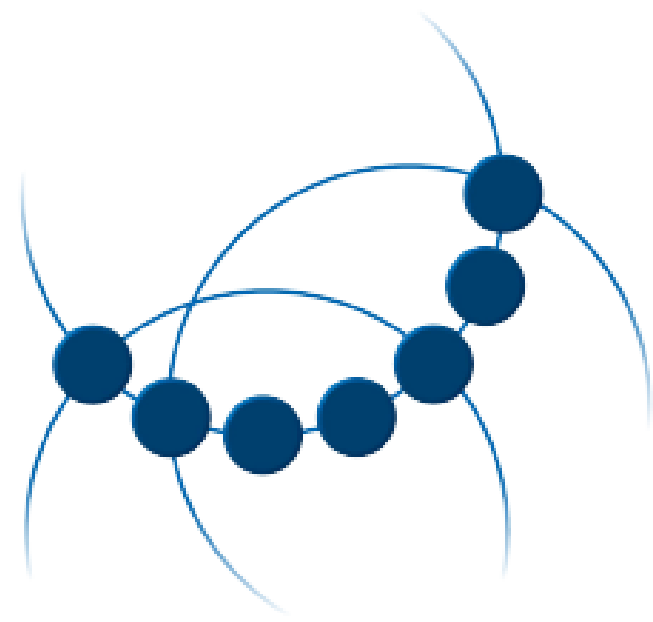
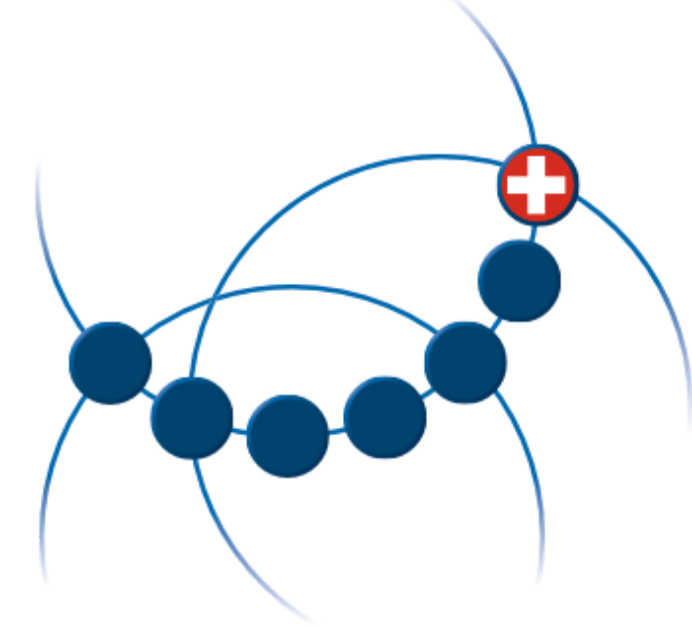


CLARIN
Common Language Resources and
Technology Infrastructure



CLARIN-CH
Common Language Resources and
Technology Infrastructure



Common Language Resources and
Technology Infrastructure and its Swiss
node CLARIN-CH: serving the national
applied linguistics community

VALS/ASLA Forum, 18 February 2022

The research infrastructure for language as social and cultural data

CLARIN



CLARIN is a digital infrastructure offering data, tools and services to support research based on language resources.

SOME FACTS:

- Designed and implemented between 2008-2012 by 9 founding countries.
- Significantly developed until 2016 and fully operational ever since.
- Currently, 26 European countries are members of CLARIN.
- Switzerland is in the process of joining the CLARIN Consortium.



