

The CORLI Consortium: CORpus, Languages and Interaction

Christophe Parisse (cparisse@parisnanterre.fr), INSERM/Modyco, CNRS/U. Paris Nanterre, CORLI, France and Carole Etienne (carole.etienne@ens-lyon.fr), ICAR/CNRS, CORLI, France and Céline Poudat (celine.poudat@unice.fr), U. Nice Sophia Antipolis, CORLI, France

1. What is CORLI?

CORLI (CORpus, Languages and Interaction) is a consortium of Huma-Num (<https://huma-num.fr>) dedicated to the sharing of methodological approaches, tools and software, best practices and training within the community of linguists building and investigating corpora. Organising such a community is particularly complex because people have different theoretical and methodological approaches, practices and needs. The development of digital humanities and the widespread access to data have introduced significant changes in linguistics, both on the methodological and theoretical levels. Tools and methods are developing very quickly and in many places. CORLI is a response to these changes and a means to offer actual solutions to the use, share and reuse of linguistic data.

CORLI was created in 2016, following 5 years of experience as previous linguistic consortiums. Its target is written, oral and multimodal language. CORLI is a self-organised network stemming from the linguistic community. It gathers researchers and engineers from as many as possible French linguistic laboratories. The steering committee represents more than 22 different laboratories and CORLI involves more than 180 participants representing various research fields.

2. The goals of CORLI

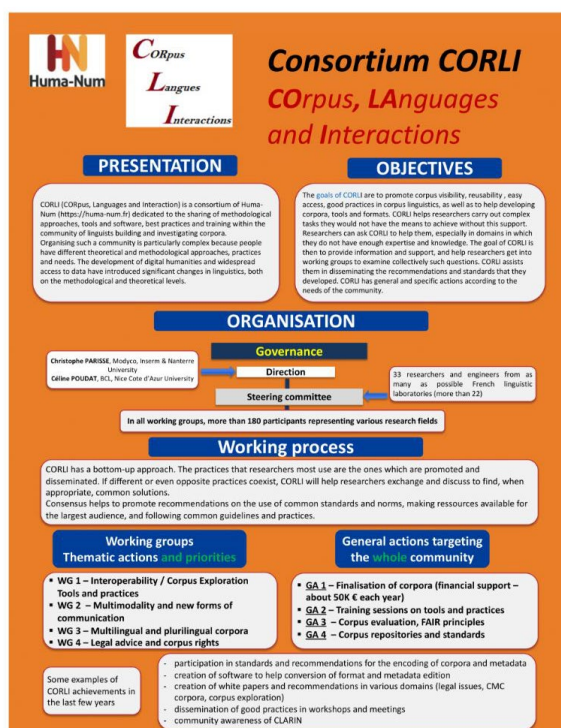
CORLI has a bottom-up approach. It is the practices that researchers use that are promoted to be used by the other groups. If there is a competition between different views, CORLI try to help people to discuss this, but doesn't decide for others what is good or not.

The goals of CORLI are to promote visibility, reuse, easy access, good practices in corpus linguistics, as well as to help developing corpus, tools and formats. CORLI help researchers in performing complex tasks they would not have the means to achieve without this support.

Researchers can ask help to CORLI, especially in domains where they feel underpowered or not sufficiently knowledgeable. Then the goal of CORLI is to find a way to answer this or to find information about this - or organise people to work about this in the community. CORLI has general and specific actions according to what is needed by the community.

General actions mainly concern financial help of people having corpora at their disposal that were not published by lack of financial resource, technical knowledge or adequate support from the infrastructure. CORLI helps projects to finalise research or organise data so that new corpus are available. Gathering and choosing data is up to the researchers.

General actions also concern recommendations about the evaluation of resources according to relevant criteria, the availability of data, legal information, and formation to the use of software for data creation (15 sessions each year), promoting international standards, editing both data and annotations,



and search tools. Moreover, CORLI contributes to CLARIN and DARIAH international infrastructures.

Specific actions are on the other hand carried out within three workgroups which have been created in that respect: (i) the first one focuses on interoperability, standards and corpus exploration; (ii) the second group concentrates on complex corpora (e.g. multimodal, CMC, sign language corpora...) whereas the third group works on multilingual corpora. The three groups work on three scientific challenges of corpus research which are basically related to different levels of complexity: oral and written corpus communities have developed separately, adopting different standards and tools. One of the goals of CORLI is to reunite the two communities, developing common tools and standards, as well as interoperable software; how to standardise and analyse complex corpora? Another goal of CORLI is to provide researchers the means and the infrastructure to develop adjusted standards and to share methods and options.

Realisations made by CORLI in the last few years were:

- about 50K € each year to help finalise projects
- participation in norm and recommendation for coding corpus and metadata
- creation of software to help conversion of format and metadata edition
- creation of white papers and recommendation for juridical problems
- workshops and meetings to discuss actual technics or difficulties
- information to the community about CLARIN
- continuing education for tools and practices

Usually, meetings and recommendations are in French because they target French speaking people. Some projects had an international public, so in this case, English was used. One of the goal for 2019 is to create a bilingual (French and English) website so as to extend our work with other communities and other countries.
