist.

Die Berechnung der Qualitätsbewertung erfolgte in drei Schritten, die von der MQM für die nicht kalibrierte Methode vorgeschlagen wurden:

- 1) Im ersten Schritt wurde die Anzahl der Strafpunkte für die ausgewählten Fehlerarten – Genauigkeit, Geläufigkeit/Flüssigkeit und Terminologie – unter Berücksichtigung ihrer Schweregrade (1,5 und 25) und Gewichtungsfaktoren (1) berechnet. Schließlich wurden die Strafpunkte addiert, um die Gesamtzahl der Strafpunkte zu ermitteln.

$$\text{Gesamtstrafpunkte} = [(\text{geringfügiger Fehler} \times \text{Schwerepunkt}) + (\text{signifikanter Fehler} \times \text{Schwerepunkt}) + (\text{kritischer Fehler} \times \text{Schwerepunkt}) \times \text{Gewichtungsfaktor}]$$

- 2) Im zweiten Schritt wurde die Anzahl der Strafpunkte pro Wort berechnet, indem die Gesamtzahl der erhaltenen Strafpunkte durch die Anzahl der Wörter in der zu bewertenden Übersetzung geteilt wurde.

$$\text{Wortstrafpunktzahl} = \text{Gesamtstrafpunkte} / \text{Anzahl der Wörter in der zu bewertenden Übersetzung}$$

3) Im dritten Schritt wurde die Qualitätsbewertung in Prozent berechnet.

$\text{Qualitätspunktzahl} = 100\% - (\text{Strafpunkte pro Wort} \times 100\%)$
--

Angenommen, eine Übersetzung enthält 1000 Wörter und der Prüfer findet 10 geringfügige Fehler, 2 signifikante Fehler und 0 kritische Fehler. Dazu kommt der Gewichtungsfaktor 1, der allen Fehlertypen zugeschrieben wird. Darüber hinaus erhalten geringfügige Fehler einen Schweregrad von 1, signifikante Fehler 5 und kritische Fehler 25:

$\text{Gesamtstrafpunkte} = [(10 \text{ geringfügige Fehler} \times 1) + (2 \text{ signifikante Fehler} \times 5) + (0 \text{ kritische Fehler} \times 25)] \times 1 = 20 \text{ Strafpunkte}$
--

$\text{Wortstrafpunktzahl} = 20 \text{ Strafpunkte} / 1000 \text{ Wörter} = 0,02$

$\text{(Rohe) Qualitätspunktzahl} = 100 \% - (0,02 \times 100 \%) = 98 \%$
--

Anhand der Qualitätspunktzahl lässt sich schließlich feststellen, ob die Übersetzung akzeptabel ist.

In meiner Forschung wurden sowohl quantitative als auch qualitative Methoden verwendet. Der quantitative Teil der Forschungsmethode beschränkte sich auf die statistische Untersuchung der Ergebnisse der Fehleranalyse, während sich der qualitative Teil auf die Identifizierung, Klassifizierung und Schwere der Fehler anhand der MQM-Qualitätsmetrik konzentrierte. Die von mir angewandte fehlerorientierte Metrik bestand darin, alle von mir festgestellten Fehler zu markieren, ihre Anzahl zu zählen und sie den entsprechenden Fehlerkategorien zuzuordnen. Daraus wurde schließlich ein prozentualer Gesamtwert errechnet, der in meiner Arbeit als (rohe) Qualitätspunktzahl bezeichnet wird. Anhand dieser Punktzahl konnte dann festgestellt werden, ob die Übersetzung akzeptabel war oder nicht, indem die Punktzahl mit einer vordefinierten Akzeptanzschwelle verglichen wurde.

Die Ergebnisse der quantitativen Analyse

Die quantitative Analyse der Daten führte zu folgenden Ergebnissen: Es wurden insgesamt 61 Fehler in der Übersetzung gefunden, von denen 41 die Genauigkeit, 13 die Flüssigkeit und die restlichen 7 die Verwendung der Terminologie betrafen. Von diesen 61 Fehlern waren 0

kritisch, 22 signifikant und 39 geringfügig. Von allen für die Analyse ausgewählten Fehlerkategorien fiel die höchste Anzahl von Fehlern auf die Kategorie Genauigkeit, die nächsthöchste auf die Kategorie Geläufigkeit/Flüssigkeit und die niedrigste auf die Kategorie Terminologie.

Die folgenden Daten wurden zur Berechnung der Qualitätsbewertung herangezogen: Die Gesamtzahl der Wörter in der Übersetzung DeepL betrug 2501. Es gab 41 Fehler bei der Genauigkeit, von denen 18 signifikant und 23 geringfügig waren; es gab keine kritischen Fehler. Es gab 13 Fehler in der Geläufigkeit/Flüssigkeit, von denen 2 signifikant und 11 geringfügig waren; es gab keine kritischen Fehler. Es gab 7 Terminologiefehler, von denen 2 signifikant und 5 geringfügig waren. Für alle Fehlertypen wurde ein Gewichtungsfaktor 1 verwendet. Geringfügige Fehler erhielten 1 Schweregradpunkt, signifikante Fehler 5 Schweregradpunkte und kritische Fehler 25 Schweregradpunkte.

