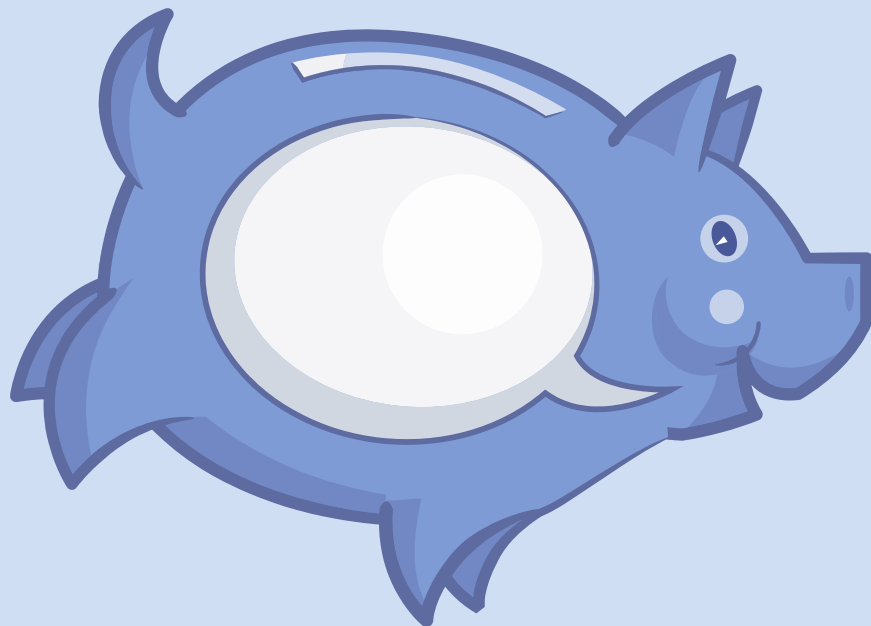
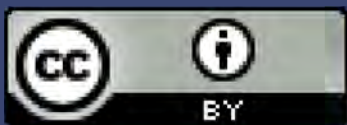




KIELIPANKKI
The Language Bank of Finland

Kielipankki ja FIN-CLARIN

Mietta Lennes



Tämä dokumentti on lisensoitu Creative Commons Attribution 4.0 Kansainvälinen -lisenssillä. Sisältöjä ovat yhteistyössä tuottaneet FIN-CLARINin Kielipankki-työryhmän jäsenet (Mietta Lennes, Krister Lindén ja Tero Aalto).

FIN-CLARIN

KIELIPANKKI

The Language Bank of Finland

KIELIPANKKI KÄYTTÄJÄKSI AINEISTOT TYÖKALUT ORGANISAATIO TUKI

IN ENGLISH PÅ SVENSKA

Käyttäjäksi



Hae kielivarojemme käyttöoikeuksia.

Aineistot



Tutustu aineistotarjontaan.

Työkalut

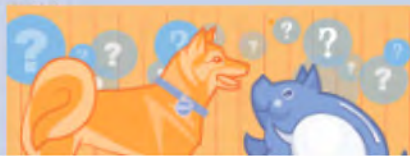


Kokeile työkalujamme.

Organisaatio



Tuki



Hae Kielipankki-portaalista:



Kuukauden tutkija: Eliisa Uotila

Kielipankin käyttäjät

- Kaikkien alojen tutkijat ovat tervetulleita!
- Monet aineistot ovat saatavilla myös ilman kirjautumista.
- FIN-CLARIN voi auttaa aineistosi säilytyksessä ja välityksessä tutkimuskäyttöön.





CLARIN ERIC

European Research Infrastructure Consortium
perustettu 29.2.2012

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



SWE-CLARIN



INSTITUT FÜR
DEUTSCHE SPRACHE



[...]



*Kansainvälinen yhteistyö
ja resurssien yhteiskäyttö*

Jäsenmaat (21):

Alankomaat

Bulgaria
Islanti
Italia
Itävalta
Kypros
Latvia
Liettua
Kreikka
Kroatia
Norja
Puola
Portugali
Ruotsi
Saksa
Slovenia
Suomi
Tanska
Tshekin tasavalta
Unkari
Viro

Tarkkailijajäsenet (3):

Etelä-Afrikka
Iso-Britannia
Ranska

Kolmantena osapuolena:

CMU (USA)

(päivitetty 25.8.2020)



CLARIN-keskukset

Jäsenmaat (21):

Alankomaat

Bulgaria
Islanti
Italia
Itävalta
Kypros
Latvia
Liettua
Kreikka
Kroatia
Norja
Puola
Portugali
Ruotsi
Saksa
Slovenia
Suomi
Tanska
Tshekin tasavalta
Unkari
Viro

Tarkkailijajäsenet (3):

Etelä-Afrikka
Iso-Britannia
Ranska

**Kolmantena
osapuolena:**
CMU (USA)

(päivitetty 25.8.2020)





<https://vlo.clarin.eu>

CLARIN Virtual Language Observatory

Welcome to the VLO!

Use the **search bar** below to start searching through hundreds of thousands of language resources, or [continue](#) to browse everything and use **facets** to narrow down to your area of interest or discover new resources.

See all records

Take a quick tour

Search

Showing all 1640603 records

Results per page: 10

noin
1,6 miljoonaa
tietuetta

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

Language

Collection

Resource type

Modality

Format

Keyword

<< < 1 2 3 4 5 6 7 8 9 10 > >>

EXMARaLDA Demo corpus

(Part of Hamburger Zentrum für Sprachkorpora (HZSK))

A selection of short audio and video recordings in various languages to be used for instruction or demonstration of the EXMARaLDA system.; HIAT (simplified); HIAT; free comment; suprasegmental information; accentuation/stress; English translation; Standard German translation; German translation; English translation; code-switch



The Hamburg MapTask Corpus (HAMATAC)

(Part of Hamburger Zentrum für Sprachkorpora (HZSK))

Audio and two video recordings of map tasks with adult L2 users of German and one L1 speaker. The speakers' L1 and their L2 proficiencies vary. The maps used for the tasks are available.; orthographic transcription/simplified HIAT; Fine-grained part of speech tagging using TreeTagger and the STTS tagset.; superordinate...



