

Suomi–ruotsi-neuroverkkokääntimen  
räätälöinti Suomen Pankin  
vuosikertomuksen kääntämisen tueksi

Kristian Appelberg  
Maisterintutkielma  
Kääntämisen ja tulkkauksen maisteriohjelma  
Käännös- ja tulkkausviestintä, kielipari ruotsi–suomi  
Humanistinen tiedekunta  
Helsingin yliopisto  
Toukokuu 2022

## Tiivistelmä

**Tiedekunta:** Humanistinen tiedekunta

**Koulutusohjelma:** Kääntämisen ja tulkkauksen maisteriohjelma

**Opintosuunta:** Käännös- ja tulkkausviestintä

**Tekijä:** Kristian Appelberg

**Työn nimi:** Suomi–ruotsi-neuroverkkokääntimen räätälöinti Suomen Pankin vuosikertomuksen kääntämisen tueksi

**Työn laji:** Maisterintutkielma

**Kuukausi ja vuosi:** Toukokuu 2022

**Sivumäärä:** 51 s., ruotsinkielinen lyhennelmä 8 s.

**Avainsanat:** neuroverkkokääntäminen, OPUS-CAT, Suomen Pankki, käännösteknologia, konekääntäminen

**Ohjaaja:** Ritva Hartama-Heinonen

**Säilytyspaikka:** Helsingin yliopiston kirjasto

**Tiivistelmä:** Konekääntäminen on viime vuosina saanut yhä enemmän jalansijaa niin yhteiskunnassa yleisesti kuin ammattikäntäjienkin työssä. Tämä johtuu suurelta osin merkittävistä parannuksista konekäännösten laadussa, jonka on sanottu jo vastaavan ihmiskääntäjien tuottamaa laatua. Pintapuolisesti sujuvissa konekäännöksissä on kuitenkin usein vielä paljon parannettavaa. Nykyisellään tehokkaimmat syväoppiviin neuroverkkoihin perustuvat konekääntimet pitää räätälöidä, eli kouluttaa erikoiskielen lähtöteksteillä ja käännöksillä, jotta ne oppivat käyttämään vakiintuneita vastineita ja ilmaisutapoja.

Tutkielma käsittelee OPUS-CAT-neuroverkkokääntimen räätälöintiä Suomen Pankin vuosien 2019 ja 2020 vuosikertomuksilla ja niiden käännöksillä sekä vertaillaan Suomen Pankin vuosikertomuksen 2021 ruotsinnosta räätälöidyllä konekääntimellä tuotettuun ruotsinnokseen. Tutkimuksen tavoitteena on selvittää, miten räätälöinti vaikuttaa konekäännökseen, kuinka konekääntäminen suoriutuu niistä vuosikertomuksen osista, joita ei löydy käännösmuistista, ja kuinka työlästä konekääntimen koulutus on.

Tutkielman teorialuvussa käydään läpi käännösmuistien käyttöä, konekääntämisen historiaa ja viime vuosien kehitystä, konekäännösten laadunarviointia sekä neuroverkkokääntimen räätälöintiä yleisesti. Aineisto- ja menetelmäluvussa esitellään Suomen Pankin vuosikertomusta, vuosikertomusta tekstityyppinä sekä OPUS-CATin räätälöinnin työvaiheita. Lisäksi luvussa esitellään tutkielmassa käytettyä konekäännösten automaattista arviointimittaria (BLEU), manuaalisen arvioinnin pohjaksi muodostettua virheluokittelua sekä valittujen arviointitapojen soveltamista.

Analyysin pohjalta voi todeta, että rajallisellakin aineistolla räätälöinti parantaa OPUS-CATilla tuotettujen konekäännösten ja Suomen Pankin kielipalveluiden tuottaman käännöksen vastaavuutta merkittävästi. Toisaalta analyysi osoittaa myös, että koulutusaineiston rajallisuus tulee vastaan jo yhdestä vuosikertomuksesta kerätyn kaksikielisen aineiston jälkeen, sillä konekäännösmallin jatkoräätälöinti toisesta vuosikertomuksesta kerätyllä aineistoilla ei enää paranna vastaavuutta suuresti. Työprosessin kuvauksesta käy myös ilmi, että OPUS-CAT-neuroverkkokääntimen räätälöinti on helppoa eikä vaadi suurta teknistä perehtyneisyyttä.

Tutkielman johtopäätöksissä todetaan, että räätälöinnin optimoimiseksi aineiston keräys tulisi mahdollisuuksien mukaan ulottaa myös muihin keskuspankin julkaisuihin. Koska OPUS-CATin räätälöinnin voi kuitenkin tehdä suoraan käännösmuistiohjelmasta tuodulla tmx-tiedostolla, pelkällä edeltävän vuoden vuosikertomuksen käännökselläkin saa nopeasti ja helposti hyviä tuloksia.

## Sisällysluettelo

<b>1 Johdanto.....</b>	<b>1</b>
1.1 Tutkimuksen tavoite ja tutkimuskysymykset.....	2
1.2 Tutkielman rakenne.....	3
<b>2 Käännösteknologinen konteksti.....</b>	<b>4</b>
2.1 Tietokoneavusteinen kääntäminen (CAT) ja käännösmuistit.....	4
2.2 Konekääntäminen .....	6
2.2.1 Konekääntämisen historia 1950-luvulta 2000-luvulle .....	7
2.2.2 Modernit konekääntimet.....	10
2.2.3 Konekäännösten laadun arviointi.....	13
2.2.4 Neuroverkkokääntimen räätälöinti .....	14
<b>3 Aineisto ja tutkimusmetodi.....</b>	<b>15</b>
3.1 Vuosikertomus tekstityyppinä .....	16
3.2 Suomen Pankin vuosikertomuksen erikoiskieli .....	17
3.3 Aineisto ja sen keruu.....	18
3.4 OPUS-CAT-neuroverkkokääntimen räätälöinti.....	23
3.5 BLEU ja SacreBLEU.....	24
3.6 Konekäännösten manuaalinen arviointi .....	27
<b>4 Analyysi.....</b>	<b>28</b>
4.1 Käännösmallien ja konekäännösten BLEU-pisteytys.....	29
4.1.1 Ensimmäinen räätälöintivaihe.....	29
4.1.2 Toinen räätälöintivaihe.....	31
4.2 Konekäännöksen arviointi virheluokittelun pohjalta .....	33
<b>5 Päätelmät ja kokoavaa tarkastelua .....</b>	<b>41</b>
<b>Lähdeluettelo.....</b>	<b>43</b>

Sammanfattning på svenska

## Luettelot ja kuvat

### Kuvat

Kuvio 1:	Vauquoisin kolmio (Koehn 2017: kohta 10:14).	9
Kuvio 2:	Odotukset eivät aina vastaa todellisuutta (Koehn 2017: kohta 8:08).	12
Kuvio 3:	Tutkielmassa käytetty konekääntimen koulutuksen aineiston keruun työprosessi.	22
Kuvio 4:	OPUS-CATin räätälöimisen työvaiheet.	24
Kuvio 5:	Mallin OC2 räätälöintiprosessin graafinen kuvaus.	30
Kuvio 6:	Mallin OC3 räätälöintiprosessin graafinen kuvaus.	32

### Taulukot

Taulukko 1:	Suomen Pankin vuosikertomuksen rakenne vuosina 2019–2021.	21
Taulukko 2:	Tutkimusaineiston laajuus.	21
Taulukko 3:	Googlen pääsääntö BLEU-pisteille (Google 2022, käänös Appelberg).	25
Taulukko 4:	Konekäännösten manuaalisen arvioinnin virhematriisi (Appelberg 2021).	27
Taulukko 5:	Geneerisen OC1-mallin ja ensimmäisen räätälöidyn OC2-mallin käännösten BLEU-pisteet.	30
Taulukko 6:	Räätälöidyillä OC2- ja OC3-malleilla tuotettujen käännösten BLEU-pisteet.	32
Taulukko 7:	Konekäännösten BLEU-pisteet.	33
Taulukko 8:	Esikäntämisen (pre-translate-toiminnon) tuottamat osumat muistista.	34
Taulukko 9:	Esimerkkejä käännösmuistin osumista eri kynnyksiarvoilla.	35
Taulukko 10:	Virheanalyysin tulos.	36

### Kaavat

Kaava 1:	BLEU-mittarin modifioitu precision-kaavio (Papineni ym. 2002: 313).	25
Kaava 2:	Lyhyysakko (en. brevity penalty, BP) (Papineni ym. 2002: 315).	26
Kaava 3:	BLEU-kaavio kokonaisuudessaan (Papineni ym. 2002: 315).	26

# 1 Johdanto

Suomi on perustuslain (731/1999) §:n 17 mukaan kaksikielinen maa, ja siksi Suomessa tarvitaan ruotsia ja ruotsin kääntäjiä niin valtionhallinnossa kuin kuntatasolla. Kielitaito yhdistää Suomen myös muihin Pohjoismaihin, ja ruotsin taidosta on hyötyä esimerkiksi kaupankäynnissä Ruotsin kanssa. Suomalaisten ruotsin kielen taito on kuitenkin viime vuosina heikentynyt, muun muassa siksi, että ruotsin kirjoittaminen ylioppilaskokeissa ei ole enää pakollista (Rossi ym. 2017: 108). Myös ruotsin kääntäjistä on jo pitkään ollut pulaa (Antinjuntti 2018: 5; Gorschnik ym. 2021), eikä pulaa helpota viime vuosina kasvanut ruotsinnosten kysyntä. Tämä on nähtävissä esimerkiksi valtionhallinnon lisääntyneessä ajankohtaistekstien ruotsinnosten kysynnässä, eikä kasvavaan kysyntään tahdo löytyä riittävästi uusia ammattilaisia korvaamaan suurta määrää pian eläköityviä valtionhallinnon ruotsin kääntäjiä (Petrell 2019). Yksi apuväline tämän käännettävän aineiston kasvun ja ammattikäntäjien pienenevän lukumäärän välisen ristiriidan ratkaisuun on konekääntäminen. Omat kokemukseni konekääntämisestä ovat olleet kannustavia, ja haluan selvittää, miten voisin käyttää niitä tehokkaammin työssäni kielenkääntäjänä Suomen Pankin kielipalveluissa.

Viime vuosien käännosteknologinen kehitys on tuottanut suuren yleisön ja ammattilaisten käyttöön konekääntimiä, joiden väitetään saavuttaneen ihmiskääntäjän tason, ainakin tietyillä mittareilla ja rajatulla aineistolla (Hassan ym. 2018). Dzimitry Bahdanaun ym. (2014) ja Ashish Wasvanin ym. (2017) esittelemiä neuroverkkopohjaisia konekääntimiä (en. *neural machine translation, NMT*), eli koneoppimiseen perustuvia tietokoneohjelmia, pidetään nykyisellään tehokkaimpana konekääntämisen muotona (Haque ym. 2021: 1). Suuren yleisön silmissä muutos näkyi esimerkiksi 2016, kun Google esitteli neuroverkkokääntimeksi muuttamansa uuden Translate-konekääntimen (Wu ym. 2016). Suurista otsikoista huolimatta yksikään konekäännin ei kuitenkaan vielä kykene tuottamaan ihmiskääntäjien veroisia käännoksiä kokonaisista teksteistä (Koponen 2018: 8), ja ongelmat korostuvat pienten kielten kohdalla, kun tekoälyn riittävään kouluttamiseen ei riitä aineistoa (Haque ym. 2021: 1).

Puutteistaan huolimatta konekääntimet kuitenkin helpottavat myös kääntäjien työtä, sillä vaikka niiden tuottama raakakäännös ei vielä sellaisenaan olisi julkaisukelpoinen, raakakäännöstä voi parantaa jälkieditoinilla. Jotta konekääntimen tuottama raakakäännös vaatisi mahdollisimman vähän jälkieditointia, konekäännin täytyy kuitenkin *räätälöidä* tietyn rajatun tekstityypin tai erikoiskielen

kääntämiseen (en. *domain adaptation*<sup>1</sup>), sillä sanojen merkitys ja kielen esitystapa vaihtelevat eri konteksteissa. Siksi geneerinen neuroverkkokääntäminen tuottaa kieliopillisesti sujuvia mutta epätarkkoja käännöksiä erikoiskielisistä teksteistä. (Koehn & Knowles 2017: 28.)

## 1.1 Tutkimuksen tavoite ja tutkimuskysymykset

Tässä tutkielmassa tarkastelen neuroverkkokääntimen räätälöimistä työnantajani<sup>2</sup> Suomen Pankin 2019 ja 2020 vuosikertomusten käännöksillä sekä räätälöimäni konekääntimen tuottamaa ruotsinkielistä käännöstä vuoden 2021 vuosikertomuksesta vertaamalla sitä Suomen Pankin kielipalveluiden maaliskuussa 2022 julkaisemaan käännökseen samasta tekstistä. Tutkimuksen tavoitteena on selvittää, millä tavoin geneerisen OPUS-CAT-neuroverkkokääntimen käännös Suomen Pankin vuosikertomuksesta muuttuu, kun konekääntintä kouluttaa kahden aiemman vuoden julkaistuilla vuosikertomuksilla ja niiden ruotsinnoksilla. Mittaan konekäännöksen laadun kehitystä koulutusprosessin aikana OPUS-CATin integroidulla SacreBLEU-mittarilla. Tämän jälkeen arvioin konekäännöksiä erillisellä BLEU-työkalulla vertaamalla niitä Suomen Pankin kielipalveluiden tuottamaan käännökseen. Lopuksi arvioin parasta konekäännöstä virheanalyysimallin perusteella. Näiden perusteella pyrin vastaamaan seuraaviin kysymyksiin:

- mikä vaikutus räätälöinnillä on konekäännöksen laatuun (BLEU-arviointi),
- kuinka konekääntäminen suoriutuu niistä vuosikertomuksen osuksista, joita ei löydy käännösmuistista (manuaalinen arviointi) ja
- kuinka työlästä konekääntimen koulutus on (yleinen prosessin arviointi).

Tutkimuksella pyrin myös osoittamaan, että konekääntimen räätälöinti OPUS-CATilla on helppoa ja vaivan arvoista, mikä voisi yleisesti laskea kääntäjien kynnystä käyttää omin voimin koulutettavia konekääntimiä lisäresurssina oman työmääränsä helpottamiseksi.

Tutkimuksessa käyttämäni OPUS-CAT on käännösohjelmistojen Trados Studio, MemoQ ja OmegaT kanssa yhteensopiva kokoelma ammattikäyttäjille suunnattuja ei-kaupallisia konekäännöstyökaluja (Nieminen 2021: 288). OPUS-CAT koostuu Windows-käyttöjärjestelmään paikallisesti asennettavasta konekääntimestä ja sen lisäosista, joilla ohjelman voi integroida käännösohjelmistoihin ja joiden avulla käyttäjä pystyy räätälöimään neuroverkkokääntimen erikoisalan tekstien kääntämiseen (mp.) Käytän

---

<sup>1</sup> Termistä on useita suomennoksia, kuten *uudelleen koulutus*, *hienosäätö* tai pelkkä *koulutus*, mutta tässä tutkielmassa käytän termille pääasiassa suomennosta *räätälöinti*.

<sup>2</sup> Työskentelen itse ruotsin ja englannin kääntäjänä Suomen Pankin kielipalveluissa.

tutkimuksessa tukena myös Trados Studion käännösmuistiominaisuutta, sillä vuosikertomus sisältää runsaasti aiempien vuosikertomusten kanssa yhteneviä ja siksi toistuvia osia ja ilmaisia, eikä kaikkea kannata kääntää konekääntimellä, sillä useat käännökset ovat automaattisesti saatavilla muistista. Käännösmuisti tarjoaa kuitenkin osumia vain riittävän yhteneväisiin kohtiin, ja konekääntimen räätälöinnillä pyrin täyttämään käännösmuistin jättämiä aukkoja.

Olen rajannut koulutusaineiston vuosien 2019–2020 vuosikertomuksiin, sillä ne vastaavat rakenteeltaan ja sisällöltään pankin nykyistä verkkovuosisikertomusta. Käyttämäni käännösmuistit sisältävät saman aineiston, jota käytän OPUS-CAT-konekääntimen räätälöimiseen. Vaikka konekääntimen räätälöinnin kannalta olisi optimaalista käyttää mahdollisimman laajaa aineistoa, sillä koulutusaineiston määrä on suoraan verrannollinen konekääntimen kielenoppimisen kanssa (Koehn & Knowles 2017: 28), tutkimuksen toistettavuuden kannalta olen päättänyt käyttää rajattua avointa aineistoa, joka koostuu kielen ja terminologian osalta tarkastetuista käännöksistä. Aineiston määrän lisäksi konekääntimen tuottaman käännöksen laadun parantamiseen vaikuttaa kaikkein eniten räätälöintiin käytettävän aineiston laatu (Tiedemann 2020: 4). Sen perusteella, että räätälöintiin käyttämäni tekstit on tarkastettu sekä lähtö- että kohdekieleltään ja valmisteltu perusteellisesti asiantuntijoiden yhteisponnistuksena, koulutusaineistoa voi pitää laadukkaana. Tällä tavoin rajatun aineiston pohjalta pystyn myös esittämään suuntaa antavan arvion siitä, millaista käytännön hyötyä räätälöidystä konekääntimestä on jälkiedoitavuuden näkökulmasta juuri tässä tapauksessa verrattuna pelkän käännösmuistiohjelman käyttöön.

Tutkielman hypoteesi on, että rajallisesta aineistosta huolimatta räätälöinti parantaa konekääntimen tuottaman käännöksen laatua ja vähentää jälkiedoitotarvetta, sillä OPUS-CAT on suunniteltu toimimaan pienellä koulutusaineistolla (Nieminen 2021: 290). Koska vuosikertomus on yli 200-sivuinen, se sitoo paljon kääntäjäresursseja, ja sitä ajatellen räätälöity konekääntin voisi helpottaa kääntäjien työtaakkaa.

## 1.2 Tutkielman rakenne

Luvussa 2 asetan tutkielmalle käännösteknologisen kontekstin, eli kerron käännösmuistien yleisistä toimintaperiaatteista sekä käyttämästäni Trados Studio -käännösmuistiohjelmasta (alaluku 2.1), konekääntämisestä (2.2) sen historiasta, periaatteista ja tutkimuksesta (2.2.1 ja 2.2.2), konekäännösten laadunarvioinnista (2.2.3) sekä neuroverkkokääntimen räätälöinnistä (2.2.4). Luvussa 3 käyn läpi aineistoa ja tutkimusmetodia, aloittaen vuosikertomuksesta tekstityyppinä (3.1). Seuraavaksi esittelen Suomen Pankin vuosikertomuksen erikoiskieltä (3.2), mistä etenen aineiston

keruuseen (3.3), OPUS-CAT-konekääntimen räätälöintiin (3.4), mitä mittareita käytän käännösten automaattiseen arviointiin (BLEU & SacreBLEU; 3.5) sekä millä tavoin arvioin konekäännöksiä manuaalisesti, eli mitä virheluokitteluja käytän analyysissä (3.6). Luku 4 koostuu tutkittavan aineiston eri analyyseistä ja tuloksista. Luvussa 5 käyn läpi tutkimuksen johtopäätöksiä ja pohdin tutkimuksen käytännön hyötyjä ja ansioita sekä havaittuja rajallisuuksia.

## 2 Käännösteknologinen konteksti

Seuraavassa esittelen tietokoneavusteista kääntämistä ja käännösmuisteja (luku 2.1), konekäännintä toimintaa yleisesti (2.2), konekäännintä historiaa (2.2.1 ja 2.2.2), konekäännösten laadun arviointia (2.2.3) ja neuroverkkokääntimien räätälöintiä (2.2.4). Käännösmuistien ja eri konekääntimien toimintaperiaatteiden sekä niiden rajallisuuksien selkeyttäminen asettaa tutkielmani kontekstiin, joka helpottaa ymmärtämään mitä hyötyä konekääntämisestä ja niiden räätälöinnistä voi olla ammattikäntäjän työssä laajojen käännösmuistien rinnalla.

### 2.1 Tietokoneavusteinen kääntäminen (CAT) ja käännösmuistit

Tietokoneavusteinen kääntäminen ja siihen liittyvät käännöstyökalut tai CAT-työkalut (en. lyhenteestä *CAT, computer-assisted translation*) ovat olennaisia osia nykyaikaisessa ammattikäntämisessä. CAT-työkalujen ominaisuuksista tärkein lienee käännösmuisti (en. *translation memory, TM*), jonka avulla kääntäjä voi hakea lähtötekstejä ja niiden käännöksiä sisältävästä kaksikielisestä tietokannasta aiemmin käännettyjä sanoja, ilmaisuja tai kokonaisia virkkeitä. Kun puhutaan tietokoneavusteisesta kääntämisestä, sillä yleensä tarkoitetaan tietokoneohjelmaan sisällytettyä käännösmuistia, sillä käännösmuistien käyttö on ammattilaiskontekstissa tietokoneavusteisen kääntämisen keskeisin ominaisuus. Tietokoneavusteisella kääntämisellä voidaan viitata myös muihin kääntäjälle hyödyllisiin ominaisuuksiin ja käsite kattaa myös mm. verkkopohjaiset käännösmuistit, termityökaluja, oikolukutoimintoja ja konekääntämistä (en. *machine translation, MT*). (Garcia 2014: 68–69.) Tässä tutkielmassa keskityn käännösmuistiin ja konekääntämiseen, joita käytän osana Trados Studio-käännöstyökalua.

Ensimmäiset tietokonepohjaiset käännösmuistit kehitettiin 90-luvulla vastauksena kasvavaan lokalisoitintarpeeseen, tavoitteena mahdollistaa suurien tekstimassojen kääntäminen tietokoneilla useamman kääntäjän voimin yhtä aikaa niin, että sanasto ja tyyli kuitenkin pysyvät yhdenmukaisena (Garcia 2014: 68–69). Kun kääntäjillä on reaaliaikainen pääsy kollegoidensa käännöksiin, työ nopeutuu



ja termistö pysyy helpommin yhdenmukaisena. Ennen tietokoneavusteisia työkaluja kääntäjät joutuivat turvautumaan paperiarkistoihin, ja esimerkiksi EU-kääntäjät joutuivat käyttämään arvokasta työaikaansa samojen tekstien uudelleenkääntämiseen (Christensen & Schjoldager 2017: 2). Tietokonekäyttöisen käännösmuistin tietokanta kuitenkin laajenee työn ohessa sitä mukaa kuin kääntäjä tallentaa sinne käännöstyökalulla tekemiään käännöksiä. Tietokantaa voi laajentaa myös käyttämällä kohdistustoimintoa (en. *alignment tools*) (mp.), jonka avulla kääntäjä syöttää käännösmuistiin dokumentteja ja niiden käännöksiä, jotka ohjelma ja tarvittaessa kääntäjä yhdistää toisiinsa virketasolla.

Tutkielmassa käyttämäni käännöstyökalu Trados Studio on yksi pisimpään markkinoilla olleista CAT-työkaluista ja alan markkinajohtaja yli 70 %:n markkinaosuudella (ProZ 2013). Trados Studio perustuu käännösmuistien käyttöön, mutta sisältää myös muita työkaluja. Uutta käännöstä aloittaessa ohjelma pilkkoo sille syötetyn käännettävän lähtötekstin yksi kerrallaan käännettäviin *segmentteihin*, eli käytännössä virkkeistä koostuviin merkkijonoihin, mutta myös epätäydelliset virkkeet, kuten otsikot, luettelokohdat ja taulukoiden solut, jakautuvat omiin segmentteihinsä. Kun kääntäjä on kääntänyt segmentin ja siirtyy seuraavaan segmenttiin, käännös tallentuu valittuun käännösmuistiin. Muistin sisältämistä segmenteistä käytetään Tradosissa nimitystä *translation unit (TU)*, josta käytän tutkielmassa nimitystä *segmentti*. Muistiin tallennettu segmentti käsittää paitsi lähtötekstin ja sen käännöksen, myös käyttäjän asettamia metatietoja, esimerkiksi tallennuspäivämäärän, lähdedokumentin nimen ja kääntäjän nimen. Kun vastaava kohta esiintyy myöhemmin samassa tekstissä tai vuoden päästä toisessa tekstissä, sen käännös löytyy muistista välittömästi, mikäli käyttäjä käyttää kyseistä muistitiedostoa. Muisteja voi ja kannattaa olla useita eri töitä varten – lääketieteellisten käännösten muistista ei esimerkiksi ole suuresti hyötyä retkeilyvarusteita käsittelevän kuvaston kääntämisessä. Trados tallentaa muistisegmentit sdtm-tiedostoon tai pilveen ja käännösmuistin voi jakaa muille kääntäjille perinteisesti tiedostonsiirtona tai serverimuistina pilvessä. Muistin sisältämät käännösyksiköt voi viedä myös tmx-tiedostoon (*Translation Memory eXchange*), joka on XML-pohjainen muistitiedosto ja yhteensopiva myös muiden käännösohjelmien kanssa.

