

KIELIPANKKI
The Language Bank of Finland

Welcome to the Language Bank of Finland!

Krister Lindén, FIN-CLARIN
krister.linden@helsinki.fi



This document is licensed under the Creative Commons Attribution 4.0 International license. Contents were produced by members of the Language Bank team in FIN-CLARIN (Mietta Lennes, Krister Lindén, Tero Aalto, Erik Axelson, Sam Hardwick, Martin Matthiesen, Jyrki Niemi, Maria Palolahti and Jussi Piitulainen).

FIN-CLARIN

<https://www.kielipankki.fi>

KIELIPANKKI

The Language Bank of Finland

LANGUAGE BANK ACCESS CORPORA TOOLS ORGANIZATION SUPPORT

SUOMEKSI PÅ SVENSKA

Access



Apply for rights to use our language resources.

Corpora



Browse our corpora.

Tools



Try our tools.

Organization



Who are the Language Bank?

Support



Help and instructions.

Search the Language Bank Portal:

Haku Hae



Researcher of the Month: Niko Partanen

News

- Researcher of the Month: Niko Partanen (15.7.2019)
- RDHum2019, August 14-16, preliminary program and abstracts now available (26.6.2019)
- Researcher of the Month: Olli Silvennoinen (14.6.2019)
- CSC news: One of the most competitive



CLARIN ERIC

European Research Infrastructure Consortium
founded on February 29, 2012

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



SWE-CLARIN



INSTITUT FÜR
DEUTSCHE SPRACHE



[...]



*International cooperation
and sharing of resources
for Humanities and Social Sciences*

Member countries:
The Netherlands

Austria
Bulgaria
Croatia
Czech Republic
Denmark
Dutch Language Union
Estonia
Finland
Germany
Greece
Hungary
Italy
Latvia
Lithuania
Norway
Poland
Portugal
Slovenia
Sweden

Observers:

France
Iceland
South Africa
United Kingdom
CMU (USA)



Member countries:

The Netherlands

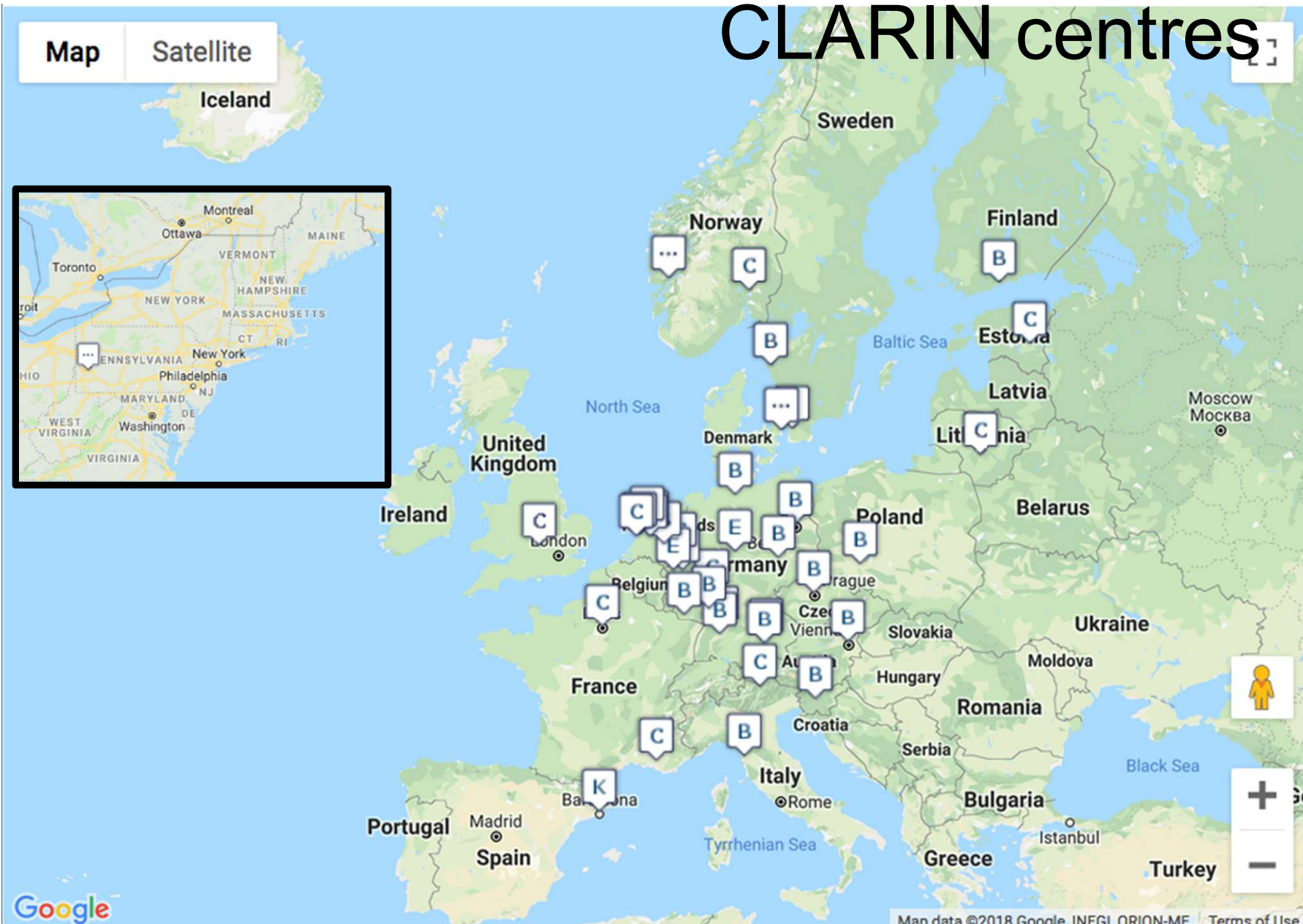
- Austria
- Bulgaria
- Croatia
- Czech Republic
- Denmark
- Dutch Language Union
- Estonia
- Finland
- Germany
- Greece
- Hungary
- Italy
- Latvia
- Lithuania
- Norway
- Poland
- Portugal
- Slovenia
- Sweden

Observers:

- France
- Iceland
- South Africa
- United Kingdom
- CMU (USA)

10.3.2020

CLARIN centres





Online Corpora

Gw = billion words, Mw = million words, h = hours

Resources	2018	2022
Text		
Magazines and newspapers 1770- (NLF and Web publ.)	12 Gw	20 Gw
Social media and similar sources 2000- (Suomi24, Ylilauta, ...)	4 Gw	10 Gw
Literature and manuscripts (Gutenberg, Fennica, archives)	60 Mw	70 Mw
Speech		
News broadcasts (YLE)		10000 h
Video sessions from the Finnish Parliament 2008-2016	500 h	1000 h
Dialect and everyday speech (Kotus, Turku)	500 h	1000 h
Sign language resources (Aalto, Kuurojen liitto)	20 h	500 h
Multilingual and Other Resources		
Multilingual Resources (EuroParl, laws, Bible, subtitles, ...)	3 Gw	10 Gw
Learner's resources (Oulu, Jyväskylä, Kotus, Aalto)	2 Mw	5 Mw
Open source lexicons and terminologies (Helsinki, Tromssa)	200 Kw	400 Kw

Currently, FIN-CLARIN has approx. 19 GW in >1400 databases



Search input field with a magnifying glass icon.

Resource families

Introduction

The aim of the CLARIN Resource Families initiative is to provide a user-friendly overview of the available corpora in the CLARIN infrastructure for researchers from digital humanities, social sciences and human language technologies. The overviews are organized according to the types of data in the corpora and include listings of corpora sorted by language.

The listings include the most important metadata and descriptions on corpus size, text sources, time periods, annotations and licences as well as links to download pages and concordancers, whenever available. In addition to the corpora found in the CLARIN infrastructure we provide an overview of other existing valuable corpora which have not yet been integrated in the infrastructure.

We also provide hyperlinks to other relevant materials such as the thematic CLARIN workshops and tutorials and their accompanying videolectures, as well as a list of key publications on the corpora surveyed.

We currently offer overviews of 8 resource families:

- [Computer-mediated communication corpora](#)
- [Historical corpora](#)
- [L2 learner corpora](#)
- [Newspaper corpora](#)
- [Parallel corpora](#)
- [Manually annotated corpora](#)
- [Parliamentary corpora](#)
- [Spoken corpora](#)



In the future, we plan to include other resource families, such as lexical resources, as well as add tutorials on how to query, annotate and analyse the data.

The overviews have been prepared by Darja Fišer and Jakob Lenardič and have received funding from the European Union's Horizon 2020 research and innovation programme for projects CLARIN-PLUS and PARTHENOS. We would like to thank all the User Involvement coordinators, National Coordinators, workshop participants and other individuals who have participated in the survey and have provided information about the

Resource families

- [CMC corpora](#)
- [Historical corpora](#)
- [L2 corpora](#)
- [Manually annotated corpora](#)
- [Newspaper corpora](#)
- [Parallel corpora](#)
- [Parliamentary corpora](#)
- [Spoken corpora](#)

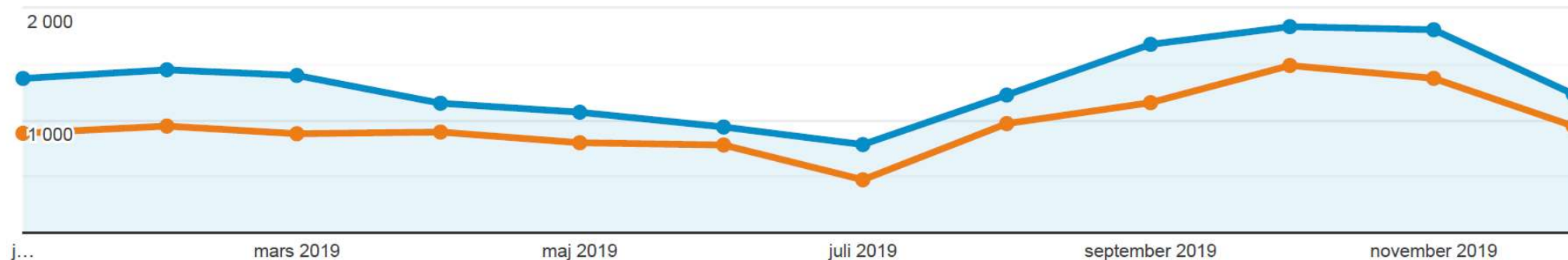
Users of the Language Bank

- Researchers from all fields are welcome!
- Many corpora are available even without signing in.
- FIN-CLARIN can help you with storing and distributing your own corpora.



User Statistics Development

2019-jan-01 - 2019-dec-31: ● Användare
2018-jan-01 - 2018-dec-31: ● Användare



Användare
41,68 %
14 368 kontra 10 141



Nya användare
40,21 %
14 055 kontra 10 024

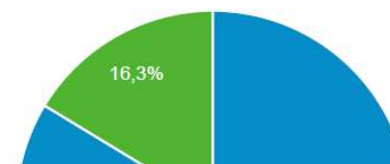


Sessioner
35,10 %
23 130 kontra 17 121




■ New Visitor ■ Returning Visitor

2019-jan-01 - 2019-dec-31



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



KIELIPANKKI
The Language Bank of Finland


LANGUAGE BANK ACCESS CORPORA TOOLS ORGANIZATION SUPPORT SUOMEKSI PÅ SVENSKA

The Language Bank of Finland's Researchers of the Month

An archive of the previous months' Researcher of the Month interviews.

	2019	2018	2017	2016
1	Johanna Kalja-Volma	Krista Lagus	Risto Turunen	
2	Tero Alstola	Matias Tamminen	Jani Marjanen	
3	Ilija Moshnikov	Zsuzsi Máthé	Tommi Jantunen	Päivi Pasanen
4	Emma Nortio	Mari Sirolinen	Jarmo Jantunen	Anna Dannenberg
5	Tim Reus	Ivana Kováčová	Ilmari Ivaska	Marko Pantermöller
6	Olli Silvennoinen	Maximilian Murmann	Juho Härme	Mihail Kopotev
7		Mietta Lennes	Katja Västi	Kirsi-Maria Nummilla
8		Maria Huttu-Hiltunen	Paul-Thor Holmberg	Antti Kanner
9		Annika Liti	Laura-Maija Suur-Askola	Tuija Määttä
10		Tatu Huovilainen	Markus Juutinen	Auroora Vihervalli
11		Niklas Edenmyr, Ali Basirat and Marc Tang	Eero Voutilainen	Markus Hamunen
12		Lari Vainio	Heli Tissari	Hanna Westerlund

Search the Language Bank Portal:



Researcher of the Month: Niko Partanen

News

- Researcher of the Month: Niko Partanen (15.7.2019)
- RDHum2019, August 14-16, preliminary program and abstracts now available (26.6.2019)
- Researcher of the Month: Olli Silvennoinen (14.6.2019)
- CSC news: One of the most competitive supercomputers in the world to be placed in Finland (7.6.2019)
- Researcher of the Month: Tim Reus (15.5.2019)



Researcher of the Month: Heli Tissari



Photo: Kathrin Kaufhold

Kielipankki – The Language Bank of Finland is a service for researchers using language resources. Heli Tissari, docent of the University of Helsinki, tells us about her research on The Suomi 24 Sentences Corpus.

Who are you?

My name is Heli Tissari. At the moment, I am working as a substitute lecturer of English at the University of Stockholm. I am also a docent at the University of Helsinki.

What is your research topic?

I am interested in the meanings of words – in other words, in semantics. Most of all, I am interested in English historical semantics. I am also curious about the relationship between language and the mind, which is why I do cognitive linguistics. I have researched, among all, English words for emotions and their conceptual metaphors. In

Search the Language Bank Portal:

Haku: Hae



Tweets by @FinClarIn

Kielipankki – The Language Bank of Finland



Heli Tissari

Researcher of the Month 12/2017

Substitute lecturer of English at University of Stockholm in 2017; Docent at University of Helsinki

Research interests:

- English historical semantics
- Hate speech in Suomi 24 in collaboration with
 - Mari Siirainen (UHEL)
 - Ulla Vanhatalo (UHEL)

Use of the Finnish emotion words *viha* (hatred, anger), *vihata* ('to hate'), *vihainen* ('angry')

Aller Media Oy (2014). *Suomi 24 virkkeet -korpus (2016H2)* [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-2017021505>

Results

- Conference presentations and interviews
- In an interview by Swedish radio, Heli Tissari has mentioned that people tend to talk rather little about their own anger but more about other people hating or being angry.

The following examples were inspired by Heli Tissari's research topic...

"HATE WORDS" IN THE SUOMI24 SENTENCES CORPUS

Aller Media Ltd. (2017). *The Suomi 24 Sentences Corpus (2017H2)* [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:2019021101>



The Suomi 24 Sentences Corpus has grown

2016H2 version



2017H2 version (current)

- 1.1.2001–24.9.2016
- 2 663 114 497 word tokens
- 256 745 360 sentences

- 1.1.2001-31.12.2017
- 4 132 665 850 word tokens
- 351 985 002 sentences

Aller Media Ltd. (2014). *The Suomi 24 Sentences Corpus (2016H2)* [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2017021505>

Aller Media Ltd. (2017). *The Suomi 24 Sentences Corpus (2017H2)* [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2019021101>

https://korp.csc.fi/#?lang=fi&stats_reduce=word&corp=[]&corpus=s24_2001,s24_2002,s24_2003,s24_2004,s24_2005,s24_2006,s24_2007,s24_2008,s

Suomi | Ruotsi | Muut kielet | Rinnakaistekstit

Viittaa Korpiin Suomi | Svenska | English Kirjautu ulos lennes@helsinki.fi

KORP

17 / 949 korpusta valittuina — 4,13G / 12,87G sanetta

Yksinkertainen Laaje

myös alkuosa loppuos

virkkeistä jotka sisä

Konkordanssi: osumia sivulla:

1600 1700 1800 1900 2000

Valitse kaikki

Tyhjennä

- Agricola (beta) (9)
- E-thesis (24)
- FinnTreeBank: suomen puupankki (3)
- Kansalliskirjaston lehtikokoelman (KLK) suomenkieliset lehdet (179)
- Kirjallisuutta (21)
- Juridisia tekstejä (3)
- Internet-keskusteluaineistoja (30)
 - Suomi24 (10)
 - Suomi24 virkkeet -korpus (2017H2) (beta) (17)
 - Suomi24 2001–2014 (näyte)
 - Yliäuta
 - HS.fi [ACA]
 - 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä (369)
 - Suomen kielen tekstikokoelma (SKTP/FTC) (50)
 - Muita tekstejä (11)
 - Puhuttua kieltä (tekstiksi litteroituna) (137)
 - Suomenoppijoiden kieltä (suomi toisena tai vieraana kielenä) (4)
 - Vanhan kirjasuomen korpus (12)
 - Varhaisnykysuomen korpus (93)
 - Demo- ja testiaineistoja (1)
 - Karjalansuomi [ACA]

sana Näytä sanakuva Näytä kartta

Suomi24 virkkeet -korpus (2017H2) (beta)


Suomi24-keskustelupalvelun keskustelut vuosilta 2001–2017 (1.1.2001–31.12.2017). Aineistossa näkyy kaikkien keskustelujen sisältö enintään kappaletasolla. Aineisto on jaettu osakorpuksiin vuosittain. Tutkijat voivat myös ladata käyttöönsä koko Suomi24-aineiston Kielipankin latauspalvelusta (lisenssi).

Huomaa, että korpus on beetatestausvaiheessa, joten siihen voi tulla muutoksia ilman ilmoitusta.

