

**KIELIPANKKI**  
The Language Bank of Finland

Welcome to  
the Language Bank of Finland!

Krister Lindén, FIN-CLARIN  
[krister.linden@helsinki.fi](mailto:krister.linden@helsinki.fi)

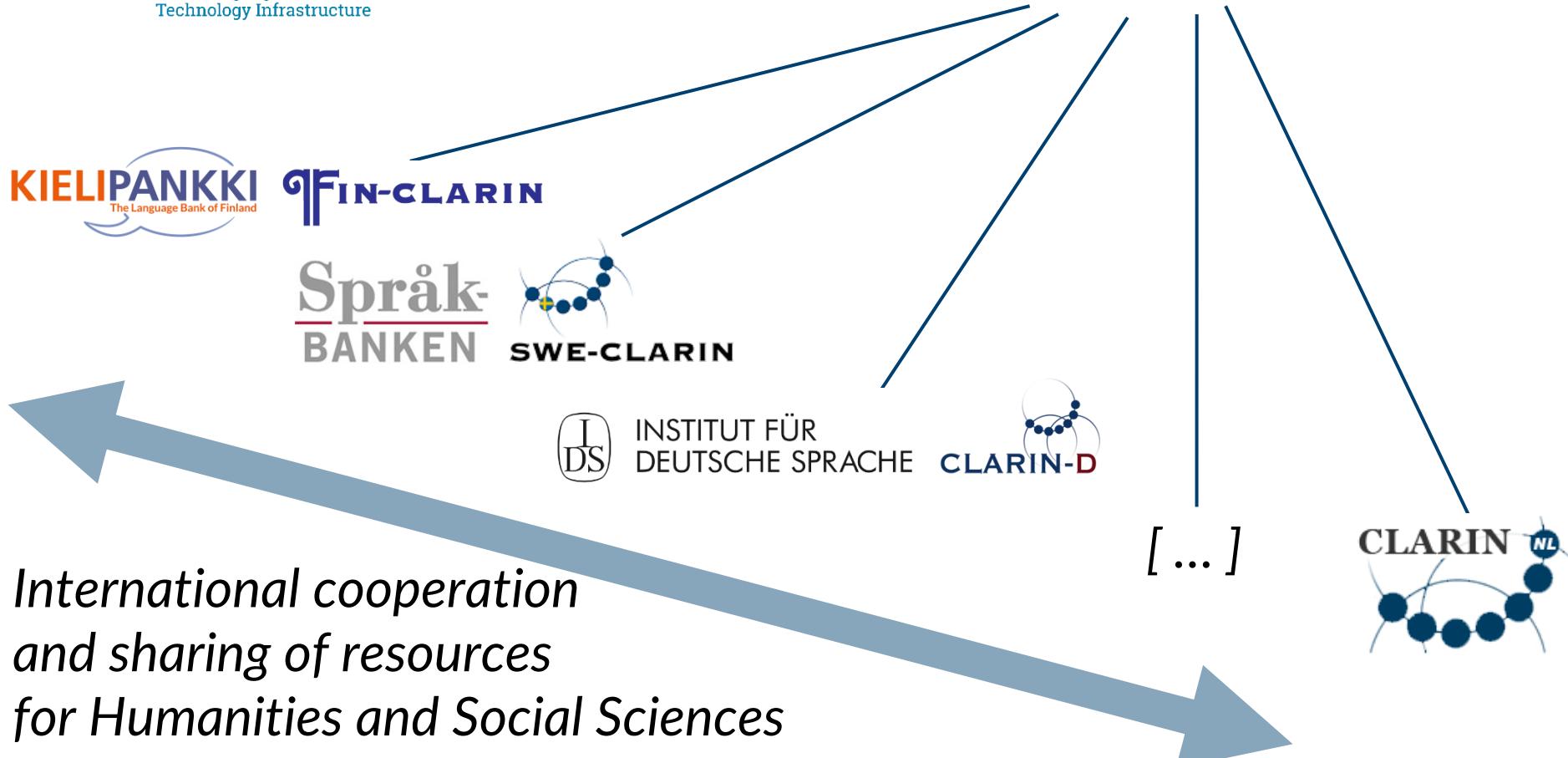


Common Language Resources and  
Technology Infrastructure

# CLARIN ERIC

European Research Infrastructure Consortium  
founded on February 29, 2012

<https://www.clarin.eu> / NL



Member countries (23):

Austria  
Belgium  
Bulgaria  
Croatia  
Cyprus  
Czech Republic  
Denmark  
Estonia  
Finland  
Germany  
Greece  
Hungary  
Iceland  
Italy  
Latvia  
Lithuania  
The Netherlands  
Norway  
Poland  
Portugal  
Slovenia  
Spain  
Sweden

Observers (3):

South Africa  
Switzerland  
United Kingdom

Third party (1):

Carnegie Mellon  
University (USA)