Uuteen segmenttiin siirryttäessä Trados Studio näyttää kääntäjälle *identtisiä vastaavuuksia* (en. *exact match*) tai *osittaisia vastaavuuksia* (en. *fuzzy match*), mikäli muistista löytyy lähtötekstiä vastaavaa tekstiä. Identtinen vastaavuus edellyttää, että vanha lähtökielen segmentti vastaa käännettävänä olevaa tekstiä täydellisesti. Osittainen vastaavuus tarkoittaa, että ohjelma arvioi käännettävän tekstin vastaavan muistista löytyviä merkkijonoja 70–99 %:n varmuudella. Jos vastaavuus jää alle 70 %:n, ohjelma ei tarjoa valmista käännöstä. Koska osumien tunnistaminen perustuu merkkijonon samankaltaisuuteen, semanttisesti, pragmaattisesti ja tekstiyhteydeltään hyvinkin samanlaiset tekstit

voivat jäädä 70 %:n alapuolelle. (Christensen & Schjoldager 2017: 3.) Myös Tradosissa raja on lähtökohtaisesti 70 %. Kääntäjä voi kuitenkin hakea vastaavuuksia muistista myös itse. Jos esimerkiksi käännettävänä oleva virke on aiemmin jaettu kahteen erilliseen virkkeeseen, erilliset muistiin tallennetut segmentit vastaavat käännettävänä olevaa tekstiä vain 50 % osalta, mutta molemmat aiemmin käännettyt segmentit löytyvät manuaalisella haulla hänen käänösmuististaan, eikä hänen tarvitse palautella mieleensä vanhoja käännöksiään.

Edellä kuvatun lisäksi käänösmuistien käytössä tulee huomioida myös sen muut rajallisuudet. Kokemukseni mukaan edes sataprosenttinen vastaavuus ei itsessään tarkoita, että muistiin tallennettu käänös on hyvä, vaan ainoastaan, että lähtöteksti on sama kuin muistiin tallennetussa segmentissä. Tallennettu käänös voi olla huono, kontekstiin sopimaton tai, jos kyseessä on esimerkiksi yhden sanan käänös, yksinkertaisesti väärä sana. Kun kääntäminen tapahtuu segmenteittäin, kääntäjän huomio voi myös siirtyä tekstistä kokonaisuutena virke- ja lausetasolle, mikä voi vähentää syntaktista koheesiota (Pym 2011: 3) ja tehdä käännöksestä tyyliltään epäjohdonmukaisen (Bowker 2005: 16).

Käänösteknologiaa käyttäessään kääntäjän täytyy siis olla tietoinen työkalujensa puutteista. Trados Studiossa, kuten monissa muissakin käänösmuistiohjelmista, ohjelman voi muistiosumien sijaan tai niiden lisäksi asettaa tarjoamaan kääntäjälle myös ohjelmaan liitetyn konekääntimen tuottamia käännöksiä, eli konekäännöksiä. Näitä käyttäessään kääntäjä ei kuitenkaan operoi enää aiemmin hyväksi havaittujen käännösyksiköiden varassa, sillä konekäännökset eivät perustu suoraan ihmiskääntäjien tekemiin käännöksiin. Kokemukseni mukaan konekäänöksissä käännösten epäjohdonmukaisuus korostuu entisestään, sillä konekäännin ei käännä tekstejä lainkaan kokonaisuutena, vaan segmenteittäin.

## 2.2 Konekääntäminen

Altavistan Babel Fish, internetin ensimmäinen ilmainen konekäännin, julkaistiin jo vuonna 1997, ja Google Translate on ollut olemassa jo vuodesta 2006, mutta vasta viimeisen kuuden vuoden aikana konekääntämisestä on tullut todella arkipäivää myös kuluttajille, sillä 2016 esitellyn neurooverkkokääntämisen myötä konekääntimien tuottama käänöslaatu on parantunut merkittävästi. Monet sosiaalisen median palvelut ovat integroineet konekääntimiä suoraan käyttöliittymiinsä, mikä mahdollistaa täysin vieraalla kielellä kirjoitetun tviitin tai Facebook-päivityksen ymmärtämisen, ainakin suuntaa antavasti.

Ammattikäntäjät ovat vuosien saatossa suhtautuneet konekäntämiseen vaihtelevasti. Yhtäältä konekäntäminen on pitkään sen välttävän laadun takia nähty korkeintaan kurioositeettina ja toisaalta kuitenkin myös uhkana koko ammattikunnalle (Meijer 1993: 1), sillä sen varjolla käntäjiltä on vaadittu mm. entistä halvempaa työpanosta (Garcia 2009: 202). SDL:n vuonna 2020 teettämän kyselyn perusteella 26 % käntäjistä piti konekäntämistä aidosti uhkana ammatilleen (SDL 2020: 14). Jari Juutisen opinnäytetyönä suorittama kysely osoitti samansuuntaista osuutta suomalaisten englannista suomeen ja suomesta englantiin käntävien parissa, mutta yleisesti asenteet vaihtelevat vielä paljon (Juutinen 2020: 44, 47). SDL:n kyselyn perusteella käy kuitenkin ilmi myös, että käntäjien suhtautuminen konekäntämiseen muuttuu positiivisemmaksi käyttökokemusten myötä (SDL 2020: 14), mistä voi tulkita, että käntäjillä on vielä jonkin verran ennakkoluuloja tuntematonta teknologiaa kohtaan.

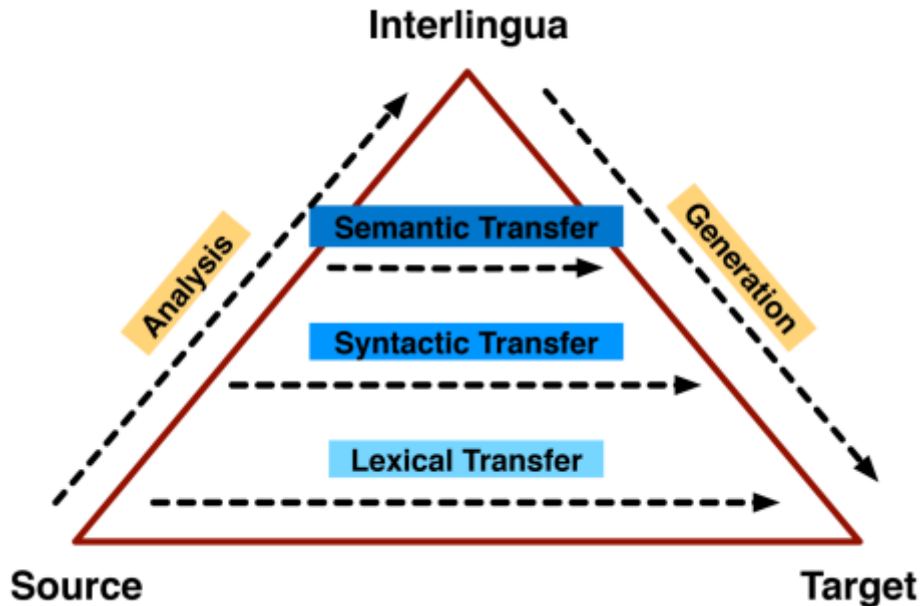
Teknologisesta edistymisestäään huolimatta konekäntöksen laatu ei vielä riitä tuottamaan sellaisenaan julkaisukelpoista käntöstä. Jotta konekäntöksestä saa julkaisukelpoisen, sitä pitää *jälkieditoita*. Standardin ISO 18587 (2017) määritelmän mukaan jälkieditoinnilla viitataan prosessiin, jossa ihminen, käytännössä käntäjä, kielentarkastaja tai muu kieliasiantuntija, korjaa konekäntöksen ja muokkaa siitä julkaisukelpoista tekstiä. Jälkieditoinnin on tutkimuksissa havaittu nopeuttavan erikoisalatekstien käntämistä merkittävästi verrattuna perinteiseen käntämiseen sekä vähentävän työn vaatimaa kognitiivista kuormitusta (Jia ym. 2019: 62).

### 2.2.1 Konekäntämisen historia 1950-luvulta 2000-luvulle

Seuraavaksi esittelen lyhyesti konekäntämisen historiaa 50-luvulta nykypäivään. Kuvauksessaan konekäntämisen historiasta John Hutchins (1995: 432–433) ajoittaa ensimmäiset haaveet konekäntämisestä 1600-luvulle, mutta käytännössä siitä tuli mahdollista vasta tietokoneiden myötä 1900-luvulla. Ensimmäiset patentit kahdelle tietokoneita edeltäneelle ”käntöskoneelle” haettiin jo 1933, jolloin ranskalais-armenialainen Georges Artsrouni ja venäläinen Petr Smirnov-Trojanski hakivat tahoillaan patentteja keksinnöilleen. Artsrounin nimitti keksintöään mekaanisiksi aivoiksi (fr. *cerveau mécanique*), ja kyseessä oli käntämisen kannalta mekaaninen sanakirja. Taas Trojanskin patentti koski ehdotusta kolmen kohdan käntöprosessiksi: ensimmäisessä vaiheessa lähtökielen osaaja suorittaa ”loogisen analyysin” sanoista ja niiden merkityksistä, toisessa vaiheessa kone käntää perusmuotoiset sanajonot ja niiden merkitykset toiselle kielelle ja kolmannessa vaiheessa kohdekielen osaaja muotoilee käntöksestä luonnollista kieltä. (Mp.)

Varsinaista käytännön toteutusta alettiin tutkia 1950-luvun vaihteessa, kun yksi konekääntämisen pioneereista, yhdysvaltalaisen Rockefeller-säätiön luonnontieteiden osaston johtaja ja matemaatikko Warren Weaver heinäkuussa 1949 julkaisi muistion otsikolla *Translation* (Weaver 1949). Muistiossa hän esitti, että konekääntämistä tulisi lähestyä neljästä näkökulmasta; sanojen monimerkityksellisyys tulisi ratkaista niiden lauseyhteyden perusteella; kääntämisen voisi nähdä loogisesti ja matemaattisesti ratkaistavana ongelmana; kääntämiseen tulisi suhtautua kuin koodinmurtamiseen, eli kryptologiaa voisi soveltaa käännostarkoituksiin; ja kaikki kielet saattavat perustua tiettyihin lingvistisiin universaaleihin, joita tulisi käyttää hyväksi konekääntämisessä (Weaver 1949: 8–11; Hutchins 1995: 433). Näkisin, että Weaverin esittämistä huomioista kolme ensimmäistä ovat edelleen relevantteja moderneissa konekääntimissä.

1950-luvulla konekääntämisen tutkimus lähestyi kysymystä kolmesta perusnäkökulmasta: ensimmäinen malli oli suora kääntäminen tietyn lähtökielen ja tietyn kohdekielen välillä mahdollisimman pienin syntaktisin muutoksin (en. *direct translation*). Toinen lähestymistapa, ns. transfer-malli, koostui kolmesta vaiheesta: ensimmäisessä vaiheessa lähtöteksti muunnetaan abstrakteiksi käsitteiksi, toisessa vaiheessa käsitteet siirretään kohdekielille ja kolmannessa vaiheessa kohdekieliset käsitteet muunnetaan luonnolliseksi kohdekieleksi. Kolmas malli perustui käytetyistä kielistä riippumattomien koodien tai symbolien käytölle. Tässä lähtökieli ensin koodataan välikielelle (en. *interlingua*, esimerkiksi esperanto tai pidgin), minkä jälkeen välikieli puretaan kohdekielille. (Hutchins 1995: 433.) Ensimmäisessä mallissa koneelta vaaditaan paljon syvällistä osaamista, mutta vain vähän lähtötekstin analyysiä, kun taas interlingua-malli edellyttää syvällistä lähtötekstin analyysiä, mutta ei niin syvällistä kielten erojen tuntemista. Bernard Vauquois visualisoi nämä kolme lähestymistapaa kolmioon vuonna 1968 ja loi näin perinteiseksi käsitteeksi muodostuneen Vauquoisin kolmion (kuvio 1).



Kuvio 1: Vauquoisin kolmio (Koehn 2017: kohta 10:14).

Ensimmäisenä konekääntimenä pidetään yleisesti IBM:n ja Georgetownin yliopiston yhteistyössä kehittämää ja vuonna 1954 esittelemää venäjä–englanti-konekäännösjärjestelmää. Se perustui rajalliseen sanastoon ja vain muutamaankielioppisääntöön (Hutchins 2014: 120) ja edusti siten alkeellista versiota yhdestä konekääntämisen kolmesta pääsuuntauksesta, eli sääntöpohjaista konekääntämistä (en. *rule-based machine translation, RBMT*). Näytösluontoisuudestaan huolimatta se toimi alkusysäyksenä konekääntämisen kehitystyölle, ja unelmat nopeasti kehitettävistä täysin automaattisista konekääntimistä keräsivät runsaasti rahoitusta seuraavan vuosikymmenen ajan.

Alkuinnostuksesta huolimatta mullistavat tulokset antoivat kuitenkin odottaa itseään, ja lunastamattoman optimismin jälkeen vuonna 1964 Yhdysvaltain hallitus asetti ALPAC-komitean (*Automatic Language Processing Advisory Committee*) arvioimaan kehitystä. Vuonna 1966 julkaisemassaan raportissa komitea totesi automaattisen konekääntämisen olevan hitaampaa, vähemmän tarkkaa ja kalliimpaa kuin kääntäjien käyttäminen ja suositteli yliopistoja kohdistamaan tutkimusrahat kääntäjien työvälineiden kehittämiseen, mikä jähmetti tutkimuksen Yhdysvalloissa vuosikymmeneksi. (Hutchins 2014: 122) Kehitys jatkui kuitenkin muualla maailmassa, ja myös ALPAC-raportin jäljiltä rahoituksensa säilyttäneet projektit jatkoivat työtään, keskiössään suoran kääntämisen sijaan interlingua- ja transfer-mallit. 1970–1989 kehitystyö keskittyi sääntöpohjaisiin konekääntimiin, ja yhtenä tunnettuna virstanpylväänä voidaan nostaa esille SYSTRANin konekääntäminen. Tätä USA:n ilmavoimat ryhtyi käyttämään käännössuunnassa venäjä–englanti vuonna 1970 ja EU-komissio käännössuunnassa englanti–ranska vuonna 1976. (Mp.) Sääntöpohjaiset konekääntimet perustuvat

sanakirjojen ja kielioppisääntöjen yhdistämiseen, mutta tuottavat käytännössä hyödyllisiä käännöksiä vain, jos sanastoa ja ilmaisutapaa on rajattu, sillä järjestelmään pitää ohjelmoida jokainen kielioppisääntö erikseen, eikä sanakirjakääntäminen usein tuota kovinkaan luonnollista kieltä (Maučec & Donaj 2019: 3). Sääntöpohjaisille konekääntimille on kuitenkin paikkansa, mistä esimerkkinä voi mainita 1976 kehitetty Météo-järjestelmän, jota käytettiin Kanadan Quebecissä menestyksekkäästi säätiedotusten kääntämiseen kieliparissa englanti–ranska vuosina 1981–2001.

Seuraava merkittävä kehitysaskel oli siirtymä tilastollisiin konekääntimiin (en. *statistical machine translation, SMT*), kun IBM:n tutkimuslaitoksen tutkijat Peter Brownin johdolla vuonna 1988 esittelivät alkuperäisen ja vuonna 1993 parannellun puhtaasti tilastollisen konekäännösmenetelmän. Malli perustui suurten kaksikielisten korpusten käyttöön ilman mitään lingvistisiä sääntöjä, mutta tuotti hyväksyttävän käännöksen lähes puolessa tutkimusaineistosta. Käytännössä se oli paluu Weaverin vuonna 1949 ehdottamaan matemaattisen ongelmanratkaisun lähestymistapaan, mutta tehokkaammilla tietokoneilla. (Hutchins 2014: 128.)

Vuonna 1993 esitellyssä IBM:n SMT-mallissa laajan kaksikielisen korpuksen sisältö kohdistetaan kielestä toiseen sanatasolla (en. *word alignment*). Yhdistelmien perusteella käännin muodostaa käännösmallin (en. *translation model*), jossa se määrittää jokaiselle sanalle todennäköisimmän käännöksen. Mitä useammin sana esiintyy korpuksessa, sitä todennäköisempänä kone pitää, että käännös on oikea. Koska SMT perustuu tilastollisille menetelmille eikä ymmärrä kieltä, se tarvitsee tuekseen kielimallin (en. *language model*), eli yksikielisen kohdekielisen korpuksen, jonka perusteella kone arvioi, mikä on todennäköisin kohdekielen sana kyseisessä yhteydessä. (Hutchins 2014: 128.) En mene tässä tarkemmin SMT:n taustalla vaikuttaviin monimutkaisiin tilastotieteellisiin menetelmiin, mutta on kuitenkin mainittava, että perusmuotoinen tilastollinen konekäännin käyttää kaksikielistä korpusta kääntämiseen ja yksikielistä kohdekielen korpusta luonnollisen kielen muodostamiseen. Pysyväksi ongelmaksi muodostui kuitenkin todennäköisimmän sanajärjestyksen valinta, sillä eri kielet eivät ole keskenään symmetrisiä. Ongelmaa on pyritty ratkomaan kehittämällä erilaisia sanajärjestysmalleja (en. *reordering model*), mutta toistaiseksi tähän ei ole löydetty yleisesti toimivaa ratkaisua (Bisazza & Federico 2016: 163).

## 2.2.2 Modernit konekääntimet

Bisazza ja Federico (2016: 166) viittaavat 2000-luvun taitteessa julkaistuihin tutkimuksiin (Och 1999; Zens, Och & Ney 2002; Koehn, Och & Marcu 2003) ja katsovat, että ne paransivat IBM:n alkuperäistä sanapohjaista mallia esittelemällä korpusten kohdistamista yksittäisten sanojen sijaan

sanayhdistelmien tasolla (en. *phrase-based statistical machine translation*, *PBSMT* tai *PSMT*), mikä osaltaan paransi myös oikean sanajärjestyksen muodostusta, sillä esimerkiksi sanaparit ja yhteenkuuluvat ilmaisut yhdistyvät todennäköisemmin oikein. Englanninkielinen *phrase* on kuitenkin hieman harhaanjohtava nimitys, sillä kyse ei ole kielioppisääntöjen mukaisista lausekkeista, vaan pelkistä sanayhdistelmistä ja vakiintuneista ilmaisuista. IBM/Brown-mallista lähtien SMT oli lupaavin ja käytetyin käytännön lähestymistapa konekääntämiseen 2010-luvun puoliväliin asti (mp.). Esimerkiksi Googlen konekäännin oli SMT-pohjainen julkaisustaan vuonna 2006 vuoteen 2016. Nykyaikaiset PBSMT-konekääntimet käsittävät tyypillisesti usean sana- tai sanayhdistelmäpohjaisen käännösmallin, kohdekielen kielimallin, sanajärjestyksen muutossakon (en. *distortion penalty*) sekä muita käyttötarkoituksen mukaan määriteltäviä sääntöjä (mas. 167), eli ne eivät ole enää pelkästään tilastollisia.

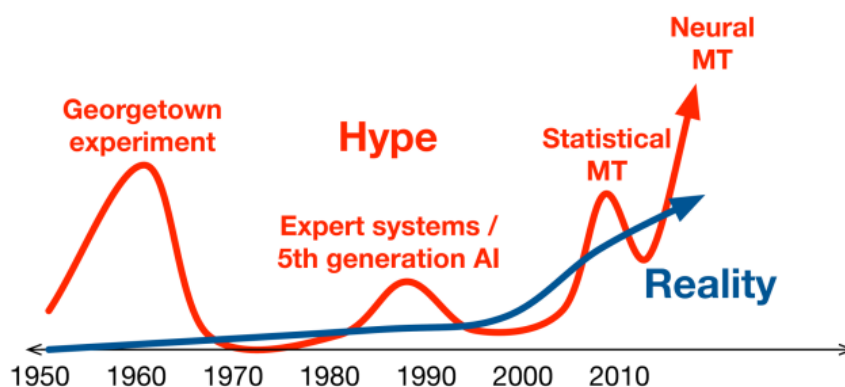
Konekääntämisen viimeisimmän vallankumouksen juuret ajoittuvat vuoteen 2013, kun Nal Kalchbrenner ja Phil Blunsom (2013: 1700) esittelivät syväoppivia neuroverkkoja hyödyntävän *encoder-decoder*-järjestelmällä toimivan konekäänninmallin, eli ensimmäisen neuroverkkokääntimen (en. *neural machine translation*, *NMT*). Tässä kahden eri neuroverkon muodostamassa neuroverkoarkkitehtuurissa ohjelma koodaa ensin lähtötekstin vektoreiksi (eli konekielelle) käyttäen konvoluutioneuroverkkoa (en. *convolutional neural network*, *CNN*), ja purkaa sitten vektorit kohdekieliseksi tekstiksi käyttämällä takaisinkytkettyä neuroverkkoa (en. *recurrent neural network*, *RNN*). Tilastollinen konekäännin arvioi käännöksen luotettavuuden sanojen esiintymistiheyden perusteella, mikä aiheuttaa ongelmia etenkin harvinaisten sanojen kääntämisessä (mp). Tätä Kalchbrenner ja Blunsom pyrkivät kompensoimaan käyttämällä neuroverkkoja, jotka yhdessä muodostavat erillisten käännös- ja kielimallien sijaan yhden suuren käännösmallin. Tämän mallin pullonkaulaksi muodostui lauseensisäisten riippuvuussuhteiden hahmottaminen, sillä jo kymmenen sanan lauseissa toisistaan kaukaisten sanojen riippuvuussuhde häviää. Ilmiötä kutsutaan englanniksi nimellä *vanishing gradient*. (Bahdanau ym. 2014: 1.)

Edellä mainitun vallankumouksen alun jälkeenkin neuroverkkokääntimien (NMT) ja tilastollisten konekääntimien (SMT) kehitys jatkui vielä rinnakkain, sillä NMT ei ollut vielä ylittänyt parhaita SMT-järjestelmiä. Vuonna 2014 Sutskever, Vinyals ja Le esittelivät mallin, jossa SMT:n koodaus ja koodinpurku tapahtuvat RNN-arkkitehtuurilla ja jossa toisistaan kaukaisten sanojen riippuvuussuhteiden havainnointia parannettiin lisäämällä arkkitehtuuriin LSTM-muistikerros (en. *long short-term memory*). Lisäksi Bahnadau ym. (2014: 4) esittelivät vielä samana vuonna tästä edelleen parannellun version, lisäämällä malliin huomiovektorin (en. *attention*). Mallissa käännös muodostuu sana kerrallaan niin, että käännin jokaisen muodostamansa sanan jälkeen palaa hakemaan tietoa

lähtötekstistä, muodostaen joka sanan kohdalla uuden arvion parhaasta seuraavasta sanasta ennen sen lyömistä lukkoon. Myös tätä edeltäneessä mallissa käänös muodostui sana kerrallaan, mutta jos yksikin sana menee väärin, käänin jatkaa kerran valitsemansa rakenteen perusteella. Näin etenkin lauseen pidentyessä käänös menee entistä enemmän väärin. Bahnadaun ym. vuonna 2014 esittelemä huomiomekanismin sisältävä malli paransi SMT:n suorituskykyä, mutta viimeksi mainittua pitkien lauseiden ongelmaa ei ole vielä kukaan ratkaistu täydellisesti, vaan se on edelleen yksi konekääntämisen suurista haasteista (Koehn & Knowles 2017: 1). Tämän jälkeen kehitys on ollut nopeaa, ja Valintina Ragnin ja Lucas Nunes Vieiran meta-analyysissä todetaankin, että tutkimukset vuosien 2015–2019 tukevat väitettä, että neuroverkkokääntäminen on tällä hetkellä tehokkain konekääntämisen muoto (Ragni & Vieira 2022: 142).

