

Database of Southern Dutch Dialects: demonstrating the opportunities for digital lexicographical research

Veronique De Tier, Jacques Van Keymeulen, Roxane Vandenberghe, Lien Hellebaut
Department of Linguistics, Ghent University Department of Linguistics, Ghent University and
Sally Chambers Ghent Centre for Digital Humanities, Ghent University

Jesse de Does, Katrien Depuydt, Tanneke Schoonheim, Dutch Language Institute

Keywords: dialectology, lexicography, linguistic data, language atlases, geo-visualisation

The Database of the Southern Dutch Dialects (DSDD) aims to aggregate and standardise three existing comprehensive dialect lexicographic databases of the Flemish, Brabantic and Limburgian dialects into one integrated dataset. The project (2017-2020) will result in a harmonised dataset of concepts, an API for researchers and a user-friendly search engine in which the outcomes of the project will be available to a broad audience.

Although the dictionaries were explicitly set up in parallel in order to make an aggregation possible, there are some methodological and lexical differences between the dictionaries. There is also a significant degree of heterogeneity from a technical point of view, in terms of file formats and logical structure. In order to undertake the linguistic integration of these concepts the Dutch Language Institute (INT)'s Lex'it tool was used.

In the demonstration we will show how a workflow to harmonise the data structures and to interconnect the dictionaries by adding an overarching layer of concepts has been set up. Once integrated, the database will enable innovative research, especially in the field of quantitative lexicology and dialect-geographical analysis.

More specifically a number of aspects will be demonstrated:

- **the integrated database with the use of Lex'it:** this part of the demonstration will allow participants to explore the data model for the integrated database, both from a technical and a linguistic point of view. Participants will be able to see how the linking capabilities within Lex'it have enabled dialectologists to address methodological problems of dealing with closely related but non-equivalent concepts that appear across the three separate dictionaries.
- **the beta-version of the user-friendly search engine** which enables researchers and the general public to search the integrated dataset. Participants will be able to try out the

various search functions for finding the different dialect words for one particular concept, all the dialect words used in one place/region and the multiple meanings of a particular dialect word. Additionally, participants will be able to display their results on a flexible dialect map using the integrated cartographic tools. They will be able to experiment with customising the maps according to their research needs, such as: selecting base maps with different regions (e.g. a province, larger or smaller dialect areas, ...), changing the legend of the map or selecting the colours of the symbols to visualise, for example, etymological clusters of words.

- **the prototype API to enable geo-visualisation of dialect data** which builds on the feedback received following our presentation at DH Benelux 2018¹ and the results of a research workshop which took place in Leiden in March 2019. Participants will be able to test the prototype application which makes it possible to automatically generate linguistic maps with dialect data provided directly from the integrated database via the API.

Through this demonstration it is intended that the interaction with the DH Benelux community will provide valuable feedback on the prototype applications being developed in the DSDD project.

¹ Chambers S., Hacigüzeller, P., Van Keymeulen, J., De Tier, V. Triest, L., Vandenberghe, R. (2018) [Geo-visualising dialect data: the experience of the Database of Southern Dutch Dialects \(DSDD\)](#)