Gabmap exercise I

Measuring Dutch dialect variation

### USB version

Introduction

We measure distances in the sound components among 348 Dutch dialect varieties. On the basis of these distances we will classify the dialect varieties by means of cluster analysis and multidimensinal scaling. We use GpsPrune, LibreOffice Calc and Gabmap, a computer program which has been developed by Peter Kleiweg (University of Groningen, Faculty of Arts, Humanities Computing).

1. Inspecting the base map

The base map consists of two components: 1) the contour of the area under investigation (in our case: The Netherlands and Flanders) and 2) the locations (in our case 348 locations of the Dutch dialect varieties). Useful manuals that explains how to create the base maps are found in the folder /Gabmap/manuals/: 'Creating maps with Google Earth.pdf' (written by Peter Kleiweg) and 'Creating maps with Google Maps.pdf' (written by Cagri Cöltekin). You can view the documents with the Document viewer which is found on the desktop.

In this course we do not need to make the base map yourself. We will use the map 'DutchRND.kml' in the folder /Gabmap/datasets/. Open this file with GpsPrune which is found on the desktop. Start GpsPrune and make the window as large as possible. Choose 'File' and subsequently 'Open'. Set 'File of Type' to 'KML files'. Browse to the folder /Gabmap/datasets/, choose 'DutchRND.kml' and click on 'Open'.

On the left you find the list of locations. You can find a location on the map when you click on the name of the location.
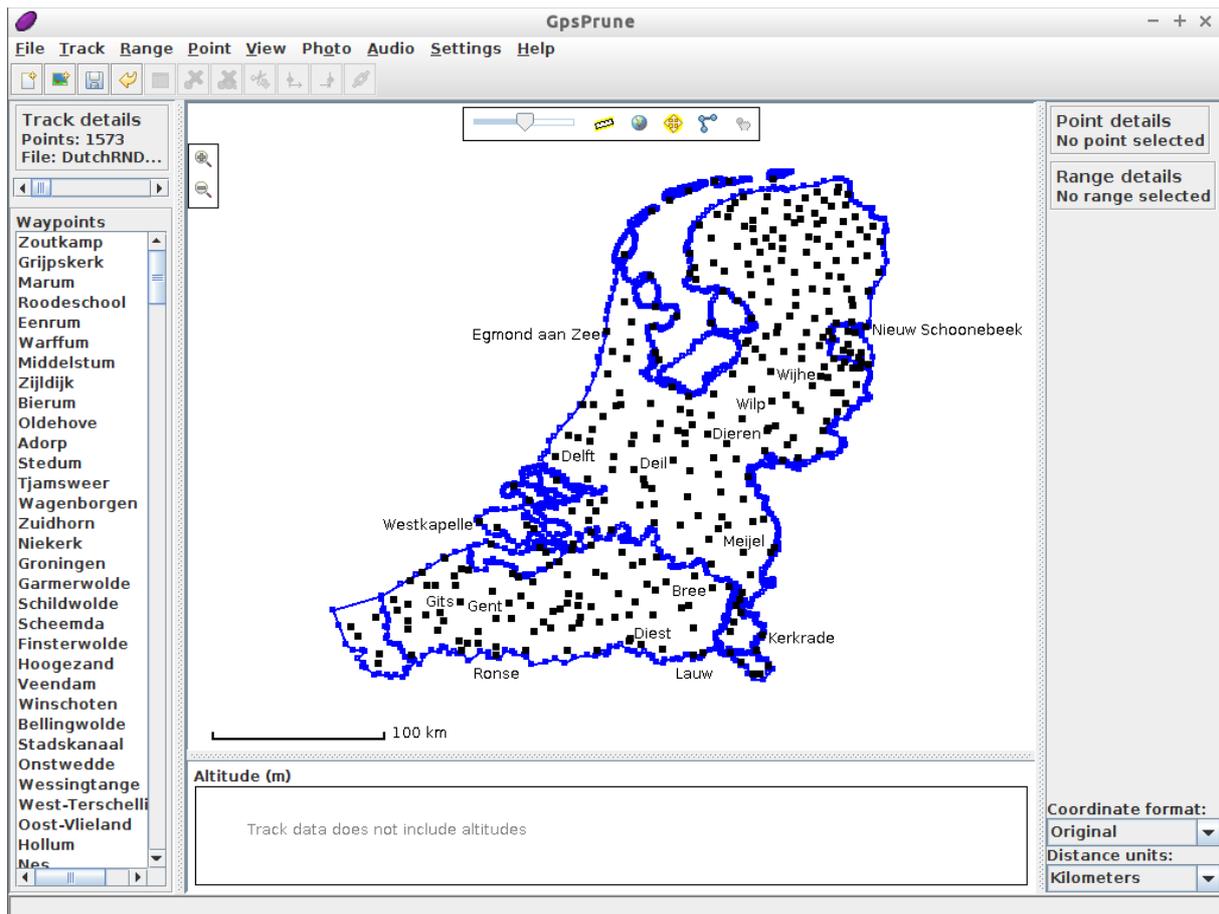
2. Inspecting the database

