Aspects of visual(ized) dialectology in the DMW project
Kai-Uwe Carstensen, Univ. Siegen, Germany

In 2016, the Academy of Sciences of Northrhine-Westfalia started funding the 17-year DMW (“Dialektatlas Mittleres Westdeutschland”) project comprising scientists of the Universities Bonn, Münster, Paderborn and Siegen. It aims at developing a digital, web-oriented, “speaking” dialect atlas for non-standard linguistic varieties in the middle western region of Germany (Nordrhein-Westfalen with parts of Niedersachsen and Rheinland-Pfalz).

The atlas will be based on 3000+ explorations in ~1000 different places, each exploration consisting of 800+ tasks of the questionnaire, where each task is related to possibly many dialect phenomena (of phonology, morphology, lexicology and syntax).

Unlike most other dialect atlases, the maps presented according to some query will be dynamically generated. Disregarding the necessary organization of data collection, modelling, storage, analysis and querying, this especially raises important questions in the domain of visualization. In general, these questions concern the conciliation of traditional dialectology’s requirements and habits with the possibilities and opportunities of modern web technology and geo-referenced content, paving the way for a new visual dialectology that merges classical and (dialectometric) computational dialectology with visual analytics.

While the maps of a dialect atlas are traditionally conceived of as static, mono-scale presentations of (analyzed) data, with a determinate iconography (e.g., symbols/graphical elements, size, color) and with restricted use for experts, modern web technology allows dynamic multi-scale explorations of dialect data in which the user (even a layperson) can take action by setting and modifying the parameters of both query and presentation. With simple means of selectively showing/hiding aspects, or with new means of, for example, smaller-scale presentations like clustered data views (indicating distribution) or heat maps (indicating quantity), different perspectives on the data are possible, allowing a gradual shift from querying to exploring, and therefore from visual presentation to visual analysis.

Algorithmically, much more than simple presentation is possible. For example, levels of relations between data may be computed and visualized (sameness, (amounts of) differences), as well as statistical inter-exploration estimation. At the same time, much more consideration will have to be spent on aspects of user interfaces: once the user has more options, the set of reasonable co-options will have to be discovered, and their effective (and efficient) use will have to be secured. Visualization can then be a guiding factor of what to explore next.

On the implementational level, it will be interesting to find out whether all classical-map features will be reproducible on a dynamic map, which new features turn out to be relevant, and which programming frameworks serve the purposes of the project best.

In the talk, I will present and discuss these and other aspects of visualization appearing in a dynamic digital dialect atlas. Especially in this early phase of the project, feedback and input are highly appreciated.

Dr. Kai-Uwe Carstensen once studied Linguistische Datenverarbeitung at the university of Trier, worked in linguistic, computational linguistic and AI departments, and is author and co-editor of introductions to computational linguistics and language technology. He is interested in visual cognition and attention and in (computational) linguistic interfaces, and is expert in cognitivist attention-based spatial semantics. He is now the technical coordinator (aka CTO) of the DMW project.