Updated: 29.8.2023



# CLARIN centres

## Member countries (23):

Austria  
Belgium  
Bulgaria  
Croatia  
Cyprus  
Czech Republic

Denmark

Estonia

Finland

Germany

Greece

Hungary

Iceland

Italy

Latvia

Lithuania

**The Netherlands**

Norway

Poland

Portugal

Slovenia

Spain

Sweden

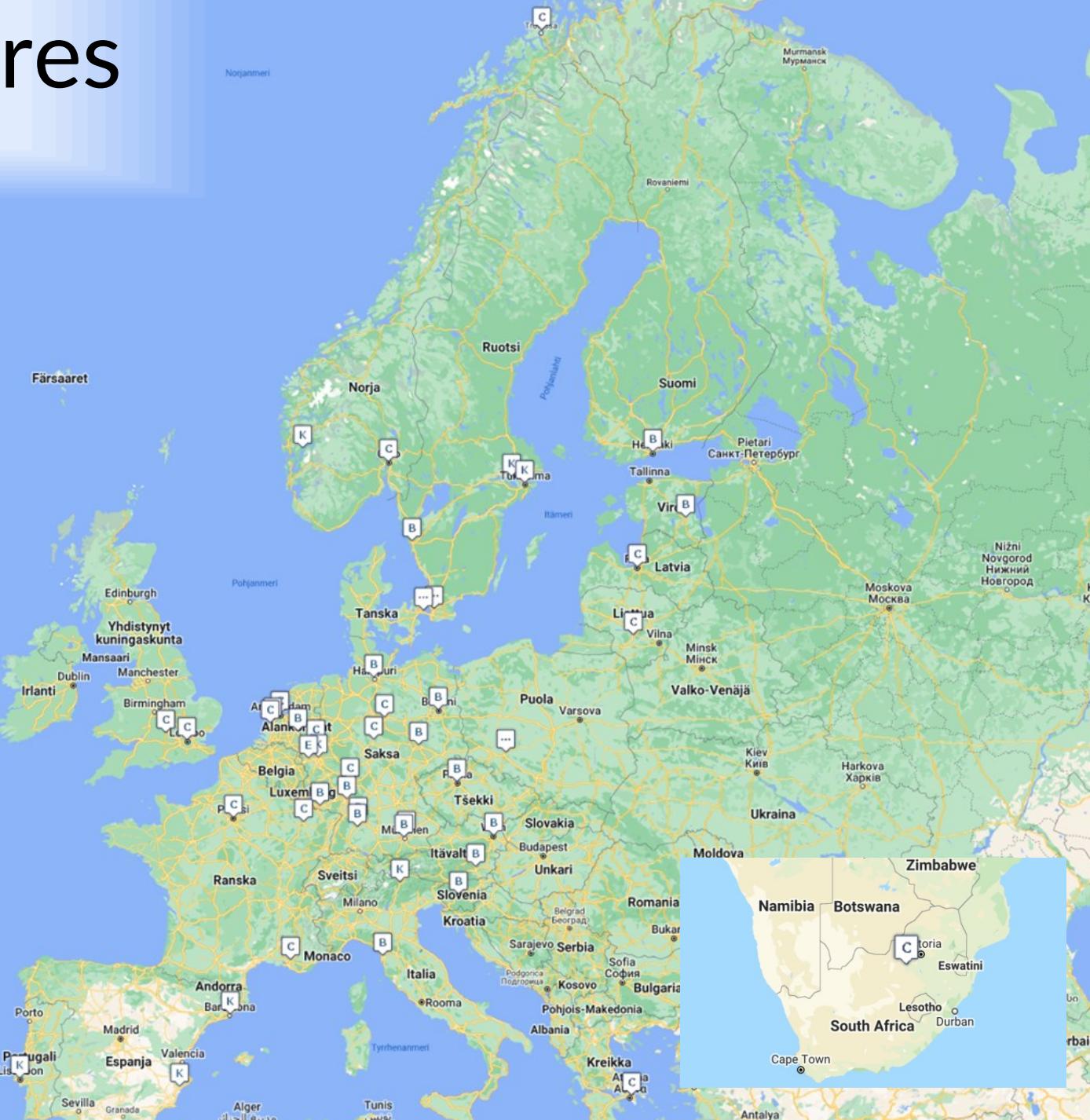
## Observers (3):

South Africa  
Switzerland  
United Kingdom

## Third party (1):

CMU (USA)

Updated: 29.8.2023





## Member countries (23):

Austria  
Belgium  
Bulgaria  
Croatia  
Cyprus  
Czech Republic

Denmark

Estonia

Finland

Germany

Greece

Hungary

Iceland

Italy

Latvia

Lithuania

The Netherlands

Norway

Poland

Portugal

Slovenia

Spain

Sweden

## Observers (3):

South Africa  
Switzerland  
United Kingdom

## Third party (1):

CMU (USA)

