

Opetus- ja kulttuuriministeriölle

Kotimaisten kielten keskuksen lausunto Tiedon saatavuus -työryhmän raportista Open access -periaatteet

Raportissa ehdotetaan tieteellisen tiedon saatavuudelle visio tavoitetilasta sekä toimenpiteitä vision saavuttamiseksi. Lausuntopyyntöön mukaan esitämme ensin näkemyksemme ehdotettujen toimenpiteiden priorisoinnista ja toiseksi kuvaamme sitä, miten Kotimaisten kielten keskus (myöhemmin Kotus) nykyisin tarjoaa julkaisunsa ja aineistonsa ja mihin se on valmis sitoutumaan tässä vaiheessa.

Toimenpiteiden priorisointi

Monet ehdotetuista toimenpiteistä limittyvät toisiinsa ja tukevat toisiaan, mutta haluamme painottaa erityisesti kolmea toimenpidettä, jotka tällä hetkellä ovat ajankohtaisia tai jotka jo kauan ovat odottaneet ratkaisuaan.

Keskeisin toimenpide tutkimustyön ja asiantuntijatyön kannalta on raportin ehdotus 5.A, jonka mukaan *tutkimustiedon hyödyntämistä tulee edistää rajoittamalla tekijänoikeutta ns. tutkimuspoikkeuksella, jonka tulisi koskea myös kulttuuriperintöaineistojen hyödyntämistä*. Perustelemme tätä priorisointia Kotuksen toimialan ja työn näkökulmasta, mutta *mutatis mutandis* nämä perustelut koskevat yhtä lailla monia muita ihmistieteiden aloja.

Kielentutkimus ja kieliteknologia tarvitsevat käyttöönsä tekijänoikeudesta vapaiden historiallisen aineistojen lisäksi myös uusia aineistoja, esimerkiksi nykykielen eri lajeja edustavia teksti- ja puhekorpuksia. Erityisesti kieliteknologisessa tutkimus- ja kehitystyössä tarvitaan laajoja, vähintään satojen miljoonien sanojen aineistoja, jotta voidaan tehokkaasti edistää kielentarkistuksen, tiedonhaun, puheentunnistuksen ja puhesynteessin sekä konekääntämisen teknologioita.

Laajojen korpusten kokoaminen ei voi useinkaan perustua siihen, että tekijänoikeuksista sovitaan suoraan tekijänoikeuksien haltijan kanssa. Esimerkiksi 1900-luvun kaikkia vuosikymmeniä edustavan sanomalehti- tai tekstikorpusten osalta tekijänoikeudenhaltijoita voi olla tuhansia ja näistä suuri osa perikuntia. Nykyinen lainsäädäntö mahdollistaa tekijänoikeusjärjestön kanssa sovittavan lisensoinnin. Ongelmaksi muodostuu kuitenkin lisensoinnin hinta. Tähänastisten kokemusten perusteella laajojen tekstikorpusten lisensointi tulisi maksamaan kielitieteen alalle vähintään kymmeniä tuhansia euroja vuodessa, vaikka korpusten käyttö rajattaisiin tutkimuskäyttöön ja vaikka teknisin ratkaisuin estettäisiin tekstien näkyminen virkettä laajempina yhtenäisinä kokonaisuuksina (mikä puolestaan osaltaan rajoittaa tutkimusmahdollisuuksia). Hallitun tutkimuskäytön ei voi ajatella vaarantavan tekijänoikeudenhaltijoiden tosiasiallisia taloudellisia intressejä, joten lisensointia tehokkaampi tapa turvata tieteen toimintaedellytyksiä olisivat tekijänoikeuslakiin sisältyvät tutkimuskäytön mahdollistavat poikkeukset. Se että tällaisia tutkimuspoikkeuksia ei ole tähän mennessä ollut, on suoraan vaikuttanut esimerkiksi

siihen, että 1900-luvun kirjoitetun suomen kehitystä ei ole voinut tutkia systemaattisesti eikä tarpeellisten laajojen aineistojen varassa. Kun 1900-luku on tässä mielessä musta aukko, ei myöskään oman ajan kirjoitetusta kielestä kiinnostuneille kansalaisille ole voitu tarjota kokonaiskuvaa viime vuosikymmenten kehityksestä. Siten tekijänoikeudet ovat kahlinneet tutkijoiden toimintavapautta ja suorastaan estäneet laajoihin korpuksiin perustuvan tutkimuksen ja kielen kehityksen seurannan.

Nykykielen korpusten lisäksi tekijänoikeuslainsäädäntö vaikuttaa myös muiden tutkimusaineistojen käytettävyyteen ja käyttöön. Esimerkiksi analogiseen muotoon kerättyjen ääniteaineistojen osalta tekijänoikeuslaki sallii arkistoille digitoinnin, mutta rajoittaa tekijänoikeuden suojaaman digitoitun aineiston käytön laitoksen tiloihin, mikä puolestaan merkittävästi rajoittaa mahdollisuuksia aineiston tutkimukselliseen käsittelyyn eri sovelluksin ja työkaluin ja sulkee kokonaan pois mahdollisuuden tarjota tekijänoikeuden alaista aineistoa tutkijoille verkossa luvanvaraisena annotoituna korpuksena. Tekijänoikeudenhaltijoiden tavoittaminen vuosikymmeniä aineistonkeruun jälkeen on käytännössä mahdotonta, semminkin kun monia valtavia aineistokokonaisuuksia ylläpidetään minimaalisin resurssein. Merkittävä ongelma on myös teostason tulkinnanvaraisuus. Monet kielitieteen arkistokokonaisuudet ovat luonteeltaan sellaisia, että pääosin teostason alapuolelle jäävän sisällön joukossa on myös teostason ylittävää sisältöä. Teostason määrittelyyn liittyviltä ongelmilta välttyttäisiin ja selvittelyyn kuluvat resurssit voitaisiin käyttää tuloksellisemmin, jos tekijänoikeuslaki sallisi tutkimuskäytön nykyistä paremmin. Lisensointimallia on hyvin vaikea soveltaa aineistoihin, joiden sisällön teoksellisuus on vaihtelevaa ja tulkinnanvaraista.

