Dialectologia 24 *(2020), 177-203.*
ISSN: 2013-2247

Received 20 February 2019.
Accepted 25 May 2019.

A GIS-BASED APPLICATION FOR DOCUMENTING AND ANALYSING PERCEPTIONS ABOUT LANGUAGE VARIATION

Xulio SOUSA[1], Soraya SUÁREZ[1], Rosa M. CRUJEIRAS[2] & Laura CALAZA[2]

[1]Instituto da Lingua Galega, Universidade de Santiago de Compostela* / [2]Departamento de Estatística, Análise Matemática e Optimización,
Universidade de Santiago de Compostela*

xulio.sousa@usc.es / sorayasuarezquintas@gmail.com / rosa.crujeiras@usc.es / laura.calaza@usc.es

Abstract

This article presents PDMapping, a new tool developed to assist the collecting and processing of data in research projects in Perceptual Dialectology. It allows information to be stored in a structured database and processed easily using either quantitative analysis-based statistical software or Geographic Information Systems (GIS). After presenting the background for development of PDMapping, we will explain how it works, show examples of it in use and provide some useful tips for using the application in other perceptual dialectology projects. Although specifically designed for an ongoing study based on a dialect identification task, PDMapping is easily adapted to other perceptual dialectology studies. The tool is available to the research community in an open source code repository and also in a demonstration web version on the project's website.

Keywords

folk linguistics, perceptual dialectology, mental maps, geographic information systems, digital humanities

©Universitat de Barcelona

UNA APLICACIÓN GIS PARA REGISTRAR Y ANALIZAR PERCEPCIONES SOBRE VARIACIÓN LINGÜÍSTICA

Resumen

En este artículo se presenta PDMapping, una nueva herramienta desarrollada para ayudar a la recopilación y procesamiento de datos en proyectos de investigación en dialectología perceptiva. Esta aplicación permite el almacenamiento de la información en una base de datos estructurada y su procesamiento mediante un programa que combina el análisis estadístico con la aplicación de un sistema de información geográfica (SIG). Después de presentar los antecedentes para el desarrollo de PDMapping, se explica su funcionamiento, se muestran ejemplos de uso y se proporcionan algunos consejos útiles para utilizar la aplicación en proyectos de dialectología perceptiva. A pesar de haber sido diseñado para un estudio en curso basado en una tarea de identificación de dialectos, PDMapping se adapta fácilmente a otras investigaciones de dialectología perceptiva. La herramienta está disponible para la comunidad académica en un repositorio de código fuente abierto y también en una versión web de demostración.

Palabras clave

lingüística popular, dialectología perceptiva, mapas mentales, sistemas de información geográfica, humanidades digitales

1. Introduction

Since their inception, within dialect studies there has been interest in the study of speakers' views about variation. Earlier dialectologists identified language varieties based on features chosen for both their linguistic relevance and their salience, that is to say, recognisability for ordinary speakers. According to Chambers & Trudgill (1980: 98), many of the features used by scholars to identify and delimit language varieties are precisely those which are most readily and consistently levelled or corrected by speakers.

The remarkable growth in interest in the study of non-specialists' perception of language variation in the late twentieth century led to a new field within the discipline of folk linguistics, called Perceptual Dialectology (Preston 1981; 1989; henceforth PD). From the start, PD was linked to dialectology while having methodological and theoretical connections with sociolinguistics. The approach focuses on how people perceive

©Universitat de Barcelona

geolinguistic variation, its main objective being to analyse non-linguists' knowledge, beliefs and attitudes concerning the geographical varieties of a language. It forms part of a broader area within language studies known as language regard, "a cover term for all approaches to the study of non-specialist belief about and reaction to language use, structure, diversification, history, and status" (Preston 2011a: 10; cf. Preston 2017).

Perceptual Dialectology is concerned with what people think about the character and distribution of linguistic facts in space. The first studies in this field were carried out in the middle of the twentieth century and examined the degree of similarity or difference of the speech of surrounding areas (Rensink 1955; Grootaers 1959; Sibata 1959). Following a period of relative neglect, interest in PD returned in the eighties on account of the work of the American scholar Dennis R. Preston, who proposed a mental mapping approach to explore the intersection of perception and language following a new trend in cultural geography studies. Fresh interest in PD led to a reworking of the field's theoretical foundation and a redefinition of its aims (Cramer 2016). The research programme proposed by Preston at that time was based on the collection and analysis of five types of data which entail different objectives and methods, as follows (Preston 1999, 2010b, 2011b):

a) Draw-a-map: the identification of dialect areas on a map. The aim of this method is to assess where respondents believe regional speech areas exist.

b) Degree-of-difference: evaluation of the degree of similarity and difference between varieties. Respondents are asked to gauge the linguistic distance between a given speech variety and their own, qualifying it as the same, not very different, different, etc.

c) Correct and pleasant: evaluation of the correctness and pleasantness of regional speech varieties. Respondents rate local speech varieties in terms of dimensions of correctness and agreeability.

d) Qualitative data: analysis of statements by speakers about matters related to language variation. Respondents give information about speech varieties through open-ended conversations and questionnaires.

©Universitat de Barcelona

e) Dialect identification: classification and evaluation of aural stimuli pertaining to different regional varieties. Respondents hear speech fragments and attempt to identify the place they are from (placement task) and evaluate features of speakers' voices.

A few decades on, these five lines of research remain valid and have guided studies in a variety of languages. Hand in hand with the spread of PD studies in different languages, there have been advances in theoretical and methodological perspectives which have contributed to the field's recognition within the linguistic discipline.

The methods used to obtain information in PD studies are based on standard procedures in dialect research and sociolinguistics, especially as regards operational and experimental techniques (Albury 2017). Together with the use of questionnaires, which are characteristic of traditional dialect studies, other procedures were introduced such as hand-drawn maps and aural stimuli presented for identification and evaluation by respondents. The incorporation of computer technology into linguistic research since the early nineties has had a notable impact on the way in which the information collected in PD studies is documented and analysed. Owing to the use of quantitative methods and GIS tools for spatial analysis, it has become progressively easier to process more and more data, making the results more robust and reliable (Montgomery & Stoeckle 2013).

In this article we present a new GIS-based application which was developed to facilitate Perceptual Dialectology data documentation, processing and analysis. The application, called PDMapping, is especially useful as a complement to more traditional methods of data collection. Moreover, it is a helpful tool for dealing with large amounts of information, in particular geolinguistic data. PDMapping was designed to assist a research project concerned with dialect identification. Its simple design allows it to be adapted to other kinds of study in which a draw-a-map task figures as one kind of documentation.

The rest of the paper is organised as follows. Section 2 provides a review of the first attempts to systematise perceptual dialectology data collection and analysis. In Section 3, the application is presented, describing its components and showing a particular example of its use. Section 4 shows how information is stored and analysed. To illustrate the process of processing and analysing data, examples from an actual, ongoing research

©Universitat de Barcelona

project are given. Finally, Section 5 presents conclusions and thoughts about the need to use standard, sharable formats in PD work.