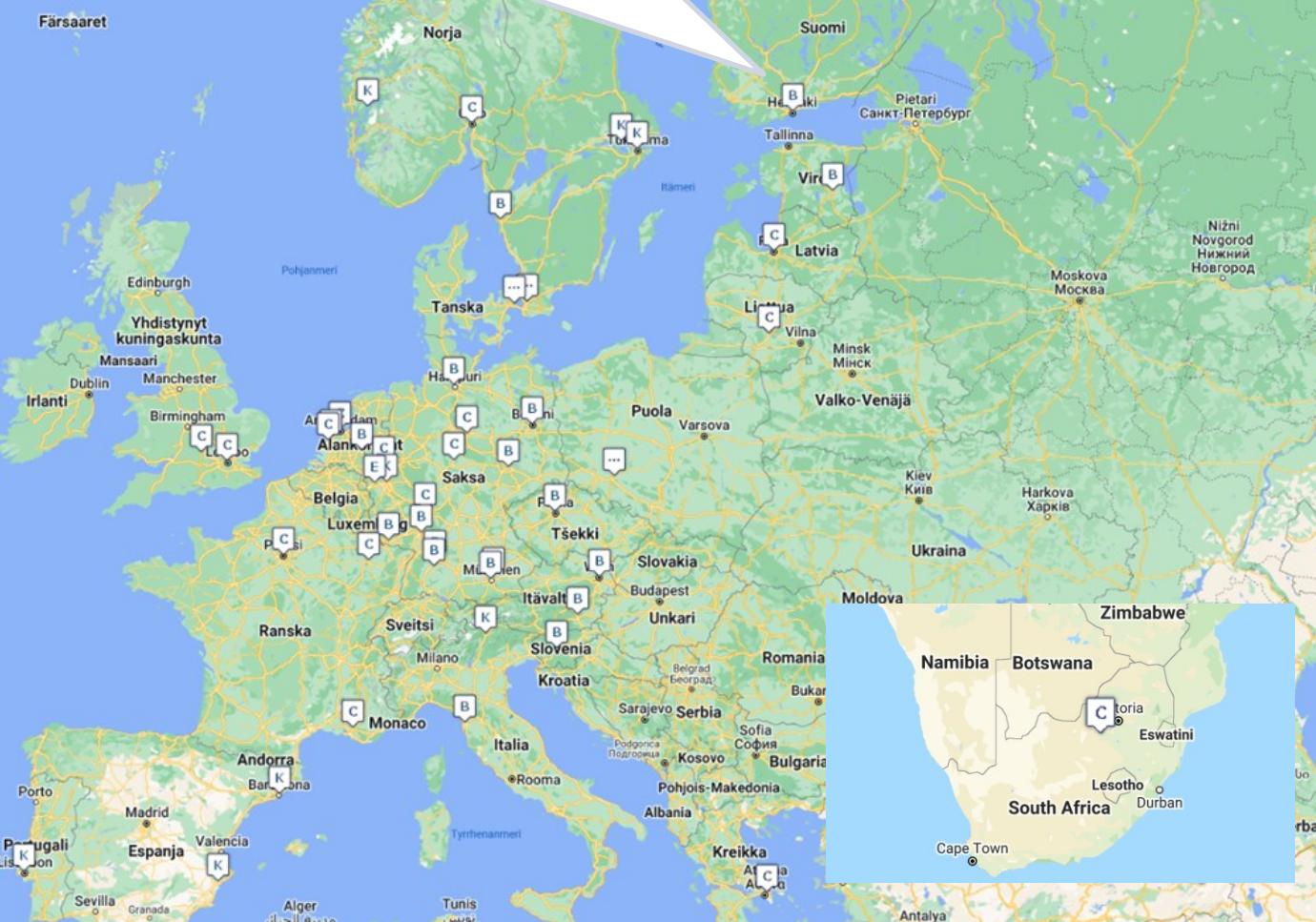
Updated: 29.8.2023

# CLARIN centres



Kielipankki –  
The Language Bank of Finland

B-centre  
FIN-CLARIN



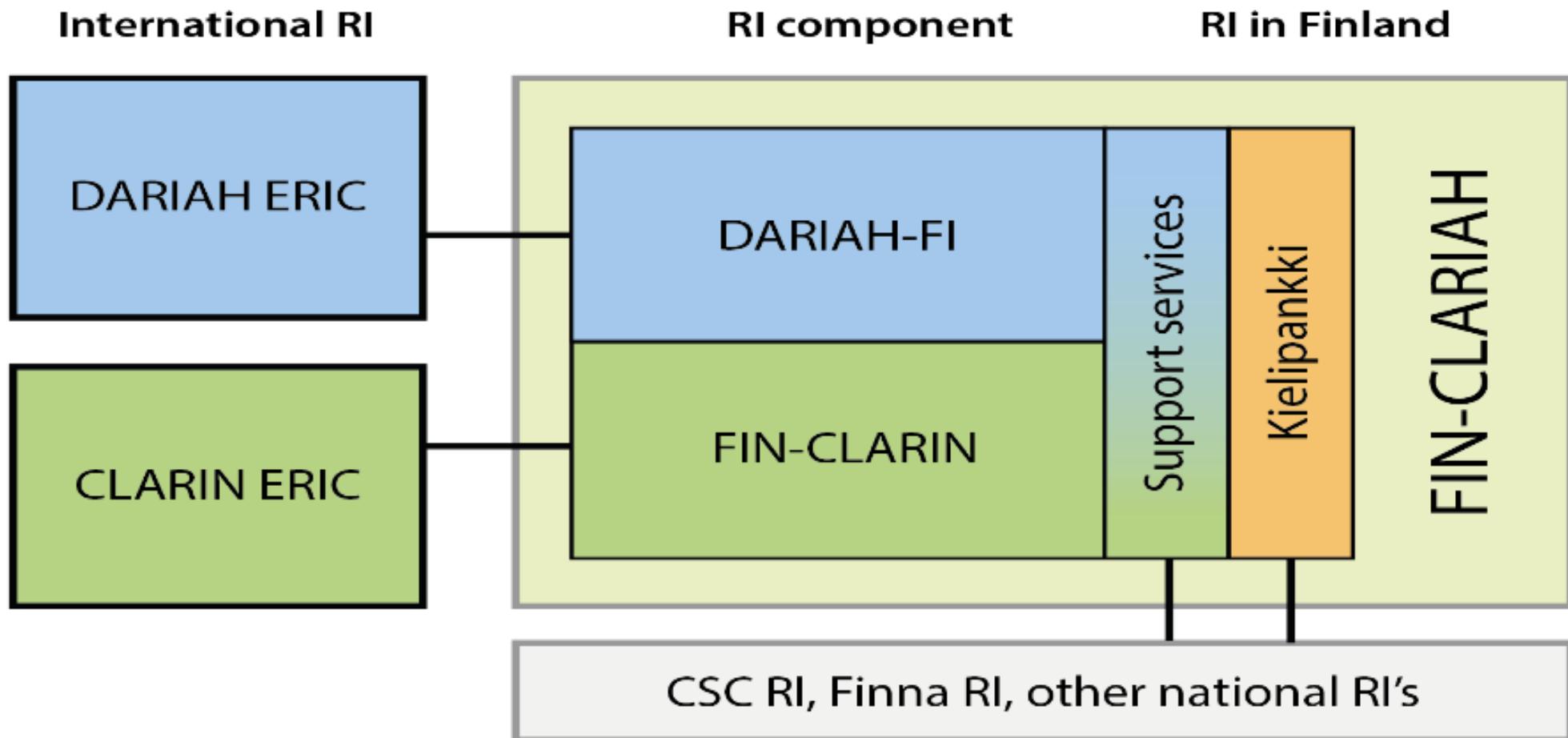


OFIN-CLARIN

**KIELIPANKKI**

The Language Bank of Finland

# FIN-CLARIAH



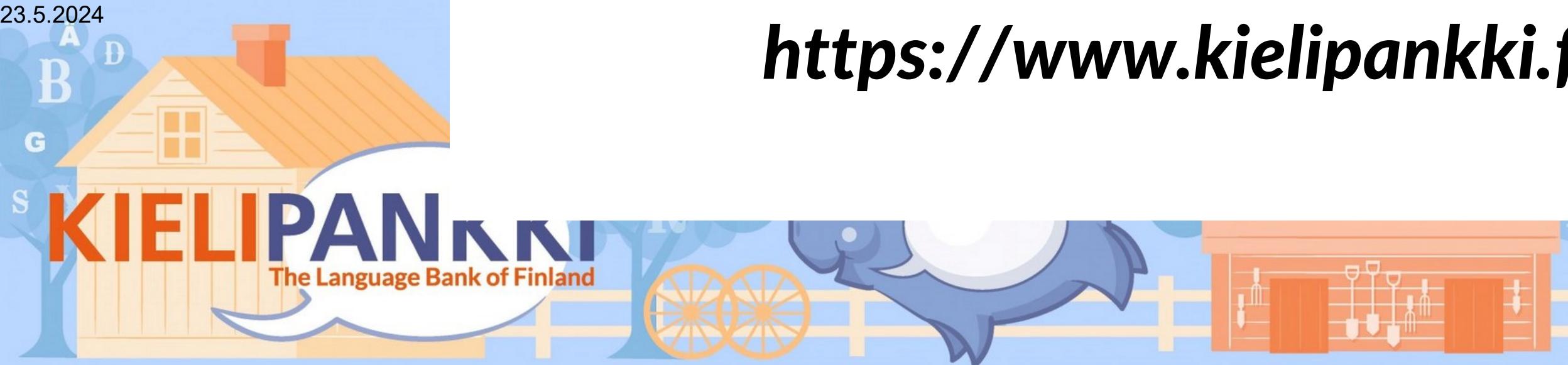
# Kielipankki Core Teams

## University of Helsinki:

- Erik Axelson
- Ute Dieckmann
- Wilhelmina Dyster
- Tommi Jauhainen
- Mietta Lennes
- **Krister Lindén, Research Director**
- Jyrki Niemi
- Jussi Piitulainen
- Jack Rueter

## CSC - IT Center for Science:

- Sam Hardwick
- Anni Järvenpää
- Martin Matthiesen
- **Katri Tegel, *Development Manager***



LANGUAGE BANK ACCESS CORPORA TOOLS ORGANIZATION SUPPORT

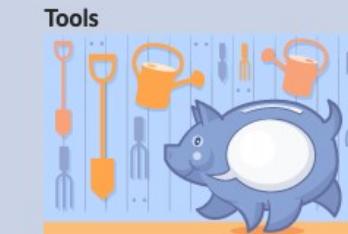
SUOMEKSI PÅ SVENSKA



Apply for rights to use our language resources.



Browse our corpora.



Try our tools.

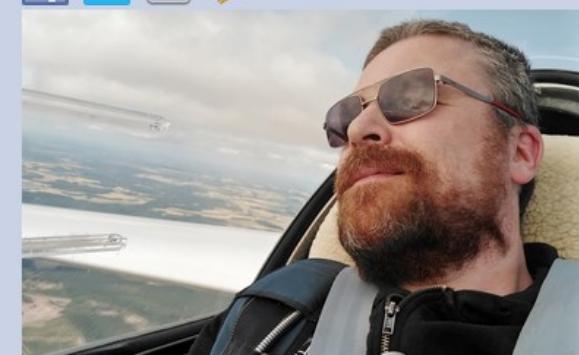


Who are the Language Bank?



