Journal of Linguistic Geography (2013), **1**, 52–85. © Cambridge University Press 2013
doi:10.1017/jlg.2013.4

ORIGINAL RESEARCH

Geographic information systems and perceptual dialectology: a method for processing draw-a-map data

Chris Montgomery,[1]* and Philipp Stoeckle[2]

[1] *School of English, University of Sheffield, Sheffield, UK*
[2] *Deutsches Seminar, Universität Zürich, Zurich, Switzerland*

This article presents a new method for processing data gathered using the "draw-a-map" task in perceptual dialectology (PD) studies. Such tasks produce large numbers of maps containing many lines indicating nonlinguists' perceptions of the location and extent of dialect areas. Although individual maps are interesting, and numerical data relating to the relative prominence of dialect areas can be extracted, an important value of the draw-a-map task is in aggregating data. This was always an aim of the contemporary PD method, although the nature of the data has meant that this has not always been possible. Here, we argue for the use of geographic information systems (GIS) in order to aggregate, process, and display PD data. Using case studies from the United Kingdom and Germany, we present examples of data processed using GIS and illustrate the future possibilities for the use of GIS in PD research.

1. Introduction

Aggregating data in perceptual dialectology (PD) is something which has occupied researchers since the earliest research was undertaken in the field (Weijnen, 1946; Mase, 1999). Modern approaches to PD use methods designed to assess respondents' mental maps of language variation and "dig deeply into the conceptual world, not only for the concepts of dialect areas but for the associated beliefs about speakers and their varieties" (Preston, 2010:11). Such methods, involving the use of hand-drawn maps (termed "draw-a-map" tasks, cf. Preston, 1982) have at their heart the aim of arriving at aggregate composite maps of dialect areas from respondents' maps (Preston & Howe, 1987:363). Such aggregate maps can be used to give an account of where respondents perceive dialect areas to exist, along with the extent of these areas. In this way, the methods of PD extend our knowledge of speech communities (Kretzschmar, 1999:xviii) by exploring the social space (Britain, 2010:70) of these communities.

PD research can also play a role in looking afresh at the results of production studies. Indeed, the ability of the discipline to challenge assumptions made from such studies has been noted as one of its strengths (Butters, 1991:296). In order to do this effectively, data must be aggregated in order to produce composite maps of perceptual dialect areas. Perceptual geographers, who provided the impetus for contemporary approaches to PD, knew this (see Gould & White, 1986). The power of an aggregate is that it gives a generalized "picture" of perception which has more explicative power than single images of mental maps produced by individual respondents (cf. Lynch, 1960; Orleans, 1967; Goodey, 1971).

Data from PD studies can be processed simply by counting the number of areas drawn on a number of maps in order to arrive at the recognition level for each area. However, to stop at this stage as some have done (e.g. Bucholtz et al., 2007) is to neglect much of the data supplied by respondents. This geographical data relating to the placement and extent of dialect areas is a valuable resource that, once properly processed, can be used for direct comparison with data from other studies (linguistic and beyond).

Despite this, it is clear why some linguists have not attempted to produce aggregate maps. This is due to the lack of a stable and useable method for completing this type of analysis for maps from large numbers of respondents. This is due to the lack of a stable and useable method for completing this type of analysis for maps from large numbers of respondents, in spite of being one of the aims for PD (Preston & Howe, 1987:363). In Bucholtz et al.'s (2007) study, for example, maps were drawn by 703 respondents. Manually processing and aggregating data from such a large number of respondents is simply not possible given that the most widely available technique is line tracing using overhead transparencies (see Montgomery, 2007:61–68).[1]

In order to work with maps from large numbers of respondents there is a need for an up-to-date, portable,

*Address for correspondence: Chris Montgomery, The School of English Literature, Language and Linguistics, 2.27 Jessop West, 1 Upper Hanover Street, Sheffield S3 7RA, UK.
(Email: c.montgomery@shef.ac.uk)

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 1. Preston & Howe's map aggregation technique – map shows southern Indiana-based respondents perception of a "South" dialect area (1987:373).

accessible, computerized method of processing and aggregating PD data. Attempts at creating such a method have been made in the past. The first was made by Preston & Howe (1987), who developed a technique involving the use of a digitizing pad and custom software. This allowed the storage of digitized line information relating to a dialect area, along with the demographic data of the respondent who drew it. Many lines could be traced using the digitizing pad with the result that aggregate maps of the dialect area could be displayed. These areas could also be queried on the basis of the demographic information. A map created using Preston and Howe's technique can be seen in Map 1.

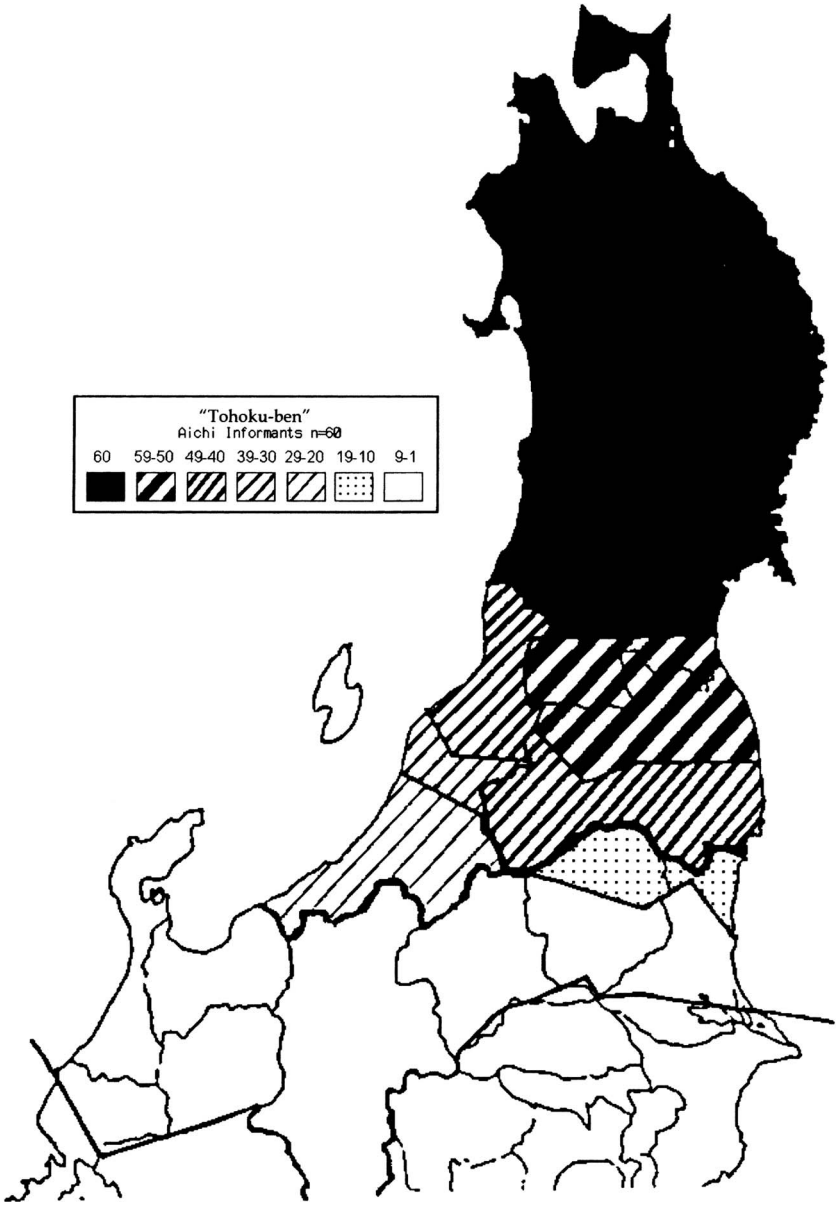
Preston & Howe's (1987) method ensured that there was a method for producing aggregate maps, which also meant that they would be able to be queried. This is a major advantage over a noncomputerized technique, as it did not require separate aggregation techniques for each social variable one wished to examine. This approach was built upon by Onishi & Long (1997) as they updated Preston & Howe's (1987) technique for use with Windows computers. The resulting software, entitled Perceptual Dialectology Quantifier (PDQ) for Windows, processed data in the same way as Preston & Howe's (1987) technique. A digitizing pad was again used to input area line data, and the software package did the rest of the data processing. Map 2 shows an aggregate map produced using PDQ.

Although the methods developed by Preston & Howe (1987) and Onishi & Long (1997) made working with draw-a-map data easier, there were problems with their approaches. The most pressing problem was the lack of "future proofing" built into the technology. The technology used by Preston & Howe (1987) quickly became obsolete, as did the technology used by Onishi & Long (1997). Thus, although PDQ for Windows is still functional to some extent, there are major problems with it. It is not portable and is only available for use in Japan (running on three increasingly elderly computers). A second issue is the low resolution of the maps produced by the program (as can be seen in Map 2), which renders them less suitable for publication. A third problem is the way in which the program permits the display of only one area on a map, which makes it unsuitable for producing composite maps showing multiple perceptual areas on one map (e.g. Preston, 1999a:362).

More recent studies (e.g. Purschke, 2011) have used simple overlay techniques in vector (cf. section 3.1) graphics programs (such as CorelDraw, Adobe Illustrator, etc.). Such programs can yield quite impressive results and an example can be seen in Map 3, which shows a summary of subjective dialect areas in Germany drawn by informants from Northern (left map) and Eastern (right map) Hessian informants.