Tools

Expertise

Resources

Network

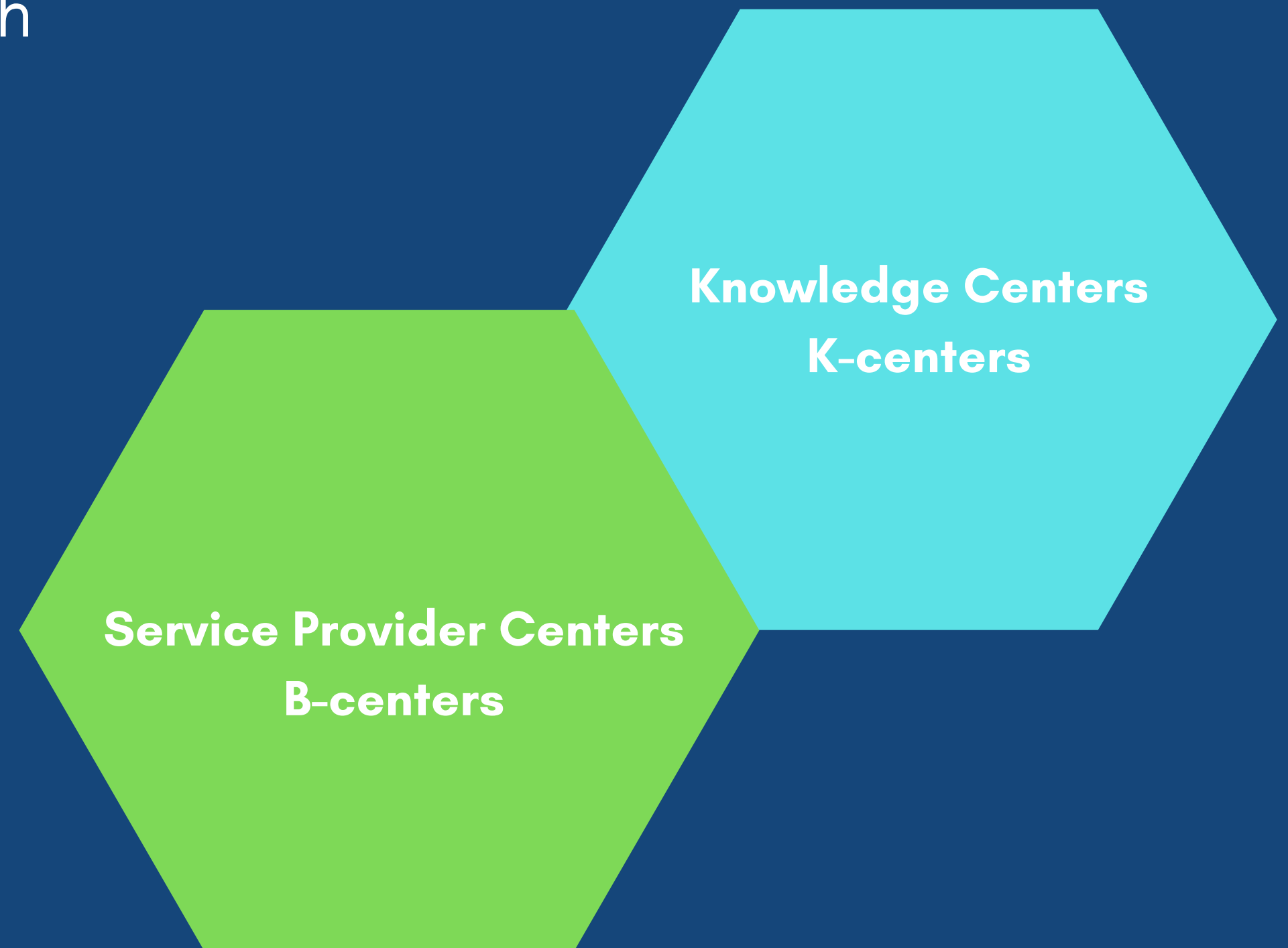
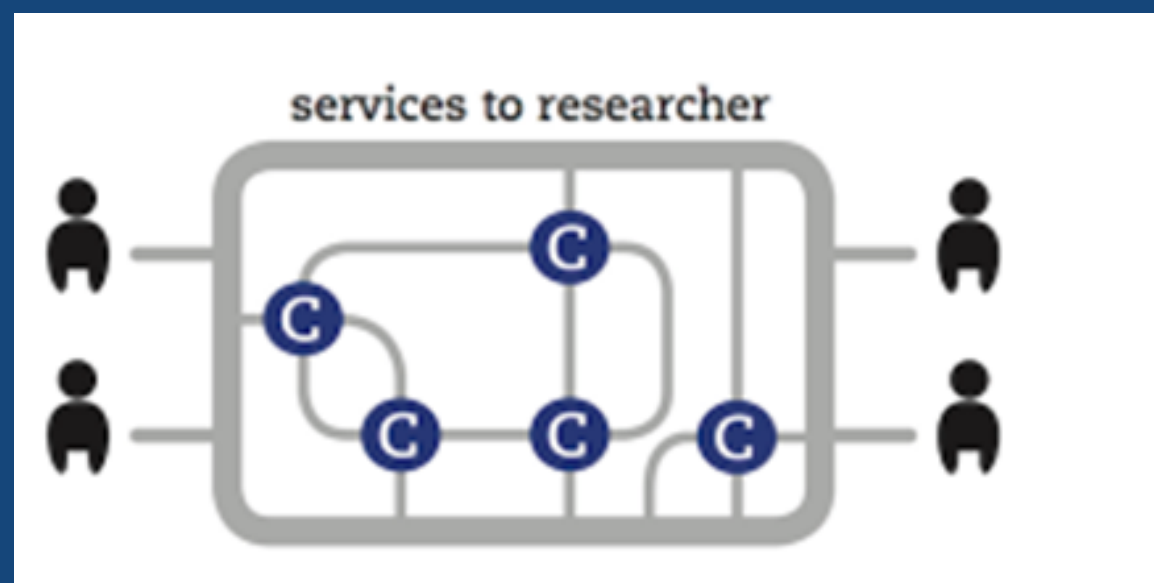
Services

