

Korpuksia (UHLCS) koskevista sopimuksista (viimeinen päivitys 10.9.2007)

/corp/bible	
	../BhebraicaS, ../in_future, ../Nestle-Aland26
Hakemistot ovat suljettuja. Anssi Yli-Jyrä tietää, voidaanko korpus siirtää CSC:lle. Anssi: korpuksat siirretään, ja ne pidetään suljettuna siihen saakka, kunnes korpusten sopimukset ovat kunnossa.	
Anssi Yli-Jyrä	../KRaamattu38
Hakemiston README-tiedostosta saatavan information perusteella voidaan päätellä, että koska Raamattu hakemistossa KRaamattu38 ei ole tekijänoikeuslain alainen, se voidaan siirtää CSC:lle. The Finnish Church Bible 1933/1998 is not under copyright. README: Morfologisesti analysoitu aineisto: Academic license.	
Anssi Yli-Jyrä	../KRaamattu92
Sopimuksesta: Vuoden 1992 käännöstä koskeva sopimus on tehty yleisen kielitieteen laitoksen ja Kirkon keskusrahaston välillä. Sijainnista: "... raamatunsuomennoksen sijoitettavaksi laitoksen tietokoneille tutkimus- ja opetustarkoitusta varten asianomaisten tutkijoiden ja opiskelijoiden käyttöön". "Laitos huolehtii raamatunsuomennoksen tekstin tietosuojasta asianmukaisilla ja kohtuullisilla toimenpiteillä"... Anssi: korpuksen saa siirtää. (README-tiedostossa todetaan vain Kirkon keskusrahaston tekijänoikeudet.) (Huom. UHLCS ei ole kustantaja eikä kustantajan edustaja.)	
/corp/eng	
	../brown
<p>Brown Corpus</p> <p>"The Brown Corpus of Standard American English was created at the Brown university by by W. N. Francis and H. Kucera. It contains one million words of written American English, taken from publications from 1961. The texts are all appr, 2,000 words long and grouped into 15 categories. More information about the content of the corpus can be found in the <i>Brown Corpus Manual</i> by Francis and Kucera, available on the ICAME webpage.</p> <p>This version of the Brown corpus has word-class annotation and has been converted into XML and indexed to be used with the XAIRA program. It is distributed on the BNC Baby CD together with the <i>BNC Baby</i> and <i>BNC Sampler</i> corpora. How to order "</p> <p>***</p> <p>Yllä oleva korpuksen esittely on BNC:n verkkosivulta. http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=products: (ks. myös http://www.natcorp.ox.ac.uk/quick.xml; ICAME: http://icame.uib.no/newcd.htm)</p> <p>British National Corpus, sopimuksesta: British National Corpus: osa BNC:sta on luovutettu yleisen kielitieteen laitoksen käyttöön sopimuksella, jossa seuraavat ehdot ovat tärkeitä: 1) Korpus annetaan instituution/ryhmän käyttöön. Korpusta voi käyttää tutkimuksen ma-</p>	

terialina, tai sitä voidaan käyttää valmistettaessa ”language products”. Sopimuksen solmijat ovat laitos ja Oxford University Computing Services.

2) Sitä, missä korpus fyysisesti sijaitsee, ei ole rajattu.

3) Korpusta ei saa siirtää kolmannelle osapuolelle. Tämän kohdan liitettynä edelliseen kohtaan voi tulkita niin, että korpus voi olla CSC:n koneella, jos sen käyttö tapahtuu korpuksen vastaanottajan hallinnassa (within its control) ja on esim. ko. laitoksen tutkijoiden käytössä. Korpuksesta saa olla varakopioita (backup).

4) Korpuksen sopimuksessa puhutaan tutkimusryhmästä, mutta en löytänyt kohtaa, jossa sen käyttö olisi rajattu jollekin tietylle tutkimusryhmälle.

5) Korpusta koskevat kaikki tavalliset copyright-oikeudet, kuten esim. julkaisemista ja erilaista korpuksen levittämistä koskevat rajoitukset.

6) Sopimukseen kuuluu erikseen kohta, joka koskee korpuksen ylläpitoa ja tietoturvaa ja korpuksen vastaanottajan vastuuta.

7) Sopimuksen kesto on viisi vuotta. Sen jälkeen se pitää uudistaa. Sopimus on allekirjoitettu heinäkuussa 1995. Jos sopimusta ei ole uudistettu, se ei siis ole enää voimassa.

../grolier

Atro Voutilainen, 5.6.2007:

”Grolier-data on ekstrahoitu laitokselle hankitusta G.El.Encyclopediasta, alunperin ihan omaan käyttööni. Sen levityksestä ei ole tietääkseni tehty sopimusta.”

