

# **Gabmap exercise III**

## **Measuring Norwegian dialect variation**

### **WEB version**

#### **Introduction**

We measure distances in the sound components among twenty Norwegian dialect varieties. On the basis of these distances we will classify the dialect varieties by means of cluster analysis. We use Google Earth and Gabmap, a web application which has been developed by Peter Kleiweg (University of Groningen, Faculty of Arts, Humanities Computing). Make sure that Google Earth is installed on your computer. Additionally Libre Office Calc or Microsoft Excel should be available.

#### **1. Making the base map with Google Earth**

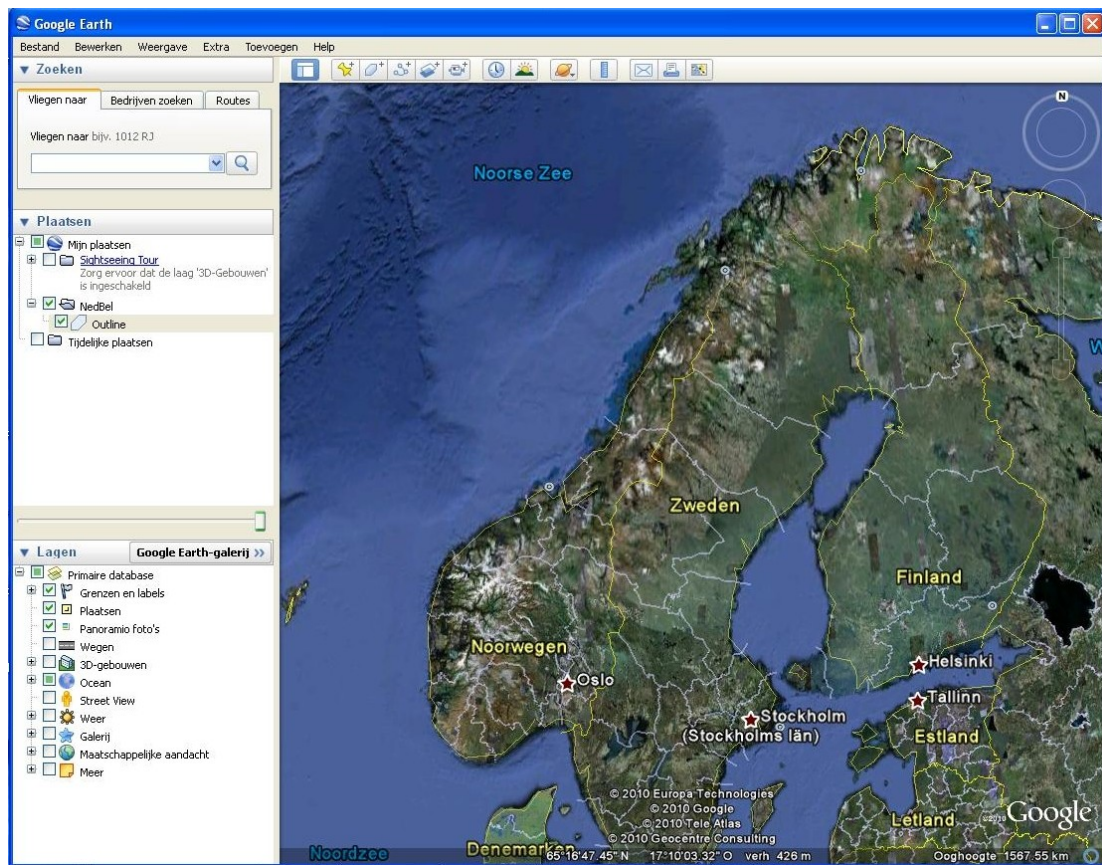
The base map consists of two components: 1) the contour of the area under investigation (in our case: Norway) and 2) the locations (in our case twenty locations of the Norwegian dialect varieties). Peter Kleiweg has written a useful manual which is available at: <http://www.let.rug.nl/~kleiweg/L04/kml/manual.html>. We partly follow the same steps, but here and there a bit simplified. Please start up Google Earth now.