Die Ergebnisse der auf Grundlage dieser Informationen durchgeführten Berechnungen waren wie folgt:

Schritt 1:

Strafpunkte für Genauigkeit = 113 Strafpunkte.
Strafpunkte für Flüssigkeit = 21 Strafpunkte.
Strafpunkte für Terminologie = 15 Strafpunkte.
Gesamtstrafpunkte = $113 + 21 + 15 = 149$

Schritt 2:

Strafpunkte pro Wort = 149 (Gesamtzahl der Strafpunkte) / 2.501 (Gesamtzahl der Wörter im zu bewertenden Text) = $0,0596$

In Prozent ausgedrückt war es: $0,0596 \times 100 \% = 5,96 \%$

Schritt 3:

(Roh-)Qualitätspunktzahl = $100 \% - 5,96 \% = 94,04 \%$.
--

Gerundet betrug der Rohqualitätswert/die Rohqualitätspunktzahl 94 %. Dabei handelt es sich um die Punktzahl, die angibt, wie hoch der Anteil der bewerteten Übersetzung ist, der fehlerfrei ist. Vor der Analyse wurde eine Akzeptanzschwelle von 85 % für die Übersetzung festgelegt. Das Ergebnis 94 % könnte daher als akzeptables Ergebnis angesehen werden. Bei

der Bewertung des Ergebnisses musste jedoch beachtet werden, dass lediglich die Kategorien Genauigkeit, Geläufigkeit/Flüssigkeit und Terminologie zu bewerten waren.

Die Ergebnisse der qualitativen Analyse

Eine qualitative Analyse der Daten ergab, dass die im August 2023 erstellte DeepL-Übersetzung nicht völlig fehlerfrei war, die Fehler jedoch nicht kritisch waren und nicht zur Ablehnung der Übersetzung führten. Die Schwere der Fehler reichte von signifikant bis gering. Signifikante Fehler, wie falsche Wortübersetzungen, Begriffe und Wortreihenfolgen, erschwerten das Verständnis der Sätze, verfälschten die Bedeutung der Sätze oder verhinderten die Übertragung semantischer Bedeutung. Geringfügige Fehler hingegen hingegen beispielsweise mit der inkonsistenten Verwendung von Begriffen zusammen, die das Leseerlebnis schwächten, oder es wurden unwesentliche Dinge in den Sätzen hervorgehoben.

Die maschinelle Übersetzung wies die meisten Genauigkeitsfehler auf, die fast immer als falsche Übersetzungen auftraten. Fehlübersetzungen waren oft auf eine Fehlinterpretation der Bedeutung des ursprünglichen Wortes zurückzuführen. Dies führte dazu, dass der Zielsatz nicht genau mit dem Inhalt des ursprünglichen Ausgangssatzes übereinstimmte. Eine große Zahl von Genauigkeitsfehlern war von Bedeutung, weil sie beispielsweise das Verständnis der Bedeutung von Sätzen störten. Diese semantischen Verzerrungen erschwerten die Interpretation des Übersetzungssatzes, machten aber nicht den gesamten ausgewerteten Text unbrauchbar. Bei der menschlichen Übersetzung wurden keine Genauigkeitsfehler festgestellt.

Nach Genauigkeitsfehlern traten die meisten Sprachfehler bei der maschinellen Übersetzung auf, die meist auf eine falsche Wortreihenfolge zurückzuführen waren. Der DeepL-Übersetzer versuchte, der Wortstellung des Ausgangstextes zu folgen, was manchmal zu schwer verständlichen Ausdrücken in der finnischen Sprache führte. Solche Fehler führten zu Ungenauigkeiten und Mehrdeutigkeiten in der Bedeutung der Sätze, was das Verständnis der ursprünglichen Botschaft erschwerte, die für die Interpretation der Botschaft erforderliche Zeit erhöhte und das Lesen verlangsamte. Allerdings gab es in der maschinellen Übersetzung keine geringfügigen oder signifikante Rechtschreib- oder Zeichensetzungsfehler und der Text war nicht unklar. Die von einem menschlichen Übersetzer angefertigte Übersetzung folgte der Standardwortreihenfolge der finnischen Sprache.

Die maschinelle Übersetzung wies in der Kategorie Terminologie die wenigsten Fehler auf. Diese Fehler hingen fast immer mit der Wahl falscher Begriffe zusammen. Die übersetzten Begriffe entsprachen oft nicht dem Original, weshalb der Sinn des Satzes nicht genau wiedergegeben wurde oder völlig unklar blieb. Die meisten Begriffe wurden von DeepL korrekt übersetzt. Bei mehrdeutigen Begriffen stieß der maschinelle Übersetzer jedoch auf Probleme. In diesen Fällen gelang es dem maschinellen Übersetzer nicht immer, einen für den wirtschaftlichen Sprachkontext geeigneten Begriff auszuwählen. Diese Beobachtung zeigte, dass die Fähigkeit des maschinellen Übersetzers, die richtigen Begriffe zu verstehen und zu verwenden, eingeschränkt sein kann, insbesondere wenn es um doppelte oder polyseme Wörter geht. Falsch übersetzte Begriffe behinderten das Verständnis und die Verwendung des Satzes, wenn der Begriff den Kernpunkt des Satzes bildete.

