



HAL
open science

A CLARIN Resource Family for Sign Languages

Henk van den Heuvel, Onno Crasborn, Marie-Anne Sallandre, Eleni Efthimiou, Stavroula-Evita Fotinea, Johanna Mesch

► To cite this version:

Henk van den Heuvel, Onno Crasborn, Marie-Anne Sallandre, Eleni Efthimiou, Stavroula-Evita Fotinea, et al.. A CLARIN Resource Family for Sign Languages. CLARIN Annual Conference Bazaar, Oct 2022, Prague (Czech Republic), Czech Republic. . hal-03955866v1

HAL Id: hal-03955866

<https://hal.science/hal-03955866v1>

Submitted on 25 Jan 2023 (v1), last revised 25 Dec 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A CLARIN Resource Family for Sign Languages

Henk van den Heuvel & Onno Crasborn Radboud University, K-Centre ACE

Marie-Anne Sallandre, Université Paris8 & CNRS SFL, CORLI-K-Centre

Eleni Efthimiou & Stavroula-Evita Fotinea, ILSP/ATHENA RC, K-Centre NLP:EL

Johanna Mesch, Stockholm University, K-Centre CLARIN-SMS



CONTRIBUTION TO CLARIN GOALS

The CLARIN Resource family page for SL lists corpora and lexica for Sign Languages. A substantial number of these can be found in the CLARIN VLO but deserve a special podium in the spotlight. Sign language (SL) corpus resources contain transcriptions/annotations of spontaneous or elicited dialogues and narratives. All resources are in a video format because of the gestural/spatial-visual modality, a vital characteristic of signed languages (sign languages, used by Deaf-blind signers, can be received in tactile modality). SL corpora are crucial resources for various types of linguistic research, such as lexicography, phonology, syntax, and pragmatics, as well as for language typology.

This page also provides access to SL lexical resources. Some of them are connected to SL corpora. There also are independent lexical resources that were primarily created for language learning and teaching.

Method and outcomes

Method:

This CRF page was constructed by harvesting metadata for SL corpora in three different ways:

- By making an inventory of the material (datasets and resources) offered by CLARIN K-Centres with expertise in SL.
- By making an inventory of other datasets in the VLO which may qualify as members of the new CRF by contacting the right holders.
- By making an inventory of any other material (e.g., new datasets, annotation tools, manuals) not yet accessible through the CLARIN Infrastructure by sending out questionnaires to SL communities.

Results:

The SL resource family connects you to about 50 SL corpora and 25 SL lexica

A new component ModalitySL was introduced in the CMI Component Registry, with the values:

- Sign language
- Sign language (tactile)
- Spoken language
- Sign system
- Silent gesture
- Embodied speech
- Sign-supported speech
- Unspecified
- Unknown
- Other

Where to go:

Visit the website of the new Resource Family:

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

-> Sign Languages

henk.vandenheuvel@ru.nl

onno.crasborn@ru.nl

marie-anne.sallandre@univ-paris8.fr

[\[eleni_e; Evita}@athenarc.gr](mailto:{eleni_e; Evita}@athenarc.gr)

johanna.mesch@ling.su.se



Acknowledgements

We thank CLARIN ERIC for bringing the SL centered Knowledge Centers together and funding this Resource Family project

CLARIN