##### **1.1 Making a new map**

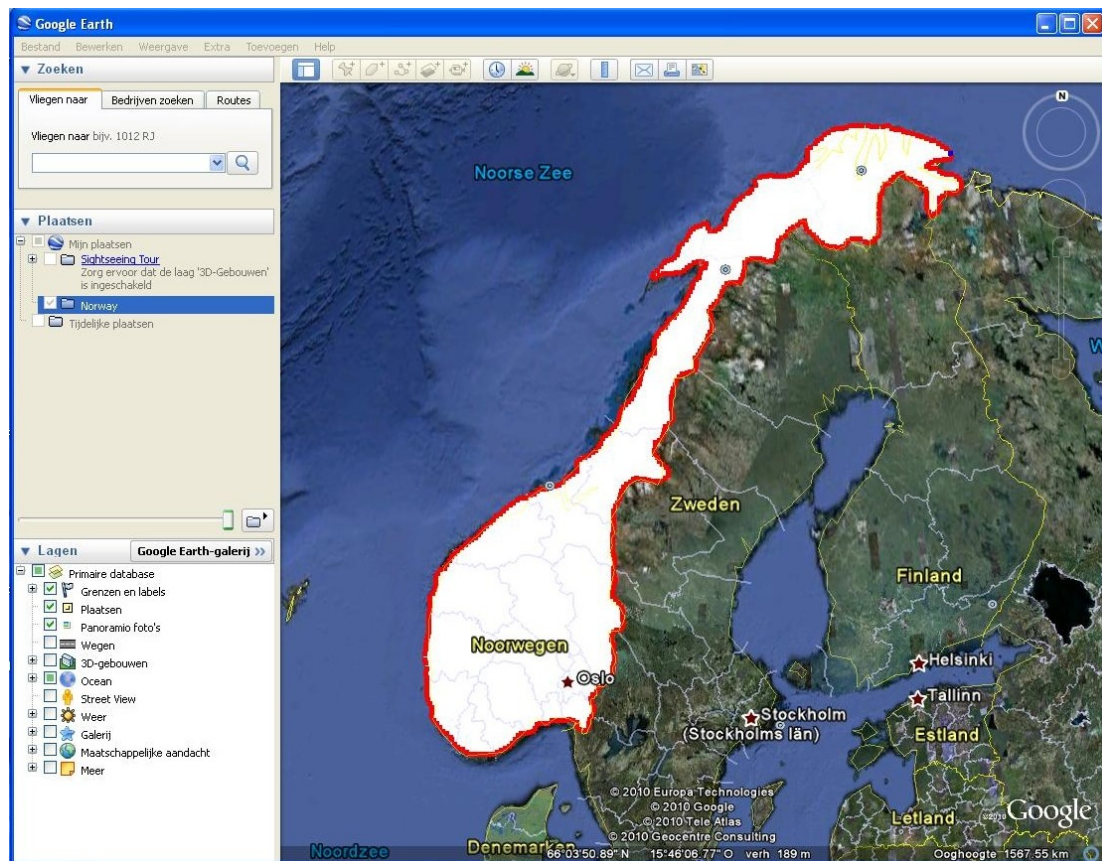
Right click on 'My places'. Choose 'Add' and next 'Folder' in the main menu on top of the screen. A little window 'Google Earth New Folder' will appear. Enter a name after 'Name': Norway. Click on 'OK'. Now you have an active map within the map 'My places'.