CLARIN's mission:

Create and maintain an infrastructure to support the sharing, use and sustainability of language data and tools for research in the SSH.



CLARIN is a **networked federation of centers**– academic and research institutions that offer the scientific community expertise, resources, tools and services.



Researchers are service-users and service-providers



Corpora

- Computer-mediated communication corpora
- Corpora of academic texts
- Historical corpora
- **L2 learner corpora**
- Literary corpora
- Manually annotated corpora
- Multimodal corpora
- Newspaper corpora
- Parallel corpora
- Parliamentary corpora
- Reference corpora
- Spoken corpora

Resources



Lexical resources

- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools to discover, explore, exploit, annotate, analyse or combine language data.

Families of tools:

- Text normalization
- Named entity recognition
- Part-of-speech tagging and lemmatization
- Sentiment analysis and opinion mining

<https://www.clarin.eu/content/tools>



CLARIN **Knowledge** Centers (K-centers) are academic or research institutions which share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure.

K-centers have their own specific areas of expertise, which can belong to many different categories, such as:

- **Individual languages** (e.g. Danish, Czech, Portuguese)
- **Language families** (e.g. South Slavic)
- **Groups of languages** (e.g. morphologically rich languages, the languages of Sweden)
- **Linguistic topics** (e.g. language diversity, language learning, diachronic studies)
- **Language processing topics** (e.g. speech analysis, building treebanks, machine translation)
- **Data types other than corpora** (e.g. lexical data, word nets, terminology banks)
- **Using or processing families of language data** that will exist for most languages (e.g. newspapers, parliamentary records, oral history)
- **Generic methods and issues** (e.g. data management, ethics)



Services

- Online courses and training materials
- **Best-practice** documents
- **Guidance** in getting access to and using data and tools
- **Digital Humanities Course Registry** (CLARIN & DARIAH): an online platform for announcing and finding DH courses across Europe.
- **Training events** carried out by a group of experts at prominent summer schools, conferences, COST Actions, etc. in disciplines relevant for CLARIN.
- **Depositing Services**: researchers can store language resources in a sustainable repository at a CLARIN centre.
- An **online library of talks and tutorials** from the CLARIN training and academic events: the VideoLectures.net video channel.
- **Technical webservice**s

1. Funding opportunities:

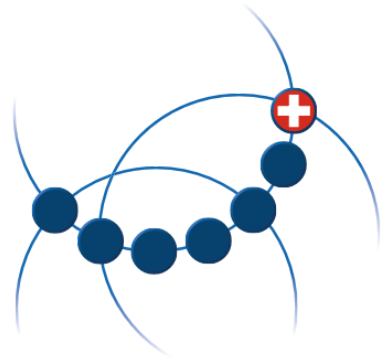
- Bridging Gaps Call
- Resource Families Project Funding
- Workshop Funding
- User Involvement Funding
- CLARIN Training Calls
- CLARIN Seed Grants
- Mobility Grants

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

2. Support for preparing EU-funded Projects:

- Help for *networking* and *documentation*
- *Assistance* in writing the proposal
- *Seed grants*: a small grant that can be used to cover certain costs coming with the preparation of a project proposal
- Facilitation services regarding the Open Science and FAIR Data requirements.