The different colors in Map 3 indicate aggregate perceptions of different dialect areas, and the color densities show different degrees of agreement. This method clearly improves on the quality of visualization, and the researcher is able to get an impression of which dialect areas are the most prominent and where they are located. However, the use of this type of technology does not allow any further analyses such as the exact calculation of agreement levels, area sizes, or distances (e.g. to the next political border). Also, due to an inability to "anchor" the

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 2. "Tohoku-ben" area, data processed in PDQ (Long, 1999:183).

visualization in the real world (cf. section 3.2), it is difficult to merge PD data with other kinds of datasets (such as streets, topography, etc.).

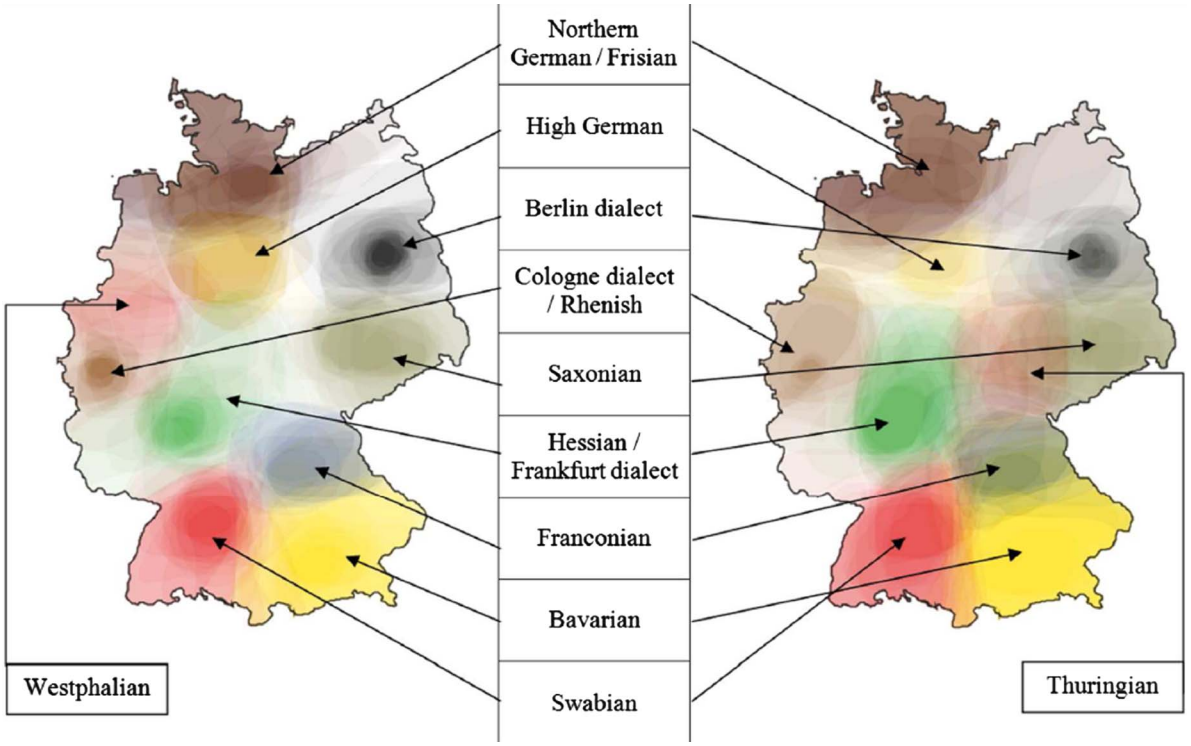
Given the difficulty of processing and aggregating geographical data from draw-a-map tasks without the use of a computer, and the general insufficiency of useable computerized techniques, there is a pressing need for new technology which can be used in this area. In this article, we discuss the role Geographic Information Systems (GIS) may play in filling the gap.

After a short review of the use of the draw-a-map task in PD (section 1.1) we will introduce the surveys and methods of data collection our analyses are based on (section 2). Following that, the principles of GIS will be presented and how they can be applied to PD data discussed (section 3). We will then demonstrate some examples of the possibilities of GIS to visualize and analyze geospatial data (section 4) before summarising our findings and arguing for a more extensive use of this technology (section 5).

1.1. The draw-a-map task in PD

One of the aims of PD research, as mentioned above, is to assess where respondents believe dialect areas to exist (Preston, 1988:475–6). The technique used to investigate this is the draw-a-map task (Preston, 1982). Respondents undertaking the task are asked to "draw

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 3. Prominent large-scale regional language areas for Northern Hessian (left) and Eastern Hessian (right) informants (Purschke, 2011:99).

boundaries on a … map around areas where they believe regional speech zones exist" (Preston, 1999b:xxxiv). An example of a completed draw-a-map task, from one of the studies considered here, can be seen in Map 4.

Data gathered via the draw-a-map task has a twofold usefulness (Garrett, 2010:183): "Firstly, it provides some insight into what and where dialect regions actually exist in people's minds.…Secondly, the task generates attitudinal comment alongside more descriptive data." We are interested in this article in the first use of the data (the spatial aspect). We focus on how we might best process these data in order that we can better understand what respondents think of regional variation, as well as "how concentrated or extensive" (Garrett, 2010:183) respondents think dialect regions are.

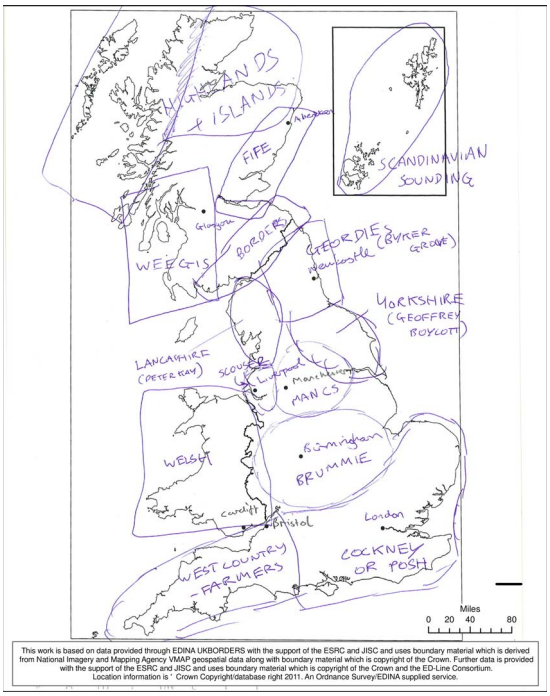
The draw-a-map task has been used in very large countries such as the United States (e.g. Preston, 1986) and Canada (McKinnie & Dailey-O'Cain, 2002), as well as in individual states (Bucholtz et al., 2007; Bucholtz et al., 2008; Anders, 2010; Evans, 2011; Cukor-Avila et al., 2012) and smaller countries (Long, 1999; Montgomery, 2007; Jeon & Cukor-Avila, 2012). While this PD research is interested mainly in the question of how nonlinguists classify large-scale dialect areas, other studies focus on the subjective construction of local dialect areas in the speakers' immediate neighborhood. Questions of this kind were especially of interest in the early years of PD [see studies conducted in the Netherlands (Weijnen, 1946) or in Japan (Mase, 1999; Sibata, 1999)]. Indeed, the draw-a-map task is based on those used by perceptual geographers in both small and large areas [see Gould & White (1986) for more discussion of such methods].

This paper uses data from two studies that took different approaches to the investigation of the perception of language variation. The first (Study 1[2]) is a large-scale survey, whose aim was to look at the national "picture" of language variation. The second (Study 2[3]) took a small-scale approach, with the aim of investigating local perceptions of variation. In the next section we discuss the datasets we will consider in this article.

2. Methods

The two studies considered here used the draw-a-map task. Both gathered data in Europe, although in different countries, and investigate perceptions of variation in different languages. Study 1 investigated the large-scale perceptions of dialects in Great Britain. The data presented from Study 2 deal with the subjective construction of local dialect areas in the southwest of Germany as well as in some places in Switzerland and in France (for first results, see Stoeckle, 2010, 2011, 2012).

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 4. Completed draw-a-map task (Montgomery, 2011).

Map 4 shows a completed hand-drawn map from Study 1, while Map 5 shows dialect areas drawn by a respondent from Study 2.

Study 1 took a large-scale approach, with the aim of gathering data relating to the national "picture" of perception in Great Britain from five survey locations around the Scottish-English border. In this way, the study aimed to investigate the impact of the Scottish-English border on the perception of language variation in English (see Montgomery, forthcoming). Map 6 shows each of the survey locations and the survey area (Scotland, Wales, and England).

Respondents in Study 1 were given a minimally detailed map containing country borders and some city location dots.[4] In all locations, they were asked to complete the paper map with a pen or pencil in the following fashion:

(1) Label the nine well-known cites marked with a dot on the map.[5]
(2) Do you think that there is a north-south language divide in the country?[6] If so, draw a line where you think this is.
(3) Draw lines on the map where you think there are regional speech (dialect) areas.
(4) Label the different areas that you have drawn on the map.
(5) What do you think of the areas you have just drawn? How might you recognize people from these areas? Write some of these thoughts on the map if you have time.

A location map which contained a number of cities and towns in England, Scotland, and Wales was projected for respondents (who completed the task as part of a class) for the first five minutes of the task, which lasted for ten minutes overall. One hundred and fifty-one respondents in total completed the fieldwork, seventy-six on the Scottish side of the border, and seventy-five on the English side. The mean age of the respondents was sixteen years and six months. Respondents drew 970 lines delimiting seventy-nine separate areas (an average of 6.4 areas drawn per map).

