

From: noreply@lausuntopalvelu.fi
Sent: 27. huhtikuuta 2020 12:35
To: oikeusministerio_om.fi
Subject: Lausuntopalvelu.fi: Uusi lausunto jätetty
Attachments: 20200427_VN-8830-2019_76502.pdf

Uusi lausunto lausuntopyyntöön 'Lausuntopyyntö selvityksestä usean kielen merkitsemisestä väestötietojärjestelmään'

Lausunnonantaja: Keskitalo Esa-Pekka

Organisaatio: Kansalliskirjasto, Kotoistushanke

Voit tarkastella lausuntoa linkin kautta:

<https://www.lausuntopalvelu.fi/FI/Proposal/GetReportViewForResponse?proposalId=d6647485-7b09-44a8-bd82-faf3eb5497f2&responseId=77dc9c63-3a59-48e7-b9b6-55aedbdd395f>

Asia: VN/8830/2019

Lausuntopyyntö selvityksestä usean kielen merkitsemisestä väestötietojärjestelmään

Lausunnonantajan lausunto

Selvityksessä on arvioitu vaikutuksia kieliryhmiin. Mitä muita mahdollisia vaikutuksia usean kielen merkitsemisellä väestötietojärjestelmään olisi kieliryhmiin. Pyydämme teitä erottelemaan vastauksessanne malli A:n ja malli B:n arvioidut vaikutukset.

Suomen väestötietojärjestelmään voi toistaiseksi merkitä vain yhden äidinkielen sekä asiointikieleksi joko suomen tai ruotsin. Kuten selvityksessä todetaan, järjestelmä ei sen vuoksi anna kattavaa kuvaa suomalaisten puhumista kielistä. Kasvavan maahanmuuton ansiosta suomalaisten kielellinen osaaminen on kasvanut nopeasti, mutta väestötietojärjestelmä antaa tästä vain kalpean aavistuksen, ja ellei järjestelmää kehitetä, ero todellisen ja raportoidun kielitaidon välillä tulee jatkossa vain kasvamaan.

Selvityksessä ehdotetaan kahta ratkaisumallia tilanteen parantamiseksi. Vaihtoehto A on usean äidinkielen merkitseminen siten, että asiointikieleksi valitaan edelleen aina joko suomi tai ruotsi. Vaihtoehto B on yhden äidinkielen ja asiointikielen merkitseminen entiseen tapaan, mutta väestötietojärjestelmään lisätään mahdollisuus lisätä yksi tai useampi kotikieli.

Asiointikielen merkitseminen säilyy molemmissa vaihtoehtoissa samana, mutta useiden kielten tallennusmahdollisuus varmistaa sen, että väestötietojärjestelmään voidaan tallentaa kielitaitoa koskeva tieto kattavasti, mutta vain jos tallennuksessa voidaan tarvittaessa käyttää ISO 639-3 –standardin kattavaa kielikoodistoa, ja kaikki asianosaiset osaavat soveltaa ISO 639 –standardin eri osien kielten, makrokielten ja kieliperheiden koodeja asianmukaisesti.

Kotoistushanke kannattaa lämpimästi väestötietojärjestelmän kehittämistä siten, että se antaa kattavamman kuvan suomalaisten kielitaidosta. Tämä on kuitenkin tehtävä niin, että mahdolliset kielteiset sivuvaikutukset minimoidaan.

Jos äidinkieliä olisi useita, jokin niistä pitäisi voida rekisteröidä ensisijaiseksi, ja väestötietoja hyödyntävien tahojen pitäisi pystyä ottamaan tämä huomioon. Virheiden mahdollisuus olisi tässä vaihtoehdossa suurempi kuin silloin jos äidinkieliä on vain yksi. Mutta jos kaksikielinen suomenruotsalainen henkilö ilmoittaa äidinkielekseen uudistetussa väestötietojärjestelmässä ruotsin ja kotikielekseen suomen, tämä ei vaikuta kuntalaisten kielijakaumaan, sillä se tilastoidaan äidinkielen perusteella.

