

PELASTAAKO TEKOÄLY SUOMEN KIELEN?

TEIVO TEIVAINEN

Onko suomi syrjäytymässä tutkimuksen kielenä, kun paineet englanniksi julkaisemiseen kasvavat? Teknologian harppaukset ovat usein vaikuttaneet historian kulkuun, ja ehkä niin käy tässäkin asiassa. Käännösalgoritmit avaavat uusia mahdollisuuksia sille, että voimme tulevaisuudessa julkaista nykyistä mielekkäämmin omalla kielellämme. Siksi yliopistoissakin on syytä kyseenalaistaa viime aikoina voimistunut ajatus, että englanti jyrää tulevaisuudessa yhä vääjäämättömämmin. Vertauskohtia tulevaisuuden käännöskoneiden merkitykseen voi etsiä esimerkiksi siitä, miten Mikael Agricolan reformoitu käsitys Jumalasta mahdollisti kielellistä tasa-arvoa.

Suomalaisessa nationalismissa omalla kielellä on ollut poikkeuksellisen tärkeä merkitys. Melko marginaalisesta kansankielestä tuli 1800-luvun aikana eliitinkin vähitellen käyttämä kirjakieli. Esimerkiksi Perussa olen kuullut hiukan huteraa kansallistunnettani hivelevää ihailua suomen kielen tarinaa kohtaan. Minulta on kaivattu näkemyksiä siihen, voisiko heidänkin maassaan koloniaalisen valtiokielen espanjan varjoon jäänyt intiaanikieli ketsua kokea joskus samankaltaisen arvonnousun. Tähän asiantuntemukseni ei tietenkään riitä, mutta samalla mielenkiintoni suomen kielen tulevaisuutta kohtaan on voimistunut.

Olen myös tullut tietoisemmaksi oman yliopiston roolin tärkeydestä suomen kielen tarinassa. Harvalla opinahjolla on maailmassa ollut yhtä suurin vaikutus kansallisen kielen kehitykseen ja siihen, mitä joskus kutsumme kansalliseksi sivistykseksi kuin työnantajallani Helsingin yliopistolla. Silloin kun Carl Ludvig Engelin suunnittelema yliopiston pääarakennus valmistui vuonna 1832, sanaa *sivistys* ei juurikaan tunnettu. Edellisellä vuosikymmenellä se oli vilahtanut *Turun Viikko-Sanomissa*. Sanan vakiinnutti kieleemme vasta hiukan myöhemmin yliopiston entinen suomen kielen professori Elias Lönnrot, joka muistetaan parhaiten Suomen kansalliseepoksen *Kalevalan* kokoajana.

Lönnrotin myöhemmin eläkepäivillään vuon-

na 1880 julkaisema *Suomalais-Ruotsalainen Sanakirja* vakiinnutti *sivistyksen* lisäksi sellaisia pian käytökelpoisia termejä kuten *itsenäisyys*, *kansallisuus*, *äänioikeus*, *kieltolaki* ja *tasavalta*.

Kun suomi alkoi Lönnrotin kaltaisten sanaseppojen myötä saada tukevampaa jalansijaa eliitin keskuudessa, kielikiistat kiihtyivät. Niitä käytiin myös Helsingin yliopistossa. Tieteen kielenä siellä käytettiin ruotsin ohella joskus latinaa, ja viime vuosisadan alkupuolella moni julkaisi tutkimuksiaan esimerkiksi saksaksi. Jotkut meistä julkaisevat nykyisin vaikkapa espanjaksi. Historiamme tunnetuimmat kielikiistat liittyvät kuitenkin suomen ja ruotsin asemaan.

Yliopistossa suomen kielen asemaa alettiin 1800-luvun puolivälin jälkeen korostaa hiukan keskenään poikkeavista lähtökohdista. Britanniasta vaikutteita saaneet liberaalit korostivat yksilönvapautta. Kansallinen vapaus olisi keino oman kielen vapaalle kehitykselle. Saksalaisesta filosofiasta vaikutteita saaneille Johan Vilhelm Snellmanin hengenheimolaisille yksilönvapaus oli usein sivuroolissa. He korostivat kansan yhtenäisyyttä. Eroista huolimatta molemmat tahot pitivät suomen kielen ja kansallisen itsemääräämisoikeuden vahvistamista tärkeänä.

