

 Content archived on 2024-06-18



Common Language Resources and Technology Infrastructure

Results in Brief

Web-based language data and processing tools

EU-funded scientists brought together available digital speech and language data and processing tools in a web-based architecture. It is equally easy to use by non-technical humanities experts and those in more historically technical fields.



HEALTH



© Thinkstock

Language is what separates humans from the rest of the animals and organisms. Understanding the structure, use and evolution of language and speech provides insight into topics as diverse as historic population migration patterns and criteria in developing web-based search engines.

Language and speech processing is a multidisciplinary field that encompasses not only linguistics but psychology, neural processing and cognition, computer science, electrical and computer engineering, biomedical engineering and mathematics.

European researchers in the humanities and social sciences (HSS) initiated the 'Common language resources and technology infrastructure' (Clarin) project to develop a unified language data and tools infrastructure.

The main objective was not to generate new knowledge. Rather, the team sought to

build on a wealth of national and European resources already available. The goal was to lay the foundations to unite existing data and tools under a common umbrella accessible to the entire research community.

The distributed data architecture was designed to provide web-based services to researchers as well as allowing non-expert users to perform complex tasks exploiting the wealth of language and speech processing tools developed in recent years.

A Virtual Language Observatory (VLO, <http://www.clarin.eu/vlo/> ) was created, making available analysed and summarised data on all language resources and tools from Clarin partners.

Many HSS communities are unfamiliar with linguistic processing tools and technology as it has historically been a less technology-oriented field. Building bridges to such communities is an important Clarin accomplishment.

In addition to technical considerations, Clarin also addressed issues of future governance and funding including investigation of possible legal, financial and organisational models. Project partners successfully mobilised a large HSS research community to lay the foundations for a unified language resources and tools infrastructure. Having access to such a huge variety of data and tools should now help scientists ask old questions in new ways. In addition, researchers will be able to ask new questions which, due to limited data and technology, they were previously not able to address.

The Clarin infrastructure has no doubt paved the way for exciting new interpretations in the field of language and speech processing.

Discover other articles in the same domain of application



New portal a key tool for better global pathogen surveillance

8 September 2023





Drones to transform hospital services

8 June 2023



The good and the bad of COVID-19 response in multi-level governance

30 November 2022



Delaying or refusing COVID-19 vaccines: The effects of misinformation

30 June 2022



Project Information

CLARIN

Grant agreement ID: 212230

Project closed

Start date

1 January 2008

End date

30 June 2011

Funded under

Specific Programme "Capacities": Research infrastructures

Total cost

€ 5 632 085,00

EU contribution

€ 4 100 000,00

Coordinated by



Last update: 23 October 2012

Permalink: <https://cordis.europa.eu/article/id/89627-webbased-language-data-and-processing-tools>

European Union, 2025