

Chat-GPT 5.0 LAAJAN KIELIMALLIN ONNISTUMINEN KÄÄNTÄJÄNÄ SUOMESTA ENGLANTIIN JA RUOTSIIN

1. Johdanto ja projektin kulku

Projektissa tutkittiin laajan kielimallin käyttöä kaunokirjallisuuden käännösnäytteiden tekemisessä. Halusimme nähdä, voiko laaja kielimalli nopeuttaa käännösnäytteen tekoa ja näin vaikuttaa käännösprosessin kustannuksia alentavasti.

Kääntämisen ja ohjelmistokäännösten arvioinnin ovat tehneet kääntäjä Minna Jefferey (suomi-englanti) sekä loppuvaiheen kielenopiskelija Eri Lassila Oulun yliopistosta (suomi-ruotsi).

Käännösohjelmaksi valittiin OpenAI:n ChatGPT 5.0. Se on maailmanlaajuisesti yksi käytetyimmistä laajoista kielimalleista ja siten myös yksi koulutetuimmista. Vaikka senkin tiedetään sotkeutuneen epäeettiseen koulutusmateriaalin, kuten piraattikirjastojen, käyttöön, voi sitä pitää melko luotettavana. ChatGPT -ohjelmasta tuli elokuussa 2025 uusi versio, joten päädyimme ottamaan juuri sen ohjelmaksemme. Ohjelmistopakettissa on useita eri versioita, esimerkiksi mini, nano ja 5.1., mutta testatessa käännös ei vaihdellut eri versioiden välillä. Käytimme siis 5.0 -versiota.

Valitsimme käännettäväksi kuusi pohjoispohjanmaalaisten kirjailijoiden tekstiä, joita he olivat tarjonneet projektiin. Tekstilajeiksi valikoituivat lastenrunot, moderni vapaamittainen runous, visuaalisia elementtejä ja proosarunoutta hyödyntävä runous, tyyliltään omaperäinen murreproosa, faktaa ja fiktiota yhdistävä proosa sekä draamaa lähestyvä proosa.

Halusimme kääntää tekstit niin sanotussa käännöskuplassa (eli API:n avulla), jossa ohjelmistoa käytetään irrallaan emo-ohjelmasta. Tällöin tekstistä saatu tieto eikä teksti itsessään tule osaksi ChatGPT:n tietoa. API-käännöstä varten piti kustakin tekstistä teettää kääntäjillä noin sivun mittainen ”maistiainen”, joka annettiin ohjelmistolle malliksi. Ohjelman niin sanottu promptaus eli ohjeistus oli hyvin yksinkertainen: käännä kielelle englanti tai käännä kielelle ruotsi.

Kun kaikista teksteistä oli sekä englannin että ruotsinkielinen käännös, lähetin ne kääntäjille alkuperäisten tekstien kanssa. Alkuperäiset tekstit lähetin muokkaamattomina, samoin käännökset, paitsi että tein runoihin perusasettelun samalla tavalla kuin runoilija oli sen tehnyt (esimerkiksi laitoin runot kukin omalle sivulleen, laitoin riviväliksi 1pt ja tummensin otsikot, jos kirjailija on niin alkuperäisessä tekstissä tehnyt). Varsinaisiin laajan kielimallin tekemiin muutoksiin asettelussa en puuttunut.

Ohjeistin heitä paitsi etsimään suoranaisia virheitä, myös arvioimaan käännöksiä laajemmin. Ohjeena oli käydä jokaisen tekstin kohdalla läpi, kuinka onnistuneelta käännös ylipäättään vaikuttaa. Pyysin myös tekemään muistiinpanot kohdista, jotka herättävät huomion: esimerkiksi kielelliset, sisällölliset tai kulttuuriset yksityiskohdat, jotka eivät osu kohdalleen;

suomen kielen rakenne tai sanonta, joka paistaa läpi käännöksestä; tyylikot tai heikot kielikuvat tai asettelun ja visuaalisuuden onnistuminen käännöksessä. Pyysin myös merkitsemään onnistuneet käännöksen kohdat tai luovat ratkaisut.

Kävin kaikki tekstit läpi kääntäjien kanssa keskustellen joko kasvokkain (Lassila) tai Teamsin välityksellä (Jeffery). Keskustelut olivat arvokkaita ja lisäsivät käsitystäni näiden tekstien kääntämisestä sekä kääntäjän työn monimutkaisesta prosessista ylipäätään. Tämä esitys on kuvaus hankkeesta tehdystä työstä, omasta ajatteluprosessistani ja haastatteluista sekä niistä tekemästäni päätelmistä. Se ei ole akateeminen tutkimus.

