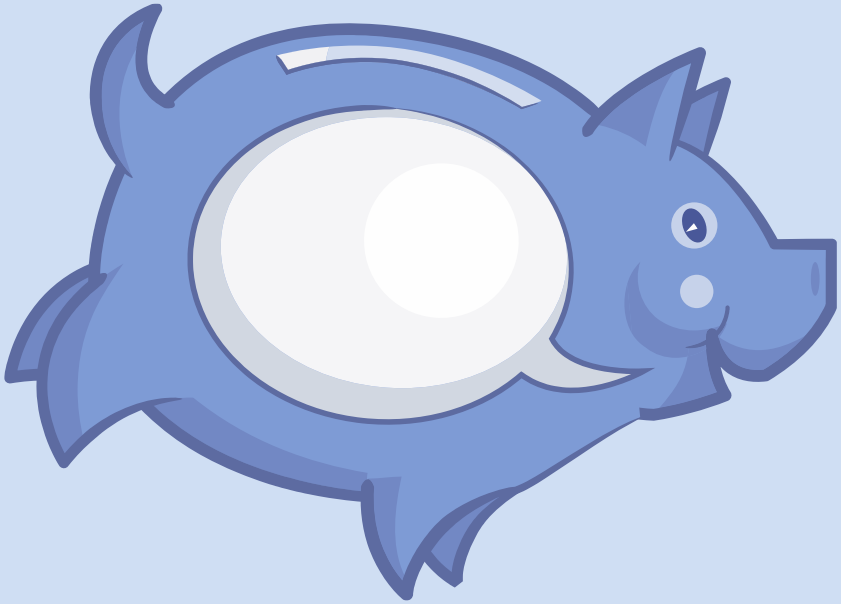


KIELIPANKKI
The Language Bank of Finland

Kielipankki pähkinäkuoreessa
The Language Bank of Finland in a nutshell

CSC User Support Coffee 30.11.2022

Martin Matthiesen, CSC

FIN-CLARIN

At a glance

- Storage and sharing of language resources
- Tools for researchers and teaching
- 14+ billion tokens of text, 1000 hours of audio/video data.
- Licensing
- Access control (Open data done right)

www.kielipankki.fi

Language Resources

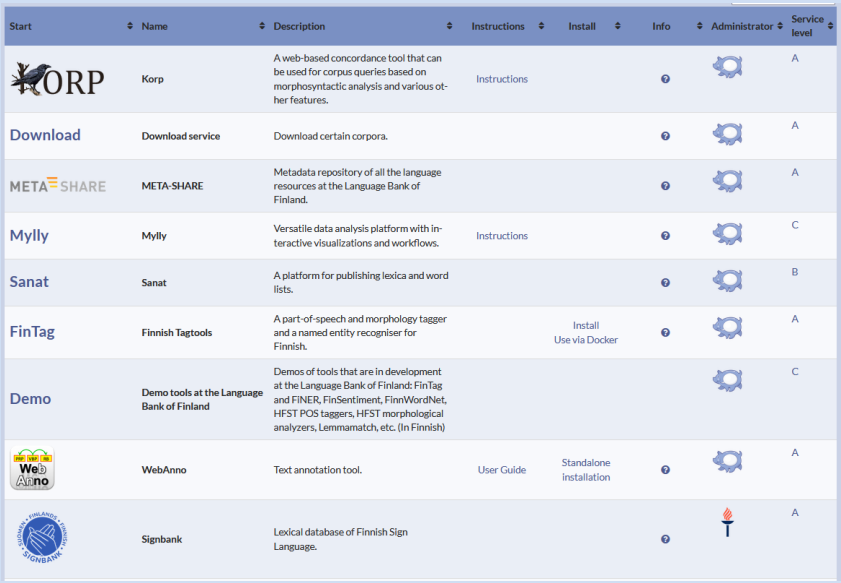
www.kielipankki.fi/corpora

Abbreviation	Name and metadata	License	Apply	Location	Service level	Help	Cite
acquis-ftb3	The Finnish Sub-corpus of the JRC-Acquis Multilingual Parallel Corpus	PUB		Korp	B		
acquis-ftb3-dl	Finnish Sub-corpus of the JRC-Acquis Multilingual Parallel Corpus, Downloadable Version	PUB		Download	B		
agricola-v1-1-korp	The Morpho-Syntactic Database of Mikael Agricola's Works version 1.1, Korp	PUB		Korp	B		
ai2d-rst-v1-1	AI2D-RST: A multimodal corpus of 1000 primary school science diagrams version 1.1	PUB		Download	B		
aku-egg-dl	Speech and EGG (Electroglottography) Simultaneous Recordings, downloadable version	ACA		Download	B		
amph	amph-Corpus	ACA		Download	B		
ArkiSyn-korp	ArkiSyn Database of Finnish Conversational Discourse, Helsinki Korp Version	PUB		Korp	B		
AVOID	Corpus of Age-related Voice Disguise (AVOID)	RES		Download	B		
BeserCorp	The Corpus of Beserman Udmurt	PUB		Korp	B		
ccmh-src	Corpus Cyrillo-Methodianum Helsingiense: Corpus of Old Church Slavonic Texts, source	PUB		Download	B		
ceal-dl	The Downloadable Version of Classics of English and American Literature in Finnish	RES		Download	A		
ceal-o	Classics of English and American Literature in Finnish, Sentences and Paragraphs in the Original Order	RES		Korp	A		
ceal-par-korp	Classics of English and American Literature as translated by Kersti Juva, English-Finnish parallel corpus, Korp	RES		Korp	A		
ceal-par-s-dl	The Downloadable Version of Classics of English and American Literature as translated by Kersti Juva, English-Finnish parallel corpus, scrambled	ACA		Download			
ceal-par-s-korp	Classics of English and American Literature as translated by Kersti Juva, English-Finnish parallel corpus, scrambled, Korp	ACA		Korp	A		
ceal-s	Classics of English and American Literature in Finnish, Scrambled Paragraphs	ACA		Korp	A		
cfinsl-conv-dl	Corpus of Finnish Sign Language: conversations, Download version	RES		Download	B		