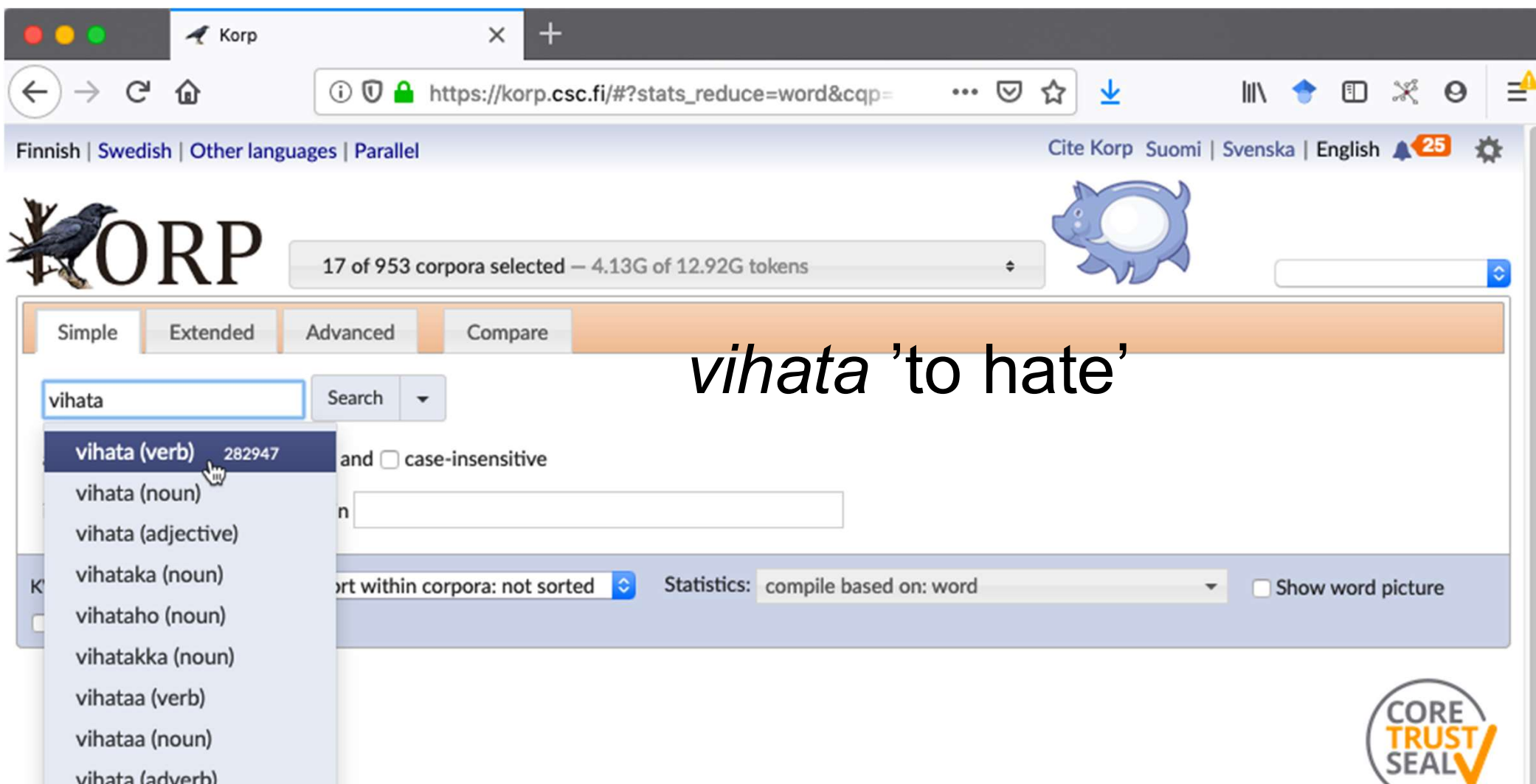
Kuvailutiedot
 Lisenssi: CC BY-NC (CLARIN PUB)
 Viittaa korpukseen
 Linkki korpukseen Korpissa: [to be added]
 Kotisivu

17 korpusta, joissa:

4 132 665 850 sanetta
351 985 002 virkettä

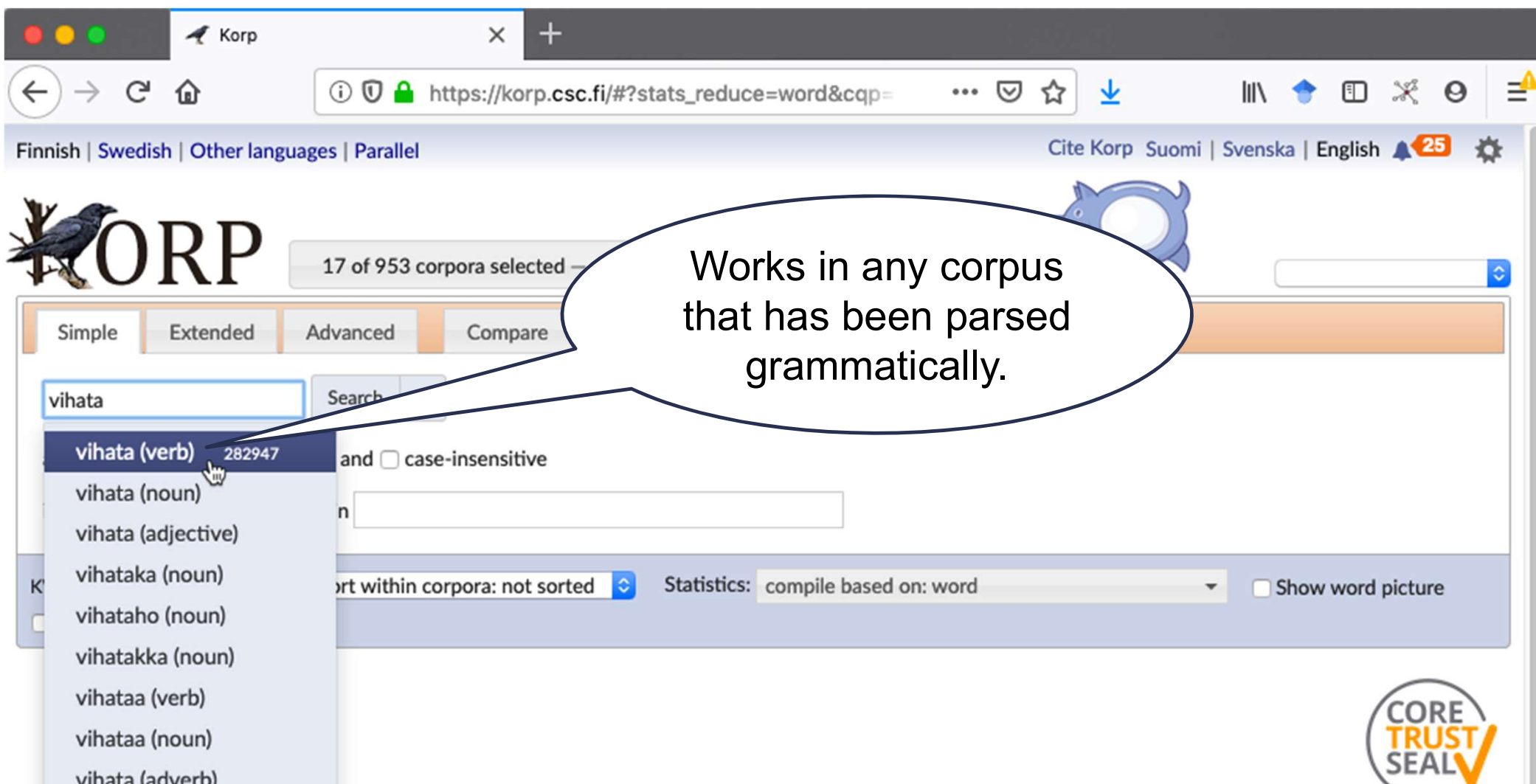


Simple search: All forms of a word



The screenshot shows the Korp search interface. The search bar contains the word "vihata". A dropdown menu is open, listing the following forms: "vihata (verb) 282947", "vihata (noun)", "vihata (adjective)", "vihatakka (noun)", "vihataho (noun)", "vihataa (verb)", "vihataa (noun)", and "vihata (adverb)". The "Simple" search mode is selected. The text *vihata* 'to hate' is overlaid on the search results area. The interface also shows a search bar with "vihata", a "Search" button, and a "case-insensitive" checkbox. The "Statistics" section shows "compile based on: word" and a "Show word picture" checkbox.

Simple search: All forms of a word



The screenshot shows the Korp search interface. The search term "vihata" is entered in the search box. A dropdown menu displays the following grammatical forms and their counts:

- vihata (verb) 282947
- vihata (noun)
- vihata (adjective)
- vihataka (noun)
- vihataho (noun)
- vihatakka (noun)
- vihataa (verb)
- vihataa (noun)
- vihata (adverb)

A callout bubble points to the "vihata (verb) 282947" entry with the text: "Works in any corpus that has been parsed grammatically."

The interface also shows search options like "Simple", "Extended", "Advanced", and "Compare". The "Simple" option is selected. The search results are displayed in a table with columns for "Port within corpora" (not sorted) and "Statistics" (compile based on: word). There is a checkbox for "Show word picture".



Simple search: All forms of a word

Simple Extended Advanced Compare

vihata (verb) Search

also as initial part final part and case-insensitive

in sentences which contain

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC Statistics Word picture Map

Results: 282,947

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Show context

SUOMI24 2017H2: 2001

kokemuksilla voin sanoa, että näin ei tosiaankaan ole,	vihaisin	koirani on ollut suomesta ..
akelpoisemmiksi ,kun jokatapauksessa meidän Mustia	vihataan	koska lööpissä luki että joku yks koira puri.Se nyt v
se et näe mitään muuta kuin oman ajatusmaailmasi ja	vihasi	.
äilytettyä oman sairautesi ja tuskasi toisille, niille joita	vihaat	.
hankala kytätä mutta ikävä kyllä täällä riittä niitä koiria	vihaavia	puskamummoja...
"Jeesusta	vihaavien	"kommentteja varmaan tulee, mutta sallinette min
Onhan niitä ohjeita kuten "älä anna auringon laskea	vihasi	ylle "vastaava löytyy varmasti myös anteeksipyynt
Itse pelkään koiria ja	vihaan	sitä, että niiden omistajat antaa niiden tulla haiste
, joka vastoin tahtoaan asustelee paikkakunnalla jota	vihaa	!
Ensiksikin sille, joka väitit että Suomessa koiria	vihataan	enemmän... voihan olla noinkin MUTTA... muualla r
Minäkin	vihasin	urheilua vähintään yhtä kiihkeästi kun jotkut koiria,
Minulla oli aijemmin 40kg koira, joka	vihasi	yli kaiken toisia uroksia, minulla oli valtavia ongelm
	Vihaan	sydämeni pohjasta sanontoja typ ' se mikä ei tapa v
Kysymys on vain, miksi ihmiset	vihaavat	nin paljon.

Corpus

Suomi24 2017H2: 2001

Subcorpus of: Suomi24 2017H2
Persistent Identifier (metadata):
urn:nbn:fi:lb-2019021101
[Cite corpus](#)
Licence: CC BY-NC (CLARIN PUB)
[Link to corpus in Korp](#)

Text attributes

title: Puolesta vai vastaan ??
date: 2001-01-10
time: 12:26:00
writer nickname: essi
user logged in: no
registered nickname: no
topic: Lemmikit > Kissat > Yleistä kissoista
topic adult-only: no

Word picture: *vihata* (verb) Suomi 24 Sentences Corpus (2016H2)



10 of 953 corpora selected – 2.66G of 12.92G tokens



vihata..vb.1

Simple Extended **Advanced** Compare

vihata (verb)

Search



also as initial part final part and case-insensitive

in sentences which contain

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map



Word picture: *vihata* (verb)

Simple Extended Advanced Compare

vihata (verb) Search

also as initial part final part and case-insensitive

in sentences which contain

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC Statistics Word picture Map

Show word class
Show up to 50 hits

Subject	vihata	Object	Adverbial
1. maailma	1378	1. homo	2062
2. Jumala	3100	2. veli	1550
3. mä	3334	3. juutalainen	1348
4. lietso	137	4. tee	2179
5. muslimi	435	5. islam	848
6. kansa	707	6. nainen	4205
7. vihaaja	130	7. suomalainen	1216
8. saatana	332	8. synti	2148
9. Esauta	47	9. hyljätä	403
10.feministi	155	10.kristitty	687
11.tyhmä	178	11.toinentaka	438
12.ateisti	330	12.minä	7798
13.juutalainen	217	13.vihollinen	522
14.Hitler	157	14.kukaan	1947
15.108.vai	27	15.venäläinen	419
16.vuodata	27	16.Israel	471
17.kohdistaa	35	17.ihminen	4373
18.sielunvihollinen	70	18.toinen	2208
19.paholainen	100	19.muslimi	542
20.ääri-misantrooppi	24	20.mies	3436
21.pakolainen	117	21.vääryys	496
22.persut	149	22.vihamies	221
		1. miksi	4963
		2. poliisi	3817
		3. suorastaan	710
		4. siksi	1919
		5. eniten	978
		6. viha	483
		7. syvästi	430
		8. silmittömästi	226
		9. syy	1226
		10.itteeni	211
		11.rakastaa	391
		12.ylikaiken	135
		13.täynnä	353
		14.nimi	899
		15.uusiakuitanne	57
		16.sydän	410
		17.avoimesti	247
		18.isä	464
		19.katkerasti	101
		20.itteäni	88
		21.pohja	451
		22.verisesti	85

The most prominent objects of 'to hate', in Suomi 24:

1. gay (person)
2. brother
3. jew
4. tea
5. islam
6. woman

Suomi 24: Link to the original message

KWIC
Statistics
Word picture
Map

Results: 186,952

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Show context

SUOMI24 2016H2: 1/10

Hänen kauttaan sainkin tämän työn ja vaikka rehellisesti	vihaan	tätä, niin en anna näkyä vaan teen työni uskollisesti.
Mielestäni on liioiteltua sanoa että	vihaisin	manneja.
JA TEUTA ESIINTYY ENIMMÄKSEEN JEESUSTA	VIHAAVANA	JA JEESUKSEN NIMEEN VALEHTELEVANA JOKA SUORAAN; AVOI
apsille kolmanteen ja neljänteen polveen, niille, jotka minua	vihaavat	- 2. Moos.
Jos vaikka isoisäni	vihaisi	jumalaa, tarkoittaako se sitä, että minä saan siitä rangaistuksen?
Heh, ennen täällä	vihattiin	ruotsinkielisiä, jotka kuulemma puhuivat ratikassa liian äänekkäästi.
Venäjää	vihataan	ja pelätään.
ntely kiinostaa taas mua siksi koska pidän niistä työajoista (vihaan	aikasin aamuhuheräämisä...
Joka	vihaa	Herraa sebaotin ja ei usko raamattuun eikä jeesuksen/jeshua olevan J
Elä siinä sitten	vihasi	kanssa ;)
Mutta kaikkihan huomaa että	vihasi	ajaa sinua tyhmiin sanoihin ;)
On hirveän tärkeää, ettei ketään jätetä yksin	vihaamaan	venäläistä julmuutta.
16 Ja ne kymmenen sarvea, jotka sinä näit, ja peto, ne	vihaavat	porttoa ja riisuvat hänet paljaaksi ja alastomaksi ja syövät hänen lihan
Pelkkä kansakoulupohja ja	vihaa	ruotsalaisia.
ovat kokoomuslaisia ja kokoomuslaiset ajavat Nato-asiaa ja	vihaavat	venäläisiä ja motoristeja.
Mä ainakin	vihaan	sua ja jos et sä lähde, niin se sun musiikillinen päänaukominen loppuu
sa jo täysin varma siitä, että kaikki suomalaiset ovat Venäjää	Vihaan	olla yötä vieraassa paikassa, enkä tiedä miksi edes suostuin, nyt hävel
He	vihaavat	raisteja, jotka kohtelevat omaa venäläisvähemmistöään koirina.
en	vihaavat	muslimeja ja pyrkivät tappamaan ja hävittämään heidät Israelista.
- Jumalaa	vihaa	miestäni.
Meillä oli yks sijaisluokanvalvoja, joka vihas muo ja mä	vihaa	sua ja jos et sä lähde, niin se sun musiikillinen päänaukominen loppuu
oita täällä on muutama, jotka yrittävät kaikkensa, että heitä	vihaavat	raisteja, jotka kohtelevat omaa venäläisvähemmistöään koirina.
On se rankkaa kun haluaisit huutaa,	vihaa	muslimeja ja pyrkivät tappamaan ja hävittämään heidät Israelista.
ta kiinnostuneitten porukoista joita juuri tämä pieni persuja	vihaava	pikku nassukka tuntuu edustavan.

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > »

Download KWIC as Annotations Excel (XLS)

[JSON](#)

Corpus

Suomi24 2016H2: 1/10

Subcorpus of: Suomi24 2016H2
 Persistent Identifier (metadata):
 urn:nbn:fi:lb-2017021505
[Cite corpus](#)
 Licence: CC BY-NC (CLARIN PUB)
[Link to corpus in Korp](#)

Text attributes

title: En halua lähteä
 title word lemmas: ei halua lähteä
 date: 18.06.2015
 text_time: 23:01
 discussion thread id: 13653541
 suomi24fi_cid: unspecified
 section: Nuoret
 subsection: Vapaa-aika
 username: Eimunjuttu
 username word lemmas: Eimunjuttu

Word attributes

base form: vihata
 baseform (compound boundaries): vihata
 part of speech: verb
 msd: PRS_Sg1|VOICE_Act|TENSE_Prs|
 MOOD_Ind|CASECHANGE_Up
 dependency relation: root
 word position in a name: outside (O)

[original thread](#)
[original message](#)
[Show Dependency Tree](#)



Aihealueet

Etsi aihealuetta



Kaikki aihealueet

Nuoret

Aikuistuminen

Bailut

Itsetunto

Keho

Kiusaaminen

Koulu

Pulinat

Rakkaus

Keskustelu24 > Nuoret > Vapaa-aika

En halua lähteä



Eimunjuttu

18.6.2015 23:01

kaverin kanssa leirille.

Leiri kestäisi viikon ja maksaisi satasen, ajalle ja rahalle olisi parempaakin käyttöä.

Vihaan olla yötä vieraassa paikassa, enkä tiedä miksi edes suostuin, nyt hävettää.

Olen aikaisemmin ollut leireillä, mutta lähtenyt kesken pois.

Miten kerron kaverille ja vanhemmille, etten halua lähteä.



JAA

ILMOITA

2 Vastausta

KIRJOITA VASTAUS



Keskustele

KESKUSTELE AIHEESTA

SEISKA.FI

Sofia Belórf etsi napapaitaa Tuurin kyläkaupasta - onko tämä kuvissa o

Tässä on Martinan ja Stefanin uusi 1,3 miljoonan euron koti: hulppeassa as

IS: Sami Kuronen avoimena parisuhteesta
Michaela Söderholmiin: "Mitä sitä

” I hate to spend the night in a strange place, [...]

Do 'I' hate less often than 'he/she/it' does?

...ä kansakoulupohja ja	vihaa	ruotsalaisia.
...t ajavat Nato-asiaa ja	vihaavat	venäläisiä ja motor
Mä ainakin	vihaan	sua ja jos et sä lähc
	Vihaan	olla yötä vieraassa
...malaiset ovat Venäjää	vihaavia	fasisteja, jotka koh
He	vihaavat	muslimeja ja pyrkiv
en	vihaa	miestäni.

...maa venäläi
...naan ja häv
...ina myötä.
...tavan.

Word attributes

base form: vihata
baseform (compound boundaries): vihata
part of speech: verb
msd: PRS_Sg1|VOICE_Act|TENSE_Prs|
MOOD_Ind|CASECHANGE_Up
dependency relation: root
word position in a name: outside (O)

[original thread](#)
[original message](#)



Simple Extended **Advanced** Compare

base form is

or

and

part of speech is

or

and

msd begins with

or

Search within

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC **Statistics** Word picture Map

Results: 88,054

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Show context

SUOMI24 2016H2: 1/10

Hänen kautaan sainkin tämän työn ja vaikka rehellisesti **vihaan** tätä, niin en anna näkyä vaan teen työni uskollisesti.
 Mielestäni on liioiteltua sanoa että **vihaisin** manneja.
 Jos vaikka isoisäni **vihaisi** jumalaa, tarkoittaako se sitä, että minä saan siitä rangaistuksen
 Ravintola-ala ja baarissa työskentely kiinnostaa taas mua siksi koska pidän niistä työajoista (**vihaan** aikasin aamuheräämisiä...
 Joka **vihaa** Herraa sebaotin ja ei usko raamattuun eikä jeesuksen/jeshua ol
 Elä siinä sitten **vihasi** kanssa ;)
 Mutta kaikkihan huomaa että **vihasi** ajaa sinua tyhmiin sanoihin ;)
 Pelkkä kansakoulunohja ja **vihaa** ruotealaisia

Corpus

Suomi24 2016H2: 1/10

Subcorpus of: Suomi24 2016H2
 Persistent Identifier (metadata):
 urn:nbn:fi:lb-2017021505
[Cite corpus](#)
 Licence: CC BY-NC (CLARIN PUB)
[Link to corpus in Korp](#)

Text attributes

title: En halua lähteä

The verb 'to hate'
 in any person:
88,054 hits



Simple Extended **Advanced** Compare

base form is Aa

or

and

part of speech is

or

and

msd begins with

or

The verb 'to hate'
in 1st person singular:
30,469 hits

Search within

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC **Statistics** Word picture Map

Results: 30,469

« < **1** 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Show context

SUOMI24 2016H2: 1/10

Hänen kauttaan sainkin tämän työn ja vaikka rehellisesti **vihaan** tätä, niin en anna näkyä vaan teen työni uskollisesti.
Mielestäni on liioiteltua sanoa että **vihasin** manneja.
tola-ala ja baarissa työskentely kiinnostaa taas mua siksi koska pidän niistä työajoista (**vihaan** aikasin aamuhieräämisiä...
Mä ainakin **vihaan** sua ja jos et sä lähde, niin se sun musiikillinen päänaukominen loppuu.
Vihaan olla yötä vieraassa paikassa, enkä tiedä miksi edes suostuin, nyt hävettää.
Meillä oli yks sijaisluokanvalvoja, joka vihas muo ja mä **vihasin** sitä.
Minäpä kerron teille nyt miksi **vihaan** maalaisia niin kovasti.
Tästä svstä **vihaan** maalaisia. vihaan niin maan perkeleesti!

