

Introducing the CLARIN Knowledge Centre for Learner Corpora

De Cock, Sylvie; Damien De Meyere, Liesbeth Degand, Thomas François, Gaëtanelle Gilquin, Sylviane Granger, Danqing Huang, Marie-Aude Lefer, Hubert Naets, Magali Paquot, Kristel Van Goethem, Patrick Watrin with the collaboration of Jennifer Carmen-Frey, Alexander König and Egon Stemle

The main objective of this poster is to introduce the CLARIN Knowledge Centre for Learner Corpora (CKL2CORPORA), present its core missions and report on current projects.

CKL2CORPORA builds on more than 30 years of expertise in learner corpus research at the Centre for English Corpus Linguistics (CECL). CKL2CORPORA also relies on the expertise of staff from two additional research centres from the Institute for Language and Communication, UCLouvain, i.e. CENTAL and VALIBEL. Together, CKL2CORPORA members have expertise in learner corpus design (e.g., metadata, transcription, file formatting, ethics), annotation (e.g., POS tagging, parsing, error annotation), and analysis (e.g., error analysis, contrastive interlanguage analysis, transfer studies) in a range of languages (English, French, Dutch, etc.).

CKL2CORPORA's sharing of expertise can take various forms, from answering (theoretical, methodological, technical) questions sent via the helpdesk to providing training services. For example, CKL2CORPORA members organized the fifth edition of the Learner Corpus Research Summer School last summer. Additionally, they actively maintain a webpage aiming to list and classify existing learner corpora in various languages (Learner Corpora around the World) and have contributed to the development of corpus query tools such as Corpor@uclouvain, an online interface that provides access to corpora developed by UCLouvain members, and the UCLouvain Error Editor (UCLEE), a tool specifically designed to facilitate the insertion of error tags and corrections into learner texts, as well as their subsequent processing (Granger et al., 2023).

Among the other projects undertaken by CKL2CORPORA members is the creation of FABRA (Wilkens et al., 2022), initially designed for readability research but versatile enough to compute an extensive array of linguistic complexity measures for L2 French. Furthermore, the CKL2CORPORA members have actively contributed to the development of a Core Metadata Schema for Learner Corpora that can be used to document a learner corpus (Paquot et al., 2023).