

Introduction to CLARIN

Martin Wynne

Senior Researcher in Corpus Linguistics
Faculty of Linguistics, Philology and Phonetics

University of Oxford

National Coordinator, CLARIN-UK

martin.wynne@ling-phil.ox.ac.uk



CLARIN in eight bullets

- **CLARIN** is the Common Language Resources and Technology Infrastructure
- CLARIN has since 2012 been an **ERIC** (European Research Infrastructure Consortium)
- ERICs are not short- or fixed-term projects, and CLARIN does not have an end date. It is intended to be a long-term infrastructure supporting and connecting research in the **humanities and social sciences** and beyond
- CLARIN aims to provide easy and sustainable access for scholars
 - to **digital language data** (in written, spoken, video or multimodal form)
 - and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they are located
 - through a **single sign-on** environment
 - that serves as an ecosystem for **knowledge exchange**
 - some services **integrated in EOSC** (European Open Science Cloud; [link](#))

See more in the CLARIN Value Proposition: <https://www.clarin.eu/content/value-proposition> (link to [pdf](#))





CLARIN in countries and centres

Official membership

- 23 members
- 3 observers
- 1 linked party

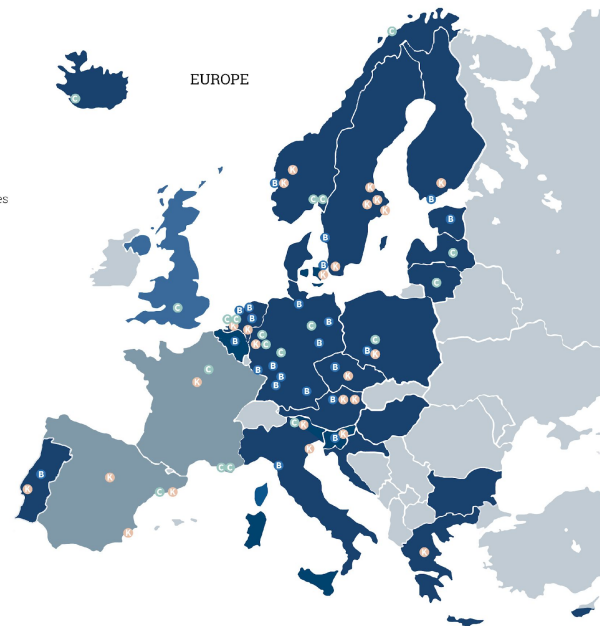
A distributed network of >60 centres

25 CTS certified data centres,
strong focus on FAIRness & interoperability

- federated login: 
- central metadata harvesting for easy discovery: 
- chained services: 
- language data - in written, spoken, video or multimedial for 
- advanced tools - to discover, explore, exploit, analyse, and combine data sets, *wherever they are located*



- ERIC members
- Observers
- Countries with participating centres
- ⓑ Centre Providing Data
- Ⓒ Centre Providing Metadata
- Ⓚ Knowledge Centre



CLARIN in services



CLARIN portal

Get an example-based impression of what's currently available



Depositing services

Store language resources in a sustainable repository at a CLARIN centre



Virtual Language Observatory

Discover language resources using a faceted browser or a map



Easy access to protected resources

Get easy access to protected resources, with your institutional username and password.



Language Resource Switchboard

Explore and analyze language data with a wide variety of tools



Virtual Collections

Create your own digital bookmarks, ideal for citing data sets.



Language Resource Inventory

Submit and access information about language resources relevant to your research.



Content Search (prototype)

Search different corpora with a single search engine



Questions & Answers

Searching for a specific data set or application? Wondering how CLARIN can assist your research? Feel free to contact us!

CLARIN in data types

- Newspaper archives
- Literary texts
- Social Media data
- Parliamentary records
- Historical letters
- Oral History data
- Disciplinary libraries
- Institutional archival data
- Broadcast archives
- ...