(ks. <http://go.grolier.com/orl/pages/gmfeatrs.htm>)

Korpuksesta ei ole sopimusta. Korpus on suljetun ryhmän käyttöön, joten laitos päättää, saako tämä suljettu ryhmä käyttää korpusta CSC:llä. Kysytty Frediltä sopimuksesta 18.8.2007.

../gutenberg

<http://promo.net/pg/>:

About Project Gutenberg

Thank you for your interest in Project Gutenberg!

Project Gutenberg is the Internet's oldest producer of FREE electronic books (eBooks or eTexts).

[What books will I find in Project Gutenberg?](#)

Project Gutenberg is the brainchild of [Michael Hart](#), who in 1971 decided that it would be a really good idea if lots of famous and important texts were freely available to everyone in the world. Since then, he has been joined by hundreds of volunteers who share his vision.

Now, more than thirty years later, Project Gutenberg has the following figures (as of November 8th 2002): 203 New eBooks released during October 2002, 1975 New eBooks produced in 2002 (they were 1240 in 2001) for a total of 6267 Total Project Gutenberg eBooks. 119 eBooks have been posted so far by [Project Gutenberg of Australia](#).

Click [here](#) for the full PG story and [here](#) for the latest [News](#), and learn about the [Stockholm Challenge Award](#) recently won by Project Gutenberg in the category "Culture".

Katso myös verkkosivut <http://gutenberg.hwg.org/index.html> ja

<http://gutenberg.hwg.org/checkdoc1.html>.

Korpus voidaan siirtää. Tästä korpuksella ei tarvita sopimusta.	
	../lob
<p>The Lancaster/Oslo-Bergen Corpus (LOB) Approximately 1,000,000 words of British written English dating from 1960. The corpus is made up of 15 different genre categories. Available as orthographic text, and tagged with the CLAWS1 part-of-speech tagging system. The Leeds-Lancaster Treebank and Lancaster Parsed Corpus are analyzed subsamples of the LOB corpus. For further information see the corpus manual (1978) and the tagged corpus manual (1986). (There is a local on-line copy of the tagged corpus manual at Lancaster.) Lähde: http://www.comp.lancs.ac.uk/computing/research/ucrel/corpora.html#lob. ICAME: http://icame.uib.no/newcd.htm</p> <p>Tästä korpuksella ei ole löytynyt sopimusta. Kysytty Frediltä sopimuksesta 18.8.2007 ja Kimmolta 20.8.2007.</p>	
	../london-lund
<p>Editor : Svartvik, Jan Format : Unknown markup Source : London-Lund corpus of spoken English. — [Lund] : [Lund University Press], s.d. — Catalogued on RLIN. — Archival copy is at BE. Size : 0 files : ca. 0 kilobytes Language : English Distribution : Available for non-commercial use only with prior permission of the depositor and on condition that this header is included in its entirety with any copy distributed Transcriber : Svartvik, Jan Lähde: http://ota.ahds.ac.uk/search/info.perl?show=ID0168.</p> <p>ICAME: http://icame.uib.no/newcd.htm</p> <p>Tästä korpuksella ei ole löytynyt sopimusta. Kysytty Frediltä sopimuksesta 18.8.2007 ja Kimmolta 20.8.2007.</p>	
	../misc
<p>Näistä ei ole löytynyt sopimuksia. Kysytty Frediltä sopimuksesta 18.8.2007 ja Kimmolta 20.8.2007. Mukana on aineistoja, jotka ovat tekijänoikeuslain ulkopuolella.</p>	
Atro Voutilainen	../susanne
<p>Lähde (korpuksen verkko-osoite oli muuttunut elokuussa): http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/public/susanne.html</p> <p>INTRODUCTION The SUSANNE Corpus has been created, with the sponsorship of the Economic and Social Research Council (UK), as part of the process of developing a comprehensive NLP-oriented taxonomy and annotation scheme for the (logical and surface) grammar of English.... Korpuksen kopioinnista: korpus on vapaasti kopioitavissa, joten se voidaan siirtää muun</p>	