##### **1.2 Digitize the contour**

Visualize Norway as large as possible but so that it just fits within the window. You need to get a picture like this one:



Click in the toolbar on top of the map on the button `Add Polygon' in order to draw the contour. A small window 'Google Earth New Polygon' will appear. Enter a name after `Name': outline, but do NOT click yet on 'OK'. Move the little window aside. Move the mouse over the contour of Norway, so that you finally end where you started. Please see to it that the movement from begin to end is just ONE move, made continuously (this may be difficult!). Right now click on 'OK'. The result will look like this:

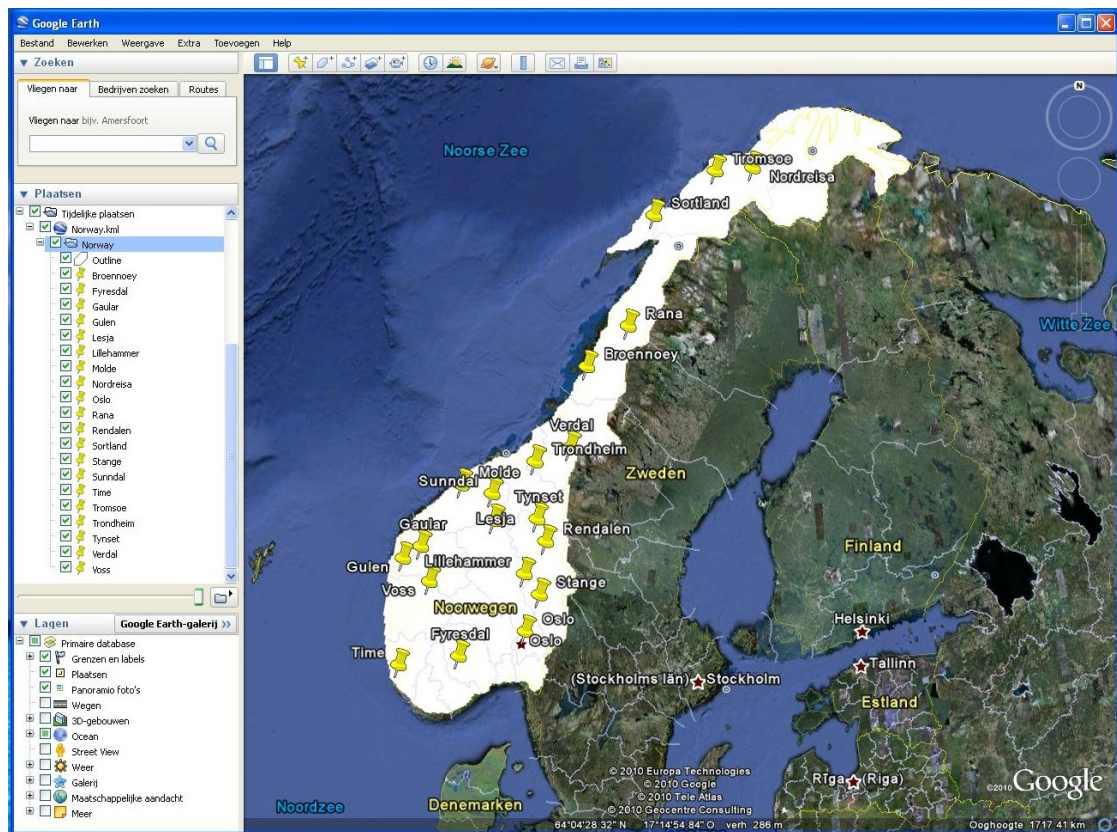


### 1.3 Digitize the locations of the dialect varieties

The following locations are involved into the investigation: Broennoey (enter as: Brønnøy, the ø is: alt + 0248), Fyresdal, Gaular, Gulen, Lesja, Lillehammer, Molde, Nordreisa, Oslo, Rana (enter as: Mo i Rana), Rendalen, Sortland, Stange, Sunndal, Time, Tromsoe (enter as: Tromsø), Trondheim, Tynset, Verdalen en Voss. Now we will enter the locations. For each location do the following:

Enter the name of the location at 'Fly to' and click on the magnifier. Now Google Earth will locate the location. Click on the button 'Add placemark' at the top. You get a new small window 'Add new placemark'. Enter the name of the location again after 'Name' in the spelling as given above (thus using 'oe' instead of 'ø'). Next click on 'OK'. The yellow place mark is put on the map now. The location should be put into the map Norway, which is the same map that contains outline!