2. Laajan kielimallin hyvät puolet kääntämisessä

Laajan kielimallin käyttö selkeän, suorasanaisten asiatekstin muokkaamisessa, tuottamisessa ja kääntämisessä on yleistynyt ohjelmistojen kehittyessä. On itsestään selvää, että ihmiset haluavat käyttää työkaluja ja välineitä, jotka tekevät heidän elämästään helpompaa ja vähentävät työhön tai aktiviteettiin käytettävää aikaa ja vaivaa.

Henkilölle, joka ei ole tutustunut laajoihin kielimalleihin, voi olla suuri ihmetyksen aihe, kun ohjelma käännöstä pyydetessä tuottaa muutamassa sekunnissa luettavaa tekstiä toisella kielellä. Sillä sen ChatGPT lupaa - ja sen se myös tekee. Englanniksi käännettäessä käännöskupla rajoitti kääntämistä niin, että yhdellä kertaa ohjelma käänsi suomenkielisestä alkuperäistekstistä englanniksi noin kaksi sivua ja ruotsiksi noin sivun. Tekstilajit eivät juuri sellaisinaan vaikuttaneet käännöksen nopeuteen, pikemminkin kyse oli tekstimassasta: runoja, joissa sanoja on vähemmän kokonaisuutta kohti, kääntyi samalla kerralla useampi sivu kuin tiheäsanaista proosaa. Vaikka työ piti palastella parin sivun kokonaisuuksiksi, sain parissa työpäivässä käännettyä ohjelman avulla satakunta sivua.

Laajan kielimallin käyttö on siis nopeaa. Mitä lähempänä teksti on yleiskielistä asiatekstiä, sitä oikeakielisempää ja sujuvampaa käännös myös on. Tämä ei ole ihme eikä kumma, onhan ohjelmisto koulutettu lähinnä asiateksteillä, mainoksilla, ohjeilla ja muilla Internetistä vapaasti löytyvillä teksteillä.

Mitä tulee pyydettyihin onnistuneisiin käännöksiin tai luoviin ratkaisuihin, molemmat kääntäjät totesivat, että muutamissa kohdin kielimalli oli tehnyt luovia valintoja. Esimerkiksi *kortteli kerrallaan* on käännetty *block by block*, mikä on kivan idiomaattinen ilmaus. Ruotsiksi käännettäessä ohjelma on vaihtanut kirosan *kyrpä* sanaksi *fitta*. Koska kyseessä on vuorosana ja huudahdus, valinta toimii paremmin kuin oletuksenmukainen *kuk*, Samaten löytyi uusi partisiipin preesensmuoto *ihopflätade* – yhteen letitettyinä.

Toisaalta ohjelma on hämmentynyt sanan *käypäläinen* edessä ja kääntänyt sen englanniksi sanalla *insect* eli hyönteinen, vaikka tekstissä ei puhuta mistään siihen viittaavasta. Ruotsiksi käännettäessä mukaan on otettu muutamia sanoja toisista skandinaavisista kielistä,

esimerkiksi yksi tanskasta ja toinen norjasta. Tällainen luovuus ei ole tekstin onnistuneisuuden kannalta hyväksi.

ChatGPT 5.0:a on kehitetty sitä edeltävistä malleista karsimaan juuri edellä kuvattua niin sanottua hallusinoointia eli faktojen tai käsitteiden ”keksimistä” tai yhdistelyä. Hallusinointi on tehnyt ohjelmistojen teksteistä epäluotettavia, eli ihmisen on pitänyt tarkistaa tietojen paikkansapitävyys. Tämän minimoimiseksi ohjelman versio 5.0 kysyttäessä neuvoa perustelee valintojaan laajasti ja antaa vaihtoehtoja – ja kääntäessään valitsee yleisimmän ja asiallisimman sanan mahdollisista. Tämä ominaisuus karsii kielen variaatiota. Kaunokirjallisuuden kääntämisen kannalta yleisin vaihtoehto ei kuitenkaan usein ole osuvin.

