

# CLARIN

Using language research infrastructure

[www.clarin.eu](http://www.clarin.eu)

Prof. Koenraad De Smedt

University of Bergen / CLARIN Board of Directors /  
Chair of the National Coordinators' Forum

Aberdeen, Digital Humanities Workshop, June 15-16, 2017



# CLARIN MISSION

“CLARIN is committed to establishing an integrated and interoperable research infrastructure of language resources and its technology. It aims at lifting the current fragmentation, offering a stable, persistent, accessible and extendable infrastructure and therefore enabling eHumanities.”

*Making it easier to use digital language data in your research.*

# Digital language data, tools and services

- Textual editions and historical archives
- Linguistically annotated corpora
- Lexical and terminological resources
- Grammars and parsers
- Sociolinguistic survey data
- Psycholinguistic experimental data
- etc.

Not a simple digital library, but language data coming from research or enriched by and for research

# Example of language data for research

Elan - CNGT0429.eaf

File Edit Annotation Tier Type Search View Options Window Help

Video Recognizer Metadata Controls

Volume: 100

Rate: 100

00:00:50.247 Selection: 00:00:10.460 - 00:00:10.670 210

Selection Mode Loop Mode

0:43.000 00:00:44.000 00:00:45.000 00:00:46.000 00:00:47.000 00:00:48.000 00:00:49.000 00:00:50.000 00:00:51.000 00:00:52.000

GlosL S1 [60]

GlosR S1 [129]

GlosL S2 [105]

GlosR S2 [338]

SS\_Head S1 [0]

SS\_Head S2 [0]

SS\_Mouth S1 [0]

SS\_Mouth S2 [0]

SignSpeakZin S1 [23]

SignSpeakZin S2 [46]

WANN NEDE DOC ONDE OF BO HELPEN HO IX: NI

WANN NEDE GEWO HO DOC K ONDE OF BO SCH LO HELPEN LOG PR HO IX: NI

Wanneer schrijven van het Nederlands (aan bod komt), kan In NmG of door op het bord te schrijven. Logopedie helpt ook. Bij logopedie leer je spreken

# Example of language research data

[ Facs ]

[ms: fol. 1r]

1 her hæfr bzeḟ þætt fem magnus k̄n̄gr fende till alra  
2 nozogs māna vm ipan loghbōkr oc um græinir  
hennaz

3 fua vattande fem her fægir nu

3 M-  
4 AGnVS

7 mEð guðs miſkūn nozogs k̄n̄gr

8 fun hakonar kn̄gs funar funar

9 fuærris kn̄gs fænder Jone ærkib̄pi

10 oc allum aḏzum guðs v̄inum oc ſinū

11 i nozogs kn̄gs rikí Q · Gf oc ſína · þer vituð at hi-

12 nir ſkilrikaſto mēn j nozeghe haḟa iðulegha geteð þ̄

13 off at þer haḟet ſpurt at ver haḟum lutt i att at bō-

[ Dipl ]

[ms: fol. 1r]

1 her hæfr bzeḟ þætt fem magnus *konongr* fende till alra  
2 nozogs *manna* vm ſkipan loghbōkr oc um græinir  
hennaz

3 fua vattande fem her fægir nu

3 M-  
4 AGnVS

7 mEð guðs miſkun̄n nozogs *konongr*

8 fun hakonar *konongs* funar funar

9 fuærris *konongs* fænder Jone ærkibiskupi

10 oc allum aḏzum guðs v̄inum oc ſinum

11 i nozogs *konongs* rikí Q · Gf oc ſína · þer vituð at hi-

12 nir ſkilrikaſto menn j nozeghe haḟa iðulegha geteð þ̄irir

13 off at þer haḟet ſpurt at ver haḟum lutt i att at bō-

# Some common challenges for language data (and other data)

- Not always available for reuse (not accessible to all researchers or unclear licensing conditions)
- Not easy to find because not sufficiently catalogued
- Not sufficiently documented in terms of structured metadata
- Coded and annotated in a multitude of formats
- Not compatible with analysis tools or services
- Not long-time archived and curated (i.e. may disappear)
- Not properly citable

*No trust relation between researchers and repositories*

# User communities working with language data and tools

- Digital humanities
- All branches of linguistics and philology
- Political and Social Sciences
- Culture and Folklore, Anthropology, Archaeology
- Literature
- History
- Translation and Lexicography
- Speech therapy
- Information science and media studies
- Teachers and general public

# CLARIN value proposition

- Coordination at European level of national efforts
- Long term solutions for deploying, connecting and sustaining digital language data and technology services
- Lower the barriers to entry into cutting edge research
- Address the needs of researchers producing data and those wanting to use data
- Towards a connected and multilingual European Research Area for digital research in the humanities and social sciences.

# Emerging multidisciplinary research agendas

Europe's multilinguality as a basis for comparative research of societal and cultural phenomena, as reflected in language use, for instance:

- Migration patterns
- Spreading of ideas and terminology
- Intellectual history
- Linguistic and cultural heritage
- Language variation across periods and regions
- Discourse in social, political, legal etc. spheres

# Example: Networks of texts and people (NTAP)

## Corpuscle :: NTAP :: Concordance

Corpuscle Home  
Getting started  
Documentation  
FAQ  
Publications  
Links

Corpus list

Query

[Concordance](#)

Collocations

Distribution

Word List

Text

Metadata

Overview

Corpus doc.