Study 2 is a small-scale survey dealing with the question of how nonlinguists construct dialect areas on a local level. The data collection took place in the south-west of Germany as well as in some places in France and in Switzerland. Map 7 gives an overview of the research area and the thirty-seven investigated locations.

As demonstrated in Map 7, thirty-two survey locations are found in Germany, three in Alsace (France) and two in Switzerland. It was the aim in each location to interview six respondents, differentiated by the sociodemographic variables of age, sex, and profession. In some locations it was not possible to find speakers for all categories, and the total number of interviews was therefore 218 (instead of 222, the number originally aimed for).

As part of the interview, respondents were asked to complete a draw-a-map task where they were given a map and asked to draw:

(1) their own local dialect area, and
(2) all other surrounding dialect areas they knew of

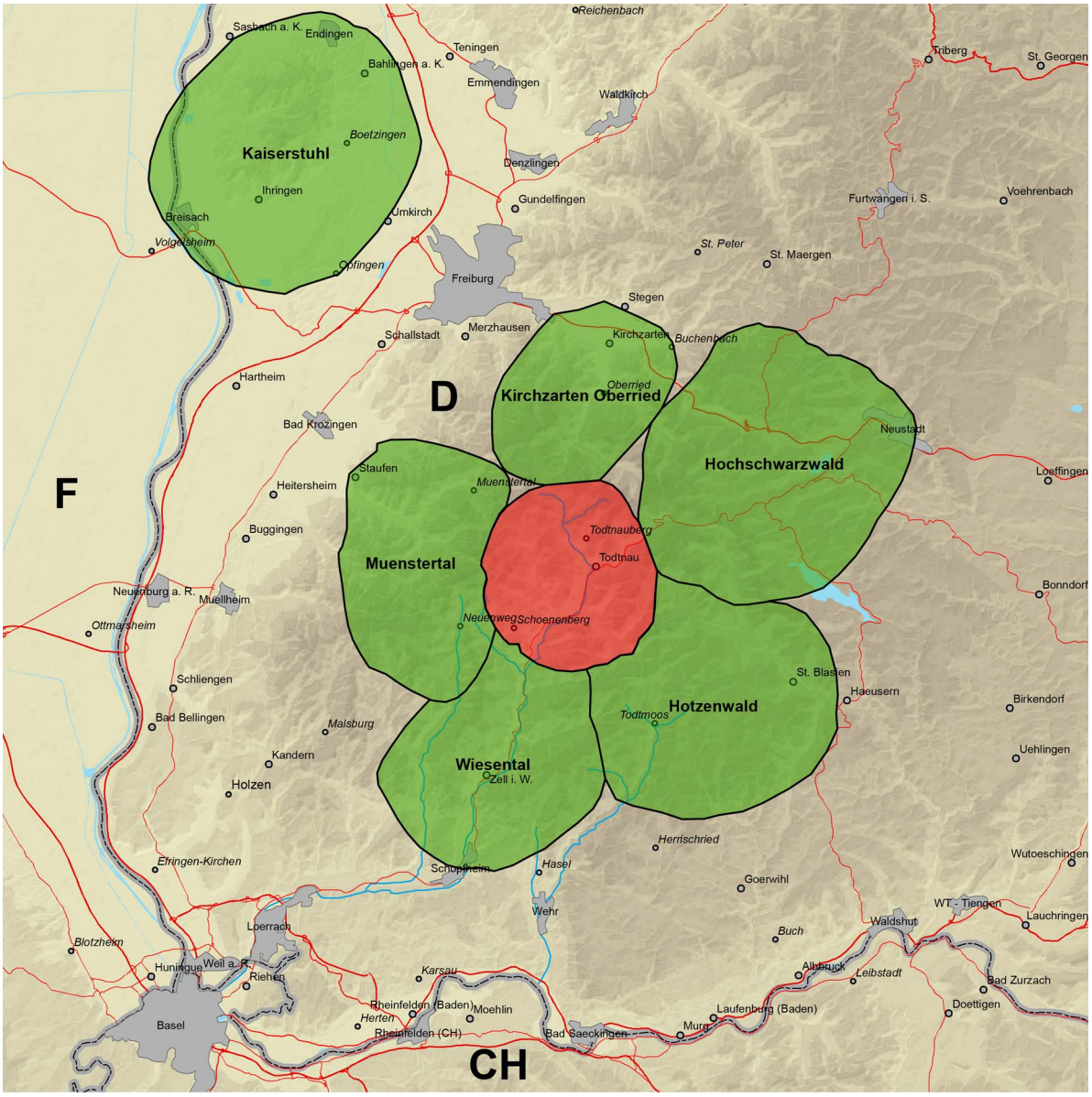
Once they had completed the initial task,[7] the map served as a starting point for further characterizations of the dialect areas. These concerned:

(3) dialect features or stereotypes,
(4) similarities/differences with regard to the respondents' own dialect,
(5) evaluations of the intensity of dialect of the identified areas and
(6) judgments about the most (and least) pleasant dialects.

The data generated in the interviews were subject to both qualitative and quantitative analyses. In this paper we will focus on the latter.

Studies 1 and 2 take slightly different approaches to the study of large- and small-scale perceptions. However, their similar use of a draw-a-map task in order to gather spatial data relating to the mental maps of dialect area boundaries (seen in Maps 3 and 4) means that, although the cognitive concepts may differ in each case, the data generated in both types of research are very similar and thus require the same type of digital processing.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 5. Hand-drawn local dialect areas by a respondent from Todtnauberg from Study 2 already entered into a GIS system.

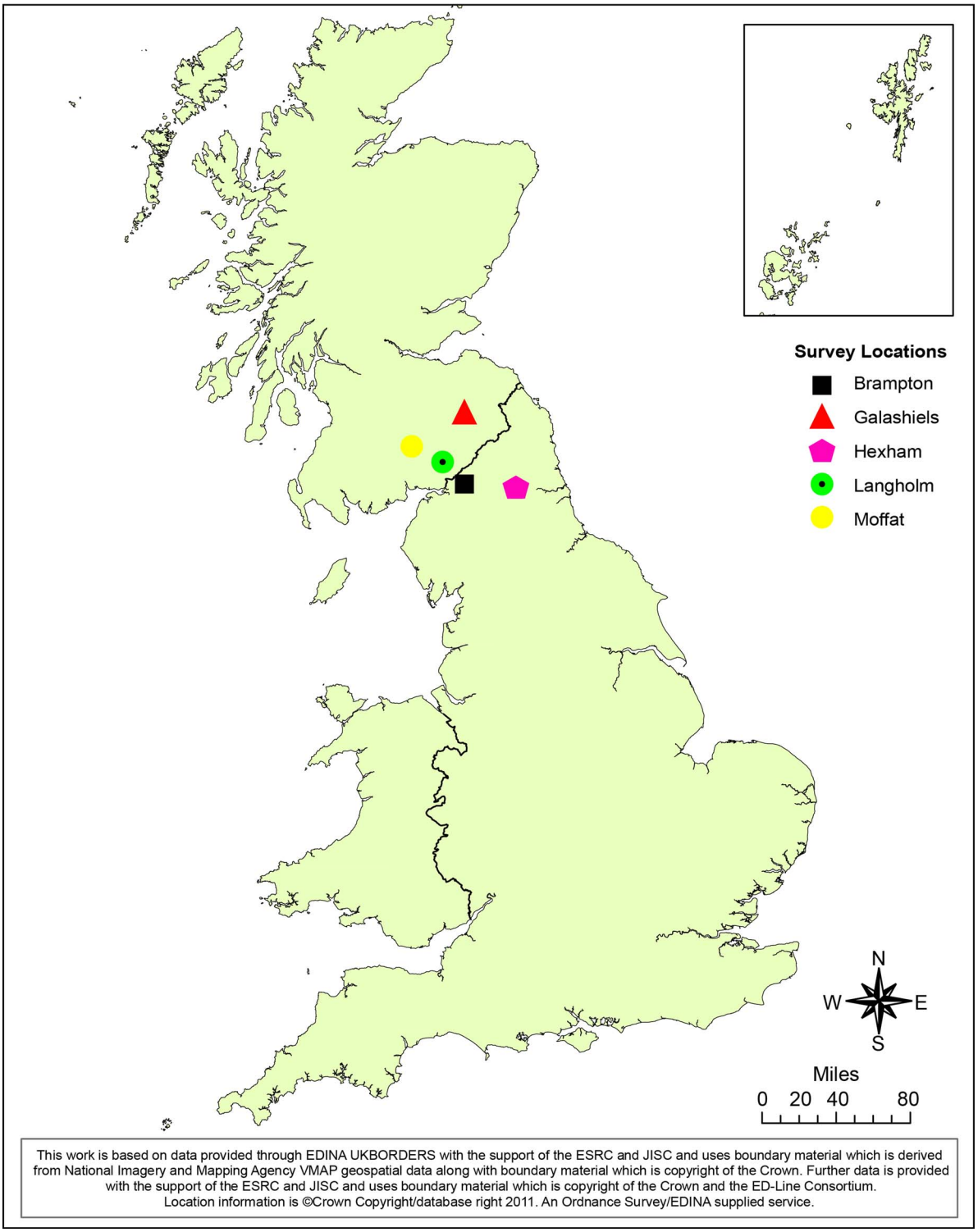
3. What is a GIS, what does it do, and why should we use one?