Selvityksessä on arvioitu vaikutuksia viranomaisten toimintaan. Mitä muita mahdollisia vaikutuksia usean kielen merkitsemisellä väestötietojärjestelmään olisi viranomaisille. Pyydämme teitä erottelemaan vastauksessanne malli A:n ja mallin B:n arvioidut vaikutukset.

Merkittävä kotikielitiedon soveltamisalue on peruskouluissa annettava äidinkielen tai kotikielen opetus. Esimerkiksi venäjää tai thaita kotikielenään puhuvat voivat pyytää sen opetusta, mutta kunnat eivät saa väestötietojärjestelmästä ennakoon lukuja siitä, miten moni lapsi saattaisi tarvita minkäkin kielen opetusta. Lapsen elämässä voi olla useita kieliä, joista pitäisi valita kotikieleksi rekisteröitävä tai rekisteröitävät. He voivat puhua kunkin vanhemman kanssa tämän ensisijaista kieltä, vaikka perheen yhteinen kieli olisi jokin kolmas kieli (usein englanti, joskus suomi tai ruotsi). Kullekin perheenjäsenelle rekisteröitävien kielien valinta ei tuolloin välttämättä suju heidän etunsa mukaisesti ilman hyvää opastusta ja ymmärrystä päätöksen seurauksista.

On tärkeää, että kotikieli määritellään riittävän tarkasti ja että sen tallentaminen ohjeistetaan esimerkein. Jos kotikieli rajataan kotona pääasiallisesti käytetyksi kieleksi, väestötietoihin voi tulla kotikieleksi ja äidinkieleksi esimerkiksi perheen yhteinen kieli englanti, vaikka lapsi käyttäisi sen lisäksi myös vanhempiensa äidinkieliä. Kotoistuksen näkemyksen mukaan lapsen kotikieleksi tulisi voida ilmoittaa mikä tahansa kotona arkipäiväisissä tilanteissa jommankumman vanhemman kanssa käytetty kieli. Laajentuvasta kielivalikoimasta ei tällöin seuraisi velvoitetta järjestää kotikielen ja kirjallisuuden perusopetusta opetuslainsäädännön mukaisesti, mutta kunnan koulutoimi hyötyisi siitä, että se saisi väestötietojärjestelmästä kattavamman kuvan koululaisten kielitaidosta.

Mitä taloudellisia vaikutuksia usean kielen merkitsemisellä väestötietojärjestelmään olisi teidän kannaltanne? Pyydämme teitä erottelemaan vastauksessanne malli A:n ja mallin B:n arvioidut vaikutukset.

Väestötietojärjestelmään esitetyillä muutoksilla ei ole taloudellisia vaikutuksia Kotoistus-hankkeen tai Kansalliskirjaston toimintaan.

Mikä selvitetystä vaihtoehdoista olisi teidän mielestänne paras malli toteutettavaksi?

Malli B

Kirjoittakaa tähän, jos teillä on muuta lausuttavaa

Ehdotetuista vaihtoehdoista Kotoistus pitää parempana ratkaisua B, koska sillä ei ole merkittäviä sivuvaikutuksia koulutoimeen, kuntien kielisyyteen tai valtioneosuusjärjestelmään. Vaihtoehto A johtaisi äidinkielten määrän suureen kasvuun, jolloin esimerkiksi äidinkielen ja kirjallisuuden opetuksen järjestäminen perusopetuksessa omalla äidinkielellä olisi nykyistä oleellisesti vaikeampi ja kalliimpi järjestää.

Muuta lausuttavaa

ISO 639 –standardista ja sen soveltamisesta

Väestötietojärjestelmään nojaavan väestötilaston mukaan Suomessa puhuttiin vuoden 2017 alussa äidinkielenä 159 kieltä. Väestötietojärjestelmässä kieli ilmaistaan ISO 639-2 –koodeina, joita on 487. Näennäisesti asiat ovat siis kunnossa. Mutta Digi- ja väestötietoviraston mukaan monet asiakkaat, mukaan lukien Tilastokeskus, haluavat kielitiedon kuitenkin edelleen 639-1 –koodeina, joita on vain 184. Konversio on häviöllinen, koska 639-1 sisältää yli 300 koodia vähemmän kuin 639-2. Henkilön ilmoittamalla kielellä on siis koodi, mutta se ei ole tarkka tai oikea. Se voi olla esimerkiksi kieliperheen (suomalais-ugrilaiset kielet) tai makrokielen (saamen kielet) koodi.

