

Konekäännös: mitä sillä tehdään?

Konekääntäminen on teknologia, jonka on jo 50 vuoden ajan ennustettu tekevän kääntäjät ja kielten opiskelun tarpeettomaksi "seuraavan kymmenen vuoden sisällä". Konekääntämisen historiaan kuuluu paljon lunastamattomia, mutta yhä enemmän myös lunastettuja lupauksia. Viime aikoina konekääntäminen on kehittynyt suurin harppauksin. Vaikka konekäännösten laatu on selvästi ammattilaiskääntäjien laatua heikompaa, silti maailman kaikista käännöksistä yli 99% tehdään nykyään konekääntimillä. Ammattilaiskääntäjät tekevät siis vain alle prosentin maailman kaikista käännöksistä. Ammattilaiskääntäjien ja tulkkien tarve kasvaa huomattavan nopeasti, mutta konekäännösten tarve kasvaa räjähdysmäisesti.

Julkaistu: 14. joulukuuta 2017 | Kirjoittanut: Niko Papula

Suomi on merkittävästi jäljessä enemmän puhutuista kielistä

Konekäännöksen tasosta ja hyödyllisyydestä puhuttaessa on erittäin tärkeää erottaa suomi ja yleisimmin puhutut, niin sanotut "suuret" eurooppalaiset kielet. "Suurten" kielten konekäännös on olennaisesti paremmalla tasolla kuin suomen useista ymmärrettävistä syistä. Ensinnäkin suomi on rakenteeltaan hyvin erilainen verrattuna useimpiin muihin kieliin. Esimerkiksi germaaniset kielet norja, ruotsi, englanti ja saksa ovat huomattavasti lähempänä toisiaan, ja siksi konekääntäminen niiden välillä on helpompaa. Toiseksi konekäännöstä kehitetään nykyään opettamalla konekäännösjärjestelmä kaksikielisen aineiston avulla. Tällaisessa aineistossa on tyypillisesti miljoonia käännöspareja, ts. sama virke kahdella eri kielellä. Englantia tai espanjaa puhuvia on satoja miljoonia, suomea puhuvia viitisen miljoonaa. On selvää, että englannin ja espanjan välillä on olemassa ja saatavilla huomattavasti enemmän käännöspareja kuin esimerkiksi suomen ja espanjan välillä. Kolmanneksi kielen puhujien määrä vaikuttaa myös kyseisen kielen tutkimukseen ja konekäännökseen käytettävissä olevien resurssien määrään. Espanjan ja englannin konekäännöksiä tutkii ja kehittää paljon runsaslukuisempi joukko kuin suomen konekäännöksiä.

Konekäännöksen kehityksessä suomen kieli on vähintään kuusi vuotta suuria eurooppalaisia kieliä jäljessä, mikä antaa meille erinomaisen mahdollisuuden ennakoita tulevaa kehitystä – katsomalla mallia kehityksessä suomea edellä olevista kielistä. Seuraavassa esittelen konekääntämistä suurten eurooppalaisten kielten näkökulmasta. Lukijan tehtäväksi jää arvioida, seuraako kehitys Suomessa muiden maiden latuja.

Tekstin konekäännös suurilla eurooppalaisilla kielillä

Konekääntämisen määrät ovat aivan valtavia ja yli satakertaiset ammattilaiskäännösten määrään verrattuna (Mesropy, 2017). Konekäännös on siis jo nyt valtavirtaa merkittävistä puutteistaan huolimatta.

Konekäännöksen avulla käännettyjen tekstien laatu on useimmiten olennaisesti huonompi kuin ammattilaiskääntäjän laatu. Tästä johtuen konekäännettyjä tekstejä julkaistaan varsin vähän sellaisenaan. Konekäännöksen tyypillisin käyttötilanne onkin kääntäminen omaan tarpeeseen, ts. vieraskielisestä tekstistä omalle kielelle.