In the following we present some characteristics of a GIS. Since these systems are very complex in nature, the literature contains many different approaches to the topic. Some deal with detailed explanations of the workings of the technology whereas others discuss specific aspects and tools provided by it. We wish to give a more basic outline here, focussing on what a GIS is and what it can be used for in relation to PD work.

A GIS is defined as a system that integrates the three basic elements of hardware, software, and data "for capturing, managing, analyzing, and displaying all forms of geographically referenced information" (ESRI, 2011b). In this article we use ArcGIS[8] (cf. Evans, 2011) to process and display our data, although we will attempt to explain the steps undertaken for data processing in a general fashion so that they can be adapted for other types of GIS software.

The main way in which a GIS works is by combining different types of data (see section 3.1) by linking them to the earth's surface. This technique is termed "georeferencing" and it permits a GIS to "combine semantic and geometrical information" (Gomarasca, 2009:481). Georeferencing uses coordinate systems in order to tie data to a set position on the earth's surface. Spatial data are usually stored in GIS

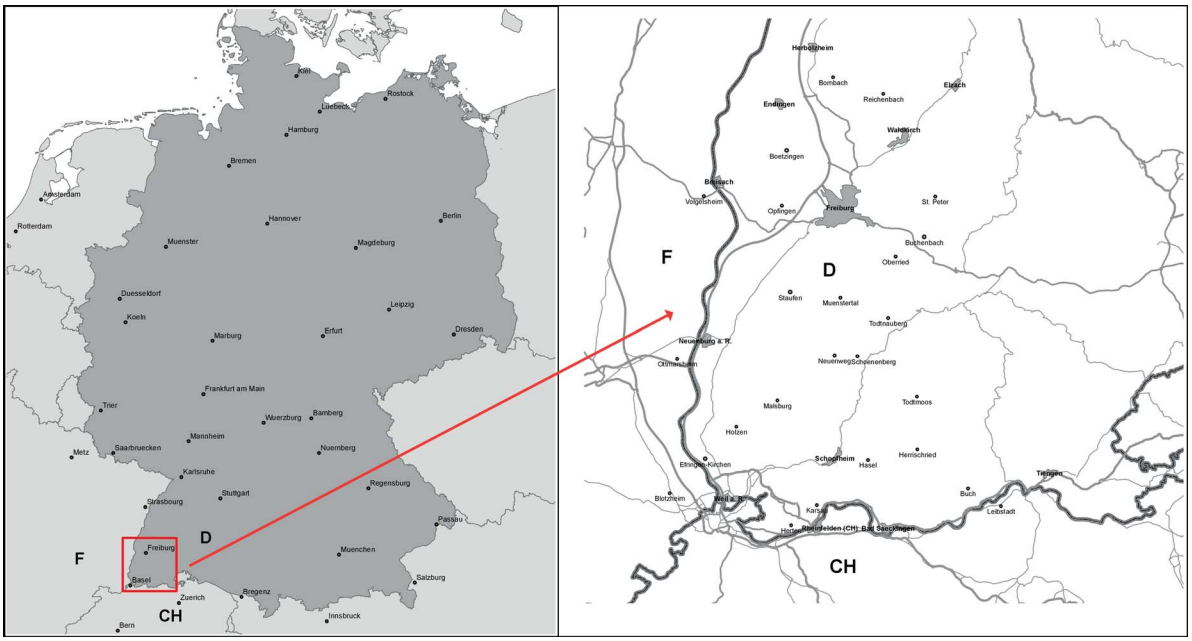
https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 6. Map of research area – Study 1.

by using latitude and longitude as the common referencing system, and they are displayed using a projection system (such as Lambert's conformal conic for mid-latitude areas or the Universal Transverse Mercator). The choice of the projection may depend on location of the area or by the need to minimize distortion in size, area, direction, etc., depending on the shape of the area. However, the use of different projections can cause some confusion for users of GIS programs, although in most cases the national grid

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 7. Map of research area – Study 2.

projection of the users' home country should be used in georeferencing. We discuss georeferencing in relation to PD data in more detail below.

Once data has been georeferenced, a GIS offers many possibilities for advanced data processing (known as "geoprocessing"). Many geoprocessing tools are designed for commercial or environmental ends, although they can also be used for other purposes such as working with linguistic data. In addition to various possibilities offered by geoprocessing tools, a GIS also provides different ways of visualizing data or creating maps. Thus, maps are georeferenced and therefore spatially meaningful, unlike conventional maps which contain only visual information (i.e. they consist of pixels of different colors). Moreover, all geographical data can have or be linked to many different types of attributes (metrical, numerical, descriptive, complex; cf. Gomarasca, 2009:484).

In summary, a GIS enables a user to process, analyze, and visualize all kinds of models of the earth's surface. This makes the technology attractive not only for geographers and geologists, but also for researchers of other disciplines (like archeology, forestry, architecture, or civil engineering) as well as administrative applications (like urban planning or traffic control) (Saurer & Behr, 1997:10). In (perceptual) dialectology, however, such technology has been used very rarely so far (exceptions being Kirk & Kretzschmar, 1992; Labov et al., 2006; Lameli et al., 2008; Evans, 2011; Cukor-Avila et al., 2012; Jeon & Cukor-Avila, 2012). This is despite the fact that dialectological questions and problems are by definition related to geographical space. Generally speaking, much simpler technologies have been used to create maps, the aims of which were not necessarily "spatially sensitive" (Britain, 2009:144).

In dialect production studies, all necessary geographical information is selected by the researcher in advance (e.g. the survey locations). Geographical space then serves as a template (or "blank canvass"; Britain, 2009:144) onto which different linguistic features can be assigned to predefined places. In PD, however, geographical data do not only serve as background. They also present the object of study as they are the data given by the respondents though their completion of hand-drawn maps. The enormous advantage of GIS lies in its ability to process, analyze and visualize these data and to combine them with reference to other geography-related data such as topography, political boundaries, population statistics or dialect isoglosses (cf. section 4.1).

3.1. How a GIS works with data

Computers "require unambiguous instructions on how to turn data about spatial entities into geographical representations" (Heywood et al., 2006:77), and as a result a GIS works with data in specific ways. Understanding the different ways in which a GIS deals with data from the real world is important if we wish to use the technology to process data from PD (Heywood et al., 2006:77).

A GIS works with data in "layers," overlaying them in order to produce composite maps. These layers of