Kotoistus suosittaa siirtymistä äidinkielen ja kotikielen tallennuksessa ISO 639-2 –standardin soveltamiseen väestötietojärjestelmässä ja muissakin julkishallinnon tietojärjestelmissä siten, että tarvittaessa käytetään kattavaa ISO 639-3 -standardia. Näin toimitaan jo nyt esim. Maahanmuuttovirastossa ja kirjastoissa. Edellinen tarvitsee eksaktin tiedon tulkkaukspalveluja varten, jälkimmäinen aineistojen kuvailuun.

Ellei hyväksyttävää tarkkuutta tavoitella, lähes kaikkien suomalaisten äidinkieli voidaan ilmaista ISO 639-1 –koodina. Tilastokeskuksen mukaan noin 22.000 henkilöllä äidinkieli on ilmaistu väestötietojärjestelmässä 639-1 –koodin tai muun koodimuotoisen tiedon asemesta kielen nimellä. Valtaosa näistä nimistäkin voidaan muuttaa 639-1 –koodeiksi. Tilastokeskuksen arvion mukaan vain 7 255 henkilön äidinkielelle joudutaan käyttämään epäinformatiivista muu kieli -koodia.

Eläviä kieliä on noin 6400. On epätodennäköistä, että monikulttuurisessa Suomessa 99.85 % kansalaisista löytäisi ISO 639-1 –standardin 184 kielestä äidinkieltä ja/tai kotikieltä, eikä edes ISO 639-2 -standardi ole riittävä. Siitä ei löydy esimerkiksi arabian, kiinan tai muiden makrokielen variantteja. Realistinen kuva suomalaisten käyttämistä kielistä saadaan vain sallimalla periaatteessa kaikki elävät, kuolleet ja keinotekoiset kielet sisältävän ISO 639-3 –koodiston käyttö silloin kun 639-2 ei riitä. Vielä tärkeämpää on tietenkin kaikkien väestötietojärjestelmän tietoja hyödyntävien julkishallinnon tietojärjestelmien päivittäminen 639-1:sta 639-2:een, josta on vain verraten lyhyt askel 639-3 –tukeen, koska sekä 639-2 – että 639-3 –koodit ovat kolmen merkin mittaisia.

Tilastokeskuksen mukaan ISO 639-3 ei ratkaisisi ongelmaa, sillä sen mukaan vain noin tuhannen ihmisen äidinkielelle saataisiin koodi ISO 639-3:sta, suurimpina ryhminä ingush, manding ja cebuano (kukin runsas 100 hlöä). Yllättäen siis noin 6000 henkilöä puhuisi siis äidinkieltä, jolle ei löydy koodia koodistosta, jonka pitäisi kattaa kaikki kielet.

Professori Fred Karlsson sai vuonna 2017 Tilastokeskukselta kategorian ”muut kielet” kieli-ilmoitukset tutkittavaksi. Hän kertoo (Karlsson):

Aineisto jakaantui tuhanneksi ortografisesti erilaiseksi tyyppiä. Merkinnoissa oli paljon vaihtelua. Esimerkiksi kaukaasialaiskieli inguušia oli kirjoitettu 25 eri tavalla. Toinen ongelma oli kielen ja sen murteiden välinen suhde – yksi kieli vai monta kieltä? Kolmas ongelma oli Afrikan ja Aasian monikielisten murre-jatkumoiden erikieliset paralleelitermit samalle kielimuodolle.

Osoittautui, että Tilastokeskuksen ilmoittamien 159 kielen lisäksi Suomessa puhutaan ainakin 350:tä kieltä. Siis yhteensä ainakin 500:aa.

”Muiden kielten” suurin ryhmä on tiibetiläis-burmalainen kayah–karen-kielten murrejatkumo (1 250 puhujaa).