2. Drawing perceptual boundaries on a map: methods and techniques

2.1 Background

Two of the methods proposed by Preston for examining linguistic perceptions involve maps: the draw-a-map task and the dialect identification task. Draw-a-map is the technique used in the earliest DP studies, consisting of giving respondents a (usually minimally detailed) printed map on which they are asked to draw in borders of the language areas which they recognise according to instructions from the researchers (Long 1999; Preston 2010b; Cramer 2010; Jeon 2012). The task typically includes labelling the identified areas. Examples showing the kind of maps produced by respondents in different parts of the United States can be found in Preston's papers. In many of them the areas identified are labelled using names in popular use for the area or speakers in question; for example: Eskimoes, Hillbillies, Southerners, British, Average Normal and Sunny Side, in a Michigan respondent's hand-drawn map of USA dialect areas (Preston 2010a: 93, Figure 1). In the dialect identification or voice placement task, respondents are given a map and hear an assortment of dialect samples in random order. The purpose of the aural stimuli is to evoke the respondent's perception of dialect variation. After hearing each voice sample, the interviewee is asked to associate it with a particular point or area on the map (see Cramer & Montgomery 2016 for some examples). There are two different types of placement task: forced-choice tasks and free-choice tasks. In forced-choice tasks listeners must link voice samples to predetermined locations (Clopper & Pisoni 2004; Plichta & Preston 2005; Clopper & Pisoni 2006). In free-choice tasks respondants have to group similar voice samples (Clopper & Bradlow 2009) or allocate voices on a empty map (Montgomery 2011; Lawrence 2015). Our research into Galician dialect variations is carried out using the latter type of dialect identification task.

©Universitat de Barcelona

Consequently, the version of the application that can be downloaded in the GitHub repository and consulted at the project website was designed for this type of invocation in PD. The use of these two methods of data collection results in a set of paper maps in which each informant has drawn boundaries or allocated places manually with a pen or pencil. The maps so produced contain data of interest, but the information collected is difficult to process and analyse. In recent years, computer software for processing graphic information has helped to simplify the task, as Montgomery observes: "the remedy to this problem is the use of computerisation in order to trace, mark, and calculate the amount of agreement of informants from a survey location over the placement and extent of dialect areas" (Montgomery 2010: 590-591).

2.2 State of the art

Automation of the process of documenting and analysing perceptual dialectology data began several decades ago. Towards the end of the twentieth century, Preston and Long designed special computer programmes for the treatment of data gathered in their research into the perception of dialects (Preston 1989; Long 1999; Long & Yim 2002), but the maps thus obtained were of low quality and could not be utilised for complex quantitative processing (Montgomery & Stoeckle 2013).

Subsequently some of the issues related to the low quality and storage format of the information collected were solved through the use of computer tools. At first, overlay techniques were used in vector graphics programmes such as CorelDraw or Adobe Illustrator to create maps (Purschke 2011). Around the same time, the use of GIS tools to process and analyse PD data started to become widespread. In 2013, Montgomery & Stoeckle, focusing on the development of an automated data processing system, proposed the use of Geographic Information Systems (GIS) tools to treat geographical perceptual data. Those tools enabled them to compose advanced heat maps (Montgomery & Stoeckle 2013).

Thanks to the incorporation of GIS tools into PD studies, the processing and analysis of digitalised data gathered by means of a draw-a-map task started to become easier and more effective, resulting in improved analysis techniques for information collected in

©Universitat de Barcelona

perceptual studies (Montgomery 2010). For example, GIS software allows all the responses obtained to be displayed on a single map; this was not possible through a visual inspection of aggregate data. Furthermore, systematic comparison of linguistic information with other kinds of data (e.g. historical, demographic, genetic or socio-economic) was made possible. The resulting aggregate data can be queried and visualized using GIS software in order to discover patterns that would have been hard to identify by traditional techniques.

Due to such improved technology, GIS came to be more widely used in dialectological studies in general, especially in the area of PD (Evans 2011, 2013; Cukor-Avila et al. 2012; Jeon & Cukor-Avila 2016; Montgomery & Stoeckle 2013). However, progress in the computerisation of data analysis did not always go hand in hand with improvement in information gathering methods. Documentation continued to be performed using paper maps on which informants drew and wrote with a pencil; while analysis was assisted by GIS, data gathering was still carried out manually as in earlier PD studies. The main weakness of that is that researchers must spend a long time scanning and digitalising manually produced input before they can analyse all the information. The process yields satisfactory results but it takes a lot of time and effort.

The first attempts to develop a computerised, portable method of documenting and processing aggregate PD data compatible with different operating systems took place towards the end of the twentieth century. Preston & Howe (1987) were the first to develop a technique involving a digitising pad and custom software that allowed geographical PD data to be processed with relative ease. Respondents were asked to draw their mental maps on the digitising pad, after which researchers used special software to produce aggregate maps with all the input from all the informants, yielding information that could be analysed in the usual manner.

Onishi & Long (1997) improved this technique ten years later and adapted it for use with computers running on Windows. Again, geographical data were gathered by means of a digitising pad for subsequent processing by means of a customised software package called *Perceptual Dialectology Quantifier* (PDQ). The display of data resulting from this

©Universitat de Barcelona

method is a noticeable improvement over the results of the earlier technique in both precision and wealth of information (Long 1999: 183).

Despite the progress that these programmes represent, because the programming systems used to develop them limited their portability they were doomed to becoming obsolete fairly quickly. Furthermore, the maps obtained through these techniques were still of low quality, and since they were formatted as images they could not be used in a straightforward manner to perform a quantitative analysis using standard software (Montgomery & Stoeckle 2013).

2.3. A new computerized proposal

Attempts to improve documentation and analysis techniques for perceptual information have parallel advances in digitalisation technology and geographical data processing. However, while PD studies have made the most of new tools for analysis, until now they have not taken full advantage of the potential of this technology for documenting geographical information with ease, precision and efficiency. Therefore we have developed a software platform that combines various techniques of Perceptual Dialectology and automates the process of data collection and information storage.

PDMapping was created as part of a research project to explore perceptions of Galician linguistic variation in Galician territory and Portugal, Asturias, León and Zamora, surrounding regions in the northwest of the Iberian Peninsula where other Romance varieties are spoken. The initial idea was to develop a tool supporting the data collection methods proposed by Preston so as to facilitate efficient documentation and analysis (Calaza 2017). The tool's general characteristics will now be presented.