Upload corpus

Localization

**Basic search** | switch to [Advanced search](#)

green

[Run Query](#)

| [Saved queries ...](#)

| [Save Query](#)

Done. Running time: 0.81 sec. (0.22 CPU sec)

Hit 1 - 30 of 142372

[Previous](#)

[Next](#)

| Go to:

| [Download](#)

(☐ Excel mode)

| Type:

| ☐ Show line filter | A

| Context size:

**match**

age industries to reduce their levels of pollution and develop  
v conventional and alternative technologies, public transport,  
n countries – particularly the theses that relate to the role of  
yday travel behaviour is insignificant. • Thesis 9: People with  
er harm. It would be for those communities already having a  
sport from roads to rail to ports. The goal would be to reduce  
ranges, and politicians everywhere are eager to accrue some  
Boris Johnson's electric cars will not be as  
wable transport through electric vehicles and smart grids will  
apacity of the electric vehicles. This technology is part of the  
clists, and swore to put every resident in walking distance of  
acades, aggressive steps must be taken to increase access to  
I all around the world. It is a symbol of advanced, civic and a  
one I did on Disney theme parks a couple of weeks ago. It is  
as cash and a few environmental initiatives, while WWF gives  
ne hard questions asked about this relationship among other  
olade with Ikea. for services rendered. • Do you know of any

green technologies. This is encouraging news ahead of next month's  
green attitudes, and land-use planning. The theses are related to v  
green attitudes. Some of the fourteen theses are well known within  
green attitudes cast aside those attitudes in their leisure-time travel  
green vision, but lacking the means to carry out critical funding....  
gre [Show context](#)  
gre  
gre word: green  
gre blog-post-id: 50f7399a8f3fbe43aef14643  
gre blog-url: newmobilityagenda.blogspot.com  
gre group-id: 7  
gre date: 2009-01-11  
gre sc-class: a  
gre path: //text/s//  
gre corpus: NTAP  
green claims that deserve closer examination? Email your example:

"green" "alternatives?"

Run Query

Save Query

Done. Running  
CPU sec.)

Show distribution type: relative | counts only ☐ | include structures ☐

of word ☐ ignore case, Δ: 0 filter:

relative to sc-class ☐ ignore case, Δ: 0 filter:

group by - ☐ ignore case, Δ: 0 filter:

and - ☐ ignore case, Δ: 0 filter:

Relative numbers are matches per 100000000 corpus positions matching the vertical value(s).

Page 1/1 of 1×1. | [Download](#)

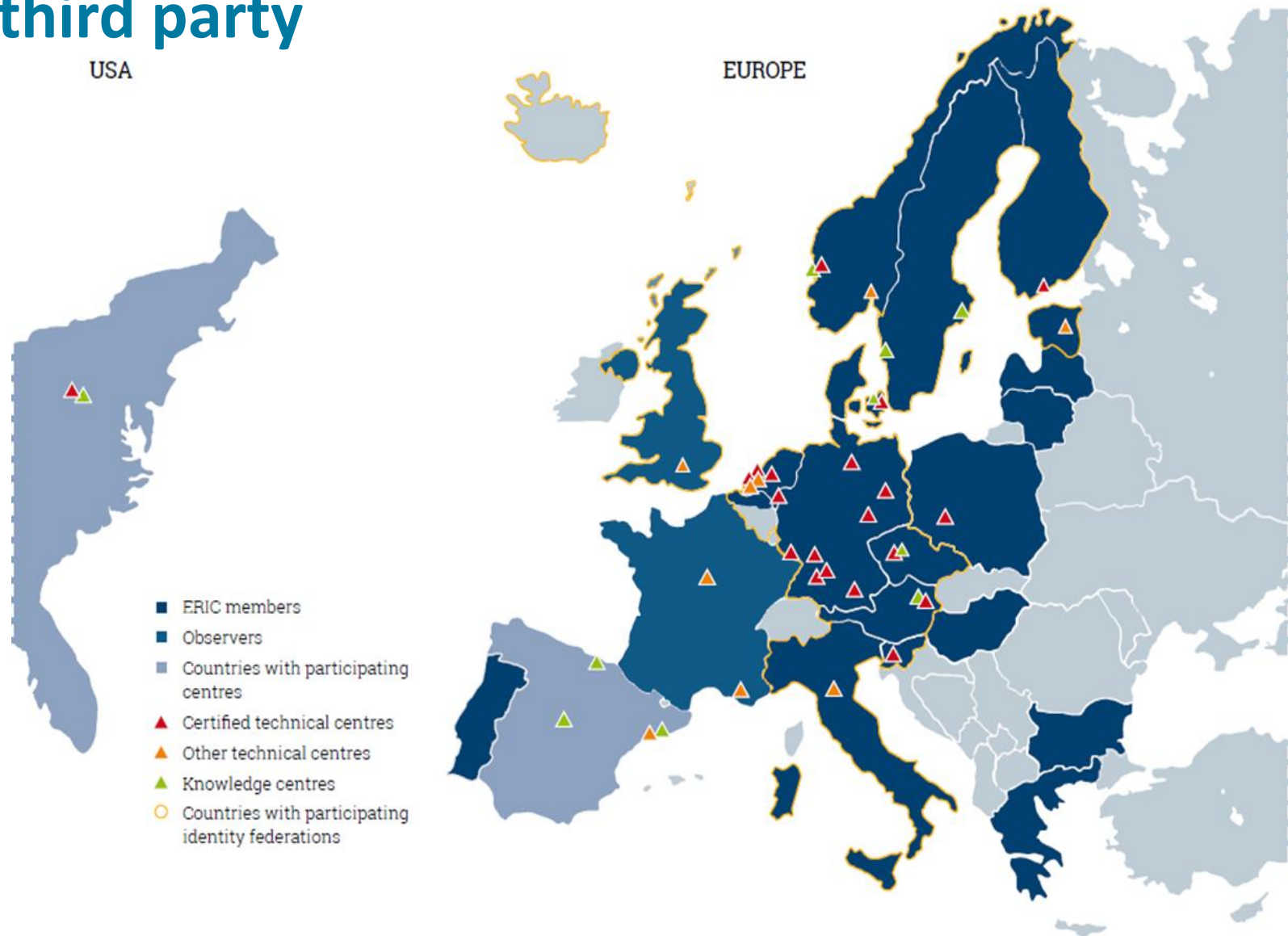
	(sum)	<input checked="" type="checkbox"/> green alternative	<input checked="" type="checkbox"/> green alternatives
(sum)	<b>490</b>	73	417
	<b>490</b>	<b>73</b>	<b>417</b>
Not Classified (125723779)	35 (27,8)	16 (12,7)	19 (15,1)
a (267899110)	414 (154,5)	38 (14,2)	376 (140,4)
n (73250966)	10 (13,7)	3 (4,1)	7 (9,6)
s (177385206)	28 (15,8)	15 (8,5)	13 (7,3)
u (15267439)	3 (19,6)	1 (6,5)	2 (13,1)