Itsenäisyyden ja sisällissodan jälkeen kysymys ruotsin kielen asemasta jakoi eliitin mieli-

teitä. Tiukimmat suomalaisuusintoilijat taipuivat hyväksymään ruotsin maan toiseksi kansalliskieleksi. Vuoden 1937 yliopistolaki vahvisti kaksikielisen yliopiston suomenkielisyyttä. Sen jälkeen on otettu uusia askeleita kohti kielijakoa, jossa sujuvan kaksikielisyyden sijaan ruotsi on paljolti jäänyt suomenruotsalaisten erilliskieleksi.

Kaksikielisyyden merkitys tulee nykyisin yliopistossa vastaan esimerkiksi silloin, kun haluamme rekrytoida ulkomailla opiskelleita ja tieteellistä uraa tehneitä suomalaisia. Osa heistä on unohtanut kouluruotsin lähes kokonaan. Toisinaan tämä johtaa eriskummallisiin tilanteisiin.

Erään yliopistonlehtoraatin täytössä piti tehdä päätös paljon julkaisseen ja montaa kieltä sujuvasti osaavan suomalaishakijan sekä vähemmän julkaisseen ja lähinnä vain englantia puhuvan ulkomaalaisen kilpakumppanin välillä. Edellinen oli useimmilla mittareilla pätevämpi.

Pätevämpää hakijaa ei kuitenkaan voitu valita, sillä hän oli menestyksekkään ulkomaanuran aikana paljolti unohtanut ruotsin. Kilpahakijaakaan ei osannut ruotsia, eikä myöskään suomea, mutta ulkomaalaisena kaksikielisyyden vaatimus ei koskenut häntä. Jahkailun jälkeen koko hakuprosessi raukesi.

Enimmäkseen ruotsin kielen asemasta yliopiston sisällä keskustellaan nykyisin sulassa sovussa. Puhe ruotsin pakollisuuden höllentämisestä on kuitenkin voimistunut myös yliopiston ulkopuolisessa Suomessa. Ruotsinkielisten lukumäärän asteittaisen vähenemisen lisäksi tärkeänä perusteluna mainitaan toisinaan kansainvälistyminen.

Nyt yliopistossa kytee uusi kielikiista. Moni tuntee huolta siitä, että englantia syrjäyttää suomen korkeakoulutuksen pääkielenä. Paineet englanniksi opettamiseen ja etenkin julkaisemiseen kasvavat. Tieteessä on yhä tärkeämpää voida kommunikoida maailman eri kolkissa olevien kollegoiden kanssa. Kansainvälistymisen kolikon kääntöpuoli on se, että samalla suomi on vaarassa hiipua korkeakoulutuksen kielenä. Englanninkielisyyttä korostavien tulostittareiden näkökulmasta tällaisenkin artikkelin kirjoittamista suomeksi voi pitää lähes arvottomana.

Suomen puhujien määrä on riittävän suuri, että pelkoa sen täydellisestä katoamisesta ei ole näköpiirissä. Lisäksi maailman sähköisessä vies-

tinnässä Suomi on puhujiensa lukumäärää suurempi kieli. Vauraus tuo digitaalisia etuoikeuksia. Professori Janne Saarikiveltä olen oppinut, että esimerkiksi kurdin tai uiguurin puhujia on selvästi enemmän kuin suomalaisia, mutta kirjallisena tai tietoverkoissa liikkuvana kielenä ne ovat selvästi pienempiä.

Voisiko suomesta silti tulla eräänlainen kyökikieli? Olemmeko kulkemassa kohti tulevaisuutta, jossa korkeakoulutettu eliitti keskustelee yhä enemmän termeillä, joille ei luoda suomenkielisiä vastineita? Voisiko suomesta joskus tulla hiukan kuin guaraní, jota moni paraguaylainen saattaa puhua useissa arkipäivän tilanteissa? Koulutusjärjestelmässä ja virallisissa yhteyksissä espanja on kuitenkin Paraguayn selkeä pääkieli.