Suurelle yleisölle näkyvin muutos tapahtui 2016 kun Google julkaisi neuroverkkokääntimensä (en. *Google Neural Machine Translation, GNMT*) (Wu ym. 2016). Tämän jälkeen suurin yksittäinen edistysaskel on ollut RNN-arkkitehtuurin korvaaminen kevyemmällä Transformer-arkkitehtuurilla, jossa RNN- ja CNN-neuroverkot jätettiin kokonaan pois ja korvattiin pelkällä huomiomekanismilla (Vaswani ym. 2017). Tällä hetkellä suurin haaste on konekäännösten laadun parantaminen rajallisella aineistoilla, etenkin pienissä kielissä, sillä nykyisellään niiden kouluttaminen vaatii todella suuria ja raskaita datamääriä (Koehn & Knowles 2017; Sennrich & Zhang 2019).

Vuonna 2018 Microsoft julisti, että konekäännin on saavuttanut ihmiskääntäjän tason (Hassan 2018). On kuitenkin huomioitava, että tutkimuksen kielet, kiina ja englanti, ovat maailman suurimmat kielet, joista löytyy kaikkein eniten koulutusaineistoa, joten tulokset eivät ole yleistettävissä kaikkien kielten kääntämiseen. Ja kuten alla olevasta Philipp Koehnin kuviosta 2 käy ilmi, konekääntämisen historia ollut täynnä suuria odotuksia 50-luvulta lähtien, mutta odotuksiin kannattaa suhtautua varauksella.



Kuvio 2: Odotukset eivät aina vastaa todellisuutta (Koehn 2017: kohta 8:08).



Ottaen huomioon, että ihmistasoista konekäännöstä ei vielä ole kehitetty, eikä varmaankaan lähitulevaisuudessa tulla kehittämään (Ragni & Vieira 2022: 148), voidaan todeta, että Smirnov-Trojanskin vuonna 1933 esittämät ajatukset olivat lähes 90 vuotta aikaansa edellä – toistaiseksi kolmivaiheisesta prosessista on poistettu vasta ensimmäinen vaihe. Tiivistäen muotoilisin asian niin, että konekäännökset ovat toistaiseksi vain matemaattisten reunaehtojen pohjalta tuotettuja parhaita arvauksia todennäköisimmistä oikeista sanoista, oikeasta sanajärjestyksestä ja sopivasta tyylistä.

### 2.2.3 Konekäännösten laadun arviointi

Kuten edellä luvussa 2.2 esitin, yksi konekääntämisen suurimmista teknologisista lupauksista on ollut siirtyminen neuroverkkopohjaisiin konekääntimiin (NMT), sillä niiden on osoitettu tuottavan parempaa laatua kuin tilastollisten konekääntimien (SMT) monilla yleisillä mittareilla (Koehn & Knowles 2017: 28) ja ne tuottavat yhä luonnollisempaa kieltä (Toral ym. 2017: 1071), joskin suurten kielten osalta paremmin kuin pienten. Mutta mistä oikeastaan puhutaan, kun puhutaan konekäännöksen laadusta ja miten sitä mitataan? Pyrin vastaamaan kysymyksiin luvussa 2.2.3, jossa käsittelen konekäännösten laadun automaattisesti ja ihmisvoimin tehtävää arviointia, sekä luvuissa 3.5 ja 3.6, joissa käsittelen tutkielmaan valittua automaattista arviointimittaria. Yksi tapa tuottaa tarkempia käännöksiä erikoiskielisistä teksteistä on konekääntimen räätälöinti, jota käsittelen luvussa 2.2.4.

Hyvän käännöksen voi määritellä usealla tavalla teoreettisesta näkökulmasta riippuen, esimerkiksi Nidan ekvivalenssiteorian mukaan (käännös on hyvä, jos se saa kohdekielen lukijassa aikaiseksi saman reaktion kuin lähdekielen teksti sai lähdekielen lukijassa, Pym 2014: 46) tai Vermeerin ja Reissin skoposteorian ja muiden funktionaalisten käännösteorioiden mukaan (käännös on hyvä, jos se täyttää funktionsa, mts. 46). Käännöksen laatu on kuitenkin subjektiivinen kysymys ja oikeita käännöksiä on yhtä monta kuin kääntäjiä. Konekäännösten laadusta puhuttaessa asiaa tulee ainakin vielä lähestyä karkeammasta kulmasta, eli sisältääkö käännös oikeat sanat ja välittykö alkuperäinen merkitys.

90-luvulle asti konekäännöksiä arvioitiin vielä ihmisvoimin ja mittareina käytettiin käännöksen adekvaattisuutta, informatiivisuutta ja sujuvuutta (White & O'Connell 1996: 107), eli noudattaako käännös tulokielen normeja ja onko siinä otettu huomioon käännöksen lukijoiden tarpeet ja odotukset (en. *adequacy*), välittykö siitä sama tieto kuin lähtötekstissä (en. *informativeness*) ja onko se kieliopillisesti hyvää ja sopiiko se kontekstiin (en. *fluency*). Tarve konekäännösten automaattisen arvioinnin mittareiden kehittämiseksi kumpusi ihmisarvioiden korkeasta hinnasta,

kertakäyttöisyydestä, subjektiivisuudesta ja hitaudesta (Dorr ym. 2009: 805) sekä halusta mahdollistaa konekääntimien tehokas hienosäätö (Och 2003: 160).

Useimmat automaattisen arvioinnin mittarit perustuvat konekäännöksen ja ihmiskäännöksen vertaamiseen. Näistä suurimman suosion on saavuttanut Kishore Papinenin ym. vuonna 2002 esittelemä BLEU, *bilingual evaluation understudy*, joka heikkouksistaan huolimatta on säilyttänyt asemansa konekäännösten automaattisen arvioinnin standardimittarina, ja sen perusteella uusia mittareita kehitetään ja siihen niitä verrataan. (Dorr 2009: 818.) BLEU:n lähtökohtana on, että mitä lähempänä konekäännös on ammattikäntäjän tekemää käännöstä, sitä parempi se on (Papineni ym. 2002: 3011). Olen valinnut BLEU:n tutkielmani automaattiseksi mittariksi, ja avaan sen toimintaperiaatteita tarkemmin luvussa 3.5.

Banerjee ja Lavie esittelivät vuonna 2005 BLEU:n pohjalta kehittämänsä METEOR-mittarin. Julkaisussaan he argumentoivat, että automaattiselle mittarille tärkeintä on korreloida ihmisen laatukäsityksen kanssa. Tällä he tarkoittavat, että mittarin tulisi antaa hyvä arvio käännökselle, jota ihmisarvioija pitäisi hyvänä, ja huono arvosana käännökselle, jota ihmisarvioija pitäisi huonona. (Banerjee & Lavie 2005: 66.) Heidän mukaansa tämän minimilähtökohdan lisäksi mittarilta voi lisäksi edellyttää neljää muuta ominaisuutta: tarkkuutta, eli kykyä huomata laatueroja eri konekääntimien välillä, johdonmukaisuutta, eli kykyä tuottaa johdonmukaisia arvioita samankaltaisen lähdemateriaalin käännöksistä, luotettavuutta, eli eri konekääntimien arvioiden tulee olla verrannollisia keskenään, ja yleistettävyyttä, eli mittarin tulee olla sovellettavasti laajasti eri tekstityyppeihin (en. *correlation, sensitivity, consistency, reliability & generality*) (mp.). BLEUta on vuosien mittaan kritisoitu monesta puutteesta, mutta toistaiseksi se on säilyttänyt suosionsa ks. luku 3.5.

#### 2.2.4 Neuroverkkokääntimen räätälöinti

Internetin ilmaiset konekääntimet, esim. Google Translate, Microsoftin Bing Translator ja DeepL Translator, ovat voittopuolisesti neuroverkkopohjaisia ja niin sanotusti geneerisiä, eli niitä ei ole suunnattu minkään tietyn tekstityypin kääntämiseen, vaan niiden pyritään selviytyvän mahdollisimman hyvin kaikesta kääntämisestä. Ammattimaisessa kääntämisessä asiakas onkin lopulta se, joka määrittää mikä on käännökselle hyväksyttävä laatu – sitä saa, mistä on valmis maksamaan, ja jos ei ole valmis maksamaan mitään, keskimäärin ymmärrettävä konekäännös voi olla riittävä. Viranomaisviestinnässä erikoiskieliseltä käännökseltä tulee kuitenkin odottaa parempaa laatua, ja silloin yleiskielinen konekäännin ei tarjoa riittävän laadukasta lopputuotetta. Erikoiskielen kääntämisessä yleiskielisestä konekääntimestä ei silti ole hyötyä enää kääntäjällekään, sillä mitä

enemmän konekäännös vaatii jälkieditointia, sitä hyödyttömämmäksi se jää, ja lopulta konekäännöksestä voi olla enemmän haittaa kuin hyötyä (Koponen 2016b: 3). Kuten johdannossa kerroin, geneerinen neuroverkkokäännin voi tuottaa kieliopillisesti sujuvia käännöksiä erikoiskielisistä teksteistä, mutta niistä tulee epätarkkoja, sillä konekäännin ei tunne erikoiskielen käytänteitä (Koehn & Knowles 2017: 28). Ratkaisuna tähän on konekääntimen räätälöinti mahdollisimman suurilla (mp.) ja laadukkailla (Tiedemann 2020: 4) tekstimassoilla. Suomessa konekääntimien räätälöintiä tietyn erikoisalan teksteillä ovat viime vuosina tutkineet opinnäytetöissään mm. Maria Mäkinen (2019) ja Saara Salminen (2020).

Neuroverkkokääntimen räätälöinnissä yksi ongelmista on, että räätälöintiä ei pysty hallitsemaan sanatasolla kovinkaan tarkasti, eli harvinaisten sanojen opettaminen on hankalaa (Koehn & Knowles 2017: 1). Koneen oppiminen vaatii runsaasti aineistoa, ja liian rajatulla aineistolla oppimisprosessi voi johtaa ”katastrofaaliseen unohtamiseen” (Saunders 2022: 1), minkä seurauksena koulutuksessa täytyy palata edeltävään vaiheeseen ja tehdä se uudestaan eri aineistoilla tai parametreillä. Toinen suuri ongelma, etenkin rajatulla koulutusaineistolla, on liian homogeenisen aineiston käyttäminen kouluttamiseen ja koneen validointiin, jolloin kone ei opi soveltamaan kielitaitoaan, vaan oppii kääntämään ainoastaan käytetyn koulutusaineiston (mas. 2).

### 3 Aineisto ja tutkimusmetodi

Tässä luvussa ja alaluvuissa esittelen vuosikertomusta tekstityyppinä (3.1) Suomen Pankin vuosikertomuksen erikoiskieltä (3.2) ja aineiston rakennetta ja keruuta (3.3), OPUS-CATin räätälöintiä (3.4), sekä konekäännösten arviointia automaattisesti (3.5) ja virheluokittelun pohjalta (3.6). Automaattisena mittarina (en. *automatic evaluation*) käytän vuonna 2001 esiteltyä ja konekäännösten automaattisen arvioinnin standardiksi nousutta BLEU-mittaria. BLEU vertaa konekäännöstä ihmisen tekemään käännökseen siitä lähtökohdasta, että mitä lähempänä käännös on ihmisen samasta tekstistä tekemää käännöstä, sitä parempi se on (Papineni ym. 2001). BLEU-mittari sisältyy myös OPUS-CATiin, jossa se toimii konekääntimen kehitystyökaluna (Nieminen 2021: 288). Konekäännösten arvioinnin vertailukohtana käytän Suomen Pankin kielipalveluiden käännöstä vuoden 2021 vuosikertomuksesta.

Koska BLEU-pisteitys tarjoaa vain suuntaa antavan arvion käännöksen laadusta, eikä sitä tulisi käyttää ainoana käännöksen laadun mittarina tieteellisessä tutkimuksessa (Reiter 2018: 399–400), arvioin käännökset myös käyttäen konekäännösten arviointiin määrittelemiäni virheluokitteluja. Näiden

kahden arviointitavan pohjalta esitän arvion siitä, missä määrin Suomen Pankin kielipalveluille räätälöimistäni konekääntimestä voisi olla hyötyä pankin vuosikertomuksen kääntämisessä.

### 3.1 Vuosikertomus tekstityyppinä

Kirjanpitolain (1336/1997) 3. luvun 1. §:n mukaan yrityksen ja yhteisön pitää laatia tilikaudelta *tilinpäätös*, joka sisältää tuloslaskelman ja taseen sekä niiden liitetiedot sekä toimintakertomuksen. *Vuosikertomus* taas on julkaisu, joka sisältää toimintakertomuksen ja tilinpäätöksen sekä muut lain vaatimat pakolliset osat, mutta siinä yritys tai yhteisö voi myös kertoa toiminnastaan päättyneellä katsauskaudella sekä suunnitelmistaan ja tulevaisuudennäkymistään (Juholin 2017: 236). Vuosikertomus tarjoaa siten yritykselle tai yhteisölle viestintäkanavan, jolla toimija voi kertoa omasta toiminnastaan omistajille ja sidosryhmille omalla tavallaan. Yleensä vuosikertomus sisältää vähintään seuraavat osat (mts. 238–239):

1. Yhtiön tai yhteisön luonnehdinta ja toiminta-ajatus
2. Yhteenveto toiminnasta, avainluvut kertomusvuodesta ja edeltävältä vuodelta
3. Toimitusjohtajan katsaus
4. Toimintakatsaus liiketoimintayksiköittäin
5. Virallinen toimintakertomus ja tilinpäätös
6. Tilinpäätös tarvittaessa kansainvälisen (IFRS tai US GAAP) käytännön mukaan
7. Lukusarjat 5–10 vuoden ajalta sekä muuta tilastotietoa harkinnan mukaan
8. Johdon esittely
9. Organisaation esittely
10. Tietoja osakkaille
11. Yhteystiedot

Laki tilinpäätösten julkisuudesta tuli voimaan jo 1928, mutta minimivaatimukset tilinpäätöstiedoille asetettiin vasta vuoden 1974 kirjanpitolaissa. 70-luvulla Suomessa alettiin kiinnittää huomiota myös vuosikertomusten ulkoasuun ja julkaisuun alettiin suhtautua yrityksen esittelymateriaalina ja käyntikorttina sidosryhmille. 80-luvulla painetut ja näyttävät vuosikertomukset yleistyivät ulkomaisen mallin mukaan, ja laman jälkeen 1990-luvun loppupuolella alettiin järjestää vuosikertomusten ”kauneuskilpailuja”. (Roine 2012: 213.) Talousasiantuntijat vastaavat virallisesta sisällöntuotannosta, mutta yhä suurempi osa prosessista lepää nykyään viestinnän ja ulkoisten mainostoimistojen ja graafisten suunnittelijoiden työpanoksen varassa (mp.), niin myös Suomen Pankissa, jossa konsulttiyhtiö on ollut mukana prosessissa vuodesta 2018 lähtien (ks. luku 3.3).

Kääntäjän näkökulmasta vuosikertomus on tekstityyppinä monipuolinen ja haastava, sillä vuosikertomus sisältää yllä olevan kuvauksen mukaisesti sekä kovaa talousasiaa että markkinointiviestintää. Käännöksessä pitää siis pyrkiä lukijaystävälliseen ja ymmärrettävään kieleen, mutta samalla tekstissä on paljon yksityiskohtia, joita säädellään toimintakertomuksen ja tilinpäätöksen osalta laeissa ja asetuksissa. Organisaation käytikorttina vuosikertomus saa usein myös laajaa näkyvyyttä (Delingua 2019) ja on se olennaisesti sidoksissa aiempien vuosien vuosikertomuksiin, lehdistötiedotteisiin ja niissä käytettyyn kieleen. Monelta osin kääntäjän tulee olla tietoinen tiettyjen vakiintuneiden ilmausten käännöksistä (esim. visio, tavoitteet, arvot), ja johdonmukaisen viestinnän kannalta kääntäjän tulisi noudattaa vuoden aikana julkaistujen lehdistö- ja pörssitiedotteiden käännöksissä käytettyjä sanavalintoja, eikä esimerkiksi johdon tarkkaan harkittuja lausuntoja ole hyvä kääntää usealla eri tavalla.

### 3.2 Suomen Pankin vuosikertomuksen erikoiskieli

Erikoiskielen (en. *language for special purposes, LSP*) voi Päivi Pasasen mukaan määritellä esimerkiksi seuraavalla kahdella tavalla: sen voi nähdä olevan kieltä, jota käytetään ”jonkin erikoisalan asiantuntijoiden keskinäisessä viestinnässä tai viestinnässä alan asiantuntijoilta tuleville asiantuntijoille eli opiskelijoille tai harvemmin maallikoille” (Pasanen 2015: 179–180). Toiseksi erikoiskielen voi määritellä sen kielellisten piirteiden pohjalta, ”joista tärkein on tietyyntyyppisen sanaston sisältyminen kieleen”. Jäljemmän määritelmän perusteella termien käyttö tarkoittaa, että kyseessä on erikoiskieli, kun taas yleiskielessä ei käytetä termejä. Termistä Pasanen käyttää Kansallisen tieteen termipankin mukaista määritelmää ”erikoisalalla käytettävän yleiskäsitteen nimitys”. (Mp.) Christer Laurénin mukaan talouden erikoiskieli on yhdistelmä useampaa toisiaan sivuavaa teknolektiä (Laurén 1993: 45–46).

Koska Suomen Pankin käyttämä kieli on paikoin erittäin termipainotteista ja usein suunnattu asiantuntijoille, sen voi nähdä edustavan talouden erikoiskieltä. Suomen Pankin vuosikertomus sen sijaan sisältää sekä asiantuntijoiden toisille asiantuntijoille kirjoittamia tekstejä (esim. tilinpäätös) että maallikoille suunnattuja tekstejä pankin toiminnasta ja roolista viranomaisena, joten sen voisi sijoittaa jonnekin talouden ja markkinointiviestinnän genrejen välimaastoon. Vuosikertomuksessa käsitellään lyhyesti myös muita pankin toiminnan kannalta olennaisia aiheita, kuten vastuullisuutta, ympäristöä ja tietoturvaluottuutta, joiden osalta kieli ja termistö poikkeavat merkittävästi yleiskielestä. Koska talous liittyy lähes kaikkiin yhteiskunnan toimintoihin, laskisin nämäkin osaksi keskuspankkien käyttämää talouden erikoiskieltä.

Suomen Pankki on itsenäinen julkisoikeudellinen laitos ja Suomen rahaviranomainen, ja pankin ”virkakielestä on voimassa, mitä kaksikielisen virka-alueen virkakielestä säädetään kielilaisissa” (laki Suomen Pankista luku 4, § 24). Suomen Pankki on kielilain (432/2003) 6. §:n mukaisesti kaksikielinen keskushallintoviranomainen. Viranomaistehtävää hoitavana laitoksena pankin käyttämää kieltä sitoo myös hallintolain (434/2003) 9. §:n vaatimus asiallisesta, selkeästä ja ymmärrettävästä kielestä. Osana Euroopan keskuspankkijärjestelmää (EKPJ) ja eurojärjestelmää pankin kieltä ohjaa EKP:n ja eurojärjestelmän käyttämä kieli, termit sekä niiden käännökset. Suomen Pankissa käytetyn ruotsin kielen osalta ohjenuorana voi pitää valtioneuvoston kanslian julkaisua *Svenskt lagspråk i Finland* (SLAF), jossa annetaan suosituksia ruotsin kielestä lain ja hallinnon piirissä. Yksi SLAFin keskeisimmistä ohjeista on, että Suomessa käytetyn ruotsinkielisen lakikielen tulisi mahdollisuuksien mukaan olla yhtenevää Ruotsissa käytetyn lain ja määräysten kielen kanssa (SLAF 2017: 52). Tästä johtuen Suomen Pankin kieltä ohjaa epäsuorasti myös Ruotsin Finanssivalvonnan (*Finansinspektionen*) ja keskuspankin (*Sveriges riksbank*) käyttämä termistö ja muu sanasto. Täydellistä yhtenäisyyttä kohti ei kuitenkaan kannata pyrkiä väkisin, sillä se ei aina ole mahdollista tai tarkoituksenmukaista, sillä maiden välillä on vakiintuneita termieroavaisuuksia (Laurén 1993: 46).

### 3.3 Aineisto ja sen keruu

Tutkimusaineistona toimii Suomen Pankin vuosikertomus 2021 ja sen ruotsinkieliset konekäännökset, joita vertaan Suomen Pankin 25.3.2022 julkaisemaan ruotsinnokseen. Konekääntimen räätälöimiseen käytän Suomen Pankin vuosikertomuksia vuosilta 2019–2020. Koska koulutusmateriaalin laatu korreloi suoraan konekääntimen tuottamaan laatuun, päädyin rajaamaan koulutusmateriaalin Suomen Pankin kielipalveluiden kääntämään ja tarkastamaan aineistoon. Käyttämällä räätälöintiin ainoastaan vuosikertomuksen käännöksiä ankkuroin tutkimuksen myös käytännön työelämään, sillä kääntäjällä ei usein ole aikaa etsiä ja tarkastaa kaikkien työn kannalta hyödyllisten käännösten oikeellisuutta ja viedä niitä muistiin. Yksittäisenä laajana dokumenttina vuosikertomus on kuitenkin helppo löytää, ja koska se jo lähtökohtaisesti kannattaa viedä muistiin, saman aineiston käyttäminen konekääntimen koulutukseen ei tuo valmistautumiseen montaa työvaihetta lisää.

Suomen Pankin käännöstoiminnasta vastaa Suomen Pankin kielipalvelut, joka on Suomen Pankin hallinto-osaston yksikkö. Kielipalvelut tarjoa käännöstöitä, kielenhuoltoa ja kielikonsultaatiota Suomen Pankille ja Finanssivalvonnalle. Vuoden 2022 keväällä yksikössä työskenteli viisi kääntäjää ja kielipalvelupäällikkö. He kaikki osallistuivat vuoden 2021 vuosikertomuksen kääntämiseen omien työkieltensä osalta. Suomen Pankin vuosikertomus käännetään suomesta ruotsiksi ja englanniksi ja julkaistaan vuosittain alkukevästä, kuitenkin viimeistään huhtikuun loppuun mennessä. Tekstit

valmistuvat käännettäväksi helmikuussa ja suomen- ja ruotsinkielinen versio julkaistaan yhtä aikaa maaliskuun vaihteessa, kuitenkin viimeistään huhtikuun loppuun mennessä (laki Suomen Pankista, luku 3 § 19 mom. 2). Vuoden 2021 vuosikertomuksen ruotsinkielisen version kääntäminen alkoi helmikuussa 2022, ja lopullinen taitettu käännös julkaistiin samanaikaisesti suomenkielisen kanssa 25. maaliskuuta 2022. Kielipalveluyksikön toiminta perustuu oman työn ja ostetun palvelun yhdistämiseen, ja vuonna 2021 kielipalveluiden ruotsinnoksista Suomen Pankille ja Finanssivalvonnalle suuri osa tuotettiin ostopalveluina. Vuosikertomuksen 2021 ruotsinnoksesta vastasivat yksi pankin oma ruotsin kääntäjä, yksi sopimuskääntäjä sekä kielipalvelupäällikkö, joka myös tarkasti ja yhdenmukaisti käännöksen. Lopuksi kirjoittajat ja asiantuntijat tarkastivat käännöksen vielä sisällön ja kielen osalta.