The data we use is taken from the *Reeks Nederlands Dialectatlassen*, a series of atlasses compiled by E. Blancquaert and W. Pée and published in the period 1925-1975. The data we use is digitized in Groningen (University of Groningen, Faculty of Arts, Information Science).

We use data which has been made suitable for our research. The table is found in the file 'DutchRNDPro.csv' in the folder /Gabmap/datasets. The table can be used unmodified later in Gabmap, but for now we have a quick look at the table by loading it in LibreOffice Calc. Start LibreOffice which is found at the desktop. Choose 'File' and subsequently 'Open'. Browse to the folder /Gabmap/datasets/ and select the file 'DutchRNDPro.csv'. Click on 'Open' (in the lower right corner). Choose the 'Tab' under 'Separator Options'. Click on 'OK'.

At the left you find the names of the 348 Dutch dialects. On top you find the column names which represent the 125 items. Per item the transcriptions of the pronunciations of the 348 locations are given. If you want to make this kind of table yourself, please consult the manual 'Preparing dialect data for Gabmap.pdf' found in the folder /Gabmap/manuals/.

3. Performing analyses with Gabmap

Switch to Firefox by clicking on the Firefox tab in the taskbar at the bottom of your screen. Nearly on top of your screen you find 'Your project'. Enter a title under 'Description'. Under 'Map' upload the map file 'DutchRND.kml' from the folder /Gabmap/datasets/. Under 'Data' upload the data table file 'DutchRNDPro.csv' in the folder /Gabmap/datasets/. Choose under 'Type of data' for 'string data'. Choosing this causes an extra option to appear: 'Type of processing'. Choose 'string edit distance – user defined'. Upload the feature definition file 'IPA.def' in the folder /Gabmap/datasets/. Click on 'Create project' at the bottom of the window. Right now we can start with the real analyses!

3.1 Differences: statistics and difference maps

Click after 'Differences' on 'statistics and difference maps'. You get a page which shows several statistics (Cronbach's Alpha and Local incoherence) and a bar graph. When scrolling down you find two maps, in which the 348 locations are connected to each other by lines. Locations which are similar (i.e. which have small pronunciation distances) are connected by darker lines, and linguistically distant locations are connected by lighter lines. In the first map only neighbouring locations are connected, in the second map geographically more distant locations are connected by lines too.

If you want, save each of the maps by clicking on 'png' top right of each map, and choose 'Save'. Choose a proper file name and spot on your hard disk or usb stick. Describe the pattern the two maps show.

3.2 Discrete clustering: cluster maps and dendrograms

Click on 'project 1' in order to return to the main menu. Click on 'cluster maps and dendrograms' after 'Discrete clustering'. You arrive at a page which shows a map and a dendrogram at the top. Right under the (first) dendrogram you find two parameters. You need to change both of them.
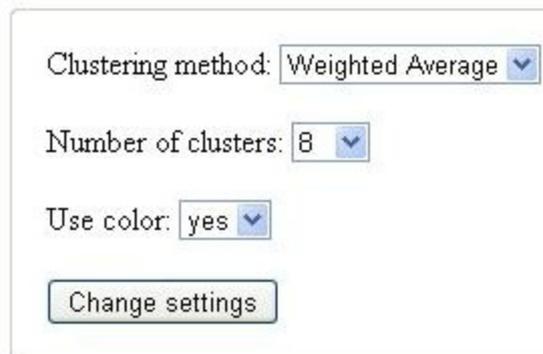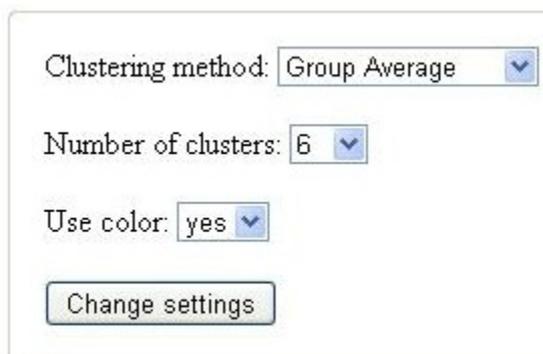

Choose 'Groups Average' as clustering method and choose six clusters.

