

UHLCS:n korpuukset, joiden sopimusten selvitys on kesken.¹
 Jos erikseen ei muuta ole mainittu, nämä korpuukset ovat CSC:n corpus-koneella hakemistossa /l/mrc-uhlcs/ADM.

/corp/bible	
	../BhebraicaS, ../in_future, ../Nestle-Aland26
Anssi hoitaa korpuksia koskevat sopimukset kuntoon korpusten omistajien kanssa.	
corp/eng	
	../brown
Minulla ei ole ollut käytössä luettelo aineistosta, jotka sisältyvät BNC:n sopimukseen. BNC:n verkkosivuilla esitellään Brown-korpus osana BNC:ta. Sopimuksen kesto BNC:n ja yleisen kielitieteen laitoksen välillä on viisi vuotta (verkossa oleva sopimus: ”Ten years” (http://www.natcorp.ox.ac.uk/docs/licence.html)). Sopimus on allekirjoitettu heinäkuussa 1995. Jos sopimusta ei ole uudistettu, se ei siis ole enää voimassa ja korpus täytty poistaa UHLCS:ltä. Atron README: Brown.334233.cg.tagged contains the first 334233 words of the Brown Corpus as disambiguated by ENGCG, the Constraint Grammar Parser of English. The analyses have also been checked manually. Ko. tiedostoa ei ole hakemistossa.	
	../grolier
Atron mukaan Grolier-data on laitokselle hankitusta G.El.Encyclopediasta. Sen levityksestä ei tiettävästi ole tehty sopimusta. Tämä tarkoittanee sitä, että korpusta ei voida ottaa enää käyttöön CSC:llä. Korpus on ollut laitoksen tutkijoiden käytössä (ryhmä y_kiel). Korpuksen sopimuksesta on tiedusteltu Frediltä ja Kimmolta.	
	../lob, ../london-lund
Minulla ei ole käytettävissäni näitä korpuksia koskevia sopimuksia. Korpuukset ovat olleet laitoksen tutkijoiden käytössä. Kummankin korpuksen sopimuksista on tiedusteltu Frediltä ja Kimmolta.	
	../misc
Minulla ei ole käytettävissäni näitä tekstejä koskevia sopimuksia. Muutamia teksteistä on myös BNC:ssa. Teksteissä on mukana myös englanninkielinen Raamattu, joka on saatavissa verkossa, sekä esim. Moby Dick ja Anna Karenina. Arvioni on, että ainakaan Raamattu ja Anna Karenina eivät enää ole tekijänoikeuslain alaisia. Aineistojen sopimuksista on tiedusteltu Frediltä ja Kimmolta. Aineisto on ollut laitoksen tutkijoiden käytössä.	
/corp/espanol	
Teksti on vapaasti saatavilla osoitteessa http://eprintweb.org/S/authors/All/ch/CHANDLER-BURNS . Voidaan siirtää tai olla siirtämättä. Materiaalia ei ole siirretty CSC:lle.	
/corp/fin	
	../bin, ../pub
Kimmo tietää, voidaanko skriptit siirtää CSC:n koneelle. Näiden siirto corpus-koneelle ei	

¹ Hakemistot ja tiedostot, jotka ovat UHLCS:ssa suljettuja, säilytetään suljettuna myös CSC:llä.

onnistunut.	
	../fintag/, ../skh, ../ksk, ../otava,../wsoy, ../sk87, ../misc, ../kuvaristikot, ../hs, ../sk75
Hakemistoissa olevia aineistoja on työstetty yleisen kielitieteen laitoksella. Hakemistossa /fintag on mm. tekstejä, jotka ovat hakemistoissa ../ksk, ../otava ja ../wsoy, ja jotkin aineistot esiintyvät myös muissa korpuksissa.	
Ehdotus: aineistot jäävät laitoksen tutkijoiden käyttöön niin kuin ne ovat olleet ennenkin. Alkuperäisten aineistojen sopimuksista on tiedusteltu Fredltä ja Kimmolta.	
	../hkv
Korpuksen omistajat Auli Hakulinen, Fred Karlsson ja Maria Vilkuna. Korpuksen tekstiaineistot ovat suhteellisen lyhyitä. Korpuksen siirrosta tiedustelu Maria Vilkunalle 6/8/2007. Maria Vilkuna hyväksyy siirron. Kysytty Frediltä korpuksen siirrosta 17.8.2007 ja Kimmolta 20.8.2007.	
Korpus on siirretty, ja se on hakemistossa /l/mrc-uhlcs/general-linguistics/uralic-lgs/finnish. Hakemisto on suljettu niin kauan kunnes Fred ja Kimmo ovat kommentoineet korpuksen käyttöoikeuksia.	
	../incoming
Hakemisto on suljettu.	
Ks. myös erillinen luettelo hakemistoista, joita ei ole siirretty (hakemistot-joita-ei-ole-siirretty-8-9-2008.pdf). Hakemistoissa on useita hakemiston /corp/fin alihakemistoihin kuuluvia tiedostoja.	
/corp/ger	
Korpuksesta ei ole README-tiedostoa, eikä muutenkaan ole ollut saatavilla mitään tietoa korpuksen alkuperästä. Korpuksen ryhmänä on uhcorp1.	
Ehdotus: korpus jää ryhmän li-adm hallintaan. Yritetään samalla selvittää tietoja korpuksen alkuperästä. Ehkä Kimmo voisi sanoa, onko korpus saatu saksan laitokselta. Minun muistini liitti korpuksen alkuperän Maria Vilkunaan ja Lauri Carlssoniin, mutta kun kysyin asiasta Marialta, hän ei muista tästä korpuksesta mitään.	
/corp/multi-ling	
Hakemisto on suljettu, sen omistaja on root.	
Onko tarpeen siirtää hakemisto CSC:lle?	
/corp/rus	
	../journal, ../tass
../journal: Korpus on valmistettu Lingsoftilla. README-tiedostossa ei kerrota mitään lähteestä. Tiedostot olivat Liisa Vilkin nimissä. Liisa ei kuitenkaan muistanut korpuksen olemassaoloa.	