Kieliä, joilla on Suomessa satoja puhujia, ovat myös Länsi-Afrikan mandinka (416), inguuši (393), Irakin ja Syyrian assyria (ml. uusaramea, 280), Etelä-Sudanin dinka (258), Filippiinien cebuano (252),

Nigerian edo (eli bini, 249), Vietnamin malaijilais-polynesialainen jarai (239), Pohjois-Afrikan tamazight-berber (169), karjala (161), Sudanin moro (145) ja Irakin kaldea (117).

Suuruusluokan 50–100 puhujan kieliä on kymmenkunta (mm. inkeroinen 53), 10–49 puhujan kieliä 60 (mm. udmurtti 16), 2-9 puhujan kieliä 140 (mm. mordva 6, vepsä 4, hanti 2) ja yhden puhujan kieliä noin 120 (mm. lyydi).

Nelisensataa ihmistä on ilmoittanut tunnistamattoman ensikielen, joko pelkän maantieteellisen nimen ("cameroon"), kielityypin ("kreoli"), kielen nimen, jota ei löydy standardilähteistä eikä internetistäkään, tai tokaisun. Tästä joukosta löytyy sellaisiakin ilmoituksia kuin häme, savo ja jalokivenä vieläpä kalevala.

Tämän analyysin perusteella vaikuttaa siltä, että vähemmistökielten tallentamista äidinkieleksi kartetaan. Suomen valinta äidinkieleksi ja kotikielen tallentamisen mahdottomuus kadottaa väestötietojärjestelmästä lähes 30.000 karjalaa puhuvaa (<https://www.kotus.fi/kielitieto/kielet/karjala>), joista jää jäljelle vain 161 karjalan kotikieleksi ilmoittanutta suomalaista.

Väestötietojärjestelmässä on 30.000 arabiaa äidinkielenään puhuvaa henkilöä, mutta he joutuvat valitsemaan ISO 639-2:sta arabian makrokielen koodin. He eivät voi ilmaista sitä arabian varianttia (esim, Marokon arabia), jota he puhuvat, koska sen koodi on vain ISO 639-3:ssa. Tämä on ongelma, koska arabian kielessä ja muissakin ISO 639-2:n makrokielissä kielten eri varianttien puhujat eivät ymmärrä toisiaan.

Variantin tallentamisen kotikieleksi väestötietojärjestelmään pitäisi olla sallittua, jos äidinkieleksi tallennetaan peruskieli. Tällöin esimerkiksi väestötilastossa voidaan käyttää makrokieltä (eli ISO 639-2 -koodia), mutta esimerkiksi tutkimuksen tai tulkkauksen tarpeita varten väestötietojärjestelmästä voidaan tuottaa myös puhutun kielen eli kotikielen tai –kielten koodi.

Kielten tarkentuvassa kuvaamisessa on haasteensa, eikä ISO 639 ole virheetön eikä välttämättä helppokäyttöinenkään standardi. Alla lisätietoa standardista ja ehdotuksia siitä, miten Suomessa perinteisesti tai yleisesti käytettyjä kieliä voitaisiin koodata ISO 639:stä soveltaen.

ISO 639 -standardista

ISO 639 on kansainvälinen standardi, joka määrittelee kielten kaksi- ja kolmikirjaimiset tunnuksat. Standardin ensimmäinen laitos hyväksyttiin vuonna 1967. Se jaettiin kahteen osaan vuonna 2002, jolloin alkuperäisten kaksikirjaimisten tunnusten päivitetystä versiosta tuli ISO 639-1. Toinen osa, joka määrittelee kolmikirjaimiset tunnuksat, sisältää kaikki ensimmäisen osan kielet sekä noin kolmesataa muuta kieltä. Yksittäisten kielten lisäksi ISO 639-2 määrittää tunnuksia myös makrokielille (esim. saamen kieli). Joillain niistä ei ole vastaavaa ISO 639-1 –koodia.