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

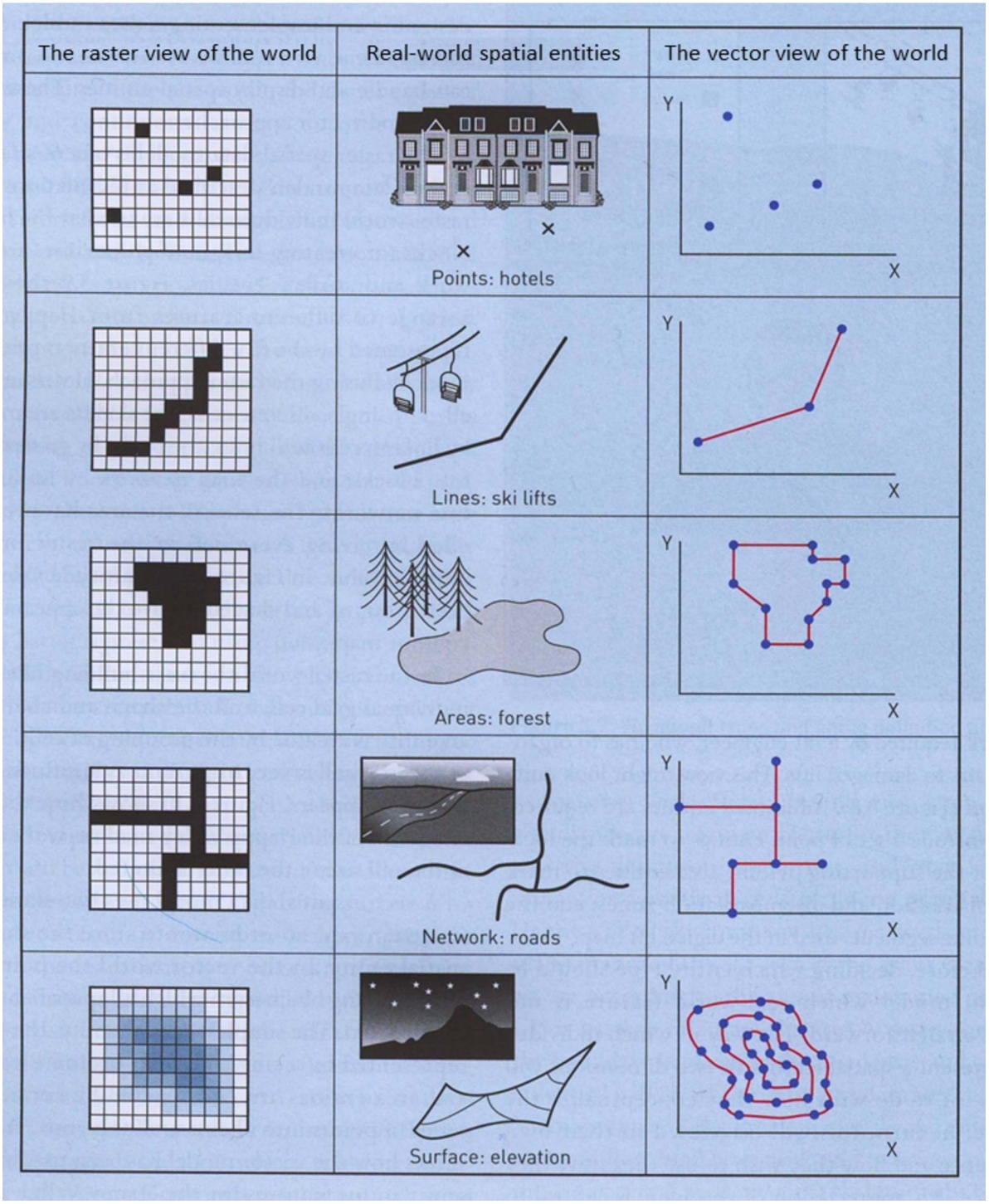
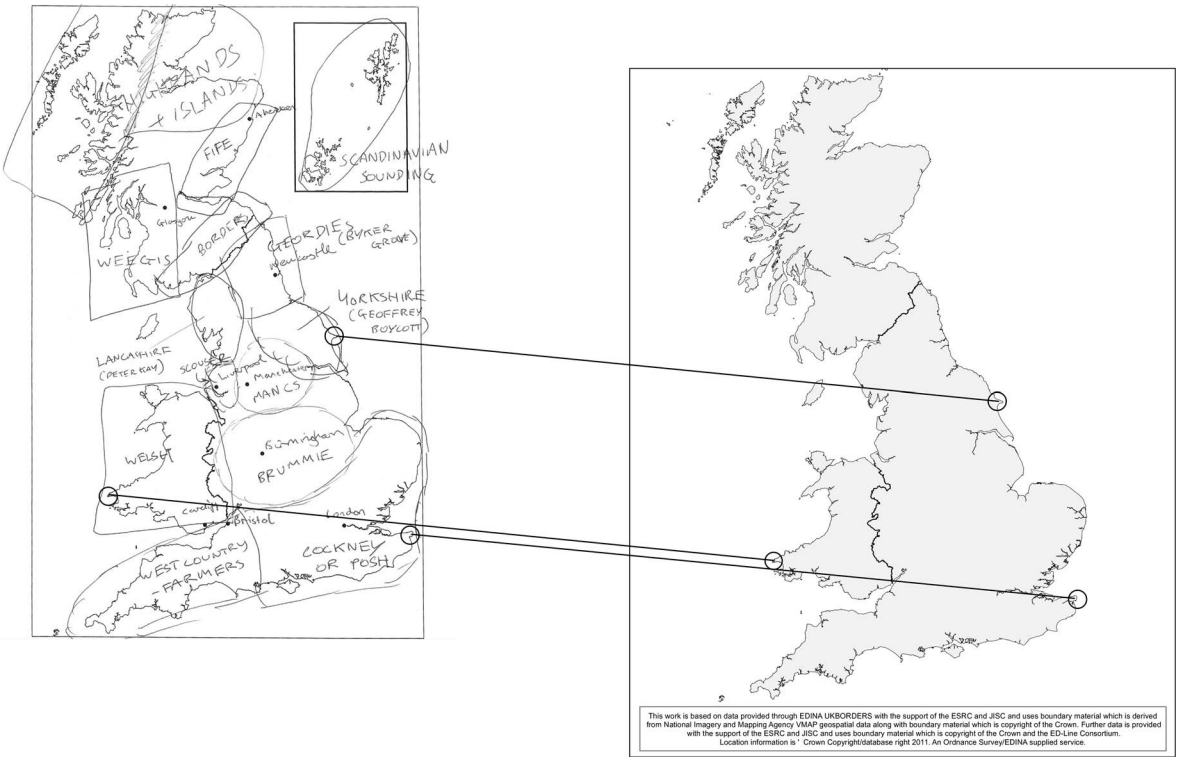


Figure 1. Vector and raster data (adapted from Heywood et al., 2006:78).

data can be queried and manipulated, and the relationships between them investigated. This makes GIS technology particularly attractive for multi-layered data such as that gathered in PD research. A GIS works with different types of data, and we wish to draw readers' attention to the distinction between the two primary types of (spatial) data: raster and vector.

Raster data can be imagined as a grid, or as consisting of cells. Each of these cells has a certain value that is "mirrored by an equivalent row of numbers in the file structure" (Heywood et al., 2006:79). A real-world object mapped as a raster will therefore "fill" some of the cells in the grid, which will correspond to its shape in the real world. Since every

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 8. Georeferencing and control points.

pixel – the smallest element that can be visualized—has a value, raster datasets can become very large regarding storage requirements. In this respect the pixel size is also an important factor. While a smaller pixel (or cell) size implies a higher resolution of the presented surface and may therefore give a more detailed view of the phenomena to be dealt with, it also leads to a larger file size and slower processing of the data. So the choice of the pixel size should always depend on the level of detail to be captured from the real world. For these reasons, raster datasets are especially useful as surface models of small geographic areas. Unlike vector datasets, they cannot be connected to or contain attribute tables (see below), which limits their usefulness if a user wishes to query the data at a later stage.

Vector data use co-ordinates to map real world objects, as opposed to the grid and cell method used by raster datasets. The file structure of a vector dataset is a series of co-ordinate points. These points can be connected in order to form lines or polygons. Unlike areas in raster datasets there is no information stored about surface characteristics (i.e., the individual points within an area). Figure 1 shows the different ways in which vector and raster data are represented in a GIS.

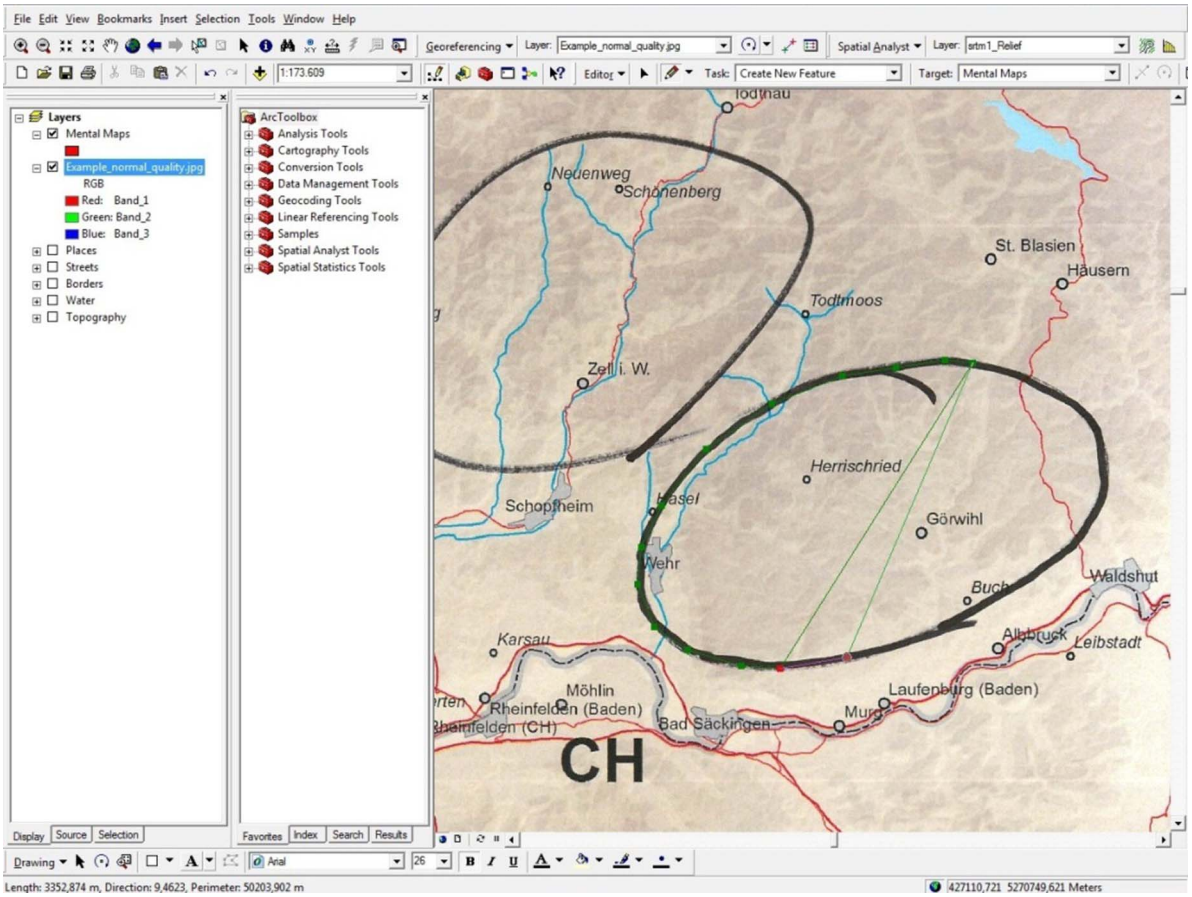
Attribute data are a third type of data (Nash Parker & Asencio, 2008:xvi), and they are also important for GIS processing. This data type provides descriptive information linked to the map data by the GIS. It can contain information about the name of an individual piece of the map data, for example, but can also contain a good deal more information about the map data (such as population size, statistical information, etc.). We will demonstrate the use of both raster and vector datasets in this article, along with attribute data, which assists in querying processed data.

3.2. *General steps involved in processing data from hand-drawn maps*

The steps involved in processing data from hand-drawn maps described below do not differ significantly from those used by Preston & Howe (1987) or Onishi & Long (1997). Data relating to dialect areas still need to be extracted from maps, attribute data (in the form of demographic information) added, and the data processed. Only then can aggregate maps of dialect areas be displayed. The ArcGIS-based method we detail below follows these steps relatively closely, although it does not use technology designed specifically for the task. This means that what we describe can at first seem daunting; however, the advantage of using a widely used and available "off-the-shelf" program will be demonstrated as we proceed.