En ole koskaan pitänyt itseäni kovin teknologiauskoisena. Viime vuosina päätäni on kuitenkin vaivannut uusi kysymys. Voisiko Google Translate tai jokin muu tekoälyä käyttävä konekääntäjä pelastaa suomen korkeakoulujen käyttökielenä? Ensireaktio viime vuosina joskus esittämäni kysymykseen on yleensä hämmentynyt. Joku räplää hetken kännykkäänsä ja näyttää hätäisen käännösoperaation tuloksen. Eihän Google Translate nyt todellakaan käännä riittävän hyvin. Ei kai kielen tulevaisuus voi riippua teknologiafirmojen tuotteista?

Käännösohjelmien kehittäminen perustuu osittain saatavilla olevan kielimassan suuruuteen. Maailmankielten rinnalla suomi on kaikesta huolimatta suhteellisen pieni. Vaikka sen rakenteen johdonmukaisuus voi olla kielikoneiden kannalta joissain tilanteissa myös etu, suomi poikkeaa useimmista muista kielistä. Niinpä Google Translate ei ole vielä oppinut tulkitsemaan sitä erityisen hyvin. Tämä vaikuttaa siihen, että yleiset mielikuvat koneellisen kääntämisen tasosta eivät ole Suomessa kovinkaan haltioituneita. Silti ohjelmien algoritmit ovat kehittyneet valtavasti viime vuosina. Mikä mahtaa olla tilanne vaikkapa kymmenen vuoden kuluttua? Entä tämän vuosisadan jälkipuoliskolla?

Käännös on aina tulkinta. Vaatimusta täydellisestä vastaavuudesta kahden kielen välillä ei ole järkevää edes esittää. Koneilta ei myöskään voi olettaa kaikkien kaunokirjallisten koukeroiden tai käyttöyhteyksien hallintaa. Koneellisen kääntämisen keskusteluissa yksi toisinaan korostettu

esimerkki on se, miten koneilta puuttuu ironian taju. Täydellisyys ei kuitenkaan ole yhtä kuin riittävän hyvä.

On aivan mahdollista, että tulevaisuudessa koneet kykenevät kääntämään tieteellistä tekstiä niin hyvin, että ne auttavat murtamaan monia kielimuureja. Seurauksena saattaa olla, että yliopistoni tulevat tutkijat voivat viedä ajatuksensa kansainväliseen keskusteluun parhaiten hallitsemallaan kielellä. Lukijoiden apuna ovat algoritmit kääntävät kirjoituksen salamannopeasti heidän omille kielilleen.

Vaikka villeimmät visiot käännöskoneiden tulevasta tehokkuudesta eivät toteutuisi, koneet helpottavat ihmiskääntäjien työtä. Sekin vie kohti tilannetta, jossa suomenkielinen tieteellinen artikkeli kääntyy aiempaa vaivattomammin muille kielille. Suomeksi kirjoittaminen siis muuttuu tältä osin joka tapauksessa helpommin käännettäväksi ja samalla mielekkäämmäksi. Siksi yliopistoissakin on syytä kyseenalaistaa viime aikoina voimistunut ajatus, että englantia jyrää tulevaisuudessa yhä vääjäämättömämmin.

Toki on olemassa myös mahdollisuus, että siten kun teknologia saavuttaa tason, jolla kääntäminen muuttuu merkittävästi helpommaksi, akateeminen eliitti on jo sopeutunut englanniksi kirjoittamiseen niin hyvin, että suomeksi kirjoittaminen tuntuu hankalalta tai rahvaanomaiselta. Kielellinen identiteetti voi ottaa kulttuurisia harppauksia, jotka vaimentavat teknologisten harppausten avaamia mahdollisuuksia. Siksikin suomen kielen ystävien on hyvä keskustella tulevaisuutta ennakoiden, jotta vääjäämättömyysretoriikka ei entisestään vahvista englannin ylivaltaa.

Käännösteknologia tulee mukaan myös puheen tulkkaukseen. Ihmisten tarve vieraiden kielten opetteluun saattaa sen myötä vähentyä. Yksi versio tällaisesta laitteesta on kehitetty Japanissa. Sen kaupallinen nimi *Megahonyaku* on japanin uudissan, joka yhdistää megafonin ja kääntämisen.