Help and instructions.

Search the Language Bank Portal:



Researcher of the Month: Filip Ginter

News

# List of tools available via the Language Bank

[www.kielipankki.fi/tools](http://www.kielipankki.fi/tools)

<input type="text" value="Etsi:"/>						
Start	Name	Description	Instructions	Install	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.	<a href="#">Instructions</a>			A
<b>Download</b>	Download service	Download certain corpora.				A
 META-SHARE	META-SHARE	Metadata repository of all the language resources at the Language Bank of Finland.				A
<b>Myilly</b>	Myilly	Versatile data analysis platform with interactive visualizations and workflows.	<a href="#">Instructions</a>			C
<b>Sanat</b>	Sanat	A platform for publishing lexica and word lists.				B
<b>FinTag</b>	Finnish Tagtools	A part-of-speech and morphology tagger and a named entity recogniser for Finnish.		<a href="#">Install</a>		A
<b>Demo</b>	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	Text annotation tool.	<a href="#">User Guide</a>	<a href="#">Standalone installation</a>		A



**Valitse korpukset:** 0 / 1246 korusta valittuina — 0 / 37,30G sanetta

Yksinkertainen

Laajennettu

Edistynyt

Vertailu

Hae

järjestyksessä ja myös  alkuosa  keskiosa  loppuosa ja  samaista pien- ja suuraakkoset

Konkordansi: osumia sivulla: 25

järjestää korpuksen sisällä: järjestämätön

Tilastoja: laskentaperuste: sana

Näytä tilastot

# Korp: queries within resources

## korp.csc.fi

The screenshot shows the Korp search interface. At the top, there's a header with a crow icon, search fields for 'Koru' and 'Korppi', and a date range from '1600' to '2020'. Below the header is a timeline chart showing the distribution of texts over time. To the right of the timeline are two buttons: 'Valitse kaikki' (Select all) with a checkmark and 'Tyhjennä' (Clear) with a cross.

The main content area displays search results. A red circle highlights the first result under 'Tietokonevälitteistä viestintää': 'Suomi24 2001–2020 (20)'. This result is checked. Other results in this category include 'Suomi24 2016H2 (10)' and 'Suomi24 2001–2014 (näyte)'. Below this category, there are more categories: 'Ylilauta' (unchecked), 'HS.fi [ACA]' (locked, unchecked), 'Akateemisia tekstejä (23)', 'Historiallisia aineistoja (115)', 'Oppijansuomen aineistoja (4)', and 'Kirjallisuusaineistoja (26)'.

At the bottom right, there are checkboxes for 'Näytä tilastot' (Show statistics) and 'Näytä sanakuva' (Show document image). The bottom left shows the page number '1 / 1' and a 'Näytä kaikki' button.

# Korp: queries within resources

## korp.csc.fi

The screenshot shows the Korp search interface. At the top, there's a navigation bar with a crow icon, a search input field containing 'ja', and buttons for 'Edistynyt' and 'Vert'. Below the search bar is a histogram showing the distribution of documents over time from 1600 to 2000. A red oval highlights the search results area.

**Search results for 'ja':**

- 20 / 1246 korposta valittuna — 4,58G / 37,30G sanetta**
- 20 korposta, joissa:**
- 4 582 558 555 sanetta**
- 391 965 356 virkettä**

**Result details:**

- Tietokonevälitteistä viestintää (32)**
  - Suomi24 2001–2020 (20)**
  - Suomi24 2016H2 (10)**
  - Suomi24 2001–2014 (näyte)**
  - Ylilauta**
  - HS.fi [ACA]**
- Akateemisia tekstejä (23)**
- Historiallisia aineistoja (115)**
- Oppijansuomen aineistoja (4)**
- Kirjallisuusaineistoja (26)**

**Bottom right controls:**

- Näytä tilastot    Näytä sanakuva

Lahjoita puhetta -kampanja on p X Korp X Helsinki Region Infoshare - Suo X +

https://www.kielipankki.fi/korp/#?cqp=%5B%5D&corpus=s24\_2001,s24\_2002,s24\_2003,s24\_2004,s24\_2...

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit Kirjaudu Suomi | Svenska | English 7 Viittaa Korpiin VALIKKO

KORP v9  20 / 1246 korputta valittuna — 4,58G / 37,30G sanetta 

Yksinkertainen Laajennettu Edistynyt Vertailu

Suodatin:  lisää aihealueen ylin taso ja  lisää aihealue Helsinki Region Infoshare Hae ▾

Janestä keskösä ja myös  alkiosa  keskiosa  loppiosa ja  samaista pien- ja suuraakkoset

Konkordansi: osumia sivulla: 25 ▾ järjestää korpuksen sisällä: järjestämätön ▾ Tilastoja: laskentaperuste: sana ▾  Näytä tilastot  Näytä sanakuva

Konkordansi Tilastoja Sanakuva

Tuloksia: 1

« < **1** > » Siirry sivulle  / 1 Näytä konteksti

SUOMI24 2001–2017: 2011

**Helsinki Region Infoshare**

« < **1** > » Siirry sivulle  / 1 Lataa konkordansi ▾

JSON

KORPUS  
Suomi24 2001–2017: 2011  
Osa aineistoa: Suomi24 2001–2020  
Kuvailutiedot  
Viittaa aineistoon  
Linkki aineistoon Korpissa  
Pysyvä tunniste:  
[urn:nbn:fi:lb-2020021803](#)  
Lisensi: CC BY-NC (CLARIN PUB)

TEKSTIN PIIRTEET  
otsikko: Helsinki Region Infoshare  
päiväys: 2011-03-18  
kellonaika: 17:23:13  
Kirjoittajan nimimerkki: Itämeren tytär  
Kirjautunut käyttäjä: ei  
rekisteröity nimimerkki: ei  
aihealue: Tiede ja teknologia >  
Tietotekniikka > Käyttöjärjestelmät  
> Soda  
aihealue vain aikuisille: ei  
viestin tyyppi: keskusteluketjun  
aloitus  
viesti on täysin tyhjä: ei  
tunniste: 9735406:0  
keskusteluketjun tunniste: 9735406

Items: 18 Unread: 1 All folders are up to date. Connected to: Microsoft Exchange

Lahjoita puhetta -kampanja on p X Korp X Helsinki Region Infoshare - - Suo X +

https://www.kielipankki.fi/korp/#?cqp=%5B%5D&corpus=s24\_2001,s24\_2002,s24\_2003,s24\_2004,s24\_2...

