ourowncorpus.de/stayintouch/hch19

HEIDELBERG SCHOOL OF EDUCATION

LOCALLY COMPILED (LEARNER) CORPORA IN FOREIGN LANGUAGE TEACHING

RESEARCH- AND USAGE-BASED SOFTWARE DEVELOPMENT

Ingo Kleiber | kleiber@heiedu.uni-heidelberg.de | @KleiberIngo

Teacher training
Unavailability of data
Lack of didactic foundations
Lack of educational software
…

CORPORA IN FLT AND ELT

(Learner) **Corpora**, systematically compiled collections of (learner) texts, can be fruitfully used in and applied to **Foreign/English Language Teaching**. However, despite many successful cases, there "remains a wide gap between the wide range of corpus-based activities that have been suggested […] and the relatively limited extend to which corpora are actually used in the ELT classroom" (Mukherjee 2006: 6). One very promising approach is the **use of locally compiled (learner) corpora** which are collected and analyzed by teachers (and learners) as "part of their normal teaching activities" (Granger 2012: 11).

PROJECT AIMS

This project has two **overarching aims**:

(1) Developing a comprehensive model for the fruitful integration and application of locally compiled (learner) corpora integrating (usage-based) linguistics as well as didactics/pedagogy.

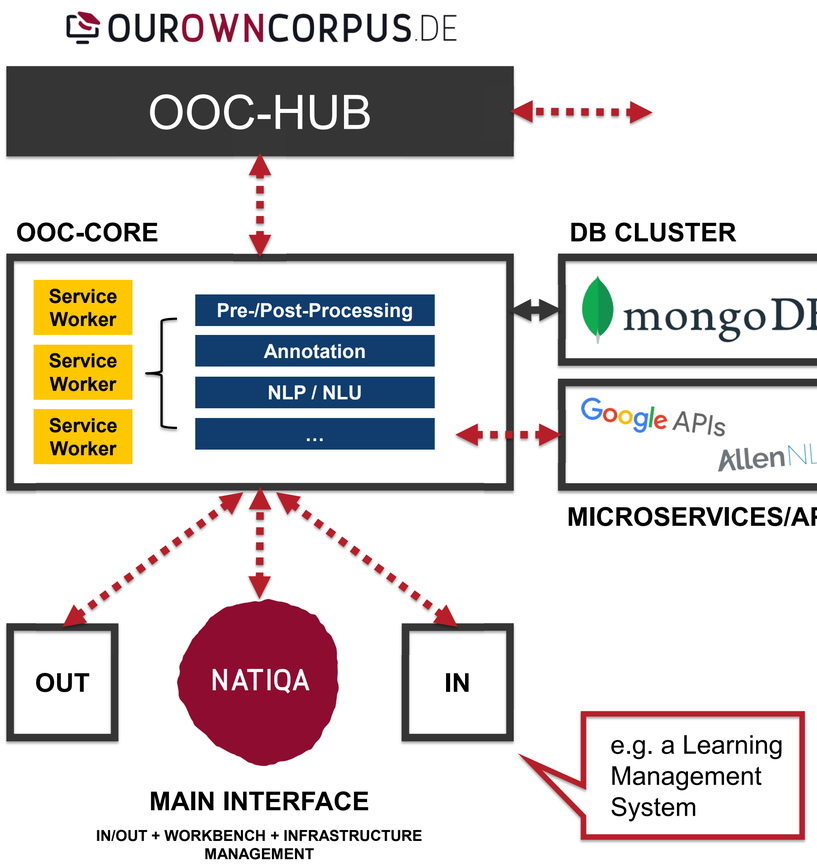
(2) Developing a pedagogically, didactically, and linguistically informed software platform specifically designed for the above.

THIS POSTER

Outlining the **software (and infrastructure) architecture** of the project: a highly modular and flexible **corpus compilation and analysis platform** designed with educators, learners (and researchers) in mind.

ARCHITECTURE - *OUROWNCORPUS*

python™ ····· JSON REST API



OOC-HUBs are optional infrastructure components which **facilitate data-sharing and transfer** between *OOC-COREs* (e.g. moving a learner and their data to another core instance).

Middleware (*OOC-CORE*) which does all of the heavy lifting (**computation**) as well as the actual **data management**. Data send to a core by an **ingress channel** (e.g. essays written by learners uploaded via *Natiqa*) is **processed, annotated, and stored** in a (interchangeable) database cluster. Complex tasks, e.g. updating or training a model, can be handed of to (distributed) service workers.

In order to be as flexible and compatible as possible, data (in and out) is exchanged via **REST (B/JSON) interfaces**. This allows us to use a **variety of data sources** (ingress channels) as well as **analysis and teaching tools**. *Natiqa*, the web-based main interface to the infrastructure, works both as a reference implementation (client) as well as a collection of basic ingress, teaching/learning, and analysis tools (e.g. document, voice, image upload; data-driven learning; basic corpus linguistic analyses).

→ Allowing for collaboration between institutions as well as between teachers/learners and researchers

→ Hiding, but not obfuscating, complexity
→ Homogenizing data/NLP processing
→ Future-proof architecture; e.g. introducing new NLP models without any changes to the input/output applications
→ Using B/JSON documents instead of, e.g. relational databases → data mobility; comprehensibility of data; flexibility; …

→ Maximizing flexibility and modularity
→ Easy, non-intrusive (teaching), collection and analysis of all possible types of learner (language) data
→ Making tool and plugin development as easy and painless as possible

Mukherjee, Joybrato. 2006. "Corpus Linguistics and Language Pedagogy: The State of the Art - and Beyond." In *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, edited by Sabine Braun, Kurt Kohn, and Joybrato Mukherjee, 5–24. Frankfurt am Main: Lang.

Granger, Sylviane. 2012. "How to Use Foreign and Second Language Learner Corpora." In *Research Methods in Second Language Acquisition: A Practical Guide*, edited by Alison Mackey and Susan M. Gass. 7–29. Chichester et al.: Wiley Blackwell.

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Pädagogische Hochschule HEIDELBERG University of Education

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386