

## Tapaaminen CSC:llä 24.8.2004: muistio

1. Tapaamisessa mukana Eero Vitie ja Pirkko Suihkonen.
2. Tapaamisen tarkoitus: päätettävä korpusten siirtämiseen liittyvistä yksityiskohdista.
3. Ryhmien nimissä voi olla vain pieniä kirjaimia (ja numeroita), ja nimien pituus voi olla vain kahdeksan merkkiä. Muutettava ryhmien nimet tämän vaatimuksen mukaiseksi. Ryhmien oltava kunnossa, ennen kuin loput korpuksista siirretään. Uusi versio ryhmistä laitetaan KitWikiin.
4. Korpusten ryhmistä:  
CSC:llä olevien ja sinne tulevien UHLCS:n aineistojen käyttöoikeuksien pitäisi vastata /corp-hakemistossa olevia käyttöoikeuksia. Tämä järjestetään ensisijaisesti ryhmien avulla. Selvitetään, miten ryhmien merkitseminen korpuksiin tehdään mahdollisimman automaattisesti. Tämän näkökulman pitäisi olla mukana myös käyttölupia myönnettäessä. Kaikissa korpuksissa hakemisto /originals ja mahdollisesti /tmp tai vastaava hakemisto, jossa on korpusten työversioita, saavat ryhmän, jossa on \*-adm.
5. Metadata-tiedostot kopioidaan CSC:lle korpushakemistojen yhteyteen alkuperäisen suunnitelman mukaan.
6. Korpusten käyttäjätunnuksen käyttöä seurataan. Käyttöluvan jatkoa pyydetään CSC:ltä. Jos kysymyksessä on projekti, jossa on useita henkilöitä, projektin johtajalle tiedotetaan käyttölupien umpeutumisesta.
7. Korpusten siirto:  
README-tiedostoissa on tarpeen olla tieto siitä, mistä versioista on kysymys: työversioista, vanhoista versioista, jne. README-tiedostossa on tarpeen olla tieto myös päivämäärästä, jolloin korpus luovutetaan/on luovutettu käyttöön. Tämä tieto on myös metadata-tiedostoissa.

Huom. CSC:llä on useita korpuksia, joiden osia on UHLCS:ssä. Aina ei suihkaan ole kysymys vanhoista tai työversioista, vaan eri tavalla toimitetuista korpuksista tai korpuksista, jotka sisältävät erilaisia aineistoja. Korpusten omistaja tai korpusten hallinnasta vastaava kuvaa korpukset README-tiedostossa (tästä ovat esimerkkeinä Arvi Hurskainen ja Jan K. Lindström).

Korpukset, joiden siirtoon on saatu lupa korpusten toimittajalta, mutta joista ei ole saatu virallista sopimusta, laitetaan ryhmän \*adm hallintaan (\* = luvanantajat: li, lik, lipr, ld).

Loput korpuksista, joiden sopimukset ovat selvinneet 31.8.2007 mennessä, siirretään CSC:lle perjantaina (31.8.2007 (siirto alkaa klo 9)). Sitä ennen on selvitettävä, tarvitaanko siirrossa adm.ling-henkilökunnan apua (Pirkko: asiasta on kerrottu ennakkoon Jukka Huhdalle 24.8.2007).

8. Uudet korpuukset:

UHLCS:iin tulee jatkuvasti uusia korpuksia. Ennakkotietoja on seuraavista korpuksista:

- a. Raamatunkäännösinstituutti: useiden eri kielten korpuksia,
- b. Tapani Salminen: metsänenetsi,
- c. Olavi Korhonen: uumajansaame,
- d. André Hesselbäck: mongolilaiskielten bibliografia (laitetaan mongolilaiskielten hakemiston yhteyteen),
- e. Mihail Kopotev (?): slaavilaiset kielet,
- f. Burgel R M Faehndrich (Havaijin yliopisto): mongolilaiskielet.

Aineistot liitetään UHLCS:n alaisuuteen, mutta mistä siirtoon tarvittavat resurssit? Olisi selvitettävä, miten voidaan varmistaa, että korpusten vastaanottamisessa, dokumentoinnissa ja asennoinnissa tarvittava työ voidaan rahoittaa. Onko mahdollista perustaa projekti, jolle voidaan hakea rahoitusta?

9. Eero selvittää, miten luodaan CSC:n kieliaineistojen käyttäjien ryhmä, jolla on yhteinen hakemisto (tällaisia on esim. angarakilla). Ryhmään kuuluvat saamassa projektissa työskentelevät käyttäjät.