

## **VerbaAlpina - Collecting and Connecting Linguistic Data**

The long-term research project VerbaAlpina investigates the Alpine lexis which consists of a multitude of languages and dialects belonging to the large language families found in the Alpine region (Germanic, Romance and Slavonic). It utilises data from three different kinds of sources: (1) Traditional linguistic atlases and dictionaries, (2) existing digital data from various partner projects and (3) new material collected by means of a custom-developed crowdsourcing tool. Each one requires its own strategies and procedures to deal with the considerable differences in encoding and representation and convert the data into a uniform structure and format. One main focus lies on creating a representation as structured and fine granular as possible to make it (re-)usable in many different contexts, both within the virtual research environment VerbaAlpina provides and outside of it.

The research environment itself features two different views on the underlying data, mirroring the two traditional ways of publishing linguistic data, namely an interactive map and a lexicon-type surface (currently under construction). Both allow access from different angles (e.g. semasiological vs. onomasiological), whereas the map additionally supports multiple visualization forms (geographic vs. abstract and qualitative vs. quantitative).

In order to reach a high degree of interconnectedness with other online resources the project pursues several strategies. One is to link whenever possible to existing online dictionaries and well-established knowledge data bases, mainly for the project's three main categories (lexemes, meanings and geographical entities). A half-yearly versioning of the entire stored information makes it easy to quote parts of the linguistic core data as well as explanatory texts. Furthermore, with each version a complete export is published. The corresponding meta data is added to the indexes of different scientific search engines and therefore can be used for other research questions in related fields.

The project's methodology is documented in detail on the VerbaAlpina homepage. The complete source code for all developed tools is published online and updated periodically. This includes modularized tools for cartographic visualization and entering and converting phonetic phrases in different transcription systems or encodings.