UHLCS:n aineiston mukana (ks. http://www.grsampson.net/Resources.html , http://www.cogs.susx.ac.uk/users/grs2/ftpBufPageSue.html).		
		../wsj
<p>Linguistic Data Consortium: Laitos on sopimuksen LDC:n kanssa v. 1997. Sopimuksen mukaan laitos saa LDC:lta TIPSTER-korpuksen (vol. 1), = wsj, ks. http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93T3A.</p> <p>1) Sopimus koskee korpuksen käyttöä tutkimuksessa ja NPL:iin liittyviä töitä. 2) Korpusta ei saa levittää. 3) Korpusta saavat käyttää henkilöt, joiden käyttöä korpuksen vastaanottaja voi kontrolloida (vrt. UHLCS:llä käyttöluja ja korpuksen yhdyshenkilö). 4) Käyttäjistä: "Individuals may be allowed access to the Information only after completing a copy of the <u>Application of an Individual to Use the TIPSTER Information-Retrieval Text Research Collection, volume 1</u>, as provided with application. The access is to be terminated when the conditions of application no longer apply." Organisaatio pitää listaa henkilöistä, jotka "with current and recently-terminated access to the information". 5) Korpuksia koskevaa informaatiota voidaan välittää vain niiden henkilöiden kesken, joilla on korpuksen käyttöoikeus. 6) Korpusta koskevat kaikki tekijänoikeuksia koskevat rajoitukset. 7) Korpuksen levitys: "Except as specifically permitted above and as necessary to use and maintain the integrity of the information on computers used by the organization, the display, reproduction, transmission, distribution or publication of the information is prohibited." 8) Korpuksen käyttö on rajattu koskemaan noin 20 henkilöä. 9) Sopimus on solmittu LDC:n ja laitoksen kesken. Sopimuksessa ei ole määritelty, missä korpus saa sijaita, ja mikä on korpuksen viimeinen käyttöpäivä. Tämän perusteella korpuksen voi siirtää CSC:n koneelle, kunhan vain laitos kontrolloi korpuksen käyttöä.</p>		
Timo Järvinen		/espanol
Teksti on vapaasti saatavilla osoitteessa http://eprintweb.org/S/authors/All/ch/CHANDLER-BURNS . Ehdotus: ei siirretä.		
Kimmo Koskenniemi		/fin
<p>Suomen kielen aineistoista: Yl. kielitieteen laitos & Kotus (xx.4.2000), pykälä: 2.7. "Osapuolet luovuttavat toisilleen käsittelemänsä tekstit niin kuin Suomen kielen tekstipankin aiemmassakin yhteistyössä (noudattaen kuitenkin liitteenä olevan käyttöoikeussopimuksen ehtoja). Tekstipankista pidetään turvallisuussyistä identtiset kopiot kummankin koostajaosapuolen hallussa..." 2.8. "...yleisen kielitieteen laitoksen ja Kotimaisten kielten tutkimuskeskuksen tutkijat saavat käyttää Kielipankkiin sijoitettavaa tekstipankkia maksutta...Aiemman LE-PAROLE-nimisen hankkeen aineistot (5,1 milj. Sanaa (HY) 10,4 milj. Sanaa (KTKK) voivat osapuolten puolesta sijaita sanassa CSC:n kielipankissa kuin TIETOKANTAHANKKEESSA tehdyt aineistot." 3.4. "TIETOKANTAHANKKEEN tuloksena syntyvien CSC:n kielipankkiin tulevien aineistojen koostajat myöntävät näiden aineistojen käyttöluja TIETOKANTAHANKKEEN johtoryhmän tarkemmin päättämällä ja suosittamalla</p>		

tavalla...”	
<p>Kohdassa 2.7. viitataan erilliseen käyttöoikeussopimukseen, missä täsmennetään käyttöoikeudet (minulla on käytössäni vain kopio CSC:n kanssa tehdystä sopimuksesta). Siinä sopimuksessa määritellään kolme eri tasoa käyttöoikeuksissa: osaa teksteistä voidaan käyttää tutkimuksessa, osaa tutkimuksessa ja siihen perustuvat kieliteknologiset sopimukset, ja osalla teksteistä on vapaa käyttö. Mutta kohta 3.4. voidaan tulkita niin, että aineistoja voidaan käyttää tutkimuksessa, opetuksessa ja kieliteknologisten ohjelmien valmistamisessa.</p> <p>Aikaisemmassa sopimuksissa (199?), jotka koski tekstipankin osaa 1a, sanotaan korpusten käytöstä seuraavaa: §2. ...Käyttöoikeus koskee tekstien hyödyntämistä tieteellisiä tutkimustarkoituksia varten ja oikeutta hyödyntää tekstejä tutkimalla selville saatuja kielen piirteitä...”, 1b: ”Käyttöoikeus koskee tekstien hyödyntämistä tieteellisiä tutkimustarkoituksia varten sekä sellaisia kaupallisia kieliteknologisia tms. sovelluksia varten, jotka eivät loukkaa tekijänoikeuslakia”, 1c: ”Tekstien käyttöoikeus on vapaa...”, ja 2a, 2b ja 2c: §4: ”Käyttöoikeus koskee tallennetekstien hyödyntämistä tieteellisiä tutkimustarkoituksia varten ja oikeutta hyödyntää tallennetekstejä tutkimalla saatuja kielen piirteitä...”</p>	
	../bin
Kimmo tietää, voidaanko siirtää skriptit CSC:n koneelle.	
	../fintag/, ../skh, ../sks, ../otava, ../misc, ../wsoy
<p>Hakemistossa olevat aineistot on laadittu yleisen kielitieteen laitoksella. Hakemistossa /fintag on eri versioita mm. teksteistä, jotka ovat hakemistoissa ../ksk, ../otava ja ../wsoy. Kysytty Frediltä alkuperäisten siirrosta ja aineistojen sopimuksista 17.8.2007 ja Kimmolta 22.8.2007. Ehdotus: korpukset voidaan siirtää, koska korpukset ovat CSC:llä olevassa fintag-korpuksessa.</p>	
	../ftc, ../hs90, ../hs95, ../hs96, ../hs97
<p>Korpukset kuuluvat Suomen kielen tietopankki -hankkeen tietokantaan. Ehdotus: hakemistot voidaan siirtää. Hakemisto /ftc sisältää yo. korpusten eri versioita. Kysytty Frediltä korpusten siirrosta ja alkuperäisten aineistojen sopimuksista 17.8.2007 ja Kimmolta 22.8.2007. Ehdotus: korpukset voidaan siirtää, koska aineistoista on sopimukset, ja aineistot ovat myös osia muista korpuksista..</p>	
	../hkv
<p>Omistajat Auli Hakulinen, Fred Karlsson ja Maria Vilku. Korpusten tekstiaineistot ovat suhteellisen lyhyitä. Korpusten siirrosta tiedustelu Maria Vilkulle 6/8/2007. Maria Vilku hyväksyy siirron. Kysytty Frediltä korpusten siirrosta 17.8.2007.</p>	
Kimmo Koskenniemi:	../hs, ../sk75, ../kuvaristikot
<p>Näitä korpuksia ei löytynyt laajemmista korpuksista. Voidaanko korpukset siirtää pienen ryhmän käyttöön?</p>	
	../parole
<p>/hs, /hs90, /sk87 sekä hakemisto /parole: Aineistot hakemistoissa /hs, /hs90 ja /sk87 ovat osa LE PAROLE -hankkeen aineistoja.</p>	

