

Työn kulumisesta:

- 1) Korpukset on tarkastettu ja jos tiedostoista on alkuperäinen tiedosto ollut saatavilla, se on kopioitu /corp-hakemistoon ko. original-tiedoston kohdalle.
- 2) Metadata-kuvaukset ovat siinä kunnossa, että ne voidaan kopioida korpusten yhteyteen. Meta-tiedostoissa oli editoinnissa tapahtuneita virheitä (ks. korpusten-siirrosta-1.doc). Yhtenäistän metadata-tiedostoja ottaen erityisesti huomioon sen, että data tulee olemaan CSC:llä.
- 3) UHLCS:n info-sivut on päivitetty. Nyt myös suomenkielinen osuus on saatavilla: aikaisemmin siellä oli vain jotakin alustavia kommentteja.
- 4) Infosivuja on yksinkertaistettu jonkin verran. Kun tein niitä aikanaan, ongelmana oli, että piti tehdä töitä sen mukaan miten dataa saatiin. Ensimmäisenä ajatuksena oli saada kerrotuksi perustietoa siitä, mitä on, ja tarve info-sivujen edustavuudesta tuli vasta myöhemmin. Ongelmana oli myös se, etteivät korpusten omistajilta saadut tiedot olleet aina yhtenäiset. Mutta mielestäni info-sivut ovat nyt selvät.
- 4) Mannerten kartat on laitettu UHLCS:n yhteyteen. Teen niistä metadata-tiedoston. Myös UHLCS:n sivulla olevan parin video-nauhan metadata-kuvaukset pitää vielä tehdä.
- 5) Suurin osa korpusten omistajista on vastannut korpusten siirtämistä koskevaan tiedusteluuni. Jotakin vastauksia vielä odotan. Myös IBT:ltä (Tukholma ja Helsinki) on otettu yhteyttä ja luvattu palata asiaan. (Odotan vielä heidän vastaustaan.) Liitän CSC:lle annettaviin dokumentteihin kaikkien korpusten omistajien vastaukset.

Tehtävää:

- 1) Info-sivuissa ja metada-tiedostoissa on jotakin oikoluettavaa (mutta vain vähän). Info-sivut ovat kaikki avoimena (tarkastan vielä, että jokainen linkki toimii). Metadata-tiedostojen hakemiston (/web/ling/uhlcs/metadata/corpus-metadata/) avaan heti, kun olen saanut IBT:ltä vastauksen.
- 2) Sopimusten kanssa on vielä jonkin verran töitä. Kirjoitan niistä oman yhteenvedon ja laitan sen TWWiki-sivulle. Sopimuksia on erilaisia, eikä riitä, että teen luettelon vain niistä sopimuksista, joita ei ole. Kirjoitan yhteenvedon tiedon siitä, mitä mikin sopimus sallii. Kimmo: niitä ”vanhojen” korpusten sopimuksia puuttuu vielä aika monta. Ymmärsinkö oikein, että Fred mahdollisesti vielä yrittää etsiä niitä?
- 3) Työstän nyt ryhmiä koskevaa osiota. Laitan tästäkin osuudesta oman kuvauksensa TWWiki-sivulle.
- 4) Päivitän CSC:n koneelle jo siirretyt korpushakemistot ja korpukset.
- 5) UHLCS:ssa olevien vanhojen korpusten siirtäminen CSC:lle.

6) Työn valmistumisen arvioitu päivämäärä:

Yritän saada kaiken mahdollisimman hyvään kuntoon viimeistään 15.8. mennessä. Käytännössä kaiken pitäisi olla kunnossa jo aikaisemmin. Ehdotan alustavasti, että sen jälkeen voitaisiin pitää kokous ja katsoa, onko vielä jotakin, mitä pitäisi muuttaa tai tehdä toisin. Olen kiitollinen kaikista kommentteista.

Parhain terveisin,
Pirkko