<p>../tass: Korpus on saatu Lingsoft Oy:n kautta Jyri (Georg?) Sooberilta. Korpusta on työstetty yleisen kielitieteen laitoksella.</p> <p>Näiden kahden korpuksen kohdalla pitäisi ensin selvittää, kenen puoleen Lingsoftilla pitää kääntyä. Lingsoftin verkkosivuilla ei ole kuitenkaan mitään tietoa henkilökunnasta. Lingsoftille lähetetään tiedustelu korpuksista 10.0.2007. Kopio kirjeestä toimitetaan Kimmolle. Päätös siitä, voidaanko korpukset laittaa ainakin tutkijoiden käyttöön, selviää sen jälkeen, kun korpusten alkuperä saadaan selvitettyksi.</p>	
	../ryscard
<p>Korpuksen alkuperäiset tekstit ovat Uppsalan yliopistosta. Korpusta on työstetty yleisen kielitieteen laitoksella. Korpuksesta tiedusteltu Uppsalasta 10.9.2007. Korpuksen README-tiedostossa on yhteystiedot. Kimmolle on toimitettu kopio kirjeestä.</p>	
	../upsala
<p>Korpus on valmistettu Uppsalan yliopistossa, mutta sitä on työstetty myös yleisen kielitieteen laitoksella. Alkuperäisiä tekstejä koskevasta sopimuksesta on kysytty Uppsalan yliopistosta 10.9.2007, koska korpuksesta ei ole löytynyt sopimusta.</p>	
/corp/somali	
<p>Korpus on laadittu yleisen kielitieteen laitoksella. Korpuksen lukuoikeus on ryhmällä fin-corp (ryhmä käsittää laitoksen tutkijoita ja opettajia). Jos korpuksen teksteistä on sopimus, joka sallii korpuksen olevan laitoksen käytössä, ei pitäisi olla esteitä korpuksen siirrolle. Korpuksen alkuperäisiä tekstejä koskevasta sopimuksesta on tiedusteltu Frediltä ja Kimmolta.</p> <p>Ehdotus: korpus voi jäädä ainakin laitoksen tutkijoiden käyttöön.</p>	
/corp/spoken ja /corp/fin/puhekieli	
<p>Yleisen kielitieteen laitos on osallistunut puhekielen korpusten laatimiseen, mutta en ole nähnyt sopimusta, jolla alkuperäiset tekstit annetaan yleisen kielitieteen laitoksen käyttöön. Kun kysymyksessä on yhteistyöhanke, laitos on yksi mahdollinen instituutio, joka vastaa korpuksen jakelusta. Alkuperäisten aineistojen sopimuksista on tiedusteltu Frediltä ja Kimmolta.</p> <p>Ks. myös erillinen luettelo hakemistoista, joita ei ole siirretty (hakemistot-joita-ei-ole-siirretty-8-9-2008.pdf).</p> <p>Ehdotus: korpus voi jäädä ainakin laitoksen tutkijoiden käyttöön.</p>	
/corp/swe	
	tiedostot swepre.bta ja betrul, hakemisto /bin jossa myös on tiedosto swe-sent.bta
<p>Näiden siirrosta on kysytty Frediltä.</p>	
	../pre ja ../snt
<p>Korpuksen siirrosta kysytty Frediltä ja Kimmolta.</p>	

/corp/bin ja /corp/tags

Kimmo: onko jokin näistä hakemistoista tai niissä olevista tiedostoista sellainen, joka olisi syytä siirtää CSC:lle UHLCS:n korpusten yhteyteen, ennen kuin angarak kaatuu?

Muista sopimuksista:

ECI/MCI: <http://www.elsnet.org/eci.html>, <http://www.elsnet.org/ecilisting.html> (List of contents of the [ECI CDROM](#))

1) mul01 = Type: Financial, Language: En/Fr/Ge, size (K words): 566, Description: Financial reports from Union Bank Switz. (most french-german)

2) fre01 = Type: newspaper, Language: french, size (K words): 4121, Description: Text from Le Monde newspaper, consisting of articles from September and October 1989, and January 1990.

Sopimuksien mukaan laitoksella on nämä korpukset. Voidaanko ne kopioida CSC:n koneelle, missä ne olisivat laitoksen tutkijoiden käytössä? Korpusten käyttöoikeudet koskevat tutkimusta.