# What is CLARIN?

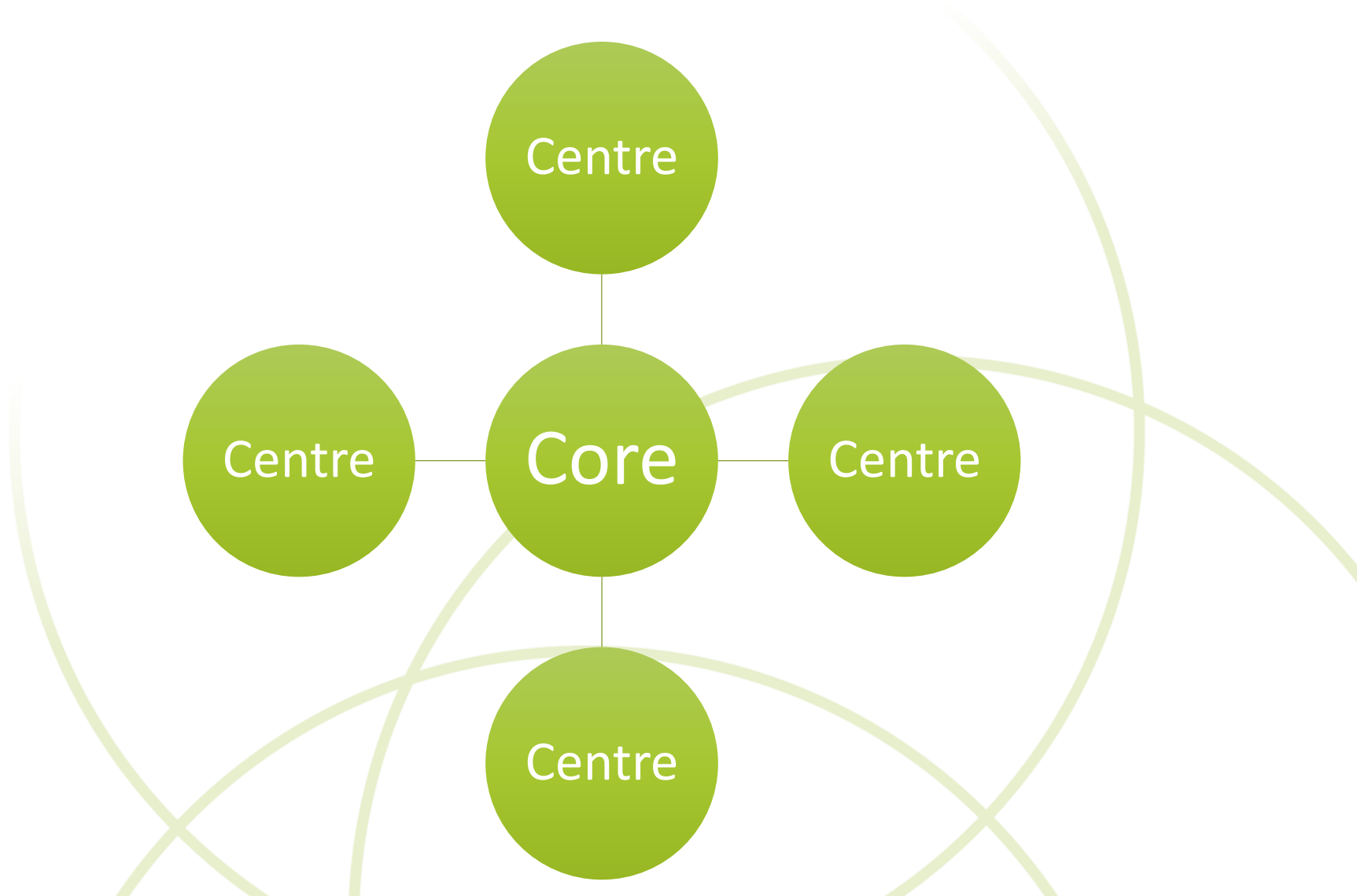
European Research Infrastructure Consortium (ERIC)

- Created by European Law
- Statutes approved by the EC
- CLARIN ERIC was the second ERIC created – in 2012
- Members are countries (not institutions)
- Countries can be observers (pre-membership)
- Third party agreements

# CLARIN ERIC: 19 members, 2 observers, 1 third party



# Network of centres and core services



# CLARIN ERIC activities and financing

- Coordination, core services, technical standards and certifications, legal framework, etc.
- Events: Workshops, CLARIN annual conference, user involvement and training events

## Financing:

- Annual membership fees
- Fee dependent on GNP
- Observers pay smaller fee

# CLARIN national activities and financing

- Network of centres providing specific language data, tools and services
- National support activities and events