3. Laajan kielimallin heikkoudet kääntämisessä

Laaja kielimalli kykenee näennäisen sujuvaan tekstin tuottamiseen vieraalla kielellä, mutta tarkemmin analysoitaessa puutteita on monenlaisia. Olen jakanut ne kolmeen kategoriaan: Suoranaiseksi virheiksi, tyylikkeiksi sekä kulttuurisen lukutaidon puutteeseen. Käsittelen sekä englanniksi että ruotsiksi tehtyjä käännöksiä rinnakkain, koska virhetyytit ja puutteet olivat molemmissa todella samankaltaisia.

3.1 Suoranaiset virheet

Laaja kielimalli teki käännöksissä monentyyppisiä kielivirheitä. Tutkimuksemme on laadullinen ja resurssien rajallisuuden vuoksi käsittelen niistä vain muutamia edustavia esimerkkikategorioita.

Yksi laajan kielimallin ongelma on, että se ei muista valintojaan eikä ole johdonmukainen niissä. Tämä tuo monenlaisia vaikeuksia: tekstin yhtenäisyys kärsii ja lukija kiinnittää huomiota eri logiikalla valittuihin sanoihin. Yksi selvä tällainen virhekattegoria on nimet. Välillä erisnimiä ei tunnisteta tai päinvastoin - laajalle kielimallinen tuntematon sana oletetaan erisnimeksi. Esimerkiksi *Veli*-sanaa isolla kirjoitettuna ei aina tunnisteta nimeksi: se käännetään sanalla *brother* tai *bror* tai sen käyttö vaihtelee. Nimeen tai runon otsikkoon liittyvä ongelma on myös, että ne saatetaan myös välillä kääntää semanttisesti väärin muodon perusteella: *nasta* on saanut englanniksi muodon *nasty* eli ilkeä – ilman että sisältö viittaisi ilkeyteen.

Toisaalta lastenrunoissa kääntäjä olisi varmasti käyttänyt tilaisuutta muokata joitain hahmojen nimiä käännöskieleen sopivammiksi: *Samppa frog* tai *Luppa hund* eivät kuulosta taidokkaalta kielenkäytöltä. Tapana on myös käyttää nimissä ruotsinkielisiä muotoja, jos sellaisia on. Hailuoto on laajan kielimallin käännöksessä sitkeästi suomeksi, vaikka *Karlö* on sen virallinen nimi toisella kotimaisella. Oma lukunsa on myös nimisanojen taivutusmuotojen siirtyminen sellaisenaan käännökseen: ”De vigdes i *Hailuodon* kyrka”.

Ruotsin kielen sanojen sukuja laaja kielimalli laittaa joskus väärin, tosin aika harvoin. Prepositioissa se tekee vähän useammin virheitä, valitsee väärän preposition tai lisää sellaisen turhaan esimerkiksi tehdäkseen tekstistä luettavamman ja oikeakielisemmän. Runokielen tahallinen aukkoisuus voi silloin kadota ja tekstin sisältö arkistua. Esimerkiksi ”*Olen vietävissä puutarhoihin/lehvät, kukinnot*”, joka on käännetty: ”*Jag är bärbar till trädgårdar/med kransar, blommor*”.

Englanniksi käännettäessä ongelmia tuottaa muun muassa futuuri, jota suomen kielessä ei ilmaista omana aikamuotonaan vaan preesensin ja sanavalintojen kautta. Esimerkiksi ”*Näin käy, lopulta*” kääntyisi paremmin englannin futuurin kautta kuin ”*This happens, ultimately*”. *Eräs nainen* on puolestaan käännetty *another woman*, eikä käytetty englannin kielen omaa keinoa ilmaista epämääräisyyttä – yksinkertaisesti valittu artikkeli *a woman*. Tässä näkyy laajan kielimallin työskentelytapa: se kääntää sanan kerrallaan ja usein pidättäytyy alkuperäisen kielen rakenteissa, sen sijaan että ajattelisi tekstin käännöskielen rakenteiden kautta.

Yksi virheeksi laskettava harvinainen (mutta tässäkin aineistossa esiin nouseva) piirre laajan kielimallin käännöksissä on sen tapa maskuliinistaa tekstiä. Kun pohjatekstissä ei ole suoraan sanottu henkilön sukupuolta, ohjelmalla on tapana valita sekä englannissa että ruotsissa maskuliininen pronomini. Näin tapahtui siinäkin tapauksessa, että suomeksi henkilöön viitataan sanalla *vaimo*.