Click on 'Change settings'. Maps and dendrograms are now changed. In the map the 348 Dutch dialect locations are classified in six areas. Right under the map we find a small visualisation of the dendrogram (without labels), and right under the box with the parameters settings we find a larger dendrogram (with labels). The colors or the branches and labels in the dendrograms correspond with the colors in the map.

Save both the map and the larger dendrogram. For each of the pictures click on 'png' found top right of each picture and choose 'save'. Choose a name and save the picture on your hard disk or usb stick. Does the

division in six groups agree with the division suggested by the line maps? Why or why not?

3.3 Multidimensional scaling: mds plots

Click on 'project 1' in order to return to the main menu. Click on 'mds plots' after 'Multidimensional scaling'. You will find a two-dimensional multidimensinal scaling plot.

If you hover the mouse over the dots, you will see the location labels. If you want to have shown particular location labels permanently in the plot, you can choose them in the list below the plot.

3.4 Multidimensional scaling: mds maps

Click on 'project 1' in order to return to the main menu. Click on 'mds maps' after 'Multidimensional scaling'. Now you get a screen which shows a map on top. Below the map you find some text:

Above: Classical MDS in three dimensions mapped onto RGB color space.

Below: Classical MDS into multiple dimensions, one map per dimension.

Below these text lines you find six maps and two graphs. In this exercise we focus on the map on the top only. Save this map, click on 'png' and choose 'Save'. Compare the map with the line maps and the cluster map. Does the MDS map suggest the same division? Which differences do you find? Which map gives the clearest picture? Which map (line maps, cluster map, MDS map) do you find the best map? Why?

3.5 Probability dendrogram and cophenetic multidimensional scaling maps

Click on 'project 1' in order to return to the main menu. Click on 'probabilistic dendrogram' after 'Fuzzy clustering'. On top you find a probabilistic dendrogram, and right under this dendrogram the settings. Look at the settings, set 'method' to 'group average' and click on 'Update settings'.

Just have global look at the dendrogram and try to find the clusters which are 100% sure.

If you scroll down, you find the cophenetic multidimensional scaling map below the settings. Compare this map with the maps you made by cluster analysis (section 3.3) and multidimensional scaling (section 3.4).

3.6 Linguistic difference versus geographic distance

Click on 'project 1' in order to return to the main menu. Click on 'linguistic difference ↔ geographic distance' after 'Differences'.

We focus on the graph and analysis on top. In the graph the blue line is the fitted line between the geographic distance and the linguistic difference. What does the shape of the fitted line tell us?

Right below the graph the relationship between the logarithmic geographic distance and the linguistic distance is analyzed by a so-called linear regression analysis. Look under 'Pr(>|t|)'. Is the relationship significant?

What are the values of the multiple R-squared and the ajusted R-squared? What is the correlation *r*? Hint: the correlation is found by taking the square root from the multiple R-squared.

3.7 Reference point maps

Click on 'project 1' in order to return to the main menu. Click on 'reference point maps' after 'Differences'. We will compare the 348 dialect varieties to the dialect of Haarlem, since Haarlem is recognized as the dialect that is linguistically closest to standard Dutch. Select the place 'Haarlem' after 'Place'. Keep 'linear distances' after 'Method'. Choose the option 'Reverse colours'. Click on 'Show map'.

You get a map in which Haarlem is marked as a star. All dialects are shown as small areas, polygons, and in each polygon the exact location is represented by yellow dots. Ech polygon has a color. Darker polygones represent dialect varieties which are linguistically distant to Haarlem, and lighter one represent dialects which are linguistically close to Haarlem.

Save the map. Click on 'png' found top right of the picture and choose 'save'. Choose a name and save the picture on your hard disk or usb stick. Describe what the map reveals.

Below the map you find a graph, in wich linguistic dialect distances compared to Haarlem are plotted against the geographic distances compared to Haarlem. Click on 'png' again and save the graph. Answer the question whether there is a relationship between linguistic dialect distance and geographic dialect distance.