Konekäännöstä käytetään paljon ammattilaiskääntäjän apuna. Konekäännösten laatu vaihtelee virkkeittäin. Osa koneen tekemistä raakakäännöksistä voi olla täydellisiä, osa taas täysin virheellisiä. Ammattilaiskääntäjä tarkistaa ja korjaa konekäännöksen tekemän raakakäännöksen. Tällä työskentelytavalla saavutetaan sama laatu kuin perinteisellä kääntämisellä, ja samalla ammattilaiskääntäjän työskentelynopeus kasvaa olennaisesti, usein 40% tai jopa enemmän verrattuna kääntämiseen perinteisellä tavalla kokonaan ilman konekäännöstä (Kubovsky & Szczepaniak, 2015).

Konekäännöksen laatu riippuu olennaisesti käännettävästä tekstistä. Voidaan puhua "tekstispektristä", johon jokainen teksti sijoittuu eri kohdalle sen mukaan, kuinka vaikea se on konekäännökselle. Tekstispektrin helpompaan päähän sijoittuvat niin sanotut "rajoitetut kielet". Tällä tarkoitetaan kapean aihealueen tekstejä, joiden kirjoittamisessa käytetään vain suppeata sanastoa ja myös suppeaa valikoimaa eri lauserakenteita. Kun konekäännin on optimoitu kääntämään kyseistä rajoitettua kieltä, konekäännös tekee parhaimmillaan yhtä vähän virheitä kuin keskimääräinen ammattilaiskääntäjä (METEO System, 2017).

Tekstispektrin toiseen, vaikeampaan päähän sijoittuvat esimerkiksi runous ja kaunokirjallisuus. Kaunokirjallisuudessa ollaan vielä kaukana siitä, että konekäännöstä voitaisiin hyödyntää edes raakakäännöksenä ammattilaiskääntäjän apuna (Taivalkoski-Shilov, 2017).

Konekäännöksen laadussa kaikki tekstit sijoittuvat edellä mainittujen tekstispektrin ääripäiden väliin. Tekstispektri on erilainen jokaiselle kieliparille ja samoin konekäännöksen laatu on erilainen jokaiselle eri kieliparille. Lähellä toisiaan olevilla kieliparilla (kuten espanja ja katalaani) konekäännöstä käytetään sanomalehtien kääntämiseen siten, että raakakäännös julkaistaan sellaisenaan (La Vanguardia ... , 2016). Esimerkiksi suomen ja englannin välillä konekäännös on vielä hyvin kaukana tästä tasosta.

Palvelujen ja tuotteiden vaatima tekninen käyttötuki voidaan suurilla eurooppalaisilla kielillä usein hoitaa konekäännöksen avulla. Tällöin teknistä käyttötukea tarjoava henkilö osaa itse tyypillisesti englantia ja hän voi hoitaa teknisiä tukitapauksia koskevaa kirjeenvaihtoa konekäännöksen avulla esimerkiksi espanjaksi, portugaliksi, italiaksi ja ranskaksi. Näin on varsinkin

tilanteissa, joissa teknisen käyttötuen ratkaistavaksi tulevat ongelmat ovat useimmiten saman tyyppisiä ja suhteellisen yksinkertaisia.

Facebook esittää konekäännettyjä kommentteja sellaisenaan. Myös tuotteiden ja palvelujen arvostelujen konekäännöksiä julkaistaan sellaisenaan. Näissä käyttötilanteissa huono konekäännös ei useimmiten aiheuta juurikaan muuta haittaa kuin ainoastaan jää ymmärtämättä. Hyvä konekäännös taas antaa lukijalle hyödyllistä tietoa. Muutenkin hyvin suuri osa konekäännöksen käytöstä perustuu siihen, että huonon konekäännöksen aiheuttama haitta on pieni ja hyvän konekäännöksen tuoma hyöty on suuri. Tämä tietysti vaikuttaa olennaisesti tilanteisiin, joissa konekäännöstä voidaan käyttää.