Zusammenfassung und Schlussfolgerungen

Nach dem Vergleich der fehlerhaften Passagen in der DeepL-Übersetzung mit der menschlichen Übersetzung fielen mir deutliche Unterschiede auf. Der menschliche Übersetzer erstellte einen fließenderen und natürlicheren Text, der den sprachlichen Normen der Zielsprache besser entspricht. Der maschinelle Übersetzer hingegen folgte manchmal zu genau der Wortreihenfolge des Ausgangstextes, was zu unnatürlichen Sätzen führte. In Bezug auf die Genauigkeit vermittelte die menschliche Übersetzung die Bedeutung der Sätze korrekt und berücksichtigte auch die Nuancen, während DeepL insbesondere bei Wörtern mit Doppel- und Mehrfachbedeutungen Fehler machte. Der menschliche Übersetzer wusste, wie er die richtigen Begriffe basierend auf dem Kontext der Wirtschaftssprache auswählt, DeepL hingegen übersetzte einige Begriffe ungenau. Insgesamt war die menschliche Übersetzung genauer und besser an die Zielsprache angepasst, während die DeepL-Übersetzung weniger genau war. Der DeepL-Übersetzer ist bei der Übersetzung von Sachbuchtexten noch nicht auf dem gleichen Niveau wie ein menschlicher Übersetzer. Man kann sagen, dass die DeepL-Übersetzung noch einer Nachbearbeitung bedarf, damit sie in Qualität und Verständlichkeit der Qualität einer menschlichen Übersetzung entspricht.

Die MQM-Qualitätsmetrik erwies sich als nützliches Werkzeug zur Klassifizierung von Fehlern und zur Bestimmung ihrer Art. Die Anwendung der Fehlerkategorien, also die Klassifizierung der in der Übersetzung auftretenden Fehler, war meiner Meinung nach einfach und klar, die Definition der Schweregrade der Fehler bereitete jedoch Schwierigkeiten.

Obwohl die MQM-Qualitätsmetrik klare Definitionen der Schweregrade liefert, war die Klassifizierung kritischer und signifikanter Fehler nicht so einfach, wie ich erwartet hatte – insbesondere bei falschen Übersetzungen und Wortreihenfolgen. Ich war zeitweise unsicher, ob eine Fehlübersetzung als kritischer oder erheblicher Fehler einzustufen ist. Bei der Beurteilung der Schwere falscher Übersetzungen war deren Auswirkung auf den gesamten Text und den Leser zu berücksichtigen. Darüber nachzudenken – welche Auswirkungen falsche Informationen auf den Leser haben können und welche Folgen sie haben können – war zeitaufwändig und mühsam.

Obwohl die MQM-Metrik einen objektiven Bewertungsrahmen bietet, kann die Subjektivität des Bewerter die Generalisierbarkeit und Zuverlässigkeit der Ergebnisse schwächen. Die Vorkenntnisse des Forschers über das zu bewertende Material und seine Erfahrung beeinflussen die Klassifizierung von Fehlern – derselbe Fehler kann je nach seiner Sichtweise entweder geringfügig oder signifikant sein. Wenn der Forscher mit der Thematik des zu bewertenden Textes vertraut ist, ist es einfacher, den Text zu verstehen und Fehler zu korrigieren, und daher kann er Fehler als geringfügig einstufen. Dementsprechend kann ein unerfahrenerer Forscher, der auch über weniger Fachkenntnisse verfügt, Fehler als die schwerwiegendsten einstufen, da das Verstehen des Textes eine größere Herausforderung darstellt. Dasselbe gilt auch für den Leser: Wenn der Leser der Übersetzung sich gut mit der Thematik auskennt, stören Fehler in der Übersetzung nicht unbedingt die Verständlichkeit des Textes, denn der Leser kann die Fehler leicht in seinem Kopf korrigieren. Wenn der Leser hingegen zum ersten Mal mit einem Buch über Wirtschaftswissenschaften arbeitet, kann die Übersetzung schwer verständlich erscheinen.

Bei meiner eigenen Beurteilung habe ich einige Sprachfehler als geringfügig eingestuft, obwohl ein anderer Prüfer sie möglicherweise als signifikant eingestuft hätte. Dies lag daran, dass ich den Ausgangstext bereits vor der Bewertung gelesen hatte und auch über allgemeine Kenntnisse in Wirtschaftsthemen verfügte. Dies könnte meine Objektivität beeinträchtigt haben, weshalb ich viele Sprachfehler als geringfügig einstufte, weil ich sie leicht zu beheben fand. Die Objektivität der Bewertung könnte verbessert werden, wenn der Prüfer eine Person wäre, die kein Deutsch spricht, sich aber mit Wirtschaftsthemen auskennt. Meine Forschungsergebnisse waren indikativ und es wäre gut, das Thema weiter zu untersuchen.

Liite 2. DeepL-konekäännös

punainen väri = tarkkuusvirhe

sininen väri = sujuvuusvirhe

vihreä väri = terminologiavirhe

Luku 6 Kapitalismi ei ole markkinatalous

Monet ihmiset **epäilevät** vaistomaisesti, että kapitalismi ei ole sama asia kuin markkinatalous. Kyselytutkimuksissa 48 prosenttia saksalaisista sanoo, että kapitalismi ei enää sovi meidän aikaamme, mutta vain 24 prosenttia sanoo näin markkinataloudesta. Tähän sopii, että markkinatalouteen suhtaudutaan myönteisemmin. Se **liitetään** vahvemmin edistykseen, vapauteen, kasvuun ja vastuuntuntoon, kun taas kapitalismi **liitetään** enemmän eriarvoisuuteen, riistoon ja ahneuteen.

Markkinatalous, se kuulostaa **hellyttävältä**. Se tuo mieleen viihtyisät **maanviljelijöiden markkinat**, joilla omenat punnitaan edelleen yksi kerrallaan ja samalla vaihdetaan kuulumisia naapuruston **uusimmista uutisista**. Tätä markkinataloutta ei kuitenkaan ole olemassa. Elämme kapitalismissa, joka on jotain aivan muuta - ja se on jo johtanut moniin **vääriin poliittisiin päätöksiin, että markkinatalous ja kapitalismi sekoitetaan niin sitkeästi**.