Corpus

Suomi24 2016H2: 1/10

Subcorpus of: Suomi24 2016H2

Persistent Identifier (metadata):

urn:nbn:fi:lb-2017021505

Cite corpus

Licence: CC BY-NC (CLARIN PUB)

[Link to corpus in Korp](#)

Text attributes

title: En halua lähteä



Simple Extended **Advanced** Compare

Aa

or

and

or

and

Aa

or

Search within

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC **Statistics** Word picture Map

Results: 46,053

« < **1** 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > » Show context

SUOMI24 2016H2: 1/10

Jos vaikka isoisäni	vihasi	jumalaa, tarkoittaako se sitä, että minä saan siitä rangaistuksen?
Joka	vihaa	Herraa sebaotin ja ei usko raamattuun eikä jeesuksen/jeshua olevan Jumalan p
Elä siinä sitten	vihasi	kanssa ;)
Mutta kaikkihan huomaa että	vihasi	ajaa sinua tyhmiin sanoihin ;)
Pelkkä kansakoulupohja ja	vihaa	ruotsalaisia.
Meidän oma psykomme	vihaa	tänne maalta-ja maahanmuuttajia.
hän on jopa osallistunut Pyhien surmaamiseen, sillä Lutherin kirje paljastaa, että hän	vihasi	sellaisia ihmisiä, jotka opettivat, että ihmisen tulee tehdä hyvää pelastuakseen,
Etkö itsekin huomaa. että olet kanavoinut	vihasi	ihan väärään kohteeseen?

The verb 'to hate'
in 3rd person singular:
46,053 hits

Corpus

Suomi24 2016H2: 1/10

Subcorpus of: Suomi24 2016H2
Persistent Identifier (metadata):

urn:nbn:fi:lb-2017021505

[Cite corpus](#)

Licence: CC BY-NC (CLARIN PUB)

[Link to corpus in Korp](#)

Text attributes

title: En halua lähteä

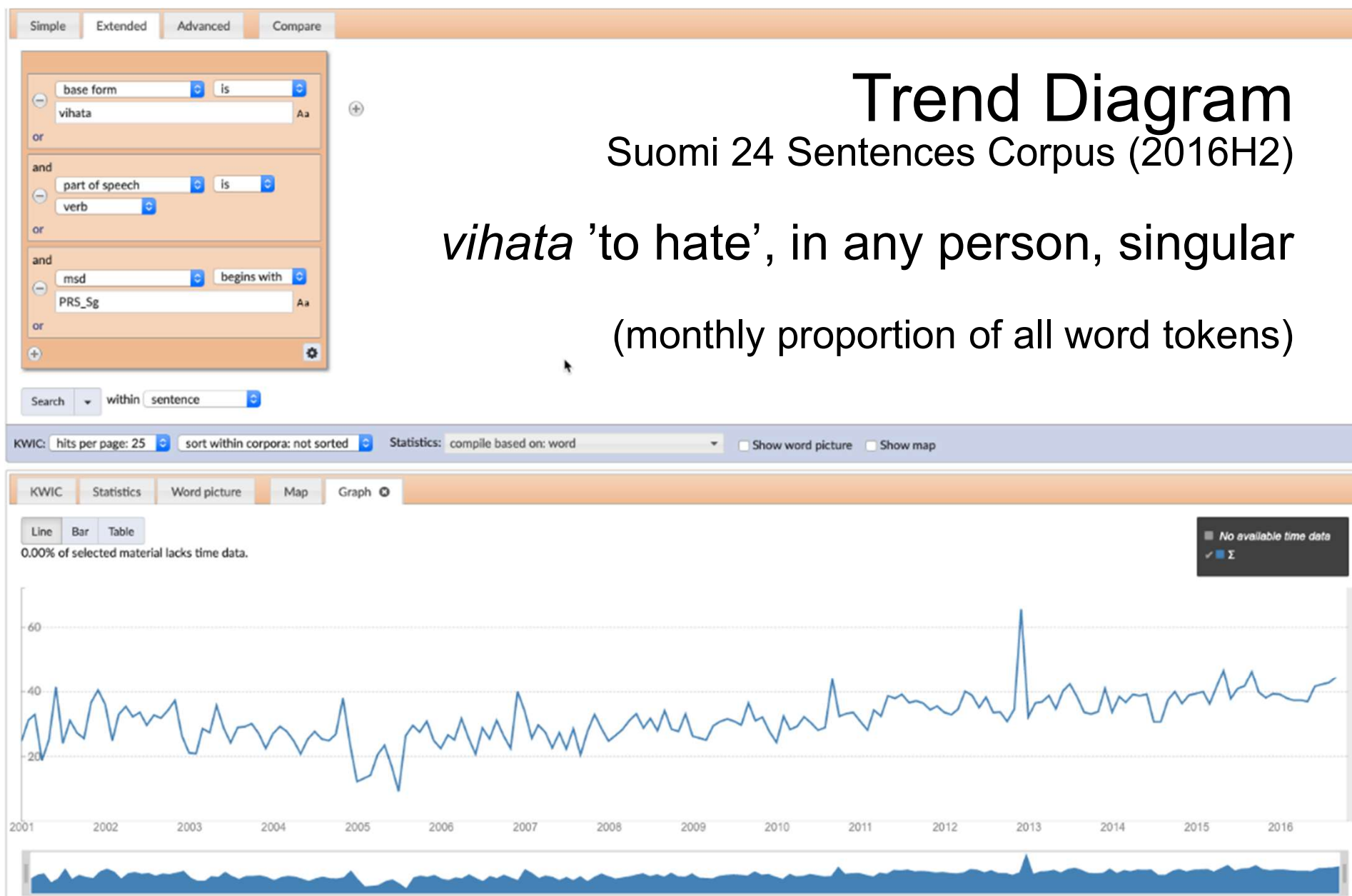
Do people talk about "hating"
more than before?



Trend Diagram

Suomi 24 Sentences Corpus (2016H2)

vihata 'to hate', in any person, singular
(monthly proportion of all word tokens)



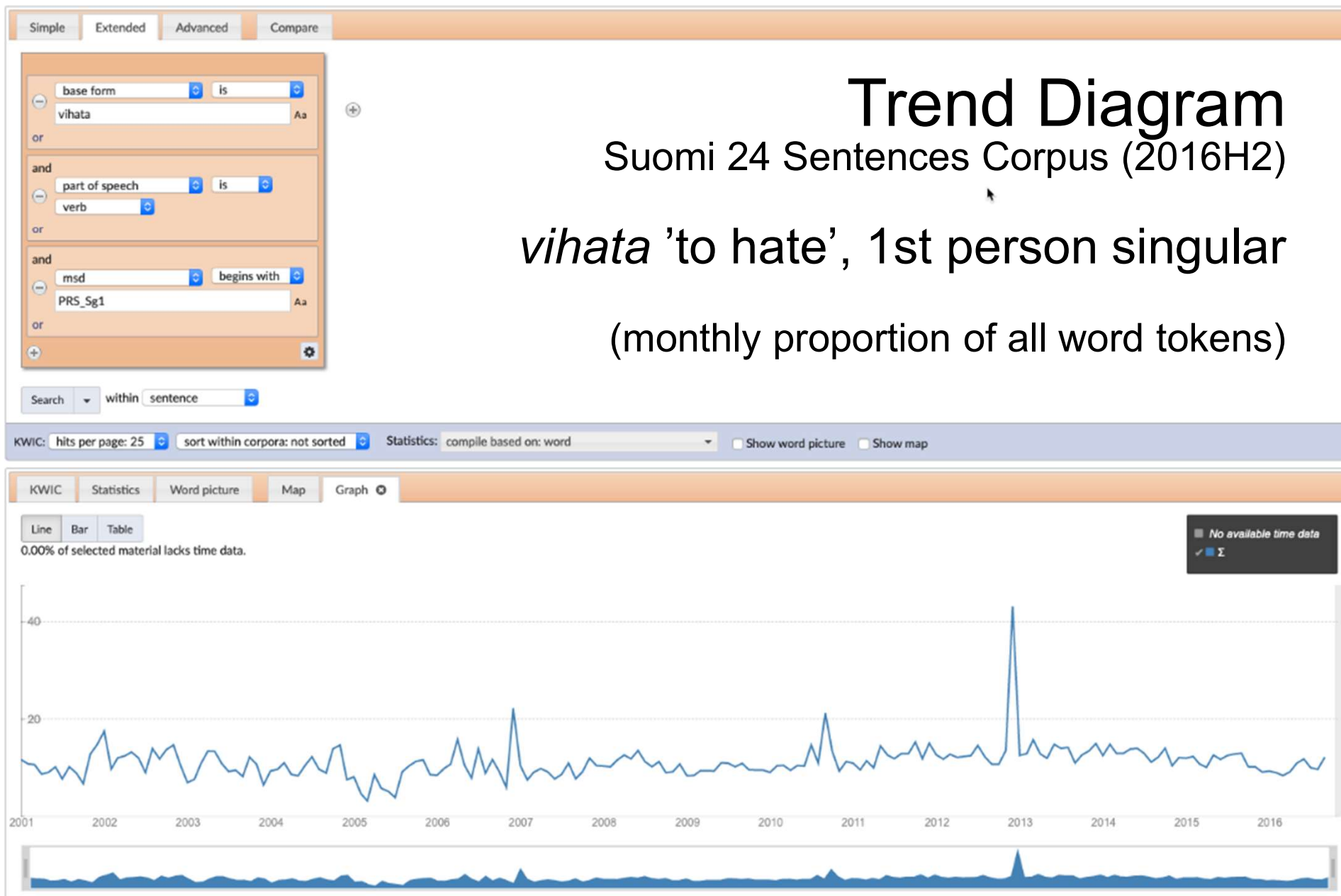


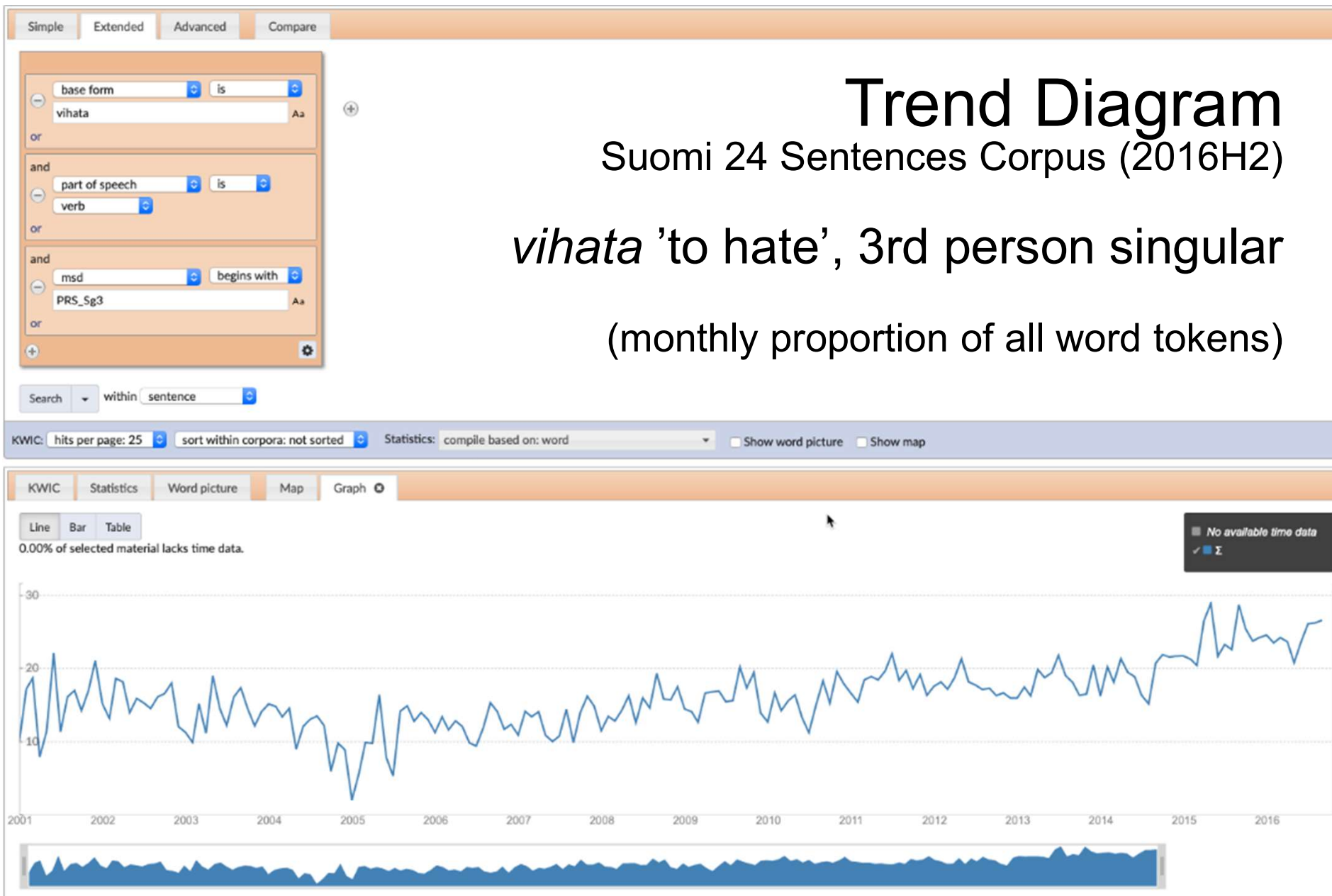
Trend Diagram

Suomi 24 Sentences Corpus (2016H2)

vihata 'to hate', 1st person singular

(monthly proportion of all word tokens)





Trend Diagram

Suomi 24 Sentences Corpus (2016H2)

vihata 'to hate', 3rd person singular

(monthly proportion of all word tokens)



Simple Extended Advanced **Compare 3**

Compare **vihata - SG1** with **vihata - SG3** and compile based on **section**

Compare

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture

KWIC Statistics Word picture Map Graph Compare Compare Compare

Prominent in *vihata - SG1*

Terveys	Health	4490
Nuoret	Young People	2321
Suhteet	Relationships	7033
Ajanviete	Pastime	838
Tiede ja teknologia	Science & Tech	729
Muoti ja kauneus	Fashion & Beauty	223
Työ ja opiskelu	Work and Studies	432
Viihde ja kulttuuri	Entertainment & Culture	1676
Koti ja rakentaminen	Home & Building	246
Harrastukset	Hobbies	309
Ruoka ja juoma	Food and Drink	76
Urheilu ja kuntoilu	Sports	222

Prominent in *vihata - SG3*

Yhteiskunta	Society	25948
Paikkakunnat	Places/Municipalities	1908
Lemmikit	Pets	902
Tori	Marketplace	2
Suomi24		4

Compare searches

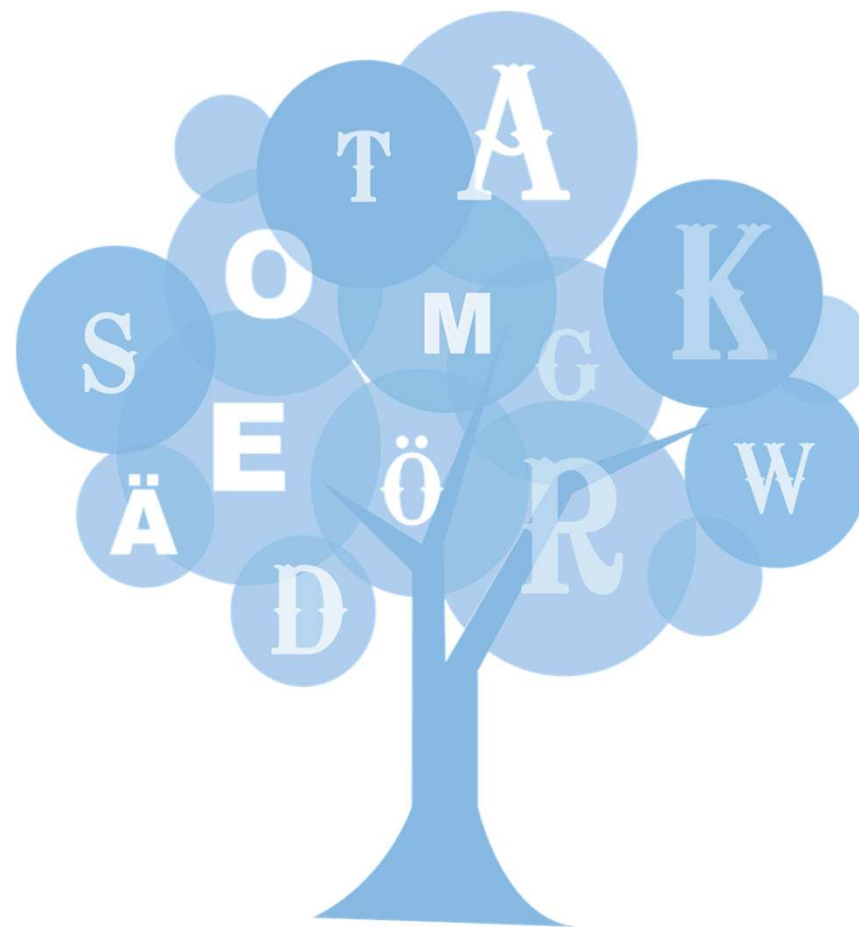
vihata 'to hate':
1st person singular
vs.
3rd person singular

Compare by section
within Suomi 24

(The comparison is based on the log-likelihood measure)

USE OF "HATE WORDS"

PART 2: SENTIMENT ANALYSIS





17 / 949 korpusta valittuina — 4,13G / 12,87G sanetta

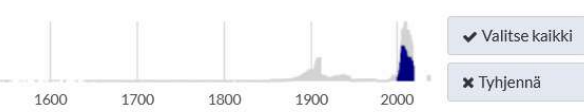
Yksinkertainen Laaje

viinainen (adjektiivi)

myös alkuosa loppuos:

virkkeistä jotka sisält

Konkordanssi: 50 järj



- Agricola (beta) (9)
- E-thesis (24)
- FinnTreeBank: suomen puupankki (3)
- Kansalliskirjaston lehtikokoelman (KLLK) suomenkieliset lehdet (179)
- Kirjallisuutta (21)
- Juridisia tekstejä (3)
- Internet-keskusteluaineistoja (30)
 - Suomi24 (10)
 - Suomi24 virkkeet -korpus (2017H2) (beta) (17)
 - Suomi24 2001-2014 (näyte)
 - Yliilauta
 - HS.fi [ACA]
 - 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä (369)
 - Suomen kielen tekstikokoelma (SKTP/FTC) (50)
 - Muita tekstejä (11)
 - Puhuttua kieltä (tekstiksi litteroituina) (137)
 - Suomenoppijoiden kielitä (suomi toisena tai vieraana kielenä) (4)
 - Vanhan kirjasuomen korpus (12)
 - Varhaisnykysuomen korpus (93)
 - Demo- ja testiaineistoja (1)

Konkordanssi Tilasto

Tuloksia: 101 859

« < 1 2 3 4 »



viinainen..jj.1

Näytä sanakuva Näytä kartta

2001 (BETA)

ja katkeroituneita sinua kohtaan, he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi kipeä

en ?

ksi "

lta vaikuta.

en kenellekään, siis anteeksi antamisen tarpeessa.

en

ksi .

en tuhauksen ja lamauttavan katseen.

ia vaan välinpitämättömiä.

en " mun ja mun jätkän takia...

en ,närkästynyt meidän (sinun kuvittelemiisi) valintoihimme?

en yhteiskunnalle, joka pakottaa molemmat vanhemmat käymään pitkiä päiviä töissä.

na puraisee niin sitten jo voi itseltä kysyä, mikä sai kanin niin pahalle tuulelle, ja lyödä vaikkapa itseään

en) ja komentaa tavallista topakemmin meidän 7 kk ikäistä kolkikissaa (ei leikkattu).

en .