Jos laaja kielimalli ei tunnista sanaa, se saattaa korvata sen jollain saman kuuloisella sanalla. Esimerkiksi sana *kippuu* (yleiskieli: kipua) on käännetty englanniksi sanalla *spill*, jonka voitaisiin ajatella olevan sanan *tippuu* käännös. Klassikkovirhe on sanan *kuusi* eri nyansseilla leikittelyn katoaminen, kun laaja kielimalli kääntää englanniksi *kuusii kuusii kuusii* pelkästään numeroilla, suorastaan pahaenteisesti *six six six*. Sana *lassii* (yleiskieli: lasia) puolestaan on saanut ruotsinnoksessa hauskan tulkinnan nimenä *Lassie*, *kirvata* (yleiskieli: kiroilla) puolestaan käännetty kirjoittamiseksi, *skriva*.

Ongelma ei koske vain murre sanoja, vaan myös muuta sanojen epätavanomaista tai kielikuvallista käyttöä. Esimerkiksi sanan *Elävä* (tekstissä merkitsee kaikkea elollista, kuoleman vastakohtaisuutta) kielimalli on kääntänyt ruotsiksi sanalla *däggdjur* eli nisäkäs. Aika paljon merkitystä katoaa ja muuttuu tällaisessa käännöksessä. Tavallista on myös haparoida erikoissanastossa: yhden tekstin ruotsinnoksessa armeijan sotilasarvot heittelivät maavoimista merivoimiin, vaikka kyse on rajavartioista.

Suoranaiseksi virheeksi voidaan laskea myös sanan kääntämättä jättäminen. Laaja kielimalli saattaa ohittaa vaikean tai murteellisen sanan tai lauseen ja jättää sen kokonaan pois. Näin tapahtuu esimerkiksi murteellisten ilmausten ”*Mie asun täs kuvas*” (yleiskieli: kuvassa) tai *muiski* (yleiskieli: muissakin) yhteydessä. Ongelmia on myös esimerkiksi omistusliitteiden kääntämisessä: *tätis* ei käänny *aunt* vaan oikeammin *your auntie*. Myös liitepartikkelit jätetään usein kääntämättä: esimerkiksi *-kin* -liitepartikkelin ohittaminen muuttaa sisältöä ratkaisevasti. Varsinkin runoissa sanojen pois jättäminen luo suuria merkityseroja. Ohjelma

on myös lisännyt yhden runon englanninkieliseen käännökseen fraasin: *and you seek to remember*.

Yhden käännettävän tekstin erityispiirteen eli proosatekstin ja asiatekstiä sisältävien alaviitteiden ongelman kielimalli ratkaisi jättämällä alaviitteet kääntämättä. Kun ajoin ne erikseen ohjelman läpi, se toki osasi kääntää ne hyvin, koska teksti ei ollut kaunokirjallisuutta, vaan asiatyylisiä yleiskieltä. Niiden paikoilleen laittaminen ”käsityönä” lisäisi tekstiä tarkistavan ja hiovan kääntäjän työtä suuresti.

3.2. Tyylin puutteet

Tyylin puutteet voi jakaa rakenteellisiin ja sanastollisiin virheisiin tai kömpelyyksiin. Tyyli-rikot vaihtelevat kielen vivahteiden katoamisista tai vaihtumisista, ”aika ok käännöksistä” tahattoman koomisiin ja kolhoihin valintoihin. Suuri osa kääntäjien tekemistä huomautuksista laajan kielimallien käännöksiin koski tyylin puutteita.

Kuten sanottua, laaja kielimalli matkii alkukielen rakenteita. Taipumus kääntää sana sanan jälkeen – sen sijaan että sisältö ilmaistaisiin toisella kielellä mahdollisimman identtisesti mutta tämän kielen rakenteiden kautta – luo kummallisia ja kömpelöitä ilmauksia. Esimerkiksi suomen *Voin vannoa* on käännetty englanniksi *I can swear*, kun oikeampi muoto olisi esimerkiksi *I swear*, *I promise* tai *I can assure you*. Tämän tyyppisiä tyylin notkahduksia oli jokaisessa tekstissä kummallakin kielellä huomattavan paljon.