## Financing:

- Mostly through national funding mechanisms
- Normal mode: national consortium applies for national funding

# CLARIN data, tools and services

- VLO catalog: almost 900000 metadata records pointing at data sets and tools
- Federated authentication and authorization
- Many specialized services to search in data (corpora, archives etc) including Federated Content Search
- Tools and workflows for language processing
- CLARIN licenses and license categories

# Certified CLARIN Data Centres

- Metadata (CMDI, an ISO standard)
- PID (Handle/DOI),
- Federated single sign-on (eduGAIN)
- Quality of data handling policies (DSA)
- Licensing policies (based on Creative Commons)

# OTA: a CLARIN centre in the UK



## University of Oxford Text Archive

Oxford Text Archive: [Home](#) | [Search](#) | [Catalogue](#) | [TCP](#) | [Deposit](#) | [Oxonly](#) | [Advice](#) | [FAQ](#) | [About](#)

TEI Texts

Corpora

Legacy formats

Search:

Show

10

entries

First

Previous

1

2

3

4

5

Next

Last

ID	Title	Author	Date	Language	Availability	Genre
<a href="#">3193</a>	The Lampeter Corpus of Early Modern English Tracts	Josef Schmied, Claudia Claridge, Rainer Siemund (ed.)	1800	English	CC BY-SA	Linguistic corpora
<a href="#">2563</a>	A Spoken Corpus of Cameroon Pidgin English: pilot study	Dr. Melanie Green, University of Sussex; Dr. Miriam Ayafor, University of Yaounde I; Dr. Gabriel Ozon, University of Sheffield	2016-09-14	Cameroon Pidgin	CC BY-NC-ND 4.0	Linguistic corpora
<a href="#">2562</a>	The Nottingham Corpus of Early Modern German Midwifery and Women's Medicine (ca. 1500-1700)	Richard J Whitt	2016-09-09	German	CC BY-NC-ND 4.0	Linguistic corpora
<a href="#">2561</a>	Arabic Speech Corpus	Nawar Halabi	2016-06-09	Arabic	CC BY-NC-ND 4.0	Linguistic corpora
<a href="#">2555</a>	French Oral Narrative Corpus	Janice Carruthers	2013-08	French	Free for non-commercial use	Linguistic corpora
<a href="#">2554</a>	British National Corpus, XML edition	BNC Consortium	2007-03-06	English	BNC User Licence	Linguistic corpora