Korpus

Suomi24 (2017H2) 2001 (beta)

Kuvailutiedot

Lisenssi: CC BY-NC (CLARIN PUB)

Viittaa korpukseseen

Linkki korpukseseen Korpissa: [to be added]

Tekstin piirteet

otsikko: Lapsen kurittaminen

päiväys: 2001-08-24

kellonaika: 15:37:00

kirjoittajan nimimerkki: No Mercy?

kirjautunut käyttäjä: ei rekisteröity nimimerkki: ei aihealue: Ryhmät > Ikäryhmät > 30 plus aihealue vain aikuisille: ei viestin tyyppi: kommentti viesti on täysin tyhjä: ei tunniste: 7540:33239 keskusteluketjun tunniste: 7540 keskusteluketjun alkuaikaleima: 2001-07-25 15:38:00

Olet hyvin aggressiivinen, viinainen ja itsekeskeinen.

Edessä hämmötti iso lasi-ikkuna ja takana iso ja viinainen voimalapääliikkö.

Katkerana ja viinainen ihminen tekee pahaan jälkeä sekä kertoo ehkä sen vielä kaikille.

näille telepaatikoillekaan ei siis kannata olla viinainen , puhumattakaan, että heitä kannattaisi tuomita, sillä välitön ensiapu ihmiselle tulee jo siitä, että ilm



17/249 korpusa valittuina — 4,13G / 12,87G sanetta



vihainen..jj.1

Ksinkertainen Laajennettu Edistynyt Vertailu

vihainen (adjektiivi) Hae

myös alkuosa loppuosa ja samaista pien- ja suuraakokset

virkkeistä jotka sisältävät

Konkordanssi: 50 Järjestä korpuksen sisällä: satunnaisesti Tilastoja: laskentaperuste: sana Näytä sanakuva Näytä kartta

vihainen adj. 'angry, full of hate'

Konkordanssi Tilastoja Sanakuva Kartta

Tuloksia: 101 859

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... » Näytä konteksti

SUOMI24 (2017H2) 2001 (beta)

Siten kun lapsesi huomaavat olevansa	vihaisia	ja katkeroituneita sinua kohtaan, he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi kipeä
Sinä olet ilmeisesti uskova koska olet minulle	vihainen	?
Eli taas sarjassamme " yritänpä v.....ssa saada porukan	vihaiseksi	".
Ei Kaizun kiemurtelu nyt kovin	vihaiselta	vaikuta.
- Ei tuo ihminen ole välttämättä katkera satuttajalleen tai ei hän koe olevansa	vihainen	kenellekään; siis anteeksi antamisen tarpeessa.
noista perusteluistasi selvemmin, muuten me muut emme ollenkaan tajua miksi olet niin närkästynyt ja	vihainen	.
Minua kammottaa, tulen toivottoman	vihaiseksi	.
miehestä tarinoiva rouva näytti siltä kuin olisi aikonut sanoa siihen jotain mutta päästikin sen sijaan vain	vihaisen	tuhauksen ja lamauttavan katseen.
Vieraista eivät perusta, eivät ole	vihaisia	vaan välinpitämättömiä.
Tai siis en mä tiä onx mun bestis mulle sillee "	vihainen	" mun ja mun jätkän takia...
Kateellinen,	vihainen	, närkästynyt meidän (sinun kuvittelemiisi) valintoihimme?
Olen	vihainen	yhteiskunnalle, joka pakottaa molemmat vanhemmat käymään pitkiä päiviä töissä.
jos se taas kunnolla	vihaisena	puraisee niin sitten joi vo itseltä kysyä, mikä sai kanin niin pahalle tuulelle, ja lyödä vaikkapa itseään
a menee vaisuksi, maukuu hiljaa, kellii lattialla, purra natustaa ihmisten käsiä jos sitä silitetään (ei siis ole	vihainen) ja komentaa tavallista topakemmin meidän 7 kk ikäistä kollikkisaa (ei leikattu).
No eihän sulle voi olla	vihainen	.
Olet hyvin aggressiivinen,	vihainen	ja itsekeskeinen.
Edessä hämmötti iso lasi-ikkuna ja takana iso ja	vihainen	voimalapäällikkö.
Katkera ja	vihainen	ihminen tekee pahaa jälkeä sekä kertoo ehkä sen vielä kaikille.
näille telepaatikoillekaan ei siis kannata olla	vihainen	, puhumattakaan, että heitä kannattaisi tuomita, sillä välitön ensiapu ihmiselle tulee jo siitä, että ilm
Voin olla harmissani tilanteesta, mutta hänelle en voi olla	vihainen	siksi, ettei hän tehnyt niin kuin MINÄ halusin.
Se oli todella kauhuissaan ja	vihainen	.
Tuli suuri tuulen puuska, iästä	vihainen	ilma; meren kuohuille kohotti, lainehille laikahutti.
Mutta rankaisun jälkeen pitää osoittaa ettei ole enää	vihainen	...
Niiden sisällä kituu	vihaisia	pulisonkipappoja liikennettä ja omaa turvallisuuttaan vaarantamassa.
Niiden sisällä kituu	vihaisia	pulisonkipappoja liikennettä ja omaa turvallisuuttaan vaarantamassa.
Sä varmaan vaan halut nähdä kuinkapaljon saat kannanottoja, jos saat tyydytyksen siitä, että voit lukee	vihaisia	vastauksia!!

Korpus

Suomi24 (2017H2) 2001 (beta)

Kuvailutiedot

Lisenssi: CC BY-NC (CLARIN PUB)

Viiittaa korpuksen

Linkki korpuksen Korpissa: [to be added]

Tehtävien piirteet

otsikko: Lapsen kurittaminen

päiväys: 2001-08-24

kellonaika: 15:37:00

kirjoittajan nimimerkki: No Mercy?

kirjautunut käyttäjä: ei

rekisteröity nimimerkki: ei

aihealue: Ryhmät > Ikäryhmät > 30 plus

aihealue vain aikuisille: ei

viestin tyyppi: kommentti

viesti on täysin tyhjä: ei

tunniste: 7540:33239

keskusteluketjun tunniste: 7540

keskusteluketjun alkuaikaleima:

2001-07-25 15:38:00

kommentin tunniste: 33239

kommentin vanhemman tunniste: 33224

kommentin vanhemman aikaleima:

2001-08-24 13:14:00

lainatun kommentin tunniste: 33238

tiedoston nimi: s24_2001_02.vrt

kappaleen tyyppi: kappale



Download concordance

miehestä tarinoiva rouva näytti siltä kuin olisi aikonut sanoa siihen jotain mutta päästikin sen sijaan vain	vihaisen	tuhauksen ja lamauttavan katseen.
Vieraista eivät perusta, eivät ole	vihaisia	vaan välinpitämättömiä.
Tai siis en mä tiä onx mun bestis mulle sillee "	vihainen	" mun ja mun jätkän takia...
Kateellinen,	vihainen	, närkeystynyt meidän (sinun kuvittelemiisi) valintoihimme?
Olen	vihainen	yhiteiskunnalle, joka pakottaa molemmat vanhemmat käymään pitkiä päiviä töissä.
jos se taas kunnolla	vihaisena	puraisee niin sitten jo voi itseltä kysyä, mikä sai kanin niin pahalle tuulelle, ja lyödä vaikkapa itseään
a menee vaisuksi, maukuu hiljaa, kellii lattialla, purra natustaa ihmisten käsiä jos sitä silitetään (ei siis ole	vihainen) ja kommentaa tavallista topakammin meidän 7 kk ikäistä kollikissaa (ei leikattu).
No eihän sulle voi olla	vihainen	.
Olet hyvin agressiivinen,	vihainen	ja itsekeskeinen.
Edessä hämmötti iso lasi-ikkuna ja takana iso ja	vihainen	voimalapäällikkö.
Katkeria ja	vihainen	ihminen tekee pahaa jälkeä sekä kertoo ehkä sen vielä kaikille.
näille telepaatikoillekaan ei siis kannata olla	vihainen	, puhumattakaan, että heitä kannattaisi tuomita, sillä välitön ensiupu ihmiselle tulee jo siitä, että ilm
Voin olla harmissani tilanteesta, mutta hänelle en voi olla	vihainen	siksi, ettei hän tehnyt niin kuin MINÄ halusin.
Se oli todella kauhuissaan ja	vihainen	.
Tuli suuri tuulen puuska, lästä	vihainen	ilma; meren kuohuille kohotti, lainehille laikahutti.
Mutta rankaisun jälkeen pitää osoittaa ettei ole enää	vihainen	...
Niiden sisällä kituu	vihaisia	pulisonkipappoja liikennettä ja omaa turvallisuuttaan vaarantamassa.
Niiden sisällä kituu	vihaisia	pulisonkipappoja liikennettä ja omaa turvallisuuttaan vaarantamassa.
Sä varmaan vaan halut nähdä kuinkapaljon saat kannanottoja, jos saat tyydytyksen siitä, että voit lukee	vihaisia	vastauksia!!!
ouse pois koiran tasolta, mene vaikka pöydän ääreen istumaan niin ettei koira pääse sinuun käsiksi, ja ole	vihaisen	oloinen, tiukka ja vakava.
mutta toisaalta ei kala ihmiselle pysty olemaan	vihainen	..niin hellyyttävä hän on.
Eräskin opettaja sanoi, että hän ei tiedä asiasta mitään, kerran se erityisope sanoi vain, että olet niin	vihaisen	näköinen, että minä hoidan nämä asiat ilman sinua.
Tänään sitten kolmen korvilla poika soitti töihin,	vihaisena	.
Mielialani vaihtelee tunteellisuuteni takia, jos näen töissäni hakutan naisen--> olen surullinen ja	vihainen	, jos näen taas nuoren, jonka asiat ovat vihdoinkin kunnossa--> iloinen ..
Tämän ei ole tarkoitus herättää ihmisissä	vihaisia	mielipiteitä, vaan auttaa Johanna aukaisemaan silmänsä omille kirjoituksilleen...
Ollaanko tyytymättömiä kumppanin seksitaitoihin tai vain	vihaisia	hänelle muusta syystä ja kostetaan?
Eräskin opettaja sanoi, että hän ei tiedä asiasta mitään, kerran se erityisope sanoi vain, että olet niin	vihaisen	näköinen, että minä hoidan nämä asiat ilman sinua.
Isä sanoi	vihaisena	, ettei ole rahaa ostaa jäätelöä ja sai lapsen itkemään.
Tyypillistä on että hän saattaa herätä lauantiaamuna todella masentuneena ja	vihaisen	, ja yksilöiden valittaa jokaisesta yksittäisestä pölyhiukkasestakin, mitä löytää, puhumattakaan sitte
Sinä laittoi siihen erittäin	vihaisen	vastineen.
Ostin myös varuilla pari	vihaista	dobermannia, joita pidän yöt auton sisällä.
välillä se koettelee hermojasi, mutta takaan ettet voi olla sille kauaa	vihainen	.. oma koirani " halasi " ikäänkuin anteeksipyyynnöksi ..
Kyse ei siis ollut vihaisesta Jumalasta, joka olisi pitänyt lepyttää, vaan kyse oli	Vihaista	koiraahan ei puistoon saa viedä.
Onko se	vihaisesta	ihmisestä, jonka Jumala tahtoi lepyttää.
Ja kun kerroin boksu hyökkää nopeasti leikkimään ja toinen voi ottaa sen	vihainen	tai pelokas?
Siksi näin on, Äiti ei jaksa enää ja on	vihaisena	hyökkäyksenä.
uksia kuten meissäkin, joku poika- rotta saataa olla kovin vilkas ja tyttö rauhallinen, jokunen rotta jopa	vihainen	.
Hän puhuu todella halveksuvaan ja	vihainenkin.	sävyyen, kuin pahaiselle kakaralle.
Vaimo on aina vaan enemmän	vihaisen	ja kakarat muistuttavat aina vaan enemmän napurin setää.
Oletteko perehtyneet koirien kieleen?), jolloin omistaja saa kuulla kunniansa	vihaisesta	koirasta, kun se todellisuudessa vain tekee niinkuin vaistot ohjaavat.
Olin ehtinyt lähettää hänelle joulun alla aika	vihaisen	kortin, että lyhennä nyt velkaasi.
Osoitit että puppelpoikakin voi olla	vihainen	, ja äityä oikein väkivaltaiseksikin!

Korpus

Suomi24 (2017H2) 2001 (beta)

Kuvailutiedot

Lisenssi: CC BY-NC (CLARIN PUB)

Viittaa korpukseen

Linkki korpukseen Korpissa: [to be added]

Tekstin piirteet

otsikko: Lapsen kurittaminen

päiväys: 2001-08-24

kellonaika: 15:37:00

kirjoittajan nimimerkki: No Mercy?

kirjautunut käyttäjä: ei

rekisteröity nimimerkki: ei

aihealue: Ryhmät > Ikäryhmät > 30 plus

aihealue vain aikuisille: ei

viestin tyyppi: kommentti

viesti on täysin tyhjä: ei

tunniste: 7540:33239

keskusteluketjun tunniste: 7540

keskusteluketjun alkuaikaleima:

2001-07-25 15:38:00

kommentin tunniste: 33239

kommentin vanhemman tunniste: 33224

kommentin vanhemman aikaleima:

2001-08-24 13:14:00

lainatun kommentin tunniste: 33238

tiedoston nimi: s24_2001_02.vrt

kappaleen tyyppi: kappale

Sanan piirteet

perusmuoto: vihainen

perusmuoto (yhdyssanarajat): vihainen

sanaluokka: adjektiivi

morfologinen analyysi: NUM_P|CASE_Par|

CMP_Pos

dependenssisuhde: pää

saneeseen liittyvät tyhjämerkit: _

Näytä alkuperäinen keskusteluketju

Näytä alkuperäinen viesti

Näytä dependenssipuu

Olet hyvin aggressiivinen, **vihainen** ja itse
Edessä häämötti iso lasi-ikkuna ja takana iso ja **vihainen** voima
Katkera ja **vihainen** ihmin
näille telepaatikoillekaan ei siis kannata olla **vihainen**, puhu
Voin olla harmissani tilanteesta, mutta hänelle en voi olla **vihainen** siksi,
Se oli todella kauhuissaan ja **vihainen**.
Tuli suuri tuulen puuska, iästä **vihainen** ilma;
Mutta rankaisun jälkeen pitää osoittaa ettei ole enää **vihainen** ...
Niiden sisällä kituu **vihaisia** pulisc
Niiden sisällä kituu **vihaisia** pulisc
on saat kannanottoja, jos saat tyydytyksen siitä, että voit lukee **vihaisia** vasta
dän ääreen istumaan niin ettei koira pääse sinuun käsiksi, ja ole **vihaisen** oloine
mutta toisaalta ei kala ihmiselle pysty olemaan **vihainen** ..niin l
rdä asiasta mitään, kerran se erityisope sanoi vain, että olet niin **vihaisen** näköi
Tänään sitten kolmen korvilla poika soitti töihini, **vihaisena** .
eni takia, jos näen töissäni hakatun naisen--> olen surullinen ja **vihainen**, jos nä
Tämän ei ole tarkoitus herättää ihmisissä **vihaisia** mielip
Ollaanko tyytymättömiä kumppanin seksitaitoihin tai vain **vihaisia** hänel
rdä asiasta mitään, kerran se erityisope sanoi vain, että olet niin **vihaisen** näköi
Isä sanoi **vihaisena**, ettei ole rahaa ostaa jäätelöä ja sai lapsen itkemään.
tä hän saattaa herätä lauantaiaamuna todella masentuneena ja **vihaisena**, ja yksilöiden valittaa jokaisesta yksittäisestä pölyhiukkasestakin, mitä löytää, puhumattakaan sitten pöydä
Sinä laitoit siihen erittäin **vihaisen** vastineen.
Ostin myös varuilla pari **vihaista** dobermannia, joita pidän yöt auton sisällä.
ä se koettelee hermojasi, mutta takaan ettet voi olla sille kauaa **vihainen** .. oma koirani " halasi " ikäänkuin anteeksipyyynnöksi ..
Vihaista koiraahan ei puistoon saa viedä.
ut vihaisesta Jumalasta, joka olisi pitänyt lepyttää, vaan kyse oli **vihaisesta** ihmisestä, jonka Jumala tahtoi lepyttää.
Onko se **vihainen** tai pelokas?
roin boksu hyökkää nopeasti leikkimään ja toinen voi ottaa sen **vihaisena** hyökkäyksenä.
Siksi näin on, Äiti ei jaksa enää ja on **Vihainen** .
saataa olla kovin vilkas ja tyttö rauhalinen, jokunen rotta jopa **vihainenkin**.
Hän puhuu todella halveksuvaan ja **vihaiseen** sävyyn, kuin pahaiselle kakaralle.
Vaimo on aina vaan enemmän **vihainen** ja kakarat muistuttavat aina vaan enemmän napurin setää.
htyneet koirien kieleen?), jolloin omistaja saa kuulla kunniansa **vihaisesta** koirasta, kun se todellisuudessa vain tekee niinkuin vaistot ohjaavat.
Olin ehtinyt lähettää hänelle joulun alla aika **vihaisen** kortin, että lyhennä nyt velkaasi.

Opening korp_kwic_vihainen_jj_1_20190225_105846.xls



You have chosen to open:



korp_kwic_vihainen_jj_1_20190225_105846.xls

which is: Microsoft Excel 97-2003 Worksheet (49.5 KB)

from: <https://korp.csc.fi>

What should Firefox do with this file?

Open with

Microsoft Excel (default)

Save File

Do this automatically for files like this from now on.

