Call for Interns – CLARIN K-centre for Learner Corpora

The CLARIN K-centre for Learner Corpora (CKL2CORPORA) is looking for motivated interns to contribute to the development of innovative infrastructure supporting learner corpus research.

Context
The CLARIN Knowledge Centre for Learner Corpora offers expert knowledge on the collection and use of learner corpora (i.e. electronic collections of language data produced by second or foreign language learners) for theoretical and applied purposes. A key recent development is LC-meta, a core metadata schema for learner corpora, which provides a structured description of learners (e.g., linguistic background, proficiency, educational context), tasks (e.g., type, modality, domain), and learning situations (Paquot et al., 2024). Initially released only as a spreadsheet, LC-meta was difficult to use efficiently across different applications. Ongoing work now focuses on converting the schema into a fully machine-readable format.

Internship description
Interns will contribute to the transformation of LC-meta into a **machine-readable format** that will serve as the backbone of future services offered by the K-centre. The work will feed directly into the development of an **interactive interface** to support two main use cases:

1. **Collection of learner metadata** – assisting corpus builders and researchers in documenting learner data in a structured, standardized way.

2. **Search and discovery of learner corpora** – enabling users to search for existing corpora based on specific metadata fields (e.g., learner L1, proficiency level, task type, educational context).

Main tasks

- Finalize the conversion of the LC-meta schema into a machine-readable format.

- Contribute to the design of a prototype interface for metadata entry and search based on LC-meta.

- Test the schema and interface with sample metadata from existing learner corpora.

- Design and implement tasks with CLARIN interoperability in mind.

Required skills

- Knowledge of web technologies including
    - HTML, CSS, CSS frameworks (e.g., Bootstrap, Tailwind),
    - JavaScript, frontend frameworks (e.g., Vue.js, React),

    - and backend frameworks (e.g., Node.js, Python/Flask).

- Knowledge of Semantic Web/Linked Data technologies (XML, RDF, OWL, SPARQL, SHACL).

- Strong analytical skills and attention to detail.

- Ability to work both independently and collaboratively in a multilingual, interdisciplinary environment.

Desirable additional skills

- Background or interest in **linguistics, corpus linguistics, or language technologies**.

- Experience with **version control systems** (e.g., GitHub).

- Familiarity with **FAIR principles** and/or research data management.

What we offer

- Hands-on experience in a collaborative research infrastructure project at the interface of linguistics, data science, and digital humanities.

- The opportunity to contribute to resources and tools that will benefit the international learner corpus research community.

- Supervision and mentoring by experts in learner corpus research and metadata.

How to apply

Interested candidates are invited to send a short motivation letter and CV to Magali Paquot (magali.paquot@uclouvain.be). Please indicate your relevant skills, background, and availability. Applications will be reviewed on a rolling basis until the positions are filled.