Introduction to Dialectometry I

Wilbert Heeringa

German Academic Exchange Service – DAAD

University of Bielefeld, Faculty of Linguistics and Literary Studies

Frisian Academy

Abidjan, December, 19–23, 2016

Overview

Cluster analysis – Multidimensional scaling

Classification of dialects

- Cluster analysis
  - Find dialect **groups**.
  - Introduced by Goebl ($\pm$ 1982) in dialectometry.
- Multidimensional scaling
  - Visualize dialect **continuum**.
  - Introduced by Embleton (1993) in dialectometry.
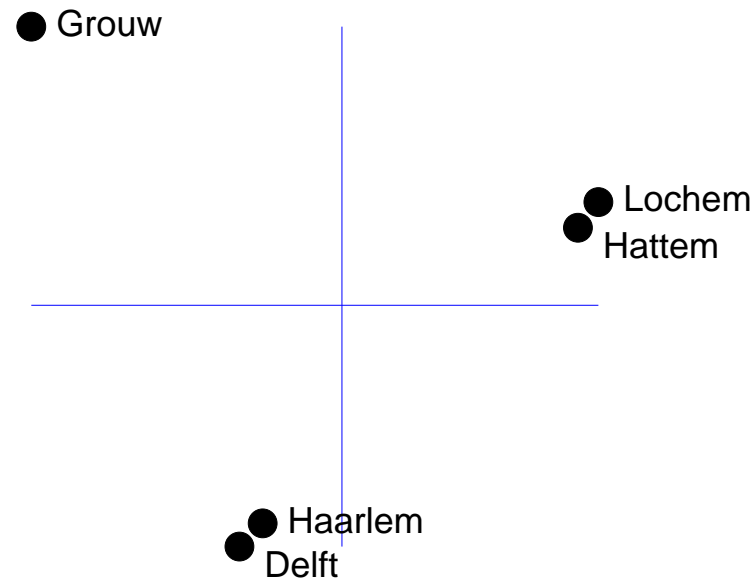
Multidimensional scaling

- Given a geographic map, distances between locations can be measured.

- Multidimensional scaling: given distances, locations on a map can be inferred.

- In our case: from $n \times n$ distances we infer coordinates in 2- or 3-dimensional space. So n dimensions are reduced to two or three.

- Types of multidimensional scaling: classical, Kruskal's non-metric multidimensional scaling (used in the examples here), Sammon mapping.

Multidimensional scaling

|	Grouw	Haarlem	Delft	Hattem	Lochem
Grouw	0	41	44	45	46
Haarlem	41	0	16	34	36
Delft	44	16	0	37	38
Hattem	45	34	37	0	20
Lochem	46	36	38	20	0

Put the five local dialects on a map so that the distances in two-dimensional space reflect the distances in the matrix as close as possible.

Multidimensional scaling

Using MDS the 5 dimensions are reduced to 2. X-coordinates represent the first and Y-coordinates represent the second dimension. The two dimensions explain 98.0% of the variance in the original distances.