OK

Cancel

PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	hit number	corpus	tokens	lemmas	text_title	text_date	text_time	text_autho	text_autho	text_autho	text_topic	text_topic	text_topic	text_topic	text_msg
2	0	S24_2001	Siten kun	l siten kun k	Lapsen ku	2001-08-2	15:37:00	No Mercy?	n	n	Ryhmät >	Ryhmät	30 plus	n	comment
3	1	S24_2001	Sinä olet il	sinä olla ilr	Ei-kristittyj	2001-10-2	16:04:00	Keppo	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	comment
4	2	S24_2001	Eli taas sa	eli taas sai	Järjen käy	2001-09-1	00:14:00	Lapintyttö	n	n	Harrastuks	Harrastuks	Yleistä me	n	comment
5	3	S24_2001	Ei Kaizun	ei Kaizun l	Viikon vira	2001-06-2	11:21:00	Pakana(M	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	comment
6	4	S24_2001	- Ei tuo ihr	- ei tuo ihn	Tuo outo	2001-06-2	16:02:00	halaajanai	n	n	Ryhmät >	Ryhmät	40 plus	n	comment
7	5	S24_2001	kerro tosia	kertoa tosi	Mikko on e	2001-03-2	18:07:00	selvennä a	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	comment
8	6	S24_2001	Minua kam	minä kamr	Norsunluu	2001-05-1	03:42:00	kapinoitsije	n	n	Tiede ja te	Tiede ja te	Yleistä tiet	n	comment
9	7	S24_2001	Naimin mie	naida mies	EclipseE:lle	2001-06-2	22:19:00	Äijä -85	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	comment
10	8	S24_2001	Vieraista e	vieras ei p	Saluki	2001-04-0	15:54:00	Minkku	n	n	Lemmikit >	Lemmikit	Yleistä koi	n	comment
11	9	S24_2001	Tai siis en	tai siis ei n	Bestiksen	2001-07-0	12:10:00	GiRL87	n	n	Nuoret > F	Nuoret	Rakkaus	n	comment
12	10	S24_2001	Kateelliner	kateellinen	vastauksia	2001-11-2	22:44:00	putte	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	comment
13	11	S24_2001	Olen vihair	olla vihaine	mitä mieltä	2001-02-0	23:29:00	Äityli	n	n	Lemmikit >	Lemmikit	Yleistä koi	n	comment
14	12	S24_2001	jos se taas	jos se taas	hei sinä ke	2001-09-1	15:32:00	v	n	n	Lemmikit >	Lemmikit	Yleistä jyr	n	comment
15	13	S24_2001	Meillä on k	minä olla k	Leikatun k	2001-08-1	21:53:00	Jippo	n	n	Lemmikit >	Lemmikit	Yleistä kis	n	thread_s
16	14	S24_2001	No eihän s	no ei sä vc	USA saa s	2001-10-1	13:49:00	Ripa	n	n	Yhteiskunt	Yhteiskunt	Maailman i	n	comment
17	15	S24_2001	Olet hyvin	olla hyvin	kahvissa o	2001-09-3	00:26:00	Jolanta(ka	n	n	Ajanviete >	Ajanviete	Kahvikesku	n	comment
18	16	S24_2001	Edessä hä	edessä hä	tulin koulu	2001-08-0	10:02:00	vekanen	n	n	Ryhmät >	Ryhmät	40 plus	n	comment
19	17	S24_2001	Katkerä ja	katkerä ja	Eräs pettäj	2001-05-0	16:00:00	minulta.	n	n	Suhteet >	Suhteet	Lapset ja p	n	comment
20	18	S24_2001	näille telep	tämä telep	sinäkö mu	2001-12-2	00:43:00	pilot of imp	n	n	Yhteiskunt	Yhteiskunt	Yleistä raja	n	comment



corp_kwic_vihainen_jj_1_20190225_110003 [Read-Only] [Compatibility Mode]

File Home Insert Page Layout Formulas Data Review View PDF-XChange V6 Tell me what you want to do...

Cut Copy Paste Format Painter Clipboard

Arial 10 A A Font

Wrap Text Alignment Merge & Center

General Number Conditional Formatting Table

Normal Neutral

C1 tokens

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	hit number	corpus	tokens	lemmas	text_title	text_date	text_time	text_author	text_author	text_author	text_topic	text_topic	text_topic	text_topic	text
2	0	S24_2001	Siten kun	siten kun k	Lapsen ku	2001-08-2	15:37:00	No Mercy:	n	n	Ryhmät >	Ryhmät	30 plus	n	com
3	1	S24_2001	Sinä olet il	sinä olla ilr	Ei-kristittyj	2001-10-2	16:04:00	Keppo	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	com
4	2	S24_2001	Eli taas sa	eli taas sa	Järjen käy	2001-09-1	00:14:00	Lapintyttö	n	n	Harrastuks	Harrastuks	Yleistä me	n	com
5	3	S24_2001	Ei Kaizun	ei Kaizun l	Viikon vira	2001-06-2	11:21:00	Pakana(M	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	com
6	4	S24_2001	- Ei tuo ihr	- ei tuo ihr	Tuo outo	2001-06-2	16:02:00	halaajanai	n	n	Ryhmät >	Ryhmät	40 plus	n	com
7	5	S24_2001	kerro tosia	kertoa tosi	Mikko on e	2001-03-2	18:07:00	selvennä a	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	com
8	6	S24_2001	Minua kamr	minä kamr	Norsunлуу	2001-05-1	03:42:00	kapinoitsije	n	n	Tiede ja te	Tiede ja te	Yleistä tiet	n	com
9	7	S24_2001	Naimin mie	naida mies	Eclipse:lle	2001-06-2	22:19:00	Äijä -85	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	com
10	8	S24_2001	Vieraista e	vieras ei p	Saluki	2001-04-0	15:54:00	Minkku	n	n	Lemmikit >	Lemmikit	Yleistä koi	n	com
11	9	S24_2001	Tai siis en	tai siis ei n	Bestiksen	2001-07-0	12:10:00	GiRL87	n	n	Nuoret > F	Nuoret	Rakkaus	n	com
12	10	S24_2001	Kateellinen	kateellinen	vastauksia	2001-11-2	22:44:00	putte	n	n	Yhteiskunt	Yhteiskunt	Yleistä usk	n	com
13	11	S24_2001	Olen vihain	olla vihaine	mitä mieltä	2001-02-0	23:29:00	Äityli	n	n	Lemmikit >	Lemmikit	Yleistä koi	n	com
14	12	S24_2001	jos se taas	jos se taas	hei sinä ka	2001-09-1	15:32:00	v	n	n	Lemmikit >	Lemmikit	Yleistä jyr	n	com
15	13	S24_2001	Meillä on k	minä olla k	Leikatun ki	2001-08-1	21:53:00	Jippo	n	n	Lemmikit >	Lemmikit	Yleistä kis	n	thre
16	14	S24_2001	No eihän s	no ei sä vc	USA saa s	2001-10-1	13:49:00	Ripa	n	n	Yhteiskunt	Yhteiskunt	Maailman	n	com
17	15	S24_2001	Olet hyvin	olla hyvin	kahvissa o	2001-09-3	00:26:00	Jolanta(ka	n	n	Ajanviete >	Ajanviete	Kahvikesku	n	com



KIELIPANKKI KÄYTTÄJÄKSI AINEISTOT TYÖKALUT ORGANISAATIO TUKI IN ENGLISH PÅ SVENSKA

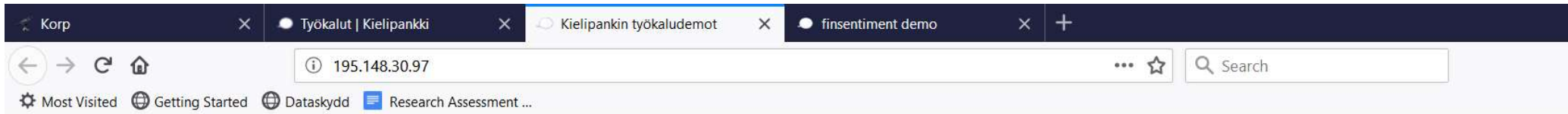
Työkalut

Käynnistä työkalu	Nimi	Kuvaus	Ohjeet	Asentaminen	Ylläpitäjä
	Korp	Selainpohjainen konkordanssityökalu, jolla voi tehdä hakuja tekstiaineistoista morfosyntaktisen jäsennyksen perusteella.	Ohjeet		
	LAT (Language Archive Tools)	Annotoitujen ääni- ja videoaineistojen selailun, hakujen ja jakamisen työkalupakki.	Ohjeet		
Lataus	Latauspalvelu	Lataa aineistoja omalle koneellesi.			
	META-SHARE	Kielipankin kiellivarojen kuvailutietovaranto.			

www.kielipankki.fi/tools > Demo

Start	Name	Description	Instructions	Install	Administrator
	Korp	A web-based concordance tool that can be used for corpus queries based on morphosyntactic analysis and various other features.	Instructions		
	LAT (Language Archive Tools)	A toolkit for browsing and querying annotated speech and video corpora.	Instructions		
Download	Download service	Download certain corpora.			
	META-SHARE	Metadata repository of all the language resources at the Language Bank of Finland.			
Myly	Myly	Versatile data analysis platform with interactive visualizations and workflows.	Instructions		
Sanat	Sanat	A platform for publishing lexica and word lists.			
FinTag	Finnish Tagtools	A part-of-speech and morphology tagger and a named entity recogniser for Finnish.		Install	
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)			
	WebAnno	Text annotation tool.			

finsentiment (sentiment analysis)



Kielipankin työkaludemot

Tällä sivustolla on kokeiltavissa erilaisia työkaluja ja resursseja joita jalostetaan virallisempaan käyttöön. Virheraportteja, kehitysehdotuksia ja avunpyyntöjä voi lähettää osoitteeseen sam.hardwick@iki.fi

Monikielinen tekstintunnistus

Eristä teksti kuvatiedostoista (jpg, png, pdf, ...)

fintag

Tekstin perusmuotoistus, taivutuksen ja nimettyjen ilmausten tunnistus

finsentiment

Tekstin sentimenttianalyysi

lemmarank

Tekstille leimallisten sanojen ja nimettyjen ilmausten löytäminen ja vertailu

lemmamatch

Tekstihaku perusmuotojen ja sanasemantiikan avulla

- Tieteen termipankkin sanalistoihin perustuva esimerkki
- Sanasemantiikkaesimerkki
- Esimerkki sääntöjen kirjoittamisesta

Latauspalvelu

Kehitysvaiheessa olevien aineistojen (sanavektorit) latauspalvelu

fintok

like-unlike (beta)

Korp x Työkalut | Kielipankki x finsentiment demo x finsentiment demo x +

← → ↻ 🏠 ⓘ 195.148.30.97/cgi-bin/predict/finsent.py ⋮ ☆ 🔍 Search

⚙ Most Visited 🌐 Getting Started 🌐 Dataskydd 📄 Research Assessment ...

finsentiment demo

Split running text into tokens.

[Show help](#)

Running text goes here.

Or populate with demo text

Or No file selected.

Page generated in 0.00 seconds

Korp x Työkalut | Kielipankki x finsentiment demo x finsentiment demo x +

← → ↻ 🏠 ⓘ 195.148.30.97/cgi-bin/predict/finsent.py ... ☆ 🔍 Search

⚙️ Most Visited 🌐 Getting Started 🌐 Dataskydd 📄 Research Assessment ...

finsentiment demo

Split running text into tokens.

[Show help](#)

kakaralle .
Vaimo on aina vaan enemmän vihainen ja kakarat muistuttavat aina vaan enemmän napurin setää .
Oletteko perehtyneet koirien kieleen ?) , jolloin omistaja saa kuulla kunniansa vihaisesta koirasta , kun se todellisuudessa vain tekee niinkuin vaistot ohjaavat .
Olin ehtinyt lähettää hänelle joulun alla aika vihaisen kortin , että lyhennä nyt velkaasi .
Osoitit että puppelipoikakin voi olla vihainen , ja äityä oikein väkivaltaiseksikin !

Or populate with demo text

Or No file selected.

Processing time is 0.00 seconds

positive	Itselläni on molempia ja kaikki ovat eri persoonia , tottakai on myös poikkeuksia kuten meissäkin , joku poika- rotti saataa olla kovin vilkasin ja tyttö rauhalinen , jokunen rotti jopa vihainenkin .
negative	Hän puhuu todella halveksuvaan ja vihaiseen sävyyn , kuin pahaiselle kakaralle .
neutral	Vaimo on aina vaan enemmän vihainen ja kakarat muistuttavat aina vaan enemmän napurin setää .
positive	Oletteko perehtyneet koirien kieleen ?) , jolloin omistaja saa kuulla kunniansa vihaisesta koirasta , kun se todellisuudessa vain tekee niinkuin vaistot ohjaavat .
negative	Olin ehtinyt lähettää hänelle joulun alla aika vihaisen kortin , että lyhennä nyt velkaasi .
neutral	Osoitit että puppelipoikakin voi olla vihainen , ja äityä oikein väkivaltaiseksikin !
negative	tokens Siten kun lapsesi huomaavat olevansa vihaisia ja katkeroituneita sinua kohtaan , he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi kipeäksi ja sitten saat luultavasti ehdollista , milloin et saa koskea kenéenkään pitkällä tikullakaan , Näin minä tein kun fucistä kekkemättömyuttani viimeksi laukattiin ja niin tekevät

finsentiment demo

Split running text into tokens.

[Show help](#)

tokens
Siten kun lapsesi huomaavat olevansa vihaisia ja katkeroituneita sinua kohtaan , he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi ki... ja sitten saat luultavasti ehdollista , milloin et saa koskea kenéenkään pi... tikullakaan , Näin minä tein kun fyysistä koskemattomuuttani viimeksi loukattiin , ja niin tekevät sinunkin lapsesi.Poliisilla ei oikein riitä ymmär... piiskaleikkejäso kohtaan ...
Sinä olet ilmeisesti uskova koska olet minulle vihainen ?
Eli taas sarjassamme " yritänpä v.....ssa saada porukan vihaiseksi " .
Ei Kaizun kiemurtelu nyt kovin vihaiselta vaikuta .
- Fi tuo ihminen ole välttämättä katkera satuttaialleen tai ei hän koe ole

Or populate with demo text Or **Browse...** No file selected.

Submit

Result:

Download TSV **Download Excel spreadsheet**

Sentiment	Text
positive	tokens
negative	Siten kun lapsesi huomaavat olevansa vihaisia ja katkeroituneita sinua kohtaan , he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi kipeäksi ja sitten saat luultavasti ehdollista , milloin et saa koskea kenéenkään pitkällä tikullakaan , Näin minä tein kun fyysistä koskemattomuuttani viimeksi loukattiin , ja niin tekevät

Opening finsentiment_result.xlsx

You have chosen to open:

finsentiment_result.xlsx
which is: Microsoft Excel Worksheet (8.7 KB)
from: http://195.148.30.97

What should Firefox do with this file?

Open with Microsoft Excel (default)

Save File

Do this automatically for files like this from now on.

OK **Cancel**

Continue analysis manually in Excel

finsentiment_result [Protected View] - Excel

File Home Insert Page Layout Formulas Data Review View PDF-XChange V6 Tell me what you want to do...

PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. [Enable Editing](#)

A1

	A	B	C	D	E	F	G	H	I	J
1	positive	tokens								
2	negative	Siten kun lapsesi huomaavat olevansa vihaisia ja katkeroituneita sinua kohtaan , he tekevät rikosilmoituksen ja sinä saat ensin maksaa itsesi kipeä								
3	negative	Sinä olet ilmeisesti uskova koska olet minulle vihainen ?								
4	neutral	Eli taas sarjassamme " yritänpä v.ssa saada porukan vihaiseksi " .								
5	negative	Ei Kaizun kiemurtelu nyt kovin vihaiselta vaikuta .								
6	neutral	- Ei tuo ihminen ole välttämättä katkera satuttajalleen tai ei hän koe olevansa vihainen kenellekään ; siis anteeksi antamisen tarpeessa .								
7	negative	kerro tosiaan noista perusteluistasi selvemmin , muuten me muut emme ollenkaan tajua miksi olet niin närkästynyt ja vihainen .								
8	neutral	Minua kammottaa , tulen tovottoman vihaiseksi .								
9	neutral	Naimin miehestä tarinoiva rouva näytti siltä kuin olisi aikonut sanoa siihen jotain mutta päästikin sen sijaan vain vihaisen tuhuksen ja lamauttava								
10	negative	Vieraista eivät perusta , eivät ole vihaisia vaan välinpitämättömiä .								
11	negative	Tai siis en mä tiä onx mun bestis mulle sillee " vihainen " mun ja mun jätkän takia ...								
12	neutral	Kateellinen , vihainen , närkästynyt meidän (sinun kuvittelemiisi) valintoihimme ?								
13	neutral	Olen vihainen yhteiskunnalle , joka pakottaa molemmat vanhemmat käymään pitkiä päiviä töissä .								
14	negative	jos se taas kunnolla vihaisena puraisee niin sitten jo voi itseltä kysyä , mikä sai kanin niin pahalle tuulelle , ja lyödä vaikkapa itseään , kun vihdoin t								
15	negative	Meillä on keväällä steriloitu tyttökissa , jolla on nyt muutaman viikon aikana ollut 8-10 päivän välein outoa käyttäytymistä , siis kissa menee vaisu								
16	negative	No eihän sulle voi olla vihainen .								
17	negative	Olet hyvin aggressiivinen , vihainen ja itsekeskeinen .								
18	negative	Edessä hämmötti iso lasi-ikkuna ja takana iso ja vihainen voimalapäällikkö .								
19	positive	Katkera ja vihainen ihminen tekee pahaa iälkeä sekä kertoo ehkä sen vielä kaikille								



Researcher of the Month: Katja Västi



photo: Valtteri Airaksinen

Kielipankki – The Language Bank of Finland is a service for researchers using language resources. Katja Västi, lecturer of Finnish Language at the University of Oulu, tells us about her research on the [Finnish Text Collection](#).

Who are you?

I'm Katja Västi, lecturer of Finnish Language at the University of Oulu.

What is your research topic?

In my doctoral dissertation I discuss the semantic and syntactic properties of Finnish verbless expressions in the light of two examples, the SOURCE construction (e.g. Oikeusasiamieheltä huomautus, lit. 'From the ombudsman a complaint') and the GOAL construction (e.g. Pakistanille rangaistus lit. 'For Pakistan a punishment'). These constructions express dynamic events in a way comparable to finite constructions. However, they are not elliptic structures but genuinely verbless con-

Search the Language Bank

Portal:

Haku:

Hae



Tweets by @FinClarín

Kielipankki – The
Language Bank of
Finland



Katja Västi

Researcher of the Month 7/2017

Lecturer of Finnish Language at the University of Oulu

Research interests:


- Verbless constructions in Finnish and their semantic and syntactic properties
- Examples
 - SOURCE construction (e.g. "Oikeusasiamieheltä huomautus", 'From the ombudsman a complaint')
 - GOAL construction (e.g. "Pakistanille rangaistus", 'For Pakistan a punishment')

Tool and method

- In 2012, Katja Västi could use the Finnish Text Collection (FTC) via the old concordancing tool *WWW-Lemmie 2.0*
- Now, the corpus is called **The Helsinki Korp Version of the Finnish Text Collection** (ftc-korp), and it is available via *korp.csc.fi*, <http://urn.fi/urn:nbn:fi:lb-2016050207>



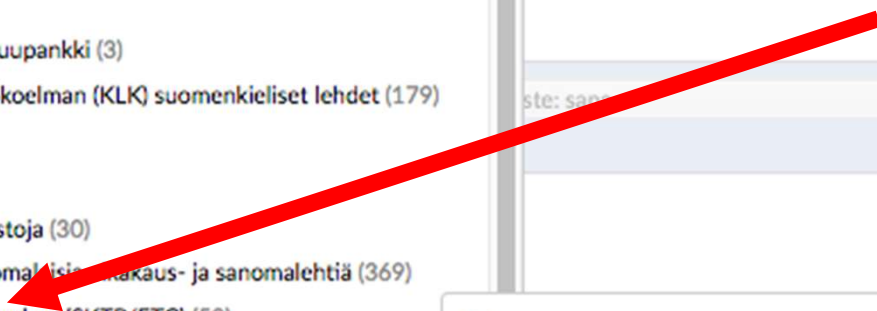
0 / 949 korpusta valittuina — 0 / 12,87G sanetta



Valitse kaikki
 Tyhjennä

- Agricola (beta) (9)
- E-thesis (24)
- FinnTreeBank: suomen puupankki (3)
- Kansalliskirjaston lehtikokoelman (KLK) suomenkieliset lehdet (179)
- Kirjallisuutta (21)
- Juridisia tekstejä (3)
- Internet-keskusteluaineistoja (30)
- 1990- ja 2000-luvun suomalaisia viikkokäytäntö- ja sanomalehtiä (369)
- Suomen kielen tekstikokoelma (SKTP/FTC) (50)
 - Lehdet (49)
 - Kustannusosakeyhtiö Otava 1993 (SKTP) [RES]
 - Muita tekstejä (11)
 - Puhuttua kieltä (tekstiksi litteroituna) (137)
 - Suomenoppijoiden kieltä (suomi toisena tai vieraana kielenä) (4)
 - Vanhan kirjasuomen korpus (12)
 - Varhaisnykysuomen korpus (93)
 - Demo- ja testiaineistoja (1)
- Karjalansuomi [ACA]
- Paul Sinebrychoffin kirjeenvaihto
- Eduskunta - transkriptiot

corp.csc.fi: FTC
(locked!)