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 Daria Fišer and Jakob Lenardič and have received funding from the European Union's Horizon 2020

"Aineistoperheitä" CLARINin kautta

Resource families

[CMC corpora](#)

[Historical corpora](#)

[L2 corpora](#)

[Manually annotated corpora](#)

[Newspaper corpora](#)

[Parallel corpora](#)

[Parliamentary corpora](#)

[Spoken corpora](#)

Gw = miljardia sanaa, Mw = miljoonaa sanaa, kw = miljoonaa sanaa, h = tuntia

Aineistot	2018	2022
<i>Tekstiä</i>		
Sanoma- ja aikakauslehtiä 1770- (Kansalliskirjasto ym.)	12 Gw	20 Gw
Verkkokeskustelut ja muu internet 2000- (Suomiz4, Ylilauta ym.)	4 Gw	10 Gw
Kirjallisuutta ja käsikirjoituksia (Gutenberg, Fennica, arkistot)	60 Mw	70 Mw
<i>Puhetta</i>		
Uutislähetys (YLE)		10000 h
Suomen eduskunnan täysistunnot 2008-2016	500 h	1000 h
Murre- ja arkikielen aineistoja (Kotus, Turku, Helsinki)	500 h	1000 h
Viittomakielten aineistoja (Jyväskylän yliopisto, Kuurojen Liitto)	20 h	500 h
<i>Monikieliset ja muut aineistot</i>		
Monikielisiä aineistoja (EuroParl, laki, Raamattu, tekstitykset, ...)	3 Gw	10 Gw
Oppijankielen aineistoja (Oulu, Jyväskylä, Turku, Aalto)	2 Mw	5 Mw
Avoimet sanakirjat ja terminologiat (Helsinki, Tromssa)	300 kw	400 kw

Gw = miljardia sanaa, Mw = miljoonaa sanaa, kw = miljoonaa sanaa, h = tuntia

Aineistot	2018	2022
<i>Tekstiä</i>		
Sanoma- ja aikakauslehtiä 1770- (Kansalliskirjasto ym.)	12 Gw	20 Gw
Verkkokeskustelut ja muu internet 2000- (Suomiz4, Ylilauta ym.)	4 Gw	
Kirjallisuutta ja käsikirjoituksia (Gutenberg, Fennica, arkistot)	60 Mw	
<i>Puhetta</i>		
Uutislähettyksiä (YLE)		
Suomen eduskunnan täysistunnot 2008-2016	500 h	1000 h
Murre- ja arkikielen aineistoja (Kotus, Turku, Helsinki)	500 h	1000 h
Viittomakielten aineistoja (Jyväskylän yliopisto, Kuurojen Liitto)	20 h	500 h
<i>Monikieliset ja muut aineistot</i>		
Monikielisiä aineistoja (EuroParl, laki, Raamattu, tekstitykset, ...)	3 Gw	10 Gw
Oppijankielen aineistoja (Oulu, Jyväskylä, Turku, Aalto)	2 Mw	5 Mw
Avoimet sanakirjat ja terminologiat (Helsinki, Tromssa)	300 kw	400 kw

Tällä hetkellä
FIN-CLARINin kautta on
tarjolla n. 19 GW
(> 1400 tietokannassa)

<https://www.kielipankki.fi/aineistot>

Etsi:

Lyhenne	Nimi ja kuvailutiedot	Lisenssi	Hae	Sijainti	Tukitaso	Ohje	Viite
acquis-ftb3	JRC-Acquis-monikielisen rinnakkaiskorpuksen suomenkielinen osakorpus			Korp	B		"
agricola-v1-1-korp	Mikael Agricolan teosten morfosyntaktinen tietokanta versio 1.1, Korp			Korp	B		"
ai2d-rst	AI2D-RST: multimodaalinen korpus peruskoulutason luonnontieteiden oppimateriaaleissa esiintyvistä diagrammeista			Lataus	B		"
aku-egg	Puheen ja EGG:n samanaikaiset tallenteet			LAT	B		"
amph	Ajatella, miettiä, pohtia, harkita -korpus			Lataus	B		"
ArkiSyn-korp	Suomenkielisten arkikeskustelujen morfosyntaktinen tietokanta, Helsinki-Korp-versio			Korp	B		"
AVOID	Muunnellun puheen korpus			Lataus	B		"
BeserCorp	Udmurtin besermanin murteen korpus			Korp	B		"
ceal-dl	Englantilaisen ja amerikkalaisen kirjallisuuden klassikoita suomeksi, ladattava versio			Lataus	A		"
ceal-o	Englantilaisen ja amerikkalaisen kirjallisuuden klassikoita suomeksi, alkuperäisessä järjestyksessä olevat virkkeet ja kappaleet			Korp	A		"



CLARIN -lisenssiluokat



julkisesti saatavilla



akateemisten, kirjautuneiden käyttäjien
käytettävissä



vaatii henkilökohtaisen käyttöluvan



Lisenssiehtojen yksityiskohdat



+BY

alkuperäisen tekijän nimi mainittava

+NC

vain epäkaupallinen käyttö

+ID

vaatii kirjautumisen



+PLAN

vaatii tutkimussuunnitelman

+PRIV

sisältää henkilötietoja



+NORED

ei uudelleenjakelua

+DEP

muokattuja versioita voi jakaa CLARINissa

sekä tarvittaessa muut aineistokohtaiset
lisäehdot

Viittausohjeen löytäminen

Digilib-1920-dl	Kansalliskirjaston sanoma- ja aikakauslehtikokoelman OCR-korpus (1875-1920)	ACA	Lataus	 Viittausmuoto How to cite
Digilib-Pub-1874-dl	Kansalliskirjaston sanoma- ja aikakauslehtikokoelman OCR-korpus (1771-1874)	PUB	Lataus	

Viittausohje: Digilib-1920-dl

Viittaa kielivaraan näin:

Kansalliskirjasto (2011). *Kansalliskirjaston sanoma- ja aikakauslehtikokoelman OCR-korpus (1875-1920)* [tekstikorpus]. Kielipankki. Saatavilla <http://urn.fi/urn:nbn:fi:lb-201801192>

[Bibtex] [Zotero]

Kuvailutiedot

 META SHARE

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

 302  8  2

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

► 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](#)



KIELIPANKKI KÄYTTÄJÄKSI AINEISTOT TYÖKALUT ORGANISAATIO TUKI

IN ENGLISH PÅ SVENSKA

Twitter #lib_suomi24

Suomi 24

Viimeisimmät versiot:

Suomi24 virkkeet -korpus 2001-2017, Korp-versio 1.1

📄 Kuvailutiedot ja lisenssi

📄 Viittausohje tähän versioon

Suomi24-korpus 2001-2017, VRT-versio 1.1

📄 Kuvailutiedot ja lisenssi

📄 Viittausohje tähän versioon

Etsi muut saatavilla olevat versiot

🔗 Avaa aineisto Korp-palvelussa ?