Hakemistot voidaan siirtää, koska ne kuuluvat yleisen kielitieteen laitoksen hallintaan.		
raw -> /proj/parole/corp/hs95/raw, raw -> /proj/parole/corp/hs96/raw, raw -> /proj/parole/corp/hs97/raw :		
Hakemistot on suljettu. Tiedostot hakemistossa /raw sisältävät alikorpusten hs95, hs96 ja hs97 raakadataa. Aineistot voidaan siirtää, koska ne kuuluvat osana parole-korpukseen, ja kuuluvat siis yleisen kielitieteen laitoksen hallintaan. Hakemisto /corp/parole on Sari Salmisuon hallinnassa. Tiedusteltu Sarilta, onko hakemistossa Sarin omia tiedostoja, koska hakemisto on suljettu.		
Kimmo Koskenniemi	../incoming, ../pub (suljetut hakemistot)	
Voidaanko hakemistot siirtää?		
Antti Arppe	../sfnet2003	
Hakemiston siirrosta tiedusteltu Antti Arpelta 6.8.2007. Vastaus 7.8.2007: korpusta ei siirretä CSC:n koneelle.		
Kimmo Koskenniemi		/ger
Korpuksesta ei ole README-tiedostoa. Muistitietojen perusteella (kuulemani mukaan) korpuksen hakivat Maria Vilkuja ja Lauri Carlsson, mutta muistitieto ei ole varmaa. Korpus on rajoitettu ryhmän käyttöön. Olen tiedustellut 22.8.2007 Maria Vilkunalta, tietääkö hän mitään korpuksen alkuperästä, mutta hän ei muista asiasta mitään. Ehdotus: korpus voidaan siirtää. Se jää ryhmän li-adm alaisuuteen, jos sitten joskus selviäisi, keltä kysyä sopimuksesta ja korpuksen käyttöluvasta.		
Martti Nyman:	/lat	
Corpus: American Philological Association / LAT-1 (http://www.apaclassics.org/) Martti Nyman on hankkinut korpuksen. Korpuksesta ei ole käytettävissä paperilla olevaa sopimusta. Tiedusteltu Martilta 16.8.2007, voidaanko korpus siirtää. Martilta myöntävä vastaus 21.8.2007.		
	../multi-ling	
Hakemisto on suljettu, sen omistaja on root. Mitä hakemisto sisältää, ja voidaanko hakemisto siirtää CSC:lle?		
Pirkko Suihkonen		/quechua
Aineiston skannaus on maksettu yl. kielitieteen laitoksen varoista. Pirkko: aineisto on saatu Perussa työskenneeltä kulttuurimaantieteilijältä (Kimmo Juvas; mukana myös Tarja Keisteri; Tarja on kuollut, Kimmo Juvas työskentelee tietääkseni SPR:ssä.). Aineistosta ei ole tehty virallista sopimusta. Korpus on minun työhakemistossani CSC:llä, kunnes se voidaan antaa suljetun ryhmän käyttöön, niin kuin se nytkin on.		
Eri omistajia		/rus
	../tampere	
Tampereen korpuksesta kysytty Hannu Tommolalta 21.8.2007. Vastaus 25.8.2007. Sopimus kattaa korpuksen sijainnin UHLCS:n yhteydessä. Korpus on tutkimuskäyttöön. Korpuksen käyttöön ei tarvitse pyytää erillistä lupaa venäjän kielten korpusten yhdys-		

