

Mapping the (Digital) Linguistic Atlas of Scotland

Pluschkovits, Markus

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

Kirk, John

john.kirk@univie.ac.at
University of Vienna, Austria

The Linguistic Atlas of Scotland was an atlas project carried out by Mather & Speitel, with the first paper questionnaires being sent out in the 1950s until the eventual publication of the individual volumes throughout the 1970s and 1980s. Covering 42 counties of Scotland, England and Ulster, with over 200 000 individual responses to 153 lexical items from 1787 localities, the lexical section of the atlas is an impressive piece of research. This lexical section of the atlas, comprised of two volumes (volume 1 [1975] and volume 2 [1977]), is now the object of a digitization effort of a small team of researchers based in Vienna, aiming to showcase the value-added in the digitization of existing, analogue data.

This digitization effort follows three main goals: The first goal is to digitize the original atlas source material. As has been pointed out by reviewers of the original works, the word lists given in the appendices of the volumes offer considerable more detail and data than the actual maps themselves (Macaulay 1977: 226 & 1979: 226, McClure 1976: 230). These maps, for example, sometimes do not include all main responses to an individual item. The starting point of the digitization process was therefore these word lists, and not the actual maps themselves. The second effort of the digitization process was an enhancement of the actual maps. While the original maps were drawn carefully and are generally very legible, they suffer the drawback of being area-maps, indicating variants with different cross-hatchings. In some areas where four or more of these overlap, they become somewhat difficult to read (McClure 1976: 230). Additionally, area maps suffer from the epistemological drawback that they suggest continuity and homogeneity, even where heterogeneity is present in the data. Furthermore, the crosshatching sometimes covers areas where no data was actually collected. In the Digital Lexical Atlas of Scotland (henceforth DLAS), the data is therefore mapped as georeferenced data points. A third major goal of the digitization process concerns the linguistic aspect: the DLAS additionally aims to re-analyze the original data and employ a consistent method based on lemmatization of the individual variants (see e.g. Kirk & Hessel 2020, Hessel & Kirk 2020, Kirk & Pluschkovits [forthcoming]). Taken together, these three goals provide the basic guidelines of the digitization process. The digitization, which is a collaborative effort with project part 11 of the Special Research Programme “Deutsch in Österreich” (‘German in Austria’, see DiÖ 2022a), was carried out using a relational PostgreSQL database and several tools developed for the project, such as an OCR-tool optimized to the layout of the word lists and a mapping tool to locate informants easily (all of which has been developed open source and is available via GitHub, see DiÖ 2022b), as well as the help of student assistants to correct data. The data model employed not only allows for the creation of maps of individual variables, but also the mapping of different variants of the same variable on one map, and even the mapping of the original responses. All the completed maps ba-

sed on the reanalysed material are publicly available (see DLAS 2022).

The poster focuses on goals one and two of the digitization effort, detailing the process of the digitization of the original word lists and the new mapping possibilities. A special emphasis is put on our utilization of the layout of the printed pages during the OCR process to automatically classify data as either informants, responses (‘variants’) or variables and the data model this is grounded in. A similar strategy of using layout and punctuation for data classification is e.g. used by Zacherl for the digitization of the *Romanische Etymologische Wörterbuch* (see Zacherl 2022). Furthermore, an additional focus is put on the design of the maps and the advantages of interactive, digital mapping over paper maps. In addition to the maps based on the reanalyzed material, the original responses can be mapped by users themselves using the “Explore”-tab, enabling them to create custom maps based on the original data of the Linguistic Atlas of Scotland.

Overall, the project aims not only to offer a re-analysis of the rich corpus of the lexical section of the original Linguistic Atlas of Scotland, making the original data more accessible and offer an easy way to map it, but also to showcase the opportunities arising from digitization of (linguistic) data. Presenting these opportunities of digitization – and the path to them – is the central objective of this poster.

Bibliography

DiÖ (2022a): *The Special Research Programme ‘German in Austria’*. <https://www.dioe.at/en/about-dioe/overview> [04.11.2022].

DiÖ (2022b): *Github Repositories of the Special Research Programme Deutsch in Österreich (‘German in Austria’)*. Github Repository <https://github.com/german-in-austria> [06.04.2023].

DLAS (2022): *The Digital Lexical Atlas of Scotland*. <https://lasdb.dioe.at/> [04.11.2022].

Hessel, Christian / Kirk, John (2020): “Digitising Collections of Historical Linguistic Data: The Example of the Linguistic Atlas of Scotland”, in: *Journal of Data Mining and Digital Humanities. Special Issue Special Issue on Visualisations in Historical Linguistics*: 1–17.

Kirk, John / Hessel, Christian (2020): “Towards a Digital Lexical Atlas of Scotland”, in: *Scottish Language* 39: 1–39.

Kirk, John / Pluschkovits, Markus (forthcoming): “Introducing the Digital Lexical Atlas of Scotland”, in: *The languages of Scotland and Ulster in a global context, past and present: Proceedings of FRLSU13, München, 28-30 October 2021 (= Publications of the Forum for Research on the Languages of Scotland and Ulster 6)*.

Macaulay, Ronald (1977): “Review of Mather, J. Y.; Speitel, H. H. (eds.). 1975. The Linguistic Atlas of Scotland. Scots Section. Volume 1 (London: Croom Helm, Ltd.)”, in: *Language* 53, 1: 224–228.

Macaulay, Ronald (1979): “Review: The Linguistic Atlas of Scotland: Scots Section, Vol. 2 by J. Y. Mather and H. H. Speitel”, in: *Language* 55, 1: 224–228.

Mather, J. Y. / Speitel, H.-H. (1975): *The Linguistic Atlas of Scotland. Scots Section. Volume 1*. London: Croom Helm.

Mather, J. Y. / Speitel, H.-H. (1977): *The Linguistic Atlas of Scotland. Scots Section. Volume 2*. London: Croom Helm.

McClure, Derrick J. (1975): “Review of Mather, J. Y.; Speitel, H. H. (eds.). 1975. The Linguistic Atlas of Scotland”, in: *American Speech* 53, 3–4: 223–234.

Zacherl, Florian (2022): “Linguistische Online-Ressourcen auf Basis traditioneller Werke. Anforderungen und digitale Möglichkeiten am Beispiel des Romanischen Etymologischen Wörterbuchs“, in: *apropos [Perspektiven auf die Romania]* 9: 254–275.