JSON

TEKSTIN PIRTEET

otsikko: Helsinki Region Infoshare  
päiväys: 2011-03-18  
kellonaika: 17:23:13  
kirjoittajan nimimerkki: Itämeren tytär  
kirjautunut käyttäjä: ei  
rekisteröity nimimerkki: ei  
aihealue: Tiede ja teknologia >  
Tietotekniikka > Käyttöjärjestelmät > Soda  
aihealue vain aikuisille: ei  
viestin typpi: keskusteluketjun aloitus  
viesti on täysin tyhjä: ei  
tunniste: 9735406:0  
keskusteluketjun tunniste: 9735406  
keskusteluketjun alkuaikaleima: 2011-03-18 17:23:13  
kommentin tunniste: 0 [ketjun aloitusviesti]  
kommentin vanhemman tunniste: 0 [ketjun aloitusviesti]  
kommentin vanhemman aikaleima: [tyhjä]  
lainatun kommentin tunniste: 0 [ei lainausta]  
tiedoston nimi: s24\_2011\_05.vrt  
kappaleen typpi: otsikko  
virkkeen polaarisuus: neutraali

SANAN PIRTEET

perusmuoto: Helsinki  
perusmuoto (yhdyssanarajat): Helsinki  
sanaluokka: substantiivi  
morphologinen analyysi:  
SUBCAT\_Prop|NUM\_Sgl  
CASE\_Nom|CASECHANGE\_Up  
dependenssisuhde: monisanainen nimi  
saneeseen liittyvät tyhjämerkit: \_  
Näytä dependenssippu  
Näytä alkuperäinen keskusteluketju  
Näytä alkuperäinen viesti

Notes

Slide 12 of 28 English (United Kingdom) Accessibility: Investigate

Lahjoita puhetta -kampanja on p | Korp | Helsinki Region Infoshare - - Suo +

https://keskustelu.suomi24.fi/t/9735406/helsinki-region-infoshare

A Etsi Kirjaudu Rekisteröidy

SUOMI 24 Valikko Etsi Kirjaudu Rekisteröidy

Hyödynnä alennukset tästä!

↓ Aihealueet

... / Käyttöjärjestelmät / Sodat

## Helsinki Region Infoshare

Itämeren tytär  
2011-03-18 17:23:13

Ilmoita

Helsinki on yksi avoimen datan edelläkävijöistä:  
[http://www.tietokone.fi/uutiset/paakaupunkiseutu\\_avaa\\_tietoja\\_reilusti](http://www.tietokone.fi/uutiset/paakaupunkiseutu_avaa_tietoja_reilusti)

"Avoimen datan pohjalta voi rakentaa esimerkiksi uusia kaupunkilaisia hyödyttäviä verkkosovelluksia. Toisaalta datan avaamisella halutaan edistää myös tutkimusta ja kehitystoimintaa sekä lisätä kansalaisten aktiivisuutta ja ymmärrystä omasta asuinalueestaan."

<http://www.hri.fi/fi/>

0 <50

↑ Äänestää Kommentoi Jaa

SUOMI 24 Osallistu keskusteluun

Venäjä aikoo yksipuolisesti mu... 151  
Venäjä aikoo muuttaa rajalinjauskia S...

Farmi Suomi: Sami Jauhojärvi ... 4  
Oi, onpa tunteikas hetki... Farmi on ky...

Asuntojen reaalihinnat alimmil... 67  
Sääliksi käy niitä jotka pari vuotta sitt...

Marin on ehkä maailman kaun... 345  
Marinin julkaisu on saanut yli 68 500 ...

Ellen Jokikunnas paljastaa - Ta... 11  
Oletko seurannut Unelmia Italiassa -s...

SUOMI24 VIIHDE

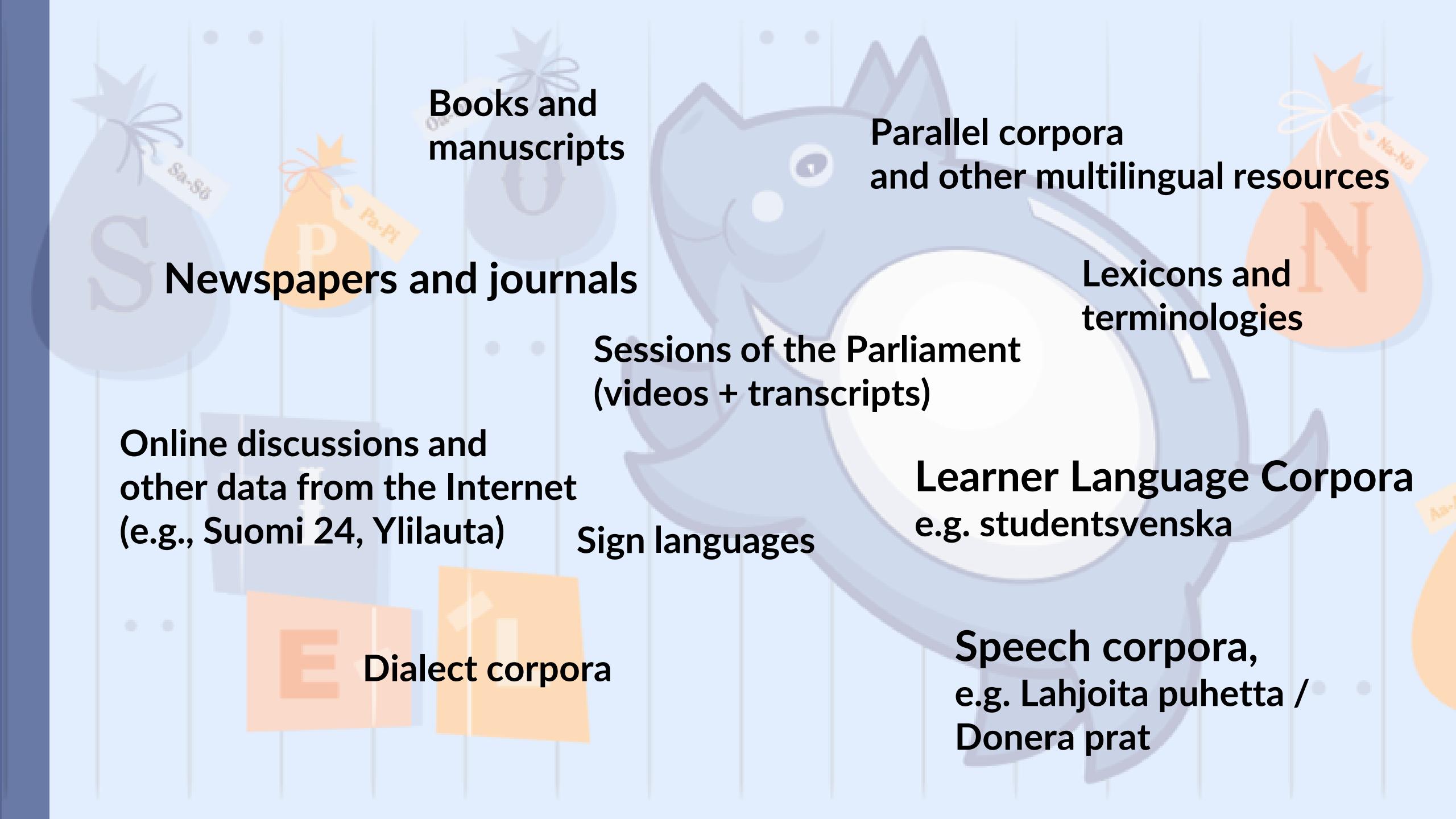
Notes

Slide 11 of 28 English (United Kingdom) Accessibility: Investigate

51%

# Research using newspapers (~19 %)

- **Political history**  
(Mats Fridlund)
- **Conceptual history**  
(Jani Marjanen, Risto Turunen)
- **Language modelling**  
(Filip Ginter, Simon Hengchen)
- **Terminology**  
(Antti Kanner, Anita Nuopponen)
- **Cognitive science**  
(Tatu Huovilainen)
- **Cognitive linguistics**  
(Szuszi Máthé)
- **Linguistics**  
(Jussi Ylikoski, Juho Härmä)
- **Language change**  
(Marko Pantermöller, Jutta Salminen, Kirsi-Maria Nummila)
- **Special language**  
(Bejamin Schweizer, Paul-Thor Holmberg)
- **Language learning**  
(Ilmari Ivaska)



