

Korpusten sijainti CSC:n koneella

Pirkko Suihkonen (12.10.2007)

UHLCS:n korpusten hakemistorakenteen ensimmäinen jako CSC:n koneella perustuu siihen, mikä instituutio myöntää käyttöluvan. Hakemistojen sisällä aineistot on järjestetty kieliperheiden mukaan. Korpusten yhteytten on liitetty ko. metadata-tiedostot. Kieliperheiden noodeja muodostettaessa on pyritty ottamaan huomioon tieto siitä, että UHLCS kasvaa, ja että UHLCS:een saadaan uusia aineistoja.

1) Hakemisto /ADM: hakemistoon on sijoitettu kaikki sellaiset aineistot, joiden sopimusten selvitys on vielä kesken. Sopimuksista on tiedusteltu Fred Karlssonilta ja Kimmo Koskenniemeltä. Aineistot siirretään ko. hakemistoihin sitä mukaa, kun niitä koskevat sopimukset saadaan ajan tasalle. Mikäli sopimuksia ei uusita tai niiden suhteen on muita epäselvyyksiä, aineistot poistetaan. Erikseen on huomioitava aineistot, jotka ovat osia korpuksista, joiden sopimukset ovat kunnossa. Nämä voitaneen jättää yleisen kielitieteen laitoksen tutkijoiden käyttöön.

2) Hakemisto /general-linguistics: hakemistossa on sellaiset korpukset, joiden käyttöluvan myöntää Helsingin yliopiston yleisen kielitieteen laitos. Käyttöluvan myöntämiseksi ei tarvitse ottaa yhteyttä korpusten omistajiin.

3) Hakemisto /general-linguistics-kotus: hakemistossa on ja sinne tulee sellaiset korpukset, joiden käyttöluvan voi myöntää yleisen kielitieteen laitos tai Kotus.

4) Hakemisto /language-departments: hakemistossa on sellaiset korpukset, joiden käyttöluvan myöntämiseksi tarvitaan lupa siltä laitokselta/laitoksen edustajalta, jonka korpuksen käyttölupaa haetaan. Toistaiseksi hakemistossa on swahilin ja ruotsin korpuksia. Kaikista venäjän korpuksista on neuvoteltu sopimus, jonka mukaan käyttöluvan myöntämiseen ei tarvitse ottaa yhteyttä korpuksen omistajaan tai omistajan edustajaan.

5) Hakemisto /linguistics-permission: hakemistossa on kaikki ne UHLCS:n korpukset, joiden käyttöluvan saamiseksi on otettava yhteyttä korpusten omistajaan tai omistajan edustajaan. Henkilöiden nimet ja yhteystiedot on annettu UHLCS:n verkkosivulla.

6) README-tiedostot: a) README: lyhyt kuvaus UHLCS:n arkistosta. b) README-GROUPS: Käyttäjärühmien luettelo englanniksi. c) README-RYHMÄT: käyttäjärühmien luettelot suomeksi.

7) UHLCS:n aineistojen hakemistorakenne: tilanne 12.10.2007

drwxr-x--- 14 suihkone li-adm 4096 1. loka 16:45 ADM

drwxr-x--- 7 suihkone li-adm 4096 12. syys 08:07 general-linguistics

drwxr-x--- 4 suihkone li-adm 4096 11. syys 10:30 general-linguistics-kotus

drwxr-x--- 5 suihkone lipr-adm 4096 5. syys 18:26 language-departments

drwxr-x--- 11 suihkone lipr-c3 8192 3. syys 10:06 linguistics-permission

```
-rw-r--r-- 1 suihkone uhles 150 4. syys 19:08 README  
-rw-r--r-- 1 suihkone uhles 6435 4. syys 19:02 README-GROUPS  
-rw-r--r-- 1 suihkone uhles 5872 4. syys 20:00 README-RYHMÄT
```