# Example resource at OTA

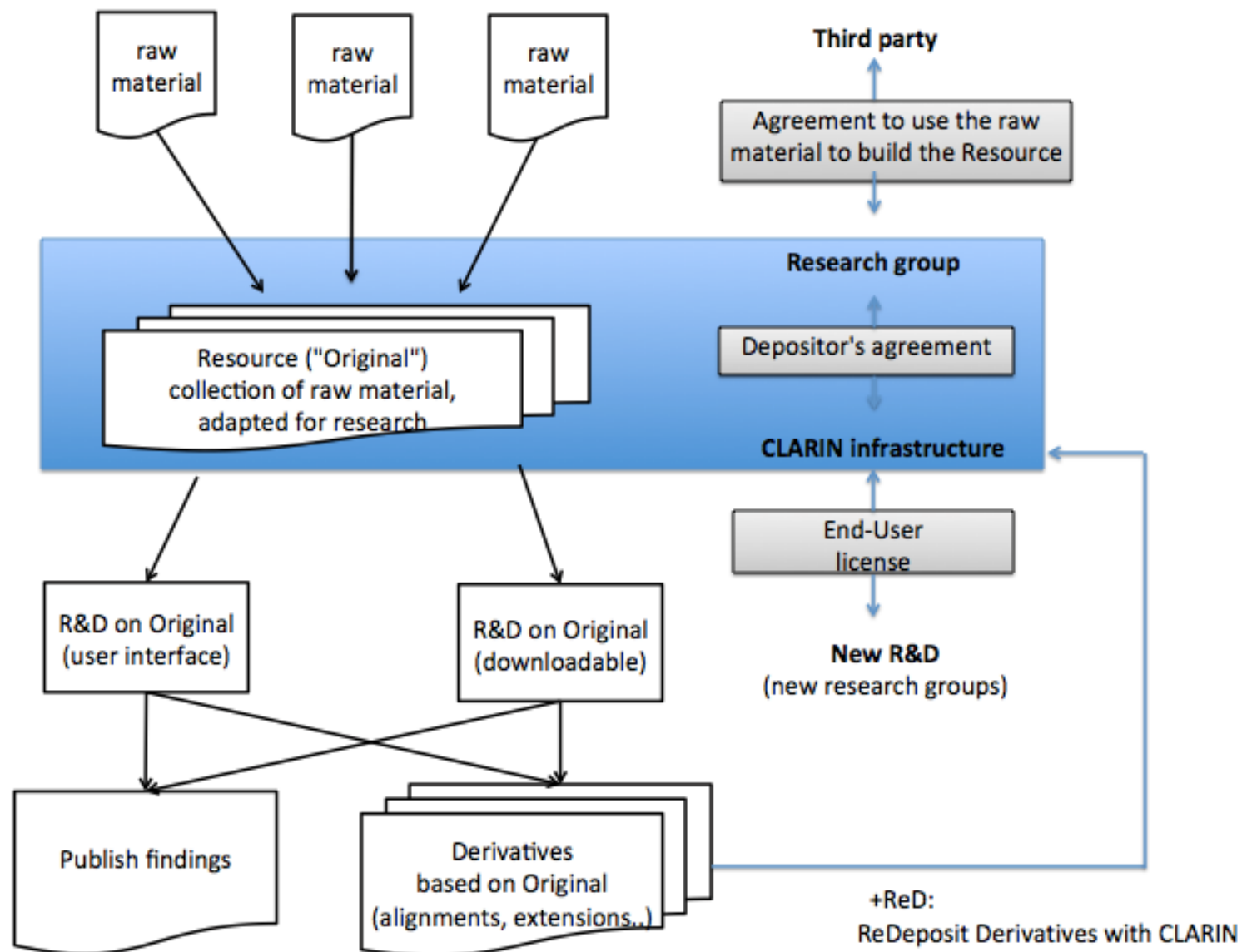


## University of Oxford Text Archive

Oxford Text Archive: [Home](#) | [Search](#) | [Catalogue](#) | [TCP](#) | [Deposit](#) | [Oxonly](#) | [Advice](#) | [FAQ](#) | [About](#)

<b>Title</b>	The Lampeter Corpus of Early Modern English Tracts
<b>Editor</b>	<a href="#">Josef Schmied</a> , <a href="#">Claudia Claridge</a> , <a href="#">Rainer Siemund</a> (ed.)
<b>Availability</b>	<a href="#">Distributed by the University of Oxford under a Creative Commons Attribution-ShareAlike 3.0 Unported License</a>
	Download: <a href="#">zip</a>
<b>Languages</b>	English
<b>OTA keywords</b>	<a href="#">Linguistic corpora</a>
<b>LC keywords</b>	<a href="#">Tracts</a> <a href="#">Anthologies -- England</a> <a href="#">Linguistic analysis (Linguistics)</a>
<b>Source Description</b>	17th- and 18th-century collections of the Founders' Library, University of Wales, Lampeter (formerly Saint David's College), especially the Tract Collection, cf. Saint David's University College. 1975. A catalogue of the tract collection of Saint David's University College, Lampeter. London: Mansell Information Publishing. <i>Revised version of</i> <a href="http://ota.ox.ac.uk/id/2400">http://ota.ox.ac.uk/id/2400</a>
<b>Permanent URL</b>	<a href="http://purl.ox.ac.uk/ota/3193">http://purl.ox.ac.uk/ota/3193</a>

# Depositor's Licence Agreements



# Corpuscle :: Talk Of Norway :: Metadata

Corpuscle Home  
Getting started  
Documentation  
FAQ  
Publications  
Links

Corpus list

Query  
Concordance  
Collocations  
Distribution  
Word List  
Diagram  
Text

**Metadata**

Overview  
Corpus doc.

Upload corpus

Localization

## Talk Of Norway

### Full metadata record:

[hdl:11495/DBCC-90CB-570D-9](https://hdl.handle.net/11495/DBCC-90CB-570D-9)

### Persistent identifier for the resource:

[hdl:11495/DBCC-915D-C2D9-0](https://hdl.handle.net/11495/DBCC-915D-C2D9-0)

### Links:

<http://clarino.uib.no/korpuskel/landing-page?identifier=ton&view=short> (landing page @ Corpuscle)

### Contact Person: ,

### This resource is licensed under the following terms:

[Norwegian Licence for Open Government Data \(NLOD\)](#)



Please click on the link to read the license terms.

**By accepting the terms of the license you will be granted access to the resource.**

### Attribution:

Please use the following text to cite this resource:

Talk Of Norway. Distributed by the [CLARINO UiB Portal](#): [hdl:11495/DBCC-915D-C2D9-0](https://hdl.handle.net/11495/DBCC-915D-C2D9-0)

# Corpuscle :: Talk Of Norway :: Collocations

**Advanced search** | [switch to Basic search](#)

[Query history ...](#)

[lemma = "grønn"]

[Run Query](#) | [Saved queries ...](#) |

[Save Query](#) as

Done. Running time: 0.01 sec. (0.01 CPU sec.)

Show collocations by lemma , left context: , right context: , sorted by MI \* log(Freq)  | [Download](#)

846 collocations calculated; page 1 of 17. [Previous](#) [Next](#) | [Show concordance for selection](#) | ☒ show freq. | ☒ show rel. freq. | ☒ show MI | ☒ show LL

Freq.	Rel. freq.	MI	LL	Delta	Collocate
975	0.7101	13.2130	25697.6070	1 <input type="checkbox"/>	grønn sertifikat
146	0.7487	13.2894	32136.8520	1 <input type="checkbox"/>	grønn skatteskitte
91	0.9891	13.6911	39047.7300	1 <input type="checkbox"/>	grønn sertifikat-marked
85	0.9551	13.6405	39245.2700	1 <input type="checkbox"/>	grønn utviklingsmekanisme
158	0.4463	12.5431	25773.2230	1 <input type="checkbox"/>	grønn sertifikatmarked
389	0.1250	10.7069	6560.9062	1 <input type="checkbox"/>	grønn skifte
63	0.4145	12.4362	33162.9920	1 <input type="checkbox"/>	grønn skattekommisjon
42	0.3750	12.2919	35962.6680	1 <input type="checkbox"/>	grønn klimafond
17	1.0000	13.7069	54437.7800	1 <input type="checkbox"/>	grønn sertifikat-ordning
37	0.3426	12.1615	36258.9400	1 <input type="checkbox"/>	grønn datasenter
29	0.3258	12.0891	38089.4840	1 <input type="checkbox"/>	grønn lunge
17	0.5484	12.8402	48480.7100	1 <input type="checkbox"/>	grønn skatteveksling
51	0.1496	10.9657	24320.4530	1 <input type="checkbox"/>	grønn karbon
8	1.0000	13.7069	61593.3830	1 <input type="checkbox"/>	grønn sertifikat-system
24	0.2727	11.8324	38128.3120	1 <input type="checkbox"/>	grønn gren
23	0.2190	11.5162	36306.3300	1 <input type="checkbox"/>	grønn sertifikatordning
33	0.1447	10.9184	28368.7910	-1 <input type="checkbox"/>	pliktig grønn
6	0.7500	13.2919	61550.3750	1 <input type="checkbox"/>	grønn sertifikat-marknad
13	0.2766	11.8527	44274.9960	1 <input type="checkbox"/>	grønn sertifikatsystem