Vielä pari vuotta sitten Megahonyakun valmistaja Panasonic toivoi tuotteensa saavan kansainvälistä näkyvyyttä viimeistään Tokion olympialaisissa vuonna 2020. Itsekin odottelin mielenkiinnolla laitteen mahdollisesti herättämää keskustelua, mutta koronakriisin myötä tämä haave on nyt hiukan lykkääntynyt. Megahonyaku on aluksi kyen-

nyt vain helppoihin tehtäviin kuten tulkkaukseen, että ”sushi on loppu”. Panasonic aikoo kuitenkin kytkeä laitteen aiempaa tiiviimmin pilvipalveluiden kautta toimiviin tehokkaisiin käännösohjelmiin. Samankaltaista teknologiaa ohjelmoidaan yhä enemmän myös megafonia helpommin mukana kulkeviin laitteisiin.

Konekääntämisen maailmanpolitiikkaa

Kirjoituksen otsikossa esittämälleni kysymykselle suomen kielen kohtalosta tärkeintä on se, miten hyvin koneiden algoritmit kykenevät murtamaan käytännön kielimuureja. Todennäköisesti niin hyvin, että meidän on nyt vaikea kuvitella kaikkia niihin liittyviä uhkia ja mahdollisuuksia. Kielestä puhuttaessa ei tässäkään yhteydessä ole kyse vain kielestä.

Esimerkiksi pitkään tutkimieni poikkikansallisesti verkostoituneiden kansalaisliikkeiden keskuudessa konekääntämistä käytetään aiempaa enemmän helpottamaan eri kieliryhmien välistä viestintää. Seurauksena on tietenkin myös harmillisia väärinymmärryksiä, kun Google Translaten tulkinta sähköpostista, WhatsApp-ryhmän kommentista tai mielenosoituskutsun perusteluista menee metsään eikä käsillä ole algoritmin tuottamaa raakiletta korjaavaa ihmiskääntäjää.

Kaikkine puutteineenkin koneet auttavat rikokomaan kielikuntien välisiä rajoja. Samalla ne avaavat kysymyksiä maailmanhallinnasta ja ihmisen olemuksesta. Kansainvälisen tekoälytutkimuksen kärkinimiin kuulunut äskettäin edesmennyt professorikollegani Timo Honkela korosti intohimoisesti tekoälyn avaamia mahdollisuuksia. Hän ehti kehittää huikeita visioita kaavailemastaan Rauhankoneesta. Sen algoritmien olisi tarkoitus auttaa maailmanpoliittisten konfliktien osapuolia ymmärtämään toisiaan. Minun on vaikea kuvitella, että nimenomaan algoritmien avulla voisi löytää tyydyttävää ratkaisua esimerkiksi kamppailuihin Israelin ja Palestiinan alueen hallinnasta. Kielellisen ja maailmankuvallisen keskinäisen ymmärryksen lisääminen voi silti tuoda uusia välineitä myös rauhantyöhön.

Toiset puolestaan korostavat tekoälyn uhkakuva. Osa uumoilee jopa koko ihmiskunnan tuhoa, kun koneet pääsevät niskan päälle. Tulevaisuudentutkijoiden piirissä kuulee joskus puhetta

teknologisesta singulariteetista. Fysiikan tieteenalalta lainattu käsite viittaa uudiskäytössään usein tilanteeseen, jossa itseään kehittävä tekoäly astuu uuteen ja arvaamattomaan vaiheeseen. Oma ymmärrykseni ei yllä tämän kiistellyn käsitteen hyödyllisyyden arviointiin, mutta vaikka pitäisi tällaista tulevaisuutta vain tieteisfantasiana, tekoäly saattaa muuttaa maailmaa rajusti lähivuosisikymmeninä.

Suuryritysten hallitsemat kielikoneet pääsevät käsiksi yhä suurempaan osaan maailman teksteistä. Monelle älypuhelimien käyttäjälle saattaa tulla yllätyksenä, missä määrin ja millaisissa tilanteissa pöydällä tai taskussa näennäisen toimittomana lojuva pienoistietokone voi nauhoittaa puhetta. Käännösohjelmien kehittäjät tarvitsevat suuria teksti- ja puhemassoja. Niiden hankkimisen tarkoitusperät on yleensä muotoiltu mahdollisimman viattomiksi. Seurauksena voi silti olla uudentyypinen algoritmi-imperialismi.

