Further CLARIN-D services

- Sample use cases of some of our favourite tools: www.clarin-d.net/en/help/short-guides
- Blog entries on various Digital Humanities topics: www.clarin-d.net/en/blog-clarin-d-en
- Our ‘CLARIN Services’ with FAQs and help with legal questions: www.clarin-d.net/en/help

Do you have questions about CLARIN-D? Would you like to contribute to our blog? Do you have data that need archiving? Would you like to join the CLARIN-D infrastructure? Then please get in touch!

The Working Group 2 ‘Foreign Philologies’ is coordinated by Prof. Dr. Christian Mair at the University of Freiburg
christian.mair@anglistik.uni-freiburg.de
0761-203-3336

facebook.com/clarindeutschland
twitter.com/CLARIN_D
youtube.com/user/CLARINGermany

Working Groups of CLARIN-D

WG1: German Philology

WG2: Foreign Philologies

WG3: Linguistic Fieldwork, Anthropology, Language Typology

WG4: Human Speech Processing

WG5: Speech and Other Modalities

WG6: Applied Linguistics, Computational Linguistics

WG7: Content Analysis in the Social Sciences

WG8: History

www.clarin-d.net/en/disciplines
CLARIN-D Resources for “Foreign Languages”

CLARIN-D supports research on European languages other than German by providing services for exploring language data, analysing written and spoken texts, and archiving corpora and research results. The core of the work is in linguistics, but we invite cooperation with literary and cultural studies, with a particular interest in new developments in the Digital Humanities. Between them, the members of the working group combine expertise in English and the major Romance and Slavic languages.

1. Language Data

The Virtual Language Observatory (VLO) facilitates the search for research data, such as corpora.
https://www.clarin-d.de/en/accessing/vlo-search-for-resources

Explore the Language Index – the Place for Learning about the Languages of the World, developed and curated by working-group member Jürgen Handke (Marburg).
https://languageindex.online.uni-marburg.de/

2. Tools

Weblicht is a service orchestration and execution environment for incremental automatic annotation of text corpora, built upon Service Oriented Architecture principles. Its main components are:

- a set of services for data processing, hosted on servers distributed throughout Europe to allow full control of the tools by their developers
- a repository containing metadata about the services, such as input requirements and output
- a web application with a user-friendly graphical interface for building chains of services and executing them

A number of sophisticated search and visualisation options are currently being developed.

WebAnno is a flexible, web-based and visually supported system for distributed annotations. The tool can be used for a wide range of linguistic annotations and includes the following features:

- various layers of morphological, syntactical, and semantic annotations
- multi-user support for different roles (annotator, curator, and project manager)
- progress and quality monitoring and measuring in terms of inter-annotator agreement
- parallel annotation of multiple projects.
https://www.clarin-d.de/en/analysing/web-based-annotation

WebMAUS is a web-based segmentation and labelling tool for speech signals. It allows the user to automatically align speech recordings to their corresponding text form. Features include:

- production in SAMPA of the standard citation pronunciation of the content
- statistically weighted graphs of all possible realisations (pronunciation variants) within the selected language
- orthographic and phonetic alignment (segmentation and labelling, S&L) of the recorded speech
- rendering into BPF, EMU, or TextGrid via a web browser.

3. The Working Group’s “Use Cases”

Indexing of digital text archives through metadata and lemmas:
Learn more about the project and see the prototype corpus based on Polish data:
https://fedora.clarin-d.uni-saarland.de/poldilemma/

Old Bailey Corpus 1720-1913 in CLARIN compatible format:
Access the proceedings of London’s Central Criminal Court in a user-friendly search interface making possible comprehensive and sophisticated text searches at:
https://fedora.clarin-d.uni-saarland.de/oldbailey/

A team at the University of Freiburg helps the further development of Weblicht by testing the software in a number of teaching and research scenarios. The entire group is active in a CLARIN-D ad-hoc working group on the challenges of nonstandard and multilingual language data for corpus-linguistic and computational analysis.
CLARIN-D Resources for “Foreign Languages”

CLARIN-D supports research on European languages other than German by providing services for exploring language data, analysing written and spoken texts, and archiving corpora and research results. The core of the work is in linguistics, but we invite cooperation with literary and cultural studies, with a particular interest in new developments in the Digital Humanities. Between them, the members of the working group combine expertise in English and the major Romance and Slavic languages.

1. Language Data

The Virtual Language Observatory (VLO) facilitates the search for research data, such as corpora.
https://www.clarin-d.de/en/accessing/vlo-search-for-resources

Explore the Language Index – the Place for Learning about the Languages of the World, developed and curated by working-group member Jürgen Handke (Marburg).
https://languageindex.online.uni-marburg.de/

2. Tools

WebLicht is a service orchestration and execution environment for incremental automatic annotation of text corpora, built upon Service Oriented Architecture principles. Its main components are:
- a set of services for data processing, hosted on servers distributed throughout Europe to allow full control of the tools by their developers
- a repository containing metadata about the services, such as input requirements and output
- a web application with a user-friendly graphical interface for building chains of services and executing them.
A number of sophisticated search and visualisation options are currently being developed.

WebAnno is a flexible, web-based and visually supported system for distributed annotations. The tool can be used for a wide range of linguistic annotations and includes the following features:
- various layers of morphological, syntactical, and semantic annotations
- multi-user support for different roles (annotator, curator, and project manager)
- progress and quality monitoring and measuring in terms of inter-annotator agreement
- parallel annotation of multiple projects.
https://www.clarin-d.de/en/analysing/web-based-annotation

WebMAUS is a web-based segmentation and labelling tool for speech signals. It allows the user to automatically align speech recordings to their corresponding text form. Features include:
- production in SAMPA of the standard citation pronunciation of the content
- statistically weighted graphs of all possible realisations (pronunciation variants) within the selected language
- orthographic and phonetic alignment (segmentation and labelling, S&L) of the recorded speech
- rendering into BPF, EMU, or TextGrid via a web browser.

3. The Working Group’s “Use Cases”

Indexing of digital text archives through metadata and lemmas: Learn more about the project and see the prototype corpus based on Polish data:
https://fedora.clarin-d.uni-saarland.de/poldilemma/

Old Bailey Corpus 1720-1913 in CLARIN compatible format: Access the proceedings of London’s Central Criminal Court in a user-friendly search interface making possible comprehensive and sophisticated text searches at:
https://fedora.clarin-d.uni-saarland.de/oldbailey/

A team at the University of Freiburg helps the further development of WebLicht by testing the software in a number of teaching and research scenarios. The entire group is active in a CLARIN-D ad-hoc working group on the challenges of nonstandard and multilingual language data for corpus-linguistic and computational analysis.
Do you have questions about CLARIN-D? Would you like to contribute to our blog? Do you have data that need archiving? Would you like to join the CLARIN-D infrastructure? Then please get in touch!

The Working Group 2 ‘Foreign Philologies’ is coordinated by Prof. Dr. Christian Mair at the University of Freiburg
christian.mair@anglistik.uni-freiburg.de
0761-203-3336

facebook.com/clarindeutschland
twitter.com/CLARIN_D
youtube.com/user/CLARINGermany