3. PDMapping: mapping linguistic perceptions on the screen

PDMapping was specifically designed for PD research and can be used to gather and store geographical, socio-demographic and linguistic data digitally. Its main advantage over earlier software applications lies in its support of the documentation of geographical

©Universitat de Barcelona

information associated with speakers' choices and its suitability for working both on-line and off-line. Data gathered using the application are stored in a format which allows statistical and spatial analyses to be performed directly by means of tools commonly used in dialect work such as R project (2018) and QGIS (2019). Storing information in standard formats also facilitates comparison of collected data with linguistic information (traditional dialectal studies, dialectometric analyses, etc.) and geographical, historical, cultural and other kinds of information (Kendall & Fridland 2016). The advantages of using PDMapping for documentation and analysis are particularly noticeable when dealing with a large volume of data.

3.1 Technical features

PDMapping:

-is a web app which can be installed on Unix systems and used on computers and other devices through a website.

-is accessible from any operating system through the Internet via normal web browsers.

-can be used off-line on Unix or through a virtual Unix system (virtual machine) permitting access through a local IP (web browser).

-has a user-friendly, self-explanatory, intuitive user interface, which is easily adapted to different research needs.

is highly effective when working with maps given that the system is designed to link all information to a Spatial Reference System.

-uses a storage system implemented in an object-relational PostgreSQL database system, allowing quantitative and qualitative information to be stored in relational databases. PostgreSQL is an object-relational database management system which allows complicated queries to be carried out on large databases making efficient use of available computing resources.

-makes it easy to use quantitative or qualitative data from different information-processing programmes.

©Universitat de Barcelona

These features make PDMapping a particularly useful tool for PD studies that involve geo-referenced data. An example follows showing the use of PDMapping with data from the project for which it was originally designed.

3.2. Collecting data

PDMapping was designed as part of a research project in perceptual dialectology in the Galician language area (Suárez 2017) in which the dialect identification technique was used, where respondents are asked to identify the dialect of a variety of voice samples in regional and standard language varieties. PDMapping's simple structure and open code mean that it is easy to adapt for use in other studies employing draw-a-map or placement tasks techniques either exclusively or together with dialect identification (see, for example, Williams, Garret & Coupland 1999).

The aim of our study was to obtain and analyse information about Galician speakers' perceptions regarding dialect variation in the Galician language area. Informants give value judgements after hearing auditory stimuli and indicate their geographical origin. As a pioneer study within the area of Iberian Romance, it drew on methods previously used in research about other languages. The perceptual data was obtained by means of 180 interviews in 45 different places within the Galician language area. Data collection was performed in a controlled manner through structured interviews lasting between fifteen and twenty minutes. This information was registered using the PDMapping platform. In some areas, the hilly terrain and poor telephone coverage limit the quality of internet access available, so the application was used in off-line mode, saving data in a database residing on the notebook used during fieldwork. The aural stimuli employed were obtained from a previous study which entailed selection of informants, analysis of recorded segments and edition of sound files. All the voice samples are of spontaneous speech with a maximum length of two minutes.

©Universitat de Barcelona

Dialectologia 24 *(2020), 177-203.*
ISSN: 2013-2247

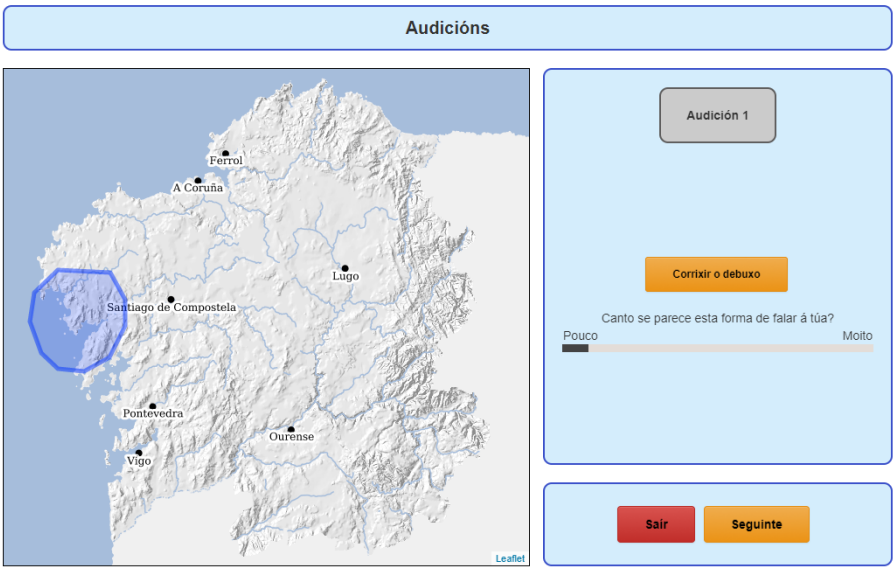


Figure 1. Collecting dialect identifications and perceptions of voice samples.

Three steps were involved in the data collecting task:

1. Background information. The informant provides background information such as age, sex, education, place of birth, residence and language profile. This information is important because such factors may influence the informant's evaluations (Baker, Eddington & Nay 2009; Calaza et al. 2015; Jones et al. 2017).
2. Perceptual test. The informant takes the perceptual test, listening to aural stimuli and assigning them to an area on a map of Galicia. The app displays a map of the area under study with a minimum of labels for identification (the names of the seven main towns). Seven different voice samples are randomly shown next to the map: in the study we are considering, they represented six dialect varieties and standard Galician. Each sample contains phonetic, prosodic, grammatical and lexical clues about where it came from.

After listening to each sample all the way through, the informant was asked to indicate the area where the variety spoken by the speaker comes from by drawing in a closed area on the map using the mouse (the informant is allowed to make corrections if necessary). The application automatically converts the region indicated by the informant into a polygonal shape defined by geographical co-ordinates (Figure 1).

©Universitat de Barcelona

Informants are next asked to estimate the degree of similarity of the voice sample they have just heard to their own speech variety by answering the question *How similar is this way of speaking to your own?* by means of a bar presenting a scale, at the left end of which it says *Not very* (for less similarity) and at the right end *Very* (for use when the two speech varieties are very similar).

3. Traditional questionnaire. A more traditional questionnaire is used to gather informants´ responses about their perceptions of the difference or similarity between various dialects, which informants are asked to describe in terms of correctness, pleasantness and the characteristics they associate with each variety and its speakers (Figure 2). Some examples of questions are *Where do you think the best Galician is spoken? Where do you think the worst Galician is spoken? Where in Galicia do you think the language is most different from the way you speak?*



Figure 2. The variation perception form.

©Universitat de Barcelona

Dialectologia 24 *(2020), 177-203.*
ISSN: 2013-2247

4. Storing and processing information with PDMapping

The information that is collected by means of the web application is stored in a number of structured, interrelated tables on a PostgreSQL database server. The tables can be queried through a database management system such as pgAdmin (Figure 3). The use of a relational database system provides multiple advantages when working with complex data of various kinds, as is the case in perceptual dialectology projects. This kind of database offers clear advantages: a) data can be organised according to category (textual data, quantitative data, geographical information, etc.); b) the storage capability is high; c) the data are easy to export to other analytical tools such as QGIS (through a BBDD administrator plug-in) and project (through an RPostgreSQL package; Conway et al. 2017).