See also the info on the **CLARIN Resource Families** initiative: <https://www.clarin.eu/resource-families>

CLARIN Resource Families

Corpora

- Computer-mediated communication corpora
- Corpora of academic texts
- Historical corpora
- L2 learner corpora
- Legal corpora
- Literary corpora
- Manually annotated corpora
- Multimodal corpora
- Newspaper corpora
- Oral History Corpora
- Parallel corpora
- Parliamentary corpora
- Reference corpora
- Sign Language Resources
- Spoken corpora

Lexical Resources

- Language Models
- Lexicons
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

- Corpus Query
- Normalization
- Named entity recognition
- Part-of-speech tagging and lemmatization
- Sentiment analysis

Coming soon:

- Tools for Lexical Semantic Change Research

<https://www.clarin.eu/resource-families>

CLARIN in communities of use

- Digital Humanities
- Linguistics and Philology
- Translation and Lexicography
- Literary Studies
- History
- Political and Social Sciences
- Media Studies
- Culture, Folklore, Anthropology
- Speech therapy
- Teachers
- General Public

•

In the design of CLARIN interoperability is key









- to overcome fragmentation in the ecosystems of language resources
- to enable the exchange and harvesting of metadata
- to identify suitable exchange formats for the output of analytic tools for language data
- to the options for supporting comparative research;
example: [ParlaMint](#), a CLARIN initiative aimed at interoperability of parliamentary data and public debate on COVID-19

CLARIN corpus resources and tools

Corpora: at least 4130 - see VLO (<https://vlo.clarin.eu/>) !

Online interfaces:

- [Corpuscle](#) 
- [Korp](#) 
- [KonText](#) 
- [NoSketch Engine](#) 
- [D*](#) ([Diacollo demo](#)) 
- [TEITOK](#) 

Federated content search: <https://contentsearch.clarin.eu/>

Diachronic collocations in a text collection: DiaCollo from the Deutsches Textarch

D*/DTA*: DiaCollo

QUERY:

DATE(s): SLICE:

SCORE: KBEST: CUTOFF:

PROFILE: FORMAT: GLOBAL:

GROUPBY: 1PASS: DEBUG:

[Home](#) [Info](#) [Help](#) [Tutorial](#)



D*/DTA* Search

Hits 1 - 10 of 498

[-HTML](#) [-Hist](#) | [Home](#) [Query Lizard](#) | [Previous](#) [Next](#) | [Help](#) [+](#) [CabErr](#)

1:	[dtae:magirus_widerlegung_1580:8]	... Altars / daß daselbst warhaffig der Leib vnnd	Blut _[1]	im Brot vnd Wein _[2] / werde mündtlich geessen...
2:	[dtae:magirus_widerlegung_1580:18]	... Brot vnd Wein _[2] / seinen warhaffigen Leib vnd	Blut _[1]	/ zuessen vñ zutrincken darreiche vnd gebe...
3:	[dtae:magirus_widerlegung_1580:18]	... die vngläubige vnd vnwürdige / den Leib vñ	Blut _[1]	Christi mit Brot vnd Wein _[2] / im heiligen...
4:	[dtae:magirus_widerlegung_1580:19]	... Wein _[2] / aber gar nicht der Leib vnd	Blut _[1]	Christi / gegenwertig dargereicht vñ empfangen werde /...
5:	[dtae:magirus_widerlegung_1580:31]	... wir vns (ob der ware Leib vnd	Blut _[1]	Christi leiblich im Brot vnd Wein _[2] sey)...
6:	[dtae:magirus_widerlegung_1580:31]	... anders gewölt / dann daß Christi Leib vnd	Blut _[1]	war haffig im Nachtmal mit Brot vnd Wein _[2] ...
7:	[dtae:magirus_widerlegung_1580:35]	... / von einer wesentlichen Gegenwertigkeit des Leibs vnd	Bluts _[1]	Christi / vnder gestalt des Brots vnd Weins _[2] ...
8:	[dtae:magirus_widerlegung_1580:36]	... Brots vnd Weins _[2] / in den Leib vnnd	Blut _[1]	Christi (also daß kein Brot vnd Wein...
9:	[dtae:magirus_widerlegung_1580:36]	... Brots vnd Weins / in den Leib vnnd	Blut _[1]	Christi (also daß kein Brot vnd Wein _[2] ...
10:	[dtae:magirus_widerlegung_1580:37]	... / vnd mit dem sichtbarn Wein _[2] / das	Blut _[1]	Christi dargereicht werde.