Vuonna 2007 julkistettu ISO 639-3 pyrkii antamaan kolmikirjaimisen tunnuksen kaikille tunnetuille kielille, olivatpa ne eläviä, kuolleita tai keinotekoisia. Kielikoodin saavalla kielellä pitää olla kirjoitettu versio ja julkaisuja. ISO 639-3 sisältää noin 7500 koodia ja on siten huomattavasti standardin muita osia laajempi. Siihen tulee myös vuosittain kohtuullinen määrä lisäyksiä ja muutoksia. ISO 639-1 –standardia ei enää muuteta, ja 639-2:een lisätään uusia kieliä vain erityisen hyvästä syystä; viimeksi montenegro vuonna 2018, koska se on Montenegron virallinen kieli. Hyvä syy oli tässä tapauksessa

poliittinen, kielitieteellisiä perusteita montenegron lisäämiseen ei juuri ole, sitä pidetään yleisesti serbian murteena.

ISO 639-2 ja 639-3 poikkeavat toisistaan siinä, että moni 639-2:n kieli on 639-3:ssa makrokieli, joka jakautuu useampiin variantteihin. Ne voivat perustua joko kielen puhuttuun tai kirjoitettuun versioon. Esimerkiksi norjan kielellä on 639-3:ssa nynorsk- ja bokmål-variantit. Arabian kielellä variantteja on 29 (<https://iso639-3.sil.org/code/ara>).

Kaikki kolme koodistoa ovat vapaasti käytettävissä verkossa:

- https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes
- <https://www.loc.gov/standards/iso639-2/langhome.html>
- <https://iso639-3.sil.org/>

ISO 639 –standardissa on ollut myös kolme muuta osaa:

- Part 4: Implementation guidelines and general principles for language coding
- Part 5: Alpha-3 code for language families and groups
- Part 6: Alpha-4 code for comprehensive coverage of language variants

Osa 4, joka määrittelee koodistojen ylläpidon periaatteet, päivitetään parhaillaan. Keskeiset periaatteet pysyvät kuitenkin samoina. Kyse on kahdesta kielestä (tai kielestä ja sen variantista) eikä kielestä ja sen murteesta, jos puhujat eivät ymmärrä toisiaan. Siten häme on Suomen murre, mutta karjala on oma kielensä. Kiina ja kantoninkiina ovat kieli ja sen variantti; niillä on sama kirjoitusasu mutta puhuttuna ne ovat hyvin erilaisia. Hämeen murre ei saa omaa 639-3 –koodia, mutta kantoninkiinalla sellainen on.

ISO 639-5 sisältää kieliperheiden koodit. Niitä on yhteensä 115 (<https://www.loc.gov/standards/iso639-5/langhome5.html>). Näitä koodeja (esim. fiu suomalais-ugrilaisille kielille) on tarpeen käyttää vain jos kielelle ei ole omaa koodia. Suomessa on esimerkiksi yksi henkilö jonka äidinkieli on Leventinèss áut d'Airö (<https://kalevalamaailmalla.kalevalaseura.fi/leventiness-aut-dairo-mika-kieli-missa/>). Kielellä ei ole koodia edes ISO 639-3:ssa, jolloin on käytettävä joko ISO 639-2 koodia muu kieli tai (mieluummin) romaanisten kielten kieliperheen koodia.

ISO 639:ssä on aiemmin ollut osa ISO 639-6, joka sisälsi koodit murteille. Sen luonti ja ylläpito osoittautui liian haasteelliseksi, eikä ko. osa enää kuulu ISO 639 -standardiin.

Väestörakenne 2018 –tilasto kertoo kielen tallennuksesta väestötietojärjestelmään seuraavaa:

Väestörekisterikeskus on siirtynyt kielikoodauksessa vuoden 1999 aikana ISO 639 -standardiin.

Tilastokeskus on saanut vuodesta 1999 lähtien Väestörekisterikeskukselta kielet valmiiksi koodattuina (ISO 639-1) ja koodaamattomat kielet selväkielisinä. Selväkielisistä kielten nimistä löytyi paljon koodaantuvia kieliä hieman väärin kirjoitettuina tai ruotsinkielellä. Näille kielille on annettu Tilastokeskuksessa koodi. Aikaisempina vuosina Tilastokeskus on koodannut kaikki vieraat kielet selväkielisistä nimistä ja antanut koodin kielelle, jos sillä on Suomessa ollut noin 15 puhujaa.