kilöltä.	
	../fowler
Korpus on vapaasti levitettävissä, joten korpuksen voi siirtää (ks. korpuksen mukana oleva README-tiedosto). Korpuksen käyttöön ei tarvitse pyytää erillistä lupaa venäjän kielten korpusten yhdyshenkilöltä.	
	../journal
M. Kopotevin mukaan korpus on valmistettu Lingsoftilla. README-tiedostossa ei kerrota mitään lähteestä, mutta tiedostot on työstänyt Liisa Vilkki, joka ei muistanut asiasta mitään. Kysytty 10.9.2007 Lingsoftilta korpuksen sopimuksesta.	
	../ryscard
Korpus on valmistettu Uppsalan yliopistossa (M. Kopotev). Kysytty UDAC:sta (Natural Language Processing Group) sopimuksesta 10.9.2007. Korpuksen README-tiedostossa on yhteystiedot. Korpusta on työstetty yleisen kielitieteen laitoksella.	
	../spoken
Korpus on valmistettu Venäjällä (the Institute of Russian Language (Moscow) (M. Kopotev)). Korpusta on työstetty yleisen kielitieteen laitoksella. Kysyttävä Hannu Tommolalta, voidaanko korpus siirtää. Korpuksen alkuperäisteksteistä on kysytty Hannu Tommolalta 25.8.2007. Hannun mukaan korpuksesta on tehty sopimus sinä aikana, kun hän oli HY:ssa laitoksen esimies. Sopimusta ei kuitenkaan löytynyt (Arja Kirvesmäki 29.8.2007). Korpuksen saa ottaa käyttöön Arto Mustajoen luvalla. Korpus on tutkimuskäyttöön.	
	../tass
Korpus on saatu Lingsoft Oy:n kautta Jyri Sooberilta. Korpusta on työstetty yleisen kielitieteen laitoksella. Kysytty 10.9.2007 Lingsoftilta korpuksen sopimuksesta.	
	../uppsala
Korpus on valmistettu Uppsalan yliopistossa, mutta sitä on myös työstetty yl. kielitieteen laitoksella. Korpuksen siirtoon pitää varmistaa lupa korpuksen valmistajalta (http://www.slaviska.uu.se/korpus.htm), koska korpuksen verkkosivulla pyydetään ”to prevent the unauthorized duplication or distribution of the data”. Prof. Lennart Lönngrénin on otettu yhteyttä 10.9.2007. Korpuksen verkkosivulla olevasta tekstistä voi päätellä, että korpuksen voi ottaa tutkimuskäyttöön ilman, että siihen tarvitsee pyytää erityistä lupaa.	
	/somali
Korpus on laadittu yleisen kielitieteen laitoksella. Korpuksen luku-oikeudet ovat ryhmällä fincorp (ryhmä käsittää laitoksen tutkijoita ja opettajia). Jos korpuksen teksteistä on sopimus, joka sallii korpuksen olevan laitoksen käytössä, ei pitäisi olla esteitä korpuksen siirrolle. Kysytty Frediltä 17.7.2007 ja Kimmolta 22.8.2007.	
	../spoken ja /fin/puhekieli
Alkuosa README-tiedostosta: Helsingin puhekielen aineisto	
”Helsingin puhekielen tutkimus on aloitettu Helsingin yliopiston suomen kielen laitoksella v. 1972 Terho Itkosen johdolla, ja Helsingin	