🔗 Lataa aineisto

Aineisto koostuu Suomi 24 -foorumilta kerätyistä keskusteluista. Sisältö on jäsennetty automaattisin menetelmin ja tallennettu VRT-muotoon.

Korpin kautta tarjottavasta Suomi 24 -korpuksella voi tehdä monipuolisia hakuja ja tilastoida tai kuvantaa hakutuloksia eri tavoin (katso Korp-palvelun ohjeet).

Hae Kielipankki-portaalista:

Hae



Kuukauden tutkija: Saana Svärd

Uutisia

- Helsingin yliopiston uutiset: Lekuri on skidien gaiffari – lahjoita sinäkin puhetta ja tekoäly ymmärtää pian myös murteita (29.10.2020)
- Save the date: 15.12.2020 tulossa ELRC- ja ELG-työpajat (27.10.2020)
- Vielä ehdit mukaan verkkokurssille Puheen

www.kielipankki.fi/aineistot

SSDC-2016	Koltansaamen dokumentaatiokorpus (2016)	RES	↗	LAT	B	”
stt-fi-1992-2018-src	STT:n uutisarkisto 1992-2018, lähdemateriaali	RES	↗	Lataus	A	”
Suomi24-2001-2014-dl	Suomi 24 2001-2014 (näyte) -korpuksen ladattava versio	ACA	↗	Lataus	A	? ”
Suomi24-2001-2014-korp	Suomi 24 2001-2014 (näyte) -korpuksen Helsinki-Korp-versio	PUB		Korp	A	? ”
Suomi24-2001-2015	Suomi 24 2001-2015 (näyte) -korpus	ACA		Lataus	A	? ”
suomi24-2001-2017-korp-v1-1	Suomi24 virkkeet -korpus 2001-2017, Korp-versio 1.1	PUB		Korp	A	? ”
suomi24-2001-2017-vrt-v1-1	Suomi24 -korpus 2001-2017, VRT-versio 1.1	ACA		Lataus	A	”
Suomi24-2015H1	Suomi 24 -korpus (2015H1)	ACA		Lataus	A	? ”
Suomi24-2016H2	Suomi 24 -korpus (2016H2)	ACA		Lataus	A	? ”
Suomi24-korp-2016H2	Suomi 24 virkkeet -korpus (2016H2)	PUB		Korp	A	? ”
Suomi24-virkkeet-2015H2	Suomi 24 virkkeet -korpus (2015H2)	ACA		Korp	A	? ”



Aineiston kuvailutiedot

The Suomi24 Sentences Corpus 2001-2017, Korp version 1.1

112 ✓ 2

▼ View resource name in all available languages

Suomi24 virkkeet -korpus 2001-2017, Korp-versio 1.1

*suomi24-2001-2017-korp-v1-1*Persistent Identifier of this resource: <http://urn.fi/urn:nbn:fi:lb-2020021803>Access location: <http://urn.fi/urn:nbn:fi:lb-2020021804>

The corpus contains all the discussion forums of the Suomi24 online social networking website from 1st January 2001 to 31st December 2017 available in the Suomi24 API.

Researchers can download the entire corpus (see <http://urn.fi/urn:nbn:fi:lb-2020021801>).

► View resource description in all available languages

« Back

Edit Resource

Aineiston sijainti

Distribution

Availability

Available - Restricted Use

Licence

CC - BY - NC

text

Monolingual text corpus

Languages

Finnish

Linguality

Linguality type: Monolingual

Metadata

Created: 02/18/2020

Last Updated: 02/18/2020

Metadata Language: English, Finnish (en, fi)

Metadata Creator

corp.csc.fi: Suomi24 (2001-2017)

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit

Viittaa Korpiin Suomi | Svenska | English 41



Yksinkertainen Laajen

myös ☐ alkuosa ☐ loppuos

virkkeistä jotka sisä

Konkordanssi:

17 / 975 korpusta valittuina — 4,13G / 13,61G sanetta



☒ Valitse kaikki
☒ Tyhjennä

☐ Agricola (9)
☐ E-thesis (24)
☐ FinnTreeBank: suomen puupankki (3)
☐ Kansalliskirjaston lehtikokoelman (KLK) suomenkieliset lehdet (179)
☐ Kirjallisuutta (22)
☐ Juridisia tekstejä (6)
☒ Internet-keskusteluaineistoja (30)

☒ Suomi24 2001-2017 (17)
☐ Suomi24 2016H2 (10)
☐ Suomi24 2001-2014 (näyte)
☐ Yliilauta
☐ HS.fi [ACA]

☐ 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä (369)
☐ Verkkouutisia (16)
☒ Suomen kielen tekstikokoelma (SKTP/FTC) (50)
☐ Muita tekstejä (13)
☐ Puhuttua kieltä (tekstiksi litteroituna) (137)
☐ Selkokieli (4)

e: sana

☐ Näytä sanakuva ☐ Näytä kartta

Suomi24 2001-2017

Suomi24 virkkeet -korpus 2001-2017, Korp-versio 1.1
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 2001-2017 -aineiston](#) Kielipankin latauspalvelusta (lisenssi).

(Tämä aineisto näkyi beetestausvaiheessa Korpissa nimellä *Suomi24 virkkeet -korpus (2017H2)* ja aineiston ensimmäinen versio nimellä *Suomi24 2017H2*.)

2020-02-20: Aineistosta on nyt käytössä korjattu versio 1.1. Se sisältää aiemmasta versiosta vuosilta 2009-2012 ja 2014 puuttuneet kirjoittajan



Yksinkertainen

Laajennettu

Edistynyt

Vertailu

1

korp.csc.fi: Suomi 24 (2001-2017)

Laajennettu haku:
terveys

(eri muotojen esiintymät)

perusmuoto on
terveys Aa
tai
+

Hae

virkkeen

sisältä

Konkordanssi: osumia sivulla: 25

järjestä korpuksen sisällä: järjestämätön

Tilastoja: laskentaperuste: aihealue

Näytä sanakuva

Näytä kartta

Konkordanssi

Tilastoja

Sanakuva

Kartta

Tuloksia: 269 273

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

SUOMI24 2001-2017: 2001

Maksavathan ihmiset monesta turhemmastakin kuin omasta terveydestään.