Jo vilkaisu historiaan osoittaa, että nykyaikaisen kapitalismin on tietenkin oltava enemmän kuin pelkkä kokoelma markkinoita. Markkinat olivat nimittäin olemassa jo antiikissa 2 500 vuotta sitten. Kuka tahansa, joka vierailee nykyään Ateenassa, voi käydä Akropolis-kirkon alapuolella sijaitsevan Agoran kaivetuissa jäännöksissä - entisellä keskeisellä kauppatorilla ja sen monilla kauppakaduilla. Markkinoita ei kuitenkaan ollut vain antiikin Kreikassa, vaan lähes kaikissa kulttuureissa. Arabeilla oli **soukinsa**, turkkilaisilla ja persialaisilla basaarinsa. Myös intiaanit ja kiinalaiset kävivät kauppaa tavaroillaan markkinoilla, mutta modernia kapitalismia ei syntynyt missään.

Jopa uusliberalistiset markkinatalouden teoreetikot tietysti ymmärtävät, että markkinat ovat kaukaa historiasta. Mutta he **tunnustavat eron** markkinoiden ja markkinatalouden välillä. Yleisen määritelmän mukaan markkinataloudessa kaikella on ensimmäistä kertaa hinta: maata, työvoimaa ja tuotteita voidaan ostaa ja myydä vapaasti.

Hinnan kaikkivoipaisuus on itse asiassa historiallinen innovaatio, jota on havaittu vasta noin 300 vuoden ajan. Keskiaikaisessa feodalismissa aateliset eivät voineet myydä maataan omasta aloitteestaan, sillä kuningas oli myöntänyt sen heille "läänityksenä". Vastaavasti maaorjat tai **orjuudessa olevat talonpojat** eivät voineet **vain** muuttaa kaupunkiin etsimään paremmin palkattua

työtä. Tämä johtui siitä, että he kuuluivat maanomistajilleen. Keskiajalla oli lukemattomia **toreja**, kuten "**Kornmarkt**", "**Gänsemarkt**", "**Tuchmarkt**", "**Fischmarkt**", "**Heumarkt**" tai "**Holzmarkt**" -nimiset katukyltit osoittavat. Nämä **markkinat** muodostivat kuitenkin melko pienen osan taloudesta, kun taas muulle tuotannolle oli ominaista feodalismi.

Nykyään taas lähes kaikella on **hinta**, ja nämä hinnoitellut alueet laajenevat jatkuvasti. Vain yksi esimerkki: hoitotyö oli ennen toimintaa, jota naiset joutuivat tekemään palkattomana perhetyönä. Viime aikoina siitä on kuitenkin tullut oma ammatti, jolla on laaja hintaluettelo. Esimerkiksi "**pieni aamu- ja iltavessa**" maksaa 12,91 euroa, kun taas "apu välipalan syömisessä" maksaa vain 4,78 euroa.

Pelkkä hinta ei kuitenkaan tee vapaita markkinoita - **kilpailu on avainasemassa**. Sanotaan, että kysynnän ja tarjonnan häiriöttömässä kohtaamisessa pitäisi syntyä oikeudenmukainen hinta. Jokainen todellinen vapaiden markkinoiden taloustieteilijä on siksi kilpailun innokas kannattaja, jota hän usein ylistää todellisena sosiaalipolitiikkana. Entinen talousministeri ja CDU:n liittokansleri **Ludwig Erhard on tyypillinen esimerkki tästä**. Kirjassaan Wohlstand für alle (Hyvinvointia kaikille) hän ei jätä sivua **lukematta** ylistämättä kilpailun sosiaalisia siunauksia. Linaan muutamia hänen **aforismejaan**: "Kilpailu on lupaavin keino saavuttaa ja turvata vauraus." Tämä on myös hänen aforisminsa. "Kilpailun avulla - sanan parhaassa merkityksessä - edistys ja voitto yhteiskunnallistuvat." "Kilpailun avulla edistys ja voitto yhteiskunnallistuvat." "Vauraus kaikille" ja "vauraus kilpailun kautta" liittyvät erottamattomasti toisiinsa. Kursivointi on muuten Erhardin itsensä tekemä, joka ei ilmeisesti luottanut lukemattomiin toistoihinsa ja turvautui siksi typografiaan varmistaakseen, että kilpailun hyvät uutiset tavoittavat lukijat.

Vuonna 1958 Erhard perusti jopa liittovaltion kartelliviraston nimenomaan kilpailun suojelemiseksi. Jos **kuitenkin** lukee sen lehdistötiedotteita, saa sen vaikutelman, että se hyökkää lähinnä melko merkityksettömiä yrityksiä vastaan. Avautuu maailma, jossa näyttää olevan kyse vain jäätelöstä, makeisista ja sementistä. **Suuryritykset eivät esiinny juuri lainkaan**. Tämä ei tarkoita sitä, että **naurettaisiin liittovaltion kartelliviraston työlle, joka työllistää noin 345 henkilöä**. On kuitenkin kysyttävä, miten **se** sopii yhteen sen kilpailuajattelun kanssa, että Porschen ja Piechin perheet hallitsevat nyt **aakkosjärjestyksessä** Audia, Bentleyä, Bugattia, Ducatia, Lamborghiniä, MANia, Porschea, Scaniaa, Seatia, Skodaa, VW:tä ja VW-hyötyajoneuvoja.