Although a complete account of every data processing stage will not be possible here for reasons of space, it is

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 9. Redrawing of mental map.

worth noting at this point that the instructions below will require some basic familiarity with the ArcGIS environment (or the equivalent environment of the GIS you wish to use). This cannot be conveyed here, although there are several useful resources available online and elsewhere.[9] We should also emphasize that the benefits of "picking up" techniques by using the software should not be underestimated.

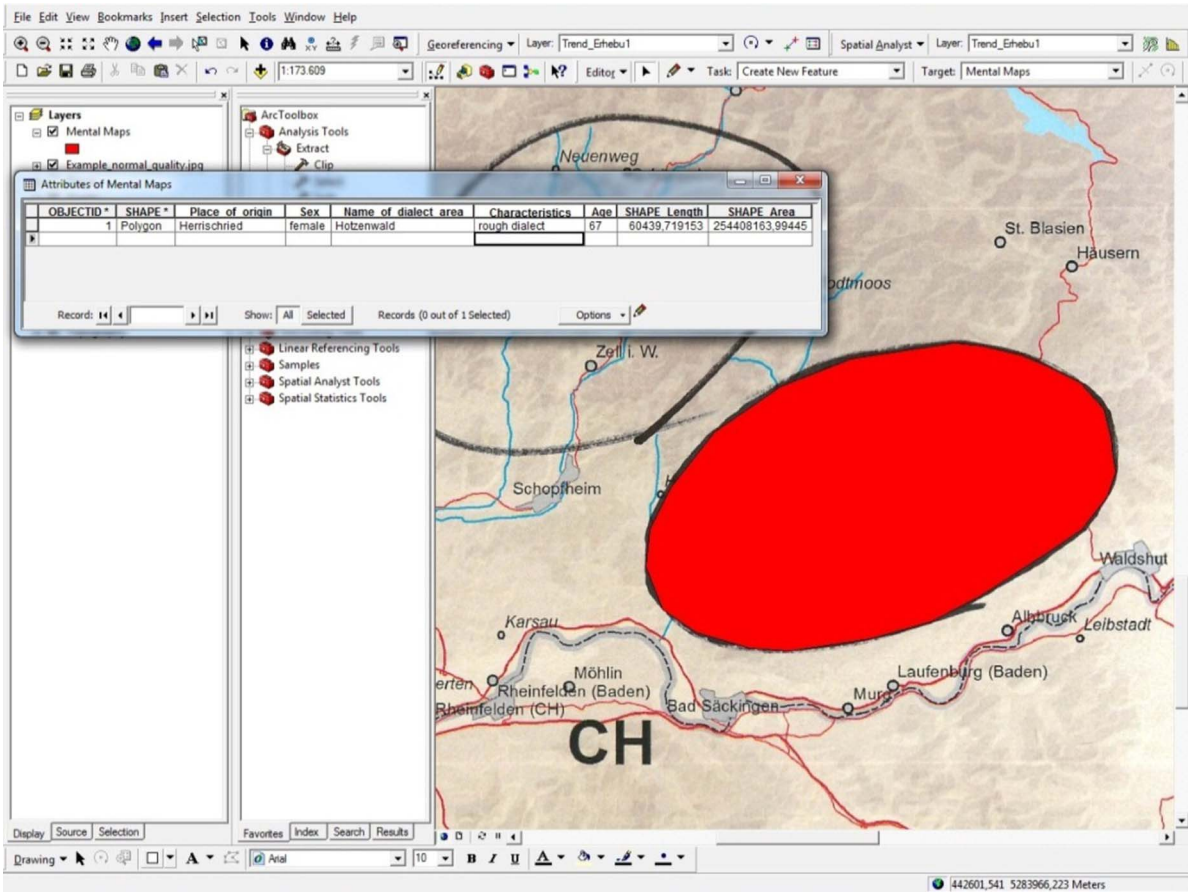
As we discuss above, the essential characteristic of a GIS is that it enables users to work with data that are georeferenced. The first data processing stage is therefore to scan all of the hand-drawn maps and to add them to an ArcGIS "project" (the term the program gives to map documents). Georeferencing can then be done using the "Georeferencing" tool by adding "control points." "Control points" are points that have been selected on a map that can be aligned with known points on another map. This means that if there is no information about a map's coordinate system, it can be georeferenced by using existing data (such as borders, rivers, etc.) as reference points that can be associated with the map with the help of the control points. Map 8 shows the principle behind georeferencing, in which three control points have been identified.[10]

The remainder of this section will use data from Study 2, in order that a clear workflow can be observed. Map 9 shows a sample of a map from this study that has been scanned, added to an ArcGIS project, and then georeferenced.

Once the map is georeferenced, the dialect areas drawn by the respondents must be digitized. In ArcGIS this can be managed by creating a polygon feature class (a vector data type; cf. section 3.2). After the creation of the polygon feature class, a file is created which as yet contains no data. Slots for attribute data can be created during this step, which will allow the user to input further data (such as demographic or attitudinal data) at a later stage.

In order to populate the new polygon feature class, the "Editor" tool is started and the tool "Create New Feature" used. This permits the dialect area indicated on the hand drawn map to be entered into the feature dataset (here named "Mental Maps") by tracing around it. As we have discussed above, respondents in both studies were not only asked to draw maps, but were also requested to label the areas and to evaluate them according to different aspects (cf. section 2). GIS offers the possibilities to add any kind

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 10. Redrawn dialect area (red oval) and attributes table.

of attributes to the datasets (cf. section 4.1). In the case of Study 2, attributes relating to the respondents (place of origin, sex, age) and the dialect area (name of dialect area, characteristics) were added to the attribute table. Map 10 shows both the redrawn dialect area as well as the table containing different attributes relating to it.[11]

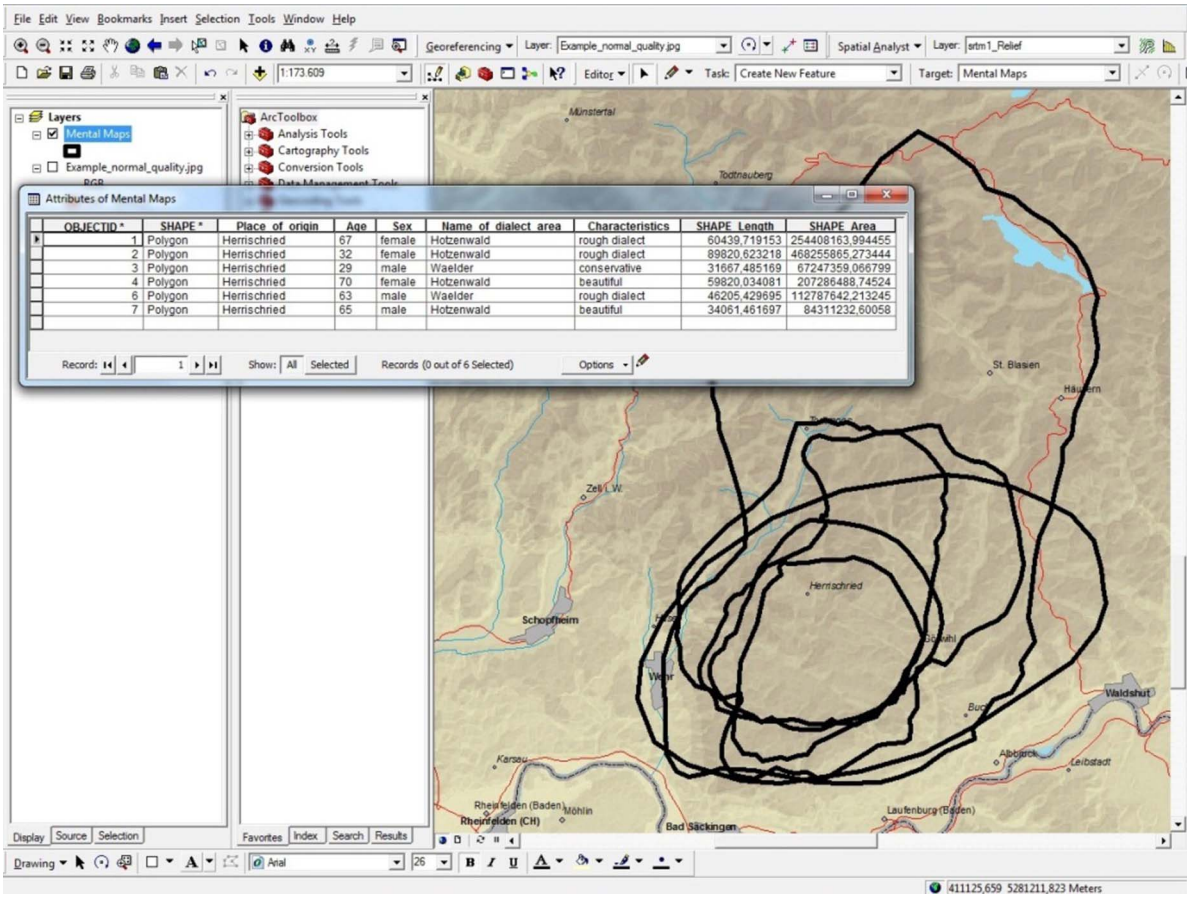
The next stage of the processing method is the hand-drawn map aggregation. The first step of this process involves adding every redrawn area to one dataset (using the same process as described above). Map 11 shows the same dataset as in Map 10, but now containing six different polygons (each representing perceptions of the same dialect area drawn by six different respondents) with their respective attributes in the table.