Sanojen valinnassa tyyli-rajun puute näkyy erityisen selvästi, kun sanaa käytetään arkikielestä poikkeavalla tavalla. Erityisesti runot, jotka toimivat usein juuri etsien uutta ja yllättävää kielen käyttöä, ovat haastavia käännettäviä laajalle kielimallille. Aineiston lastenrunoissa runsaasti esiintyvää onomatopoiettikkaa se ei osaa kääntää, vaan ohittaa sanat tai jättää ne kääntämättä. Myös rytmillisyyys katoaa, kun tekstit käännetään sanojen sisältöä painottaen ja kirjaimellisesti. Ohjelma myös muuttaa rivityksiä, mikä on ongelma runojen säkeiden kohdalla. Useissa kohdissa ChatGPT on siirtänyt yhden tai useamman sanan toiseen säkeeseen, yhdistänyt säkeitä tai jättänyt säkeen osia kääntämättä tehdäkseen tekstistä ymmärrettävämpää ja selkeämpää. Taiteilijan valinnat, huolelliset rakennelmat, ChatGPT ajattelee virheiksi. Esimerkiksi sanatoisto, jolla on paikkansa tekstin tulkinnassa, leikataan pois ylimääräisenä ja virheellisenä ilmauksena.

Lisäksi käännökseen tulisi saada mukaan kirjailijan tarkoittamia vihjeitä ja nyansseja. Kun kirjailija käyttää sanasta *hyvinvointialue* väännöstä *pahoinvointialue*, kääntää laaja kielimalli sen alkuperäisellä termillä ja mahdollisuus huumoriin on hukattu. Samoin runon kohdassa peräkkäiset sanat *sisäiset*/*sisäkkäiset* on käännetty englanniksi vain yhdellä sanalla (*internal*) ja ohitettu kielen nyanssien esille tuomisen mahdollisuus.

Joissain kohdin käännöskieleen hiipii tyyli-rikkoja eri kielistä tai kielen eri muodoista: ruotsiksi sana *kivääri* oli käännetty anglismilla *rifle*. Englanniksi puolestaan muuten on

valittu brittienglanti, mutta välillä käytetään amerikanenglannin muotoja, esimerkiksi *goddamn*. Tai kun käännös on yleiskielistä runoutta, sanoja yhdistävät muodot kuten *that's* tai *it's* vaikuttavat puhekielisyydessään tyyli-poikkeamilta. Runossa tällainen puhekielisyyys olisi korostettu kielellinen valinta, mutta sellaista kirjailija ei ole tehnyt, eikä kääntäjä sitä tekisi.

3.3. Kulttuurinen lukutaidottomuus

Kulttuurinen tieto ja sen välittäminen on suuri osa kääntäjän työtä, joka työskentelee paitsi kahden kielen, myös kahden kulttuurin välillä. Tämä aukeni itselleni hienosti juttellessani Minna Jefferyn kanssa tekstistä, jossa käytettiin paljon metsään, rantaan ja kesämökkeilyyn liittyvää käsitteistöä. Niinkin yksinkertainen asia kuin *mökkinaapurit* ei kuulosta englanniksi oikealta *the cottage neighbours* käännöksenä. Peribritti ei tunnista myöskään *mökkitietä* tai *rantametsikköä*. Niin, kesämökki on jotain hyvin suomalaista eikä englantilaisilla rannoilla kasva metsää. Nämä käsitteet tulisi avata ja selittää niin, että lukija ymmärtää ne, mutta selitys on myös kirjailijan kielen ja tyylin kanssa sopusoinnussa.

Kun teksti käsittelee leimallisesti pohjoista maaseutua, kulttuurinen tuntemus korostuu entisestään: laaja kielimalli ei löydä erityissanastoa vaikkapa hirvenmetsästykseen. Tosin keskiverto kääntäjäkin joutuu varmaan etsimään, miten kääntää *nylkeminen*, *suolestus*, *kunnan mies* tai *yhteismetsä*. Kasvillisuuden ja eläinten kääntämisessä ChatGPT on melkein aina hyvä, mutta jostain syystä yhdessä tekstissä sotki lintujen nimiä ruotsiksi käännettäessä. Osassa ohjelmistoja on kyllä mahdollisuus ajaa erityissanastoa osaksi käännösohjeita. Mietin kuitenkin, onko se kääntäjän normaalia sanojen etsintää nopeampaa. Jonkun se erityissanastojen etsintä ja syöttäminen on tehtävä ja kääntäjän on niiden merkityksen oikeellisuus tarkistettava.