Vaikka maailmassa on vielä joitakin itsenäisiä autonvalmistajia, tämä takaa maailmanlaajuisen kilpailun vain rajoitetusti. Yritykset ovat jo pitkään tehneet monenlaista yhteistyötä. VW tekee yhteistyötä Daimlerin ja Suzukin kanssa, Daimler myös Renault-Nissanin ja Fordin kanssa, BMW Toyotan ja Peugeotin kanssa ja Opel myös Peugeotin kanssa. Autoteollisuudessa toistuu se, mikä nähtiin terästeollisuudessa jo 1800-luvulla: korkeisiin investointikustannuksiin on varaa vain suurilla

yrityksillä, jotka sitten omaksuvat määräävän markkina-aseman. Tällä ei ole enää mitään tekemistä **perinteisen kilpailun käsitteen** kanssa.

Jopa Ludwig Erhard ei voinut sivuuttaa tätä keskittymistaipumusta. Hän valitti hieman kiusallisesti: "Nykyaikaisen tekniikan kehittyminen on omalta osaltaan jälleen kerran edistänyt tiettyjä monopolisoitumistaipumuksia, niin että tasapuoliset **toimintaedellytykset** ovat epäilemättä häiriintyneet kaikkialla". Mutta tästä oivalluksesta ei tullut mitään. Erhard ei ollut valmis kysymään itseltään, kuinka paljon hänen "markkinataloudestaan" jäisi jäljelle, jos hallitsevilla teollisuudenaloilla kaikilla olisi hänen pahoittelemiaan "monopolisoitumistendenssejä".

Näillä "monopolisoitumistendensseillä" ei ole ainoastaan laajoja vaikutuksia, kun yritykset sulautuvat tai tekevät yhteistyötä keskenään, vaan ennen kaikkea syviä vaikutuksia. Tekninen termi on "vertikaalinen integraatio". Suuryritykset pyrkivät hallitsemaan kaikkia tuotantovaiheita raaka-aineista myyntiin. Ne eivät aina tuota kaikkea itse, vaan usein ne "ulkoistavat", kuten nykysaksaksi sanotaan. Mutta näitä näennäisesti riippumattomia toimittajia valvotaan armottomasti ja alistetaan yritysten etujen alaisuuteen. Vertikaalinen integraatio alkoi **myös** jo 1800-luvulla, kun suuret yritykset halusivat irrottaa tuotantovaiheensa mahdollisimman paljon markkinoista.

Taloustieteen Nobel-palkinnon saanut Herbert Simon pohti kerran, miten paha aavistamaton marsilainen kuvailisi talouttamme: "Olettaisiko hän, että Maan asukkaat elävät markkinataloudessa?". Todennäköisesti ei. Sen sijaan hän luultavasti päättelisi, että Maan asukkaat toimivat järjestäytyneessä taloudessa, jossa suurin osa taloudellisesta toiminnasta koordinoidaan yritysten sisällä - **pikemminkin** kuin näiden yritysten välisten markkinasuhteiden kautta."

Suuryritykset ovat onnistuneet sementtoimaan markkinat yli 100 vuoden ajan. Lähes kaikki DAX-yritykset ovat peräisin ajalta ennen ensimmäistä maailmansotaa. Ainoa todellinen poikkeus on ohjelmistokonserni SAP, joka perustettiin vasta vuonna 1972. Tämä johtavien osakeyhtiöiden huomattava jatkuvuus osoittaa, **ettei kukaan voi kilpailla niiden pelkän koon kanssa**. Olipa kyse teräksestä, autoista, kemikaaleista tai lääkkeistä: nämä markkinat ovat suurelta osin suljetut, eivätkä uudet tulokkaat voi enää murtautua niille.

Luku 9 Raha on arvoitus - eikä se ole sama kuin pääoma

Mitä raha on? Miten se toimii? Ja miten sitä voi **säästää**? Nämä kysymykset kuulostavat hyvin yksinkertaisilta, mutta niihin on vaikea vastata. Kuuluisa taloustieteilijä Joseph Schumpeter työsti vuosia kirjaa rahasta, mutta sitä ei koskaan julkaistu. Aihe oli liian monimutkainen.

Silti se vaikuttaa ensi silmäyksellä niin yksinkertaiselta: kaikki käyttävät rahaa. Et voi ostaa edes sämpylää laittamatta tiskiinkin vähintään kymmentä senttiä. Lisäksi raha on ilmeisesti yhtä vanha kuin ihmiskunta. Olipa kyse simpukankuorista, lehmistä tai kullasta - heti kun kansat asettuivat aloilleen, ne loivat mittayksikön, jolla ne pystyivät **määrittämään** sosiaaliset ja taloudelliset velat.

Mutta niin vanha ja arkipäiväinen kuin raha onkin, se on edelleen mysteeri. Tämä käy ilmi siitä, että ei ole olemassa selkeää määritelmää siitä, mitä raha tarkalleen ottaen on. Sen sijaan luodaan "rahamääriä", jotka kukin keskuspankki määrittelee itse. Maailmanlaajuisesti ei ole selvää sopimusta siitä, mitä pitäisi laskea rahaksi. Rahan tapauksessa kansainvälinen standardointi, joka on itsestäänselvyys jokaisen kopiokoneen ja jokaisen auton kohdalla, epäonnistuu.

Raha on niin outo ja niin muuttuva, koska se on sosiaalinen konstruktio. Raha on sitä, mikä hyväksytään rahaksi. Se voi olla mitä tahansa. Näin Virginian tupakkavaltiossa amerikkalaiset keksivät, että myös tupakalla voi maksaa. Se toimi kohtuullisen hyvin noin 200 vuotta.