<https://www.clarin.eu/content/suport-eu-funded-projects>



CLARIN-CH

What is CLARIN-CH?



CLARIN-CH is the Swiss node of CLARIN Europe.

Concretely, CLARIN-CH is a network of six Swiss academic institutions, supported by the ASSH, who founded the *CLARIN-CH Consortium* in December 2020.





Our mission

Join the European CLARIN community, and build an active and impactful national network.



1. Obtain Switzerland's CLARIN membership and give Swiss researchers access to the entire CLARIN infrastructure.

2. Bring together the Swiss community using language resources and create national working groups.

3. Foster the sharing of expertise and of resources.

4. Encourage the initiation of national and international collaborations.

Foster the sharing of expertise



Dialectology
Sociolinguistics
Language ideologies
Pragmatics
Psycholinguistics
Discourse Studies
Prosody and regional variation
Semiology
Corpus Linguistics
Historical Linguistics
Bilingualism
Linguistic Anthropology
Engendered languages
Digital Humanities



Language variation
Language change
Conversational analysis
Linguistic narrativity
Spanish linguistics
Italian language and culture
Epistemology
Dialectology
Slavistics
Stylistics
Media discourse
Language sound structure
Computational linguistics
Language teaching
Digital Humanities



Cognitive Linguistics
Corpus Linguistics
Historical Linguistics
Language teaching
Dialectology
Interactional Linguistics
Discourse Studies
Pragmatics
Semantics
Developmental pragmatics
Cognitive anthropology
Cognitive sociology
Speech therapy
Language pathologies
Corpus statistics



Digital literacy
Language teaching
Political Linguistics
Legal Linguistics
Swiss German Sign Language
Institutional Communication
Social multilingualism
Professional translation
Audio description
Multimodal communication
Information design
Journalism
Media Linguistics
NLP, Machine Learning
Deep Learning

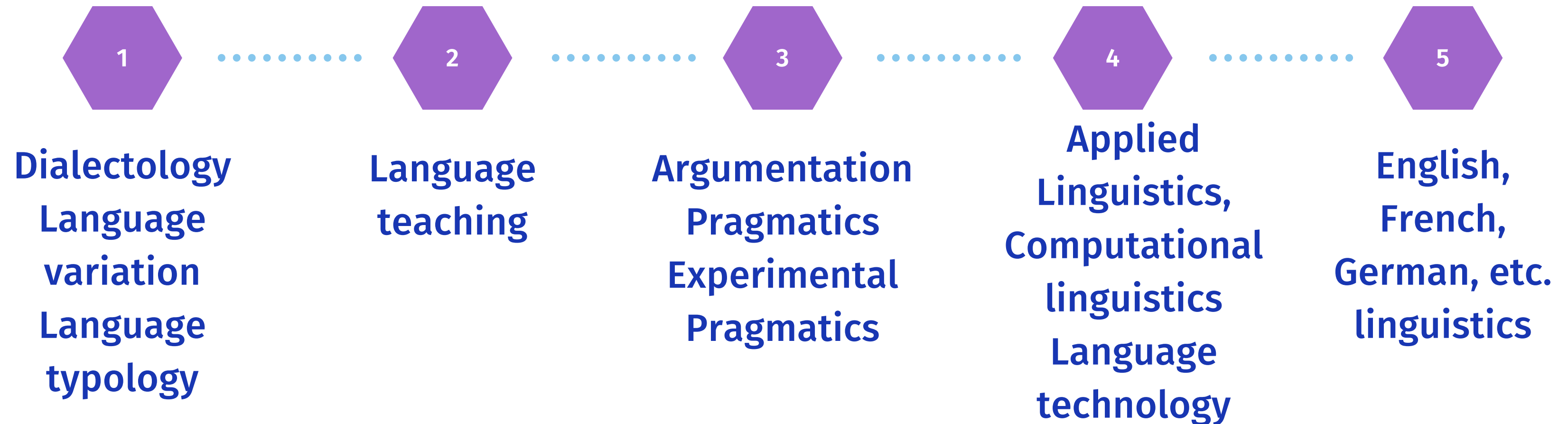


Discourse Studies
Argumentation
Persuasion
Intercultural pragmatics
Financial Communication
Semantics
Multimodality
Bilingualism
Italian Linguistics
Applied Linguistics
Stylistics