Kielipalvelujen kääntäjät kääntävät vuosikertomuksen osissa, verkkopohjaisessa työtilassa, jossa kaikki prosessiin osallistuvat kirjoittajat, kääntäjät ja eri osastojen asiantuntijat pääsevät muokkaamaan samaa Word-dokumenttia. Kääntäjät lataavat tiedoston omalle tietokoneelleen, kääntävät tekstit Trados Studiossa ja korvaavat tiedostot työtilassa valmiilla käännöksellä. Tradosin projektinhallintatyökaluja ei käytetä prosessin aikana. Valmiit tekstit tarkastetaan työtilassa, minkä jälkeen ne merkitään valmiiksi ja pyydetään kirjoittajia ja asiantuntijoita eri osastoilta tarkastamaan käännökset kielen ja sisällön osalta. Tällä varmistetaan, että tiedoston historia ja muokkaukset ovat kaikkien tiedossa.

Kielipalveluilla on ollut käytössään käännösmuistiohjelman Trados Studio vuodesta 1999. Tällä hetkellä käytössä on ohjelman vuoden 2021 versio (ks. tarkempi kuvaus Tradosista edellä luvussa 2.1). Käännösmuistiohjelman avulla merkittävästi vuodesta toiseen samaa kaavaa noudattavan vuosikertomuksen kääntämisessä, sillä käännösmuistiin tallennetuista teksteistä kääntäjät löytävät edeltävän vuoden aikana ja edeltävissä vuosikertomuksissa käytetyt termit ja muotoilut. Kielipalveluilla on käytössään myös kieliteknologia-yhtiö Tilden Suomen EU-puheenjohtajakautta varten kehittämä Presidency Translator -konekäännin, mutta konekääntimen kehitystyön rahoitus päättyi pian puheenjohtajakauden jälkeen (Papula, Eskola & Nieminen 2021) eikä käännintä voi kouluttaa omalla aineistolla.

Suomen Pankin vuosikertomus toimii lain vaatimana tilinpäätöksenä, joka käsittää taseen, tuloslaskelman ja niiden liitetiedot sekä toimintakertomuksen (laki Suomen Pankista 1998/214 § 19). Suomen Pankin mukaan vuosikertomuksia on julkaistu vuodesta 1914 lähtien, vuoteen 1996 asti nimellä *Suomen Pankki – Vuosikirja*, ja vuodesta 1997 lähtien vuosikertomus on julkaistu myös sähköisesti. (Suomen Pankki 2022.) Vuonna 2015 pankki siirtyi julkaisemaan vuosikertomuksen

painetun kirjan sijaan pdf-muotoisena verkkodokumenttina. Vuoden 2018 vuosikertomus oli ensimmäinen nykymuotoinen verkkovuosisuosikertomus, ja se julkaistiin keväällä 2019. (Hienonen 2019.) Myös Suomen Pankki on hyödyntänyt konsultteja vuosikertomuksensa laatimisessa, sillä ensimmäisestä verkkovuosisuosikertomuksesta lähtien konsulttiyritys Crasman Oy on osallistunut vuosikertomusprosessiin mm. antamalla kirjoittajille tyyliohjeita ja toteuttamalla verkkovuosisuosikertomuksen teknisen julkaisualustan. Vuonna 2018 Crasman Oy:n kanssa yhteistyössä uusittu verkkovuosisuosikertomus (Crasman 2018) arvioitiin viestintäalan ammattijärjestön ProComin kilpailussa vuoden parhaaksi omaksi mediaksi (ProCom 2019).

Vuodesta 2019 lähtien vuosikertomus on koostunut kolmesta pääosiosta, jotka kattavat keskuspankin toimintakertomuksen, kuvauksen toiminnasta ja strategiasta sekä tilinpäätöksen. Vuosikertomukseen sisältyy myös pääjohtajan tervehdys sekä organisaatiokuvaus. Osioiden järjestys ja otsikointi on vaihdellut vuositason, mutta itse sisältö noudattaa melko yhdenmukaista kaavaa vuodesta toiseen. Taulukossa 1 kuvaan vuosikertomuksen sisällöt vuosina 2019–2021.

2019	2020	2021
<b>Pääjohtajan tervehdys</b>	<b>Pääjohtajan tervehdys</b>	<b>Pääjohtajan tervehdys</b>
<b>Toimintakertomus</b>	<b>Toiminta ja strategia</b>	<b>Toiminta ja strategia</b>
Rahapolitiikka	Suomen Pankki pähkinänuoressa	Suomen Pankki pähkinänuoressa
Rahoitusvakaus	Suomen pankin strategia uudistui vuonna 2020	Suomen Pankki toteutti uutta strategiaansa vuonna 2021
Raha ja maksaminen	Johtokunnan työnjako	Johtokunnan työnjako
Varallisuuden hallinta	Vahva asiantuntemus on vakauden valuuttaa	Suomen Pankki valittiin innostavimpien työpaikkojen joukkoon vuonna 2021
Vaikuttaminen ja yhteistyö	Suomen Pankin vuoden 2021 budjetti ja operatiiviset kustannukset	Suomen Pankille ja Finanssivalvonnalle yhteinen johtamisvisio
Vastuullisuus	Ympäristövaikutusten hallinta: korona vähensi Suomen Pankin päästöjä	Suomen Pankin vuoden 2022 budjetti ja operatiiviset kustannukset
		Suomen Pankki tarkensi hiilijalanjälkensä laskentatapaa vuonna 2021
		Uusi tietotilinpäätös täydentää Suomen Pankin vuosikertomusta
		Suomen Pankin mittaritavoitteet ja niiden toteutuminen 2021
<b>Toiminta ja strategia</b>	<b>Toimintakertomus</b>	<b>Toimintakertomus</b>
Suomen Pankki pähkinänuoressa	Rahapolitiikka	Rahapolitiikka
Suomen pankki rakentaa taloudellista vakautta	Rahoitusvakaus	Rahoitusvakaus
Johtokunnan työnjako	Raha ja maksaminen	Raha ja maksaminen
Tavoite- ja tuloskehikko	Varallisuuden hallinta	Varallisuuden hallinta
	Vaikuttaminen ja yhteistyö	Vaikuttaminen ja yhteistyö
	Vastuullisuus Suomen Pankissa	Vastuullisuus Suomen Pankissa



2019	2020	2021
Tilinpäätös	Tilinpäätös	Tilinpäätös
Tase	Tase	Tase
Tuloslaskelma	Tuloslaskelma	Tuloslaskelma
Johtokunnan esitys voitonjaosta	Johtokunnan esitys voitonjaosta	Johtokunnan esitys voitonjaosta
Tilinpäätöksen laatimisperiaatteet	Tilinpäätöksen laatimisperiaatteet	Tilinpäätöksen laatimisperiaatteet
Taseen liitetiedot	Taseen liitetiedot	Taseen liitetiedot
Tuloslaskelman liitetiedot	Tuloslaskelman liitetiedot	Tuloslaskelman liitetiedot
Tilinpäätöksen liitteet	Tilinpäätöksen liitteet	Tilinpäätöksen liitteet
Riskienhallinnan liitetiedot	Riskienhallinnan liitetiedot	Riskienhallinnan liitetiedot
Tilintarkastuskertomus	Tilintarkastuskertomus	Tilintarkastuskertomus
Lausunto tilintarkastuksesta	Lausunto tilintarkastuksesta	Lausunto tilintarkastuksesta
Organisaatio	Organisaatio	Organisaatio

Taulukko 1: Suomen Pankin vuosikertomuksen rakenne vuosina 2019–2021.

Kuten yllä olevasta taulukosta käy ilmi, vuosikertomus on vuodesta 2019 lähtien sisältänyt täysin samat pääluvut toimintakertomuksen osalta. Tilinpäätös koostuu kaikkina aineiston vuosina samoista osioista. Vuoden 2020 vuosikertomuksesta lähtien ”Toiminta ja strategia” on siirretty ensimmäiseksi luvuksi ”Pääjohtajan tervehdyksen” jälkeen. Lähdekielisen aineiston laajuus käy ilmi alla olevasta taulukosta 2.

Vuosikertomus	2019	2020	2021
Sivuja	221	234	247
Käännössivuja	165	207	191
Sanoja	28 852	34 172	31 867
Uniikkeja segmenttejä	2282	3240	2907

Taulukko 2. Tutkimusaineiston laajuus.

Taulukot 1 ja 2 havainnollistavat, että vuosikertomuksen pituus on kasvanut runsaalla 20 sivulla vuoden 2019 vuosikertomuksesta vuoden 2021 vuosikertomukseen. Sivujen määrä ei kuitenkaan ole suoraan verrannollinen tekstin määrään, vaan tarkemman arvion työmäärästä saa laskemalla sanojen tai liuskojen lukumäärän. Yksi käännössivu vastaa 1560 merkkiä välilyönteineen. Myös sanojen ja kääntäjän sivujen määrä osoittaa kasvua vuodesta 2019. Tilinpäätöksen osuus vuosikertomuksesta ei kuitenkaan ole muuttunut merkittävästi tarkasteltavien vuosien osalta, vaan se oli 68 sivua vuonna 2019, 62 sivua vuonna 2020 ja 64 sivua vuonna 2021.

Käännösmuistia ja konekääntimen räätälöintiä varten kerään vuosikertomukset 2019–2020 pdf-muotoisena Suomen Pankin verkkosivuilta osoitteesta vuosikertomus.suomenpankki.fi. Sen jälkeen siirrän tekstit Microsoft Word -tiedostoihin, jotta pystyin siivoamaan niistä pois kaiken paitsi sen leipätekstin, joka löytyi molemmilla kielillä. Esimerkiksi kaikki kuvat ja taulukot eivät olleet

kopioitavissa molemmista tiedostoista, mistä johtuu vuoden 2019 pienempi käännösyksikköjen määrä. Tämän jälkeen syötän aineiston rinnakkaisina tiedostoina Trados Studion käännösmuistiin kohdennustoiminnolla (en. *align documents*), joka muodostaa lähdetekstistä ja käännöksestä pareja, eli yhdistää tekstin osat toisiinsa siten, että otsikot, virkkeet sekä taulukkojen sisällöt vastaavat toisiaan. Kohdennettuani tekstit käyn vielä läpi, että ohjelma on yhdistänyt segmentit toisiinsa oikein. Tarkistuksen jälkeen siirrän kohdennetut segmentit niille erikseen luomiini käännösmuisteihin, kunkin vuosikertomuksen omaan tiedostoonsa. Muuntamalla muistit tmx-tiedostoiksi pystyn syöttämään ne suoraan OPUS-CATIin. Työprosessi on kuvattu alla kuviossa 3.



Kuvio 3: Tutkielmassa käytetty konekääntimen koulutuksen aineiston keruun työprosessi.

Jotta esimerkiksi vuosikertomuksessa toistuvat taulukkojen solut eivät aiheuttaisi tilastollista vinoumaa automaattiseen arviointiin, jätän lopulliseen konekäännettävään lähtötekstiin ainoastaan *uniikkeja segmenttejä*.

### 3.4 OPUS-CAT-neuroverkkokääntimen räätälöinti

Käytän tutkimuksessa käännösmuistiohjelma Trados Studioon lisäosana asennettavaa OPUS-CAT-konekäännintä. OPUS-CAT on Tommi Niemisen kehittämä ei-kaupallinen neuroverkkokäännin, jonka pystyy integroimaan Trados Studioon ja muutamaaan muuhun käännösohjelmaan. Nieminen on kehittänyt OPUS-CATin Helsingin ja Turun yliopistojen sekä Kites ry:n:n Fiskmö-projektin (sv. *finsk-svensk-maskinöversättning*, Tiedemann ym. 2020) pohjalta. (Nieminen 2021: 288.)

OPUS-CATin yleiskielinen konekäännin perustuu geneerisille, eli yleiskielisille, kaksikielisille käännösmalleille (en. *OPUS-MT models*), jotka on esikoulutettu Microsoftin Translator-työryhmän kehittämällä Marian-neuroverkkokääntimellä (Nieminen 2021: 288). Marian on vuonna 2018 esitelty avoimen lähdekoodin neuroverkkokäännin (Junczys-Dowmunt ym. 2018). OPUS-CATin geneeriset käännösmallit on koulutettu aineistolla, joka on saatavilla OPUS-tietokannasta (Tiedemann & Thottingal 2020: 1). OPUS on avoin korpustietokanta, eli kokoelma lähtötekstejä ja niiden käännöksiä. OPUS-hankkeen perimmäinen tarkoitus on koota verkossa vapaasti käytettävissä olevia kaksikielisiä tekstejä helposti käytettävään muotoon, etenkin pienten kielten osalta. (Tiedemann 2012: 2214.)

OPUS-CATin käyttö aloitetaan lataamalla ohjelma osoitteesta <https://helsinki-nlp.github.io/OPUS-CAT/install> ja asentamalla se tietokoneelle. Tämän jälkeen OPUS-CATin Trados Studioon sopiva lisäosa (en. *plugin*) ladataan *RWS App Storesta*. Räätälöinti aloitetaan lataamalla OPUS-MT-tietokannasta kohdekieliparin geneerinen kaksikielinen konekäännösmalli klikkaamalla *Install OPUS model from Web* -painiketta ja valitsemalla kieliparin uusin malli. Koska malli asennetaan tietokoneelle paikallisesti, sen käyttöön asennuksen jälkeen ei tarvita verkkoyhteyttä (Nieminen 2021: 288). Tämän jälkeen varsinainen räätälöinti (ohjelmassa nimellä *fine-tuning*) tietyn erikoisalan tai tekstityypin teksteillä tapahtuu syöttämällä ohjelmaan kaksikielistä erikoisalan aineistoa tmx-tiedostona tai lähtötekstinä ja sen käännöksenä kahdessa erillisessä tiedostossa. Erillisten tiedostojen tulee vastata toisiaan rivien lukumäärän osalta, jotta OPUS-CAT tunnistaa mikä segmentti lähtötekstissä vastaa mitään segmenttiä kohdekielisessä dokumentissa. Etenkin vähemmän harjaantuneen käyttäjän on helpointa käyttää tmx-muotoista käännösmuistia.

Räätälöintiprosessin aikana konekäännin *validoi* koulutusprosessin kulkua ja näyttää tuloksen integroidulla SacreBLEU-mittarilla (lisää eri BLEU-mittareista luvussa 3.5). Validointi tarkoittaa, että OPUS-CAT konekääntää koulutusprosessin aikana säännöllisesti sille syötetyn validointitiedoston ja vertaa käännöstään ihmisen käännökseen samasta validointiaineistosta SacreBLEU-mittarilla. Valitsemalla räätälöidyn mallin päävalikossa ja klikkaamalla *Show fine-tuning progress* ohjelma näyttää

uudessa välilehdessä, kuinka lähellä koneen validointiaineistosta (en. *in-domain*) tekemät käännökset ovat ihmisen samasta lähtötekstistä tekemiä käännöksiä. (Nieminen 2021: 292.) Samassa välilehdessä SacreBLEU-mittari näyttää myös, miten räätälöintiaineiston syöttäminen vaikuttaa koneen yleiskieliseen (en. *out-of-domain*) käännöstaitoon. Yleiskieliseen validointiin OPUS-CAT käyttää useimpien kieliparien osalta omia validointiaineistoja (mp.), niin myös tässä tutkielmassa.

Käyttäjä voi valita, haluaako käyttää erikoiskieliseen validointiin omaa aineistoa vai antaa OPUS-CATin poimia aineistosta validointiin käytettävät segmentit automaattisesti. Jos käyttäjä valitsee erilliset tiedostot, hänen tulee syöttää ohjelmaan räätälöintiaineiston tavoin validointitiedostot, eli lähtötekstin ja sen käännöksen kahdessa riveittäin toisiaan vastaavaa tiedostossa. Aineiston tulee mielellään olla samalta erikoisalalta, mutta validointiaineisto ja koulutusaineisto eivät saa sisältää samoja käännöksiä, sillä koekääntämällä samoja segmenttejä kuin koulutusaineistossa, kone ei opi soveltamaan oppimaansa (ks. 2.2.4). Siksi on helpointa käyttää myös tässä oletuksena tarjottua automaattista validointiaineiston poimintaa.



Kuvio 4: OPUS-CATin räätälöimisen työvaiheet.

Automaattisten valintojen sijaan harjaantunut käyttäjä voi kokeilla käyttää validointiin omia validointitiedostoja (suosituksena n. 100–300 segmenttiä), mikä voi parantaa erikoiskielistä käännöstulosta (Balashov 2021). OPUS-CATissa on myös muita hienosäätöä mahdollistavia toimintoja, mutta tässä tutkimuksessa pitäydyn oletusasetuksissa.

### 3.5 BLEU ja SacreBLEU

BLEU on konekäännösten mittaukseen kehitetty työkalu, joka mittaa käännöksen tarkkuutta (en. *precision*, kaaviossa P) vertaamalla sitä ihmisen tuottamaan käännökseen asteikolla nollasta yhteen, missä nolla on huonoin ja yksi on paras. Tulos ilmaistaan usein asteikolla nollasta sataan. Yli 70 pisteen

tulokset viittaavat ylisovittamiseen (en. *overfitting*), eli siihen, että konekäännintä on koulutettu täysin samalla aineistolla, jota sillä käännetään, jolloin kone ei ole aidosti oppinut soveltamaan kieltä, vaan toistaa asioita ikään kuin ulkomuistista (ks. luku 2.2.4). Googlen AutoML Cloud Translationin (Google 2022) ohjeet luokittelevat BLEU-pisteet seuraavasti:

BLEU-pisteet	Selite
< 10	Lähes hyödytön
10–19	Vaikea hahmottaa merkitystä
20–29	Merkitys välittyy, mutta käännös sisältää paljon kielioppivirheitä
30–40	Käännös on ymmärrettävä tai jopa hyvä
40–50	Hyvälaatuinen käännös
50–60	Erittäin hyvälaatuinen, hyväksyttävä ja sujuva käännös
> 60	Usein ihmiskääntäjää parempi laatu

Taulukko 3: Googlen pääsääntö BLEU-pisteille (Google 2022, käännös Appelberg).

Alkuperäisessä tarkkuutta mittaavassa kaaviossa lasketaan ensin konekäännöksessä esiintyvät kokonaiset sanat (unigrammit), jotka esiintyvät myös ihmisen tekemässä referenssikäännöksessä. Tämä sanamäärä jaetaan konekäännöksen kokonaissanamäärällä, josta saadaan käännöksen tarkkuus. BLEU:ssa tätä kaavaa (kaava 1) on muokattu siten, että pelkkien sanojen lisäksi mitataan sanajonoja (n-grammeja, 1–4 sanan jonoja) (Papineni ym. 2002: 313).

#### Konekäännöksen tarkkuus $p_n$

$$= \frac{\text{konekäännöksessä esiintyvät referenssikäännöksen sanajonot}}{\text{referenssikäännöksen sanajonojen lukumäärä}}$$

Kaava 1: BLEU-mittarin modifioitu precision-kaavio (Papineni ym. 2002: 313).

Lopuksi kaavioon lisätään lyhyyssakko (en. *brevity penalty*, *BP*, kaava 2), joka alentaa pisteitä, jos käännökset ovat lyhyempiä kuin referenssikäännökset, jotta lyhyet käännökset eivät näin saisi suhteellisesti parempia pisteitä. Lyhyyssakkoa ei kuitenkaan lasketa lausekohtaisesti, vaan koko konekäännetyt tekstin laajuudelta keskimäärin, mikä mahdollistaa joustoa käännöksen pituudessa segmenttitasolla. Lyhyyssakon kaava on yksi miinus referenssikäännöksen pituus  $[r]$  jaettuna konekäännöksen pituudella  $[c]$ .

$$\text{Lyhyyssakko BP} = 1 - \frac{\text{referenssikäännöksen pituus (r)}}{\text{konekäännöksen pituus (c)}} = \begin{cases} 1 & \text{jos } |c| > |r| \\ e^{(1-\frac{|r|}{|c|})} & \text{jos } |c| \leq |r| \end{cases}$$

Kaava 2: Lyhyyssakko (en. *brevity penalty*, *BP*) (Papineni ym. 2002: 315).

Suoraviivaisesti selitettynä BLEU-pisteet lasketaan seuraavasti: lyhyysakko × (konekäännöksessä esiintyvät referenssikäännöksen sanajonot) / (referenssikäännöksen sanajonojen lukumäärä) (kaava 3).

$$\begin{aligned} \text{BLEU} &= \text{Lyhyysakko} * \text{Konekäännöksen tarkkuus} \\ &= 1 - \frac{\text{referenssikäännöksen pituus (r)}}{\text{konekäännöksen pituus (c)}} \\ &\quad * \frac{\text{konekäännöksessä esiintyvät referenssikäännöksen sanajonot}}{\text{referenssikäännöksen sanajonojen lukumäärä}} \end{aligned}$$

Kaava 3: BLEU-kaavio kokonaisuudessaan (Papineni ym. 2002:315).

Tämän kaavan pohjalta konekäännöksen tulee siis mahdollisimman tarkasti vastata referenssikäännöstä pituudeltaan (liian lyhyt käännös laskee pisteitä), sanavalinnoiltaan (oikeat sanat nostavat pisteitä) ja sanajärjestykseltään (mitä enemmän samoja 1–4 sanan jonoja, sitä paremmat pisteet). BLEU ei kuitenkaan ole hyödyllinen yksittäisten lauseiden tai edes lyhyiden tekstien käännöslaadun arvioimisessa, vaan vaatii käytännössä aina yksittäistä käännöstä laajemman tekstin tuottaakseen luotettavan arvion (Papineni ym. 2002: 311) eikä BLEU-pisteiden vertaaminen ole hedelmällistä, mikäli referenssikäännöksiä on eri määrä, sillä suurempi referenssikäännösten lukumäärä nostaa BLEU-pisteitä (Mas. 315). Esimerkiksi Tradosin omistava kieliteknologyyhtiö RWS suosittelee käytettäväksi BLEU-arviointiin vähintään tuhatta lauseparia (Vashee 2019).

BLEUta on kritisoitu myös sen yhteismitattomuudesta, sillä tulokset eivät verrannollisia keskenään. Korjausliikkeenä Matt Post (2018: 186–187) on kehittänyt SacreBLEU:n, joka on osin paranneltu ja yhteismitallisempi mittaustapa. Se on sisäänrakennettuna myös OPUS-CATIin (validoinnin arviointi). Koska perinteinen BLEU on alun perin suunniteltu kehitystyökaluksi ja tutkimukset osoittavat sen korreloivan hyvin tai melko hyvin ihmisen laatukäsityksen kanssa, se toimii kuitenkin hyvin konekäännösjärjestelmien kehityksen arvioinnissa (Post 2018: 1; Reiter 2018: 399–400). Sen sijaan sitä ei tulisi käyttää yksittäisten käännösten arviointiin tai ainoana arviointimittarina tieteellisessä tutkimuksessa, sillä kuten Reiter (mp.) osoittaa, BLEU ei välttämättä tarjoa johdonmukaisia tuloksia, vaan tulosten ja ihmisarvioijien käsityksissä on ollut suurtakin hajontaa, esimerkiksi vuosien 2007–2016 WMT-konekäännöskonferenssien vertailuissa. Reiter esittää myös, että BLEU ei välttämättä mittaa luotettavasti aitoja käännöstilanteita, eikä kaikkia sen puutteita ole vielä selvitetty esimerkiksi eri konekäännösteknologioiden mittauksessa. (Mp.) Siksi käytän BLEUta vain suuntaa antavana konekääntimen kehityksen mittarina ja arvioin käännöksiä myös käsin seuraavassa alaluvussa esittämiäni parametrien pohjalta.