<p>puhekielen aineisto on kerätty vuosina 1972-74. Helsingin puhekielen tutkimushanke liitettiin sittemmin osana Heikki Paunosen johtamaan valtakunnalliseen Valtion humanistisen toimikunnan rahoittamaan</p> <p>Nyky-suomen puhekielen murroksen tutkimus -nimiseen hankkeeseen, joka alkoi esitutkimuksena 1976 ja jatkui tutkimusprojektina 1977-80. Helsingin puhekielen aineisto on kerätty haastattelemalla kolmeen ikä- ja sosiaaliryhmään kuuluvia helsinkiläisiä tunti verran. Jokaisen 127 haastatellun henkilön puheytteestä on litteroitu puoli tuntia puheita. Nauhoitteita säilytetään Kotimaisten kielten tutkimuskeskuksen nauhoitearkistossa, josta niitä voi tarpeen mukaan tiedustella. Aineisto on saattu tietokonemuotoon Helsingin yliopiston yleisen kielitieteen laitoksella (tiedosto jatkuu korpuksen kuvauksella).”</p> <p>Korpusta koskeva sopimus: yllä olevan perusteella voitaisiin päätellä, että yleisen kielitieteen laitos on yksi mahdollinen instituutio, joka vastaa korpuksen jakelusta. Ja edelleen: yleisen kielitieteen laitoksen johto päättää, miten jakelu tehdään. Näin ollen ei pitäisi olla syytä, miksi korpusta ei voitaisi ottaa käyttöön myös CSC:n laitteilla.</p>	
Arvi Hurskainen	/swa
	../bin, ../conjoin, ../dialects, ../standard, ../swa-tmp, README-tiedosto
Tiedusteltu Arvi Hurskaiselta korpusten siirrosta 07/05/2007. Myöntävä vastaus 22.8.2007. Arvi kirjoittaa uudet käyttöohjeet sen jälkeen, kun korpuksat on siirretty CSC:lle.	
	/swe
Fred Karlsson	tiedostot swepre.bta ja betrul, hakemisto hakemisto bin jossa tiedosto swe-sent.bta
Näiden siirrosta on kysytty Frediltä 17.8.2007.	
Jan K. Lindström	../fisc
Tiedusteltu 6/8/2007 Jan Lindströmiltä, voidaanko korpus siirtää. Myönteinen vastaus 20.8.2007. Korpukselta siirretään osia Janin sähköpostitse lähettämien ohjeiden mukaisesti. Kaikkia tiedostoja ei siirretä.	
Jan K. Lindström	../hki
Hakemistossa esiintyy ryhmä fisc2. Korpuksen siirrosta kysytty Jan K. Lindströmiltä 24.8.2007. Myöntävä vastaus samana päivänä. Korpus siirretään fisc-korpuksen mukana.	
Fred Karlsson	../pre ja ../snt
Korpuksen siirrosta kysytty Frediltä 17.8.2007.	
	/viro
Kazuto Matsumura:	../viro1
Tiedusteltu korpusten siirrosta Kazuto Matsumuralta 5.8.2007. Myöntävä vastaus 9.8.2007. Alkuperäisteksteistä ei ole kirjallista sopimusta.	
Maria Vilku:	../viro2

Tiedusteltu korpusten siirrosta Maria Vilkunalta 5/8/2007. Myönteinen vastaus 6/8/2007. Alkuperäisteksteistä ei ole kirjallista sopimusta.	
root	words
Kari Pitkänen & Timo Järvinen: http://www.ling.helsinki.fi/uhlcs/data/helsinki-corpora-1.html (Information on the corpora was collected in the early 90's.) (Huom. Tämä linkki ei toimi.). Käsitykseni mukaan ei voida osoittaa, että sanaluettelot rikkovat jonkun tekijänoikeuksia (sanaluetteloita on mahdollista tehdä hetkessä mistä tahansa aineistosta). Näin ollen korpus voidaan siirtää CSC:lle.	
root	/yiddish
Jussi Karlgren	../yiddish2
Aineistosta tiedusteltu Jussi Karlgrenilta 07/09/2007. Aineisto on saatu Jussilta, eikä aineistosta ole sopimusta. Jussi on saanut korpuksen Ellen Princeltä. Suullisen sopimuksen perusteella aineisto voi olla suljetun ryhmän käytössä. Silloin ei pitäisi olla merkitystä sillä, missä aineisto sijaitsee.	
Johan G. Mångård	../yidish
Hakemisto on suljettu. Otettu yhteyttä korpuksen siirron vuoksi J. Mångårdiin 16.8.2007. Myönteinen vastaus 16.8.2007. Keskustelun jatkuessa sovittiin, että korpusta ei siirretä. John on tämän lukuvuoden hoitovapaalla ja jatkaa opintojaan ensi vuonna. Korpuksesta on useita turvakopioita Johnin omalla koneella. Olen neuvonut Johnia ottamaan ensi vuonna suoraan yhteyttä CSC:hen.	

Aineistot, joista on laadittu UHLCS:n verkkosivuilla oleva aineistojen käyttöä koskeva sopimus, tai joihin ei liity tekijänoikeuksiin liittyviä erityiskysymyksiä.	
IBT	/caucasian-lgs
IBT: tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	alihakemistot /avar, /lak, /tabassaran
IBT	/chukotko-kamchatkan-lgs
IBT: tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	alihakemistot /chukchi, /koryak
IBT	/iranian-lgs
IBT: tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	alihakemistot /kurdish, /ossete, /tajik
IBT	/mongolic-lgs
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007;	alihakemisto /kalmyk