Miksi pitää ulkomailta raahata tänne suomeen löytökoiria, joiden luonteesta ja terveydestä ei ole mitään takuita, kun kotomaassamme on kymmenittäin mukavia ja terveitä löytö

ollisen altaan alueella elävien kanssa, monet ovat eläneet vuosikausia sietämättömän unen, terveyden varmpia?

Sen olen oivaltanut, että kaikki lähtee jokaisen omasta sisimmistä ja se muovaa ihmisen ja kaiken toivon menettäneen epävarmuuden alla.

ppeelee) ja uros+naaras-yhdistelmä tuottaa niin suunnattomia määriä poikasia, että naaraan terveyden ja ympäristön.

Kiva, jos terveyteen (=oikeanlainen liikunta ja ravinto).

Kärsii kuin myös sinun mielenrauha etsiessä niille hyviä koteja.

on nuiilta osin kunnossa.

Hellitä heti tai menee loputkin terveyden rippeistäsi.

Ego =ihmisen minuus, jota ilman ihminen ei ole ihminen (joten sinnikkyys on vain terveyden merkki)

> Ego =ihmisen minuus, jota ilman ihminen ei ole ihminen (joten sinnikkyys on vain terveyden merkki) >

Terveiden määritelmät ovat pikkuisen laajempialaisia ja syvempiä

Kukaan ei ole ilmoittautunut halukkaaksi, mikä onkin järkevää heidän terveytensä kannalta.

Jokainen tervejärkinen nainen haluaa huolehtia omasta terveydestään.

Valtion rahoittama markkinointikampanja vaarantaa kansalaisten terveyttä ja manipuloi ihmiset jo elinaikanaan asumaan ruumisarkuissaan, kuten monista surullis

nen elämässään ymmärtää usein arvostaa jotakin asiaa vasta, kun on sen menettänyt (esim. terveys).

Olen sitä mieltä, että ihminen ei kertakaikkiaan voi arvostaa terveyttään kun se on ja että se on aivan oikein tehty.

Tämä ei ollenkaan sulje pois sitä, etteikö omasta ruumistaan ja terveydestään voisi hetkittäin nauttia suunnattomastikin, esimerkiksi kesäiltana saunan jälkeen verar

olen pietistiseen syyllisyyteen lapsena ympätty ja se ihan varmaan vaikuttaa tähän * arvosta tai muita elämäsi ominaisuuksia, kun sinulla ne on.

..eli en ole tosiaankaan vielä tilannut näitä Viron kiekkoja, terveyttäsi tai otan sen pois * -käsitykseen.

terveys (siis kalojen) ja mahd. pitkä elinikä menevät tietysti halvemman hinnan edelle!

Korpus

Suomi24 2001-2017: 2001

Osa aineistoa: Suomi24 2001-2017

Aineiston pysyvä tunniste (kuvaileutiedot):

urn:nbn:fi:lb-2020021803

Viittaa aineistoon

Lisenssi: CC BY-NC (CLARIN PUB)

Linkki aineistoon Korpissa

Tekstin piirteet

otsikko: Liian paksu perhoseksi(nainen)

päiväys: 2001-01-02

kellonaika: 10:35:00

kirjoittajan nimimerkki: Ex-samanlainen

-63

kirjautunut käyttäjä: ei

rekisteröity nimimerkki: ei

aihealue: Terveys > Henkinen hyvinvointi

ja mielenterveys > Yleistä

mielenterveydestä

aihealue vain aikuisille: ei

viestin tyyppi: kommentti

viesti on täysin tyhjä: ei

tunniste: 43752:211676

perusmuoto on

terveys Aa

tai

+ -

Hae virkkeen sisältä

Konkordanssi: osumia sivulla: 25 järjestä korpuksen sisällä: järjestämätön Tilastoja: laskentaperuste: aihealue Näytä sanakuva Näytä kartta

Arvoja: 1 854

Näytä trendikuvaaja

☐ aihealue	Yhteensä	Suomi24 20...	Suomi24 20...	Suomi24 20...	Suomi24 20...	Suomi24 20...
☒ Σ	65,2 (269 273)	37,3 (1 329)	44,9 (2 942)	44,4 (4 914)	49 (9 982)	55,3 (17 401)
☐ Yhteiskunta > Maailman menoa	2,8 (11 715)	0,8 (27)	1 (67)	1 (110)	0,8 (165)	1 (326)
☐ Terveys > Painonhallinta > Laihdutus > Vähähiilihydraattiset dieetit	2,1 (8 761)	0 (0)	0 (1)	0 (0)	0,1 (21)	0,9 (277)
☐ Terveys > Sairaudet > Homesairaudet	1,7 (6 956)	0 (0)	0 (1)	0 (3)	0,2 (39)	1,4 (443)
☐ Yhteiskunta > Uskonnot ja uskomukset > Kristinusko > Lestadiolaisuus > Vanh...	1,2 (4 919)	0 (0)	0 (0)	0 (0)	1,4 (280)	1,6 (518)
☐ Suhteet > Sinkut	1,1 (4 368)	0 (0)	0,3 (19)	0,2 (22)	0,2 (31)	0,3 (93)
☐ Ryhmät > Ikäryhmät > 60 plus	1 (4 267)	0,3 (11)	0,6 (37)	0,8 (84)	0,4 (78)	0,4 (122)
☐ Lemmikit > Koirat > Yleistä koirista	0,9 (3 831)	4,9 (175)	2,5 (162)	1,3 (144)	1 (196)	1,7 (529)
☐ Terveys > Terveiden hoito > Vaihtoehtoiset hoidot > Vaihtoehtoiset hoidot - yl...	0,9 (3 776)	0 (0)	0 (0)	0 (0)	0 (5)	0,2 (62)
☐ Terveys > Painonhallinta > Laihdutus > Yleistä laihdutuksesta	0,8 (3 477)	1,7 (61)	1,7 (109)	2 (217)	1,3 (258)	1,1 (348)
☐ Yhteiskunta > Uskonnot ja uskomukset > Kristinusko > Helluntailaisuus	0,8 (3 341)	0 (0)	0 (0)	0 (3)	0,4 (80)	0,9 (271)
☐ Perhe > Raskaus > Abortti	0,8 (3 148)	0 (0)	0,2 (13)	0,2 (27)	0,4 (73)	0,4 (116)
☐ Yhteiskunta > Uskonnot ja uskomukset > Kristinusko > Adventismi	0,8 (3 145)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

korp.csc.fi: Suomi 24 (2001-2017)

terveys



17 / 975 korpusta valittuina — 4,13G / 13,61G sanetta



sairaus..nn.1

Yksinkertainen

Laajennettu

Edistynyt

Vertailu 1

perusmuoto on
terveys Aa
tai

Hae virkkeen sisältä

Konkordanssi: osumia sivulla: 25 järjestä korpuksen sisällä: järjestämätön Tilastoja: laskentaperuste: aihealue

