Geo-visualising dialect data: the experience
of the Database of Southern Dutch Dialects (DSDD)

Sally Chambers and Piraye Hacıgüzeller,
Ghent Centre for Digital Humanities, Ghent University

Jacques Van Keymeulen, Veronique De Tier, Liesbet Triest and Roxane Vandenberghe
Department of Linguistics, Ghent University

Geo-visualisation of data is fundamental to dialect research. Dialectologists have been creating and interpreting language maps for many years[1]. Language maps enable dialectologists to present a large amount of data in a clear and understandable manner. Particular examples include: (a) **objective maps** enabling researchers to display the raw language data prior to interpretation, (b) **isogloss maps** which use lines (isoglosses) to mark the border between two language variants and (c) **symbol maps** where for example, the same dialect words are used in different geographical locations.

In 2016, the Research Foundation Flanders funded a medium-scale research infrastructure project, the **Database of Southern Dutch Dialects (DSDD)[2].** The aim of the DSDD project is to aggregate and standardise three comprehensive dialect lexicographic databases of the **Brabantic[3]**, the **Limburgian[4]** and the **Flemish Dialects[5]** into one database for the Southern Dutch Dialect area. Now that the integration of the three dictionaries is underway in close collaboration with the Dutch Language Institute (INT[6]) using their linguistic tool Lex-IT, the project team can turn their attention to exploring how to geo-visualise this integrated dialect dataset, using state-of-the-art web-mapping techniques.

Originally, dialect maps were created using the commercial software, MapInfo[7]. With thanks to our technical partner,[8] the MapInfo software has been adapted so that symbol maps could be automatically generated from a range of different base maps. Frequently used symbols were also directly included in the software. The symbol maps were generated on the basis of Kloeke codes included in the dialect data[9]. While this system works perfectly well for the offline creation of dialect maps, for example for printed dictionaries, the current set-up does

---

[1] See for example: http://www.dialectloket.be/tekst/dialectologie/dialectkaarten/
[2] Database of Southern Dutch Dialects (DSDD): http://www.ghentcdh.ugent.be/projects/database-southern-dutch-dialects-dsdd
[3] Dictionary of the Brabantic Dialects: https://e-wbd.nl
[4] Dictionary of the Limburgian Dialects: http://e-wld.nl
[5] Dictionary of the Flemish Dialects: https://e-wvd.be
[6] Dutch Language Institute - Instituut voor de Nederlandse Taal: http://www.inl.nl
[7] MapInfo: https://www.pitneybowes.com/us/location-intelligence/geographic-information-systems/mapinfo-pro.html
[8] Info Service Belgium (ISB): https://www.infserv.com/en/
[9] Kloeke codes are unique codes for places and hamlets in the Dutch language area. They were designed by dialectologist G. G. Kloeke (http://www.dbnl.org/tekst/anro001bioe01_01/kloe004.php) in the 1920s and are currently maintained by the Meertens Institute in Amsterdam, see: https://www.meertens.knaw.nl/kloeke/

not meet the online needs of 21ˢᵗ Century dialectology[10]. The project team is therefore looking for a new solution to geo-visualise the integrated dataset of the Southern Dutch Dialect Area.

A first option could be to visualise the data on a Google map, as is in case for the Dictionary of Flemish Dialects, as shown in Figure 1. By linking the Kloeke codes in the dialect data set with geographic coordinates, such as latitude and longitude, the dialect data can be published on a Google map. However, a disadvantage of these maps is that they cannot be easily adapted by the end-user.
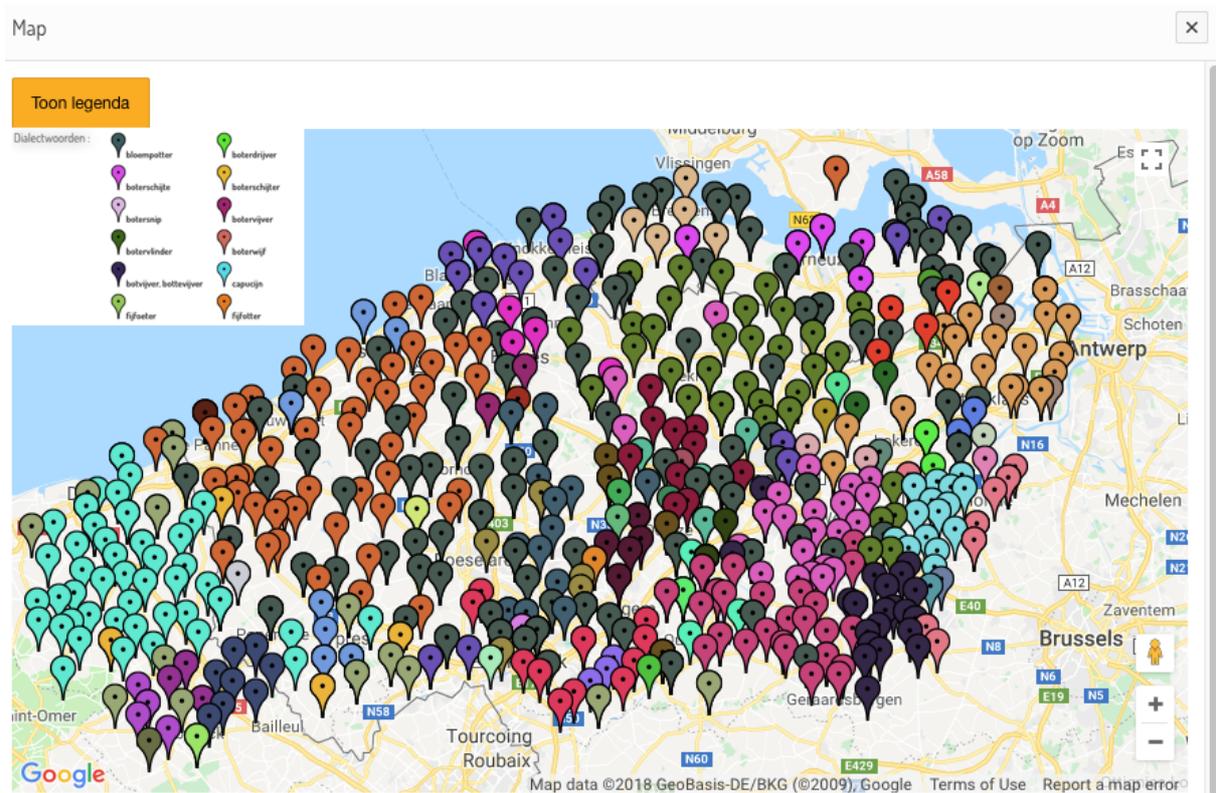

Figure 1. Dialect map visualising the dialect words for 'butterfly' on a Google Map

The aim of this short paper is to present the findings of our desk research to: a) identify and review existing digital solutions for the geo-visualisation of language / dialect data, b) identify web-mapping technologies from beyond the field of linguistics which could be adapted for geo-visualisation of dialect data and c) report on initial experiments conducted together with the Department of Geography on visualising the dialect data. By having the opportunity to present our ongoing research, we hope to be able to open up these questions for discussion and debate, drawing on the expertise of the DH Benelux community.

Keywords: geo-humanities, language atlases, geo-visualistion, dialectology, digital humanities

---

[10] For example, the adapted version of the software is based on an out-of-date version of the MapInfo software, the licence of which, means that it can only be used offline. As MapInfo is not open source, the costs of upgrading the licence are prohibitive for a small research team.