Kielitieteen tutkimus- ja arkistoaineistoissa on paljon kansan parista kerättyä materiaalia. Tällaiseen aineistoon ei ole alun perinkään liittynyt tekijän taloudellisia intressejä, joten sen käytön ohjaamiseen sopii varsin huonosti taloudellisia oikeuksia vahvasti suojaava tekijänoikeuslainsäädäntö, joka ei ota riittävästi huomioon tutkimuksen ja kulttuurin erityistarpeita. Suullisen kulttuuriperinnön luonteeseen kuuluu sen yhteisöllisyys ja jakaminen, joita nykyinen tekijänoikeuslaki estää, esimerkiksi vaikeuttaa kieli- ja kulttuurivähemmistöjen mahdollisuuksia hyödyntää omaa perintöään kielen ja kulttuurin revitalisaatioissa.

Toinen keskeinen ja vahvistamisen arvoinen toimenpide-ehdotus on näkemyksemme mukaan ehdotus 4.E, joka koskee *julkaisujen saatavuutta edistäviä käytäntöjä ja toimintatapoja*. Siinä työryhmä suosittelee käytettäväksi Vihreää OA:ta (Green Open Access), jolloin julkaisun tai sen osan rinnakkaiskopio tallennetaan tieteenala- tai organisaatiokohtaiseen julkaisuarkistoon joko heti tai ennalta määrätyn embargoajan jälkeen, tai Kultaista OA:ta (Gold Open Access), jolloin koko julkaisu on maksuttomasti käytettävissä kustantajan tai muun välittäjän verkkopalvelussa. Tällaiseen menettelyyn ohjeistetaan jo nyt monissa yliopistoissa ja tutkimuslaitoksissa, ja tällainen julkaiseminen ja arkistointi ovat edistyneet myös käytännössä. Tärkeää on kuitenkin, että tässä edetään niin, että arkistointi tulee osaksi tutkimusprosessia ja että kaikki tieteenalat noudattavat samanlaisia toimintatapoja. Näin edistetään tiedon leviämistä nopeasti ja demokraattisesti kaikkien kiinnostuneiden ulottuville.

Kolmas keskeinen toimenpide-ehdotus koskee *tutkimusjärjestelmän organisaatioiden ja toimijoiden tukea ja kannustusta*. Ehdotuksissa 2.B ja 2.C painotetaan sitä, että suomalaiset tutkimusrahoittajat ottavat huomioon OA-julkaisemisen kustannukset ja että OKM kehittää rahoitusmallin OA-julkaisemisen tukemiseksi. Tämä ehdotus on tärkeä siksi, että maksutta julkaiseminen aiheuttaa organisaatioille myös kuluja ja tulonmenetyksiä. Tämä voi koskea varsinaisia tutkimusjulkaisuja, mutta vielä enemmän asiantuntijatyössä tuotettuja, laajemmalle lukijakunnalle kuin oman alan tutkijoille tarkoitettuja julkaisuja. Kotuksen työn näkökulmasta esimerkkinä voi mainita sanakirjat ja kielenhuollon yleistajuiset oppaat ja julkaisut.

Kotimaisten kielten keskuksen sitoutuminen ja tulevaisuuden tavoitteet

Kotuksen pääjulkaisuja ovat sanakirjat ja kielenhuollon julkaisut sekä niihin liittyvät tai niitä lähellä olevat tutkimusartikkelit tai teokset. Kotuksessa on myös laajat aineistot, joiden saatavuutta ja käytettävyyttä on viime vuosina jatkuvasti parannettu. Tässä työssä on tehty yhteistyötä Fin-Clarinin kanssa ja Kansallisen digitaalisen kirjaston piirissä. Tulevaisuudessa pääosa Kotuksen kieliaineistoista on saatavilla CSC:ssä. Osa aineistosta on luvanvaraista, valtaosa vapaasti käytettävissä.

Valtaosa Kotuksen julkaisuista on jo nykyisin maksutta tarjolla Kotuksen omassa verkkopalvelussa. Kotuksen työntekijät julkaisevat artikkeleita myös alan aikakauskirjoissa (kuten *Virittäjä* ja Suomalais-Ugrilaisen Seuran julkaisut), ja ne julkaistaan verkossa useimmiten vuoden mittaisen embargoajan jälkeen. Kaupallisten kustantajien kanssa noudatetaan niiden julkaisupolitiikkaa tai tehdään maksutta julkaisemisesta tietyn embargoajan jälkeen sopimus; hyvä esimerkki tästä on *Ison suomen kielioopin verkkoversio VISK*.

Kotuksen julkaisuista vain pieni osa on julkaistu viime vuosina kirjana. On todennäköistä, että maksullisten julkaisujen osuus Kotuksen julkaisuista yhä vähenee. Tämä edellyttää kuitenkin rahoitusta, jolla yhtäältä korvataan julkaisujen tuottamista verkkoon ja niiden ylläpitoa, toisaalta korvataan se tulonmenetys, joka avoimesta julkaisemisesta väistämättä seuraa.

Helsingissä 5. joulukuuta 2013

Pirkko Nuolijärvi
johtaja

Ritva Paananen
suunnittelupäällikkö