sähköpostit 5.7. ja 8.8.2007).		
IBT		/tungusic-lgs
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	alihakemistot /even, /evenki, /nanay	
		/turkic-lgs
IBT	alihakemistot /azerbaijan, /balkar, /bashkir, /crimean-turkish, /gagauz, /khakas, /kirghiz, /tatar, /turkmen, /tuvan, /uighur, /uzbek, /yakut	
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).		
	alihakemisto /chuvash	
André Hesselbäck:	../paasonen	
Tiedustelu korpusten siirrosta André Hesselbäckiltä, Myöntävä vastaus 11.7./2007.		
Daniel Kimmage:	/uzbek, ..alihakemisto /sanakirja	
Voidaan siirtää. Uzbekin sanakirja on lahjoitettu UHLCS:iin, eikä pelkkä sijainnin muutos vaikuttaa sanakirjan käyttöoikeuksiin.		
Voidaan siirtää. Kriminturkin sanasto on lahjoitettu UHLCS:iin, eikä ole odotettavissa, että pelkkä sijainnin muutos vaikuttaa sanaluettelon käyttöoikeuksiin.	alihakemisto /crimean-turkish: sanasto	
IBT:		/ukrainian
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).		
		uralic-lgs
Pirkko Suihkonen, Eeva Ahonen	alihakemisto /corp-proj	
Hakemistossa on ECHO-hankkeen rahoituksella tehtyjä skriptejä (Eeva Ahonen), joiden pohjana on Max Planck Instituutin (Leipzig) rahoituksella (Vili Junkkari) ja vapaaehtoisena työnä (Pirkko Suihkonen) tehtyjä tiedostoja.. Vastaavia tiedostoja ja hakemistoja on kaikissa alihakemistoissa, joiden nimessä on CAP-in-preparation (CAP = kielen lyhennettä vastaava kirjainryhmä). Hakemistoissa olevat alihakemistot ja tiedostot ovat käytettävissä UHLCS:n korpuksia koskevassa työssä. Hankkeiden vastaava: Pirkko Suihkonen		
Manja Lehto: korpusten siirrosta tiedusteltu Manja Lehdolta. Myöntävä vastaus 07/05/2007.	alihakemisto /ingrian	
	alihakemisto /kamas	
Jarmo Alatalo:	texts-donner, list-of-tags	
Tiedusteltu korpusten siirrosta Jarmo Alatalolta 11.7.2007. Myöntävä vastaus 19.8.2007.		
	alihakemisto /enets	
IBT:	../New-Testament	

Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	alihakemisto / karelian
	../dvina-karelian
IBT:	../Books-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	../livvi
IBT:	../Bible-of-Children, ../Books-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Miikul Pahomov:	../lude
Korpus on lahjoitettu käytettäväksi UHLCS:ssä. Tässä tapauksessa korpuksen sijainti ei vaikuta korpuksen käyttöön.	
	alihakemisto / khanty
IBT:	../Books-of-Children
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Merja Salo:	../atlym-dialect, ../kazym-dialect, ../konda-dialect, ../nizjam-dialect, ../obdorsk-dialect, ../synja-dialect, /rugin
Tiedusteltu korpusten siirrosta Merja Salolta. Myöntävä vastaus 06/04/2007-	
	alihakemisto / komi
	alihakemisto permyak
IBT:	../Books-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	alihakemisto /zyrian
IBT:	../Books-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Paula Kokkonen:	../komi-texts, ../komi-texts-snt, ../morphologically-tagged-corpora
Tiedustelu korpusten siirrosta 11.7.2007. Myöntävä vastaus 7.8.07.	
Jack Rueter, mordvalaiset kirjailijat ja kustantamot.	../novels, ../originals-novels
Korpusten siirrosta tiedusteltu Jack Rueterilta 06/14/2007. Myöntävä vastaus 10.8.2007.	
	alihakemisto / livonian
IBT:	../Books-of-Children
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Seppo Suhonen:	../suhonen

Korpus on n. sivun mittainen näyte eikä kokonsa vuoksi riko tekijänoikeuslakia niiltä kohdin, mikä koskee uudelleen julkaisemista ja levittämistä. Korpusten käyttöä koskevis- sa ohjeissa sanotaan, että lähdetiedot on mainittava kaikissa niissä dokumenteissa, joissa niitä käytetään lähdeaineistona.	
	alihakemisto / mansi
IBT:	../Books-of-Children
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukhol- ma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	alihakemisto / mari
	alihakemisto /eastern-mari
IBT:	../Bible-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	alihakemisto /western-mari
IBT:	../Books-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukhol- ma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
André Hesselbäck:	../hill-mari-texts
Tiedusteltu korpusten siirrosta André Hesselbäckiltä. Myöntävä vastaus 11.7.2007.	
	alihakemisto / mordvin
	alihakemisto /erzya
IBT:	../Bible-of-Children, ../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukhol- ma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Jack Rueter:	../dictionary, ../journals, ../epos, ../novels, ../poetry, ../short-stories, ../morphologically- tagged-corpora
Korpusten siirrosta tiedusteltu Rueterilta 06/14/2007. Myöntävä vastaus 10.8.2007.	
Dennis Estil: Kirje Estilille 11.7.2007.	/historical-word-list
	alihakemisto /moksha
IBT:	../Books-of-Children
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukhol- ma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Dennis Estil:	../historical-word-list
Kirje Estilille 11.7.2007.	
Jack Rueter.	../novels
Korpusten siirrosta tiedusteltu Jack Rueterilta 06/14/2007. Myöntävä vastaus 10.8.2007.	
	alihakemisto / nenets
IBT	../New-Testament
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukhol- ma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	