Up to this point, the different polygons are stored in one dataset and as a result it is not possible to show different degrees of agreement, which is one of the aims of the method. This can be achieved by a two-stage process: First the self-union of the feature class containing all the polygons has to be calculated (by using the "Union" tool), and then the frequency of each of the polygons in the output has to be counted (by using the "Frequency" tool) and the output of this calculation added to the map. The frequency count of all the polygons, in this case ranging from one to six, gives the different degrees of overlap. Map 12 shows a possible visualization as a result of this process.

Above, we have outlined the steps that will produce a basic map displaying agreement about the placement and extent of a dialect area among a group of respondents. Data processing should not stop here, however, as this type of dataset (i.e. vector data) requires a large amount of memory space and is thus hard to handle.[12] Second, it is difficult to either merge the dataset with other kinds of datasets (e.g. more polygons indicating a dialect area, or neighboring regions) or to perform further analyses on it. Third, it displays all of the single values of overlap, which results in too much influence from single areas and many sharp borders. Conversion from vector to raster data is therefore helpful[13] as this data format permits processing without these drawbacks.

The process outlined above requires the use of a large dataset in order that the benefits become most apparent. To this end we have used data from Study 2 relating to the so-called "Kaiserstuhl" [literally *emperor's chair*], a small mountain and former volcano very close

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 11. Several dialect perceptions added to one dataset.

to the French border that is very well known for its viticulture. This was the most readily recognized area among respondents in Study 2. Of the total of 218 respondents, ninety-five identified and drew this area. Using the same stage of the data processing technique as shown in Map 12, Map 13 shows the perceived "Kaiserstuhl" dialect area. For comparison, Map 14 shows a raster-based map of the perception of the same area. The different colors indicate different degrees of overlap, ranging from one (green) to ninety (red).

Although containing the same data, the raster dataset shown in Map 14 gives a much better impression of respondents' perception of the "Kaiserstuhl" dialect area than that displayed in Map 13. The "Neighbourhood Statistics" tool has been used in Map 14 in order to smooth the surface of the data, which makes any sharp edges between the different degrees of overlap disappear.[14] A continuous scale has been used with contour lines added. The contour lines (unlike in topographic maps) do not indicate altitude, but degree of overlap.

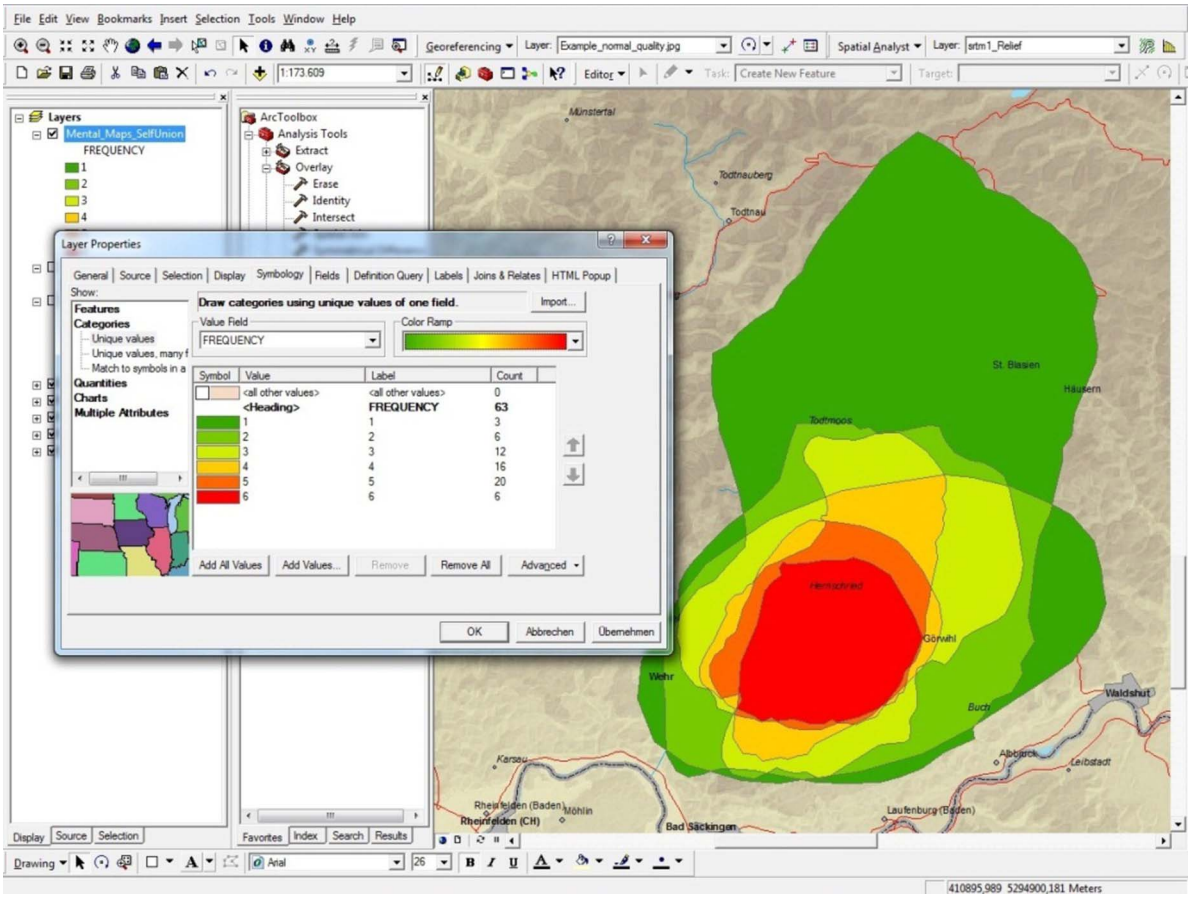
The data processing technique described above is summarized in the flow chart shown in Figure 2.[15]

There is no doubt that, in addition to improving the processing and display of PD data, the use of GIS has numerous advantages over the other processing techniques discussed above. Chief among these is the ability to make PD data more useable alongside other datasets. Other advantages include the customization of aggregate data, the ability to combine individual areas on the same map, as well as the numerous possibilities to perform calculations and statistical analyses on the data.

4. Merging different datasets on one map

GIS allows us to examine the impact of many factors on a much wider scale and in a much more efficient fashion by permitting us to merge many different datasets on the same map, as well as enabling interrogation of these datasets using tools within the GIS. This ability permits spatially sophisticated analysis of (perceptual) dialectological data (Britain, 2002:633). There are a vast number of additional datasets for Great Britain available via various sources such as data.gov (HM Government, 2011),

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 12. Map showing different degrees of overlap.

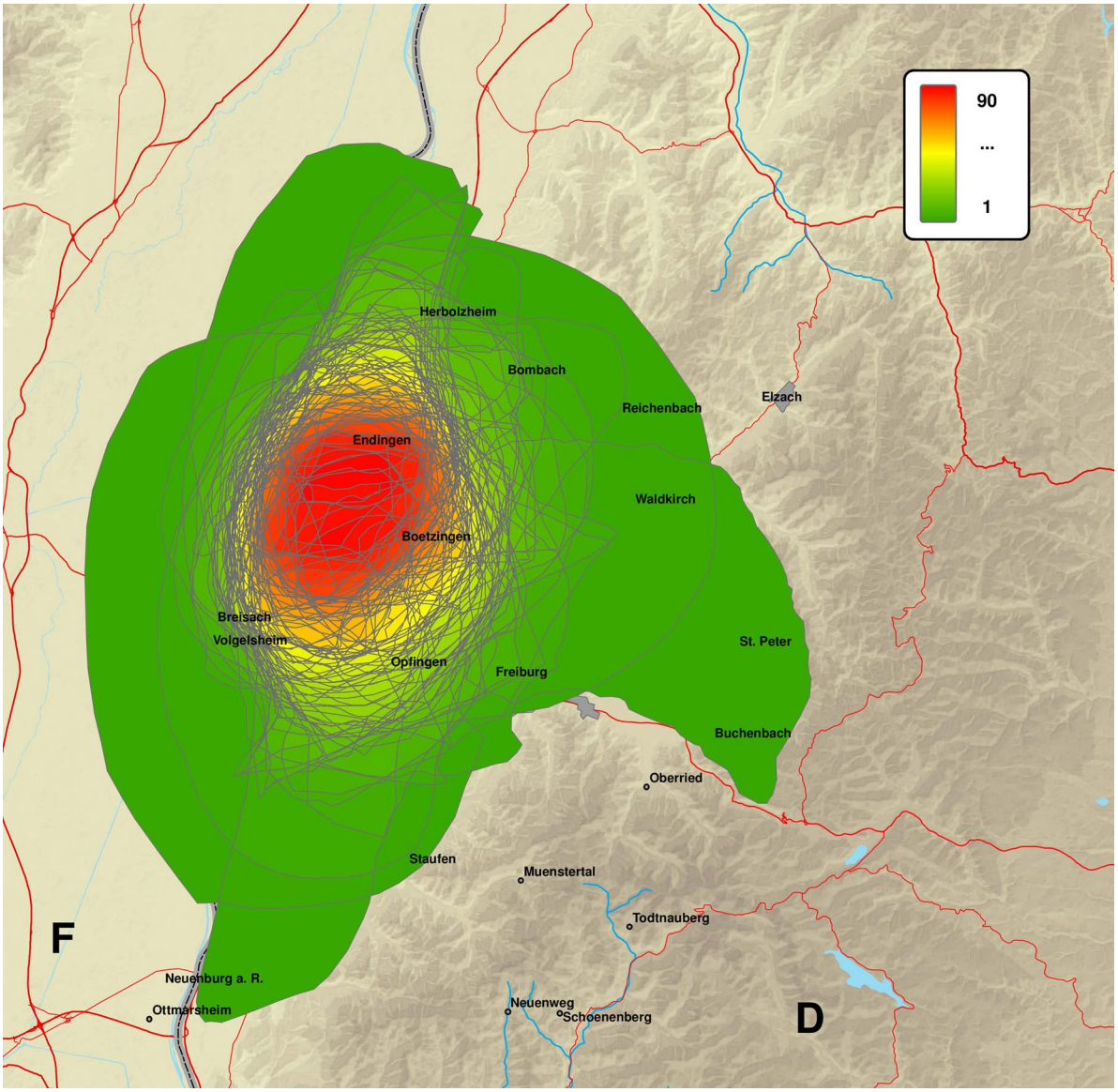
Digimap collections (Edina, 2011), OS Open Data (Ordnance Survey, 2011) and in the numerous collections gathered at census.ac.uk (U.K. Data Archive, 2011). Datasets relating to Germany can be found at the GeoDatenZentrum (Bundesamt für Kartographie und Geodäsie, 2011) or at Geofabrik (2011). Such datasets contain georeferenced data relating to a whole host of factors, and we will demonstrate some of these below.