**Online discussions and  
other data from the Internet  
(e.g., Suomi 24, Ylilauta)**

**Newspapers and journals**

**Books and  
manuscripts**

**Dialect corpora**

**Sign languages**

**Sessions of the Parliament  
(videos + transcripts)**

**Learner Language Corpora  
e.g. studentsvenska**

**Speech corpora,  
e.g. Lahjoita puhetta /  
Donera prat**

**Parallel corpora  
and other multilingual resources**

**Lexicons and  
terminologies**

# List of corpora at the Language Bank of Finland

## www.kielipankki.fi/corpora

### Cite

Corpora-specific reference instructions and citations in other papers are available by clicking the quote “” link.

If you are looking for a corpus not listed here, please have a look in [META-SHARE](#) or CLARIN Virtual Language Observatory (VLO).

Please find an overview of all our resources sorted by resource families in the [resource families of FIN-CLARIN](#).

Shortname	Name and metadata	License	Location	Cite	Resource group and help	Apply	Publication year	Support level
acquis-ftb3	The Finnish Sub-corpus of the JRC-Acquis Multilingual Parallel Corpus 		Korp	<a href="#">“”</a>	jrc-acquis			B
acquis-ftb3-dl	Finnish Sub-corpus of the JRC-Acquis Multilingual Parallel Corpus, Downloadable Version 		Download	<a href="#">“”</a>	jrc-acquis			B
agricola-v1-1-korp	The Morpho-Syntactic Database of Mikael Agricola's Works version 1.1, Korp 		Korp	<a href="#">“”</a>	agricola		2020	B
ai2d-rst-v1-1	AI2D-RST: A multimodal corpus of 1000 primary school science diagrams version 1.1 		Download	<a href="#">“”</a>	ai2d-rst		2020	B
aku-egg-dl	Speech and EGG (Electroglossography) Simultaneous Recordings, downloadable version 		Download	<a href="#">“”</a>	aku-egg		2014	B
amph	amph-Corpus 		Download Puhti	<a href="#">“”</a>	amph			B
ArkiSyn-korp	ArkiSyn Database of Finnish Conversational Discourse, Helsinki Korp Version 		Korp	<a href="#">“”</a>	arkisyn		2017	B
AVOID	Corpus of Age-related Voice Disguise (AVOID)  		Download	<a href="#">“”</a>	avoid		2019	B
BeserCorp	The Corpus of Beserman Udmurt 		Korp	<a href="#">“”</a>	besercorp		2016	B
ccmh-src	Corpus Cyrillo-Methodianum Helsingiense: Corpus of Old Church Slavonic Texts, source 		Download	<a href="#">“”</a>	ccmh		2021	B



KIELIPANKKI KÄYTTÄJÄKSI AINEISTOT TYÖKALUT ORGANISAATIO TUKI

IN ENGLISH PÅ SVENSKA

Lahjoita puhetta -kampanja on päättynyt.

Kiitos kaikille lahjoittajille!

Lisätietoja vuosina 2020-2024 toteutetuista suomen ja suomenruotsin Lahjoita puhetta -kampanjoista ja niissä kerätyin puheaineiston käytöstä



# Tools: Aalto-ASR

## Automatic speech recognition and alignment

A *demo tool*  
*is also*  
*available*



**KIELIPANKKI**  
The Language Bank of Finland

LANGUAGE BANK ACCESS CORPORA TOOLS ORGANIZATION SUPPORT SUOMEKSI PÅ SVENSKA

## Aalto-ASR - Aalto University Automatic Speech Recognition System

[Suomeksi](#)

Aalto-ASR includes two main functionalities: automatic speech recognition and automatic forced alignment. The speech recognizer allows you to obtain a transcript of an audio file that contains Finnish speech. The result of the recognition process is returned either as an annotation file (such as Praat TextGrid) or as a plain text file. In case a transcript of the audio file already exists, the words included in the text can be automatically aligned with the corresponding parts of the audio signal. The automatic forced aligner also works in several other languages in addition to Finnish.

[Aalto University Automatic Speech Recognition System, version 2.1](#)  
[Metadata, license and citation instructions](#)  
[Instructions \(in Finnish only; English instructions forthcoming!\)](#)

### Recent updates

The new, upgraded version 2.1 was installed in Kielipankki in September 2021. The tools can be used in the Puhti environment. If required, Aalto-ASR can now also be installed on other environments as a Docker container.

Some features of version 1 were previously available via the [Mylly](#) service as well. However, the old Aalto-ASR version was removed in 2020 and the new version has not yet been plugged in to Mylly, but this may happen at a later stage.

Persistent identifier of this resource group page: <http://urn.fi/urn:nbn:fi:lb-2021082325>

Search the Language Bank Portal:



Researcher of the Month: Rosa González Hautamäki

### News

- New resource (beta): Parallel Corpus of the book "Four Battles", written by L. Uspenskij; Korp version (5.6.2023)
- New resource: Finnish Dark Web Marketplace Corpus (findarc) (30.5.2023)
- New resource: The Movie Corpus (Mark

# Tools: tekstiks.ee

Automatic  
speech  
transcription  
via an online  
interface

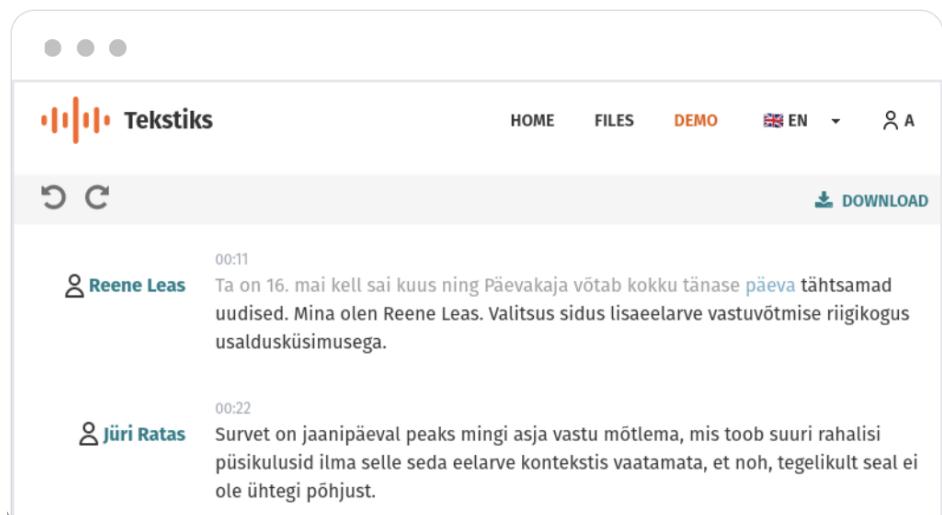
## Puheentunnistus: puheesta tekstiksi

Automaattisen puheentunnistuksen palvelu viroon ja suomen kielille sekä käyttöliittymä transkription muokkaamiseen.