metashare.csc.fi

Tools

www.kielipankki.fi/tools



Start	Name	Description	Instructions	Install	Info	Administrator	Service level
KORP	Korp	A web-based concordance tool that can be used for corpus queries based on morphosyntactic analysis and various other features.	Instructions				A
Download	Download service	Download certain corpora.					A
META-SHARE	META-SHARE	Metadata repository of all the language resources at the Language Bank of Finland.					A
Mylly	Mylly	Versatile data analysis platform with interactive visualizations and workflows.	Instructions				C
Sanat	Sanat	A platform for publishing lexica and word lists.					B
FinTag	Finnish Tagtools	A part-of-speech and morphology tagger and a named entity recogniser for Finnish.		Install Use via Docker			A
Demo	Demo tools at the Language Bank of Finland	Demos of tools that are in development at the Language Bank of Finland: FinTag and FINER, FinSentiment, FinnWordNet, HFST POS taggers, HFST morphological analyzers, Lemmamatch, etc. (In Finnish)					C
WebAnno	WebAnno	Text annotation tool.	User Guide	Standalone installation			A
SIGNBANK	Signbank	Lexical database of Finnish Sign Language.					A

https://docs.csc.fi/apps/

Language Research and Other Digital Humanities and Social Sciences

- eBay's tsv-tools Utilities for manipulating large tabular data files
- finnish-parse Dependency Parser for Finnish
- Finnish Tagtools Finnish Tagtools
- HFST Helsinki Finite-State Transducer Technology
- HFST-fi Helsinki Finite-State Technology for Finnish
- HFST-sv Helsinki Finite-State Technology for Swedish
- Kaldi Kaldi Speech Recognition Toolkit
- UDPipe UDPipe Kielipankki version

Data on HPC

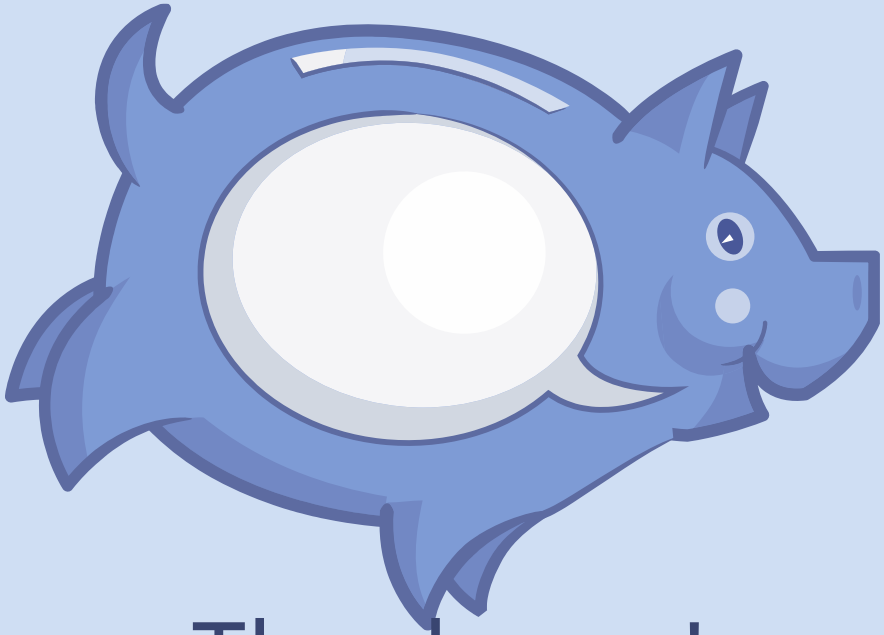
```
[matthies@puhti-login15 kielipankki]$ cd /appl/data/kielipankki
[matthies@puhti-login15 kielipankki]$ ls
acquis                Kalevala        sfnet
admin                 kksxml          SNC1
amph                  kra             suomi24-2001-2017-vrt-v1-2
bert_models           lehdet-vrt-v2   suomi24-2018-2020-vrt-beta
dspcon2013-2016-dl    mrc-uhlcs       susanne
Eduskunta             nlfcl-fi-vrt    wikipedia-fi-2017-src
FNC1                  nlfcl-sv-vrt    wordnet
hcs-a-v2-dl           ota             words
hcs-na-v2             oulu            ylilauta
[matthies@puhti-login15 kielipankki]$
```

CLARIN ERIC

- 22 Members
- 2 Observers
- 70 Centres

- Metadata Search: https://vlo.clarin.eu
- Metadata Standards: CMDI
- Federated Content Search
- Persistent Identifiers: Handles, URNs
- Federated Login
- License Framework (PUB, ACA, RES)

KIELIPANKKI
The Language Bank of Finland



Thank you!

www.kielipankki.fi