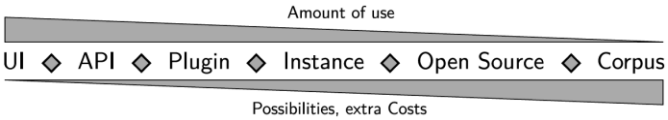
Reproducible complex corpus analyses with KorAP's R and Python libraries

Marc Kupietz, Nils Diewald, Peter M. Fischer, Eliza Margaretha
Leibniz-Institut für Deutsche Sprache, Mannheim
{kupietz,diewald,peter.fischer,margaretha}@ids-mannheim.de

Corpus data are difficult to access for simple interpretation due to their high-dimensional and opaque structure. In addition, they usually offer only limited access, as they are rarely openly available for legal reasons. The corpus linguistics group at IDS Mannheim has been pursuing a multi-level model to still make corpus data accessible and usable for as many as possible as well as for sophisticated applications, despite these challenges. At present, this model for using the German Reference Corpus DeReKo with the help of the corpus analysis platform KorAP provides for the following levels (Kupietz et al. forthcoming):



Through this (dynamic) model, as many different applications and analyses as possible should be manageable and technically supported without the need to copy and move data. The focus of this poster is on API-level access using KorAP's client libraries for R[1] and Python[2] (Kupietz et al. 2020). Basically all KorAP functions are accessible via the API, including authentication and authorization, which are also used by the KorAP web user interface (UI). Replication of complex or multi-part queries (with modified corpus sections, searches, or parameters, as well as on comparison corpora in other languages) and corresponding visualizations are an important application area for the API level. Another volatile application are experimental functions not yet fully supported by the backend, such as collocation analysis, aggregating representations, and cross-corpus analyses.

References: Kupietz, M., N. Diewald & E. Margaretha (forthcoming): Building paths to corpus data. *The CLARIN Book*, DeGruyter. • Kupietz, M., N. Diewald & E. Margaretha (2020): RKorAPClient. In: *Proc. of LREC 2020*, ELRA. pp. 7016–7021.

[1] https://github.com/KorAP/RKorAPClient
[2] https://github.com/KorAP/PythonKorAPClient