Esimerkkejä kulttuurisen kontekstin ohittamisesta on paljon. Kun *hakkuut* käännetään sanalla *cuts*, ei mukaan tule metsänhoidollista kontekstia. Sanatarkka: *Heats the sauna and goes to the benches* saa aikaan mielikuvan, että sauna lämmitetään ja mennään jonnekin penkeille ulos istuskelemaan. *Lauteet* täytyy siis ilmiönä selittää, samoin *water thrown on the stones*. Ehkä myös *kesäteatteri* on outo konsepti, ainakin *Salkkarit* täytyy selittää jatkuvajuoniseksi televisiosarjaksi. Lastenrunon nimi *heviä heviä* ei puolestaan herättänyt ChatGPT:ssä mielikuvaa raskaasta musiikista. Se tuotti virheelliset käännökset molemmissa kielissä: *heaving heaving* englanniksi ja *hästar och hästar* ruotsiksi.

Kulttuurista lukutaitoa vaatii myös sanomisen tapojen tunnistaminen. *Vuodenpäivät* ei käänny englanniksi sanalla *years*. sisältö muuttuu ratkaisevasti yhdestä vuodesta moneen. Lastenrunossa kiinalaisen kuuloinen tervehdys *hai long* on puolestaan ruotsinnettu *hej lång*, jolloin viittaus kolmanteen kulttuuriin katoaa.

Murteet ovat oma lukunsa kääntämisessä – kaikki murteellista tai edes voimakkaasti puhekielistä tekstiä tehneet tai lukeneet ymmärtävät, että sellaisia kirjoittaessa kyse ei ole

autenttisesta ilmauksesta vaan puheen representaatiosta. Autenttinen murre ei toimi kirjoitettuna ja kiinnittää liiaksi huomion sanomisen tapaan. Minna Jeffery pohtikin, että tekstien suomalaiset murreilmaukset käännettäisiin brittimurrepiirteiden keveällä sekoituksella, ei jollain valmiilla murteella. Näin saataisiin aikaan kokemus puhekielisyydestä ja paikallisuudesta, mutta kieli ei identifioituisi mihinkään olemassa olevaan paikkaan Britanniassa. Siinä säilyisi mahdollisuus myös suomalaiseen ilmaisuun ja tuntu siitä, että kirja avaa lukijalle ehkä aikaisemmin kokematonta kulttuuria.

Kaiken kaikkiaan kulttuurisen lukutaiton alla esitettävät käsitteet kurottelevat kaikkein syvimmälle kääntäjän tietotaidon hyödyntämiseen. Jos teksti käännetään vain sanoina, siitä katoaa kaikki taika ja paikallisuus, mitä hyvä kirjoittaja sinne ripottelee ja jonka hyvä kääntäjä osaa säilyttää.

4. Yhteenveto ja johtopäätökset

Kaunokirjallisen tekstin lähtökohta on toinen kuin vaikkapa mainostekstin, ohjeen tai muun selkeän asiatekstin. Kaunokirjallinen teksti on hiottu taideteos, jonka yksityiskohdilla on tarkkaan valitut merkityksensä. Sinnepäin ei ole tarpeeksi hyvä.

Tästä nousee laajan kielimallin käyttämisen suurin ongelma kaunokirjallisuuden suhteen: kääntämisen rajat ovat toisaalta hyvin tiukat, toisaalta niistä pitää osata välillä radikaalisti joustaa. Hyvä kääntäjä ei riko runon rytmiä tai käännä murteellista proosaa kuin makkaramainosta. Kaiken pohjana on kirjailijan työ, jota kunnioitetaan ja josta tehdään toiselle kielelle mahdollisimman hyvä versio. Kuten kääntäjämme Eri Lassila sanoi: ”Kääntäjä ei voi vaihtaa sisältöä, muotoa eikä tarkoitusta, jonka kirjailija on valinnut.” Silti tekstiä ei käännetä sanasta sanaan, vaan myös käännöskielen ominaisuuksia kunnioittaen.