### 3.6 Konekäännösten manuaalinen arviointi

Suurin kritiikki BLEUta kohtaan liittyy siihen, ettei puuttuvista sanoista saa sakkoa, eikä lyhyysakko riitä kattamaan puutetta (Banerjee & Lavie 2005: 67). Toinen suuri heikkous on, että BLEU-pisteytys on periaatteessa riippumaton sanajärjestyksestä eikä ota huomioon synonyymejä tai parafraseja (Cer ym. 2010: 556). Saadakseni suuntaa antavalle BLEU-pisteytykselle vahvistuksen, käyn parhaan BLEU-pisteytyksen saaneen konekäännöksen läpi myös käsin käyttämällä taulukon 4 virhematriisia. Virhematriisi on aineistolähtöinen sovellettu versio Irina Temnikovan esittelemästä konekäännösten virheanalyysistä (en. *Cognitive MT Error Ranking*, Temnikova 2010: 3487). Temnikovan malli sisältää alla olevien luokittelujen lisäksi välimerkkivirheet mutta OPUS-CAT suoriutuu niin hyvin välimerkeistä, että se ei toisi tutkimukseen lisäarvoa. Temnikovan malli sisälsi myös erillisinä lausekkeiden ja sanojen järjestyksen muutokset, jotka huomion omassa luokittelussani yhtenä luokkana, Taulukossa mainittujen virheiden ja niiden lukumäärän lisäksi merkitsen, käyttäisinkö konekäännöstä jälkieditoinnin pohjana (kyllä/en) sekä jos käyttäisin, miten paljon muokkausta se vaatisi (vähän/paljon).

Jaan virheet kolmeen pääluokkaan ja yhteensä seitsemään alaluokkaan. Pääluokat ovat morfologiset virheet, muotovirheet ja lausetason virheet. Tarkempi jaottelu alla olevassa taulukossa 4:

	Morfologiset virheet	Leksikaaliset virheet					Syntaktiset virheet
Virheluokka	Oikea sana väärässä muodossa	Väärä synonyymi	Väärä sana	Ylimääräinen sana	Puuttuva sana	Idiomivirhe	Sanajärjestysvirhe

Taulukko 4: Konekäännösten manuaalisen arvioinnin virhematriisi (Appelberg 2021).

Vuoden 2021 vuosikertomusta ei kuitenkaan ole tarpeen kääntää kokonaisuudessaan konekääntimellä, sillä osa vuosikertomuksesta noudattaa aiempien vuosien muotoiluja sanatarkasti ja nämä osat löytyvät käänösmuistista. Tämän tutkielman tarkoituksena on tutkia, mitä hyötyä konekääntimestä on käänösmuistin apuna, ja siksi vertailen virheluokittelun avulla ainoastaan niitä osia vuosikertomuksen käännöksestä, joille ei löydy *identtistä tai osittaista osumaa* (ks. tarkemmin luvussa 4.2) käänösmuistista, sillä jo kertaalleen hyväksi havaittu ja julkaistu käänös on parempi lähtökohta jälkieditoinnille kuin uusi konekäännös (Tezcan, Bulté, Vanroy 2019: 1).

## 4 Analyysi

Tässä luvussa käyn läpi analyysin eri vaiheet. Pääluvussa käsittelen dokumenttien valmistelun, eli lopullisen lähtötekstin ja referenssikäännöksen muodostamisen. Alaluvussa 4.1 kuvaan räätälöintiprosessia ja sen automaattista arviointia OPUS-CATin sisäänrakennetulla SacreBLEU-työkalulla sekä räätälöimäni konekääntimen tuottaman käännöksen ja kielipalveluiden ruotsinnoksen vertailua kieliteknologiayhtiö Tilden tarjoamalla BLEU-työkalulla. Alaluvussa 4.2. käyn vielä läpi parhaan BLEU-pisteytyksen saaneen käännöksen manuaalisesti edellä luvussa 3.6 esittelemäni konekäännösten virheluokittelun pohjalta. Virheluokittelun lähtökohta on arvioida, kuinka pienellä vaivalla konekäännöksen saa jälkieditoitua julkaisukelpoiseksi käännökseksi.

Automaattisen arvioinnin aluksi kokoan vuoden 2021 suomenkielisen vuosikertomuksen ja kielipalveluiden tuottaman ruotsinnoksen siitä kahteen toisiaan vastaavaan Word-tiedostoon käsin, sillä Suomen Pankin verkkosivuilta ladatun pdf-version suora muuntaminen Word-muotoon aiheuttaa jonkin verran häiriöitä, keskeltä katkenneita lauseita ja asettelun ongelmia. Myöskään käsin koottujen tiedostojen suora vertailu ei olisi hedelmällistä, sillä automaattinen arviointi onnistuu ainoastaan, jos kahdessa erillisessä tiedostossa on täysin sama määrä lähtötekstejä ja niiden käännöksiä omilla riveillään. Tuottaakseni kaksi riveittäin vertailtavaa tekstitiedostoa kohdistan käsin kokoamani suomenkielisen vuoden 2021 vuosikertomuksen ja pankin kielipalveluiden tuottaman käännöksen siitä Tradosissa ja tallennan kohdennetut segmentit uuteen tyhjiin tmx-muotoiseen käännösmuistiin.

Suomenkielisen vuoden 2021 vuosikertomuksen ja kielipalveluiden ruotsinnoksen kohdennus tuotti 3448 kohdennettua segmenttiä (en. *translation unit*). Uuteen käännösmuistiin näistä tallentui 2907 uniikkia segmenttiä. Uniikki viittaa tässä siihen, että esimerkiksi taulukoissa esiintyvät lyhyet selitteet, sekä sisällysluettelossa että tekstin lomassa esiintyvät otsikot tai muut toistuvat ilmaisut eivät tallennu muistiin moninkertaisesti, sillä yksi käännös riittää. Tämän jälkeen kopioin käännösmuistiin tallennetut käännösyksiköt kahteen txt-muotoiseen tekstitiedostoon: ensimmäiseen pelkät lähtökieliset tekstit, toiseen pelkät kohdekieliset tekstit. Lopputuloksena on kaksi tekstitiedostoa, jotka koostuvat omilla riveillään olevista uniikeista segmenteistä. Nämä tekstitiedostot muodostavat konekäännösten automaattiseen arviointiin käyttämäni *lähtötekstin ja referenssikäännöksen*.

Manuaalisen analyysin suoritan ainoastaan niistä vuosikertomuksen osista, joille ei löydy valmiita käännöksiä vuosien 2019 ja 2020 vuosikertomuksista, sillä kerran käännettyä ja hyväksi havaittua käännöstä ei kannata enää konekääntää. Karsin pois aiemmin käännettyt osat Trados Studion esikäännoistoiminnolla (ks. tarkemmin luvussa 5.2).



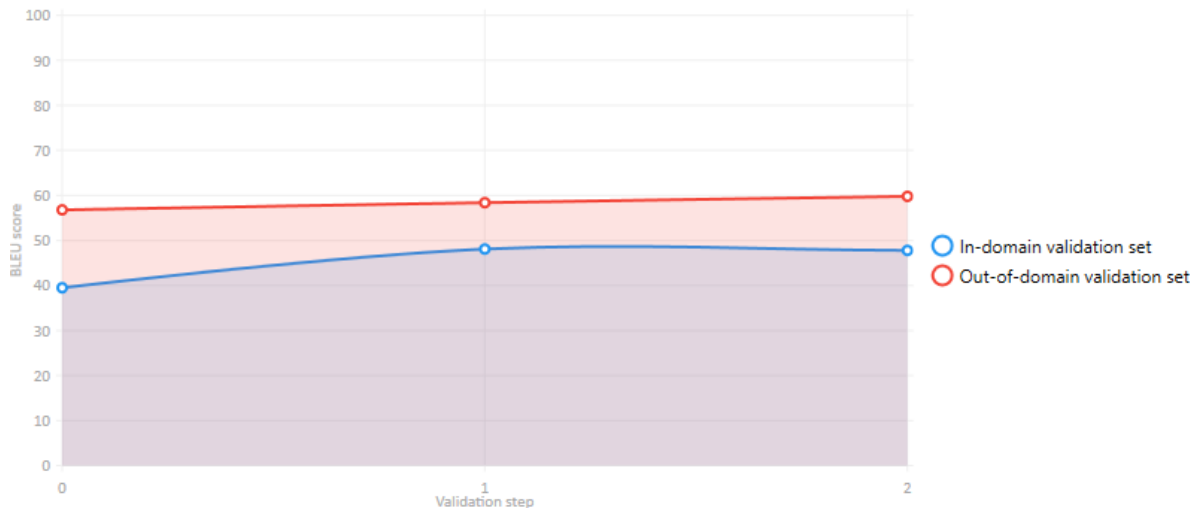
## 4.1 Käännösmallien ja konekäännösten BLEU-pisteytys

Käytän tutkimuksessa kolmea eri OPUS-CAT-mallia: ensimmäinen malli on räätälöimätön geneerinen malli, joka löytyy OPUS-MT models -arkistosta nimellä opus-2021-02-19 (uusin malli, haettu 1.5.2022). Alla käytän geneerisestä mallista lyhennettä OC1. Toinen malli on vuosikertomuksella 2020 räätälöity malli, joka pohjautuu mallille OC1; tästä mallista käytän lyhennettä OC2. Tämän jälkeen räätälöin seuraavan mallin vuoden 2019 vuosikertomuksella käyttämällä lähtökohtana mallia OC2; tähän malliin viittaan lyhenteellä OC3. Mallit ja niiden lyhenteet luettelona:

- OC1 = kouluttamaton malli (opus-2021-02-19)
- OC2 = vuosikertomuksella 2020 koulutettu malli
- OC3 = vuosikertomuksilla 2020 ja 2019 koulutettu malli

### 4.1.1 Ensimmäinen räätälöintivaihe

Seuraavaksi kuvaan mallin OC2 räätälöintiä. Käytän räätälöintiin vuoden 2020 vuosikertomuksesta tmx-tiedostoon vietyä 3240 käännösyksikköä. OPUS-CATin räätälöinnin edistymistä voi seurata ohjelman sisäänrakennetulla SacreBLEU-työkalulla. Räätälöintiprosessi 3240:llä segmentillä kestää noin 15 minuuttia ja sen aikana validointia voi seurata *fine-tuning process* -välilehdeltä. Mallin OC2 räätälöinnin edistyminen näkyy oheisessa kuviossa 5. Punainen käyrä osoittaa, millainen vaikutus räätälöinnillä on mallin yleiskielisen validointiaineiston (*out-of-domain validation set*) kääntämiseen ja sininen käyrä osoittaa vaikutusta erikoiskielisen validointiaineiston (*in-domain validation set*) kääntämiseen. Validointi ennen räätälöintiä (vaaka-akselilla 0) antaa yleiskielen osalta SacreBLEU-pisteet 56,8 ja erikoiskielen osalta pisteet 39,5. Räätälöinnin edetessä pisteet nousevat ja ensimmäinen räätälöity validointi (vaaka-akselin kohta 1) antaa yleiskieliselle käännökselle tuloksen 58,4 ja erikoiskielen kääntämiselle tuloksen 48,1. Ohjelma jatkaa räätälöintiä, ja toisen validoinnin kohdalla (vaaka-akselin kohta 2) tulokset ovat yleiskielen osalta 59,8 ja erikoiskielen osalta 47,8.



Kuvio 5: Mallin OC2 räätälöintiprosessin graafinen kuvaus.

Ensimmäisen räätälöintiprosessin kuvaus osoittaa, että mallin erikoiskielen osaaminen ei parane enää ensimmäisen validointikierron jälkeen, eli se ei enää lähene referenssikäännöstä. Tämä viittaisi siihen, että malli on oppinut aineistosta kaiken, mitä se voi siitä oppia.

Seuraavaksi käytän generistä OC1-mallia ja sen pohjalta räätälöityä OC2-mallia tutkimukseni lähtötekstin kääntämiseen. Lähtöteksti koostuu 2907:stä vuoden 2021 vuosikertomuksesta poimitusta omalle rivilleen sijoitetusta käännösyksiköstä. Tuotan konekäännökset lähtötekstistä asettamalla Trados Studion käyttämään OPUS-CATia tekstien automaattiseen kääntämiseen ilman käännösmuisteja. Tallennan käännökset nimillä **OC1\_SV** ja **OC2\_SV**. Sen jälkeen syötän käännökset ja kielipalveluiden tuottaman referenssikäännöksen Tilden tarjoamaan BLEU-sovellukseen osoitteessa <https://www.letsmt.eu/Bleu.aspx>. Ensimmäisen BLEU-mittauksen tulokset näkyvät taulukossa 5.

	<b>OC1_SV</b>	<b>OC2_SV</b>
<b>BLEU</b>	32,65	39,36

Taulukko 5: Generisen OC1-mallin ja ensimmäisen räätälöidyn OC2-mallin käännösten BLEU-pisteet.

Kuten taulukosta 5 näkyy, jo räätälöimätön versio saa melko hyvät BLEU-pisteet vuoden 2021 vuosikertomuksen käännöksestä, 32,65. Googlen pääsäännöt BLEU-pisteiden arviointiin asettavat sekä generisen mallin tuottaman käännöksen **OC1\_SV** että ensimmäisen räätälöidyn mallin tuottaman käännöksen **OC2\_SV** (BLEU 39,36) luokkaan ”ymmärrettävä tai hyvä” (kaikki arviointiluokat tarkemmin luvussa 4.3). OPUS-CATin SacreBLEU-työkalun arvio ei täysin vastannut Tilden BLEU-työkalun tulosta, mutta niillä mitataankin eri asioita: SacreBLEU mittaa tässä oppimista, kun taas BLEU mittaa, kuinka

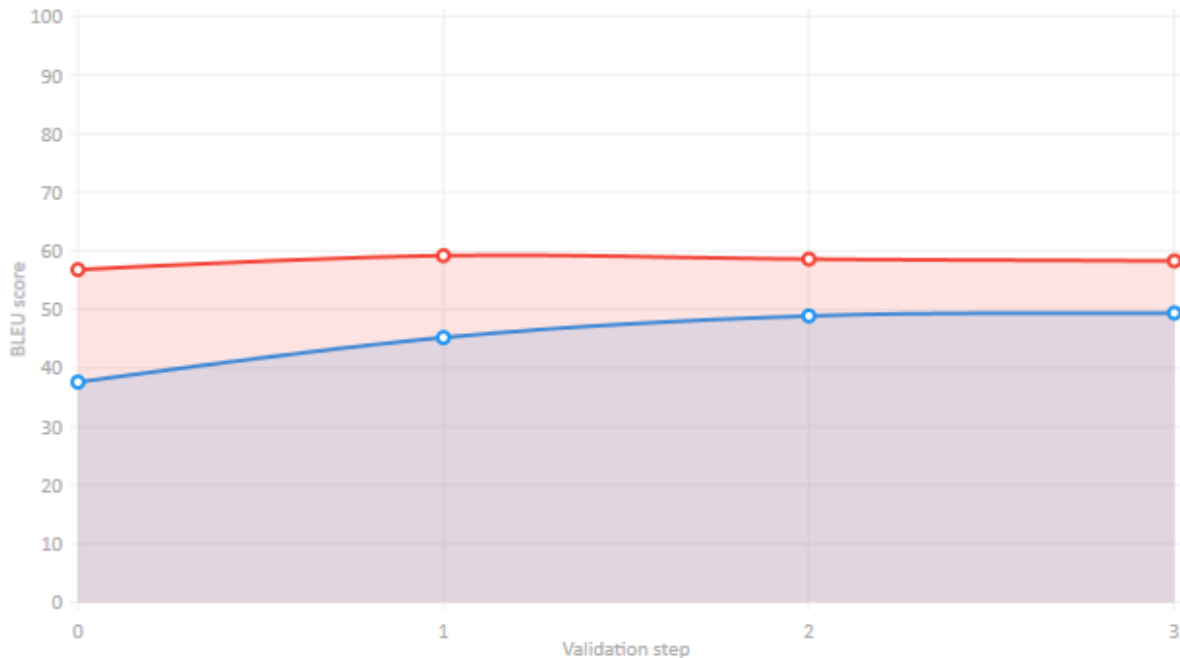
lähellä malleilla OC1 ja OC2 tuotettu konekäännös on ammattikäntäjien käännöstä samasta lähtötekstistä.

Yksittäisinä esimerkkeinä ensimmäisen räätälöintikierroksen tuloksista voi mainita, että OC2-malli ei vielä oppinut, että ainoa oikea ruotsinkielinen vastine substantiivilausekkeelle *Suomen Pankin pääjohtaja on Finlands Banks chefdirektör*, vaan tarjoaa OC1-mallin tavoin käännöstä *Finlands Banks generaldirektör*. Malli OC1 ei tuntenut substantiivia *chefdirektör* lainkaan, joten OC2 oppi substantiivin *chefdirektör*, mutta tämä oikea vastine esiintyy konekäännöksessä vain viisi kertaa, kun taas *generaldirektör* esiintyy seitsemän kertaa vastikkeena lausekkeelle *Suomen Pankin pääjohtaja* ja seitsemän kertaa muussa yhteydessä. Tämä tarkoittaa, että räätälöinti ei korvannut mallin OC1 osaamaa sanaa kokonaan, vaan sanavarasto laajeni. Esimerkkinä onnistuneesta oppimisesta voi sen sijaan mainita, että malli oppi, että *koronapandemia* on ruotsiksi *coronapandemi* eikä *koronapandemin*.

#### 4.1.2 Toinen räätälöintivaihe

Seuraavaksi kuvaan mallin OC3 räätälöintiä. Käytän räätälöinnin pohjana mallia OC2 ja räätälöin sitä vuoden 2019 vuosikertomuksesta poimituilla käännösyksiköillä. Vuoden 2019 vuosikertomus sisälsi 2282 uniikkia käännösyksikköä, joista 529 ovat yhteneviä vuoden 2020 vuosikertomuksen kanssa. Siten vuoden 2019 vuosikertomuksen kohdistaminen ja muuntaminen tmx-tiedostoksi ei tuottanut kuin 1753 uutta segmenttiä räätälöinnin jatkamiseksi. Näin ollen räätälöintiin käytettävä aineisto sisältää yhteensä 4993 uniikkia käännösyksikköä.

Mallin OC3 räätälöinnin kehitys näkyy oheisessa kuviossa 6. Punainen käyrä merkitsee jälleen yleiskielisen validointiaineiston kehitystä ja sininen käyrä erikoiskielisen validointiaineiston kehitystä. Validointi ennen räätälöintiä (vaaka-akselin kohta 0) antaa yleiskieliselle validoinnille SacreBLEU-pisteet 56,8 ja erikoiskielen validoinnille pisteet 37,6. Erikoiskielen validoinnin eri aloituspisteet johtuvat OPUS-CATin automaattisen validointiaineiston valinnasta, eli se ei käytä samaa erikoiskielen validointiaineistoa kuin ensimmäisessä koulutuksessa. Räätälöinnin edetessä pisteet nousevat ja ensimmäinen räätälöity validointi (vaaka-akselin kohta 1) saa yleiskielen osalta tuloksen 59,2 ja erikoiskielen osalta tuloksen 45,2. Ohjelma jatkaa räätälöintiä, ja toisen validoinnin kohdalla (vaaka-akselin kohta 2) tulokset ovat yleiskielen osalta 58,6 ja erikoiskielen osalta 48,9. Kolmannen validoinnin kohdalla (vaaka-akselin 3) yleiskielen osaaminen on laskenut 58,3 pisteeseen ja erikoiskielen validointi noussut 49,4 pisteeseen. Räätälöintiprosessi kestää noin 10 minuuttia.



Kuvio 6: Mallin OC3 räätälöintiprosessin graafinen kuvaus.

Toisen räätälöintiprosessin kuvaus osoittaa, että mallin erikoiskielen osaaminen ei parane enää toisen validointikierroksen jälkeen ja lisäksi mallin yleiskielisen validointiaineiston pisteet vähenevät. Tämä viittaa siihen, että räätälöinnin jatkaminen samankaltaisella aineistolla alkaa heikentää konekääntimen yleiskielen osaamista, vaikkakin vain vähän. Se ei ole toivottavaa, sillä se voi pahimmillaan johtaa yleiskielisten sanojen ja ilmaisuun unohtamiseen (ks. luku 2.2.4). Jatkoräätälöinnissä validointiaskelia on kolme, mikä johtuu suuremmasta aineistosta – OPUS-CATilla menee pidempään vertailla materiaalia.

Seuraavaksi käytän OC2-mallin pohjalta räätälöityä OC3-mallia tutkimukseni lähtötekstin kääntämiseen. Tuotan konekäännöksen samalla tavalla kuin OC2-mallilla asettamalla Trados Studion käyttämään OPUS-CATia tekstien automaattiseen kääntämiseen. Tallennan käännöksen nimellä **OC3\_SV**. Sen jälkeen syötän jälleen käännöksen ja referenssikäännöksen Tilden BLEU-sovellukseen. BLEU-mittauksen tulokset näkyvät taulukossa 6, johon jätin vertailun vuoksi myös OC2-mallilla tuotetun käännöksen tuloksen. Taulukosta näkee, että jatkoräätälöinti ei enää juurikaan parantanut koneen oppimistulosta.

	<b>OC2_SV</b>	<b>OC3_SV</b>
<b>BLEU</b>	39,36	40,68

Taulukko 6: Räätälöidyillä OC2- ja OC3-malleilla tuotettujen käännösten BLEU-pisteet.

Googlen suuntaa antavan ohjeen (ks. luku 4.3) mukaan 40–50 BLEU-pistettä viittaa hyvälaatuiseen käännökseen. Testaan tämän hypoteesin seuraavassa alaluvussa. Kuten räätälöintiprosessin kuvauksesta kuvioista 6 näkyy, oppimistulos ei kuitenkaan enää parantunut toisen validointiaskeleen jälkeen, vaan tämän jälkeen validointi osoitti, että yleiskielisen validointiaineiston pisteytys vähenee eikä erikoiskielinen enää parane. Arvioni on, että uusi aineisto on liian homogeenistä suhteessa aiempaan räätälöintiaineistoon, jotta kone voisi enää oppia siitä uutta. Myös lopullinen BLEU-pisteytys viittaa samaan, sillä mallin OC3 konekäännös parani enää yhden BLEU-pisteen verran verrattuna mallilla OC2 tuotettuun käännökseen.

	<b>OC1_SV</b>	<b>OC2_SV</b>	<b>OC3_SV</b>
<b>BLEU</b>	32,65	39,36	40,68

Taulukko 7: Konekäännösten BLEU-pisteet.

Yllä olevasta taulukosta 7 näkyy, kuinka räätälöinti vaikutti konekäännösten BLEU-pisteisiin. Ensimmäisen räätälöinnin tulos on huomattava parannus, lähes 7 pistettä. Jatkoräätälöinti samanlaisella aineistolla ei kuitenkaan enää nostanut pisteitä merkittävästi, vain 1,32 pisteen verran. Koska BLEU on vain suuntaa antava mittari, käyn seuraavassa luvussa läpi, minkä laatuinen parhaat BLEU-pisteet saanut käännös todella on.

#### 4.2 Konekäännöksen arviointi virheluokittelun pohjalta

Tässä alaluvussa analysoin OPUS-CATin mallin OC3 tekemiä käännösvirheitä mallilla tuotetussa käännöksessä vuoden 2021 vuosikertomuksen niistä osista, joille ei löydy vastaavuutta vuosikertomukset 2019 ja 2020 sisältävistä muisteista. Valmistelen analysoitavan aineiston rajaamalla lähtötekstistä ensin pois ne lauseet (segmentit), joille käännösmuisti löytää vastaavuuden vuosien 2019 ja 2020 vuosikertomuksista, sillä konekäännintä ei yleensä kannata käyttää, jos muistista löytyy jo käännös. Rajaus tapahtuu syöttämällä 2907 segmenttiä sisältävä lähtöteksti Trados Studio -käännösmuistiohjelmaan ja käyttämällä esikäntämistoimintoa (en. *pre-translate*). Esikäntämisen kynnsarvo on oletuksena 70 %, mutta sen voi asettaa myös alemmas. Suoritan rajauksen eri kynnsarvoilla, nähdäkseni kuinka monta segmenttiä sijoittuu mihinkin vastaavuusluokkaan. Lopuksi käyn aineiston läpi käsin poimiakseni käännettävästä materiaalista segmentit, jotka eivät syystä tai toisesta löydy korkeammalla kynnsarvolla, vaikka niiden sisältö vastaisi lähtötekstiä. Alla olevasta taulukosta 8 näkyy, kuinka monta esikäntettyä segmenttiä Trados Studio tarjoaa kullakin esikäntätoiminnon kynnsarvolla.