# Corpuscle :: ICAME – COLT :: Metadata

Corpuscle Home  
Getting started  
Documentation  
FAQ  
Publications  
Links

Corpus list

Query  
Concordance  
Collocations  
Distribution  
Word List  
Diagram  
Text

**Metadata**

Overview  
Corpus doc.

Upload corpus

Localization

## COLT - The Bergen Corpus of London Teenage Language (with sound)

### Full metadata record:

[hdl:11495/D9B6-1393-7E6D-1](https://hdl.handle.net/11495/D9B6-1393-7E6D-1)

### Persistent identifier for the resource:

[hdl:11495/D9B6-13F8-41BB-1](https://hdl.handle.net/11495/D9B6-13F8-41BB-1)

### Links:

<http://clarino.uib.no/korpuskel/landing-page?identifier=colt&view=short> (landing page @ Corpuscle)

<http://clu.uni.no/icame/colt/> (project page)

**Contact Person:** Jørgensen, Annette Myre

### This resource is licensed under the following terms:

[CLARIN\\_RES-PLAN-INF-PRIV-ND-DEP-\\*](#)



BY

DEP

ID

INF

ND

NORED

PERM

PLAN

PRIV

Please click on the link to read the license terms.

**You have accepted this license.**

### Attribution:

Please use the following text to cite this resource:

COLT - The Bergen Corpus of London Teenage Language (with sound). Created by *Språkkontakt og ungdomsspråk i Norden*. Distributed by the CLARINO UiB Portal: [hdl:11495/D9B6-13F8-41BB-1](https://hdl.handle.net/11495/D9B6-13F8-41BB-1)

# Corpuscle :: ICAME – COLT :: Concordance

**Basic search** | [switch to Advanced search](#)

[Run Query](#)

| [Saved queries ...](#)

| [Save Query](#)

innit

as

Done. Running time: 0.00 sec. (0.00 CPU sec.)



☒ show audio/video player

Hit 1 - 30 of 325 | [Previous](#) | [Next](#) | Go to:  | [Download](#) (☐ Excel mode) | Type:  | ☐ Show annotation | ☐

Show line filter | [Attributes ...](#) | [Structures ...](#) | Page size:  | Context size:

## match

▶ liss Miss ' s definitely better Good He's almost he's dead now	innit	? Yeah Did he die at Why's she on there? Ah ah! See who Mis
▶ ng much Ah but you're going on Tuesday at five half past five	innit	? No next Tuesday or next Monday Yeah sure What? What? G
▶ f I come on hello and they say hello and that's that It's good	innit	? That is a bit shit! Well That is not a bit shit No well I wasn't
▶ ' Yeah well there's Freddie I know this Saturday It's your dad	innit	? Yeah your dad I want you to know Oh I think it was Freddie
▶ have Hello Wesley A three-legged cat Tiger It's name's Tiger	innit	? A three-legged Tiger How many stripes has it got? Why No
▶ :s a bit of hostility here This thing's gonna be so bloody then	innit	? No but she said you could have eight people all talking at o
▶ e it to I don't give it to Miss! And Miss gives it to the student	innit	? Cassie I've gotta come in on Monday morning when there's
▶ really echoes when you bang the table Ha? Do it again Good	innit	? Ah! Probably have a go at me for smashing their microphor
▶ ha? I'll come over to you No I'm listening to the radio Wesley	innit	? Where do you live Wesley
▶ ncies Robbie out of Take That I'll tell you what It's disgusting	innit	? She likes this song Come on come on come <u>on</u> come on Tal
▶ rn turn That's was mv indicator impression Oh right It's good	innit	? Yeah I suppose so What d'va mean you suppose If you sav



## The CLARINO Bergen Centre offers:

A repository to search and deposit language data  
Online services for treebanks and other corpora  
Online editing of CMDI metadata



## Welcome to CLARINO Bergen Centre

CLARINO is a Norwegian infrastructure project jointly funded by the Research Council of Norway and a consortium of Norwegian universities and research institutions. Its goal is to implement the Norwegian part of CLARIN. The ultimate aim is to make existing and future language resources easily accessible for researchers and to bring eScience to humanities disciplines.

[Search](#)[Advanced Search](#)

### Author


[Giellatekno - Saami ... \(22\)](#)[The Divvun group at ... \(17\)](#)[Parra Escartín, Carla \(3\)](#)[Kristiansen, Marita \(2\)](#)[Olstad, Vemund \(2\)](#)[... View More](#)

### Subject

[Bilingual Lexicon \(9\)](#)[Termbase \(9\)](#)[South Saami \(8\)](#)[Terminological \(8\)](#)[Terminology \(8\)](#)[... View More](#)

### Language (ISO)

[Norwegian Bokmål \(15\)](#)[Northern Sami \(7\)](#)[Southern Sami \(7\)](#)[English \(6\)](#)[Norwegian Nynorsk \(6\)](#)[... View More](#)

 Please use the following text to cite this item or export to a predefined format:

BIBTEX

CMDI

Hareide, Lidun, 2013, *The Norwegian-Spanish Parallel Corpus*, Common Language Resources and Technology Infrastructure Norway (CLARINO) Bergen Repository, <http://hdl.handle.net/11509/73>.