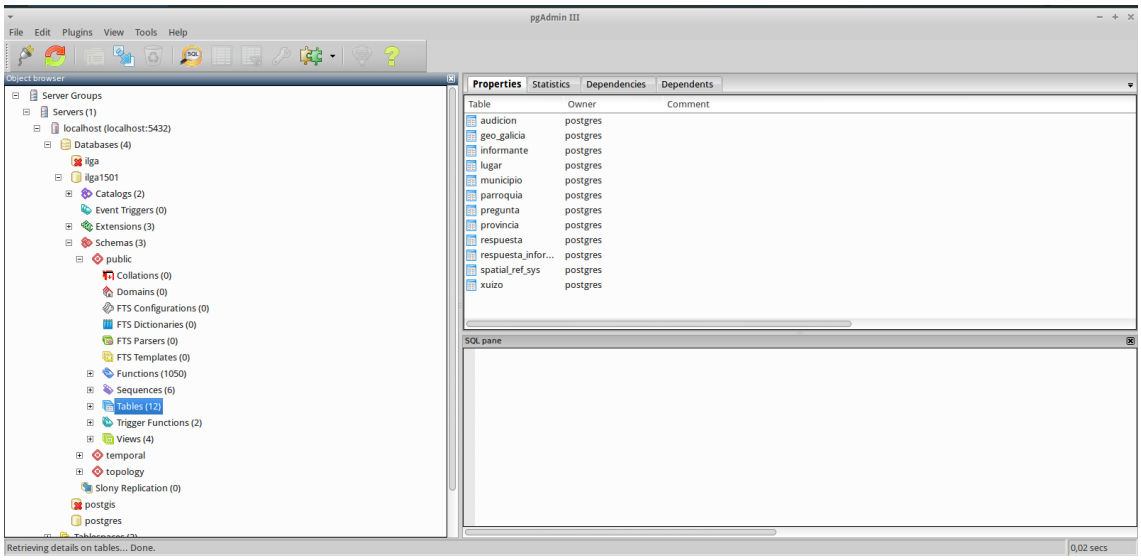


Figure 3. Database tables in pgAdmin.

4.1 Database structure and query syntax

Twelve structured tables are used to store the information held in PDMapping. The diagram in Figure 4 shows relations among the main tables. Each table contains

©Universitat de Barcelona

information that is relevant for an exhaustive study of the data. In the figure, relations between tables are represented by arrows, and linked variables are labelled *c*, *i*, and *g*.

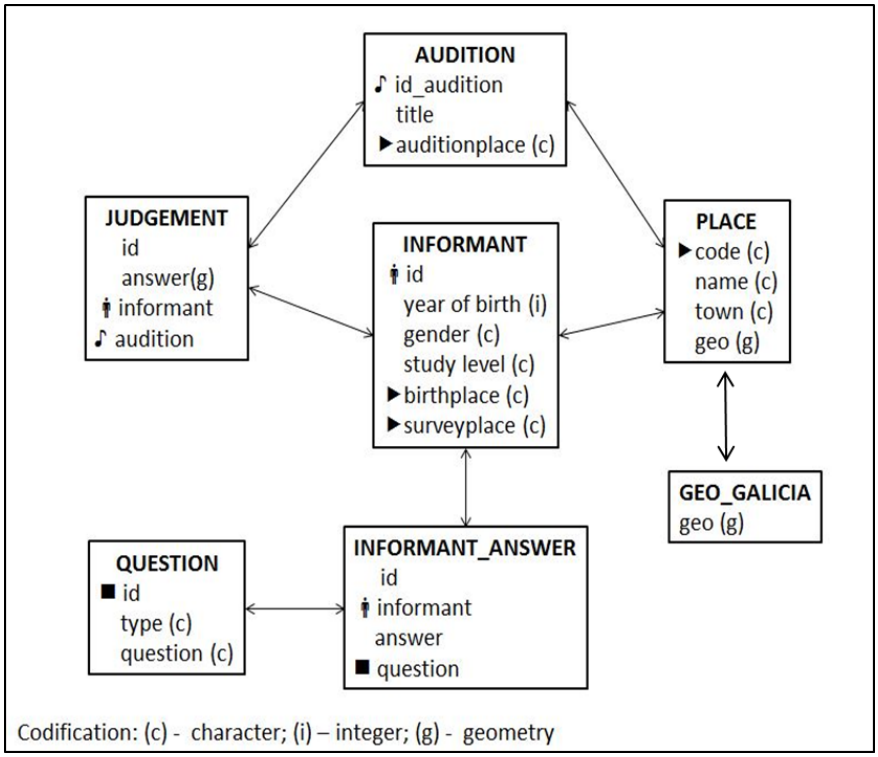


Figure 4. Relations among the main tables of the PDMapping database.

SQL (Structured Query Language)'s table structure allows information to be collected efficiently and accessed as needed. An example of a query follows. Suppose we are interested in obtaining information about the birthplace of the informants. The information is contained in two different tables, *Place* and *Informant*, linked by the place code. Those tables contain the following information of particular relevance for this purpose (Figure 5):

i. The *Informant* table contains information about the people who were surveyed, namely identification code, year of birth, gender, education, place of birth, place where the interview took place, etc.

ii. The *Place* table contains detailed information about the place where the interview occurred: identification code, name of the town, geographical coordinates etc. This information can be linked with the *Informant* table starting from the name and the geometric location of the place.

©Universitat de Barcelona

	id_informan [PK] bigint	ano_nacem integer	sexo character(1	estudos character v	outro_nacei text	outro_enqui text	lugar_nacen character(9	lugar_enqui character(9
1	2	1991	H	UNIVERSITA			360242501	150780001
2	4	1996	M	SECUNDARI(			360040210	150780001
3	5	1963	M	PRIMARIOS			320400801	320400801
4	6	1981	M	SECUNDARI(			320400801	320400801
5	7	1964	H	SECUNDARI(			320400802	320400801
6	8	1990	M	UNIVERSITA			320320101	320320101
7	9	1961	H	SECUNDARI(			320320101	320320101
8	10	1951	M	UNIVERSITA			320320101	320320101
9	11	1963	M	PRIMARIOS			320120401	320120401

	codigo character(9)	nome character varying(255)	municipio character(5)	geo geometry(Point,4258)
1	150010101	O Adro	15001	0101000020A2100000FF02839CDE9820C097941689E49B4540
2	150010102	Beade	15001	0101000020A2100000DCBF0D8C179C20C00CFB873EE99C4540
3	150010103	Borreiros de Abaix(	15001	0101000020A2100000DF1A54F21C8D20C0F31D5FEF3C9D4540
4	150010104	Calvelos	15001	0101000020A210000061E8C715CD9920C08B7D3462879B4540
5	150010105	Cancelada	15001	0101000020A2100000E52AB3946E9120C08D6C3A44D49C4540
6	150010106	A Cruz-Barreiros	15001	0101000020A2100000F635DC91FB9920C0A1766CC2459C4540
7	150010107	A Fonte	15001	0101000020A2100000D14F1F1E159720C09123DF9D639C4540
8	150010108	O Loureiro	15001	0101000020A21000008963D9A3A29320C0794EA35EAA9C4540
9	150010109	Merdil	15001	0101000020A21000004A703491E39420C0BADC3F9A5A9C4540

Figure 5. Data in Informant and Place tables.