Kun olen keskustellut tietoteknisten neuvonantajien kanssa laajalla kielimallilla kääntämisen ongelmista, esiin nousee taikasanan lailla promptaus eli kielimallin ohjeistus ja tehtävänannon rajaaminen. On mahdollista esimerkiksi promptata kielimallia pyytämällä käännöstä jonkun tunnetun henkilön tyyliä: ”Käännä runo William Shakespearen tyyliin.” Ohjeet voivat olla hienovaraisempiakin, voidaan pyytää lisää aistikkautta, puhekielisempää ilmausta jne. Yksi neuvonantaja mietiskeli, että ohjelman voisi promptata kääntämään tietyn henkilön vuorosanat tietyllä tyyliä, vaikka: ”Kirjoita vuorosanat pohjoissavolaisen, ujon ja alkoholisoituneen miehen puhekielellä”. Tällaisessa pyynnössä, ja promptauksessa ylipäätään, on eräs ongelmallinen seikka - tulkinta hahmosta olisi kääntäjän, millaiseksi hän hahmon näkee. Vaikka kääntäminen on aina tulkintaa, tässä mahdollisuus alkuperäisen tekstin ja tarkoituksen vääristymiseen on todella suuri.

Oikeastaan, jos kirjan kääntäminen vaatisi voimakasta kielimallin promptausta, kirjailijan pitäisi se tehdä tai ainakin häntä pitäisi konsultoida asiassa. Laajalla kielimallilla kääntämisen prosessin nopeus, helppous ja halpuus haihtuvat, jos kirjailijalta pitää laajasti ja toistuvasti varmistaa asioita. Lisäksi, onko kirjailijalle mielekästä ohjeistaa ohjelmaa, jos on olemassa asiantunteva kääntäjä, jonka kanssa tekstistä keskustelu on kahden kielen ja kirjallisuuden ammattilaisen välistä vuoropuhelua?

Voiko siis laajan kielimallin opettaa kääntämään kaunokirjallista tekstiä suomesta muille kielille? Uskon, että halutessaan kyllä voi. Ehkä se vaatii vielä muutaman loikan ohjelmistojen kehityksessä, varmasti se vaatii jonkun, joka opettaa laajalle kielimallille kunkin genren yleispiirteet ja tyylin variaatioita syöttämällä mallitekstejä ja promptaamalla. Mikäli mallitekstejä ei haluta varastaa – kuten useimmat ohjelmistoyritykset ovat tehneet – ja kouluttajalle ja ohjelmoijallekin halutaan maksaa työstä, tämä ei ole ilmaista. Lisäksi kielimallien käyttö kuluttaa sähköä ja erityisesti puhdasta vettä huomattavan paljon.

Onkin kyseenalaista, onko kunnollisen, kielen ja kulttuurin nyanssit ymmärtävän ”käännöskoneen” tekeminen kaunokirjallisuutta varten taloudellisesti kannattavaa. Erityisesti, kun mietimme, kuinka paljon ammattitaitoinen kääntäjä esimerkiksi kahdenkymmenen sivun käännösnäytteestä tai edes kokonaisen käsikirjoituksen kääntämisestä ansaitsee.

Nopeaan laajan kielimallin käännökseen ihastuvat asiakkaat saattavat ajatella, että kääntäjän työ olisi nopeampaa, kun pohjatyö on jo tehty. Saamani palautteen mukaan näin ei kuitenkaan ole, koska kaikki pitää kuitenkin tarkastaa. Usein työ on raskaampaa, yksitoikkoisempaa ja sitä jaksaa tehdä vähemmän kerrallaan. Minna Jeffery totesikin: ”On paljon työläämpää yrittää ymmärtää kielimallin tekemiä virheitä kuin lähteä alkuperäisestä tekstistä.” Työergonomian kärsiessä putoaa myös motivaatio ja hyviä kääntäjiä saattaa olla jatkossa vaikea löytää projekteihin, joissa hyödynnetään laajoja kielimalleja. Erityisesti, jos resurssia kääntämiselle annetaan vähemmän luullen, että ohjelmisto on jo tehnyt valtaosan työstä.

Laajan kielimallin käännöstä käytettäessä kääntäjän työ on siis vaarassa kaventua ja kielentarkistajaksi sekä tyylin parantelijaksi. Se on ironista, koska kielentarkistuksessa, tiedon etsinnässä ja vaihtoehtojen esittelyssä tekoälypohjaiset ohjelmat ovat hyviä. Ehkä kääntäjälle pitäisi sallia kielellä leikittely, luovuus ja kekseliäisyys - ja antaa ohjelmiston tehdä aputyöt?

5. Projektin jatko

Käännösten testaus laajan kielimallin avulla on ensimmäinen osa hanketta. Seuraavaksi lähдемme kohti auditiivista käännösnäytettä ja lopuksi laadimme lyhyet videomuotoiset esittelyt kirjailijoista.