**ALOITA** **KATSO DEMO**

### Kuinka tämä toimii?

Tekstiks.eeon Tallinnan teknillisen yliopiston (TTÜ) Kieliteknologian laboratorion avoin puheentunnistusjärjestelmä. Käyttöliittymään syksyllä 2022 lisätty suomen puheentunnistuksen tuki on kehitetty yhdessä Kielparkin ja Aalto yliopiston kanssa CLARIn-yhteistyönä. Järjestelmä on täysin automatisoitu ja voi käsitellä useita tiedostoja yhtäaikaisesti. Tunnistuksessa voi esiintyä jonoja kiireisinä aikoina, mutta keskimääräinen prosessointiaika on noin puolet nauhoitteen kestosta. **Tulossa: tiedoston tunnistuksen etenemisen seuranta ja odotettu valmistumisaika.**



### 1. Lataa suomenkielinen nauhoitus serverille.

Yleisimmät ääni- ja videoformaatit ovat tuettu.

### 2. Odota transkription valmistumista.

Koneoppimista hyödyntävä järjestelmä etsii äänitiedostosta puhetta ja yrittää tuottaa sitä vastaavan tekstin.

### 3. Korja puheentunnistukseen jäneet virheet käsin.

Tekstin editointi on interaktiivista. Äänen toistaminen ja transkriptio on kohdistettu toisiinsa. Kulloinkin toistettu sana korostetaan korjaamisen helpottamiseksi.

### 4. Lataa tiedosto.

Tällä hetkellä tuki on docx-tiedostoiille.

# Citation instructions

ArkiSyn-korp	ArkiSyn Database of Finnish Conversational Discourse, Helsinki Korp Version 	Korp	”	arkisyn	2017
AVOID	Corpus of Age-related Voice Disguise (AVOID) 	Download	”	avoid	 2019
BeserCorp	The Corpus of Beserman Udmurt 	Korp	”	besercorp	2016 [suomeksi] [in English]

## Reference instructions: AVOID

Please cite the language resource as follows:

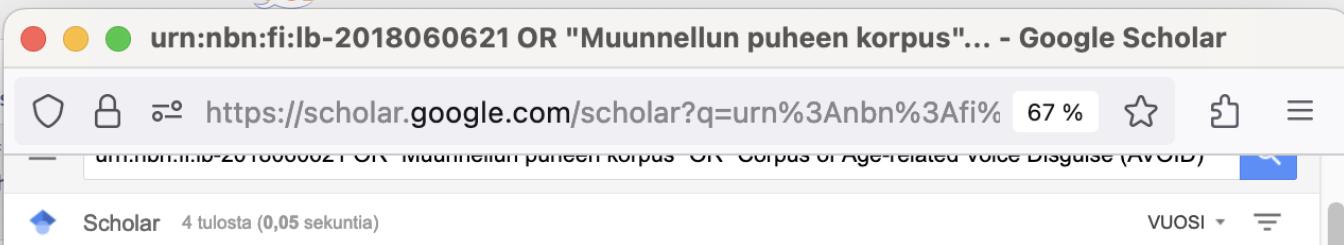
Kinnunen, T., Hautamäki, R. G., Sahidullah, M., Hautamäki, V., Werner, S., & Bentz, M. (2019). *Corpus of Age-related Voice Disguise (AVOID)* [data set]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2018060621>

Show: [Bibtex](#) [Zotero](#)

[Search for references to the language resource in Google Scholar](#)

# Systematic data citations and PIDs are helpful:

amph	amph-Corpus [file]	ACA Download Puhti	”	amph	↗	B
ArkiSyn-korp	ArkiSyn Database of Finnish Conversational Discourse, Helsinki Korp Version [file]	PUB Korp	”	arkisyn	2017	B
AVOID	Corpus of Age-related Voice Disguise (AVOID) [file] [audio]	RES Download	”	avoid	2019	B
BeserCorp	The Corpus of Beserman Udmurt [file]	PUB Korp	”	besercorp	2016	B
ccmh-src	Corpus Cyrillo-Methodianum of Old Church Slavonic Texts, s... [file]					
ceal-dl	The Downloadable Version of American Literature in Finnish [file]					
ceal-o	Classics of English and American Finnish, Sentences and Paragraph Order [file]					
ceal-par-korp	Classics of English and American literature by Kersti Juva, English-Finnish Korp [file]					
ceal-par-s-dl	The Downloadable Version of American Literature as translated into English-Finnish parallel corpus [file]			On the limits of automatic speaker verification: Explaining degraded recognizer scores through acoustic changes resulting from voice disguise <a href="#">R González Hautamäki, V Hautamäki...</a> - The Journal of the ..., 2019 - asa.scitation.org	[PDF] uef.fi Full View	
ceal-par-s-korp	Classics of English and American literature by Kersti Juva, English-Finnish, Scrambled, Korp [file]			In speaker verification research, objective performance benchmarking of listeners and automatic speaker verification (ASV) systems are of key importance in understanding the limits of ... ☆ Tallenna 9 Viittaa Viitauksen määrä 17 Aiheeseen liittyviä artikkeleita Kaikki 9 versiota		
ceal-s	Classics of English and American Finnish, Scrambled Paragraphs [file]			Can we use speaker recognition technology to attack itself? enhancing mimicry attacks using automatic target speaker selection <a href="#">T Kinnunen, RG Hautamäki, V Vestman...</a> - ICASSP 2019-2019 ..., 2019 - ieeexplore.ieee.org	[PDF] arxiv.org Fulltext - Univ.Helsinki	



# CLARIN license categories



Publicly available



Available for academic, logged in users



Access is based on an individual application

*Language Bank Rights: lbr.csc.fi*

# Safety measures for sharing personal data

- The Language Bank of Finland can offer various ways that help in protecting personal data and other types of restricted content:
  1. Access management, if needed: university login (ACA); access granted on an individual basis upon application (RES)
  2. Resource-specific data protection terms and conditions (must be accepted by the end-users)
  3. Data encryption (first resource: [findarc](#))
  4. Sensitive Data (SD) services at CSC

# Findable

Consistent metadata



**KIELIPANKKI**  
The Language Bank of Finland

[www.kielipankki.fi](http://www.kielipankki.fi)



Virtual Language Observatory

# PID

+ Access location  
PID

# Accessible

DOWNLOAD

**KORP**  
v9



**PUB**

 **eduGAIN**

**ACA**

 **Haka**

**RES**

Language  
Bank  
Rights

# FAIR data

Instructions for  
resource creators

HRT / VRT  
common formats

Common  
processing  
tools

# Interoperable



Deposition  
agreements

Support for  
versions and  
variants

Long-term  
archiving  
(?)