Näytä sanakuva

Näytä kartta

Konkordanssi

Tilastoja

Sanakuva

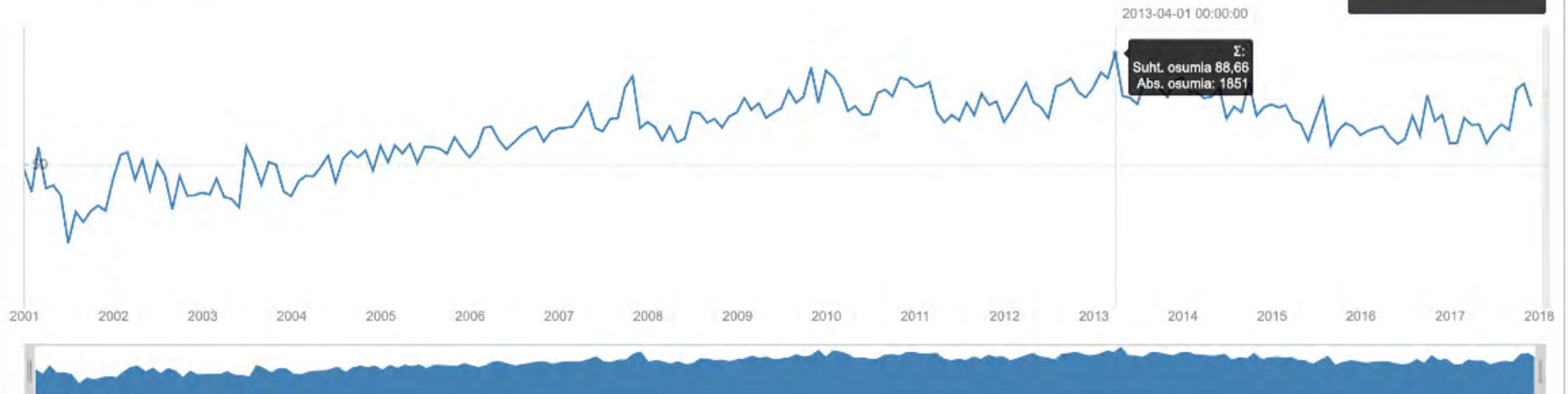
Kartta

Kuvaaja

Viiva

Pylväs

Taulukko



Yksinkertainen

Laajennettu

Edistynyt

Vertailu

2

korp.csc.fi: Suomi 24 (2001-2017)

Laajennettu haku:
sairaus

(eri muotojen esiintymät)

perusmuoto on
sairaus Aa
tai

Hae virkkeen sisältä

Konkordanssi: osumia sivulla: 25 järjestä korpuksen sisällä: järjestämätön Tilastoja: laskentaperuste: aihealue

☐ Näytä sanakuva

☐ Näytä kartta

Konkordanssi

Tilastoja

Sanakuva

Kartta

Tuloksia: 526 398

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

SUOMI24 2001-2017: 2001

Sairaus on helppo hoitaa kun siihen löytyy selvä syy kuten esim huumeiden käyttö.
sairauden huomaa ehkä liian myöhään... = (Virikkeistä... esteet on hyvä ajatus.
sairaus !
sairauden pois-sulkemiseksi.
sairaus , etenkin valkoisilla.Kysy kasvattajalta onko tätä ongelmaa suvussa.Koira on jääräpäinen
sairauksia siellä edes tutkimaan... onko eläinlääkärit niin ammattitaitoisia.
sairauden oireet niin kannattaa ehdottomasti ottaa yhteyttä eläinlääkäriin eikä kysyä " päteviä2 n
sairautesi ja tuskasi toisille, niille joita vihaat.
sairauksista , jotka lienevät perinnöllisiä.
sairauksia esiintyy paljon.
sairauksia (koirani nyt n4v.)Osta koira kasvattajalta joka on tarkistuttanut koiran vanhemmat!!!
sairauksiin , joskin kyllä tiedän niiden olemassa olost.
sairautta jolla voi olla tosi ikävät seuraukset.Koin itse sellaisen suhteen-en voinut lopulta edes h
sairauksia ja kehitysvammaisuutta.
sairaus tahi kehitysvammaisuus johtuu, ollaan eittämättä tuhoisilla ja egoistisella tiellä.
sairauksien uhka saadaan geeniterapialla poistettua.
sairaudet , auton varastaneet ym.Joillekin jo 40Km/h on liian suuri nopeus kykyihin nähden.
sairauksiin kannattaa tutustua, etteivät tule yllätyksenä.
sairaudesta ja tästä lääkkeestä mikä tappaisi collien?
sairaudesta , vaikka sähköpostitse?
sairauksista sekä voivat olla alttiita tapaturmille 6) Jousimiehet ja Kalat:ovat henkeviä ja yhteistyök

Korpus

Suomi24 2001-2017: 2001

Osa aineistoa: Suomi24 2001-2017
Aineiston pysyvä tunniste (kuvailutiedot):
urn:nbn:fi:lb-2020021803

[Viittaa aineistoon](#)

Lisenssi: CC BY-NC (CLARIN PUB)

[Linkki aineistoon Korpissa](#)

Tekstin piirteet

otsikko: Liian paksu perhoseksi(nainen)

päiväys: 2001-01-02

kellonaika: 10:35:00

kirjoittajan nimimerkki: Ex-samanlainen
-63

kirjautunut käyttäjä: ei
rekisteröity nimimerkki: ei
aihealue: Terveys > Henkinen hyvinvointi
ja mielenterveys > Yleistä
mielenterveydestä
aihealue vain aikuisille: ei
viestin tyyppi: kommentti
viesti on täysin tyhjä: ei
tunniste: 43752:211676