Ajattelimme kuitenkin vielä testata keväällä 2026 sekä ChatGPT 5.0 niin sanotun lämpötilan (temperature) nostamista että DeepL-ohjelman käyttöä käännöksissä. DeepL on logiikaltaan hieman toisenlainen kuin ChatGPT ja sen sanotaan olevan hieman luovempi sanavalinnoissaan. Sen heikkous on puolestaan suurempi hallusinointi. Mitä tulee turvallisuuteen ja koulutuksen läpinäkyvyyteen – esimerkiksi mistä kaunokirjallisuuden koulutusmateriaali on otettu – DeepL ei ole eettisesti paras valinta. Tämän takia testi ei muuta tutkimuksen lopputulosta, mutta saattaa olla mielenkiintoinen jatkotutkimuksen kannalta.

Laajan kielimallin lämpötila on suure, jonka voi - mutkat hieman oikoen - sanoa kuvaavan ohjelmiston toiminnan asiallisuutta ja varmistelua. Normaaliasetuksilla esimerkiksi ChatGPT 5.0 toimii asteikolla 0-1 noin 0,3-0,4 asetuksilla. Tällöin ohjelma pyrkii melkein aina tiivistämään, selkeyttämään ja tekemään tekstistä loogisen. Kokeilemme kääntää ainakin osaa teksteistä niin sanotusti täydellä lämpötilalla eli testata, tuoko asetuksen muuttaminen muutosta esimerkiksi sanavalintoihin. Tälläkin testillä on vaikutusta korkeintaan tuleviin projekteihin.

Suurimmat kynnyskysymykset nimittäin eivät muutu miksikään: laajan kielimallin tuottama käännös ei ole sellaisenaan kielellisesti riittävän laadukas, se ei ole käännöskielen rakenteen mukainen, kulttuurisen kontekstin huomioon ottava teksti. Ja koska kääntäjän mahdollisuus promptaamiseen on myös kapeampi kuin asiatekstissä, siitä on vaikeaa myöskään kovin laadukasta tulla. Lisäksi täytyy muistaa tiheän promptaamisen nostamat energiakustannukset ja luonnonvarojen, erityisesti puhtaan veden, suuri käyttö. Kaikki tämä huomioon ottaen toteamme, että niin projektissamme kuin kääntämisessä ylipäätään vanhassa on vara parempi ja ekologisesti, eettisesti ja taloudellisesti kääntäjän työ on laadukkaampaa kuin laajan kielimallin käyttö.

Erityisesti, koska hankkeemme tarkoituksena on tuottaa laadukkaita käännöspaketteja myynninedistämistä varten, on tärkeää, että käännökset ovat ensiluokkaisia. Minna Jeffery arvioi, että yksikään laajan kielimallin tekemistä käännöksistä ei edes pienen parantelun jälkeen sopisi kustantajalle lähetettäväksi: ”He eivät ajattelisi, että tämä on huonosti käännetty, vaan että kirja on huonosti kirjoitettu.” Tällaista viestiä emme tietenkään halua hankkeen teksteistä välittää.

Selvityksen jälkeen olemme hankkeessa päätyneet käyttämään kääntäjien tekemiä käännöksiä osana tutkimuksemme seuraavia osia. Projektin seuraavaan osaan kuuluvat kolme 20 sivun mittaista tekstiä käännetään äänikäännösnäytettä varten ja kääntäjät saavat päättää, miten he käyttävät laajan kielimallin tekemiä pohjakäännöksiä: niitä voidaan käyttää vertailukohtana tai pohjana, ne voidaan myös ohittaa kokonaan. Valinta on kääntäjän.

Auditiivista käännösnäytettä varten olemme hankkineet kääntäjille ElevenLabs -ohjelmistot. ElevenLabs on tekstistä ääneksi ja äänestä tekstiksi operoiva ohjelma. Tarkoituksena on antaa kääntäjien kokeilla, mitä apua tällaisista ohjelmista tai laajasta kielimallista on kääntämisen prosessissa: sparraajina, apurina, tiedon hakijana jne. Koska olemme kaikki erilaisia, ohjelmiston käyttö saattaa vaihdella suuresti kääntäjästä toiseen. Laajan kielimallin ja AI:n rooli on siis kääntäjien työtä tukeva.