Ei tiedetä, kuinka kauan ihmiset ovat käyttäneet rahaa. Ainoa varma tieto on, että ensimmäiset kirjalliset todisteet ovat peräisin Mesopotamiasta, ja ne ovat yli 4 000 vuotta vanhoja. Tuolloin laskentayksikkönä käytettiin hopeasekeliä, joka ei ollut vielä liikkeessä kolikkona, vaan se piti punnita, ja yksi sekeli vastasi noin 8,5 grammaa hopeaa.

Vaikka babylonialaiset ja assyrialaiset eivät vielä tunteneet kolikoita, he tiesivät jo tarkalleen, miten **korkokorko** lasketaan. On säilynyt luottosopimuksia, jotka palvelivat eri tarkoituksia: Joidenkin tarkoituksena oli rahoittaa **pitkän matkan kauppaa** esimerkiksi Anatolian kanssa. Toiset taas hyödynsivät pienviljelijöiden ahdinkoa, sillä heidän oli lainattava viljaa kuivuuden jälkeen elättääkseen perheensä seuraavaan satoon asti. Nämä lainasopimukset ja velkakirjat ovat vanhimpia ihmiskunnan tuntemia asiakirjoja. Kirjoitusta ei keksitty kirjallisuuden laatimista varten, vaan kaupallisten sopimusten kirjaamista varten.

Assyrialaiset keksivät idean, joka vaikuttaa uskomattoman nykyaikaiselta: he käyttivät velkakirjoja maksuvälineenä. **Kauppiaat ja kauppiat** eivät odottaneet lainan erääntymistä, vaan toimittivat asiakirjan eteenpäin omien **velvoitteidensa** hoitamiseksi. Ihmiset tunsivat alusta alkaen luottorahan, jota käytämme myös nykyään - hieman eri muodossa.

Kolikko ei siis ollut rahan alkuperä, vaan suhteellisen myöhäinen keksintö, joka voidaan todennäköisesti ajoittaa seitsemännelle vuosisadalle eaa. Kreikkalainen historioitsija Herodotos kirjoitti tunnetusti: "Ensimmäiset ihmiset, joista tiedämme, olivat lydialaiset, jotka **leimasivat** kolikoita kullasta ja hopeasta." Lydialaiset olivat indoeurooppalainen kansa, joka asui Vähä-Aasiassa, ja heidän

kuuluisin kuninkaansa oli nimeltään Kroisos. Hän teki vaikutuksen kreikkalaisiin naapureihinsa antamalla niin runsaita lahjoja Delfoin, Miletoksen ja Efesoksen oraakeleille, että ilmaisu "rikas kuin **Krösus**" on säilynyt tähän päivään asti.

Jo muinoin ihmisiä askarrutti kysymys, joka askarruttaa sijoittajia vielä nykyäänkin: Onko kulta- ja hopeakolikoilla itsessään arvoa, koska ne on tehty jalometallista? Vai johtuuko kolikoiden arvo ensisijaisesti niiden yhteiskunnallisesta tehtävästä toimia maksuvälineenä ja helpottaa erilaisten tavaroiden vaihtoa?

Kreikkalaiselle filosofille Aristoteleelle vastaus oli selvä: rahan arvo perustuu sosiaaliseen konventioon, jonka avulla eri tavarat ovat vertailukelpoisia. Hän toteaa kirjaimellisesti: "Siksi kaiken, mitä vaihdetaan, on oltava jotenkin vertailukelpoista. Tämän vuoksi raha tuli kuvaan: Tietyllä tavalla siitä tulee välittäjä, koska kaikkea voidaan mitata sitä vasten, myös liikaa ja liian vähän, esimerkiksi sitä, kuinka monta kenkää vastaa taloa tai ruokaa. ... Mutta raha luotiin eräänlaiseksi tarpeiden vaihdettavaksi edustajaksi, joka perustuu keskinäiseen sopimukseen. Ja se kantaa nimeä >raha< (nomisma), koska se ei ole olemassaolostaan velkaa luonnolle, vaan koska se on asetettu >arvoiseksi< (nomos), ja on meistä itsestämme kiinni, haluammeko muuttaa sitä tai asettaa sen pois käytöstä."

Rahalla ei siis ole itseisarvoa, koska se on tehty kullasta ja hopeasta, vaan se on luotu yhteiskunnallisella sopimuksella. Raha on olemassa, koska sitä tarvitaan. Tämä oivallus ei ehkä kuulosta erityisen sensaatiomaiselta, mutta Aristoteles oli jo 2 500 vuotta sitten paljon pidemmällä kuin monet kulttuurintutkijat, jotka puhuvat rahasta nykyään.

Arkeologiset kolikkolöydöt Vähä-Aasiasta osoittavat, miten oikein Aristoteles kuvasi rahan tehtävän. Historioitsijat olivat yllättyneitä havaitessaan, että kolikoiden kultapitoisuus vaihteli suuresti, vaikka kolikot oli lyöty samalla tavalla ja olivat siksi ilmeisesti samanarvoisia. Kulta kultana ei siis näyttänyt olevan ensisijaisen kiinnostava, vaan kolikoiden arvo johtui siitä, että ne olivat laillinen maksuväline, jonka kiertoa valtio valvoi. Tärkeää ei ollut kolikoiden materiaali, vaan se, että hallitsijat **varmistivat** rahan ostovoiman säilymisen **varmistamalla**, että raha ei moninkertaistunut mielivaltaisesti.