17 / 975 korpusta valittuina — 4,13G / 13,61G sanetta



sairaus..nn.1

Yksinkertainen Laajennettu Edistynyt Vertailu 2

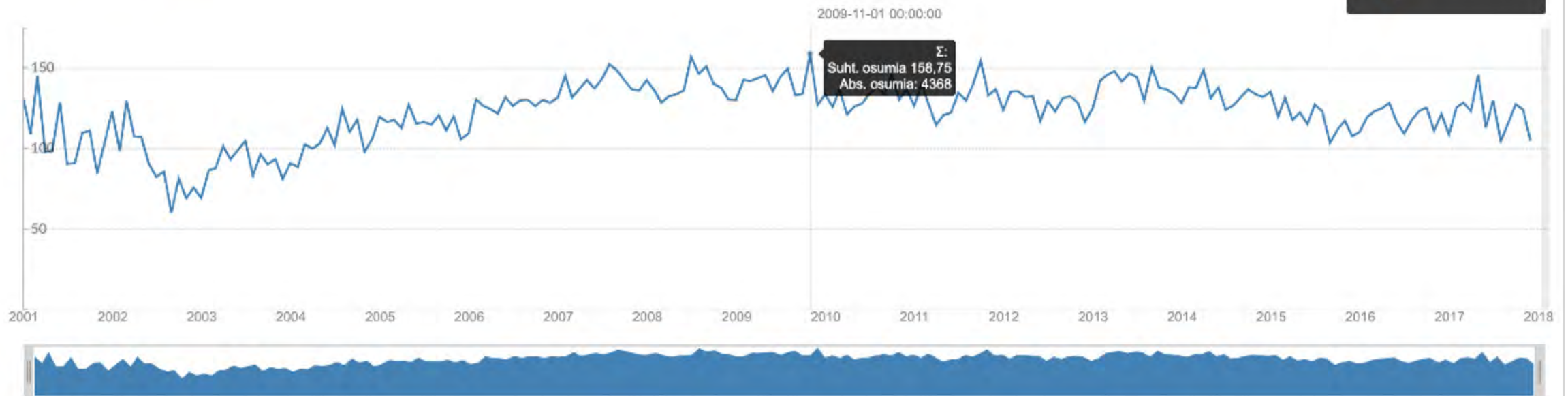
perusmuoto on
sairaus Aa
tai

Hae virkkeen sisältä

Konkordanssi: osumia sivulla: 25 järjestä korpuksen sisällä: järjestämätön Tilastoja: laskentaperuste: aihealue Näytä sanakuva Näytä kartta

Konkordanssi Tilastoja Sanakuva Kartta Kuvaaja

Viiva Pylväs Taulukko



Vertaa hakua

Vertaa hakua

Konkordanssi: Tilastoja: ☐ Näytä sanakuva ☐ Näytä kartta

Konkordanssi Tilastoja Sanakuva Kartta Kuvaaja Vertaa hakua

Korostuneita haussa *terveys*

[Terveys](#) > [Painonhallinta](#) > [Laihdutus](#) > [Vähähiilihydraattiset](#) **8761**

[Terveys](#) > [Painonhallinta](#) > [Laihdutus](#) > [Yleistä laihdutuksesta](#) **3477**

[Koti ja rakentaminen](#) > [Rakentaminen ja remontointi](#) > [Kotivaurio](#) **1981**

[Yhteiskunta](#) > [Uskonnot ja uskomukset](#) > [Horoskoopit ja astrologia](#) **1643**

[Urheilu ja kuntoilu](#) > [Voimailulajit](#) > [Kehonrakennus](#) **1003**

[Terveys](#) > [Päihteet ja riippuvuudet](#) > [Tupakointi](#) **1404**

[Suhteet](#) > [Sinkut](#) **4368**

[Yhteiskunta](#) > [Maailman menoa](#) **11715**

[Perhe](#) > [Raskaus](#) > [Abortti](#) **3148**

[Terveys](#) > [Painonhallinta](#) > [Laihdutus](#) > [Laihduttajien tukiryhmät](#) **889**

[Urheilu ja kuntoilu](#) > [Kuntoilu ja kuntosalit](#) **646**

[Terveys](#) > [Suun terveys](#) **1173**

[Terveys](#) > [Plastiikkakirurgia](#) **1042**

[Urheilu ja kuntoilu](#) > [Voimailulajit](#) > [Paininta](#) **449**

Korostuneita haussa *sairaus*

[Terveys](#) > [Henkinen hyvinvointi ja mielenterveys](#) > [Kaksisuuntainen mielialahäiriö](#) **6978**

[Terveys](#) > [Henkinen hyvinvointi ja mielenterveys](#) > [Yleiset mielenterveydestä](#) **14292**

[Terveys](#) > [Henkinen hyvinvointi ja mielenterveys](#) > [Masennus](#) **12631**

[Yhteiskunta](#) > [Uskonnot ja uskomukset](#) > [Kristinuskon kirkon laillisuus](#) **14251**

[Terveys](#) > [Sairaudet](#) > [Neurologiset sairaudet](#) > [Neurologiset sairaudet](#) **3387**

[Terveys](#) > [Sairaudet](#) > [Syömishäiriöt](#) > [Anoreksia](#) **6226**

[Yhteiskunta](#) > [Uskonnot ja uskomukset](#) > [Kreationismi ja evoluutio](#) **5399**

[Terveys](#) > [Sairaudet](#) > [Neurologiset sairaudet](#) > [Epilepsia](#) **3648**

[Terveys](#) > [Henkinen hyvinvointi ja mielenterveys](#) > [Paniikkihäiriö](#) **3596**

[Terveys](#) > [Sairaudet](#) > [Syöpä](#) **4723**

[Terveys](#) > [Sairaudet](#) > [Sisätaudit](#) > [Reuma](#) **2786**

[Terveys](#) > [Henkinen hyvinvointi ja mielenterveys](#) > [Skitsofrenia](#) **2333**

[Terveys](#) > [Kuolema ja suru](#) **3114**

[Terveys](#) > [Sairaudet](#) > [Neurologiset sairaudet](#) > [MS-tauti](#) **1894**

Hakuvertailu

terveys vs. sairaus
aihealueen mukaan

corp.csc.fi: Suomi24 (2016H2)

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit Viittaa Korpiin

 10 / 975 korpusta valittuina — 2,66G / 13,61G sanetta

Yksinkertainen Laajen

myös ☐ alkuosa ☐ loppuos

virkkeistä ☒ jotka sisä

Konkordanssi: osumia sivulla:

1600 1700 1800 1900 2000

☒ Valitse kaikki
☒ Tyhjennä

- ☐ Agricola (9)
- ☐ E-thesis (24)
- ☐ FinnTreeBank: suomen puupankki (3)
- ☐ Kansalliskirjaston lehtikokoelman (KLK) suomenkieliset lehdet (179)
- ☐ Kirjallisuutta (22)
- ☐ Juridisia tekstejä (6)
- ☒ Internet-keskusteluaineistoja (30)
 - ☐ Suomi24 2001-2017 (17)
 - ☒ Suomi24 2016H2 (10)
 - ☐ Suomi24 2001-2014 (näyte)
 - ☐ Ylilauta
 - ☒ HS.fi [ACA]
- ☐ 1990- ja 2000-luvun suomalaisia aikakaus- ja sanomalehtiä (369)
- ☐ Verkkouutisia (16)
- ☒ Suomen kielen tekstikokoelma (SKTP/FTC) (50)
- ☐ Muita tekstejä (13)
- ☐ Puhuttua kieltä (tekstiksi litteroituna) (137)
- ☐ Selkokieli (4)

e: sana ☐ Näytä sanakuva ☐ Näytä kartta

 Suomi24 2016H2

Suomi24-keskustelupalvelun keskustelut 1.1.2001-24.9.2016.