Share:   

Clarino


 **Authors** Hareide, Lidun

 **Date issued** 2013


 **Type** corpus

 **Size** 32000000 bytes

 **Language(s)** Spanish , Norwegian

 **Description** The Norwegian Spanish Parallel Corpus (NSPC) was created at the University of Bergen, Norway. The corpus is primarily constructed for research in Translation Studies, and is built to be roughly comparable to the Spanish-English P-ACTRES corpus. The NSPC is a parallel, unidirectional translation corpus of contemporary Norwegian written texts translated into Spanish, published between 2000 and 2009. It contains fiction and non-fiction, and each text is classified according to genre, the author's gender and the gender and mother tongue of the translator.

 **Publisher** University of Bergen

 **Subject(s)**

Annotation	Bilingual Lexicon Induction	Document Classification	Intra - Document Coreference Resolution
Language Identification	Language Modelling	Lexicon Access	Lexicon Acquisition From Corpora
Lexicon Enhancement	Lexicon Extraction From Lexica	Linguistic Research	Machine Translation
Named Entity Recognition	Natural Language Generation	Natural Language Understanding	Parsing

## Files in this item

[Download all files in item \(31.14 MB\)](#)

This item is **Academic Use** and licensed under:

CLARIN\_ACA



**Name** nspc-aligned.tar.gz  
**Size** 10.54 MB  
**Format** application/gzip  
**Description** Unknown



[Download file](#)

**Name** nspc-meta.xls  
**Size** 25.5 KB  
**Format** Microsoft Excel  
**Description** Unknown



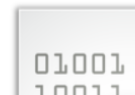
[Download file](#)

**Name** nspc-meta2.xls  
**Size** 43 KB  
**Format** Microsoft Excel  
**Description** Unknown

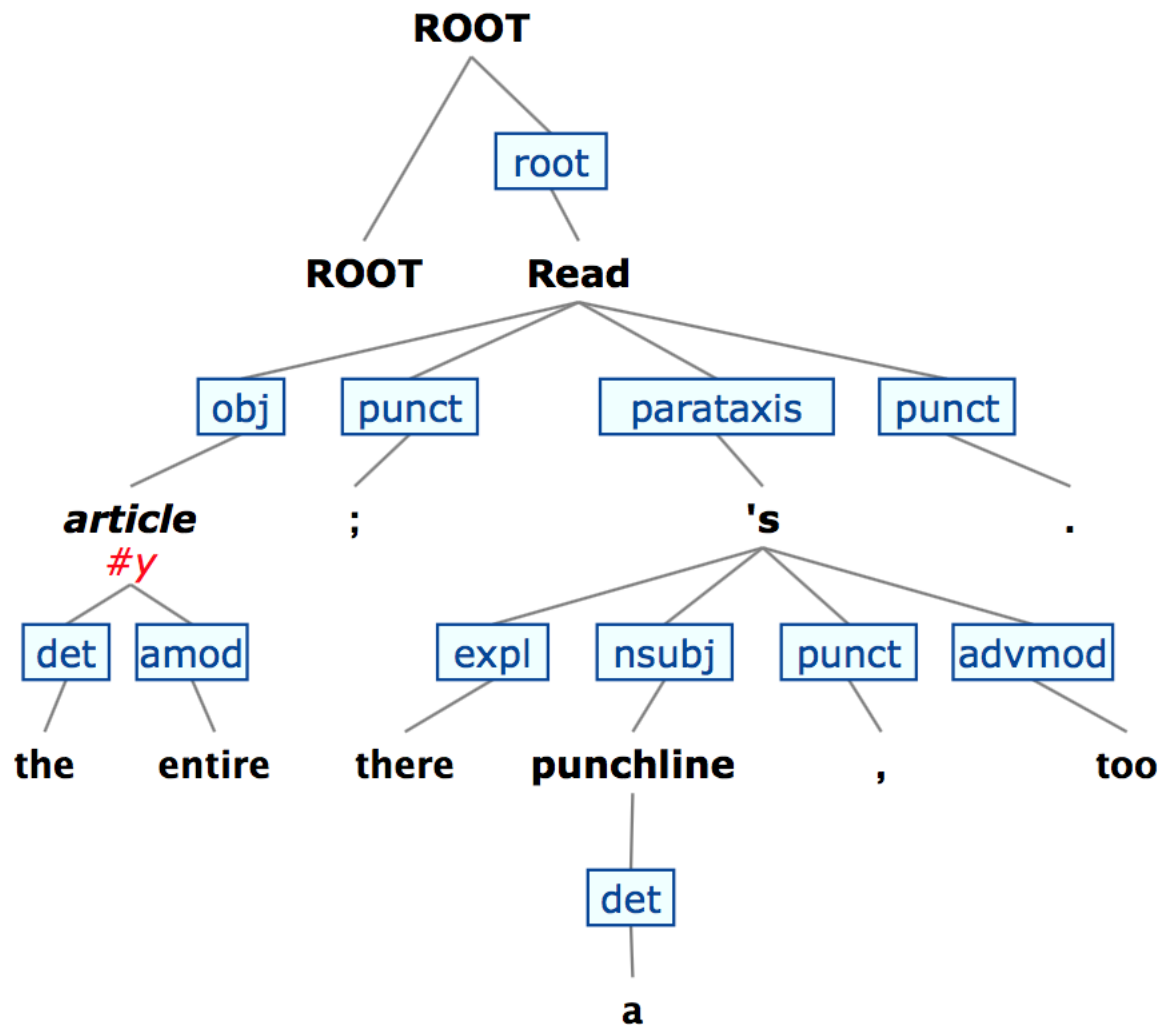


[Download file](#)

**Name** nspc-pos-tag.tar.gz  
**Size** 9.45 MB



# Treebank exploration



# Matches of syntactic relations, frequencies

Click on a row to see the matching sentences.