 Suomen kielen tekstikokoelma (SKTP/FTC)

Suomen kielen tekstikokoelma: Lemmie-palvelussa olleet osakorpuksat

Kuvailutiedot
Lisenssi: CLARIN RES +PLAN +NC +ND
Viittaa korpukseen
Linkki korpukseen Korpissa: urn:nbn:fi:lb-2014052719
Aineiston kuvaus

50 korpusta, joissa:

144 117 021 sanetta
13 037 217 virkettä



Log in to Korp

The screenshot shows the Korp website interface. At the top, there is a browser window with the URL [https://korp.csc.fi/#?stats_reduce=word&cqp=\[\]&lang=en](https://korp.csc.fi/#?stats_reduce=word&cqp=[]&lang=en). The page header includes language options: Finnish | Swedish | Other languages | Parallel, and a navigation menu with items like Cite Korp, Suomi, Svenska, English, and a notification bell with '27'. The main content area features the Korp logo (a crow) and a search bar with the text '0 of 954 corpora selected – 0 of 12.93G tokens'. Below the search bar are tabs for 'Simple', 'Extended', 'Advanced', and 'Compare'. A search input field is present with a 'Search' button. Below the search bar, there are checkboxes for 'initial part', 'final part', and 'case-insensitive', and a dropdown menu set to 'sentences'. At the bottom, there are settings for 'KWIC: hits per page: 25', 'sort within corpora: not sorted', and 'Statistics: compile based on: word'. A 'Log in' button is located in the bottom right corner of the main content area. A red circle highlights the 'Log in' button in the navigation menu on the right side of the page. The 'KIELIPANKKI' logo and 'The Language Bank of Finland' text are also visible in the navigation menu area. A 'CORE TRUST SEAL' logo is located at the bottom right of the page.

Log in to Korp



The Language Bank of Finland

Korp

Choose your organization or federation.

Login with:

University of Helsinki

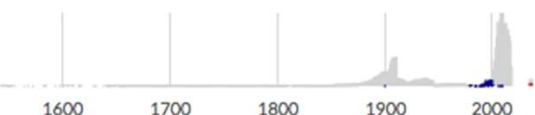
Login

Narrow down the list by typing any part of your organisation.





50 / 949 korpusta valittuina – 144,12M / 12,87G sanetta



✓ Valitse kaikki

✗ Tyhjennä

Yksinkertainen Laaje

myös alkuosa loppuos

virkkeistä jotka sisä

Konkordanssi: osumia sivulla

- Agricola (beta) (9)
- E-thesis (24)
- FinnTreeBank: suomen puupankki (3)
- Kansalliskirjaston lehtikokoelman (KLIK) suomenkieliset lehdet (179)
- Kirjallisuutta (21)
- Juridisia tekstejä (3)
- Internet-keskusteluaineistoja (30)
- 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä (369)
- Suomen kielen tekstikokoelma (SKTP/FTC) (50)
- Muita tekstejä (11)
- Puhuttua kieltä (tekstiksi litteroituna) (137)
- Suomenoppijoiden kieltä (suomi toisena tai vieraana kielenä) (4)
- Vanhan kirjasuomen korpus (12)
- Varhaisnykysuomen korpus (93)
- Demo- ja testiaineistoja (1)
- Karjalansuomi [ACA]
- Paul Sinebrychoffin kirjeenvaihto
- Eduskunta - transkriptiot

13 037 217 virkettä valituissa korpuksissa

sana Näytä sanakuva Näytä kartta

Suomen kielen tekstikokoelma (SKTP/FTC)

Suomen kielen tekstikokoelma: Lemmie-palvelussa olleet osakorpuksset

Kuvailutiedot

Lisenssi: CLARIN RES +PLAN +NC +ND

Viittaa korpukseseen

Linkki korpukseseen Korpissa: urn:nbn:fi:lb-2014052719

Aineiston kuvaus

50 korpusta, joissa:

144 117 021 sanetta

13 037 217 virkettä

How to search for sentences (max 11 words) without a verb

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit



50 / 949 korpusta valittuina – 144,12M / 12,87G sanetta



Yksinkertainen

Laajennettu

Edistynyt

Vertailu

sanaluokka ei ole
verbi
tai
virkkeen alussa ✕

sanaluokka ei ole
verbi
tai
toista - kertaa

sanaluokka ei ole
verbi
tai
virkkeen lopussa ✕

Hae sisältä

Any sentence-initial word whose part of speech is **not** a verb

Any word whose part of speech is **not** a verb, may repeat 0-9 times

Any sentence-final word whose part of speech is **not** a verb

Yksinkertainen

Laajennettu

Edistynyt

Vertailu

sanaluokka ei ole

verbi

tai

virkkeen alussa

sanaluokka ei ole

verbi

tai

toista - kertaa

sanaluokka ei ole

verbi

tai

virkkeen lopussa

Hae

virkkeen sisältä

Konkordanssi: osumia sivulla: 25

järjestä korpuksen sisällä: järjestämätön

Tilastoja: laskentaperuste: sana

 Näytä sanakuva Näytä kartta

Konkordanssi

Tilastoja

Sanakuva

Kartta

Tuloksia: 2 202 168

« ‹ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... › » Näytä konteksti

AAMULEHTI 1995 (SKTP)

Kaikki pankkiasiakkaat riskiluokkiin
SIVU 18
SIVU 14
Kalankäsittelyn mestari.
Nimismiehet Virkkilpalluun
SIVU 34
Tampereen uusi oikeustalo käyttöön
SIVU 33
SIVU 11
SIVU 2
SIVU 18
SIVU 20
Suomen uusi omakuva
HANNU OLKINUORA
menneisyyteen.
Kymmenesosa Suomen eliölajeista uhanalaisia
Uhanalaisten eläinten ja kasvien seurantatoimikunnan mietintö (1991).
MATTI LEINONEN
Kotka, komea lintu.
JARI SINKKONEN
Osanottajia enemmän kuin koskaan ennen
Metsästys sivuasias
VELI -- PEKKA LESKELÄ
Missä jänis?
Tulokset:

Korpus

Aamulehti 1995 (SKTP)

Kuvailutiedot

Lisenssi: CLARIN RES +PLAN

Viittaa korpukseseen

Linkki korpukseseen Korpissa:
2014052719**Tekstin piirteet**

otsikko: Nimismiehet virkkilpalluun

tekijä: [tyhjä]

julkaisija: Kustannus Oy Aamulehti

tekstin sanamäärä: 82

tekstin tunniste Lemmiessä:

kieli: suomi

päiväys: 1995-10-01

tiedoston nimi: aamulehti/1995-10-01

/10/aamu5551.xml

käyttöoikeusluokka: B

avustaja: Helsinki (HEL)

lähde: Aamulehti

korpuksen tunniste Lemmiessä:

aihe: [tyhjä]

kappaleen tyyppi: otsikko

sisältävät elementit: body div

Plenary Sessions of the Parliament of Finland

- *Plenary Sessions of the Parliament of Finland, Kielipankki Korp Version 1.1* contains the transcripts of the plenary sessions held during 10.9.2008 - 1.7.2016, accessible via the Korp search and analysis tool.

The Parliament of Finland (2017). *Plenary Sessions of the Parliament of Finland, Helsinki Korp Version 1.1* [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2017020202>

- Search results also contain a link to the video recording



Eduskunnan täysistunnot selected — 22.46M of 12.93G tokens



Simple Extended Advanced Compare

base form is

maahanmuuttaja Aa

or

Search within

Plenary Sessions of the Parliament of Finland: *maahanmuuttaja* 'immigrant' (all forms)

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC Statistics Word picture Map

Results: 1,618

« « 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... » » Show context

EDUSKUNNAN TÄYSISTUNNOT (does not support extended context)

in jonain päivänä näkeväni myös kuntoutusammattilaisia ja maahanmuuttajapalveluissa lisää kielitaitoisia **maahanmuuttajia** kuten esimerkiksi Lappeenrannassa on jo tällä hetkellä.
Sivistysbudjetissa uudet suomalaiset eli **maahanmuuttajat** on noteerattu yhtenä painopistealueena.
Lisävaroja tulee **maahanmuuttajien** perusopetukseen valmistavaan koulutukseen sekä maahanmuuttajataustaisten opettajien koulutta
Väestömme ikääntyessä **maahanmuuttajia** tarvitaan Suomeen yhä enemmän.
Myös **maahanmuuttajien** tarpeet ammatillisessa koulutuksessa on huomioitava nykyistä paremmin.
Maahanmuuttajien ammatilliseen koulutukseen ohjataan lisää rahaa.
Merkittävä toinen Pop panostus on **maahanmuuttajien** valmistavan koulutuksen keston pidentäminen puolesta vuodesta vuoteen.
Nuorten syrjäytymistä vastaan on tehty paljon täällä toimenpiteitä esimerkiksi **maahanmuuttajien** osalta erityisopetuksen osalta ammatillisia aloituspaikkoja on systemaattisesti lisätty ja lisätään ed
tuista varoista 15 miljoonaa kohdennettiin uudelleen siten että ne menevät ammatilliseen koulutukseen **maahanmuuttajien** koulutukseen ja nuorten syrjäytymisen ehkäisemiseen ja työllisyyden edistämiseen.
Erillinen määräraha **maahanmuuttajien** kielikoulutukseen olisi välttämätön jotta siihen ei jouduttaisi käyttämään muutenkin riittämättömiä
Me tarvitsemme **maahanmuuttajia** .
Sisäasiainministeriön vuoden 2009INE talousarviossa on tavoitteena että **maahanmuuttajien** koko koulutusjärjestelmä vastaa sekä laadullisesti että määrällisesti yksilöllisiä tarpeita.
Tavoitteena on myöskin että **maahanmuuttajien** työttömyys puolittuu vuoteen 2012INE mennessä.
e merkittäville maahanmuuttoalueille laaditaan yhdessä valtion ja seutujen kuntien kanssa pilottiohjelma **maahanmuuttajien** kotouttamisen ja työllistymisen edistämiseksi.
Samoin on koko yhteiskunnan etu myös että **maahanmuuttajat** työllistyvät mahdollisimman hyvin ja saavat tarvitsemansa koulutuksen.
Sinne tuli sekä paikallisia että näitä **maahanmuuttajia** pienen ajan päästä tuli valtavasti kumpienkin vanhempia myös mukaan.
On hieno asia se että **maahanmuuttajia** ja turvapaikanhakijoita on otettu Suomeen aiempaa enemmän.
ää että yhdistys ja järjestötoiminta on erittäin keskeisessä asemassa ja tiedämme myöskin miten tärkeitä **maahanmuuttajat** ovat meidän jalkapallolemme jos nyt vain viitataan Saksa otteluun.
eman kuin sosiaalisia yrityksiä jotka esimerkiksi olisivat kuin vuokratyöyrityksiä ja tarjoaisivat ja etsisivät **maahanmuuttajille** nimenomaan turvapaikkamenettelystä tulleille ihmisille ehkä sitten tällä tavalla työpaikkoja.
Miten tullaan varmistamaan se että suomen kielen opetusta on tarjolla ja että sitä saa silloin kun **maahanmuuttajia** on valmis sitoutumaan kielen opiskeluun.
niin tämä on todella paljon ja samaan aikaan meillä sanotaan että maahan tarvitaan niin paljon **maahanmuuttajia** ia niin paljon iotta tämä kansa voi selvitä tulevaisuudessa.

Corpus

Eduskunnan täysistunnot

Persistent Identifier (metadata):

urn:nbn:fi:lb-2017020202

Cite corpus

Licence: CC BY-NC-ND (CLARIN PUB)

Link to corpus in Korp

Text attributes

file name: 2008-09-18.vrt

date: 2008-09-19

text_time: 06:12:09

speaker_name: Outi Alanko-Kahiluoto

speaker_parl_group: parlgroup_vihv

speaker_role: [empty]

speech_type: [empty]

speaker: Outi Alanko-Kahiluoto /vihv

utterance begin time: 4797368

utterance duration: 20960

utterance end time: 4818328

utterance number: 9577

Word attributes



Eduskunnan täysistunnot selected — 22.46M of 12.93G tokens



Simple Extended Advanced Compare

base form is Aa

or

Compile statistics from search results in Korp

Search within

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: speaker role (+1) Show word picture Show map

KWIC Statistics Word picture Map

Results: 1,618

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 »

in jonain päivänä näkeväni myös kuntoutusammattilaisia ja maahanmuuttajapalvelu-
Sivistysbudjetis

Merkittävä
Nuorten syrjäytymistä vastaan on tehty paljon tällä hetkellä
tuista varoista 15 miljoonaa kohdennettiin uudelleen siten että ne menevät amm-
Sisäasiainministeriön vuoden 2009INE talousarvio-
Tavoitteena
e merkittäville maahanmuuttoalueille laaditaan yhdessä valtion ja seutujen kuntien
Samoin on koko yhteistyö
Sinne tuli s

On hieno asia se että maahanmuuttajia ja turvapaikanhakijoita on otettu Suomeen aiempaa enem-
ää että yhdistys ja järjestötoiminta on erittäin keskeisessä asemassa ja tiedämme myöskin miten tärkeää maahanmuuttajien o-
vat meidän jalkapallollemme jos nyt vain viitataan Saksa o

- word
- Word attributes**
- base form
 - part of speech
 - msd
 - dependency relation
 - name classification
- Text attributes**
- date
 - file name
 - speaker
 - speaker role
 - speaker_name
 - speaker_parl_group
 - speech_type
 - text_time
 - utterance begin time
 - utterance duration

Aa

Support extended context)

esimerkiksi Lappeenrannassa on jo tällä hetkellä.
erattu yhtenä painopistealueena.
petukseen valmistavaan koulutukseen sekä maahan-
in Suomeen yhä enemmän.
: ammatillisessa koulutuksessa on huomioitava nyky-
illiseen koulutukseen ohjataan lisää rahaa.
avan koulutuksen keston pidentäminen puolesta vu-
erityisopetuksen osalta ammatillisia aloituspaikkoja
kseen ja nuorten syrjäytymisen ehkäisemiseen ja ty-
lutukseen olisi välttämätön jotta siihen ei jouduttais

koulutusjärjestelmä vastaa sekä laadullisesti että mää-
nyys puolittuu vuoteen 2012INE mennessä.
amisen ja työllistymisen edistämiseksi.
yvät mahdollisimman hyvin ja saavat tarvitsemansa l-
ajan päästä tuli valtavasti kumpienkin vanhempia m

perusmuoto on
 maahanmuuttaja Aa
 tai
 +

Statistics (*maahanmuuttaja* / speaker role, parliamentary group)

Hae virkkeen sisältä

Arvoja: 37
 Näytä trendikuvaaja

<input type="checkbox"/>	puhujan rooli	eduskuntaryhmä	Yhteensä	Eduskunnan täysistunnot
<input checked="" type="checkbox"/>	Σ	Σ	72 (1 618)	72 (1 618)
<input type="checkbox"/>	ei tiedossa	Perussuomalaisten eduskuntaryhmä	15,2 (342)	15,2 (342)
<input type="checkbox"/>	ei tiedossa	Kansallisen kokoomuksen eduskuntaryhmä	12,5 (281)	12,5 (281)
<input type="checkbox"/>	ei tiedossa	Sosialidemokraattinen eduskuntaryhmä	10,2 (229)	10,2 (229)
<input type="checkbox"/>	ei tiedossa	Keskustan eduskuntaryhmä	8,9 (200)	8,9 (200)
<input type="checkbox"/>	ei tiedossa	Vasemmistoliiton eduskuntaryhmä	6,1 (136)	6,1 (136)
<input type="checkbox"/>	ei tiedossa	Vihreä eduskuntaryhmä	5,9 (133)	5,9 (133)
<input type="checkbox"/>	Maahanmuutto- ja eurooppaministeri	ei tiedossa	2,3 (51)	2,3 (51)
<input type="checkbox"/>	ei tiedossa	Kristillisdemokraattinen eduskuntaryhmä	1,9 (43)	1,9 (43)
<input type="checkbox"/>	ei tiedossa	Ruotsalainen eduskuntaryhmä	1,8 (41)	1,8 (41)
<input type="checkbox"/>	Työministeri	ei tiedossa	1,2 (26)	1,2 (26)
<input type="checkbox"/>	Opetusministeri	ei tiedossa	1 (23)	1 (23)
<input type="checkbox"/>	Puhemies	ei tiedossa	0,6 (13)	0,6 (13)
<input type="checkbox"/>	Oikeus- ja työministeri	ei tiedossa	0,6 (13)	0,6 (13)
<input type="checkbox"/>	ei tiedossa	Vasenyryhmän eduskuntaryhmä	0,5 (11)	0,5 (11)
<input type="checkbox"/>	Valtiovarainministeri	ei tiedossa	0,5 (11)	0,5 (11)



Eduskunnan täysistunnot selected — 22.46M of 12.87G tokens



maahanmuuttaja.nn

Map feature (Named Entity recognition)

Simple Extended Advanced Compare

maahanmuuttaja (noun) Search

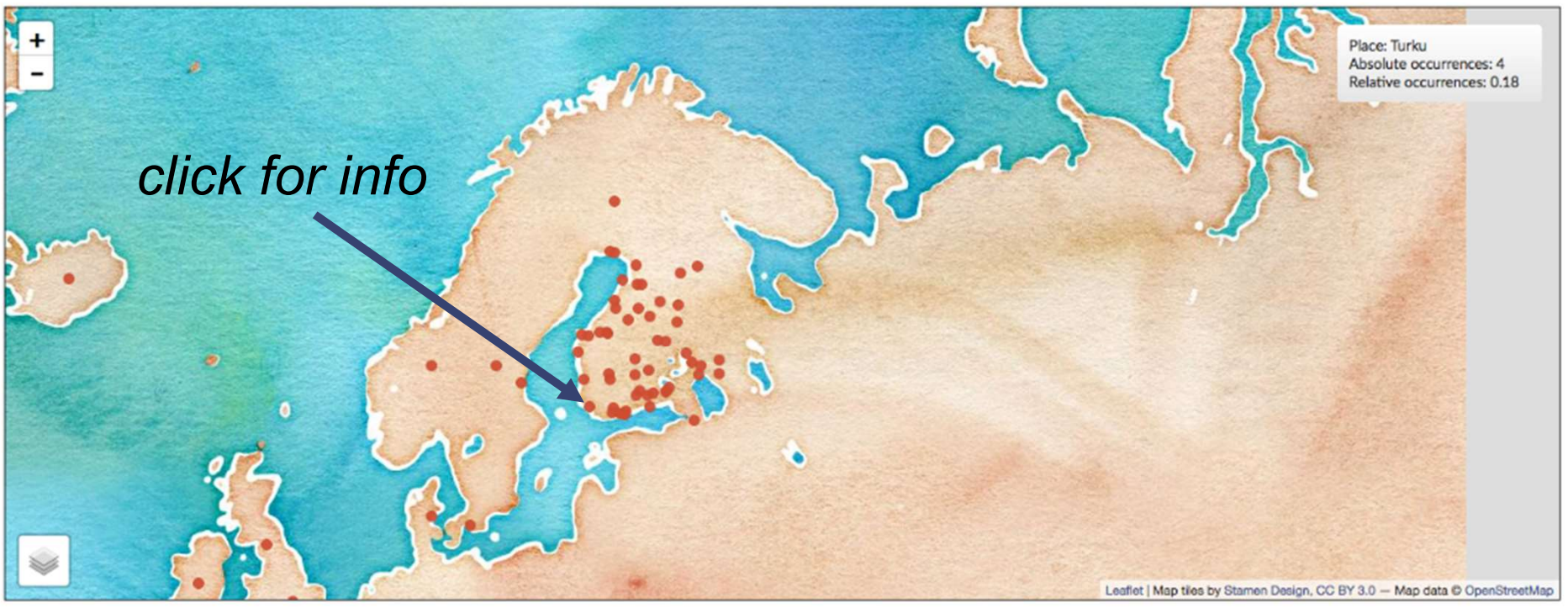
also as initial part final part and case-insensitive

in which contain

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC Statistics Word picture Map KWIC