Using SQL language, a query can be performed to obtain information about informants' birthplace, for instance:

```
SELECT place.geo,informant.id_informant, place.name
FROM informant, place
WHERE informant.birthplace=place.code
```

The table thus obtained (Figure 6) brings together information about informants' ID and the place where they were born (coordinates and name).

©Universitat de Barcelona

	geo geometry(Point,4258)	id_informante bigint	nome character varying(255)
1	0101000020A210000(	198	O Rañal
2	0101000020A210000(	196	O Rañal
3	0101000020A210000(	195	O Rañal
4	0101000020A210000(	194	O Rañal
5	0101000020A210000(	218	O Acevedo
6	0101000020A210000(	217	Queixas
7	0101000020A210000(	216	Queixas
8	0101000020A210000(	215	Rodís
9	0101000020A210000(	23	Chaín
10	0101000020A210000(	21	Chaín

Figure 6. Table with informant data.

4.2 Analysis of the data

The information that the tables of the database contain may be of different kinds (locations, numerical data, text and so on); thus, they can be analysed from various perspectives using a variety of tools. The project for which PDMapping was originally developed involved analysing collected data for the most part using two freeware applications: a) a free statistical package, R Project, which serves to analyse all the quantitative information and obtain statistical breakdowns; b) a GIS programme, QGIS, with which to display the geolinguistic analysis and show the results in maps and graphics. Both these programmes are currently widely used in a number of scientific fields including PD (Cramer & Montgomery 2016). On account of the standard format in which the information is stored in PDMapping, it is also possible to use a number of other computer programmes, some of which are free and others proprietary, in order to analyse this information and display results.

To illustrate the statistical use of the data collected with PDMapping, we will now describe the procedure whereby R Project and QGIS were connected to the database to perform statistical analyses and maps of the results.

©Universitat de Barcelona

a) Statistical analysis

The data tables can be imported into the R Project statistical programme using the RpostgreSQL package, which provides the special functions dbDriver and dbConnect for that purpose. Connection to the database can be achieved as follows:

```
drv <- dbDriver("PostgreSQL")
con <- dbConnect(drv,
host="XXX.XXX.XX.XXX",port="5432",dbname="ilga20160412",user="postgres",
password="postgres")
```

After running this function, all the information in the database can be accessed using the package's dbSendQuery function. In this way, any variable type can be read conforming to a *data.frame* object, and objects with geometrical information can also be imported to allow work with geospatial data. An example follows:

```
q_birthplace="SELECT ST_AsText(st_transform(place.geo,25829)) AS wkt_geometry,
id_informant FROM informant, place WHERE informant.birthplace=place.code"
rs_birthplace = dbSendQuery(con,q_birthplace)
```

Direct access to such georeferenced information is useful not only for displaying geographical data but also for performing specific analyses combining and comparing data from a variety of sources. It was shown in Calaza et al. (2015) that GIS tools can be used with data from distinct sources to carry out PD studies such as direct comparison with data from traditional dialectology and perceptual dialectology studies, and identification of the factors that influence dialectal variety recognition using regression models.

©Universitat de Barcelona

b) Cartographic display

A Geographic Information System (GIS) was used to display the results of our PD study on a map. We use QGIS, an easy-to-use open source GIS application that is suitable for data viewing, editing, and producing maps. One of the advantages of using QGIS is that it can easily be connected to a PostGIS database to spatially view any geometric object spatially, thus providing a simple way to access and display information from PDMapping. The statistical package R Project also allows one to create geographical representations of perceptual data that are similar to those obtained with QGIS but with more limited ability to combine data from different sources (Lansley & Cheshire 2016). Like other GIS programmes, QGIS uses a file format that is standard in map applications.

Next, we will see the procedure for creating a connection between a PostgreSQL database and QGIS. First, QGIS must be opened, and "Add PostGIS Layers" selected from the *Layer* menu. This opens a dialogue window where we can connect to the database by clicking on "New" to open another dialogue box in which the connection we wish to set up is configured, and its type identified, as in Figure 7.

Once connected, we can add whatever layers we like. QGIS has a database administration component for accessing specific information by running SQL statements. QGIS's display options make it easy to create heat maps based on respondents' answers after loading each of the layers of the geometries of their responses. The map displays can identify areas of greatest coincidence in informants' evaluations (Figure 8).

The above process requires a computer with sufficient processing power. An alternative way to obtain map displays is using the database's own management system. When running the function _SELECT exporta_raster_tiff_, the PostGIS geometries defining the responses are converted into PostGIS objects in raster format. This output can be saved as a .tiff file and loaded from QGIS later as a layer, either dragging the file directly or adding it as a raster layer.

©Universitat de Barcelona

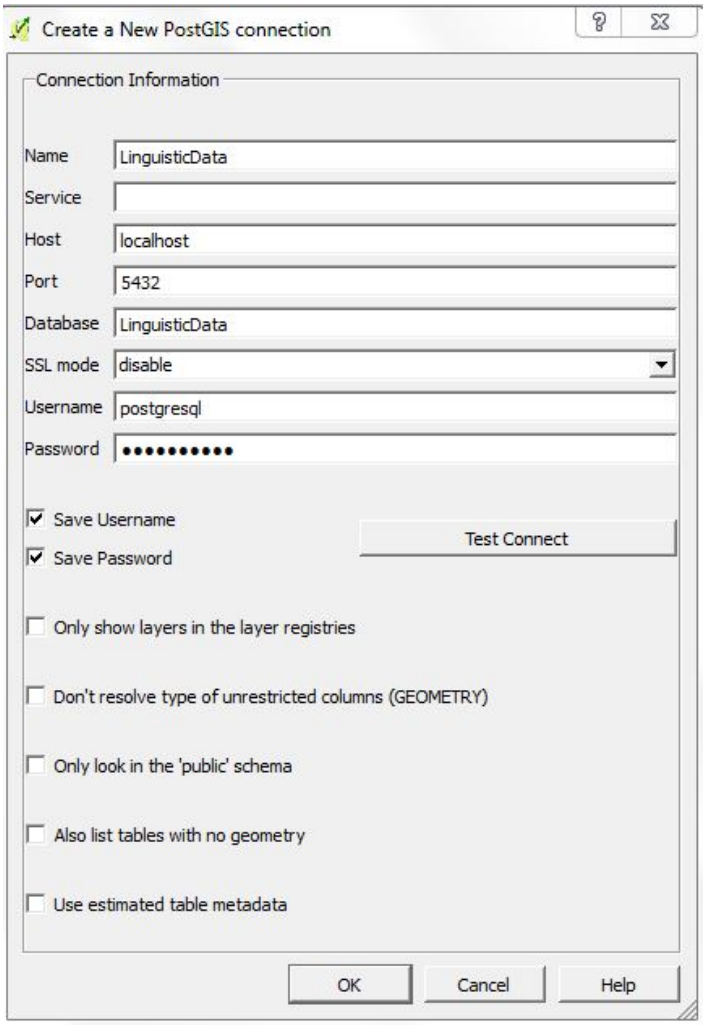


Figure 7. Creating a connection from PostGIS to QGIS.