Huomaa, että aineisto ei ole kattava mainitulta aikaväliltä. **Käytä ensisijaisesti uudempaa ja selvästi kattavampaa Suomi24 2017H2 -aineistoa**, jollet tarvitse vertailukelpoisuutta tähän aineistoon perustuvan tutkimuksen kanssa.

Aineistossa näkyy kaikkien keskustelujen sisältö enintään kappaletasolla. Aineisto on jaettu useaan osakorpukseen suuren kokonsa vuoksi. Tutkijat voivat myös ladata käyttöönsä [koko Suomi24-aineiston](#) Kielipankin latauspalvelusta (lisenssi).

corp.csc.fi > Suomi24 (2016H2)

Suomi | Ruotsi | Muut kielet | Rinnakkaistekstit Viittaa Korpiin

 10 / 975 korpusta valittuina — 2,66G / 13,61G sanetta 

Yksinkertainen Laajennettu Edistynyt Vertailu

terveys Hae ▾

- terveys (substantiivi) 174487
- terveyskeskus (substantiivi)
- terveyspalvelu (substantiivi)
- terveysasema (substantiivi)
- terveyshaitta (substantiivi)
- terveysministeriö (substantiivi)
- terveyskeskuslääkäri (substantiivi)
- terveysala (substantiivi)
- terveystarkastaja (substantiivi)
- terveysvaikutus (substantiivi)
- terveysriski (substantiivi)
- terveystarkastus (substantiivi)
- terveysongelma (substantiivi)
- terveys (adverbi)
- terveys (adjektiivi)

ouksen sisällä: järjestämätön ▾ Tilastoja: laskentaperuste: sana ☐ Näytä sanakuva ☐ Näytä kartta



Yksinkertainen

Laajennettu

Edistynyt

Vertailu

terveys (substantiivi)

Hae

myös ☐ alkuosa ☐ loppuosa ja ☐ samaista pien- ja suuraakkoset

virkkeistä jotka sisältävät

corp.csc.fi > Suomi24 (2016H2) Sanakuva

Konkordanssi: osumia sivulla: 25 järjestä korpuksen sisällä: järjestämätön Tilastoja: laskentaperuste: sana ☒ Näytä sanakuva

Konkordanssi

Tilastoja

Sanakuva

Kartta

Etumääräite

terveys

Jälkimääräite

terveys

verbi

Verbi

terveys

1. ihminen	5522	1. nähtenne	92
2. oma	9834	2. tallella	109
3. lapsi	3537	3. huolto	76
4. fyysinen	1925	4. vaara	71
5. henkinen	1331	5. koti	71
6. koira	1278	6. keskus	56
7. psyykkinen	506	7. huoli	71
8. mieli	582	8. haitalli	25
9. kansalainen	577	9. ruuma	52
10. äiti	692	10. takaa	48
11. potilas	378	11. uhallalla	28
12. kissa	487	12. kunto	67
13. väestö	288	13. etusija	24
14. vauva	386	14. viis	34
15. työntekijä	369	15. lehti	49

1. mennä	3050
2. pettää	846
3. kestää	1213
4. vaarantua	295
5. reistailla	244
6. koheta	227
7. vaarantaa	239
8. sallia	334
9. kohentua	131
10. kärsiä	405
11. säilyä	297
12. heiketä	181
13. parata	297
14. romahtaa	215
15. reistata	66

1. vaarantaa	2984
2. edistää	2702
3. riskeerata	770
4. pilata	1331
5. menettää	1612
6. tuhota	1291
7. uhata	629
8. vahingoittaa	572
9. hoitaa	1302
10. ylläpitää	425
11. uhrata	369
12. parantaa	498
13. vaalia	198
14. ajatella	922
15. turmella	154

terveys

Yksinkertainen

Laajennettu

Edistynyt

Vertailu

sairaus (substantiivi)

Hae

myös ☐ alkuosa ☐ loppuosa ja ☐ samaista pien- ja suuraakkosetvirkkeistä ☒ jotka sisältävät

corp.csc.fi > Suomi24 (2016H2) Sanakuva

Konkordanssi: osumia sivulla: 25

järjestä korpuksen sisällä: järjestämätön

Tilastoja: laskentaperuste: sana

☒ Näytä sanakuva

Konkordanssi

Tilastoja

Sanakuva

Kartta

Etumääräite

sairaus

Jälkimääräite

sairaus

verbi

Verbi

sairaus

1. vakava	11936	1. aiheuttaa1005
2. perinnöllinen	4827	2. lukuisä 276
3. psyykkinen	3831	3. parantua 404
4. krooninen	2366	4. parata 351
5. fyysinen	3287	5. parantaa 322
6. aiheuttaa	3126	6. hoitaa 553
7. neurologinen	1510	7. estää 392
8. parantumaton	1330	8. sairastaa 190
9. mieli	1932	9. vaatia 468
10. vaikea	2559	10. vaikuttaa 369
11. periytyä	624	11. johtaa 328
12. johtua	1027	12. johtua 293
13. psykiatrinen	844	13. sairastua 148
14. somaattinen	444	14. lääkitä 71
15. harvinainen	1057	15. aiheutua 106

1. aiheuttaa2042
2. edetä 895
3. puhjeta 567
4. paheta 512
5. iskeä 778
6. parata 670
7. johtua 1290
8. runnella 183
9. lisääntyä 609
10. murtaa 224
11. periytyä 233
12. ilmetä 417
13. pahentua 185
14. parantua 320
15. uuvuttaa 123

