

Putting (Linguistic) Research Data on a Map – The DiÖ Sprachatlas Tool

Pluschkovits, Markus

markus.pluschkovits@univie.ac.at
University of Vienna, Austria

Bal, Jakob

jakob.peter.bal@univie.ac.at
University of Vienna, Austria

Ever since the project of the mapping of (German) linguistic data was started in the early 19th century (cf. Rabanus 2005: 346), and arguably reached an early high point with the *Sprachatlas des Deutschen Reichs* under the lead of Georg Wenker, putting linguistic data on a map has become a staple of variationist linguistics. The advent of digitalization has had huge implications for not only the production of linguistic maps and atlases (which are often associated with considerable workloads, cf. eg. Rabanus 2005: 345), but also on the reception of them, allowing mere readers of an atlas or map to become users of a powerful tool.

The *Sprachatlas* of the Special Research Programme *Deutsch in Österreich* ('German in Austria', henceforth DiÖ) aspires to be such a tool. The *Sprachatlas*, which is currently being developed and will be completed by the end of 2023, aims to be one of the tools by which the language data collected by DiÖ will be made accessible – not only to other researchers, but to the general public as well. The data collected by DiÖ is diverse, with the language data collected ranging from urban to rural locations, native German speakers to German as a second/third language speakers, older as well as younger speakers, and with speakers from a variety of educational levels. Additionally, a variety of settings was employed for the data collection, ranging from computer-assisted language production experiments to conversations between peers in absence of a researcher, to capture different levels of formality. Overall, the data was gathered with the intention to investigate the variety, change, perception and contact of the different varieties of German in Austria (for further information, see e.g. Budin et al. 2018).

This corpus and its associated metadata, like the annotations used for linguistic research, form the basis of the DiÖ *Sprachatlas*. The *Sprachatlas* utilizes an API (= Application Programming Interface) to directly load and chart the research data from the DiÖ database to the user interface. Users can not only listen to the recordings made in the different localities, but can also search for specific tokens, tags, or settings. They can also filter for specific parameters of informants (e.g. age, gender). Furthermore, they can create legends with complex queries, which enable them to map their own research questions using the DiÖ data, which can be queried for specific annotations and filtered by a variety of parameters, like the aforementioned informant parameters. By utilizing an API to map the data, the cost of the production of individual, custom maps is virtually non-existent. In addition to this explorative (cf. Pamperl 2017: 92) or documentative (Girnth 2010: 101) approach to data visualization, pre-defined maps complement the tool. It therefore doubles as a traditional linguistic atlas where commented maps can be browsed, and the data mapped

on it can be listened to. These pre-defined maps will be created using the same method of creating legends with different parameters, with the parameters being saved on our servers.

Whilst the *Sprachatlas* is not alone in trying to create dynamic maps to chart linguistic research data (other examples include, for example, the Scots Syntax Atlas, the interactive map of the Verba Alpina project, or LiÖ, the Lexikalisches Informationssystem Österreich ['Lexical Information System Austria']), we see great potential in the freedom afforded to its users. The API allows for fully customizable queries on nearly the entirety of the DiÖ-corpus, based on annotations, tokens, settings and the likes, with options for filtering according to informant features such as age or gender, or setting.

The development of the *Sprachatlas* is open source, and the code will be published on the DiÖ Github repositories (see DiÖ 2022). As of now, the repository is not publicly available yet, as the data the API accesses has not been anonymized. Development will be done by the end of 2023, where the *Sprachatlas* will be publicly launched. The main framework used for development is Vue.js (You 2020).

We believe that employing an API for the creation of dynamic, linguistic maps has several advantages. It does not only make it possible to create a nearly infinite number of maps based on custom parameters at comparatively little cost to us, it also makes the linguistic research of the project more transparent: by utilizing the annotations used by our researchers, and making them chartable on a map, the research data becomes more accessible and open in accordance with the philosophy of open science (cf. European Commission 2016: 33–35).

Bibliography

- Budin, Gerhard / Elspaß, Stephan / Lenz, Alexandra N. / Newerkla, Stefan Michael / Ziegler, Arne** (2018): "Der Spezialforschungsbereich 'Deutsch in Österreich (DiÖ). Variation – Kontakt – Perzeption'", in: *Zeitschrift für germanistische Linguistik* 46, 2: 300–398 <https://doi-org.uaccess.univie.ac.at/10.1515/zgl-2018-0017> [03.11.2022].
- DiÖ** (2022): *Github Repositories of the Special Research Programme Deutsch in Österreich ('German in Austria')*. Github Repository <https://github.com/german-in-austria> [06.04.2023].
- European Commission** (2022): *Open innovation, Open Science, open to the world. A vision for Europe*. European Commission, Directorate-General for Research and Innovation. <https://dx.doi.org/10.2777/061652> [03.11.2022].
- Girnth, Heiko** (2010): "Mapping Language Data", in: Lameli, Alfred / Kehrein, Roland / Rabanus, Stefan (eds.): *Language and Space: An International Handbook of Linguistic Variation Volume 2*. Berlin: de Gruyter, 98–121.
- Pamperl, Beate** (2017): "Visualisierungen in den Digital Humanities - ein Überblick", in: *Maske und Kothurn* 63, 1: 90–98.
- Rabanus, Stefan** (2005): "Sprachkartographie des Deutschen: Von Schmeller bis zum Digitalen Wenker-Atlas", in: Di Meola, Claudio / Hornung, Antonie / Rega, Lorenza (eds.): *Perspektiven Eins. Akten der 1. Tagung Deutsche Sprachwissenschaft in Italien (Rom, 6.-7. Februar 2004)*. Rom: Istituto Italiano di Studi Germanici, 345–363.
- You, Evan** (2020): *Vue.js. The Progressive JavaScript Framework*. <https://vuejs.org> [03.11.2022].