SELECT exporta_raster_tiff(st_union(st_asraster(st_transform(st_intersection(g.geo, x.geo),25829), 200.0, 200.0, st_xmin(g.geo), st_ymin(g.geo), '32BUI', 1, 0), 'SUM'), '/home/'||cast(x.audicion as varchar)||'.tiff')

FROM geo_galicia g, xuizo x

group by x.audicion

©Universitat de Barcelona

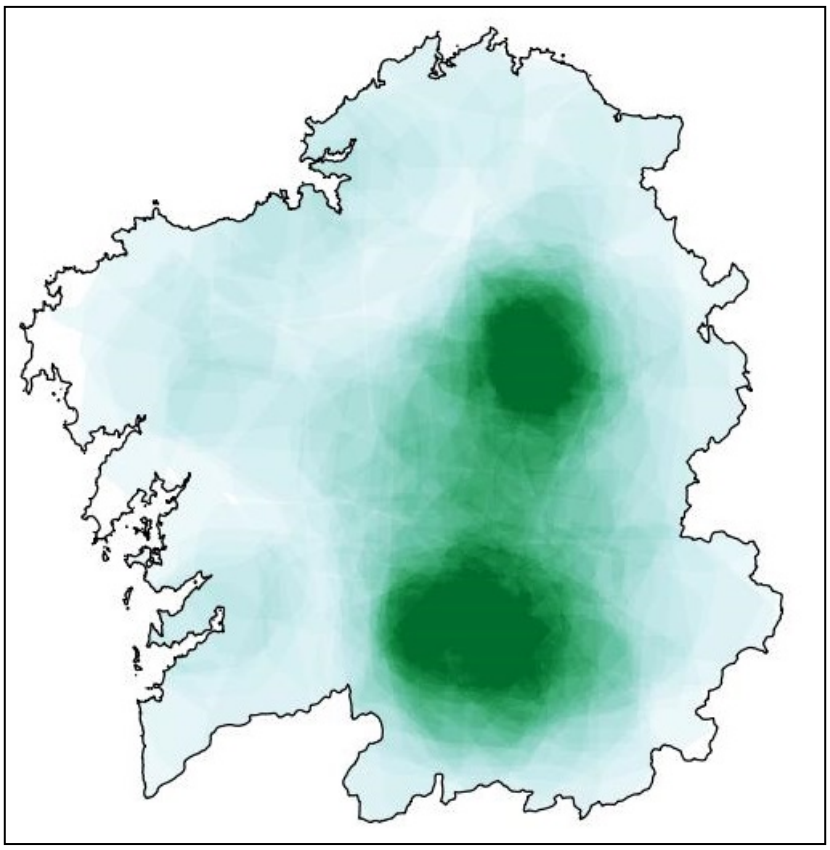


Figure 8. Heat map produced with QGIS.

4.3 Comparing data: dialectal varieties and respondants' perceptions

The data obtained through PDMapping may be used to compare data from different sources and types. The evaluation between perceptual and geolinguistics data could be of a special relevance to sociolinguistic and dialectal studies. For instance, if comparable data are accessible in a compatible format it is possible to identify links between more salient areas or features for speakers, according to perceptual analysis, and dialectal varieties described by researchers.

Taking as starting point the results of our research and using Baddeley's distance technique to measure the discrepancies between theoretical and perceptual regions (Baddeley et al. 1992; Rumpf et al. 2010), we can compare in a simple and efficient manner results from perceptual studies and information from dialectal analysis (Calaza Díaz et al. 2015). In the current case, the results found allowed us to explore the distribution of the distances between the polygons outlined by the informants and those defined by different dialectology methods, both qualitative (Fernández Rei 1990) and

©Universitat de Barcelona

quantitative (Sousa 2006). Figure 9 shows the density of the distance from traditional varieties and respondent's perceptions, for the three Galician dialectal areas which are best recognised by speakers. The red line represents the distribution of the western dialectal area and corresponds to the smallest distance values, namely, this dialectal area is the best identified by informants of perceptual experiment. In contrast, the green line represents the central varieties, in which the distances are larger and which the informants find more difficult to identify and recognize. In the descriptions of Galician dialectal varieties, it is often noted that western varieties are characterised by having phonetic and morphological characteristics which are more easily perceivable (because they are outstanding and recursive) by non-expert speakers (Fernández Rei 1990).

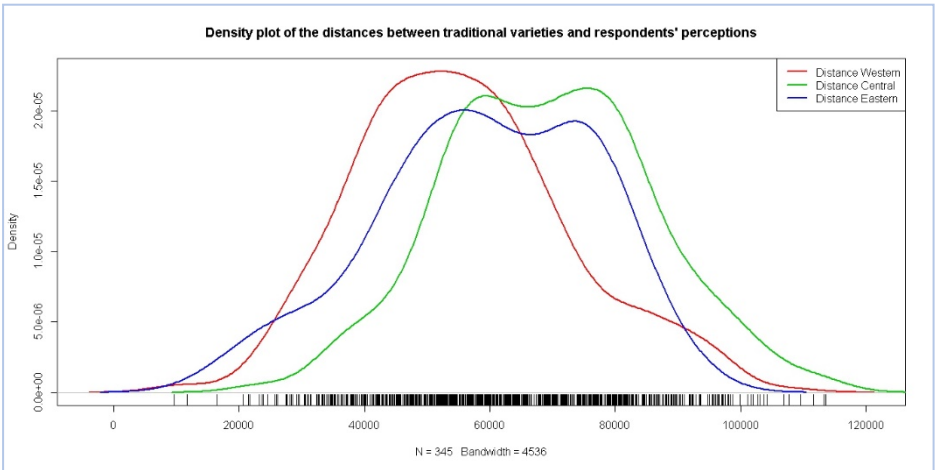


Figure 9. Main Galician varieties and respondents' perceptions.

5. Conclusions

The study of speakers' perceptions and attitudes about language varieties makes it possible to analyse linguistic ideologies and find out what influence they exert on language change. The methods designed and implemented by Perceptual Dialectology (Cramer & Montgomery 2016) constitute a useful tool for obtaining such knowledge and non-linguists' view of linguistic diversity in general. Research procedures based on a draw-a-map task can tell us about language diversity, ethnicity and cultural identity.