Multidimensional scaling
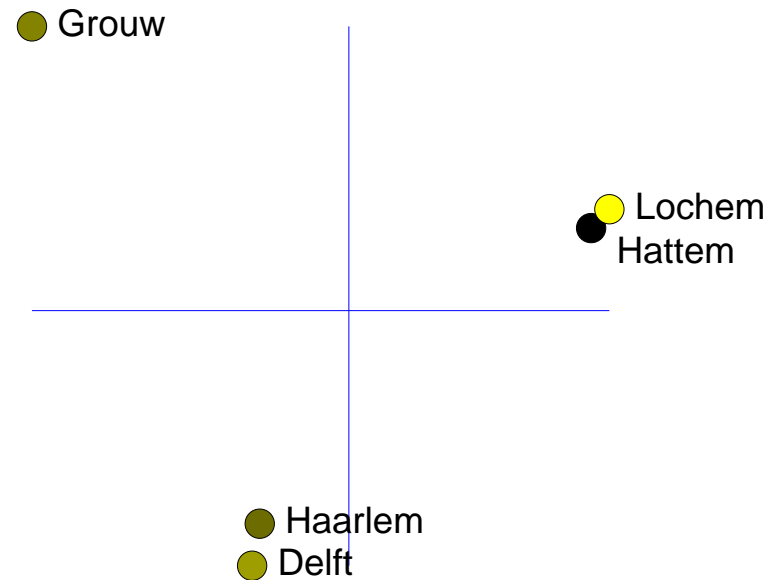
Using MDS the 5 dimensions are reduced to 3. X-coordinates represent the first, Y-coordinates represent the second, and greytone represents the third dimension. The three dimensions explain 96.4% of the variance in the original distances.
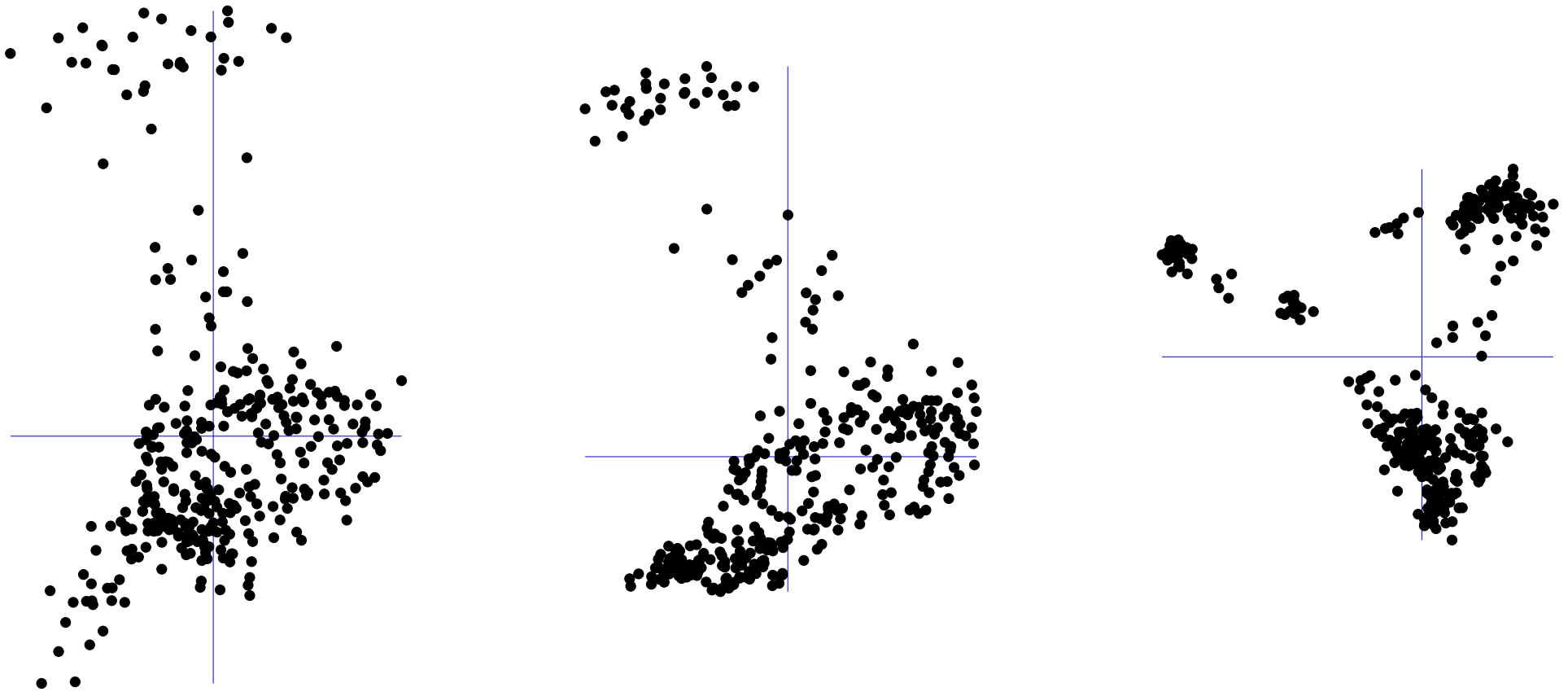
Using MDS the 361 dimensions are reduced to 2. They explain 84.3% (left: lexical relative difference values), 45.0% (middle: lexical weighted difference values) and 51.9% (right: pronunciation Levenshtein distances) of the variance in the original distances. Labels are omitted.
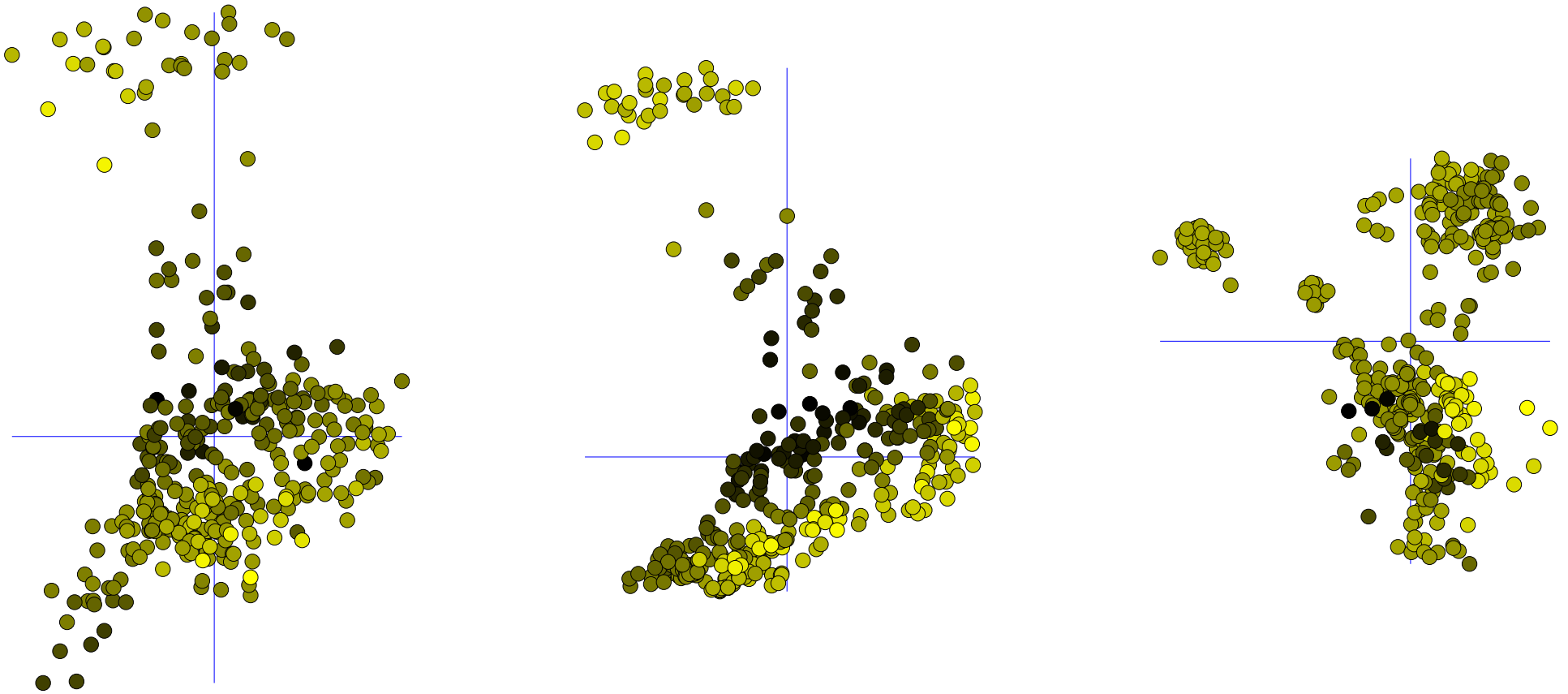
Using MDS the 361 dimensions are reduced to 3. They explain 89.5% (left: lexical relative difference values), 50.7% (middle: lexical weighted difference values) and 88.4% (right: pronunciation Levenshtein distances) of the variance in the original distances. Labels are omitted.

Final remarks

- The maps were produced with RuG/L04, developed by Peter Kleiweg, and available at: `http://www.let.rug.nl/kleiweg/L04/`.

- Gabmap is a web application made for dialectologists and students, and available at: `http://www.gabmap.nl/`.

DAAD
Deutscher Akademischer Austausch Dienst
German Academic Exchange Service

Universität Bielefeld

FRYSKE AKADEMY