Foster the sharing of expertise and encourage collaborations



Discipline specific national working groups: examples



Digital literacy in university contexts (DigLit)

As digitalisation progresses, new technologies can offer valuable support to academic writers. However, research and practice show that students and their lecturers are often not aware of the applicability of available tools and are unable to exploit their full potential. These are the issues addressed by the DigLit project, supported by swissuniversities and the partner institutions.

Leading house: ZHAW School of Applied Linguistics

Project financing: P-8 Digital Skills programme, coordinated by swissuniversities and the partner universities

Overall budget: CHF 2.06 million

Project duration: January 2021 – December 2024



swissuniversities



Foster the sharing of resources: corpora and databases



- Interrogation Programme & Supersenses Extraction (IT)
- Lyra database (IT)
- ENIAT database (Slavic and South Asian Languages)
- FLORALE database (audio, FR)
- Phonocolor.ch (FR L2)



- ORFEUS databasis (FR)
- Glossaire de la Suisse Romande (FR)
- OFROM (audio and transcribed, FR)



- Swiss Web Corpus for Applied Linguistics (multilingual)



- Ortsnamen.ch/toponymes.ch (Swiss and international toponymy)
- Schweizer Textcorpus (DE)
- Schweizerisches Idiotikon (Swiss German)
- Vocabolario dei dialetti della Svizzera italiana (IT)
- Dicziunari Rumantsch Grischun (Romansh)
- Medieval Latin Dictionary
- Thesaurus linguae Latinae (ancient Latin)
- Historical Dictionary of Switzerland (DE, FR, IT)
- Collocation dictionary (DE)
- OLdPhras (DE)

Foster the sharing of resources: corpora and databases



- BNC Dependency Bank (EN)
- Giessen-Long Beach Chaplin Corpus (EN)
- Zurich English Newspaper
- International Corpus of English
- PADLF-Les plus anciens documents linguistiques de la France (FR)
- Tzéro database (FR)
- SenS-Korpus (DE)
- Archimob (DE)
- Picture postcard corpus (DE)
- DAVADS -Digital Audio/Video Archive (DE)
- Metalanguage Discourses (DE)
- Zurich summer 1968 (DE)
- Temporal entity extraction from historical texts (DE)
- JAKOB lexicon (DE)
- SADS-Syntactic Atlas of German-speaking Switzerland
- bulletin4corpus (DE, FR, IT, EN)
- sms4science (DE, FR, IT, Romansh)
- SMULTRON-Stockholm Multilingual Treebank (Swiss German, DE, IT, FR, Romansh)
- eSSRQ-Electronic Collection of Swiss Legal Sources (FR, IT, DE, Romansh, Latin)
- Phonogram Archives (Swiss dialects)
- Text+Berg corpus (Swiss German, DE, IT, FR, EN, Romansh)
- Bullinger Digital corpus (DE, Latin)
- CoNTra: the Federal Gazette (parallel FR-DE)
- PHOIBLE database (more than 1000 languages)
- Chintang Language Corpus
- Nepali National Corpus
- SEAlang corpus (South Asian languages)
- Corporum (Latin)
- European Language Equity resource collection (~60 corpora, ~40 applications, some lexica) (DE, IT, FR, Romansh)
- Genes and Languages Together database
- AUTOTYP database
- Macedonian Spoken Corpus
- Pre-Standardized Balkan Slavic Literature
- Torlak (a language from Timok area)
- Serbian Forms of Address (Serbian)



Take home messages:

- CLARIN is a community: a networked federation of centers.
- CLARIN-CH is a consortium of scholars targetting the development of an active and impactful Swiss CLARIN community.
- CLARIN and CLARIN-CH are based on networking and sharing of expertise, resources, and tools.
- Researchers are both service-users and service-providers.
- Conducting research with digital language and text data can be challenging at the individual level, but much easier at the collective level.