©Universitat de Barcelona

The tool presented in this paper, PDMapping, assists in the gathering and processing of information using quantitative analysis software and geographic information systems. It was created within a particular project using a draw-a-map task and controlled interviews as documentation sources. Its design features and the free circulation of its code in the scientific community mean that it is easy to adapt to other projects with similar objectives. Over the past year the tool has been tested, modified and improved. Future developments should focus on incorporating other functionalities, making it possible to add to the types of perceptual data that it can incorporate.

Funding

This work was supported by the Spanish Ministerio de Economía y Competitividad [FFI2015-65208-P], European Regional Development Fund (multiannual financial framework 2014-2020]) and the Xunta de Galicia (TecAnDali research network, [ED341D R2016/011]).

Acknowledgements

We would like to thank Dr. Miguel Rodríguez Luaces, Alejandro Cortiñas (Laboratorio de Bases de Datos, Universidade da Coruña) and Noemí Varela (Instituto da Lingua Galega, Universidade de Santiago de Compostela) for client and server-side development insight and support. We would also like to thank Dr. Jose Ramón Ríos Viqueira (CITIUS, Universidade de Santiago de Compostela) and Dr. Alberto Rodríguez Casal (Departamento de Estatística, Análise Matemática e Optimización) for their suggestions.

References

ALBURY, N. J. (2017) "How folk linguistic methods can support critical sociolinguistics", *Lingua,* 199, 36-49. doi:0.1016/j.lingua.2017.07.008.

BADDELEY, A., Y. M., CHANG, SONG & R. TURNER (2012) "Nonparametric estimation of the dependence of a spatial point process on spatial covariates", *Statistics and its Interface*, 5(2), 221-236. < https://dx.doi.org/10.4310/SII.2012.v5.n2.a7>

©Universitat de Barcelona

BAKER, W., D. EDDINGTON & L. NAY, L. (2009) "Dialect identification: The effects of region of origin and amount of experience", *American Speech*, 84(1), 48-71. doi.org/10.1215/00031283-2009-004.

CALAZA DÍAZ, L. (2017) "Lingüística matemática: a información que se esconde tras as percepcións", in M. I. Borrajo, G. Castiñeira, L. López & V. Sanmartín (eds.), *Actas do Seminario de Iniciación á Investigación*. Santiago de Compostela: Instituto de Matemáticas-USC, 75-80. <http://www.usc.es/export9/sites/webinstitucional/gl/institutos/matematicas/descargas/ActasSII_2016_2017.pdf> (accessed 3 January 2019).

CALAZA DÍAZ, L., S. SUÁREZ QUINTAS, R. M. CRUJEIRAS, A. RODRÍGUEZ CASAL, X. SOUSA & J. R. RÍOS VIQUEIRA (2015) "A method for processing perceptual dialectology data", in M. J. Ginzo Villamayor, J. M. Alonso Meijide & L. Ramil Novo (eds.), *ACTAS XII Congreso Galego de Estatística e Investigación de Operacións*. Lugo: Servizo de Publicacións da Deputación Provincial de Lugo – SGAPEIO, 282-291. Retrieved from <http://eio.usc.es/pub/mjginzo/descargas/XII_SGAPEIO.pdf> (accessed 5 December 2018).

CHAMBERS, J. K. & P. TRUDGILL (1980) *Dialectology*, Cambridge: University Press.

CLOPPER, C. G. & A. R. BRADLOW (2009) "Free Classification of American English Dialects by Native and Non-Native Listeners", *Journal of Phonetics*, 37(4), 436-51. doi.org/10.1016/j.wocn.2009.07.004.

CLOPPER, C. G. & D. B. PISONI (2004) "Some Acoustic Cues for the Perceptual Categorization of American English Regional Dialects", *Journal of Phonetics*, 32(1), 111-140. doi.org/10.1016/S0095-4470(03)00009-3.

CLOPPER, C. G. & D. B. PISONI (2006) "Effects of Region of Origin and Geographic Mobility on Perceptual Dialect Categorization", *Language Variation and Change*, 18(2), 193-221.

CONWAY, J., D. EDDELBUETTEL, T. NISHIYAMA, S. KUMAR PRAYAGA & N. TIFFIN (2017) *RPostgreSQL: R Interface to the 'PostgreSQL' Database System. R package version 0.6-2*. https://CRAN.R-project.org/package=RPostgreSQL.

CRAMER, J. & C. MONTGOMERY (eds.) (2016) *Cityscapes and Perceptual Dialectology: Global Perspectives on Non-Linguists Knowledge of the Dialect Landscape*, Berlin / Boston: De Gruyter Mouton.

CRAMER, J. (2010) *The Effect of Borders on the Linguistic Production and Perception of Regional Identity in Louisville, Kentucky*, Urbana-Champaign, IL: University of Illinois dissertation.

©Universitat de Barcelona

<https://www.ideals.illinois.edu/bitstream/handle/2142/18426/Cramer_Jennifer.pdf?sequence=1> (accessed 3 September 2018).

CRAMER, J. (2016) "Perceptual Dialectology", in *Oxford Handbooks Online*, Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199935345.013.60.

CUKOR-AVILA, P., B. E. EVANS, L. JEON, D. LONG, C. MONTGOMERY, D. R. PRESTON & P. STOECKLE (2012) "A Geographical Information Systems (GIS) approach to Perceptual Dialectology data", *Paper presented at New Ways of Analyzing Variation 41*, Indiana University.

EVANS, B. E. (2011) "Seattletonian to Faux Hick: Mapping perceptions of English in WA", *American Speech* 86(4), 383-413.

EVANS, B. E. (2013) "Seattle to Spokane: Mapping Perceptions of English in Washington State", *Journal of English Linguistics,* 41(3), 268-291. doi: 10.1177/0075424213494822.

FERNÁNDEZ REI, F. (1990) *Dialectoloxía da lingua galega*, Vigo: Xerais.

GROOTAERS, W. (1959) "Origin and Nature of the Subjective Boundaries of Dialects", *Orbis,* 8(2), 355-384.

JEON, L. & P. CUKOR-AVILA (2016) "Urbanicity and language variation and change: Mapping dialect perceptions in and of Seoul", in J. Cramer & C. Montgomery (eds.), *Cityscapes and Perceptual Dialectology: Global Perspectives on Non-Linguists Knowledge of the Dialect Landscape*, Berlin: Mouton De Gruyter, 97-116. doi: 10.1515/9781614510086-011.

JEON, L. (2012) *Drawing boundaries and revealing language attitudes: Mapping perceptions of dialects in Korea*, Master's Thesis. Denton: University of North Texas.

JONES Z., Q. YAN, L. WAGNER & C. CLOPPER (2017) "The development of dialect classification across the lifespan", *Journal of Phonetics*, 60, 20-37. doi:10.1016/j.wocn.2016.11.001.