Locations: 85





Eduskunnan täysistunnot selected - 22.46M of 12.87G tokens



maahanmuuttaja..nn

Simple Extended Advanced Compare

maahanmuuttaja (noun) Search

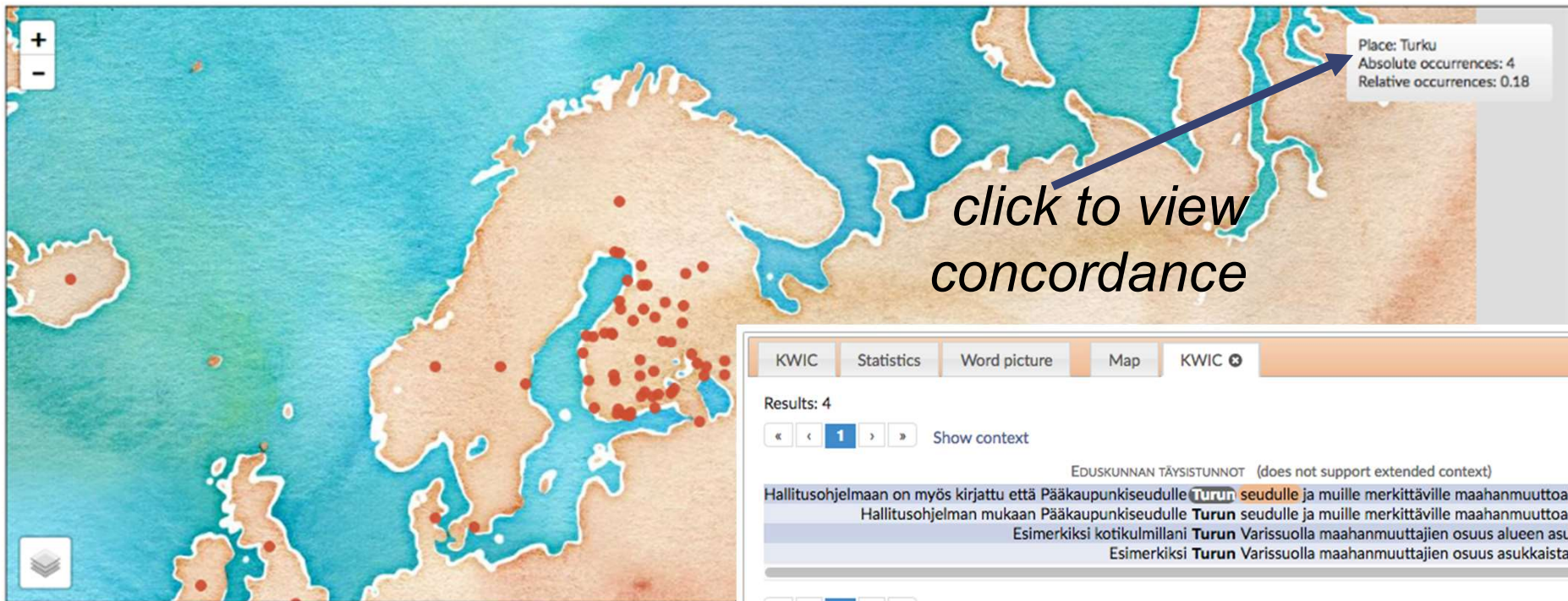
also as initial part final part and case-insensitive

in sentences which contain

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show word picture Show map

KWIC Statistics Word picture Map KWIC

Locations: 85



Map feature (Named Entity recognition)

click to view concordance

KWIC Statistics Word picture Map KWIC

Results: 4

Show context

EDUSKUNNAN TÄYSISTUNNOT (does not support extended context)

Hallitusohjelmaan on myös kirjattu että Pääkaupunkiseudulle Turun seudulle ja muille merkittäville maahanmuuttoalueille laaditaan yhdessä Hallitusohjelman mukaan Pääkaupunkiseudulle Turun seudulle ja muille merkittäville maahanmuuttoalueille laaditaan yhdessä Esimerkiksi kotikulmillani Turun Varissuolla maahanmuuttajien osuus alueen asukkaista on koko ajan kas Esimerkiksi Turun Varissuolla maahanmuuttajien osuus asukkaista on jo hieman yli 40 pros

Show context

työntekijät ja maahanmuuttajien työntekijät ja saavat tarvitessaan koulutuksen.
maahanmuuttajia pienen ajan päästä tuli valtavasti kumpienkin vanhempia myös mukaan.
maahanmuuttajia ja turvapaikanhakijoita on otettu Suomeen aiempaa enemmän.
maahanmuuttajat ovat meidän jalkapallollemme jos nyt vain viitataan Saksa otteluun.
maahanmuuttajille nimenomaan turvapaikkamenettelystä tulleille ihmisille ehkä sitten tällä
maahanmuuttajia on valmis sitoutumaan kielen opiskeluun.
maahanmuuttajia ja niin paljon jotta tämä kansa voi selvitä tulevaisuudessa.
maahanmuuttajia .
maahanmuuttajien valmistavaa opetusta jota esitetään pidennettäväksi puolesta vuodesta
maahanmuuttajien valmistavan koulutuksen lisäämistä ja suomi tai ruotsi toisena kielenä o
maahanmuuttajien valmistavan opetuksen ja kielikoulutuksen laajentamiseksi.



Excel (XLS) ▾

parliamentary group: parlgroupp_kok
speaker role: [empty]
type of speech: [empty]
speaker: Arja Karhuvaara /kok
utterance begin time (ms): 1033944
utterance duration (ms): 18709
utterance end time (ms): 1052653
utterance number: 5290
session duration: 0:21:11.600
utterance begin time: 0:17:13.944
utterance end time: 0:17:32.653
utterance duration: 0:00:18.709

Word attributes

base form: maahanmuuttaja
baseform (compound boundaries):
maahanmuuttaja
part of speech: noun
msd: NUM_PI|CASE_Par
dependency relation: coordination
word position in a name: outside (O)

Show Dependency Tree

Show video

Show video in LAT service



Play/Pause -30 s -5 s +5 s +30 s

Play selection Stop video at selection end:

Utterance: Kuntoutuspäätösten tekijänä tulevaisuudessa soisin jonain päivänä näkeväni myös kuntoutusammattilaisia ja maahanmuuttajapalveluissa lisää kielitaitoisia **maahanmuuttajia** kuten esimerkiksi Lappeenrannassa on jo tällä hetkellä .

name of speaker: Arja Karhuvaara
parliamentary group: parlgroup_kok
speaker role: speakerrole_
type of speech: speechtype_

utterance begin time: 0:17:13.944
utterance end time: 0:17:32.653
utterance duration: 0:00:18.709
utterance begin time (ms): 1033944
utterance end time (ms): 1052653
utterance duration (ms): 18709

Media source:

https://down0-80133-down0.dna.qbrick.com/80133-down0/Syysistuntokausi_2008/74._perjantaina_12._syyskuuta_2008.mp4

Corpus



Eduskunnan täysistunnot

Persistent Identifier (metadata):

[urn:nbn:fi:lb-2019101621](https://nbn-resolving.org/urn:nbn:fi:lb-2019101621)

Licence (audio and video): CLARIN PUB +BY +PRIV +ND +OTHER 1.0

Licence (text): CLARIN PUB +BY +PRIV 1.0

[Link to search in Korp](#)

Text attributes

publication date: 2008-09-12

publication time: 11:21:12

Please note that the video feature on this page is not supported by all browsers. The page is most likely to work in a Firefox browser.

The original transcript has been aligned with the media file and annotated by using automatic methods. The alignment or the annotations may thus contain errors.



Researcher of the Month: Tommi Jantunen



photo: Martti Minkkinen

The Language Bank of Finland is a service for researchers using language resources. University of Jyväskylä academy researcher in Language and Communication Studies Tommi Jantunen describes his research on the ProGram corpus.

Who are you?

I am Tommi Jantunen and I work as an academy researcher of the Finnish sign language at the Department of Languages and Communication Studies of the University of Jyväskylä

Tommi Jantunen

Researcher of the Month 3/2017

Professor of Finnish Sign Language,
University of Jyväskylä

Research interests:

- structure of sign language from various perspectives
- Applications of computer vision and motion capture technologies in the study of video corpora

Corpus of Finnish Sign Language (CFinSL)

- Finnish Sign Language material collected in the CFINSL project
- over 15 hours of video files from six camera angles
- Elicited and spontaneous signing
- annotations in ELAN format

University of Jyväskylä, Sign Language Centre (2019). Corpus of Finnish Sign Language [sign language corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2019012321>

The CFINSL corpus: elicited narratives



Text

Grid

Subtitle

Timeline

Video display



Information

General Session Technical

Resource: CFINSL2014_012_04...

Media file: CFINSL2014_012_04...

Elapsed time: 00:06:52:320

Selected chunk:

Begin time: -

End time: -

Text: -

Mini Data Frame

Tier: none

Font size: 14

Timeline

00:06:48:500 00:06:49:000 00:06:49:500 00:06:50:000 00:06:50:500 00:06:51:000 00:06:51:500 00:06:52:000 00:06:52:500 00:06:53:000 00:06:53:500 00:06:54:000

ID_1_oik	VA-]	PAINELLA	_kvkk	K	ZZz	LIKAINEN	_kvmk	_kvkk	_kvkk	Ei-	_kvkk	MENNÄ	AJAA-AU
ID_huomioita_						@y				@y			
ID_2_oik													
ID_huomioita_													
ID_1_vas	VA-]	PAINELLA	_kvak	K	ZZz		_kvmk	_kvkk		KÄMMEN-	Ei-	_kvak	AJAA-AUT
ID_2_vas													
Käännös_1			ajaa kaasu-pohjassa ohitellen.	(t)		Rokkapönttö lentää auton-törmäessä-siihen.				Ei väli:	mir	sukkuloiden eteenpäin.	(Bean) aj
Käännös_2													
Käännöshuomik													
Käännöshuomik													
@_1_oik				@h				@toisto			@n		

- Play selection
- Clear selection
- Create bookmark

Navigation buttons: |< >|, << >>, < >, + -

- Play screen by screen
- Play continually

Tier text font:

”The trashcan flies as the car crashes into it...”

How to locate and access resources



CLARIN licence categories



Publicly available



Available for academic, logged in users



Personal permission is required for access

More detailed licence conditions



+BY author must be cited
+NC non-commercial use only



+ID login is required
+PLAN research plan is required

+PRIV contains personal data



+NORED redistribution is not allowed
+DEP modified versions can be redistributed
via CLARIN

List of corpora at the Language Bank

www.kielipankki.fi/corpora

Etsi:

Lyhenne	Nimi ja kuvailutiedot	Lisenssi	Hae	Sijainti	Ohje	Viite
acquis-ftb3	JRC-Acquis-monikielisen rinnakkaiskorpuksen suomenkielinen osakorpus			Korp		”
Agricola	Mikael Agricolan teosten morfosyntaktinen tietokanta			Korp		”
aku-egg	Puheen ja EGG:n samanaikaiset tallenteet			LAT		”
amph	Ajatella, miettiä, pohtia, harkita -korpus			Lataus		”
ArkiSyn-korp	Suomenkielisten arkikeskustelujen morfosyntaktinen tietokanta, Helsinki-Korp-versio			Korp		”
BeserCorp	Udmurtin besermanin murteen korpus			Korp		”
ceal-o	Englantilaisen ja amerikkalaisen kirjallisuuden klassikoita suomeksi...			Korp		”

Citation instructions

suomi24-2001-2017-korp-v1-1	The Suomi24 Sentences Corpus 2001-2017, Korp version 1.1		Korp	”
Suomi24-2015H1	The Suomi 24 Corpus (2015H1)		Download	”
Suomi24-2016H2	The Suomi 24 Corpus (2016H2)		Download	”
Suomi24-2017H2	The Suomi 24 Corpus (2017H2)		Download	”
Suomi24-korp-2016H2	The Suomi 24 Sentences Corpus (2016H2)		Korp	”
Suomi24-virkkeet-2015H2	The Suomi 24 Sentences Corpus (2015H2)		Korp	”
susanne-uhlcs	The Susanne Corpus (UHLCS)		Taito	”
taaf	Triangle of Aspects Analysis of Frozen		Download	”
taajuussanasto9996	Frequency Lexicon of the Finnish Newspaper Language		Download	”
topling-en	The English Subcorpus of Topling - Paths in Second Language Acquisition		Korp	”
topling-fi	The Finnish Subcorpus of Topling - Paths in Second Language Acquisition		Korp	”
topling-sv	The Swedish Subcorpus of Topling - Paths in Second Language Acquisition		Korp	”
TSK-35	The Vocabulary of Safety and Health at Work (TSK 35)		Download	”
TSK-49	The Terminological Vocabulary of Kela – Benefit-related Concepts, 4th Edition (TSK 49)		Download	”
ums-seami-uhlcs	Uma Seami Corpus (UHLCS)		Taito	”

suomi24-2001-2017-korp-v1-1	The Suomi24 Sentences Corpus 2001-2017, Korp version 1.1	PUB	Korp	”
Suomi24-2015H1	The Suomi 24 Corpus (2015H1)	ACA	Download	”
Suomi24-2016H2	The Suomi 24 Corpus (2016H2)	ACA	Download	”
Suomi24-2017H2	The Suomi 24 Corpus (2017H2)			
Suomi24-korp-2016H2	The Suomi 24 Sentences Corpus (2016H2)			
Suomi24-virkkeet-2015H2	The Suomi 24 Sentences Corpus (2015H2)			
susanne-uhlcs	The Susanne Corpus (UHLCS)			
taaf	Triangle of Aspects Analysis of Frozen			
taajuussanasto9996	Frequency Lexicon of the Finnish Newspaper Language			
topling-en	The English Subcorpus of Topling - Paths in Second Language A			
topling-fi	The Finnish Subcorpus of Topling - Paths in Second Language A			
topling-sv	The Swedish Subcorpus of Topling - Paths in Second Language A			
TSK-35	The Vocabulary of Safety and Health at Work (TSK 35)			
TSK-49	The Terminological Vocabulary of Kela - Benefit-related Conce			
umo-saami-uhlcs	Ume Saami Corpus (UHLCS)			

viittaus | Kielipankki - Mozilla Firefox

https://www.kielipankki.fi/viittaus/?key=Suomi24-2016H2&lang=en 80%

[suomeksi] [in English]

Reference instructions: Suomi24-2016H2

Please cite the language resource as follows:

Aller Media Ltd. (2014). *The Suomi 24 Corpus (2016H2)* [text corpus]. Kielipankki. Retrieved from <http://urn.fi/urn:nbn:fi:lb-2017021506>

Show: [Bibtex] [Zotero]

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

suomi24-2001-2017-korp-v1-1 The Suomi24 Sentences Corpus 2001-2017, Korp version 1.1 PUB Korp

urn:nbn:fi:lb-2017021506 OR "The Suomi 24 Corpus (2016H2)" - Google Scholar - Mozil...

https://scholar.google.com/scholar?q=urn%3Anbn%3Afi%3Alb-20170215

urn:nbn:fi:lb-2017021506 OR "The Suomi 24 Corpus (2016H2)"

Scholar 11 tulosta (0,07 sekuntia) VUOSI

Tarkoititko: urn:nbn:fi:lb-2017021506 OR "The **Suomi24** Corpus (2016H2)"

[SITAATTI] Mietta Lennes: Kielipankin esittely: "Mitä voin lainata Kielipankista?"-tapahtumassa Tiedekulmassa 8.11. 2016 pidetty esitys

ME Lennes - 2016 - researchportal.helsinki.fi
... Finland. Project: Research project. Datasets. **The Suomi 24 Corpus (2016H2)**. Dataset. Research Infrastructures. CLARIN - Common Language Resource and Technology Infrastructure in Finland. Facility/equipment: Equipment ...

[SITAATTI] Keskustelutilan lämmönsäätelyä
KH Lagus - Citizen mindscapes: tutkimuskollektiivi, joka ..., 2016 - researchportal.helsinki.fi
... Related content. Projects. Citizen Mindscapes – Detecting Social, Emotional and National Dynamics in Social Media. Project: Research project. Datasets. **The Suomi 24 Corpus (2016H2)**. Dataset. Powered by Pure, Scopus & Elsevier Fingerprint Engine™ © 2019 Elsevier BV ...

[SITAATTI] SUOMI24: Muodonantoa aineistolle

ACA Download
ACA Download

elipankki.fi/viittaus/?key=Suomi24-2016H2&lang=ei 80%

[suomeksi] [in English]

itions: Suomi24-2016H2

source as follows:
Suomi 24 Corpus (2016H2) [text corpus]. Kielipankki. Retrieved from http://urn.fi/urn:nbn:fi:lb-2017021506

language resource in Google Scholar

List of corpora at the Language Bank

www.kielipankki.fi/corpora

Etsi:

Lyhenne	Nimi ja kuvailutiedot	Lisenssi	Hae	Sijainti	Ohje	Viite
acquis-ftb3	JRC-Acquis-monikielisen rinnakkaiskorpuksen suomenkielinen osakorpus			Korp		”
Agricola	Mikael Agricolan teosten morfosyntaktinen tietokanta			Korp		”
aku-egg	Puheen ja EGG:n samanaikaiset tallenteet			LAT		”
amph	Ajatella, miettiä, pohtia, harkita -korpus			Lataus		”
ArkiSyn-korp	Suomenkielisten arkikeskustelujen morfosyntaktinen tietokanta, Helsinki-Korp-versio			Korp		”
BeserCorp	Udmurtin besermanin murteen korpus			Korp		”
ceal-o	Englantilaisen ja amerikkalaisen kirjallisuuden klassikoita suomeksi...			Korp		”

META SHARE

Documentation

Statistics



digilib

Search

Filter by:

▸ Language

▸ Resource Type

▸ Media Type

2 Language Resources

Order by: Resource Name A-Z ▾

 **The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771-1874)** 

↓ 2 👁 297

Finnish Swedish

Metadata



META SHARE

[Browse Resources](#) [Community](#) [Documentation](#) [Statistics](#)

The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771-1874) 

 302  8  2

▸ View resource name in all available languages

Digilib-Pub-1874-dl

Persistent Identifier of this resource: <http://urn.fi/urn:nbn:fi:lb-2015051201>

Access location: <http://urn.fi/urn:nbn:fi:lb-201505112>

The corpus is available in Kielipankki - the Language Bank of Finland, download: <http://urn.fi/urn:nbn:fi:lb-201505112>, as well as on the Taito server (<https://sui.csc.fi/group/sui/ssh-console>), directory name: /appl/kielipankki/Digilib-pub.