**Count** #y: *word*

4	it
4	letter
4	this
4	some
3	reviews
3	what
3	article
2	which
2	chapter
2	something
2	book

Click on a row to go to the sentence.

<i>Treebank</i>	<i>Document</i>	<i>Id</i>	<i>Sentence</i>
eng-ud-2.0-dep	train	<b>2500</b>	I 've read the book - it 's a first person memoir with the ring of truth, especially since so much of the wider context has confirmed the allegations of Contra drug connections that led to Terrell becoming a target for discrediting by the TWIG counterintelligence group that ran interference for the Contra effort, staffed by covert ops types including Ollie North, Robert Owen, Vince Cannistaro and Buck Revell.
eng-ud-2.0-dep	train	<b>9981</b>	Keep his cage open and go on your computer, or read a book, etc and maybe he will come out to you.



## The CLARINO Bergen Centre offers:

A repository to search and deposit language data  
Online services for treebanks and other corpora  
Online editing of CMDI metadata



## Welcome to CLARINO Bergen Centre

CLARINO is a Norwegian infrastructure project jointly funded by the Research Council of Norway and a consortium of Norwegian universities and research institutions. Its goal is to implement the Norwegian part of CLARIN. The ultimate aim is to make existing and future language resources easily accessible for researchers and to bring eScience to humanities disciplines.





[Advanced Search](#)

### Author

Giellatekno - Saami ... (22)

The Divvun group at ... (17)

Parra Escartín, Carla (3)

Kristiansen, Marita (2)

Olstad, Vemund (2)

[... View More](#)

### Subject

Bilingual Lexicon (9)

Termbase (9)

South Saami (8)

Terminological (8)

Terminology (8)

[... View More](#)

### Language (ISO)

Norwegian Bokmål (15)

Northern Sami (7)

Southern Sami (7)

English (6)

Norwegian Nynorsk (6)

[... View More](#)

Home  
Documentation

Catalogue  
Query  
Concordance  
Collocations  
Word List

Text

Overview

Localization

## AM 162 B α fol – A fragment of Njáls saga v. 1.0.2

Previous page

Next page

| Go to: 1r ▾

| Levels: ☐ Facs | ☒ Dipl | ☐ Norm

[ Dipl ]

[ms: fol. 1r]

- 1 fem bezt fípan★ oun hann fara j fiozdu ueftur ok figm-
- 2 und2000...000 hann flytia allt feit ueftan
- 3 þat000...000 þa0 0 fiozdum ueftur ok uera j burt leingi
- 4 000...000 þa er menn rida til þess 1. 11
- 5 000...000 ooda ætla skalttu r
- 6 000...000 oerdar med þer fa
- 7 000...000 til huilu þínar
- 8 000...000 þo skalt nefna uot
- 9 000...000 feigia skilit uid ha
- 10 000...000 at Alþingif mali
- 11 000...000 laugum lika uetta skalttu hafa fræ

Features:

norm: vestr

dipl: ueft{ur}

facs: ueft̥

document: AM-162-B-alfa-fol

reference: 1r 3

# Characteristics of RI in SSH

- Multitude of pre-existing centres with data, tools and expertise
- Relatively cheap to build and operate (compare 123 M GBP to CERN)
- Gradual construction, open-ended
- Cost-free open access
- Some restrictions due to copyright and privacy considerations
- No panacea for all research needs
- Obvious European dimension: bridging the language and culture barriers, towards inclusive societies.

# Final remarks

- Coordinated CLARIN activities in 21 countries
- High number of institutions and centres involved
- Wide span of targeted researchers, multidisciplinary research agendas
- Increasing amount of data and services offered
- Targeted mostly at researchers; also potentially useful for technology and education
- Participation in H2020 infrastructure projects: PARTHENOS, CLARIN-PLUS
- Towards sustainability: technology basis, long-term national commitments, synergy with other initiatives

# CLARIN in the UK



The image shows the CLARIN-UK website. The header is dark blue with the CLARIN-UK logo on the left, which consists of a stylized network of blue dots and lines above the text 'CLARIN-UK'. To the right of the logo, the text 'CLARIN-UK' is displayed in a large, white, serif font, followed by 'Infrastructure for Digital Language Resources and Tools' in a smaller, white, sans-serif font. Further right is a search bar with the word 'Search' in a small, grey font and a magnifying glass icon. Below the header is a navigation bar with a dark blue background and white text. The 'Home' link is highlighted with a teal background. Other links include 'Centres', 'Resources', 'People', 'News', 'Events', 'Blog', and 'About'. Below the navigation bar is a large banner image of a brick building with many windows and a green lawn in the foreground. A teal semi-transparent box is overlaid on the right side of the banner, containing the text 'CLARIN-UK: Promoting Cross-disciplinary Corpus Linguistics' in a large, white, sans-serif font, followed by 'Pre-conference workshop at CL2017' and '24 Jul' in a smaller, white, sans-serif font. At the bottom right of the teal box are two white circular arrows pointing left and right, separated by a vertical green bar.

CLARIN-UK  
Infrastructure for Digital Language Resources and Tools

Search

Home Centres ▼ Resources ▼ People News Events Blog About ▼

CLARIN-UK: Promoting Cross-disciplinary Corpus Linguistics  
Pre-conference workshop at CL2017  
24 Jul

# Get in touch

<http://clarin.ac.uk>

<http://clarin.eu>

<http://clarino.uib.no>