1. aiheuttaa	4549
2. parantaa	2925
3. sairastaa	2010
4. hoitaa	3249
5. todeta	1450
6. diagnosoida	626
7. kantaa	997
8. ehkäistä	479
9. potea	381
10. laukaista	276
11. pahentaa	336
12. poissulkea	196
13. ennaltaehkäistä	103
14. hoidattaa	105
15. periytyä	84

sairaus

Eduskunnan täysistuntojen pöytäkirjat videoihin kohdistettuina



Medialähde:

https://down0-80133-down0.dna.qbrick.com/80133-down0/Syysistuntokausi_2008/77._torstaina_18._syyskuuta_2008.mp4

Korpus



Eduskunnan täysistunnot

Aineiston pysyvä tunniste (kuvaileutiedot):

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

Lisenssi (ääni ja video): CLARIN PUB +BY +PRIV +ND +OTHER 1.0

Lisenssi (teksti): CLARIN PUB +BY +PRIV 1.0

[Linkki Korp-hakuun](#)

Tekstin piirteet

julkaisupäivä: 2008-09-19

julkaisuaika: 06:12:09

Huom. Kaikki selaimet eivät tue tälle sivulle upotettua videota. Varmimmin sivu toimii Firefox-selaimella.

Alkuperäinen litteraatti on kohdistettu mediatiedostoon ja annotoitu automaattisin menetelmin. Tästä syystä kohdistuksessa ja annotaatiolissa saattaa esiintyä virheitä.

Pysäytä video alueen loppuun: ☐

Puhunnos: Uutena rokotteenä Kansalliseen rokotusohjelmaan esitetään lisättäväksi erityisesti lasten **terveyden** kannalta tärkeä rotavirusrokote.

puhujan nimi: Paula Risikko
eduskuntaryhmä: ei tiedossa
puhujan rooli: Peruspalveluministeri
puheenpuheen tyyppi: ei tiedossa

puhunnoksen alkuaika: 2:00:30,664
puhunnoksen loppuaika: 2:00:38,768
puhunnoksen kesto: 0:00:08,104
puhunnoksen alkuaika (ms): 7230664
puhunnoksen loppuaika (ms): 7238768
puhunnoksen kesto (ms): 8104



Pulistaan meille parempia palveluja

Nyt kerätään kaikenlaista puhuttua suomea! Lahjoittamasi puheen puheen avulla esimerkiksi ääniohjatut laitteet voivat oppia ymmärtämään erilaisia murteita ja puhetapoja. Valitse alta aihe, josta haluat puhua.



ELÄINYSTÄVÄT



SUKELLA KESÄÄN



KERRO KORONA-AJASTA

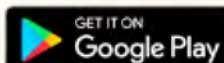
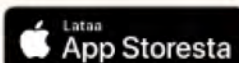


URHEILUHETKET



LÄHELLÄNI JUURI NYT

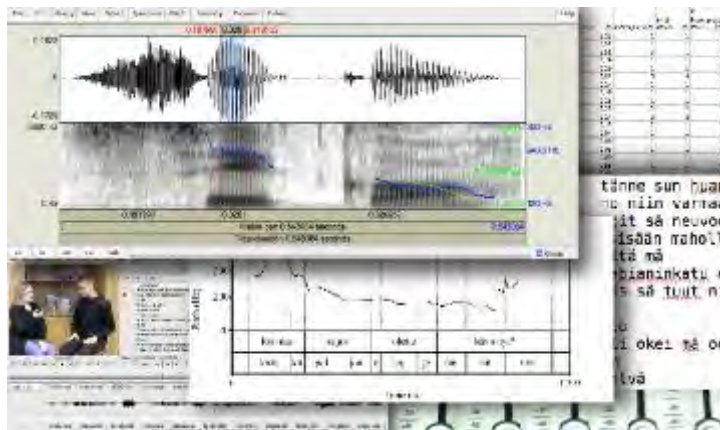
Lahjoittaminen onnistuu parhaiten mobiilisovelluksella:



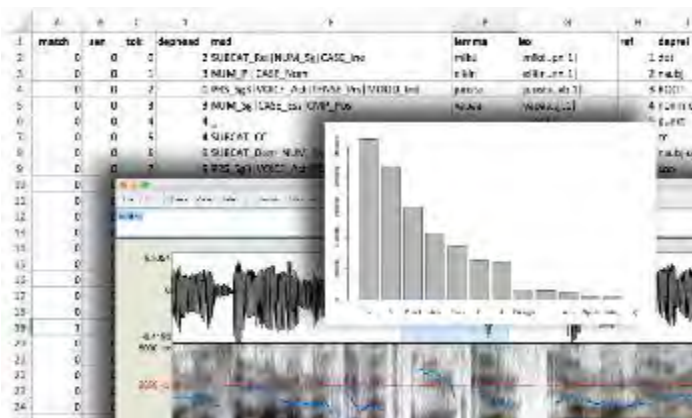
Korpuslingvistiikka ja tilastolliset menetelmät (5 op)



Puheen analyysin perusteet (5 op)

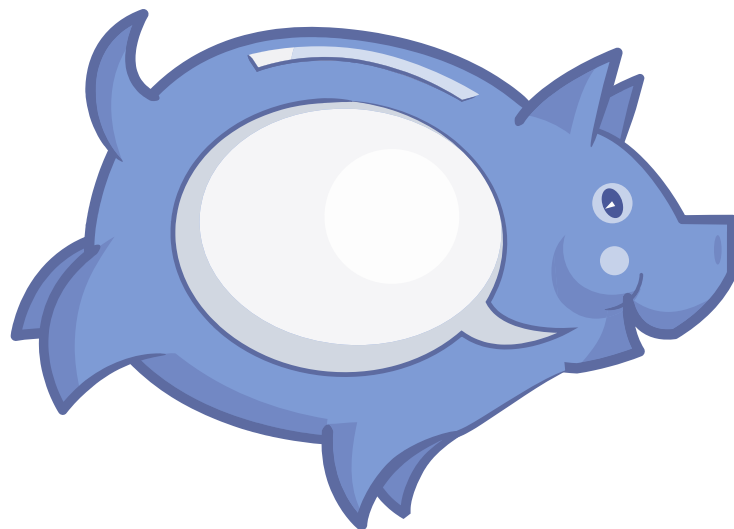


Aineistoklinikka, "Data Clinic" (5 op)



<https://www.kielipankki.fi/tuki/koulutus>

/



Kiitos! Tack! Thank you!

www.kielipankki.fi

Yleinen käyttäjätuki

fin-clarin@helsinki.fi

Tekniset palvelut ja käyttöluvat

kielipankki@csc.fi