Kun kielten tallentaminen koodattuina väestötietojärjestelmään aloitettiin, ISO 639 tarjosi vain alle 200 kielikoodia. Standardin päivittyminen ISO 639-2:ksi vuonna 2003 on otettu huomioon väestötietojärjestelmässä, mutta väestötietoja lähetetään muille toimijoille, kielikoodit konvertoidaan Digi- ja väestötietoviraston mukaan monille vastaanottajille edelleen 639-1 -koodeiksi. Tähän voi olla esimerkiksi teknisiä syitä (639-1 –koodit ovat kahden merkin mittaisia, 639-2 –koodit kolmen).

Tilastokeskuksen asiantuntija Markus Rapo täsmensi Kotoistuksen pyynnöstä tilaston kertomaa:

Ne kielet, joilla ei ole ISO 639-1 -luokituksessa koodia jäävät ilman kielikoodia. Heidän äidinkielekseen merkitään ”muu kieli.”

Tällaisia henkilöitä oli vuoden 2018 lopussa väestössä 7 255 henkilöä eli 0,13 prosenttia väestöstä. Lisäksi äidinkieli tuntematon tapauksia oli 854 henkilöä.

Väestöstä 99,85 prosentille löytyy äidinkieli ISO 639-1 -luokituksesta.

Kielet, jotka eivät löydy ISO 639-1 -luokituksesta ovat pääsääntöisesti murteita ja heimokieliä.

Kotoistuksen arvion mukaan riippuvuus 639-2 –standardista on jatkossa ongelma varsinkin jos kotikielen avulla halutaan ilmaista se ISO 639-2:n makrokielen variantti, jota henkilö puhuu ja/tai kirjoittaa. Mutta 639-2 –koodit tuskin riittävät edes äidinkielen ilmaisemiseen monikulttuuristuvassa Suomessa; siitähän puuttuu lähes 6000 elävää kieltä. Monet niistä voivat olla hyvin harvinaisia, mutta näidenkin kielten puhujille pitäisi antaa mahdollisuus tallentaa oman äidinkielen. Mitään teknistä tai juridista syytä tämän estämiseen ei saisi olla.

Annamme ISO 639-1:n ja ISO 639-2:n soveltamisen ongelmista muutamia Suomen kannalta relevantteja esimerkkejä alla. Mitään kattavaa ongelmien kartoitusta ja ohjeistusta ei tässä lausunnossa voida antaa. On toki selvää että jos kielten tallentamisvaihtoehtoja lisätään, tarvitaan myös lisää ohjeistusta etenkin ISO 639-3:n soveltamisesta. Parempaa ohjeistusta tarvitaan toki jo nytkin: lapsen kieleksi ei voi tallentaa nyt eikä jatkossakaan savoa, hämettä tai kreolia, ja jos tallennetaan koodin sijaan kielen nimi, sen pitäisi olla kirjoitettu yhdenmukaisesti eikä 20 eri tavalla.

Saame

Suomessa puhutuista saamen kielistä 639-1 kattaa vain pohjoissaamen (koodi se). Standardissa ei ole kaikki saamen kielet kattavaa koodia. Tilastokeskuksen mukaan eri saamen kieliä puhuvat kirjataan pohjoissaamen puhujiksi, koska muuten heidät luokiteltaisiin tilastoissa vieraskielisiksi.

ISO 639-2:ssa on sekä saamen makrokielen koodi (smi) että koodit pohjoissaamelle (sme), inarinsaamelle (smn) ja koltansaamelle (sms). Toisin sanoen kaikki Suomessa puhutut saamen kielet voi lisätä äidinkieleksi tai kotikieleksi soveltamalla ISO 639-2 –standardia. Mutta ISO 639-1 -konversiossa äidinkieleksi tulee pohjoissaame riippumatta siitä mitä saamen kieltä väestötietojärjestelmän mukaan puhuu.

Jos Suomeen tulee kiltinäsaamea puhuva henkilö, hänen äidinkielenensä kuvaamiseen tarvitaan ISO 639-3 –standardin koodia sjd. ISO 639-2 tarjoaa vain saamen makrokielen vaihtoehdon, ja ISO 639-1 konversiossa henkilön äidinkieleksi tulee pohjoissaame. Tätä ei voi pitää asianmukaisena.