After having entered all of the twenty locations, the results looks like this:



## 1.4 Save the map

Right click on the name of the folder *Norway* in order to open a menu, and choose 'Save Place As'. Although this may suggest that just one single location is saved, the complete map together with the outline and the twenty locations is saved.

## 2. Downloading the database

The data we use is compiled and transcribed by Jørn Almberg, phonetician at the University of Trondheim. The original data can be found at: <http://www.hf.ntnu.no/nos/> . We will use data which has been made suitable for our research.

In case you are using LibreOffice Calc go to: <http://www.wjheeringa.nl/dialect/datasets/NorwegianNOSPro.csv> and download the file. The table can be used unmodified later in Gabmap, but for now we have a quick look at the table by loading it in Libre Office Calc. Choose 'File' and next 'Open'. Browse to the location where you stored the file. Select the file and click on 'Open' (in the lower right corner). Choose the 'Tab' under 'Separator Options'. Click on 'OK'.

In case you are using Excell go to:

<http://www.wjheeringa.nl/dialect/datasets/NorwegianNOSPro.txt> and download the file. The table can be used unmodified later in Gabmap, but for now we have a quick look at the table by loading it in Excel. When opening the file in Excel, it is important to set 'File type' to 'All files' when loading the file. Choose the tab as delimiter between the columns.

Once the table is loaded, at the left you find the names of the twenty Norwegian dialects. On top you find the column names which represent the 58 items. Per item the transcriptions of the pronunciations of the twenty locations are given. If you want to make this kind of table yourself, please consult: <https://www.gabmap.nl/~app/doc/preparing/> .

### **3. Load the feature definition file**

Go to: <http://www.wjheeringa.nl/dialect/datasets/IPA.def> . Rightclick on the file and choose: Save Page As ... Store the file on your hard disk or USB drive.

### **4. Performing analyses with Gabmap**

Go to <http://www.gabmap.nl/> . Click under 'This is Gabmap' on the button 'start application'. In order to log in, you need to create an account. Enter your email address under 'Create an account' and choose a username and password (which need to be entered twice). Click on 'Create account'. An e-mail will be send to the e-mail address which you have just entered. The e-mail contains an url, click on this url in order to confirm your account. Now you arrive in Gabmap where you find the message that your account has been confirmed. Click on 'Home' in the upperleft corner in order to login in Gapmap. Enter your username and password under 'Log in'. Now you are logged in. Note that your account will be removed automatically if you do not used it during two months.

After being logged in, you enter the page 'Your project'. Enter a title under 'Description'. Under 'Map' upload the file with the base map which you have made with Google Earth. Upload the data table under 'Data'. Choose under 'Type of data' for 'string data'. Choosing this causes that an extra option appears: 'Type of processing'. Choose 'string edit distance - user defined'. Upload the feature definition file IPA.def. Click on 'Create project'. Right now we can start with the real analyses!

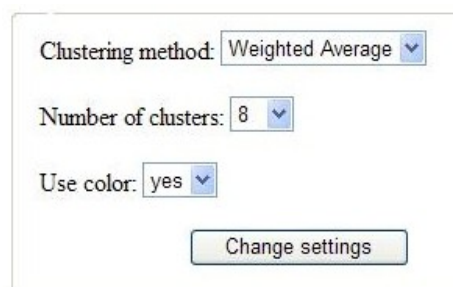
#### 4.1 Differences: statistics and difference maps

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 twenty 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.

#### 4.2 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.



Clustering method: Weighted Average ▼

Number of clusters: 8 ▼

Use color: yes ▼

Change settings

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

Clustering method: Group Average ▾

Number of clusters: 3 ▾

Use color: yes ▾

Change settings

Click on 'Change settings'. Maps and dendrograms are now changed. In the map the twenty Norwegian dialect locations are classified in three 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 three groups agree with the division suggested by the line maps? Why or why not?