Tapani Salminen:	../sample-sentences
Tiedusteltu korpuksen siirrosta Tapani Salmiselta. Myöntävä vastaus 06/06/2007.	
	alihakemisto / saami
IBT	../kildin-saami
Tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
	../northern-saami
Irja Seurujärvi-Kari:	../report
Tiedusteltu korpuksen siirrosta Seurujärvi-Karilta. Myöntävä vastaus 07/23/2007.	
Kerttu Vuolab:	../vuolab
Korpus on lahjoitettu käytettäväksi UHLCS:ssa. Sopimuksen ovat allekirjoittaneet kustantaja (Davvi Girji), kirjailija (Vuolab) ja laitos. Korpuksen sopimuksia laadittaessa oli vaikea saada yhteyttä Kerttu Vuolabiin, ja siksi en ole yrittänytkaan uusia yhteydenottoa. Kustantaja Davvi Girji oli hyvin positiivinen korpushankkeelle. Korpus on keskeneräinen, koska ei ole ollut resursseja toimittaa korpusta transliterointia kuntoon. Tässä tapauksessa ei ole odotettavissa, että korpusten siirto muuttaa korpusten käyttöoikeuksia, koska hallinnolliset suhteet säilyvät.	
Olavi Korhonen:	../ume-saami
Tiedusteltu korpusten siirrosta Olavi Korhoselta. Myöntävä vastaus 07/16/2007.	
Jarmo Alatalo	alihakemisto / selkup
Kirje Alatalolle 11.7.2007. Myöntävä vastaus 19.8.2007.	../h-dialects, ../ivankino-dialect, ../ket-dialect, ../tundra-dialect, ../tym-dialect, ../upper-ob-dialect, ../list-of-tags
	alihakemisto / udmurt
IBT	/Books-of-Children, /New-Testament,
tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma) 07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).	
Pirkko Suihkonen, Bibinur Zaguljaeva, udmurtilaiset kirjailijat ja kustantamot.	/udmurt-snt, /udmurt-texts-unmodified
Olen ottanut yhteyttä Bibinur Zaguljaevaan jo hankkeen ”Udmurt-English-Finnish Dictionary” vuoksi, mutta en ole saanut vastausta. Korpuksessa olevien tekstien kustantajiin on käytännössä mahdotonta saada yhteyttä, kun yhteyttä ei ole saatu edes Zaguljaevaan. Korpusten yhtenä omistajana luotan siihen, että korpukset voidaan siirtää CSC:lle.	
Pirkko Suihkonen.	/udmurt-statistical-data,
Korpus ja siihen liittyvät README- ja LUE TAMA -tiedostot voidaan siirtää CSC:lle. Pirkko Suihkonen.	
	alihakemisto / vepsian
IBT:	../Bible-of-Children, ../Books-of-Children, ../New-Testament
tiedusteltu korpusten siirrosta IBT:tä (Helsinki) 07/04/2007, IBT (Moskova ja Tukholma)	

07/05/2007. Myöntävä vastaus (kirje 1.8.2007; sähköpostit 5.7. ja 8.8.2007).

Hallintoon liittyviä alihakemistoja /corp-hakemistossa ja yksi tyhjä hakemisto (/indo). Onko jokin näistä hakemistoista tai niissä olevista tiedostoista sellainen, joka olisi syytä siirtää CSC:lle UHLCS:n korpusten yhteyteen?					
drwxr-x--x	3 root	root	8192	2000-02-06 10:46	adm
drwxrwxr-t	3 root	root	8192	2000-02-06 10:46	admin
drwxrwxr-t	2 root	260	8192	2000-02-06 10:47	bin
Hallintoon liittyviä hakemistoja.					
drwxr-sr-t	2 19967	274	8192	2000-02-06	indo
Tyhjä hakemisto, joka on poistettu.					
			8192	2000-02-06 10:54	info
Laitoksen toimintaan liittyviä tiedostoja ja hakemistoja.					
drwx-----	2 root	root	8192	2000-02-06 09:58	.tags
Suljettu tiedosto.					

Muista sopimuksista:

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

Monikieliset korpukset on koottu eri toimittajilta, ja eri kielistä koottujen osakorpusten käyttöehdoissa on jonkin verran eroja. Kaikkia osakorpuksia koskee tekijänoikeuksien suojaus.

Sopimuksen mukaan laitoksella on tämä monikielinen korpus. Voidaanko korpus kopioida CSC:n koneelle, missä se olisi suljetun ryhmän käytössä?