Kun hakukoneet, käännöskoneet ja erilaiset digitaaliset seurantajärjestelmät kytkeytyvät toisiinsa, syntyy aineksia ennenkuulumattoman tehokkaalle valvontakoneistolle. Google ja sen emoyhtiö Alphabet ovat jo nyt nousseet merkittävään asemaan tietoverkkojen hallinnassa. Erityisesti ranskalaisissa keskusteluissa alkoi 2010-luvulla yleistyä termi GAFa. Se viittaa Googlen, Applen, Facebookin ja Amazonin kaltaisiin yrityksiin. Niiden rinnalle nousee varmasti myös uusia tekoälyä käyttäviä suuryrityksiä, jotka vaikuttavat maailmanpolitiikan virtauksiin ja arjen käytäntöihin.

Kirjoittaisimmeko siis suomeksi?

Yhtäältä käännöskoneet voivat vapauttaa suomalaiset käyttämään kieltään vailla pelkoa impivaaralaisesta eristymisestä. Samalla ne saattavat sekä yksinkertaistaa kielellisiä ilmauksia että tukahduttaa muita vapauksia. Kun tietoverkkojen hallinta keskittyy voittoa tavoitteleville liikeyrityksille, käsitteet demokratiasta ja yksityisyydestä muuttuvat. Siihen on syytä varautua.

Ehkäpä sellaiset Lönnrotin kehittämät termit kuin tasavalta ja äänioikeus kadottavat alkuperäisen tarkoituksensa. Sen sijaan kielikoneiden muokkaamassa maailmassa Lönnrotin luoma termi kansallisuus saattaa säilyttää kielinationalistisen merkityksensä yllättävän hyvin. Suomen kieli

muuttuu, niin kuin se on aina tehnyt, mutta se ei ole katoamassa.

Tekoälyn tulevaisuutta voi havainnollistaa muistelemalla mennyttä. Kuljetaan siis hetkeksi kirjasuomen varhaisuurille. Kun Mikael Agricola antoi 1500-luvulla raamatunkäännöksellään alkusykäyksen suomen kirjakiellelle, hän vetosi kaikkivaltiaan kielitaitoon. Siihen aikaan moni ajatteli, että kristittyjen Jumalaa tulee puhutella latinaksi. Agricola oli toista mieltä. Yksi hänen elämäntyönsä perustelu oli, että ”kyllä se kuulee suomen kielen, joka ymmärtää kaikkein mielen”. Luterilaisen reformaation puolivuosituhantista taivalta muisteltaessa on hyvä ottaa huomioon, miten Agricolan reformoitu Jumala mahdollisti eräänlaisen kielellisen tasa-arvon.

Missä määrin konekääntäjä voisi vastaavasti kuulla suomen kielen? Voisiko yhdistelmästä algoritmeja tulla hiukan samankaltainen kielellinen kaikkivaltias kuin Agricolan kuvaamasta Jumalasta? Itselleni jumalvertaukset sisältävät hippusellisen hölynpölyä, mutta uskonnollisen vertauskuvan käännösalgoritmien taidoille voi löytää myös ajallisesti lähempää.

Kun kävin 1980-luvulla krakovalaisessa katedraalissa, puolalaiset katolilaiset pyysivät minulta vieraskirjaan tervehdystä paaville. Kysyin, millä kielellä he haluavat minun kirjoittavan. He pitivät kysymystäni outona, sillä tokihan Vatikaanissa istuva puolalainen paavi ymmärtäisi kaikkia kieliä. ”Kirjoita siis suomeksi”, he kehottivat. On aivan mahdollista, että sama kehoitus vielä palaa myös suomalaisten korkeakoulujen kielipolitiikkaan. Siis kirjoitamme ehkä jatkossakin suomeksi.

Kirjoittaja on maailmanpolitiikan professori Helsingin yliopistossa. Kirjassaan *Maailmanpoliittinen kansalliskävely* (2018) hän on käsitellyt Suomen historiaa ja tulevaisuutta maailman talousinstituutioiden, kielten ja koloniaalisten valtasuhteiden valossa. Tämä artikkeli perustuu osittain tuohon kirjaan.

Twitter: @teivoteivainen.