

## Liite 2 (UHEL): Helsingin yliopiston FIN-CLARIN-panostus

<https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/FinClarInSiteHY>

Helsingin yliopisto (jäljempänä jäsen) toimii FIN-CLARIN -konsortiossa panostaen FIN-CLARINin ja CLARIN ERICin perussäännön mukaisten tavoitteiden saavuttamiseksi seuraavaa:

- Jäsen toimii CLARIN ERICin teknillisen ja tieteellisen kuvauksen (*CLARIN ERIC Technical and Scientific Description, July 2011*) mukaisena R-tason CLARIN-keskuksena (*Recognized Centre*) ja tarjoaa yhdessä CSC:n kanssa B-tason keskuksen (*Service Providing Centre*).
- Jäsen vastaa yhdessä CSC:n kanssa siitä, että konsortiolla on käytettävissään asianmukainen kansallisen todentamis- ja valtuutusjärjestelmä, joka on yhteensopiva CLARIN ERICin vaatimusten kanssa sekä kehittää näitä järjestelmiä yhdessä CSC:n kanssa.
- Jäsenellä on seuraavassa taulukossa kohdissa 5 ja 6 lueteltuja olemassa olevia ja tekeillä olevia kielivaroja ja jäsen pyrkii saamaan kyseiset kielivarat FIN-CLARINin ja CLARIN ERICin käyttöön tekijänoikeuksien ym. sallimissa puitteissa.
- Jäsen koordinoi CLARIN ERICin kielivaroja koskevien standardien saamista käyttöön kansallisesti ja tukee niiden käyttöönottoa konsortion piirissä sekä edistää kyseisten standardien soveltamista uusiin kielivarioihin omassa piirissään.
- Jäsen tarjoaa FIN-CLARINille asiantuntemusta mm. seuraavilla alueilla: Euroopan ja maailman kielet, korpuslingvistiikka, englannin kielen variaatio (VARIENG huippuyksikkö), kielen universaalit, kieliteknologia mukaan lukien morfologinen ja pintasyntaktinen jäsentäminen and äärellistilaisten transduktorien teknologia (HFST), puhesynteesi.

Alla on yhteenveto organisaation liittyessä FIN-CLARINille ja CLARIN ERICille tarjottavien palvelujen, kielivarojen ja -työkalujen ym. keruun ja valmistamisen käytetty työ- tai rahamääräarvio v. 2006–2015

	Panostuksen laji	Kuvaus	KEUR
1	Tieto- ja palvelukeskus	Kielipankki (Språkbanken i Finland, The Language Bank of Finland) ohjelmistojen käyttöönotto ja aineistojen vastaanottotarkastus ja soveltaminen työkaluja varten, (jäsenen oma panostus)	1430
2	Autentikointi- ja auktorisointipalvelun ylläpito	Language Bank Rights sovelluksen kehitystyö (jäsenen oma panostus)	250
3	Autentikointi- ja auktorisointipalvelun käyttö	HAKA-tunnistuspalvelun käyttö ja aineistojen käyttöilupien myöntäminen Language Bank Rights sovelluksen kautta	130
4	Asiantuntemuksen jakamista koko CLARINille	IPR classification and CLARIN SP Federation Agreement HFST-Helsinki Finite-State Technology	710
5	Olemassa olevat kielivarat, työkalut ja palvelut	- Kielipankin 1990-luvun sanomalehtikorpuset - Multilingual Resource Collection of the UHLCS - HFST (morfologian kompiloitiohjelmat, morfologiset jäsentimet, oikoluku ja tavutus)	2160
6	Tekeillä olevat kielivarat, työkalut ja palvelut	HFST (sanaluokkajäsenin ja syntaktinen jäsenin), FinnWordNet (suomen synonyymi- ja suomi-englanti käännoissanakirjat), FinnTreeBank (määrittelykorpus ja jäsenyspankki)	1610
7	CLARIN-standardien ja käytänteiden omaksuminen ja levittäminen	EU CLARIN kehitystyötä ja META-NORD aineistojen paikannus ja kuvaus sekä seminaareja tekijänoikeudesta ja metadatan keruusta	1190
	Yhteensä		7480

Alla on yhteenveto Helsingin yliopiston arvioidusta vuosittaisesta panostuksesta liittymisen jälkeen.

	Panostuksen laji	Kuvaus	KEUR
1	Tieto- ja palvelukeskus	B-tason CLARIN-keskus yhdessä CSC:n kanssa: Kielipankki (Språkbanken i Finland, The Language Bank of Finland) ohjelmistojen käyttöönotto ja aineistojen soveltaminen ja perusannotointi (jäsenen oma panostus)	210
2	Autentikointi- ja auktorisointipalvelun ylläpito	Jäsenmaksu	25
3	Autentikointi- ja auktorisointipalvelun käyttö	HAKA-tunnistuspalvelu ja Language Bank Rights	20
4	Asiantuntemuksen jakamista koko CLARINille	HFST (Helsinki Finite-State Technology)	90
5	Olemassa olevat kielivarat, työkalut ja palvelut	Tieteen kansallinen termipankki sekä muut HY:n digihum-hankkeiden tulokset	400
6	Tekeillä olevat kielivarat, työkalut ja palvelut	Koordinaatiota varten kehitettävät FIN-CLARINin aineistot ja ohjelmistot sekä muut HY:n projektien uudet tulokset	300
7	CLARIN-standardien ja käytänteiden omaksuminen ja levittäminen	Kielipankin seminaareja kopiosuojasta, metadatan keruusta ja aineistojen annotoinnista	155
	Yhteensä		1200