Kynnysarvo	Muistiosumia
99 %	912
90 %	1120
80 %	1217
70 %	1326
60 %	1474
30 %	2330

Taulukko 8: Esikäntämisen (*pre-translate*-toiminnon) tuottamat osumat muistista.

Esikäntäminen 99 %:n kynnysarvolla, eli lähes täydellisillä vastaavuuksilla, tuottaa 912 valmista käännöstä 2907:stä segmentistä. 90 %:n kynnysarvolla Trados Studio kääntää automaattisesti 1120 segmenttiä. Yli 90 %:n vastaavuus tuottaa tässä tapauksessa todella lähellä olevia osumia. 80 %:n kynnysarvolla esikäntösten lukumäärä nousee 1217:ään, mutta tällä raja-arvolla useat muistista haetut osumat sisältävät jo puutteita. 70 %:n kynnysarvolla esikäntösten lukumäärä nousee 1326:een, mutta vastaavasti muistiosumien vastaavuuden taso heikkenee. 60 %:n kynnysarvolla osumat nousevat 1474:ään. Viimeiseksi teen esikäntöksen myös 30 %:n kynnysarvolla. Tämän seurauksena esikäntettyjen segmenttien lukumäärä nousee 2330:een, mutta erittäin suuri osa käännöksistä on enää vain suurpiirteisiä tai ne sisältävät jonkin yksittäisen saman sanan kuin lähtökielen segmentti. Koska käännösmuisti ei välttämättä tunnista esimerkiksi kahta erillistä lausetta, jotka on tallennettu muistiin yhtenä virkkeenä, tai uutta sanajärjestystä, vaikka sisältö olisi sama, käyn läpi segmentit myös käsin. 2907:stä jää 1564 segmenttiä, runsaat puolet lähtötekstistä, joille löytyy niin tyydyttävä osuma käännösmuistista, että niiden kääntämiseen ei kannata käyttää konekäännintä, sillä hyvä käännösmuistin sisältämä segmentti on usein nopeampi jälkieditoida kuin konekäännös (Tezcan ym. 2019: 1). Alla olevassa taulukossa 7 on esimerkit eri kynnysarvoilla tuotetuista esikäntetystä segmenteistä, sekä konkordanssihaulla löydetyistä osittaisista osumista. Punainen merkintä ja yliviivaus merkitsee aiemmin tallennetun segmentin sisältöä, sininen taas uutta käännettävänä olevaa segmenttiä.

Kynnysarvo	Lähtöteksti	Muistiosuma
99 %	Likviditeettilyijäämä kasvoi entistäkin suuremmaksi	Likviditetsöverskottet växte ytterligare
90 %	Suomen Pankin rahoitusvarallisuuden <del>kerke</del> sijoituksia ohjaa pankin johtokunnassa vuosittain päätettävä sijoituspolitiikka.	Finlands Banks investeringar i räntebärande värdepapper styrs av en investeringspolicy som årligen fastställs av bankens direktion.
80 %	Suomen Pankin rahoitusvarallisuuden <del>tuotte</del> tuotot kasvoivat	Avkastning på Finlands Banks finansiella tillgångar
70 %	Kevyt rahapolitiikka <del>auttaa</del> auttoi pitämään pankkiluottojen tarjonnan runsaana ja <del>helpottaa</del> helpotti rahoituksen saantia- <del>kaikilla talouden sektoreilla.</del>	En expansiv penningpolitik bidrar till ett rikligt utbud av bankkrediter och underlättar tillgången till finansiering inom alla sektorer av ekonomin.

Kynnysarvo	Lähtöteksti	Muistiosuma
60 %	<del>Rahoitussektori palveluissa</del> <del>hyödynnetään</del> Rahoitussektori hyödyntää laajasti digitalisaation luomia mahdollisuuksia.	I de finansiella tjänsterna utnyttjas i stor utsträckning de möjligheter som digitaliseringen erbjuder.
30 %	Kiintiön koko ( <del>pysyi koko vuoden</del> kuusi kertaa vähimmäisvarantovelvoite) 2020 <del>vakiona, pysyi koko kuusi kertaa</del> <del>vähimmäisvarantovelvoite</del> vuoden 2021 vakiona.	Kvoten hade samma storlek under hela året och var sex gånger kassakravet.
Konkordanssihaku	Eurokäteisen väarennöstilanne vuonna 2021	Förekomsten av euroförfalskningar [Ø]

Taulukko 9: Esimerkkejä käännösmuistin osumista eri kynnysarvoilla.

Kuten taulukosta 9 näkyy, 99 %:n osuma tarkoittaa, että täysin vastaava teksti on ollut edeltävän vuoden vuosikertomuksessa ja muistista löytyy sille suora vastine. 90 %:n osumassa vain yksi sana on muuttunut, etuliite *korko-* on tippunut pois, ja muuten teksti on täysin sama kuin aiemmin muistiin tallennetussa segmentissä. 80 %:n osumassa pitää lisätä loppuun verbi *kasvoivat*, mutta tämäkin on helposti lisättävissä. 70 %:n osumassa on vaihtunut kahden verbin aikamuoto ja lausetta on lyhennetty loppupäästä, ja lauseen hännän voi poistaa myös ruotsinkielisestä käännöksestä. 60 %:n osumassa sanamuoto on vaihdettu passiivista aktiiviin, mutta muuten lause on sama. 30 %:n osumassa koko lauseen sanajärjestys on muutettu, mutta sisältö ei ole vaihtunut merkittävästi, lauseke *kuusi kertaa vähimmäisvarantovelvoite* on vain irrotettu omaksi parenteesiksi. Hakemalla käännösmuistista tutun näköisiä lauseen osia voi löytää osumia, kuten alimmassa esimerkissä, jossa otsikon perään on suomeksi lisätty *vuonna 2021*. Muisti ei löydä sitä automaattisesti, sillä puolet sanoista puuttuu, mutta kääntäjä voi hakea sen muistista manuaalisesti.

Poistettuani segmentit, joille löydän vastineet vuoden 2019 ja 2020 vuosikertomuksen sisältävästä käännösmuistista, siirrän taulukkolaskentaohjelmaan 1343 suomenkielistä segmenttiä, joille en löydä riittävää vastaavuutta muistista. Vertailussa tarkastelen suomenkielistä lähtötekstiä, kielipalveluiden tuottamaa käännoästä ja OPUS-CATin mallilla OC3 tuotettua käännoästä OC3\_SV. Seuraavaksi käyn järjestelmällisesti läpi konekäännökset ja merkitsen niissä havaitsemani virheet luvussa 3. esittelemäni virhematriisin perusteella. Lukumääräinen analyysi näkyy taulukossa 9. Päädyin käyttämään aineistosta otantaa, sillä suuntaa antaviin tuloksiin riittää myös pienempi aineisto. Myös toisella luokittelulla suorittamani laskenta antoi saman suuntaisia tuloksia.

	Kielivirheet	Leksikaaliset virheet					Syntaktiset virheet
Virheluokka	Oikea sana mutta väärä muoto	Väärä synonyymi	Väärä sana	Ylimääräinen sana	Puuttuva sana	Idiomivirhe	Sanajärjestysvirhe
Esiintymiä	58	150	60	22	50	28	45
Yhteensä	413						

Taulukko 10: Virheanalyysin tulos.

Konekäännösten vertailu lähtötekstiin ja kielipalveluiden tuottamaan ruotsinnokseen osoittaa, että OPUS-CATilla ja vuosikertomuksilla räätälöidyllä OC3-käännösmallilla tuotetun konekäännöksen segmentit ovat erittäin laadukkaita. Ensimmäinen virheluokka on konekääntimen valitsemat **oikeat sanat väärässä muodossa** eli kielivirheet, 58 kappaletta. Näitä oli hämmästyttävän vähän, mutta kuten johdannossa lainasin Koehnia, neuroverkkokäännökset ovat usein sujuvan näköisiä mutta merkitykseltään petollista.

Lähtöteksti	Kielipalvelujen käännös	OC3_SV
(1) Väestön ikääntyminen on lisännyt <i>säästämistä</i> , likviditeetin kysyntää ja laskenut korkotasoa.	(1) Befolkningens åldrande har ökat <i>sparandet</i> och efterfrågan på likviditet och fått räntenivån att sjunka.	(1) Den åldrande befolkningen har ökat <i>sparande</i> , efterfrågan på likviditet och räntenivån.
(2) Keväällä julkaistiin seitsemän uratarinaa, ja kesätyöntekijät kertoivat kesätyöarjestaan yhdeksässä kesätyöblogissa.	(2) På våren publicerade bankens sju karriärberättelser och sommarpraktikanterna skrev sju blogginslag om sin vardag på centralbanken.	(2) Under våren publicerades sju <i>karriärhistoria</i> och sommararbetarna berättade om sin sommararbetstid på nio sommararbetsbloggar.
(3) Kansallisarkisto päättää, <i>mikä</i> keskuspankin aineistosta on pysyvästi säilytettävää.	(3) Riksarkivet bestämmer <i>vilka</i> delar av centralbankens arkiv som ska förvaras varaktigt.	(3) Riksarkivet beslutar om <i>vilket</i> av centralbankens material som ska förvaras permanent.

Toinen virheluokka on **synonyymivirheet**, 150 kappaletta. Synonyymivirhe, tai termivirhe, tarkoittaa, että konekäännin on käyttänyt väärää sanaa asiayhteyteen nähden ja tuottaa siksi käännöksen, jonka merkitys on eri kuin lähtötekstissä. Väärä synonyymi voi johtaa siihen, että käännös menee aivan väärin, mutta se voi olla helposti jälkiedoitavissa. Esimerkiksi Suomen Pankin käyttämä ruotsinkielinen vastine substantiiville *hiilijalanjälki* on *klimatavtryck*, vaikka *koldioxidavtryck* saattaa olla jopa yleisempi nimitys. Koska Suomen Pankki on päättänyt käyttämään käännöstä *klimatavtryck*, tämä lasketaan termivirheeksi. Näiden osalta kääntäjän tulee olla erityisen huolellinen, sillä käännös



voi näyttää täydelliseltä, mutta koska konekäännin ei välttämättä käytä termejä johdonmukaisesti, termi voi kääntyä saman virkkeenkin sisällä usealla eri tavalla. Termivirheen havaitseminen edellyttää kääntäjältä huolellisuutta, mutta havaittu virhe on helppo jälkieditoida. Ensimmäisessä esimerkissä kone on kääntänyt väärin substantiivin *työpaikka* sanalla *työ, jobb*. Toisessa esimerkissä kone on kääntänyt väärin termit *strateginen tavoite* ja *tiedolla johtaminen*. Kolmannessa virhe johtuu jo valmiiksi suomen kielen muodosta, sillä tavanomaisesti vaatimuksia *asetetaan, ei luoda*. Konekäännin ei kuitenkaan osaa ottaa tällaista huomioon, ja kääntää tekstin epäsovivalla vastineella *skapa*.

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) Suomen Pankki on Suomen innostavimpien <i>työpaikkojen</i> joukossa	(1) Finlands Bank bland de mest inspirerande <i>arbetsplatserna</i> i Finland	(1) Finlands Bank är bland de mest inspirerande <i>jobben</i> i Finland
(2) Tiedonhallinnan kehittämisohjelmat tukevat Suomen Pankin strategista tavoitetta tiedolla johtamisesta	(2) Utvecklingsprogrammen för informationshanteringstödjer Finlands Banks <i>strategiska prioritering kunskapsbaserad ledning</i>	(2) Utvecklingsprogram för informationshanteringstöder Finlands Banks <i>strategiska mål</i> med <i>information om ledarskap</i>
(3) Keskuspankin tiedonkäsittely <i>luo</i> suuret vaatimukset tietoturvalle	(3) Informationshanteringen vid centralbanken <i>ställer</i> höga krav på datasäkerheten	(3) Centralbankens informationsbehandling <i>skapar</i> stora krav på informationssäkerhet

Kolmas virheluokka on **väärät sana**, 60 kappaletta. Virheluokittelu käsittää lähtökielen sanan korvaamisen jollakin, mikä tarkoittaa jotakin aivan muuta, konekääntimen keksimiä sanoja (*uusmuodoste*) (ks. termin määritelmä luku 3.2). Alla esimerkkejä konekääntimen tuottamista vääristä sanoista. Toisessa esimerkissä OPUS-CAT on kääntänyt väärin substantiivin *pankkilaiset*, muodostamalla siitä keksityn sanan, *bankinvånare*, eli *pankkiasukkaat*. Kolmannessa esimerkissä on kyse konekääntimen itse keksimästä vastineesta sanalle *suojadigitointi*.

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) EKP julkaisi laajan joukon strategia-arvioon liittyviä artikkeleita <i>syyskuun jälkipuoliskolla</i> (englanniksi).	(1) ECB gav ut en stor mängd artiklar om strategiöversynen under <i>senare delen av september</i> (på engelska).	(1) ECB publicerade ett brett spektrum av artiklar om strategiöversynen <i>under andra halvåret</i> (på engelska).
(2) Suomen Pankki lisäsi vuoden 2021 aikana työnantajakuvaviestintäänsä erityisesti sosiaalisessa mediassa,	(2) Finlands Bank ökade kommunikationen om sin arbetsgivarprofil under 2021 framför allt i sociala medier, där	(2) Under 2021 tillade Finlands Bank sin arbetsgivarbildskommunikation, framför allt i sociala medier, där

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
jossa pankki viesti säännöllisesti avoimista työpaikoista ja <i>pankkilaiset</i> kertoivat uratarinoissaan työskentelystä keskuspankissa.	banken regelbundet informerade om lediga tjänster och <i>medarbetarna</i> bidrog med sina karriärberättelser om arbetet på en centralbank.	banken regelbundet kommunicerade om lediga jobb och <i>bankinvånarna</i> berättade om sina karriärberättelser om arbetet på centralbanken.
(3) Osittain digitointi kohdistuu historiallista mielenkiintoa sisältäviin kokonaisuuksiin, kuten pankkivaltuusmiesten pöytäkirjoihin, osa on puolestaan tuhoutumisvaarassa olevan aineiston <i>suojadigitointia</i> .	(3) Digitaliseringens fokus ligger i viss mån på helheter med historiskt värde, såsom bankfullmäktiges protokoll, men omfattar också <i>konsivering av handlingar</i> som löper risk att förstöras.	(3) Delvis är digitaliseringen inriktad på helheter av historiskt intresse, såsom bankfullmäktiges protokoll, medan en del är <i>skyddsdigiering</i> av material som riskerar att förstöras.

Neljäs virheluokka on **lisätty sana, 22 kpl**. Lisäykset olivat aineistossa harvinaisia ja rajanveto lisäyksen ja **väärän sanan** välillä on häilyvä. Esimerkissä 1 konekäännin on lisännyt määreen *svag* ennen *heikennystä, försämring*, mikä johtaa kaksinkertaiseen negaatioon. Esimerkeissä 2 on ratkaissut puutteen kielitaidossaan kääntämällä kaksi samantapaista mutta eri sanaa samalla vastineella. Esimerkissä 3 konekäännin on toistanut sanan *amorteringsfria*, eikä ole korvannut sillä mitään, vaan luonut sisältöä tyhjästä (en. *hallucinate*, Koehn & Knowles 2017: 30).

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) Edelliseen tutkimukseen verrattuna tuloksissa näkyi positiivista kehitystä eikä <i>juurikaan</i> heikentymistä.	(1) Resultatet visade en positiv utveckling och knappt <i>någon försämring</i> alls jämfört med föregående enkät.	(1) Jämfört med den föregående undersökningen visade resultaten en positiv utveckling och ingen <i>svag</i> försämring.
(2) Suomen Pankki on edustettuna myös Opettajien ja <i>opettajankouluttajien</i> talouskoulutus -hankkeen ohjausryhmässä.	(2) Finlands Bank medverkar också i styrgruppen för projektet Opettajien ja opettajankouluttajien talouskoulutus ("Ekonomisk utbildning för lärare och <i>lärarutbildare</i> ").	(2) Finlands Bank är också representerad i styrgruppen för projektet Utbildning för lärare och <i>lärare</i> .
(3) Suomen Pankki seurasi vuonna 2021 taloyhtiölainojen kehitystä ja kannatti ehdotuksia, joissa taloyhtiölainojen enimmäisluotto-osuus rajataan 60 prosenttiin ja uudiskohteiden taloyhtiölainojen lyhennysvapaista luovutaan.	(3) Finlands Bank följde 2021 utvecklingen av bostadsbolagslånen och understödde förslagen om att begränsa bostadsbolagslånen maximala belåningsandel till 60 % och slopa <i>amorteringsfria</i> perioder i bostadsbolagslån för nybyggen.	(3) Finlands Bank följde upp utvecklingen av bostadsbolagslånen 2021 och stödde förslagen om att begränsa den maximala kreditandelen för bostadsbolagslån till 60% och slopa <i>amorteringsfria</i> lån för bostadsbolagslån till nya objekt.

Viides virheluokka on **puuttuva sana, 50 kpl**. Näissä konekäännin on ohittanut jonkin sanan. Yksi selitys on, että sanaa edeltää liian pitkä määrätekijä, eikä kääntimen huomio riitä lauseen alkupäästä loppupäähän (*vanishing gradient*, ks. 2.2.2). Kone on myös voinut oppia, että jokin sana sisältää jo riittävän määrätteen, kuten esimerkissä 1 alla, jossa kone todennäköisesti kuvittelee, että *stimulanstaktik* käsittää myös adjektiivin *elvyttävä*. *Stimulanstaktik* ei kuitenkaan tarkoita yhtään mitään, vaan se on konekääntimen kehittämä uusmuodoste, ja *elvyttävä* jää siksi pois käännöksestä. Esimerkissä 2 konekäännin on ohittanut sanan *ennakoitu*. Oppiminen ei ole lineaarista, joten syy voi olla mikä hyvänsä. Jälkieditoinnin kannalta käänös on helppo korjata, mutta kiireessä sana, jonka sijainti lauseessa on ruotsiksi eri kuin suomeksi, on helppo unohtaa. Esimerkissä 3 sana *aikaistetusti* on hävinnyt, kenties koska etumääräite on raskas, sillä konekäännöksen lauserakenteessa ilmaus *i förtid* sijoittuisi viimeiseksi, eli melko kauas verbirakenteesta *maksaa takaisin/återbetala*.

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) Näiden tietojen perusteella EKP:n neuvosto ilmoitti pitävänsä rahapolitiikan <i>elvyttävän</i> virityksen ennallaan kesäkuussa 2021.	(1) Mot bakgrund av dessa uppgifter meddelade ECB-rådet att det håller den <i>expansiva</i> penningpolitiska hållningen oförändrad i juni 2021.	(1) På basis av dessa uppgifter meddelade ECB-rådet att det i juni 2021 kommer att hålla den [Ø] penningpolitiska stimulanstaktiken oförändrad.
(2) Syksyllä 2021 <i>ennakoitu</i> talouskasvu vahvistui	(2) Hösten 2021 stärktes den ekonomiska tillväxten <i>som väntat</i>	(2) Hösten 2021 förstärktes den ekonomiska tillväxten [Ø]
(3) Pankeilla oli myös mahdollisuus maksaa TLTRO III -operaatioista hakemaansa rahoitusta takaisin <i>aikaistetusti</i> syyskuussa ja joulukuussa 2021.	(3) Bankerna hade också en möjlighet till <i>förtida</i> återbetalning av tilldelningen från TLTRO III-refinansieringstransaktionerna i september och december.	(3) Bankerna hade också möjlighet att i september och december 2021 betala tillbaka den finansieringsom de ansökt om inom ramen för TLTRO III-transaktionerna [Ø].

Kuudes virheluokka on **idiomivirheet, 28 kpl**. Näitä ovat väärinymmäretyt idiomaattiset ilmaisut. Näitä oli aineistossa melko vähän, johtuen todennäköisesti aineiston asiategstipainotteisuudesta. Alla esimerkkejä idiomiluokan virheistä. Olen laskenut näihin myös virheet, joissa konekäännin on tulkinnut jonkin ilmauksen kirjaimellisesti, kun kyseessä on abstraktimpi käsite. Esimerkissä yksi konekäännin on tulkinnut *maksamisen kentän* kirjaimellisesti *maksukentäksi, betalningsfält*. Esimerkissä 2 konekäännin ei ole ymmärtänyt substantiivia *rima* lainkaan, vaan korvannut sen sanalla *riktlinje*, joka vastaa substantiivia *ohje* tai *suuntaviiva*. Esimerkissä 3 ongelmaksi muodostuu epäsuora ilmaus työnyhteisön jäsenistä, *pankkilaiset*. Tätä konekäännin ei ymmärrä, vaan korvaa sen sanalla *pankkitoiminta, bankverksamhet*.

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) Maksamisen kenttä muuttuu nopeasti.	(1) Förändringarna på betalningsområdet går snabbt.	(1) Betalningsfältet förändras snabbt.
(2) Talouspolitiikan rima pysyykin korkealla.	(2) Ribban för den ekonomiska politiken ligger därför alltså högt.	(2) Riktlinjen för den ekonomiska politiken fortsätter att vara hög.
(3) Pankkilaisuus kestää keskimäärin 15,7 vuotta.	(3) Den genomsnittliga anställningstiden är 15,7 år.	(3) Bankverksamheten varar i genomsnitt 15,7 år

Seitsemäs virheluokka on **virheet sanajärjestyksessä, 35 kpl**, eli lausetason virheet. Tämän tyyppin konekäännökset saattavat sisältää kaikki oikeat sanat, mutta jonkin ymmärrysvirheen takia sanat ovat väärässä järjestyksessä. Koska konekäännin muutenkin noudattaa melko vahvasti suomen lauserakennetta, se tuottaa helposti luonnollisia käännöksiä segmenteistä, joiden sanajärjestys ei poikkea suuresti suomen konventioista.

Lähtöteksti	Kielipalvelujen käänös	OC3_SV
(1) Suomen Pankin toiminnan keskiössä oli vuonna 2021 viisi painopistettä:	(1) Finlands Bank antog fem prioriteringar för sin verksamhet 2021:	(1) Finlands Banks verksamhet stod i centrum för fem prioriteringar under 2021:  <i>Fem prioriteringar stod i centrum för Finlands Banks verksamhet under 2021:</i>
(2) Pandemian puhjettua siinä tilanteessa päätetyt raha- ja muun talouspolitiikan toimet onnistuivat hyvin.	(2) De penningpolitiska och ekonomisk-politiska åtgärder som vidtogs med anledning av pandemins utbrott hade önskad verkan.	(2) Efter pandemins utbrott lyckades de penningpolitiska åtgärderna och den övriga ekonomiska politiken väl.  <i>De penningpolitiska åtgärderna och den övriga ekonomiska politiken lyckades väl efter pandemins utbrott.</i>
(3) Koronatilanteen vakiinnuttua ja laajan etätyösuosituksen päätyttyä marraskuun alussa Suomen Pankissa valmistauduttiin hybridityöhön siirtymiseen.	(3) Sedan coronaläget hade stabiliserats och rekommendationen om omfattande distansarbete upphört i början av november gjorde Finlands Bank förberedelser för en omställning till hybridarbete.	(3) Efter det att coronasituationen hade stabiliserats och den omfattande rekommendationen om distansarbete hade avslutats i början av november förbereddes Finlands Bank för övergången till hybridarbete.

Näiden lisäksi arvioin, kuinka moni tarkastelluista segmenteistä toimisi hyvänä lähtökohtana jälkieditoinnille ja kuinka paljon niitä tarvitsisi muokata. 37 (21 %) segmenttiä ei tarvinnut lainkaan korjausta, 101 (57 %) olisivat kevyitä korjata, 35 (20 %) olisi tarvinnut merkittävää korjausta ja 3 (2 %) olivat täysin hyödyttömiä. Temnikovan luokittelu asettaa korjausluokat jälkieditoinnin kannalta helpoimmasta vaikeaan, eli oikea sana väärässä muodossa (luokka 1) on helpoin korjata, kun taas sanajärjestyksen muokkaaminen vaatii suurimman kognitiivisen ponnistuksen (Temnikova 2010: 3488).

