

From Dialectology to Dialectometry 2025

Organiser: [Matthew Sung](#) (LUCL/ LUCDH PhD Candidate)

Email: [h.w.m.sung\[at\]hum.leidenuniv.nl](mailto:h.w.m.sung@hum.leidenuniv.nl)

Location: Huizinga Room 0.09 (LUCDH Digital Lab), Leiden University

Time: 10:00-12:00

Dates: Every Tuesday from 30th September to 9th December 2025

Brief description:

Dialectology, the study of language variation across geographical space, may not be a subject that is as accessible to everyone as before. This workshop series aims to introduce methods used in traditional dialectology, including synchronic comparison, historical reconstruction, data digitization and visualisation, as well as more recent computational approaches¹ in studying geographical variation of language. The workshops will consist of hands-on activities and discussions, including the creation of digital maps and calculating pronunciation distances. We will also work on a small digitisation project with a published dialect survey of English.

Workshop schedule:

	Title	Date
1	Introduction: What is a ‘dialect’? What is ‘dialectology’?	30/09/2025
2	Linguistic Atlases, Dialect Surveys & Linguistic Maps	07/10/2025
3	Data Extraction & Digitisation	14/10/2025
4	Mapping dialect data with QGIS	21/10/2025
5	Dialect Comparison and Historical Reconstruction	28/10/2025
6	What is Dialectometry? Why Dialectometry?	04/11/2025
7	Dialect Distances & Distance Matrix	11/11/2025
8	Cluster analysis & Multidimensional Scaling	18/11/2025
9	Dialectometry Beyond Classification	25/11/2025
10	Corpus Dialectometry	02/12/2025
11	Contemporary Approaches	09/12/2025

¹ No programming is required and involved.

Details:

1. *What is a 'dialect'? What is 'dialectology'?* (30th September)

The first workshop will start with a discussion on what 'dialect' and 'dialectology' are. Linguists have been arguing what a 'dialect' is for a long time, and it seems like even until now, no one has an answer. How do dialectologists get over this problem? 'Dialectology' is also not really taught at universities, what has happened to the discipline? Has it 'transformed' into something else? What counts as dialectology nowadays? After the discussion, we will look at the development of dialectology from its early days up to the present day.

2. *Linguistic atlases, dialect surveys and linguistic maps* (7th October)

This workshop will involve students going to the library to find a linguistic atlas/ dialect survey of their choice, and they will have to scan or take a picture of 2 maps from the source and bring it to class. The aim of this task is to explore the possible types of maps in atlases from languages and traditions as well as to look at how dialectologists classify and present data in these atlases.

3. *Data Extraction and digitization* (14th October)

Since the dialect data are printed on paper, we cannot really do much with it. This workshop will introduce some methods in extracting dialect survey data onto a computer (a .csv file), with examples from the *Survey of English Dialects* and another atlas. The methods introduced in this workshop will be used in our digitization project.

4. *Mapping Dialect Data with QGIS* (21st October)

Maps are the most important visualization tools for dialect geographers. After extracting the data from a linguistic atlas, we may want to plot the data on a map in another way, for example, overlaying two maps to find similarities in the distribution of the variations, or plotting the frequencies of certain variants across different localities. In this workshop, we will go over the basics of QGIS, an opensource mapping software. The maps covered in this workshop include point (symbol and text) maps, area maps and frequency maps, and you will have a hands-on experience on creating these maps yourself.

5. *Dialect Comparison and Historical Reconstruction* (28th October)

Dialectologists are interested in both the synchronic and diachronic aspects of dialectal variation. In the first half of this workshop, you will see a range of methods in dialect comparison, such as comparing vowels using different reference systems. The

geographical patterns we see on maps are shaped by history. In the second half, we will look at some of these maps and data and try to reconstruct the linguistic history.

6. What is dialectometry? Why dialectometry? (4th November)

In the first half of the work, we will first assess the pros and cons of the traditional methods used in dialect classification, followed by an introduction to what dialectometry is and an explanation to how they can overcome some of the problems traditional approaches have. The workshop will end with a brief history of the field.

7. Dialect Distances & Distance Matrices (11th November)

Instead of using isoglosses to assess the relationship between dialects, dialectometrists are interested in calculating ‘dialect distances’, i.e. how much a dialect differs from another. There are numerous ways to calculate dialect distances (depending on your data), but this workshop will focus on several methods which are often found in the literature, e.g. Relative Distance Value (Goebel 2018) and Levenshtein Distance (Heeringa 2004). Once we have obtained dialect distances, we can then perform further analyses by 1) making use of the distances directly from the distance matrix (e.g. reference-point maps and network maps), and 2) perform cluster analysis and multidimensional scaling. In this workshop, we will look at the analyses stated in 1).

8. Cluster Analysis & Multidimensional Scaling (18th November)

After obtaining the dialect distances, we can make use of computational methods on the distances we calculated to find the relationship between the dialects in our data. Cluster analysis groups similar dialects together and yield dialect groups; multidimensional scaling (MDS) on the other hand is often used to visualise dialect continua. In this workshop, we will look at how these two methods work, as well as how to read the visualisations produced with these methods, including MDS plots, dendrograms and cluster/ MDS maps.

9. Dialectometry Beyond Classification (25th November)

Dialectometrical methods are often used for dialect classification, but it can do much more than that! One of these uses includes the application in sociolinguistics; we can look at whether political borders have any impact on dialects, as well as looking at dialect change and the formation of regiolects. Other uses of dialectometry also include tracing the origin of dialects and exploring the regularity of sound changes.

10. *Corpus dialectometry* (2nd December)

Many studies in dialectometry make use of linguistic atlas data. While these studies are showing a lot of insights, some criticised that the atlas-based data are not naturalistic enough, since data were collected through elicitation (Szmrecsanyi 2013: 3-4). This motivated scholars to turn to corpus-approaches in dialectometry, as corpus data reveal more about the context and magnitude of the usage of linguistic features (Kuparinen and Scherrer 2024).

11. *Contemporary Approaches in Computational Dialectology* (9th December)

Dialectometry, whether it's Goebel's (1982, 1984) *Relative Identity Value/ RIV* or the use of *Levenshtein Distance* in measuring dialect distances (Heeringa 2004), these methods were developed 20 years ago (for RIV, 40 years ago). Beyond 'classic' dialectometry, what other methods in dialectology have been developed in the past 20 years? In this workshop, we will go on a quick tour on the latest approaches in the field, including some on-going research in Leiden.