Kuten edeltä ilmenee, konekäännöksen laadusta puhuttaessa ei ole kovin relevanttia kysyä "kuinka hyvin konekääntäminen toimii", vaan kuinka hyvin konekääntäminen toimii tietyille tekstile ja tietyille kielelle käännettäessä. Kaikilla kielillä rajoitetun kielen konekääntäminen on lähimpänä ammattilaiskääntäjää. Kaikilla kielillä runouden konekääntäminen on mahdotonta. Konekääntämisen laatu on aina tapauskohtainen ja käännettävä teksti vaikuttaa siihen olennaisesti.

Tällä hetkellä ja lähitulevaisuudessa konekääntäminen ei ole vähentämässä kääntäjien kysyntää tai työpaikkoja. Konekäännöksen laadun merkittävästä parantumisesta huolimatta ammattilaiskääntäjien tarve kasvaa jatkuvasti. Esimerkiksi Yhdysvalloissa kääntäjien ja tulkkien tarpeen ennustetaan kasvavan 17% kymmenen vuoden kuluessa (Bureau of Labor Statistics, U.S. Department of Labor, 2017). Kääntäjät ja tulkit ovat kahdenkymmenen nopeiten kasvavan ammattinimikkeen joukossa. Käännösten tarpeen suuri kasvu johtuu kansainvälisen kaupan ja kontaktien lisääntymisestä. Nykyisellään konekäännös ei niinkään korvaa ammattilaiskääntäjiä, vaan tehostaa ja täydentää näiden työtä. Suuri osa konekäännöksistä tehdään tilanteissa, joissa ammattilaiskääntäjien käytölle ei esimerkiksi kustannus- tai aikataulusyistä ole mahdollisuuksia. On olemassa myös esimerkkejä, joissa konekääntämisen käyttö jopa kasvattaa ammattilaiskääntäjien kysyntää.

Mielenkiintoinen kysymys on, missä määrin konekäännöksen käyttäminen siirtää työpaikkoja maasta toiseen (Sajari, 2017). Edellä on mainittu todellinen tilanne, jossa teknisen käyttötuen työpaikka on jo nyt mahdollista ja usein kannattavaa siirtää toiseen maahan esimerkiksi halvemman työvoiman vuoksi.

Puheen kääntäminen eli tulkkaus

Viime aikoina on markkinoille tullut ratkaisuja tulkkaukseen eli puheen kääntämiseen reaaliaikaisesti toiselle kielelle. Esimerkiksi Skype tarjoaa kaikille käyttäjilleen puhelujen kääntämistä reaaliaikaisesti kahdeksan hyvin yleisesti puhutun kielen välillä. Samoin on olemassa kännykässä toimivia ratkaisuja, joilla voidaan tulkata esimerkiksi kaupassa käytäviä keskusteluja.

Nämä ratkaisut toimivat toistaiseksi varsin huonosti, vaikka ovatkin useimmiten parempia kuin ei mitään ratkaisua.

Puheentunnistuksen ja konekääntämisen virheet itse asiassa kertautuvat eli koneellinen tulkkaus tekee enemmän virheitä kuin erillinen puheentunnistus ja konekääntäminen yhteensä. Siitä huolimatta monessa tilanteessa huonokin käännös on parempi kuin ei lainkaan käännöstä. Ja teknologia kehittyy nopeasti. Voidaan kysyä, missä vaiheessa koneellinen tulkkaus on parempi kuin huono "turistiespanja"?

Kuvassa olevan tekstin kääntäminen

Markkinoille on tullut myös ratkaisuja, joiden avulla voi ottaa valokuvan esimerkiksi opaskyltistä tai ravintolan ruokalistasta ja saada tekstin käännettynä omalle kielelleen. Kuten koneellinen tulkkaus, tämäkin on varsin uusi teknologia ja vielä varsin epäluotettava. Tässäkin teknologiassa virheet kertautuvat. Kuvassa olevan tekstin tunnistuksessa tapahtuvat virheet tekevät konekääntimen tehtävän entistäkin vaikeammaksi. Virhemäärä on kokonaisuudessaan suurempi kuin tekstin tunnistuksen ja konekäännöksen virheet yhteenlaskettuna.