Kotoistuksen kannalta kaikki Suomessa puhutut saamen kielet tulisi tallentaa omalla koodillaan, ei saamen makrokielen koodina eikä varsinkaan pohjoissaamen koodilla.

Romanikieli

Romanikielen voi tallentaa väestötietojärjestelmään äidinkieleksi. Varsin harva romani on tätä mahdollisuutta käyttänyt hyväksi. Mahdollisuus tallentaa romanikieli kotikieleksi antaisi meille realistisemman kuvan siitä, miten moni romani käyttää aktiivisesti tätä kieltä, jota ollaan elvyttämässä.

ISO 639-2:ssa romanikielillä on vain makrokielen koodi rom. Mutta ISO 639-3:ssa on koodi rmf Suomen romanikielelle (https://www.kotus.fi/kielitieto/kiellet/suomen_romanikieli, <https://www.kielikello.fi/-/suomen-romanikieli>), joka eroaa muista romanikielen varianteista niin paljon, ettei se ole muiden varianttien puhujien ymmärrettävissä.

Suomen romanikieli pitäisi Suomessa koodata sen omalla ISO 639-3 –koodilla romanikielten yhteisen koodin asemesta.

Karjalan kieli

KOTUS kertoo karjalan kielestä seuraavaa (<https://www.kotus.fi/kielitieto/kiellet/karjala>):

Karjalan kieli on suomen lähin sukukieli, jota ei pidä sekoittaa suomen kielen niin sanottuihin karjalaismurteisiin (etenkin kaakkoismurteisiin). Karjalan kielestä on myös käytetty muun muassa nimityksiä itäkarjalaismurteet ja karjala-aunus.

Karjalan kielessä on kaksi päämurretta: varsinaiskarjala ja livvinkarjala (eli aunuksenkarjala). Varsinaiskarjala jaetaan vienankarjalaan ja eteläkarjalaan. Karjalan kieleen on varsinkin aiemmin katsottu kuuluvaksi lyydin kielimuoto (eli ns. lyydiläismurteet), mutta nykyisin lyydiä pidetään tavallisesti omana kielenään.

Karjalan kielellä on ISO 639-2:ssa ja 639-3:ssa koodi krl. Edelliseen se lisättiin vasta 2005. Livvinkarjalalle on koodi olo ISO 639-3:ssa, mutta 639-2:ssa sille pitäisi käyttää fenno-ugrilaisten kielten kieliperheen koodia fiu. Lyydin tilanne on sama, sillä on oma koodi ”lud” vain 639-3 – standardissa.

Jos väestötietojärjestelmässä aletaan käyttää ISO 639-3 -luokitusta, on syytä määritellä, mitkä kielet ovat kansalliskieliä (nykyisin suomi ja ruotsi) ja mitkä mahdollisesti laskettaisiin ”kotimaisiksi kieliksi” tai ”kotimaisiksi vähemmistökieliksi.” Karjalaa mahdollisesti äidinkielenään puhuvista lähes kaikkien äidinkieli on tällä hetkellä väestötietojärjestelmässä suomi. Vain 161 henkilöä on valinnut äidinkielekseen karjalan, vaikka karjalan puhujia on Suomessa KOTUSin arvion mukaan 30.000. Syy tähän voi olla se, että he päätyvät tilastoissa nykymääritelmillä vieraskielisiksi, eikä tilanne muutu siitä, että he voisivat tallentaa karjalan äidinkieleksi ja suomen kotikieleksi. Käyttäjät voidaan ohjata tallentamaan suomi äidinkieleksi ja karjala kotikieleksi, mutta monen suomalaisen kohdalla tosiasiallinen tilanne voi olla päinvastainen. Tämä ongelma voitaisiin korjata määrittelemällä karjala kansalliskieleksi.

Kurdin kieli

Suomessa yleisesti puhutuista kielistä myös kurdin kieli on koodauksen kannalta ongelmallinen. ISO 639-2:ssa kurdin kielellä on vain makrokielen koodi kur. Kurdi jakautuu useisiin variantteihin, joista yleisimmät ovat kurmandzi ja sorani. Edellistä kirjoitetaan latinalaisin aakkosin, jälkimmäistä arabialaisin. Puhuttuina nämä kielet ovat lähellä toisiaan, mutta kirjoitetut versiot ovat täysin erilaisia, millä saattaa olla merkitystä esimerkiksi viranomaisasioidinnissa.