This corpus consists of the OCR results of the material... [Read More](#)

▸ View resource description in all available languages

[« Back](#) [Edit Resource](#)


digilib 





Showing 1 results for digilib 


Results per page: 10 


Use the categories below to limit the search results to those matching the selected value(s).

Language 

Collection 


Resource type 

Modality 

Format 

The Newspaper and Periodical OCR Corpus of the National Library of Finland (1771-1874)

(Part of CLARIN Centres)

 The corpus is available in Kielipankki - the Language Bank of Finland, download: <http://urn.fi/urn:nbn:fi:lb-201505112>, as well as on the Taito server (<https://sui.csc.fi/group/sui/ssh-console>), directory name: /appl/kielipankki/**Digilib**-pub. This corpus consists of the OCR results of the material published before 1875...





Download service

www.kielipankki.fi/download



Select corpus

Location: /download/

Logged in as: (none)

Name	Size	Description
 AMPH/	-	amph-Corpus
 avoid/	-	Corpus of Age-related Voice Disguise
 BYU/	-	The BYU corpora at Kielipankki - The Language Bank of Finland
 CEAL/	-	CEAL corpus
 Digilib/	-	Kansalliskirjaston lehtikokoelma
 DSPCON/	-	Aalto University DSP Course Conversation Corpus
 eduskunta/	-	Plenary Sessions of the Parliament of Finland

Accept license conditions



Digilib-1920-dl (Digilib 1875 - 1920) Download

Lue alle linkitetty lisenssisopimus huolella ja rasti ruutu. Sopimus on toistaiseksi vain englanniksi. Voit ladata aineiston vain, jos sinulla on Haka-tunnus.

Carefully read the license agreement linked below and select the checkbox. You need a University account to download the resource.

 I have read and understood the [Digilib-1920-dl license agreement \(CLARIN ACA NC NoRED\)](#)

Log in to Download



The Language Bank of Finland

Korp

Choose your organization or federation.

Login with:

University of Helsinki

Login

Narrow down the list by typing any part of your organisation.



Select corpus

Location: /download/

Logged in as: (none)

Name	Size	Description
 AMPH/	-	amph-Corpus
 avoid/	-	Corpus of Age-related Voice Disguise
 BYU/	-	The BYU corpora at Kielipankki - The Language Bank of Finland
 CEAL/	-	CEAL corpus
 Digilib/	-	Kansalliskirjaston lehtikokoelma
 DSPCON/	-	Aalto University DSP Course Conversation Corpus
 eduskunta/	-	Plenary Sessions of the Parliament of Finland

Restricted corpora:

Ei latausoikeutta

Valitun kielivaran lataaminen edellyttää käyttöoikeutta.

Kielipankin oikeudet: <https://lbr.csc.fi>.

Tiedustelut ja tuki: kielipankki@csc.fi.

Takaisin lataussivulle.

No download access

Downloading the selected resource requires a permission.

Language Bank Rights

Applying for access to restricted corpora

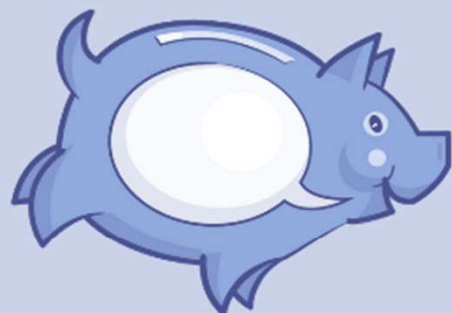


Log in

www.kielipankki.fi

[ETUSIVU](#) [TIETOSUOJASELOSTE](#)

[EN](#) [FI](#)



KIELIPANKKI
The Language Bank of Finland

Kirjaudu sisään

Kirjaudu sisään Haka-tunnuksillasi

Haka

Select corpora

Kielivarat



Kielivarat ↑

1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpus

Lisätietoja ↗

Lisää koriin

Aalto University DSP Course Conversation Corpus 2013-

Lisätietoja ↗

Lisää koriin

Ajatella, miettiä, pohtia, harkita -korpus

Lisätietoja ↗

Lisää koriin

BYU-korpukset (COCA, COHA, GloWbE), ladattava versio

Lisätietoja ↗

Lisää koriin

Corpus of Age-related Voice Disguise

Lisätietoja ↗

Lisää koriin

Digitaalinen muoto-opin arkisto

Lisätietoja ↗

Lisää koriin



Add to basket

Kielivarat

 2 kielivaraa korissa

1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä -korpus

Poista

Digitaalinen muoto-opin arkisto

Poista

Hae näitä kielivaroja (2 kpl) yhdessä

Hae



Fill in the application form

Tiedot

Tutkimusaihe lyhyesti (enintään 80 merkkiä)

Tutkimussuunnitelman tiivistelmä tai lyhyt selvitys käyttötarkoituksesta

Käyttöehdot

- [Kielipankin yleiset käyttöehdot](#)
- [CLARIN RES +PLAN +BY +NC +LOC +NORED +ND \(DMA\)](#)



Tallenna luonnos

Lähetä hakemus

Accept licence conditions

CLARIN RES loppukäyttäjän lisenssisopimus +NC +LOC +ND 1.0

Kielivara: Digitaalinen muoto-opin arkisto (URN: [urn:nbn:fi:lb-201403261](https://nbn-resolving.org/urn:nbn:fi:lb-201403261))

Oikeudenhaltija: Helsingin yliopiston suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos

Oikeudenhaltija myöntää luvansaajalle maksuttoman, ei-yksinomaisen ja pysyvän (tekijänoikeuden voimassaoloajan kestävän) luvan **käyttää ja kopioida kielivaraa luvansaajan omaan käyttöön** muutettuna, muuttamattomana tai osana yhteenliitettyä teosta. Mainitut oikeudet koskevat kaikkia tunnettuja viestintävälineitä ja muotoja ja sisältävät oikeuden tehdä sellaisia muutoksia, jotka mahdollistavat luvansaajalle kielivaran käyttämisen toisissa laitteissa ja formaateissa.

Palveluehtojen mukaiset lisenssiehdot:

Tunnistamiseen ja pääsyyn liittyvät ehdot

- ID: Käyttäjän on tunnistauduttava.

Processing and approval

Hakemus

PDF

Tila: Hyväksytty

✓ Hae

✓ Hyväksyntä

✓ Hyväksytty

www.kielipankki.fi/tools

Start	Name	Description	Instructions	Install	Administrator
	Korp	A web-based concordance tool that can be used for corpus queries based on morphosyntactic analysis and various other features.	Instructions		
	LAT (Language Archive Tools)	A toolkit for browsing and querying annotated speech and video corpora.	Instructions		
Download	Download service	Download certain corpora.			
	META-SHARE	Metadata repository of all the language resources at the Language Bank of Finland.			
Mylly	Mylly	Versatile data analysis platform with interactive visualizations and workflows.	Instructions		
Sanat	Sanat	A platform for publishing lexica and word lists.			
FinTag	Finnish Tagtools	A part-of-speech and morphology tagger and a named entity recogniser for Finnish.		Install	
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)			
	WebAnno	Text annotation tool.			
	Signbank	Lexical database of Finnish Sign Language.			
	OPUS	An interface for open source parallel corpora.			

Monikielinen tekstintunnistus

Eristä teksti kuvatiedostoista (jpg, png, pdf, ...)

fintag

Tekstin perusmuotoistus, taivutuksen ja nimettyjen ilmausten tunnistus

finsentiment

Tekstin sentimenttianalyysi

lemmarank

Tekstille leimallisten sanojen ja nimettyjen ilmausten löytäminen ja vertailu

lemmamatch

Tekstihaku perusmuotojen ja sanasemantiikan avulla

- [Tieteen termipankkin sanalistoihin perustuva esimerkki](#)
- [Sanasemantiikkaesimerkki](#)
- [Esimerkki sääntöjen kirjoittamisesta](#)

Latauspalvelu

Kehitysvaiheessa olevien aineistojen (sanavektorit) latauspalvelu

fintok

Tekstin tokenisointi

like-unlike (beta)

Haku sanavektori-aineistosta

Demo tools

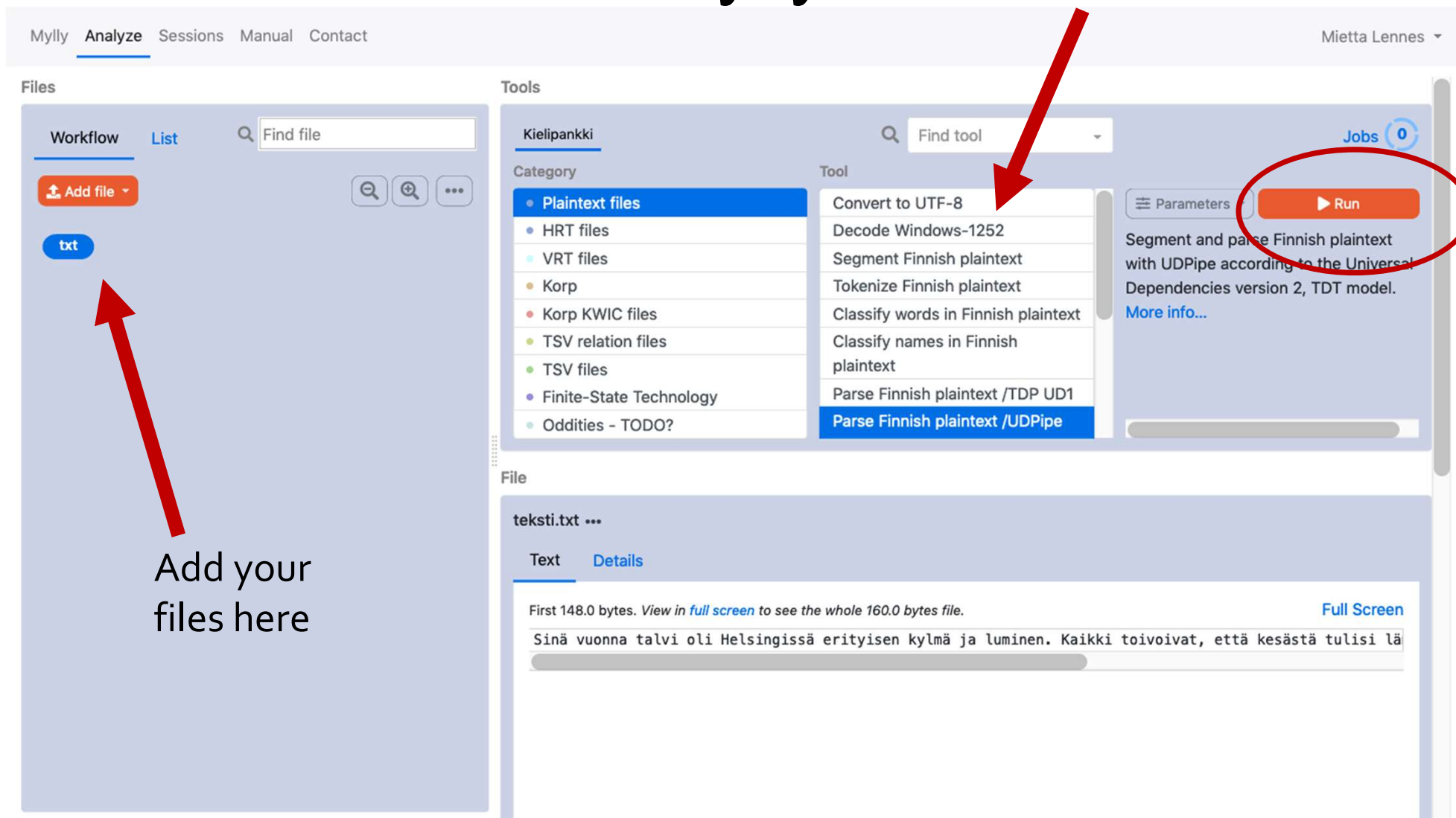
Mylly

<https://mylly.rahtiapp.fi>

- Mylly is a versatile data analysis platform with interactive visualizations and workflows.
- A variety of tools are available for building workflows of your own, e.g., parsers, converters.
- You can also import search results from Korp directly to Mylly (over Korp API).
- Mylly is under development! More tools are forthcoming: automatic speech recognition and alignment, etc.

Mylly

Select tool



The screenshot shows the Mylly web interface. At the top, there is a navigation bar with 'Mylly', 'Analyze', 'Sessions', 'Manual', and 'Contact'. The user 'Mietta Lennes' is logged in. The interface is divided into three main sections: 'Files', 'Tools', and 'File'.

Files Section: Includes a 'Workflow' tab, a 'List' view, a search bar 'Find file', an 'Add file' button, and a 'txt' file type button. A red arrow points to the 'txt' button with the text 'Add your files here'.

Tools Section: Includes a search bar 'Find tool', a 'Jobs' indicator, and a list of tools. A red arrow points to the search bar with the text 'Select tool'. The 'Plaintext files' category is selected, showing a list of tools. The 'Parse Finnish plaintext /UDPipe' tool is highlighted. To the right, the tool's parameters and a 'Run' button are visible. The 'Run' button is circled in red.

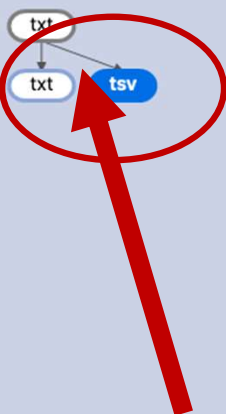
File Section: Shows a file named 'teksti.txt ...' with a 'Text' tab and a 'Details' tab. The file content is displayed: 'Sinä vuonna talvi oli Helsingissä erityisen kylmä ja luminen. Kaikki toivoivat, että kesästä tulisi lä'.

Mylly

Files

Workflow **List**

↑ Add file 🔍 🔍 ⋮



Results appear here

Tools

Kielipankki **Jobs** 0

Category Tool

- Plaintext files
 - Convert to UTF-8
 - Decode Windows-1252
 - Segment Finnish plaintext
 - Tokenize Finnish plaintext
 - Classify words in Finnish plaintext
 - Classify names in Finnish plaintext
 - Parse Finnish plaintext /TDP UD1
 - Parse Finnish plaintext /UDPipe**
- HRT files
- VRT files
- Korp
- Korp KWIC files
- TSV relation files
- TSV files
- Finite-State Technology
- Oddities - TODO?

Parameters **▶ Run**

Segment and parse Finnish plaintext with UDPipe according to the Universal Dependencies version 2, TDT model. [More info...](#)

File

teksti-ud2.fi.rel.tsv ...

Spreadsheet Text Details

First 25 rows. [View in full screen](#) to see all the rows. [Full Screen](#)

kMsentence	kMtok	id	form	lemma	upostag	xpostag	feats
1	1	1	Sinä	se	PRON	Pron	Case=Ess Number=Sing PronType=Dem
1	2	2	vuonna	vuosi	NOUN	N	Case=Ess Number=Sing
1	3	3	talvi	talvi	NOUN	N	Case=Nom Number=Sing
1	4	4	oli	olla	AUX	V	Mood=Ind Number=Sing Person=3 Tense=Past V
1	5	5	Helsingissä	Helsinki	PROPN	N	Case=Ine Number=Sing
1	6	6	erityisen	erityinen	ADJ	A	Case=Gen Degree=Pos Derivation=Inen Number=
1	7	7	kylmä	kylmä	ADJ	A	Case=Nom Degree=Pos Number=Sing

Online courses

Corpus Linguistics and Statistical Methods (5 cr)

Word attributes

- baseform: kieli
- baseform (compound boundaries): kieli
- part-of-speech: noun
- msd: NUM_SgCASE_Nom
- dependency relation: nominal subject
- word position in a name: outside (O)
- original thread
- original message
- Show Dependency Tree
- Näytä dependency puu.



Introduction to Speech Analysis (5 cr)

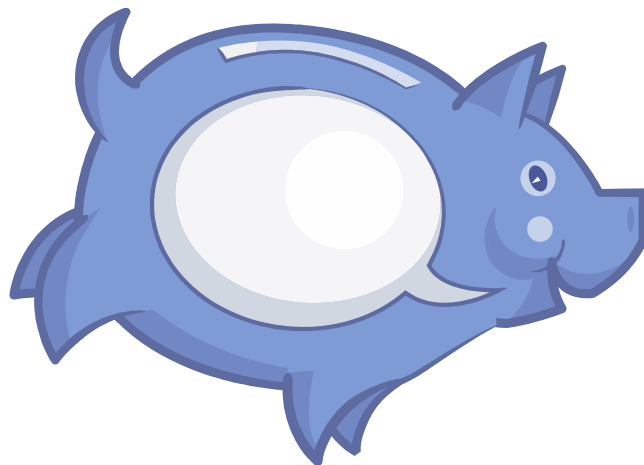
tänne sun huan
 nc niin varmaa
 it sä neuvoe
 isään rohoit:
 itä mä
 sieninkatu e
 s sä tuut ni
 i okel. tä ooi
 tuä



Data Clinic (5 cr)

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1	0	0	0	1	1	SUBCAT_0x([NUM_Sg]CASE_0x)	lamma	ho	ret	deprvl																
2	0	0	0	1	1	NUM_P_CASE_Nom	ret	ret	ret	ret																
3	0	0	0	2	1	MS_Sg(WRCH_0x)(CASE_0x)WORD_0x	paavo	paavo	ret	ret																
4	0	0	0	3	1	NUM_Sg_LAB_0x(CASE_0x)	hukka	hukka	ret	ret																
5	0	0	0	4	1																					
6	0	0	0	5	1	SUBCAT_0x(NUM_Sg)																				
7	0	0	0	6	1	SUBCAT_0x(NUM_Sg)																				
8	0	0	0	7	1	SUBCAT_0x(NUM_Sg)																				
9	0	0	0	8	1	SUBCAT_0x(NUM_Sg)																				
10	0	0	0	9	1	SUBCAT_0x(NUM_Sg)																				
11	0	0	0	10	1	SUBCAT_0x(NUM_Sg)																				
12	0	0	0	11	1	SUBCAT_0x(NUM_Sg)																				
13	0	0	0	12	1	SUBCAT_0x(NUM_Sg)																				
14	0	0	0	13	1	SUBCAT_0x(NUM_Sg)																				
15	0	0	0	14	1	SUBCAT_0x(NUM_Sg)																				
16	0	0	0	15	1	SUBCAT_0x(NUM_Sg)																				
17	0	0	0	16	1	SUBCAT_0x(NUM_Sg)																				
18	0	0	0	17	1	SUBCAT_0x(NUM_Sg)																				
19	0	0	0	18	1	SUBCAT_0x(NUM_Sg)																				
20	0	0	0	19	1	SUBCAT_0x(NUM_Sg)																				
21	0	0	0	20	1	SUBCAT_0x(NUM_Sg)																				
22	0	0	0	21	1	SUBCAT_0x(NUM_Sg)																				
23	0	0	0	22	1	SUBCAT_0x(NUM_Sg)																				
24	0	0	0	23	1	SUBCAT_0x(NUM_Sg)																				
25	0	0	0	24	1	SUBCAT_0x(NUM_Sg)																				
26	0	0	0	25	1	SUBCAT_0x(NUM_Sg)																				





Kiitos! Tack! Thank you!

www.kielipankki.fi

General support
fin-clarin@helsinki.fi

Technical support
kielipankki@csc.fi