Puutteistaan huolimatta tämäkin teknologia löytänee käyttäjäkuntansa. Jälleen kerran huonokin käännös on usein parempi kuin ei lainkaan käännöstä. Milloin näiden käännösten taso on parempi kuin heikko kielitaito?

Pari ajatusta Suomen tilanteesta

On vaikea nähdä, miksi kehitys Suomessa ei seuraisi muiden maiden kehitystä. Meillä on kuitenkin yksi erityispiirre: englannin kielen opiskelu on erittäin yleistä ja englannin kielen taito on usein parempi kuin monessa muussa maassa. Ja kuten edellä mainitsin, englanti on kieli, jolla konekäännös toimii lähes aina parhaiten.

Kun suomen kielen konekäännös toimii usein erittäin huonosti, suomalaiset osaavat hyvin englantia ja englannin kielen konekäännös toimii parhaiten, niin onko itse asiassa parasta käyttää konekäännöstä englannin kielen avulla? Onko suomalaisen helpompaa kommunikoida espanjankielisen kanssa käyttämällä huonosti espanja-suomi-kieliparilla toimivaa konekäännöstä? Vai onko suomalaisen parempi hyödyntää englannin kielen taitoaan ja käyttää englanti-espanja-kieliparilla huomattavasti paremmin toimivaa konekäännöstä?

Useimmille harvinaisillekin kielille on saatavilla konekäännös englannin kanssa. Toisaalta useimmille harvinaisille kielille ei ole saatavilla konekäännöstä suomen kanssa. Ainakin harvinaisten kielten kohdalla monen suomalaisen lienee usein helpointa kommunikoida konekäännöksellä englannin kielen avulla.

Englannin kielen vahva asema konekäännöksessä korostanee englannin kielen merkitystä ja tarvetta. Milloin on odotettavissa tilanteita, jossa hyvä englannin kielen taito yhdistettynä konekäännökseen espanjaksi on parempi kuin heikko espanjan taito?

Johtaako konekäännöksen kehitys jossakin vaiheessa siihen, että on parempi osata useita kieliä kohtuullisesti vai hyvin yhtä sellaista kieltä, jolla konekäännös toimii hyvin?

Niko Papula on kieliteknologiaan erikoistuneen ohjelmistoyritys Multilizer:in toimitusjohtaja ja alan yhdistyksen KITES:in varapuheenjohtaja. KITES on Suomen kieliklusterin eli alan yritysten, ihmisten, järjestöjen ja yliopistojen yhteistyöorganisaatio. Niko Papulalla ja Multilizer:lla on yli 10 vuoden kokemus konekääntämisestä. Multilizer:in tuotteilla on yli kaksi miljoonaa käyttäjää yli 130 maassa.

Lähteet

Bureau of Labor Statistics, U.S. Department of Labor (2017). [Occupational Outlook Handbook, Interpreters and Translators \(2016\)](#). Viitattu 9.11.2017.

Kubovsky, J. & Szczepaniak, J. (2015). [How to quadruple localization productivity with MT postediting](#). The MemSource Blog. Viitattu 9.11.2017.

[La Vanguardia celebrates 5th anniversary of Catalan edition](#) (2016). Lucy Software and Services GmbH. Viitattu 9.11.2017.

Mesropyan, V. (2017). [Machine Translation vs. Human Translation](#). The MemriseBlog. Viitattu 9.11.2017.

[METEO System](#) (2017). Wikipedia. Viitattu 9.11.2017.

Sajari, P. (2017). [Taloustieteilijä Richard Baldwin varoittaa: Suomen kieli ei pian enää suojaa palvelualojen työpaikkoja – edessä on ”virtuaalinen maahanmuutto”](#). *Helsingin Sanomat* 26.8.2017.

Taivalkoski-Shilov, K. (2017). [Ihmisen korvaamattomuudesta kaunokirjallisuuden kääntämisessä](#). Suomen kääntäjien ja tulkkien liitto. Viitattu 9.11.2017.