ISO 639-3:ssa kurmandzin koodi on kmr ja soranin ckb. Koska kurdi on kuudenneksi yleisin Suomessa puhutuista vieraista kielistä (14 054 henkilöä puhui sitä äidinkielenään vuoden 2018 lopussa, ks. https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto.html#V%C3%A4est%C3%B6%20kielen%20mukaan%2031.12.), on syytä harkita käytetäänkö kotikielen merkitsemisessä vain makrokielen koodia, vai sallitaanko myös variantin tallennus.

Arabia ja kiina

Arabia ja kiina ilmaistaan väestötietojärjestelmässä ISO 639-2:n makrokielen koodeilla zho ja ara (639-1:ssä vastaavasti zh ja ar). Jos ISO 639-3 -koodien käyttö väestötietojärjestelmässä sallitaan, niillä voitaisiin ilmaista väestötietojärjestelmässä henkilön kotikieli (esim. Marokon arabia) silloin kun äidinkieli on makrokieli (esim. arabia). Tällöin esimerkiksi tulkkauspalvelujen tarjoaminen olisi mahdollista jo väestötietojärjestelmän sisältämien tietojen avulla.

Sama kaksitasoinen tallennusratkaisu on ainakin periaatteessa sovellettavissa kaikkiin ISO 639-3:n makrokieliin, joita on 62.

Puuttuvat kielet

Tilastokeskuksen mukaan noin 6000 henkilön väestötietojärjestelmään tallennettu kieli ei ole muunnettavissa ISO 639 1 -koodiksi. Näille ”ongelmatapauksille” tallennetaan väestötilastoon äidinkieleksi koodi ”muu kieli”. Tämä vääristää tilastoja ja voi periaatteessa vaikuttaa henkilön saamiin palveluihin. Vaikka asiointikieli onkin aina joko suomi tai ruotsi, esimerkiksi tulkkauspalveluita on helpompi järjestää, jos henkilön äidinkieli on tiedossa.

Fred Karlssonin analyysin mukaan suurimmalle osalle henkilöistä joiden kieli on väestötilastossa ”muu kieli” on löydettävissä jokin ISO 639 –koodi analysoimalla väestötietojärjestelmään tallennettu kielen nimi. Todellisia ongelmia jää vain 400. Nämä ongelmatapaukset ovat tallennusvirheitä (häme, kalevala, kreoli, camerun), jotka ovat estettävissä asianmukaisella opastuksella.

Lähteet

Constable, Peter & Gary Simons: An analysis of ISO 639: Preparing the way for advancements in language identification standards. SIL International, 2002. Saatavissa: <https://www.sil.org/resources/publications/entry/7847>

Karlsson, Fred. Suomen 500 kieltä. Teoksessa: Suomen kielet 1917-2017. Saatavana: <https://suomenkielet100.fi/>

Saukkonen, Pasi. Monikulttuurisuuden tilastointi kaipaa uudistamista. Tilastokeskus, 2016. Saatavissa: <https://www.tilastokeskus.fi/tietotrendit/artikkelit/2016/monikulttuurisuuden-tilastointi-kaipaa-uudistamista/>

Usean kielen merkitseminen väestötietojärjestelmään. (Oikeusministeriön julkaisuja. Selvityksiä ja ohjeita 2020:8) [elektroninen julkaisu]. Tammenmaa, Corinna (toim.). Hki: Oikeusministeriö, 2020. Saatavissa: <http://urn.fi/urn:isbn:978-952-259-800-4> [2020-03-30].

Lisätietoja

Juha Hakala

Erityisasiantuntija

Kansalliskirjasto

ISO 639-2 Registration Authority / Joint Advisory Committeeen jäsen

juha.hakala@helsinki.fi

Keskitalo Esa-Pekka

Kansalliskirjasto, Kotoistushanke - Kotoistushanke / Kansalliskirjasto,
kirjastoverkkopalvelut