

FiLRC: FIN-CLARIN-johtoryhmän kokous

Aika: pe 20.5.2011 klo 15.15-16.15

Paikka: Kotus-sali, Kotimaisten kielten tutkimuskeskus, Vuorikatu 24, Helsinki

PÖYTÄKIRJA

Kokous alkoi klo 15.15

Osanottajat:

HY: Kimmo Koskeniemi (pj), Mietta Lennes (siht.), Krister Lindén, Jussi Piitulainen

CSC: Tero Aalto, Antti Pursula

JY: Ari Huhta

Kotus: Toni Suutari

OY: Jarmo Jantunen

TaY: Mikhail Mikhailov

ISY: Stefan Werner

1. Yleistä, ilmoitusasioita

Marketta Harju-Autti jäi tulevan eläkkeelle siirtymisensä vuoksi pois jo tästä kokouksesta ja häntä edustaa nyt virallisesti Jarmo Jantunen, varajäsenenä Harri Mantila. Tervetuloa uudet jäsenet!

2. Metadatatyöpajan satoa

Metadatatyöpaja oli hyödyllinen tilaisuus. Työpajaan osallistui useita kieliaineistojen haltijoita ja saatiin parempi näkemys siitä, minkälaista informaatiota he toivovat aineistojen kuvailutiedoissa voitavan esittää. Itse kuvailutietojen keruu ei työpajan aikana onnistunut aivan toivotulla tavalla, koska yhteisen metadataformaatin, META-NETissä kehitteillä olevan ns. komponenttimetadatan (CMDI) määrittely ei ollut vielä käytettävissä, joten metadatan rakenteesta oli jonkin verran epäselvyyttä. Työpajassa nähtiin myös lyhyt demonstraatio MPI:ssä kehitetyistä Annex- ja Trova-työkaluista, jotka on juuri asennettu CSC:n palvelimelle. Niillä voidaan selailla aineistoja ja tehdä niistä hakuja, kunhan aineistojen annotaatiot ja metadata on kuvattu muodollisesti oikein. Seuraava Metadatatyöpaja järjestetään Kotuksessa 30.9.2011 yhdessä META-NORD-verkoston kanssa.

3. CLARIN ERIC

Toukokuun lopussa Hollanti jättää alustavan CLARIN ERICin perustamista koskevan hakemuksen komissioon, jolloin prosessi lähtee virallisesti liikkeelle jo nyt ministeriöiden allekirjoittaman aiesopimuksen perusteella (ns. Memorandum of Understanding). Heinäkuussa eri maat allekirjoittavat seuraavan, sitovamman askeleen. Kunkin valtion rahoituksen olisi oltava selvillä tätä sitoumusta allekirjoitettaessa. Lokakuuhun saakka on vielä mahdollisuus päästä ns. perustajajäseneksi. Ensi vuoden tammikuussa Hollannin tiedeministeri lähettää kuitenkin vielä laajemman liittymiskutsun, ja silloin CLARIN ERICiin voi tulla mukaan tavallisena jäsenenä. OKM:n on vielä varmistettava, mikä taho Suomessa ottaa vastuun CLARIN ERICissä annettujen lupauksen käytännön toimeenpanosta. Helsingin yliopisto FIN-CLARINin nykyisenä vastuutahona olisi selkein vaihtoehto jatkamaan (sillä ministeriö ei halua itselleen tällaisia asioita). Helsingin yliopiston täytyisi siis allekirjoittaa CLARIN ERICin edellyttämät sitoumukset siitä, että Suomessa tehdään mitä CLARIN ERIC siltä edellyttää ja ministeriön pitää turvata noiden velvollisuuksien edellyttämä rahoitus (HY:n budjettiin) ainakin vielä vuosiksi 2013-2016. OKM:n ja HY:n johdon väliset neuvottelut rahoituksesta ja sen muodosta ovat käynnissä. Ministeriössä asia etenee normaalisti heinäkuun etappiin saakka. Silloin allekirjoitusta vaaditaan ministeriltä, joten asia on uuden hallituksen käsissä ja on periaatteessa mahdollista muttei todennäköistä, että CLARIN ERIC-prosessin eteneminen Suomen osalta hidastuu. Valtiolla ei ole erityistä momenttia, jolla rahoitettaisiin erilaisia infrastruktuureja, eli nyt on hankala tilanne jossa jokaista tällaista hanketta rahoitetaan erityisjärjestelyillä.

Olisi tärkeä saada eri yliopistojen tuen ilmauksia FIN-CLARINille: kun asia nähdään tärkeäksi koko maassa, on rahoituksen saamiselle paremmat lähtökohdat.

CLARIN ERICin säännöt edellyttävät, että jäsenmaalla on kansallinen CLARIN-konsortio, mutta ne eivät edellytä mitään tiettyä rakennetta konsortiolle, vaan se voidaan ratkaista kansallisesti. FIN-CLARINin tähänastinen rakenne ja rahoitusmalli kelpaavat pääpiirteittäin CLARIN ERICin sääntöjen mukaiseksi kansalliseksi CLARIN-konsortiksi. FIN-CLARINilla tulisi edelleen olemaan johtoryhmä, joka ohjaa ja valvoo konsortion toimintaa. Käytännössä nytkin FIN-CLARINin rahat tulevat tarkoitukseen myönnetystä osasta HY:n budjettia, josta HY on vastuussa. Voidaan olettaa, että jatkossa asetelma säilyy samantapaisena. Kuitenkin HY:n keskushallinto tulee olemaan tiiviimmin mukana FIN-CLARINin sitoumuksissa ja vastuissa, voidaan odottaa, että se edellyttää FIN-CLARINin johtoryhmälle muodollisesti ja juridisesti näitä vastuita vastaavia säännöksiä, eli emme voi aivan nykyisen johtoryhmän kesken sopia enemmistöpäätöksillä kaikesta.