# Re-usable

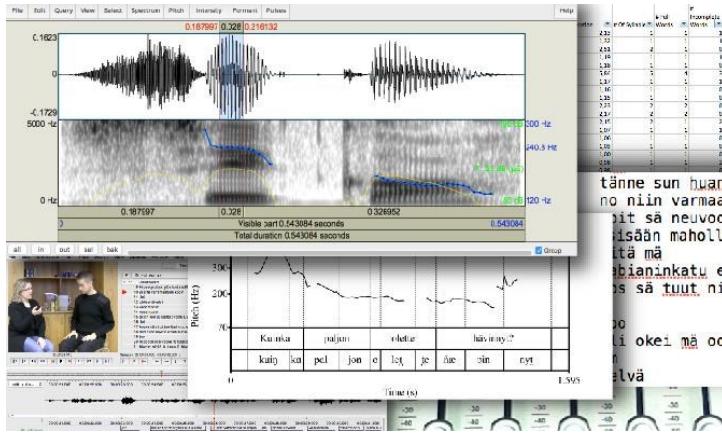
## Corpus Linguistics and Statistical Methods (5 cr)

**Word attributes**

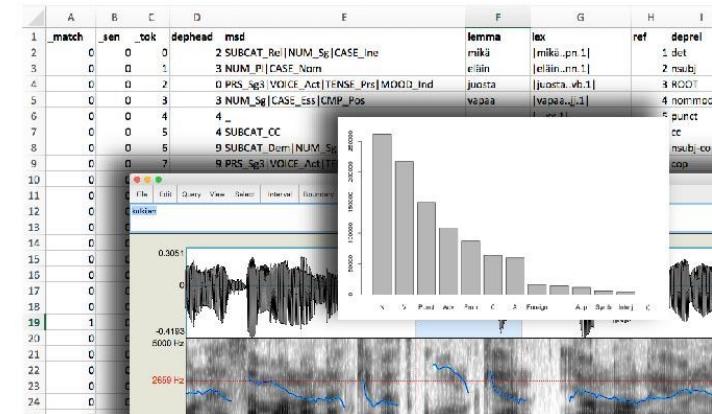
- baseform: kieli
- baseform (compound boundaries): kieli
- part-of-speech: noun
- msd: NUM\_Sg|CASE\_Nom
- dependency relation: nominal subject
- word position in a name: outside (O)
- original thread
- original message
- Show Dependency Tree
- Näytä dependenssipuu



## Introduction to Speech Analysis (5 cr)



## Data Clinic (5 cr)



# Researchers of the Month

<https://www.kielipankki.fi/language-bank/researcher-of-the-month-archive/>

## The Language Bank of Finland Researchers of the Month

Suomeksi

Do you know researchers who use the Language Bank of Finland and who might be good candidates for Researcher of the Month? Would you be one of them? [Inform us!](#)

Published	Researchers	Corpora/Tools	Publications
2023.04	Johanna Vaattovaara	Suomi24	2022a, 2022b, 2019, 2012, 2011, 2009
2023.03	Noora Hoffréen	CFINSL, snowfrog-dl	2019
2023.02	Maria Sarhemaa	Suomi24	2022, 2021
2023.01	Therese Lindström Tiedemann	topling-sv, sinebrychoff-sv, e-thesis-sv	2019a, 2019b, 2021a, 2018a, 2020, 2018b, 2004, 2022a, 2018c, 2016, 2021b, 2019c, 2022b, 2021c, 2014
2022.12	Marja-Liisa Helasvuo	la-murre, agricola-v1-1, LAS2, LAS1, ArkiSyn, Suomi24	2020, 2019, 2016a, 2016b
2022.11	Marjatta Palander	FINKA, tver-1980, tver-2020	2015, 2018, 2017a, 2017b
2022.10	Benjamin Schweitzer	KLK-fi, Korp	2019
2022.09	Mikko Laitinen	COCA, COHA, ELFA, Suomi24	2020a, 2020b, 2018, 2016, 2020c
2022.08	Filip Ginter	Turku-neural-parser, FinBERT, finnish-parse, Suomi24, KLK-fi	2020a, 2020b, 2018, 2017
2022.07	Sampsu Holopainen	HHP, EVE	2022, 2021, 2020
2022.06	Jack Rueter	ERME, Pabivus, Uralic UD, Fenno-Ugrica	2021a, 2021b, 2020a, 2021c, 2021d, 2020b, 2020c, 2020d, 2021e, 2020e
2022.05	Mika Hämäläinen	SKN, FinnWordNet, Opusparcus, FinMeter,	2020a, 2019a, 2019b, 2019c, 2020b, 2018

Search the Language Bank Portal:

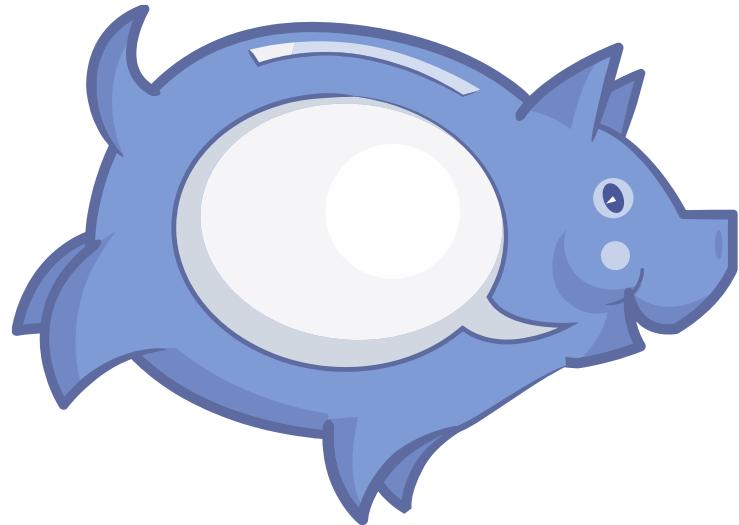
Haku ... Hae



Researcher of the Month: Rosa González Hautamäki

### News

- New resource (beta): Parallel Corpus of the book "Four Battles", written by L. Uspenskij; Korp version (5.6.2023)
- New resource: Finnish Dark Web Marketplace Corpus (findarc) (30.5.2023)
- New resource: The Movie Corpus (Mark Davies, english-corpora.org) - Kielipankki version, source (24.5.2023)
- New resource: The Coronavirus Corpus (Mark Davies, english-corpora.org) - Kielipankki version 2021-05, source (17.5.2023)
- Researcher of the Month: Rosa González Hautamäki (12.5.2023)



Kiitos! Tack! Thank you!

**[www.kielipankki.fi](http://www.kielipankki.fi)**

**General user support**  
*fin-clarin@helsinki.fi*

**Technical services, user accounts  
etc.**

*kielipankki@csc.fi*