KENDALL, T. & V. FRIDLAND (2015) "Mapping the perception of linguistic form: Dialectometry with perception data", in J. Nerbonne, M.-H. Côté & R. Knooihuizen (eds.), *The Future of Dialects*, Berlin: Language Science Press, 173-194. doi:10.17169/langsci.b81.149.

LANSLEY, G. & J. CHESHIRE (2016) *An Introduction to Spatial Data Analysis and Visualisation in R*. CDRC Learning Resources. <https://data.cdrc.ac.uk/tutorial/an-introduction-to-spatial-data-analysis-and-visualisation-in-r> (accessed 3 January 2018).

LAWRENCE, Daniel (2015) "Phonetic Variation as a Cue to Regional Identity: An Experimental Approach", *York Papers in Linguistics*, Special Issue 2, 42-63.

LONG, D. & Y-C. YIM (2002) "Regional differences in the perception of Korean dialects", in D. Long & D. R. Preston (eds.), *Handbook of Perceptual Dialectology*, vol. 2, Amsterdam: John Benjamins, 249-275.

©Universitat de Barcelona

LONG, D. (1999) "Geographical Perception of Japanese Dialect Regions", in D. R. Preston (ed.), *Handbook of Perceptual Dialectology*, vol. 1, Amsterdam: John Benjamins, 117-198.

MONTGOMERY, C. & P. STOECKLE (2013) "Geographical Information Systems and Perceptual Dialectology", *Journal of Linguistic Geography,* 1(1), 52-85. doi:10.1017/jlg.2013.4.

MONTGOMERY, C. (2010) "Sprachraum and its perception", in A. Lameli, R. Kehrein & S. Rabanus (eds.), *An international handbook of linguistic variation. Volume 2: Language mapping*, Berlin: Mouton de Gruyter, 586-606.

MONTGOMERY, Chris (2011) "Starburst Charts: Methods for Investigating the Geographical Perception of and Attitudes towards Speech Samples", *Studies in Variation, Contacts and Change in English*, 7. <http://www.helsinki.fi/varieng/journal/volumes/07/montgomery/index.html.>

ONISHI, I. & D. LONG (1997) *Perceptual Dialectology Quantifier (PDQ) for Windows*. <http://nihongo.hum.tmu.ac.jp/~long/maps/perceptmaps.htm> (accessed 3 January 2018).

PLICHTA, B. & D. R. PRESTON (2005) "The /Ay/s Have It: The Perception of /Ay/ as a North-South Stereotype in United States English", *Acta Linguistica Hafniensia*, 37(1), 107-30. https://doi.org/10.1080/03740463.2005.10416086.

PRESTON, D. R. & G. M. HOWE (1987) "Computerized Studies of Mental Dialect Maps", in K. M. Denning, S. Inkelas, J. R. Rickford & F. McNair-Knox (eds.), *Variation in Language: NWAV-XV at Stanford* (Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation), Stanford CA: Department of Linguistics, Stanford University, 361-378.

PRESTON, D. R. (1981) "Perceptual dialectology: Mental maps of United States dialects from a Hawaiian perspective (summary)", in H. Warkentyne (ed.), *Methods IV/Méthodes IV (Papers from the Fourth International Conference on Methods in Dialectology)*, [Victoria]: University of Victoria British Columbia, 192-198.

PRESTON, D. R. (1989) *Perceptual dialectology: nonlinguists' views of areal linguistic*, Dordrecht: Foris.

PRESTON, D. R. (ed.) (1999) *Handbook of perceptual dialectology*, vol. 1, Amsterdam: John Benjamins.

PRESTON, D. R. (2010a) "Language, space, and the folk", in P. Auer & J. Schmidt (eds.), *Theories and methods (Language and space: An international handbook of language variation,* volume 1, Berlin: Mouton de Gruyter, 179-201.

©Universitat de Barcelona

PRESTON, D. R. (2010b) "Perceptual Dialectology in the 21st Century", in C. A. Anders, M. Hundt & A. Lasch (eds.), *Perceptual Dialectology. Neue Wege der Dialektologie (Linguistik – Impulse & Tendenzen)*, Berlin: Mouton de Gruyter, 1-30.

PRESTON, D. R. (2011a) "The power of language regard – discrimination, classification, comprehension, and production", *Dialectologia*, Special Issue, II, 9-33. <http://www.publicacions.ub.edu/revistes/dialectologiaSP2011/>

PRESTON, D. R. (2011b) "Methods in (applied) folk linguistics: Getting into the minds of the folk", *AILA Review* 24, 15-39. doi:10.1075/aila.24.02pre.

PRESTON, D. R. (2017) "The cognitive foundations of language regard", *Poznan Studies in Contemporary Linguistics*, 53(1), 17-42. doi: 10.1515/psicl-2017-0002.

PURSCHKE, C. (2011) "Regional Linguistic Knowledge and Perception: On the conceptualization of Hessian", *Dialectologia*, Special Issue 2, 91-118. <http://www.publicacions.ub.edu/revistes/dialectologiaSP2011/>

QGIS DEVELOPMENT TEAM (2019). *QGIS Geographic Information System*. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

R CORE TEAM. (2018) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/> (accessed 5 December 2017).

RENSINK, W. G. ([1955] 1999) "Informant Classification of Dialects", reprinted in D. R. Preston (ed.), *Handbook of Perceptual Dialectology*, vol. 1, Amsterdam: John Benjamins, 3-7.

RUMPF, J., S. PICKL, S. ELSPAß, W. KÖNIG & V. SCHMIDT (2010) "Quantification and Statistical Analysis of Structural Similarities in Dialectological Area-Class Maps", *Dialectologia et Geolinguistica*, 18, 73-98.

SIBATA, T. ([1959] 1999) "Consciousness of Dialect Boundaries", reprinted in D. R. Preston (ed.), *Handbook of Perceptual Dialectology*, vol. 1, Amsterdam: John Benjamins, 39-62.

SOUSA, X. (2006) "Análise dialectométrica das variedades xeolingüística galegas", in C. Rolão Bernardo & H. Mateus Montenegro (eds.), *Encontro de estudos dialectológicos. Actas*, Ponta Delgada: Instituto Cultural de Ponta Delgada, 345-362.

SUÁREZ QUINTAS, S. (2017) "O galego non é o ghallegho que falamos nós: a percepción e as actitudes como condicionantes do cambio lingüístico", in X. L. Regueira Fernández & E. Fernández Rei (eds.), *Estudos sobre o cambio lingüístico no galego actual.* Santiago de

©Universitat de Barcelona

Compostela: Consello da Cultura Galega - Instituto da Lingua Galega, 187-213. <http://consellodacultura.gal/mediateca/extras/CCG_2017_Estudos-sobre-o-cambio-linguistico-no-galego-actual.pdf> (accessed 5 December 2017).

WILLIAMS A., P. GARRET & N. COUPLAND (1999) "Dialect Recognition", in D. R. Preston (ed.), *Handbook of Perceptual Dialectology*, Vol. 1, Amsterdam: John Benjamins, 345-358.

©Universitat de Barcelona