Kansalliselta CLARIN ERICin koordinaattorilta edellytetään suunnitelma siitä, mitä ko. maa aikoo tuoda CLARIN ERICiin.

Suunnitelmia aiotaan jatkossa tarkistaa ja päivittää kerran vuodessa (vaikka ERIC-jäsenyys onkin 5-vuotinen) ja CLARIN ERIC valvoo samalla, että lupaukset on täytetty. Koska kyseessä on yhteinen intressi, ei ongelmia ole käytännössä odotettavissa.

4. FIN-CLARIN-CONTENT-rahoituksella valmisteltavat aineistot

Kotuksen aineistojen (SKN-korpus eli *Suomen kielen näytteitä* -sarjan murrenäytteet) valmistelu etenee aikataulun mukaisesti. Kansalliskirjaston aineistoista jatketaan neuvotteluja Kopioston kanssa. Kirjastolla on miljoonia sivuja tekstiä OCR:llä tunnistettuna, ja nämä on mahdollista saada rajoitettuun käyttöön, ei kuitenkaan artikkelin kokoisina vaan "kielimassaksi silputtuna". Ratkaisun aikataulu, onnistumisen mahdollisuus ja hinta ovat epäselviä.

Mikäli katsotaan hyödylliseksi, on mahdollista selvittää käyttöoikeuksien hankintaa Celian (Näkövammaisten kirjaston) äänikirjoihin sekä niiden käännöksiin eri kielillä (sekä äänitteinä että teksteinä). Celian äänikirjoissa teksti on valmiiksi kohdistettu ääneen SGML-pohjaisessa Daisy-muodossa ja eri maiden näkövammaisten kirjastot käyttävät samaa tallennusmuotoa, jolloin olisi mahdollisuus monikieliseen puhetta ja tekstiä sisältävään aineistoon.

SKS:n kanssa on aloitettu aineistoja koskevat neuvottelut.

FIN-CLARINin johtoryhmä on nimetty myös FC CONTENTin toimintaa ohjaavaksi tahoksi.

5. META-NET- ja META-NORD-hankkeiden kuulumisia

Suomea koskeva Language Report on saatu valmiiksi.

Kieliaineistojen käyttöoikeuksista (IPR/META-SHARE) on valmisteilla kaavio, jonka perusteella aineiston tuottaja voi valita aineistolle sopivan käyttöoikeussopimustyyppin.

6. Kielipankki

Metadata nykyisistä Kielipankin aineistoista on olemassa, mutta uutta komponenttimetadatan spesifikaatiota odotellaan.

Kielipankkiin mahdollisesti tulossa olevia aineistoja on kartoitettu ja alustavasti priorisoitu, sillä tällaista tietoa piti toimittaa META-NETille. Selvitys jatkuu aineistokohtaisesti.

Suunnitteilla on lyhyt kysely, jota levitetään laajasti kaikille Suomen kielentutkijoille ja mahdollisille Kielipankin käyttäjille. Kyselyllä on tarkoitus selvittää aineistoihin ja työkaluihin liittyvää tilannetta Suomessa, Kielipankin ja sen nykyisten palveluiden tunnettua, erityyppisten uusien aineistojen tarvetta ja niihin liittyvien käyttäjäryhmien tarpeita.

Samaa kieliaineistoa käyttävillä tutkijoilla on luonnostaan tarve ottaa yhteyttä toisiinsa, sillä jokaisesta aineistosta ja siihen liittyvistä menetelmistä löytyy erikoisuuksia ja virheitä, joita aineiston metadatatassa ei pystytä kuvaamaan. Aineistojen tutkimus ja sovellukset myös kehittyvät ajan myötä. Tästä syystä CSC:n ylläpitämien aineisto- ja lupapalvelujen jatkeeksi ja tueksi olisi mahdollista ottaa käyttöön esim. blogipohjainen alusta nimeltä Kielipankki. Yhteisöalusta voisi toimia erilaisista kieliaineistoista kiinnostuneiden tutkijoiden ja opiskelijoiden foorumina, jossa esimerkiksi työkalujen käyttöohjeita ja erilaisia aineistojen keruu- ja tutkimusmenetelmiä voitaisiin kerätä, kommentoida ja kehittää kulloinkin sopivan kokoisissa, vaikkapa aineistokohtaisissa keskusteluryhmissä. Kielipankki-yhteisön ylläpitovastuu olisi Helsingin yliopistolla, kun taas aineistojen ja niihin liittyvien käyttö- ja lupapalveluiden teknisestä alustasta vastaisi CSC. Hyvä tapa selvittää tätä mahdollisuutta olisi pilotoida jonkin yhteisöpohjaisen alustan käyttöä, jolloin saataisiin tietoa jonkin ongelman ratkaisusta yhteisön avulla.

7. Muut asiat

FIN-CLARINille on luvassa pieni A4-kokoinen esite alkusyksystä.

8. Seuraava kokous

Seuraava kokous pidetään 30.9. toisen Metadatatyöpajan yhteydessä.

9. Kokouksen päättäminen

Kokous päättyi klo 16.15.

Tehtävälista

Kyselytutkimuksen suunnittelu yhdessä CSC:n Kielipankin kanssa

-- [MiettaLennes](#) - 2011-06-03