

## Liite 2 (UEF): Itä-Suomen yliopiston FIN-CLARIN-panostus

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

Itä-Suomen 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).
- 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.
- The University of Eastern Finland offers expertise, e.g., in the following areas: planning, recording and annotation of speech corpora and creation of speech and text corpora for pathological language; multi-lingual resources, in particular for translation and contrastive language studies.

Alla on yhteenveto organisaation liittyessä FIN-CLARINille ja CLARIN ERICille tarjottavien palvelujen, kielivarojen ja -työkalujen ym. keruun ja valmistamisen käytetty työmäärä (henkilötyökuukausina) **2006-2015**.

	Panostuksen laji	Kuvaus	Hlökk 2006-2015
1	Tieto- ja palvelukeskus	-	-
2	Autentikointi- ja auktorisointipalvelun ylläpito	-	-
3	Autentikointi- ja auktorisointipalvelun käyttö	HAKA-tunnistuspalvelun käyttö ja aineistojen käyttöluvien myöntäminen Language Bank Rights sovelluksen kautta	5
4	Asiantuntemuksen jakamista koko CLARINille	-	-
5	Olemassa olevat kielivarat, työkalut ja palvelut	Tässä luetellaan vain vv. 2006-2011 työn alla olleet (kaikki UEF:ssä kehitetyt FIN-CLARIN -kielivarat, ks. FIN-CLARIN -sivusto): - <b>Corpus of Spoken Southwestern Finnish</b> : audio corpus of spoken Finnish; over 300 speakers in 12 communities - <b>Finnish Broadcasting Company Corpus of Subtitles</b> : digital research material of translated subtitles; size: ca. 100 million word tokens. Languages: Finnish, Finland-Swedish and Saami - <b>Joensuu Corpus of Finnish Compounds</b> : about 52000 word tokens in total, WordMill analyzed for structure and frequency features - <b>Joensuu Corpus of Swedish Compounds</b> : computer corpus (list) of Swedish compounds in Göteborgs-Posten, data-base of 24.2 million word tokens, about 3800 compound tokens, with their WordMill variables	16 6 5 5
6	Tekeillä olevat kielivarat, työkalut ja palvelut	(1) Crosslinguistic Studies of Phrasal Idioms: Digitalized Text and Word-list Corpora of English, German, Swedish, Russian, Finnish – a joint UEF project (2) <i>Speech Examination</i> -puhetestin suomenkielinen kontrolliaineisto	12 11
7	CLARIN-standardien ja -käytänteiden omaksuminen ja levittäminen	-	2
	Yhteensä		62 htkk

Alla on yhteenveto organisaation arvioidusta vuosittaisesta panostuksesta 2016 ja eteenpäin.

	Panostuksen laji	Kuvaus	henkilötyökk per vuosi
1	Tieto- ja palvelukeskus	-	-
2	Autentikointi- ja auktorisointipalvelun ylläpito	-	-
3	Autentikointi- ja auktorisointipalvelun käyttö	HAKA-tunnistuspalvelu ja Language Bank Rights käyttö	1
4	Asiantuntemuksen jakamista koko CLARINille	Erityisesti puhekorpuksen käytön neuvonta ja patologisen kielen korpuksen suunnittelu ja laadinta.	12
5	Olemassa olevat kielivarat, työkalut ja palvelut	Ks. vastaava kohta lomakkeessa yllä. Lisäksi: olemassa olevien, erityisesti puhekorpuksen laadullinen kehittäminen	2
6	Tekeillä olevat kielivarat, työkalut ja palvelut	Ks. vastaava kohta lomakkeessa yllä. Lisäksi Suomen itäpuolisten lähialueiden kielikorpuksat ja 300 puhujan n. 30.000 tunnin spontaanin puheen aineiston keruu suomalaisilta eläkeläisiltä Espanjassa; helposti adaptoitavan automaattisen annotoinnin kehittäminen.	12
7	CLARIN-standardien ja -käytänteiden omaksuminen ja levittäminen	Standardien ja käytänteiden integrointi opetukseen	0,5
	Yhteensä		27,5 htkk