Yllä mainittujen virheiden lisäksi konekäännökset noudattivat suurilta osin suomenkielistä passiivirakennetta, mikä näkyy erona ihmiskääntäjien tekemiin referenssikäännöksiin. Subjektin lisääminen ilman kontekstia olisi tietenkin mahdotonta, tai ainakin riskialtista, sillä konekäännin kääntää segmentin kerrallaan eikä ymmärrä kuin lauseen- tai segmentinsisäisiä viittaussuhteita.

## 5 Päätelmät ja kokoavaa tarkastelua

Tässä luvussa pohdin edeltävässä luvussa 4 suorittamani analyysin tuloksia ja esitän vastaukset tutkimuskysymyksiini ja hypoteesini paikkansapitävyyteen. Käyn myös läpi tulosten ja tutkimusmenetelmäni puutteita ja ansioita sekä OPUS-CATin mahdollista soveltamista Suomen Pankin kielipalveluiden valikoimaan käännöstyökaluja. Palatakseni alussa esittämiini tutkimuskysymyksiin voin todeta, että BLEU-arvioinnin perusteella räätälöinti lisäsi merkittävästi konekäännöksen ja Suomen Pankin kielipalveluiden keskinäistä vastaavuutta. Manuaalisen arvioinnin perusteella konekäännin suoriutui jopa yllättävän hyvin niistä vuosikertomuksen kohdista, joita ei käsitelty kahden aiemman vuoden vuosikertomuksissa. Konekääntimen räätälöinti oli varsin yksinkertaista ja suoraviivaista.

BLEU-arviointi antaa vain arvion laadun kehityksestä, mutta koska BLEU edellä luvussa 3.5 esitetyn mukaisesti toimii hyvänä mittarina konekääntimen kehitystyössä, voidaan todeta, että 8,03 pisteen tulos noin 15 + 10 minuuttia kestäneellä käännösmuistilla räätälöinnillä on tehokas tulos. Prosessin suoraviivaisuus osoitti myös, että räätälöinti ei vaadi suurta teknistä osaamista, vaan harjaantuneen Trados Studion käyttäjän voi kuvitella suoriutuvan siitä ohjeita seuraamalla.

Aineiston manuaalinen läpikäynti ei myöskään ole aukoton metodi, sillä jo pelkkä eri virheluokkien päällekkäisyys ja niiden suhteellinen subjektiivisuus ja yhteismitattomuus vähentää niiden vertailtavuutta. Käännösvirheiden mittaaminen on toisaalta aina subjektiivista, ellei kyseessä ole täysin ilmiselvä väärä termi tai puute. Silloinkin on kuitenkin mahdollista, että vika on lähtöisin epäselvästä lähtötekstistä, eikä ihminen välttämättä suoriutuisi kääntämisestä konetta paremmin. Tutkimuksen

perusteella on kuitenkin selvää, että konekäännin voi tuottaa julkaisukelpoisia tai lähes julkaisukelpoisia käännöksiä tilanteessa, joissa käännösmuisti ei tuota lainkaan tuloksia.

Jatkon kannalta olisi mielekästä jatkaa mallin kouluttamista muilla pankin julkaisuilla, jotta siitä tulisi yleishyödyllinen työväline. Pienellä aineistolla kouluttamisessa on se riski, että kone oppii vain koulutusaineiston, joten en tiedä miten OC3-malli selviytyisi muista Suomen Pankin tekstien kääntämisestä. Vuosikertomuksen laajan aihe kattavuuden takia arvelen sen kuitenkin tuottavan vähintäänkin tyydyttävän lähtökohdan.

## **Kiitokset**

Suomen Pankille työnantajana kiitos työn mahdollistamisesta ja kiinnostuksesta. Toivon, että tutkielmastani on iloa sekä kollegoilleni että työnantajalleni muutenkin. Suurkiitos myös ohjauksesta, neuvoista ja kannustuksesta ohjaajalleni!

# Lähdeluettelo

## Tutkimusaineisto

Finlands Bank (2019). *Årsberättelse 2019*. Noudettu 25.2.2022 osoitteesta <https://arsberattelse.finlandsbank.fi/2019/hem/>

Finlands Bank (2020). *Årsberättelse 2020*. Noudettu 25.2.2022 osoitteesta <https://arsberattelse.finlandsbank.fi/2020/hem/>

Finlands Bank (2021). *Årsberättelse 2021*. Noudettu 25.3.2022 osoitteesta <https://arsberattelse.finlandsbank.fi/2021/hem/>

Suomen Pankki (2019). *Vuosikertomus 2019*. Noudettu 25.2.2022 osoitteesta <https://vuosikertomus.suomenpankki.fi/2019/etusivu/>

Suomen Pankki (2020). *Vuosikertomus 2020*. Noudettu 25.2.2022 osoitteesta <https://vuosikertomus.suomenpankki.fi/2020/etusivu/>

Suomen Pankki (2021). *Vuosikertomus 2021*. Noudettu 25.3.2022 osoitteesta <https://vuosikertomus.suomenpankki.fi/2021/etusivu/>

## Kirjallisuus

Antinjuntti K. (2018). Kaksikielisessä Suomessa pulaa ruotsin taitajista. *Kääntäjä – Översättaren* 2018(3), s. 5. Suomen kääntäjien ja tulkkien liitto SKTL. Noudettu 25.2.2022 osoitteesta [https://www.sktl.fi/@Bin/1773072/ka%CC%88a%CC%88nta%CC%88ja%CC%88\\_3\\_18.pdf](https://www.sktl.fi/@Bin/1773072/ka%CC%88a%CC%88nta%CC%88ja%CC%88_3_18.pdf)

Balashov, Y. (2021). OPUS-CAT: A State-of-the-Art Neural Machine Translation Engine on Your Local Computer. *The ATA Chronicle – American Translators Association*. Noudettu 22.1.2022 osoitteesta <https://www.ata-chronicle.online/highlights/opus-cat-a-state-of-the-art-neural-machine-translation-engine-on-your-local-computer/>

Banerjee, S. & Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments, *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Michigan*, s. 65-72. Noudettu 16.1.2022 osoitteesta <https://aclanthology.org/W05-0909.pdf>

- Bahdanau, D., Cho, K. & Bengio, Y. (2014). *Neural Machine Translation by Jointly Learning to Align and Translate*. Noudettu 6.4.2022 osoitteesta  
[https://www.researchgate.net/publication/265252627\\_Neural\\_Machine\\_Translation\\_by\\_Jointly\\_Learning\\_to\\_Align\\_and\\_Translate](https://www.researchgate.net/publication/265252627_Neural_Machine_Translation_by_Jointly_Learning_to_Align_and_Translate)
- Bisazza A. & Federico M. (2016). A Survey of Word Reordering in Statistical Machine Translation: Computational Models and Language Phenomena. *Computational Linguistics 2016*, 42 (2), s. 163–205. [https://doi.org/10.1162/COLI\\_a\\_00245](https://doi.org/10.1162/COLI_a_00245)
- Bowker, L. (2005). Productivity vs Quality? A pilot study on the impact of translation memory systems. *Localisation Focus*, 4 (1), s. 13–20. Noudettu 15.4.2022 osoitteesta  
<https://www.semanticscholar.org/paper/Productivity-vs-Quality-A-pilot-study-on-the-impact-Bowker/c749e045fc4cce2a689904748a525f8e1d9d87b8>
- Cer, D., Manning, C. & Jurafsky, D. (2010). The Best Lexical Metric for Phrase-Based Statistical MT System Optimization. *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics*. Noudettu 20.1.2022 osoitteesta  
[https://nlp.stanford.edu/pubs/best\\_lexical\\_metric\\_statmt.pdf](https://nlp.stanford.edu/pubs/best_lexical_metric_statmt.pdf)
- Christensen, T. & Schjoldager A., (2017). Translation-Memory (TM) Research: What Do We Know and How Do We Know It?, *HERMES - Journal of Language and Communication in Business*, 23(44), s. 89–101. Noudettu 24.1.2022 osoitteesta  
[https://www.researchgate.net/publication/320588606\\_Translation-Memory\\_TM\\_Research\\_What\\_Do\\_We\\_Know\\_and\\_How\\_Do\\_We\\_Know\\_It](https://www.researchgate.net/publication/320588606_Translation-Memory_TM_Research_What_Do_We_Know_and_How_Do_We_Know_It)
- Crasman (2018). *Yksisivuinen sisältökonsepti kannatteli Suomen Pankin vaativaa vuosikertomusuudistusta*. Noudettu 27.4.2022 osoitteesta  
<https://www.crasman.fi/blogi/suomen-pankin-vuosikertomus>
- Dorr, B., Snover, M. & Madnani, N. (2009). Part 5: Machine translation evaluation. Teoksessa B. Dorr (toim.), *DARPA GALE program report*. Noudettu 20.1.2022 osoitteesta  
<http://www.cs.cmu.edu/~alavie/papers/GALE-book-Ch5.pdf>
- Delingua Oy, 2019. *Vuosikertomus käännöstekstityyppinä: Millainen on hyvä vuosikertomus (kääntää)?* Delingua blogi 4.1.2019. Noudettu 22.1.2022 osoitteesta  
<https://www.delingua.fi/kehitys/millainen-on-hyva-vuosikertomus-kaantaa/>
- Garcia, I. (2007). Power-shifts in web-based translation memory. *Machine Translation*, 21, s. 55–68. Noudettu 24.1.2022 osoitteesta <https://www.jstor.org/stable/30219110>



- Garcia, I. (2009). Beyond Translation Memory: Computers and the Professional Translator. *The Journal of Specialised Translation*, 12, s. 199–214. Noudettu 15.4.2022 osoitteesta [https://www.jostrans.org/issue12/art\\_garcia.php](https://www.jostrans.org/issue12/art_garcia.php)
- Garcia, I. (2014). Computer-Aided Translation. Teoksessa C. Sin-wai (toim.), *The Routledge Encyclopedia of Translation Technology* (s. 68–87). <https://www.routledgehandbooks.com/doi/10.4324/9781315749129.ch3>
- Google Cloud Translation (2022). *Evaluating models > Understanding the BLEU Score > Interpretation*. Noudettu 6.5.2022 osoitteesta <https://cloud.google.com/translate/automl/docs/evaluate>
- Gorschelnik, H., Ahlblad, L., Kulmala, T. & Kavén, J. (2021, 30. syyskuuta). Ammattikäntäjä tai -tulkkii saa viestin perille – kieli- ja käänntieteilistä ei voida enää lyhytnäköisesti säästää. *Mielipidekirjoitus Aamulehdessä YK:n kansainvälisenä käntäjienpäivänä*. Noudettu 21.1.2022 osoitteesta <https://www.aamulehti.fi/lukijalta/art-2000008296879.html>
- Hallintolaki* 434/2003. Annettu Helsingissä 6 päivänä kesäkuuta 2003. Noudettu 10.4.2022 osoitteesta <https://www.finlex.fi/fi/laki/alkup/2003/20030434>
- Haque, R., Liu, C. & Way, A. (2021). Recent advances of low-resources neural machine translation. *Machine Translation* 35, s. 451–474. <https://doi.org/10.1007/s10590-021-09281-1>
- Hassan, H., Aue A., Chen C., Chowdhary, V., Clark J., Federmann C., Huang X., Junczys-Dowmunt M., Lewis W., Li M., Liu S., Liu T.-Y., Luo R., Menezes, A., Qin, T., Seide, F., Tan, X., Tian, F., Wu, L., Wu, S., Xia, Y., Zhang, D., Zhang, Z., & Zhou, M. (2018). *Microsoft AI & Research*. Noudettu 22.1.2022 osoitteesta <https://www.microsoft.com/en-us/research/uploads/prod/2018/03/final-achieving-human.pdf>
- Hienonen, M. (2019, 25. toukokuuta) *Asiaa taloudesta modernilla twistillä*. Suomen Pankin blogi. Noudettu 24.3.2022 osoitteesta <https://www.eurojatalous.fi/fi/blogit/2019/asiaa-taloudesta-modernilla-twistilla/>
- Hutchins, W.J. (1995). Machine Translation: A Brief History. Teoksessa E.F.K Koerner & R.E Asher (toim.), *Concise History of the Language Sciences: From the Sumerians to the Cognitivists* (431–445). <https://doi.org/10.1016/B978-0-08-042580-1.50066-0>
- Hutchins, W.J. (2014). Machine translation: History of Research and Applications. Teoksessa C. Sin-wai (toim.), *The Routledge Encyclopedia of Translation Technology*. Noudettu 27.3.2022 osoitteesta

<http://dspace.vnbrims.org:13000/xmlui/bitstream/handle/123456789/4402/Routledge%20Encyclopedia%20of%20Translation%20Technology.pdf>

ISO 18587:2017(en) *Translation services — Post-editing of machine translation output —*

*Requirements*. Noudettu 25.3.2022 osoitteesta

<https://www.iso.org/obp/ui/#iso:std:iso:18587:ed-1:v1:en>

Jia Y., Carl, M. & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, 31, s. 60–84. Noudettu 23.1.2022 osoitteesta

[https://www.jostrans.org/issue31/art\\_jia.pdf](https://www.jostrans.org/issue31/art_jia.pdf)

[https://www.jostrans.org/issue31/art\\_jia.pdf](https://www.jostrans.org/issue31/art_jia.pdf)

Juholin, E. (2017). *Communicare! Viestinnän tekijän käsikirja*. MIF Management Institute of Finland.

Juutinen J., (2020). *Konekäännin käytännön käännoistyössä: kääntäjien asenteet ja konekäännösten hyödyntäminen* [Maisterintutkielma, Helsingin yliopisto]. Noudettu 10.4.2022 osoitteesta

<https://helda.helsinki.fi/handle/10138/316762>

Kalchbrenner N. & Blunsom P., (2013). Recurrent Continuous Translation Models. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*, s. 1700–1709. Noudettu 7.4.2022 osoitteesta

<https://aclanthology.org/D13-1176.pdf>

<https://aclanthology.org/D13-1176.pdf>

*Kielilaki* 2003/423. Annettu Helsingissä 6.6.2003. Noudettu 10.4.2022 osoitteesta

<https://www.finlex.fi/fi/laki/ajantasa/2003/20030423>

Koehn, P. & Knowles, R. (2017). Six Challenges for Neural Machine Translation, *Proceedings of the First Workshop on Neural Machine Translation*, s. 28–39. Noudettu 27.3.2022 osoitteesta

<https://aclanthology.org/W17-3204.pdf>

Koehn, P. (2017). *Introduction to Neural Machine Translation by Philipp Koehn, Omniscien*

*Technologies Webinar*. Noudettu 8.4.2022 osoitteesta <https://youtu.be/oP-oZl8Hedw>

Koponen, M. (2016a). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25, s. 131–148. Noudettu

18.4.2022 osoitteesta [https://www.jostrans.org/issue25/art\\_koponen.pdf](https://www.jostrans.org/issue25/art_koponen.pdf)

Koponen, M. (2016b). *Machine Translation Post-editing and Effort: Empirical Studies on the Post-editing Process* [Väitöskirja, Helsingin yliopisto]. <http://urn.fi/URN:ISBN:978-951-51-1975-9>

- Koskela, M. (2020). Institutionaalista tekstilajia laatimassa: Valtionhallinnon henkilöstön näkemyksiä vuosikertomuksen synnystä. Teoksessa H. Hirsto, M. Enell-Nilsson, H. Kauppinen-Räisänen, & N. Keng (toim.), *Työelämän viestintä III* (s. 33–46). VAKKI Symposium XL 6.–7.2.2020. Noudettu 10.4.2022 osoitteesta [https://vakki.net/wpcontent/uploads/2020/12/Vakki2020\\_Koskela\\_UusiPDF.pdf](https://vakki.net/wpcontent/uploads/2020/12/Vakki2020_Koskela_UusiPDF.pdf)
- Laki Suomen Pankista* 1998/214. Annettu Helsingissä 27.3.1998. Noudettu 25.3.2022 osoitteesta <https://www.finlex.fi/fi/laki/smur/1998/19980214>
- Laurén, C. (1993). *Fackspråk. Form, innehåll, funktion*. Lund: Studentlitteratur.
- Maučec, M.S. & Donaj G. (2019). Machine translation and the Evaluation of Its Quality. Teoksessa A. Sadollah & T.S. Sinha (toim.), *Recent Trends in Computational Intelligence*. Intech Open. Noudettu 7.4.2022 osoitteesta <https://www.intechopen.com/chapters/68953>
- Meijer, S. (1993). Attitudes towards Machine Translation. *Language International*. Noudettu 15.4.2022 osoitteesta <https://aclanthology.org/www.mt-archive.info/90/LangInt-1993-Meijer.pdf>
- Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Aji, A.F., Bogoychev, N., Martins A.F.T. & Birch, A. (2018). *Marian: Fast neural machine translation in C++*. Noudettu 10.5.2022 osoitteesta <https://arxiv.org/abs/1804.00344>
- Mäkinen, M. (2019). *Domain adaptation: Retraining NMT with translation memories* [Pro gradu - tutkielma, Helsingin yliopisto]. Noudettu 27.3.2022 osoitteesta <https://helda.helsinki.fi/handle/10138/302732>
- Nieminen, T. (2021). OPUS-CAT: Desktop NMT with CAT integration and local fine-tuning. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (s. 288–294). Association for Computational Linguistics. Noudettu 17.1.2022 osoitteesta <https://aclanthology.org/2021.eacl-demos.34.pdf>
- Och, F.J. (2003). Minimum Error Rate Training in Statistical Machine Translation. *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics (ACL)*, s. 160–167. Noudettu 16.1.2022 osoitteesta <https://aclanthology.org/P03-1021.pdf>
- OPUS – an open source parallel corpus (2022). Noudettu 27.3.2022 osoitteesta <https://opus.nlpl.eu>
- Papineni, K., Roukos, S., Ward, T. & Zhu, W. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation, *Proceedings of the 40th Annual Meeting of the Association for*

- Computational Linguistics (ACL)*, s. 311–318. Noudettu 16.1.2022 osoitteesta <https://aclanthology.org/P02-1040.pdf>
- Papula, N. & Eskola, K., Nieminen, T. (2021). Fiskmö-konekäännin. *Tietolinja*, 2021(1).  
<https://urn.fi/URN:NBN:fi-fe2021061938982>
- Pasanen, P. (2015). Tavallisten kielenkäyttäjien käsityksiä termeistä. Teoksessa D. Rellstab & N. Sponkoski (toim.), *Rajojen dynamiikkaa* (s. 178–187) VAKKI-symposiumi XXXV. VAKKI Publications 4. Noudettu 17.4.2022 osoitteesta [https://www.vakki.net/publications/2015/VAKKI2015\\_Pasanen.pdf](https://www.vakki.net/publications/2015/VAKKI2015_Pasanen.pdf)
- Petrell, K. (2019). Svensköversättarna inom förvaltningen – en överblick av läget. *Språkbruk* 2019(1). Noudettu 24.3.2022 osoitteesta <https://www.sprakbruk.fi/-/svenskoversattarna-inom-forvaltningen-en-overblick-av-laget>
- Post, M. (2018). *A Call for Clarity in Reporting BLEU Scores*. Proceedings of the Third Conference on Machine Translation (WMT), Volume 1: Research Papers, s. 186–191 Noudettu 8.5.2022 osoitteesta <https://aclanthology.org/W18-6319.pdf>
- ProCom (2019). *Suomen Pankin vuosikertomus on vuoden paras oma media*. ProCom – Viestinnän ammattilaiset ry. Noudettu 7.5.2022 osoitteesta: <https://procom.fi/suomen-pankin-vuosikertomus-on-vuoden-paras-oma-media/>
- ProZ (2013). *CAT tool use by translators: what are they using?* Noudettu 10.4.2022 osoitteesta <https://go.proz.com/blog/cat-tool-use-by-translators-what-are-they-using>
- Pym, A. (2011). What technology does to translating. *The International Journal for Translation & Interpreting Research*, 3(1), 1–9. Noudettu 15.4.2022 osoitteesta <http://www.trans-int.org/index.php/transint/article/viewFile/121/81>
- Pym, A. (2014). *Exploring Translation Theories*. Routledge.
- Ragni, V., Vieira, L.N., (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, 30:1, s. 137–158.  
<https://doi.org/10.1080/0907676X.2021.1889005>
- Reiter A. (2018). A Structured Review of the Validity of BLEU. *Computational Linguistics*, 44 (3), s. 393–401. Noudettu 17.1.2022 osoitteesta <https://aclanthology.org/J18-3002.pdf>

- Roine, S-L. (2012). Talousviestintä. Teoksessa P. von Hertzen, E. Melgin, L. Åberg (toim.), *Vuosisata suhdetoimintaa – Yhteisöviestinnän historia Suomessa*. Noudettu 7.5.2022 osoitteesta [https://procom.fi/wp-content/uploads/2018/03/Talousviestint%C3%A4\\_Sirkka-Liisa-Roine.pdf](https://procom.fi/wp-content/uploads/2018/03/Talousviestint%C3%A4_Sirkka-Liisa-Roine.pdf)
- Rossi P., Ainoa A., Eloranta O., Grandell M., Lindberg M., Pasanen J., Sihvonen A., Hakola O. & Pirinen T. (2017). *Kuka opettaa ruotsia? – Ruotsin kielen opettamiseen kelpoisuuden tuottavien koulutusten arviointi*. Kansallinen koulutuksen arviointikeskus Karvi. Noudettu 24.3.2022 osoitteesta [https://karvi.fi/wp-content/uploads/2017/05/KARVI\\_1417.pdf](https://karvi.fi/wp-content/uploads/2017/05/KARVI_1417.pdf)
- Salminen, S. (2020). *Snowmobiles in the pants – Tekstiialan tuotekuvausten neuroverkkokonekääntäminen* [Pro gradu -tutkielma. Helsingin yliopisto]. <http://urn.fi/URN:NBN:fi:hulib-202006223383>
- Saunders, D. (2022). *Domain Adaptation and Multi-Domain Adaptation for Neural Machine Translation: A Survey*. Noudettu 10.5.2022 osoitteesta <https://arxiv.org/pdf/2104.06951.pdf>
- SDL (2020). *Translation Technology Insights 2020. Coping with the rise in pressure: Why humanizing technology is key to translation success*. Noudettu 16.4.2022 osoitteesta <https://www.trados.com/download/tti20-humanizing-technology-to-cope-with-pressure/165258/>
- Suomen Pankki 2022, Vuosikertomus-verkkosivu. Noudettu 27.3.2022 osoitteesta <https://www.suomenpankki.fi/fi/media-ja-julkaisut/julkaisut/vuosikertomus/>
- Suomen perustuslaki 1999/731*. Annettu Helsingissä 11.6.1999. Noudettu 10.4.2022 osoitteesta <https://www.finlex.fi/fi/laki/ajantasa/1999/19990731>
- Sutskever I., Vinyals O. & Le Q.V. (2014). Sequence to sequence learning with neural networks. Teoksessa *Advances in neural information processing systems*, s. 3104–3112. <https://dl.acm.org/doi/10.5555/2969033.2969173>
- Svenskt lagspråk i Finland – ruotsinkielinen lakikieli* (2017). Statsrådets kanslis publikationsserie 1/2017. Noudettu 10.4.2022 osoitteesta <https://julkaisut.valtioneuvosto.fi/handle/10024/79143>
- Temnikova, I., (2010): Cognitive Evaluation Approach for a Controlled Language Post–Editing Experiment. *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*. s. 3485–3490. Noudettu 23.1.2022 osoitteesta [http://www.lrec-conf.org/proceedings/lrec2010/pdf/437\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/437_Paper.pdf)

- Tezcan, A., Bulté B. & Vanroy B. (2021). Towards a Better Integration of Fuzzy Matches in Neural Machine Translation through Data Augmentation. *Informatics* 8(1), 7. Noudettu 8.5.2022 osoitteesta <https://www.mdpi.com/2227-9709/8/1/7/htm#B14-informatics-08-00007>
- Tiedemann, J., Nieminen, T., Aulamo, M., Kanerva, J., Leino, A., Ginter, F. & Papula, N., (2020). *The FISKMÖ Project: Resources and Tools for Finnish-Swedish Machine Translation and Cross-Linguistic Research, Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA), s. 3808–3815. Noudettu 23.1.2022 osoitteesta <https://aclanthology.org/2020.lrec-1.470.pdf>
- Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), s. 2214–2218. Noudettu 23.3.2022 osoitteesta <https://aclanthology.org/L12-1246/>
- Tiedemann, J. & Thottingal, S. (2020). OPUS-MT – Building open translation services for the World. A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg M.L. Forcada (toim.), *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, s. 479–480. Noudettu 27.3.2022 osoitteesta <https://aclanthology.org/2020.eamt-1.61.pdf>
- Tiedemann, J. (2020). The Tatoeba Translation Challenge - Realistic Data Sets for Low Resource and Multilingual MT. Teoksessa *Proceedings of the Fifth Conference on Machine Translation*, s. 1174-1182. Noudettu 23.3.2022 osoitteesta <https://www.aclweb.org/anthology/2020.wmt-1.139>
- Toral A. & Sánchez-Cartagena V.M. (2017). A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. *Proceedings of the 15th conference of the European chapter of the association for computational linguistics*, s. 1063–1073. Noudettu 17.1.2022 osoitteesta <https://aclanthology.org/E17-1100.pdf>
- Vashee, K. (2019, 12.huhtikuuta). *Understanding machine translation Quality: BLEU scores*. RWS:n blogi. Noudettu 10.5.2022 osoitteesta <https://www.rws.com/blog/understanding-mt-quality-bleu-scores/>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. & Polosukhin, I. (2017). Attention Is All You Need, *31st Conference on Neural Information Processing Systems (NIPS 2017)*. <https://dl.acm.org/doi/pdf/10.5555/3295222.3295349>

Weaver, W. (1949). *Translation*. The Rockefeller Foundation. Noudettu 27.3.2022 osoitteesta [http://gunkelweb.com/coms493/texts/weaver\\_translation.pdf](http://gunkelweb.com/coms493/texts/weaver_translation.pdf)

White, J.S., O'Connell, T. (1994). Adaptation of the DARPA machine translation evaluation paradigm to end-to-end systems. *Conference of the Association for Machine Translation in the Americas*, 193–205. Noudettu 18.4.2022 osoitteesta <http://aclanthology.lst.uni-saarland.de/1996.amta-1.11.pdf>

Wu Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Łukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M. & Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. <https://doi.org/10.48550/arXiv.1609.08144>

# Sammanfattning på svenska

Helsingfors universitet

Humanistiska fakulteten

Magisterprogrammet i översättning och tolkning

Översättnings- och tolkningskommunikation

Kristian Appelberg: Träning av ett finsk-svenskt neuralt maskinöversättningsverktyg till stöd för översättning av Finlands Banks årsberättelse till svenska

Magisteravhandling: 51 s., svenskspråkig sammanfattning 8 s.