Mutta jos kolikoilla ei ollut itseisarvoa, miksi ne ylipäätään valmistettiin jalometallista? Tämä itsestään selvä kysymys oli jo keskiaikaisen filosofin Nikolai Oresmen mielessä, ja hänen vastauksensa oli erittäin pragmaattinen: Kultaa ja hopeaa oli yksinkertaisesti niukasti ja niitä oli helppo kuljettaa.

Jäljelle jäi vain yksi ongelma: kultaa ja hopeaa ei ollut Euroopassa vain niukasti saatavilla, vaan usein niitä ei ollut lainkaan. Antiikin ajoista lähtien himoitut kolikot olivat kadonneet itään eivätkä koskaan palanneet takaisin. Kuten jo mainittiin, intiaanit ja kiinalaiset eivät olleet kiinnostuneita

länsieurooppalaisista tuotteista, joita he pitivät **raakana**, ja he olivat valmiita vaihtamaan silkkiä tai posliinia vain, jos saivat vastineeksi hopeaa. Samanlainen ongelma ilmeni Venäjän tai Levantin kanssa käytävässä kaupassa: Sielläkin ihmiset luovuttivat turkiksia tai mausteita vain, jos vastineeksi saatiin hopeaa.

Jos rahaa ei ole, se on keksittävä. Tämä periaate koski todennäköisesti kaikkia kulttuureja, ja myös länsieurooppalaiset noudattivat tätä mallia. Keskiajalla he käyttivät samaa ajatusta, joka oli ollut jo mesopotamialaisilla. He muuttivat luoton rahaksi - mutta tällä kertaa he maksoivat vekseleillä.

Luku 11 Velka ja korot? Kyllä, kiitos

Sana "velka" on saksan kielessä **moniselitteinen**, koska sitä voidaan käyttää moraalisesti tai taloudellisesti. Velka voi vaatia sovitusta tai tarkoittaa yksinkertaisesti vain rahan takaisinmaksua. Velka kuulostaa myös **syllisyyden tornilta, orjuuden uhalta** ja vapauden puutteelta. Mikään termi ei ole yhtä **värikäs** kuin "velka", ja siksi monet ihmiset eivät pidä siitä, että koko talousjärjestelmämme näyttää perustuvan luottoon ja että me kaikki maksamme luottorahalla.

Velan monitulkintaisuus on esitetty jo Vanhassa testamentissa, jossa velkojien moraalisenä velvollisuutena on pitää joka seitsemäs vuosi "anteeksiantovuosi" ja luopua saatavistaan: "Mutta anteeksiantovuoden kanssa on toimittava näin: jos joku on lainannut jotakin lähimmäiseltään, niin antakoon se hänelle anteeksi eikä peri sitä lähimmäiseltään eikä veljeltään, sillä Herralle on julistettu anteeksiantovuosi." Tämä on siis "anteeksiantovuosi". (5. Moos. 15:1-2)

Kuitenkin jo muinaiset heprealaiset kohtasivat ongelman, jota voitaisiin nykyaikaisella talouskielellä kutsua "moraaliseksi vaaraksi": Rahanlainaajat pystyivät laskemaan tarkalleen, milloin tulisi **toinen** anteeksiantovuosi - ja mitä lähemmäs tämä seitsemäs vuosi tuli, sitä vähemmän he olivat halukkaita myöntämään lainoja. Siksi Raamattu varoittaa: "Varokaa, ettette saa sydämeenne syntymään petollista ajatusta, joka sanoo: 'Seitsemäs vuosi, vapautusvuosi, on lähellä', ja ettette katso köyhää veljeenne epäystävällisesti ja anna hänelle mitään." (5. Moos. 15:9)

Jo tämä osoittaa, että Raamattu ei paheksu lainoja, vaan nimenomaan vaatii rikkaita **lainaamaan rahaa tai viljaa köyhille**, kun häitä uhkaa. Velkaantuminen ei ole tuomittavaa, vaan Raamattu pitää moraalisenä ongelmana sitä, että velkojat haluavat **nähdä rahansa takaisin**.

Tämä raamatullinen **tuomio** saattaa tuntua oudolta, mutta se **vastaa käytännön kokemukseen**: muinoin oli tavallista orjuuttaa velalliset, jos he eivät kyenneet maksamaan lainojaan. Heidän oli myytävä itsensä tai sukulaisensa velkaorjuuteen. Erityisesti huonojen satojen jälkeen lähes kaikki köyhät

menettivät vapautensa: selviytyäkseen heidän oli lainattava rikkailta viljaa, jota he eivät voineet maksaa takaisin seuraavina vuosina, varsinkaan korkojen kanssa, koska heillä ei ollut riittävästi maata tarvittavien lisätulojen tuottamiseen. Vapautusvuosi otettiin käyttöön, jotta kansa ei koostuisi jossain vaiheessa vain orjista. Raamattu lohduttaa rikkaita sillä, että tämä ei ollut heidän kannaltaan huono **kauppa**: "Jos veljesi, heprealainen tai heprealainen nainen, on myynyt itsensä sinulle, hän palvelkoon sinua kuusi vuotta; seitsemäntenä vuonna päästä hänet vapaaksi... Äläkä anna hänen vapauttamisensa olla sinulle raskasta, sillä hän on palvellut sinua kuusi vuotta kuin kaksi päivätyöläistä." (5. Moos. 15:12-18)

Velkaorjuus ei ollut vain heprealaisten ongelma; köyhien velkojen anteeksianto oli yleistä Mesopotamiassa jo 1700-luvulla eaa. ja Ateenassa Solon määräsi yleisen velkojen anteeksiannon vuonna 594 eaa.