Meet the founding members of the CLARIN-CH Consortium



Marianne Hundt

National Coordinator
Representative of UZH



Anita Auer

President
Representative of UNIL



Cristina Grisot

Scientific Coordinator



Sandrine Zufferey

Representative of UNIBE



Martin Hilpert

Representative of UNINE



Andrea Rocci

Representative of USI



Cerstin Mahlow

Representative of ZHAW



Beat Immenhauser

Representative of SAGW/ASSH



Information Sessions CLARIN-CH

What is CLARIN?

CLARIN is a pan-european **digital infrastructure** offering **data, tools** and **services** to support research based on language resources.

What is CLARIN-CH?

CLARIN-CH is a **national network** that gathers Swiss scholars working on language related issues and fosters the sharing of resources and expertise.

Join us for the information session and find out more about how to gain access to all CLARIN resources and how to get involved !

TOUR DE SUISSE 2022: DATES

- UNINE: TUE 8 March
- UNIL: WED 16 March
- IDIAP: WED 23 March
- UNIBE: WED 30 March
- ZHAW: TUE 26 April
- UZH: FRI 6 May
- USI: TUE 17 May
- UNIBAS: TUE 14 June



DR. CRISTINA GRISOT
Scientific Coordinator
cristina.grisot@uzh.ch

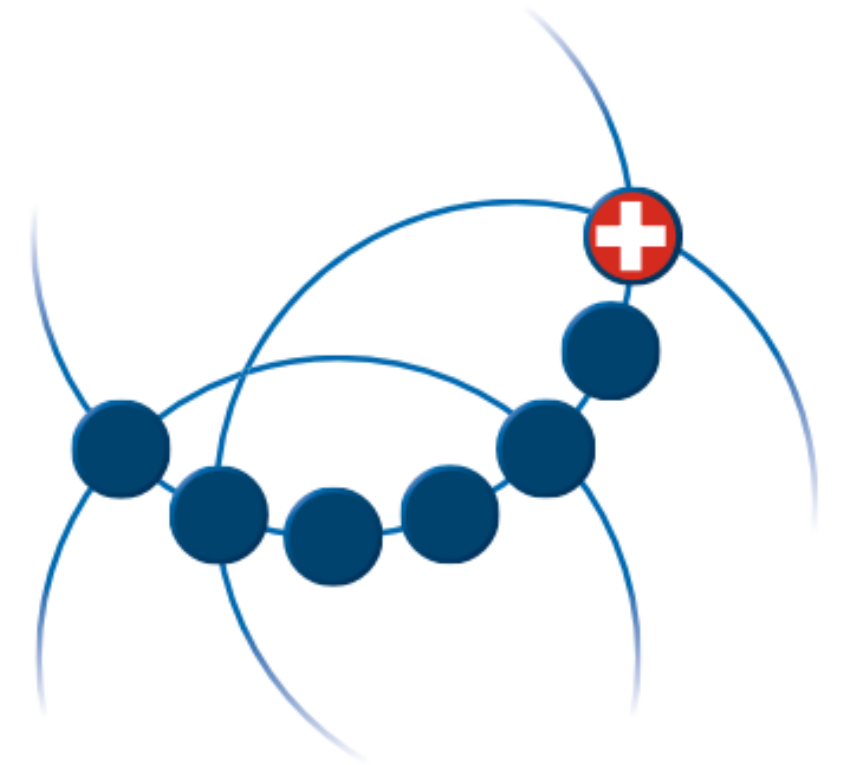
TOUR DE SUISSE 2022: DATES

- UNINE: TUE 8 March
- UNIL: WED 16 March
- IDIAP: WED 23 March
- UNIBE: WED 30 March
- ZHAW: TUE 26 April
- UZH: FRI 6 May
- USI: TUE 17 May
- UNIBAS: TUE 14 June



CLARIN-CH

Common Language Resources and
Technology Infrastructure



Thank you!

EMAIL

cristina.grisot@uzh.ch

WEB

Currently: <https://clarin-ch.linguistik.uzh.ch>

Soon will become: <https://clarin-ch.ch>