Maj 2022

## 1. Inledning

Enligt 17 § i grundlagen (731/1999) är Finland ett tvåspråkigt land. Trots det har det redan länge funnits en brist på översättare till svenska och efterfrågan på svenska översättningar har inte heller lättat under de senaste åren. Detta kan man se till exempel i den ökade efterfrågan på aktualitetstexter inom statsförvaltningen, och det finns inte tillräckligt med nya professionella översättare för att svara på den växande efterfrågan och ersätta den stora mängden svensköversättare som kommer att gå i pension (Petrell 2019). Ett hjälpmedel i detta snedvridna förhållande mellan det ökande behovet av översättningar och minskade antalet språkproffs är att anlita maskinöversättningar som stöd. De senaste årens tekniska utveckling har gett allmänheten och professionella översättare tillgång till maskinöversättningar som påstås ha uppnått kvaliteten i professionella översättares översättningar, åtminstone enligt vissa mått och med ett begränsat material (Hassan m.fl. 2018).

Denna avhandling går ut på att skraddarsy det neurala maskinöversättningssystemet OPUS-CAT med Finlands Banks årsberättelse för åren 2019 och 2020 och undersöka vilken inverkan detta har på kvaliteten på en maskinöversättning av årsberättelsen 2021. Ett ytterligare mål är att undersöka hur verktyget klarar av att översätta delar av årsberättelsen som inte finns i översättningsminnet. Det tredje målet är att bedöma hur krävande det är att skraddarsy ett maskinöversättningsverktyg. Utvecklingen av maskinöversättningens kvalitet mäts med automatiska mått. Till slut mäts maskinöversättningens kvalitet genom en manuell genomgång av de segment som inte finns färdigt i översättningsminnena som innehåller årsberättelserna för 2019 och 2020.

I undersökningen används därför också översättningsminnet Trados Studio för att automatiskt översätta vissa delar av årsberättelsen, eftersom årsberättelsen innehåller en stor mängd årligen



återkommande formuleringar och innehåll och det inte lönar sig att översätta allt från noll, utan mycket finns färdigt i översättningsminnena. De översättningsminnen jag skapar för studien innehåller samma material som jag använder för att skräddarsy OPUS-CAT. Det optimala sättet att skräddarsy ett maskinöversättningsverktyg går ut på att använda så mycket material som möjligt, eftersom träningsmaterialets mängd korrelerar direkt med hur bra ett maskinöversättningssystem lär sig språk (Koehn & Knowles 2017: 28), men med tanke på undersökningens reproducerbarhet är det ändamålsenligt att använda ett mindre men öppet material, bestående av språkgranskade och publicerade översättningar av dokumenterat god kvalitet.

Min hypotes är att kvaliteten på maskinöversättningen förbättras genom att skräddarsy maskinöversättningsverktyget, eftersom OPUS-CAT är avsett att fungera med begränsat träningsmaterial (Nieminen 2021: 290). Genom studien vill jag också visa att det är lätt och värt ansträngningen att skräddarsy ett maskinöversättningsverktyg och i allmänhet sänka tröskeln för översättare att använda maskinöversättning som en tilläggsresurs i översättningsarbetet.

## **2. Översättningsverktyg och maskinöversättning**

Datorstödd översättning och relaterade översättningsverktyg eller s.k. CAT-verktyg (eng. *computer-assisted translation*) är viktiga verktyg inom det moderna översättaryrket. Den viktigaste funktionen hos CAT-verktyg är översättningsminnen (eng. *translation memory, TM*), som gör det möjligt för översättaren att hämta ord, uttryck eller hela meningar som tidigare översatts från en tvåspråkig databas som innehåller källtexter och deras översättningar. De första datorbaserade översättningsminnena utvecklades på 90-talet som svar på det växande lokaliseringsbehovet. Målet var att möjliggöra översättning av stora textmassor på datorer med flera översättare samtidigt så att ordförrådet och stilen förblir enhetliga. I min undersökning använder jag Trados Studio, som är ett av de först lanserade CAT-verktygen och utgör marknadsledaren med en marknadsandel på över 70 % (ProZ 2013). Trados Studio är främst ett översättningsminne, men omfattar också andra verktyg. När översättaren öppnar ett nytt jobb, delas källtexten upp i *segment* som ska översättas ett efter ett – i praktiken meningar, men också ofullständiga meningar som rubriker, kataloger och tabellers celler utgör separata segment. Efter översättning sparas dessa segment i minnet, och minnet föreslår automatiska översättningar för segment som har översatts tidigare om källtexten är tillräckligt lik.

Trots att översättningsminnen är utmärkta verktyg, måste översättaren beakta också deras brister. En självklar kritik hänför sig till de översatta segmentens kvalitet, eftersom inte ens hundra procentiga motsvarigheter i minnet innebär att den sparade översättningen är god, utan endast att källtexten är densamma som i det sparade segmentet. Den sparade översättningen kan också vara dålig, olämplig i det aktuella sammanhanget eller, om det är fråga om en översättning av till exempel ett ord, helt

enkelt fel ord. När översättningen sker per segment kan översättarens uppmärksamhet också överföras från texten som helhet till meningsnivå, vilket kan minska den syntaktiska kohesionen (Pym 2011: 3) och leda till inkonsekventa översättningar (Bowker 2005: 16).

Professionella översättarens attityder gentemot maskinöversättning har varierat under åren. Å ena sidan har maskinöversättning på grund av den skiftande kvaliteten setts som högst som en kuriositet, medan andra har sett det som ett hot mot hela yrket (Meijer 1993: 1), eftersom det bland annat har lett till att uppdragsgivare betalat lägre ersättningar för översättningarna (Garcia 2009: 202). Enligt språkteknologiföretaget SDL:s enkät från 2020 ansåg 26 % av översättarna att maskinöversättning utgör ett verkligt hot mot yrket (SDL 2020: 14). Enkäten i Jari Juutinens magisteravhandling visade på en liknande andel hos översättare från finska till engelska och engelska till finska, men i allmänhet varierar attityderna fortfarande mycket (Juutinen 2020: 44, 47). SDL:s enkät visar emellertid också att översättarnas attityder gentemot maskinöversättning blir mer positiva när de bekantar sig med verktyget (SDL 2020: 14), vilket kan tolkas som att översättare fortfarande har vissa fördomar mot främmande teknik.

Under årens lopp har olika slags maskinöversättningsverktyg utvecklats: de första utvecklades redan på 50-talet och var regelbaserade (*rule-based machine translation, RBMT*). RBMT grundade sig på användningen av ordböcker och förhandsprogrammerade grammatiska regler. Eftersom språk är komplexa och reglerna dessutom har undantag, är det omöjligt att lära ett verktyg allt, och ordagranna översättningar leder lätt till fel synonymer. Nästa steg var statistisk maskinöversättning (en. *statistical machine translation, SMT*) som utvecklades på 90-talet. SMT använder stora tvåspråkiga korpusar för att lära sig matematiskt sannolika översättningar av ord och fraser, och resultaten var lovande. Eftersom maskinen dock inte "förstår" språk, finns en stor risk för att skapa ogrammatiska konstruktioner. Det senaste utvecklingssteget var övergången till neural maskinöversättning (eng. *neural machine translation, NMT*), som skedde i mitten av 2010-talet. NMT bygger på användningen av neurala nätverk, dvs. artificiell intelligens, som försöker imitera mänsklig språkinläring. Inläringen sker fortfarande med stora korpusar, men eftersom NMT också lär sig grammatiska konventioner och i sina ordval beaktar det textuella sammanhanget, dvs. de närliggande orden, leder denna teknik till mer flytande och grammatikaliskt rätta översättningar.

Maskinöversättning har emellertid ännu inte nått professionella översättarens nivå (Koponen 2018: 8), och problemen accentueras i mindre språk där det inte finns tillräckligt med källmaterial för att träna den artificiella intelligensen (Haque m.fl. 2021: 1). Trots sina brister lättar maskinöversättning dock på översättarnas arbetsbörda. Även om en maskinöversättning inte kan publiceras som sådan, kan texten användas som en utgångspunkt för efterredigering. Forskning har visat att efterredigering av facktexter är betydligt snabbare och kräver mindre kognitiva resurser jämfört med traditionell översättning (Jia

m.fl. 2019: 62). För att en maskinöversättning ska kräva minimal efterredigering måste maskinen dock skräddarsys för en viss texttyp eller ett särskilt språk (eng. *domain adaptation*), eftersom ordens betydelse och den språkliga stilen varierar i olika sammanhang.

Eftersom generiska NMT-system producerar grammatikaliskt smidiga men oprecisa översättningar av facktexter (Koehn & Knowles 2017: 28), måste maskinen skräddarsys för ändamålet. Detta sker genom att träna maskinen med så stora (ibid.) och så goda textmängder som möjligt (Tiedemann 2020: 4). Bland annat Maria Mäkinen (2019) och Saara Salminen (2020) har i sina magisteravhandlingar behandlat processen för att skräddarsy neurala maskinöversättningsverktyg inom ett visst fackområde.

### **3. Teoretisk bakgrund och forskningsmaterial**

Enligt Päivi Pasanen kan fackspråk (eng. *language for special purposes, LSP*) definieras på två sätt: 1) det är ett språk som används för att kommunicera mellan experter inom ett fackområde eller för att förmedla information till studenter inom området eller, mindre ofta, till lekmän eller 2) det kan definieras utgående från att det präglas av en viss terminologi (Pasanen 2015: 179). Eftersom språket i Finlands Banks texter innehåller stora mängder termer och ofta riktar sig till experter, kan det sägas representera ekonomiskt fackspråk. Forskningsmaterialet i studien, Finlands Banks årsberättelse, innehåller emellertid både texter som experter skrivit för andra experter (t.ex. bokslutet) och texter som riktar sig till lekmän om bankens verksamhet och roll som myndighet. Ur detta perspektiv kan årsberättelsen placeras mellan det ekonomiska fackspråket och marknadskommunikation. Enligt Christer Laurén är ekonomiska fackspråket en kombination av flera näraliggande teknolekter (Laurén 1993: 45–46).

Utifrån översättarens perspektiv är årsberättelsen en mångsidig och utmanande texttyp, eftersom den innehåller både faktatunga ekonomiska texter och marknadskommunikation. I översättningen måste man därför sträva efter ett läsarvänligt och begripligt språk, men samtidigt finns det många detaljer i texten som regleras i lagar och förordningar i fråga om verksamhetsberättelsen och bokslutet. Årsberättelsen utgör ofta ett visitkort för organisationen och får därför stor synlighet. Den är också i hög grad kopplad till tidigare årsberättelser, pressmeddelanden och de termer som används i dem. Till många delar är översättarens händer i sista hand bundna till vissa stilistiska val och ordval, eftersom hen med tanke på konsekvent kommunikation bör följa de ordval som används i översättningarna av press- och börsmeddelanden som publicerats under året (Delingua 2019.).

Mitt forskningsmaterial är Finlands Banks årsberättelse 2021 och dess svenska maskinöversättningar, som jag jämför med den svenska översättning som gavs ut av Finlands Bank den 24 mars 2022. Finlands Banks årsberättelse är en verksamhetsberättelse enligt 19 § i lagen om Finlands Bank (1998/214) och

omfattar balansräkning, resultaträkning och bilagor till dessa samt en verksamhetsberättelse. Enligt Finlands Bank har årsrapporter publicerats sedan 1914, fram till 1996 under namnet *Finlands Bank - Årsbok*. Sedan 1997 har årsberättelsen också publicerats elektroniskt (Finlands Bank 2022). Under 2015 övergick banken till att publicera boken endast i elektroniskt format och den nyaste versionen, en webbårsberättelse, publicerades första gången på våren 2019 (Hienonen 2019). Sedan 2019 består årsredovisningen av tre huvuddelar: centralbankens verksamhetsberättelse, en beskrivning av verksamheten och strategin samt bokslutet. I årsberättelsen ingår också chefsdirektörens hälsning och en organisationsbeskrivning. Följande arbetsskeden ingick i insamlingen av materialet för att skraddarsy översättningsmaskinen:

1. nedladdning av årsberättelserna som pdf-filer på Finlands Banks webbplats
2. konvertering av pdf-filer till Word-dokument
3. genomgång och rensning av Word-dokumenterna
4. sammanlänkning av segmenten i Trados
5. manuell kontroll och korrigerings av sammanlänkningen
6. exportering av segmenten till ett nytt översättningsminne
7. konvertering av översättningsminnet till en tvåspråkig tmx-fil.

#### **4. Forskningsmetod**

Min studie går ut på att skraddarsy en generisk översättningsmaskin med Finlands Banks årsberättelser för åren 2019–2020 och deras svenska översättningar, varefter jag jämför den generiska maskinöversättningen av årsberättelsen 2021 mot den skraddarsydda maskinens översättning. OPUS-CAT är ett icke-kommersiellt maskinöversättningsverktyg avsett för professionella översättare som utvecklats av Tommi Nieminen OPUS-CAT utgående från Helsingfors och Åbo universitets och Kites rf:s Fiskmö-projekt (finsk-svensk-maskinöversättning, Tiedemann m.fl. 2020). Programmet kan integreras i Trados Studio och ett antal andra översättningsprogram. OPUS-CAT utnyttjar neurala nätverk (eng. *Neural Machine Translation, NMT*). Programmet installeras på datorns hårddisk och kan till skillnad från många andra översättningsmaskiner användas utan internetuppkoppling (Nieminen 2021: 288).

OPUS-CAT använder generiska tvåspråkiga översättningsmodeller (*OPUS-MT models*) som har tränats med Microsofts neurala maskinöversättningsprogram Marian (Microsoft Github 2022). Marian bygger på Transformer NMT-arkitekturen, som presenterades av Vaswani m.fl. år 2017. De generiska översättningsmodellerna har tränats med öppna data från den öppna korpusdatabasen OPUS (Tiedemann & Thottingal 2020: 1). Det främsta syftet med OPUS-projektet är att sammanställa fritt tillgängliga tvåspråkiga texter på webben i en form som är lätt att använda, framför allt på mindre

språk (Tiedemann 2012: 2214). OPUS-MT-modellen som jag använder i undersökningen är opus-2021-02-19.

De flesta automatiska mått för utvärdering av maskinöversättningar baserar sig på att maskinöversättningen jämförs med en professionell översättares översättning. Det populäraste automatiska måttet för detta ändamål är BLEU (eng. *bilingual evaluation understudy*), som presenterades av Kishore Papineni m.fl. 2002. Trots sina brister har BLEU bevarat sin ställning som standarden för automatisk bedömning av maskinöversättningar eftersom det korrelerar bra med mänskligt omdöme vad gäller korrekthet och flyt. Utgångspunkten för BLEU är att ju närmare maskinöversättningen är till en professionell översättares översättning av samma text, desto bättre är den (Papineni m.fl. 2002: 3011).

BLEU mäter översättningens noggrannhet från noll till ett, där noll är lägsta vitsordet och ett det högsta. Detta uttrycks ofta i hundradelar, dvs. på skalan 0–100, så också i denna studie. Den matematiska ekvationen går ut på att först räkna antalet sådana ordled (N-grams, ordled på 1–4 ord) i maskinöversättningen som också förekommer i referensöversättningen som gjorts av en professionell översättare. Detta antal ord divideras med det totala antalet ord i maskinöversättningen, vilket mäter översättningens precision. Till ekvationen läggs till slut en straffkoefficient, (eng. *brevity penalty*, *BP*), som drar av poäng om maskinöversättningarna är kortare än referensöversättningarna, så att de korta översättningarna inte får relativt sett bättre poäng. Straffkoefficienten räknas dock inte ut för varje segment, utan för hela maskinöversättningen i genomsnitt, vilket möjliggör flexibilitet i översättningens längd på segmentnivå. Straffkoefficienten är ett minus referensöversättningens längd dividerat med maskinöversättningens längd. Direkt uttryckt beräknas BLEU-poängen enligt följande ekvation:

*straffkoefficienten*  $\times$  (*antalet N-grams referensöversättningen* / *N-grams referensöversättningen*)

Med denna ekvation bedöms maskinöversättningen likhet till referensöversättning utgående från att den så noggrant som möjligt borde motsvara referensöversättningen i fråga om översättningens längd (straffkoefficient), ordval (motsvarande ord) och ordföljd (bättre poäng för liknande struktur).

Faser i maskinöversättning och automatisk bedömning av översättningarna:

1. installation av generisk OPUS-modell från nätet
2. träning av modellen med den tvåspråkiga tmx-filen
3. utvärdering av inlärningsresultatet med det integrerade SacreBLEU-verktyget
4. maskinöversättning av källtexten med den skräddarsydda modellen
5. jämförelse av maskinöversättningen och referensöversättningen med BLEU.

Efter den automatiska bedömningen utförs en manuell genomgång av översättningarna utgående från en tillämpad version av Irina Temnikovas modell för felanalys av maskinöversättningar (Temnikova 2010: 3487). Den tillämpade felmatrisen analyserar felkategorierna **morfologiska fel** (rätt ord men fel form), **lexikaliska fel** (fel synonym; fel ord; för många ord; ord saknas; felöversatt idiomatiskt uttryck) och **syntaxfel** (fel ordföljd). Utöver felanalysen bedöms om det maskinöversatta segmentet utgör en god grund för efterredigering och om det kräver mycket eller lite arbete.

#### 4. Analys och resultat

Den generiska modellen som jag laddar ner får namnet OC1. Processen för att skräddarsy modellen börjar med årsberättelsen för 2020. Den första skräddarsydda modellen får namnet OC2. Efter att ha gjort den första maskinöversättningen fortsätter jag att träna modellen med årsberättelsen för 2019. Denna modell får namnet OC3. Tmx-filen med årsberättelsen för 2020 innehåller 3240 unika segment som jag använder för att skräddarsy den första modellen. Jag fortsätter att skräddarsy maskinen med årsberättelsen för 2019, som består av 2282 unika segment. Eftersom 529 av dem fanns redan i årsberättelsen för 2020, utökades träningsmaterialet med enbart 1753 nya segment.

Maskinöversättningarna som skapas med dessa modeller får namnen **OC1\_SV**, **OC2\_SV** och **OC3\_SV**. En jämförelse av maskinöversättningarna mot referensöversättningen med BLEU ger följande resultat:

	<b>OC1_SV</b>	<b>OC2_SV</b>	<b>OC3_SV</b>
<b>BLEU</b>	32,65	39,36	40,68

Tabell 1: BLEU-poäng för maskinöversättningarna.

Såsom framgår av tabellen är översättningen relativt acceptabel redan med den generiska modellen. Googles tumregel för bedömning av BLEU-poäng i AutoML Translation (Google 2022) placerar både den generiska modellen och den första skräddarsydda modellen i kategorin "förståelig eller god översättning", 30–40 poäng. Av tabell 1 framgår att detta inte förbättrade översättningsmodellen märkbart. Detta tyder på att materialet är för homogent för att maskinen ska kunna lära sig nå got nytt. Enligt Googles instruktion tyder 40–50 BLEU-poäng på att "översättningen är av god kvalitet".

Efter den automatiska bedömningen med BLEU genomförs en manuell genomgång av maskinöversättningen OC3\_SV. Den manuella granskningen gäller endast de segment som inte förekommer i de tidigare årsberättelserna. En automatisk översättning av källtexten med Trados Studios pre-translate-funktion lämnar 1353 segment utan en tillräckligt god översättning och måste översättas med den skräddarsydda OPUS-CAT-modellen OC3.

En manuell genomgång av de maskinöversatta segmenten visar på följande resultat.

	Morfologi	Lexikon					Syntax
Feltyp	Rätt ord, fel form	Fel synonym	Fel ord	För många ord	Ord saknas	Felöversatt idiomatiskt uttryck	Fel ordföljd
Antal förekomster	58	150	60	22	50	28	45
Antal fel totalt	413						

Tabell 2: Manuell genomgång av maskinöversättningarna.

## 5. Slutsatser

Av resultaten kan dras slutsatsen att kvaliteten på maskinöversättningen av årsberättelsen förbättrades avsevärt (med ca 7 BLEU-poäng) genom att skräddarsy systemet med det föregående årets årsberättelse. Fortsatt träning med årsberättelsen från ännu ett år tidigare förbättrade dock inte resultatet märkbart, vilket tyder på att materialet är för homogent och att NMT-systemet inte längre lär sig något nytt. Resultatet kan förmodligen förbättras genom att använda andra fackspråkliga texter inom samma fackområde, t.ex. pressmeddelanden, nyheter, publikationer och dylikt.

Den manuella genomgången av de maskinöversatta segmenten visar att den skräddarsydda OPUS-CAT-modellen skapar mycket goda översättningar som i regel fungerar som gott underlag för efterredigering. Felen är främst ytliga och lätta att korrigera. 37 (21 %) segment var färdiga utan redigering, 101 (57 %) skulle kräva lätt efterredigering, 35 (20 %) skulle kräva tung efterredigering och endast 3 (2 %) av de granskade 176 maskinöversättningarna var så dåliga att resultatet inte går att utnyttja alls.

Av avhandlingen framgår också att det är relativt lätt att skräddarsy ett NMT-system, speciellt om tidigare översättningar redan finns i ett översättningsminne. Eftersom professionella översättare i regel använder översättningsminnen och har tillgång till färdiga minnen, erbjuder OPUS-CAT ett förhållandevis enkelt och gratis sätt att skräddarsy ett maskinöversättningsverktyg att använda dem. Efterredigering av maskinöversättningar kräver ett annat förhållningssätt till arbetet än traditionell översättning, men kan i brådskande situationer vara till mycket stor hjälp vid översättning.