Lainojen takaisinmaksu ei kuitenkaan ollut usein pelkästään vaikeaa. Vähintään yhtä ongelmallisia olivat ylimääräiset korkomaksut. Tämän vuoksi Vanha testamentti muotoilee tiukan korkokiellon. Mooseksen kirjan 5. luvussa sanotaan: "**Älä ota korkoa veljeltäsi, älä rahasta, älä ruoasta äläkä mistään sellaisesta, mistä voi ottaa korkoa.** Voit ottaa korkoa muukalaiselta, mutta et veljeltäsi." (23, 21-22.) Tästä erottelusta **alkuasukkaiden** ja muukalaisten välillä luovuttiin kristinuskossa. Kristityt eivät yleensä saaneet **ottaa korkoa**, koska kaikkia ihmisiä oli pidettävä veljinä.

Kun Raamattu käsitteli ensisijaisesti köyhien **ahdinkoa**, Aristoteles muotoili suunnilleen samaan aikaan filosofisen ongelman, joka herättää kiinnostusta: Miten raha ylipäättään voi **luoda** rahaa? Eläinten kohdalla **lisääntymisen tie** on selvä, koska ne lisääntyvät. Siellä, missä on aluksi vain yksi uuhi, on ensi keväänä ainakin yksi karitsa lisää. Mutta rahan kanssa? Korkojen avulla tapahtuva lisääntyminen on luonnotonta, selittää Aristoteles Poliitikassaan: "Vielä inhottavampaa on koronkiskonta, **joka saa voittonsa** itse rahasta eikä siitä, mitä varten raha on. Sillä raha keksittiin vaihtoa varten, mutta koron kautta se lisääntyy itse itsensä kautta. Siitä sen nimi: Se, mikä syntyy, on sama kuin se, mikä synnyttää, ja koron (tokon) kautta raha syntyy rahasta. Tämä rahan hankkimistapa on siis kaikkein vastoin luontoa."

Tuomas Akvinolainen tarkensi tätä väitettä vielä 1200-luvulla selittämällä, että "**aikaa myydään**" **korkokäytännöllä**. Yksilöt eivät kuitenkaan voi omistaa aikaa, vaan se kuuluu yksin Jumalalle. Siksi korkoa pidettiin koronkiskontana (usura) ja syntinä.

Lainojen korko ei kuitenkaan ollut ainoa tapaus, jossa raha näytti muuttuvan lisää rahaksi. Yhtä oudolta tuntui se, että kauppiat saivat voittoa pelkästään kaupankäynnillä. Loppujen lopuksi

vaihtokaupassa vaihdettujen tavaroiden pitäisi olla samanarvoisia. Mutta jos vaihdettiin samanarvoisia tavaroita, mistä voitto tuli?

Tähän kysymykseen saattoi olla vain yksi vastaus: Kauppiaiden epäiltiin huijaavan kansalaisiaan. Siksi heihin suhtauduttiin äärimmäisen epäluuloisesti. Tuomas Akvinolainen katsoi, että kauppa "oikeutetusti paheksutaan, koska se tyydyttää voitonhalun, joka ulottuu yli kaikkien rajojen äärettömyyteen".

Kuten niin usein käy, teoria ei kuitenkaan toiminut käytännössä. Jo keskiajalla kävi ilmeiseksi, että korkojen ja voittojen yleinen kieltäminen ei ollut kovin tehokasta. Tämä johtui siitä, että kauppiaille oli käytettävissään erilaisia temppuja, joiden avulla he saattoivat laatia vekseleitä siten, että voitto salattiin. Oli myös taloudellisia syitä, joiden vuoksi voitto oli sallittava, eikä kirkko voinut jättää niitä huomiotta. Näitä olivat muun muassa tappioriski ja kauppiaan työ, joka oli korvattava. Lisäksi jopa kirkko tunnusti, että tavaroiden ostaminen ulkomailta oli yleisen edun mukaista. Eräässä 1200-luvulla laaditussa rippikirjassa todetaan: "Monissa maissa olisi suuri puute, jos kauppiaat eivät toisi yltäkylläisyyttä yhdeltä alueelta toiselle, jossa on niukkuutta. He saavat oikeutetusti työnsä hinnan.

Liite 3. DeepL-käännöksen virheiden dokumentointi

Taulukko 9. DeepL-käännöksen virheiden dokumentointi

Virheluokka ja -tyyppi	Vähäinen	Merkittävä	Kriittinen	Rangaistuspisteet ¹
Tarkkuus	23	18	0	113
Virheellinen käännös	20	17	-	105
Poisto	1	-	-	1
Lisäys	1	-	-	1
Kääntämätön	1	1	-	6
Sujuvuus	11	2	0	21
Oikeinkirjoitus	-	-	-	0
Välimerkit	-	-	-	0
Kielioppi	11	2	-	21
Epäselvä teksti	-	-	-	0
Terminologia	5	2	0	15
Virheellinen termi	4	2	-	14
Termin epäjohdonmukainen käyttö	1	-	-	1

¹Rangaistuspisteiden laskukaava:

= [(vähäinen virhe x vakavuuskerroin/vakavuuspiste) + (merkittävä virhe x vakavuuskerroin/vakavuuspiste) + (kriittinen virhe x vakavuuskerroin/vakavuuspiste)] x painotuskerroin