[-HTML](#) [-Hist](#) | [Home](#) [Query Lizard](#) | [Previous](#) [Next](#) | [Help](#) [+](#) [CabErr](#)



Hilfe

Texte ▾

Projekt ▾

Dokumentation ▾

 in den Titeldaten
 im Korpus
 in der Dokumentation

Magirus, Johannes: Gründliche Widerlegung des unwahrhaftigen zwinglischen Buches, welches Magister Ambrosius Wolf unter dem Titel (Historie der Augsburger Konfession) in öffentlichem Druck ausgehen lassen. Tübingen, 1580.

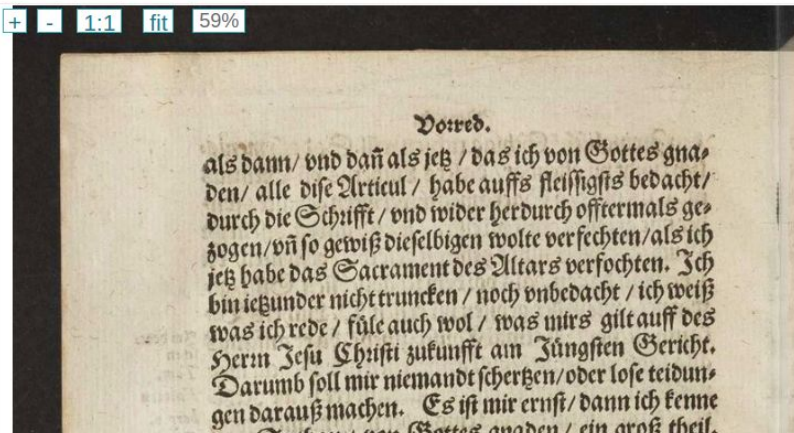
Bild: 0008 ▾

<< vorherige Seite

nächste Seite >>



+ - 1:1 fit 59%



als dann / vnd daß als jetz / das ich von Gottes gnaden / alle dise Articul / habe auffß fleissigsts bedacht / durch die Schrift / vnd wider herdurch offtermals gezogen / vñ so gewiß dieselbigen wolte verfechten / als ich jetz habe das Sacrament des Altars verfochten. Ich bin ietzunder nicht truncken / noch vnbedacht / ich weiß was ich rede / füle auch wol / was mirs gilt auff des Herrn Jesu Christi zukunfft am Jüngsten Gericht. Darumb soll mir niemant schertzen / oder lose teidungen darauß machen. Es ist mir ernst / dann ich kenne den Sathan / von Gottes gnaden / ein groß theil. Kaner Gottes Wort vnd Schrift verkeren vnd verwirren / was solt er nicht thon mit meinen / oder eins andern Worten? Vnd abermals hernacher daselbst: Eben so rede ich auch / vnd bekenne / das Sacrament des Altars / daß daselbst warhaftig der Leib vnnd **Blut** im Brot vnd Wein / werde mündtlich geessen vnd getruncken / ob gleich die Priester / so es raichē / oder die / so es

Useful Links

- CLARIN Value Proposition ([link](#))



- How to prepare for joining ([link](#))



- Newsflash ([link](#))

- Tour de CLARIN ([link](#))



- CLARIN Funding Hub ([link](#))



- Workshops ([link](#))



- Training Network ([link](#))



- Training Suite ([link](#))