We have already demonstrated merged datasets in Maps 12–14. These maps show aggregate perceptual dialect areas overlaid onto nonlinguistic datasets (like places, streets, political borders, or topography). This is of course the least that we would expect of the technology. Indeed, some of the visualizations presented in the last section (i.e. Map 13 and a simplified version of Map 14) can be achieved by using "regular" vector graphics editors (such as Corel Draw, Adobe Illustrator, etc.). However, besides the fact that all information contained within such packages is purely visual (i.e. pixels of different colors), with no attributes associated to the data, another major disadvantage is that such data cannot be used for any further processing or analyses. Thus, such tools do not move us any further past the opportunities offered by previous or existing data display/processing tools. This necessitates the use of GIS in order to undertake Gomarasca's three different types of data analysis: "Spatial Data Analysis, […] Attributes Analysis, […] and Integrated Analysis of Spatial Data and Attributes" (Gomarasca, 2009:498f).

Aggregate maps produced by perceptual dialectologists have always been examined alongside other maps in order to attempt to find correlations. Early perceptual work in Japan found that physical and political boundaries were important for respondents when completing perceptual tasks (Preston, 1993:376; Grootaers, 1999). Map 15 shows perceptual areas in the Northern part of England and the Southern part of Scotland from Study 1 with the Scottish-English border and English county boundaries superimposed. Map 16 shows aggregate data from Study 2, with religious affiliation boundaries superimposed.

Both Map 15 and 16 demonstrate that there is agreement between "official" boundaries. As discussed in more detail in Montgomery (forthcoming), the effect of the Scottish-English border is striking, with almost

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

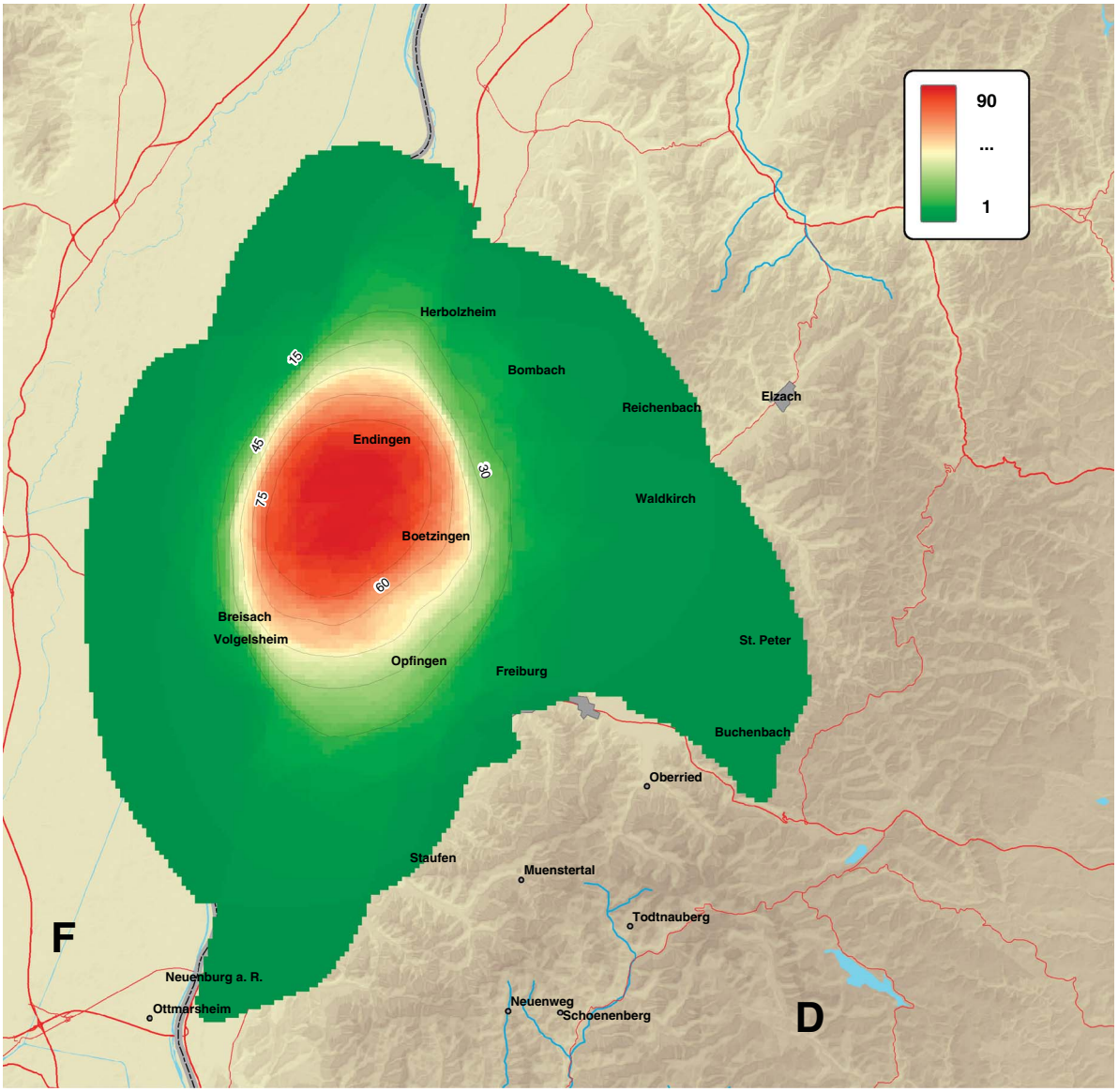


Map 13. Self-union of hand-drawn maps indicating agreement rates (vector data).

no crossing of the border for each perceptual area. The "Cumbria" dialect area in the northwest of England also fits almost entirely within the modern county of Cumbria. The "Geordie" dialect area is less respectful of modern county boundaries, although it fits well within the boundaries of the older county of Northumberland (cf. Llamas, 2000). A similar correlation between perceptual data and traditional boundaries can also be seen in Map 16. Indeed, in the interviews from Study 2, it was a striking observation that in Protestant locations many respondents explicitly referred to the traditional religious borders as the main influences on the current dialect structure (cf. Stoeckle, 2010, 2012). The ability to test qualitative statements such as this in a GIS is another factor that should recommend the use of the technology.

The use of GIS can also allow us to interrogate data in order to investigate evidence of specific linguistic phenomena. For example, regional dialect levelling is said to be having a large impact on linguistic diversity in Great Britain (Kerswill, 2003). This is underlined by maps drawn by Kerswill (The Economist, 2011; Kinchen, 2011) and Trudgill (1999:83). Such maps predict a future dialect landscape in England typified by large city-centred dialect areas. As nonlinguists' perceptions could act as a bellwether for language change of this type, a comparison between urban areas and aggregate perceptual data is appropriate. Map 17 shows this type of comparison.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 14. Rasterized version of hand-drawn maps (with contours).

Map 17 does appear to demonstrate that urban areas were important when completing draw-a-map tasks. Despite the predictions made by others (Trudgill, 1999; Kinchen, 2011), these areas have not yet been identified by dialectologists (Montgomery, forthcoming). The ability to combine PD data with that from other sources (be they datasets relating to urban areas as in Map 17 or georeferenced linguistic data) is important if we are to continue to test theories of language change.

This section has demonstrated the capabilities of a GIS in overlaying many different datasets in order to answer specific questions about the perception of dialect areas, and (in addition) it has underlined the possibilities for combining large amounts of data in the same place at the same time.

4.1. Querying and customizing the display of aggregate data

As we discussed above, the ability to query the aggregate dataset was one of the main motivations for Preston's shift to a computer-based method of working with draw-a-map data (Preston & Howe, 1987:369). The advantage of using a computer to query data and display the result is clear: The data only need to be entered once. To redraw areas by hand for each variable the researcher wishes to examine is neither desirable nor practical. To this end, query functions were built into both Preston & Howe's (1987) method as well as PDQ (Onishi & Long, 1997). PDQ's query facilities were limited to age, sex, and informant number (which could then be

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

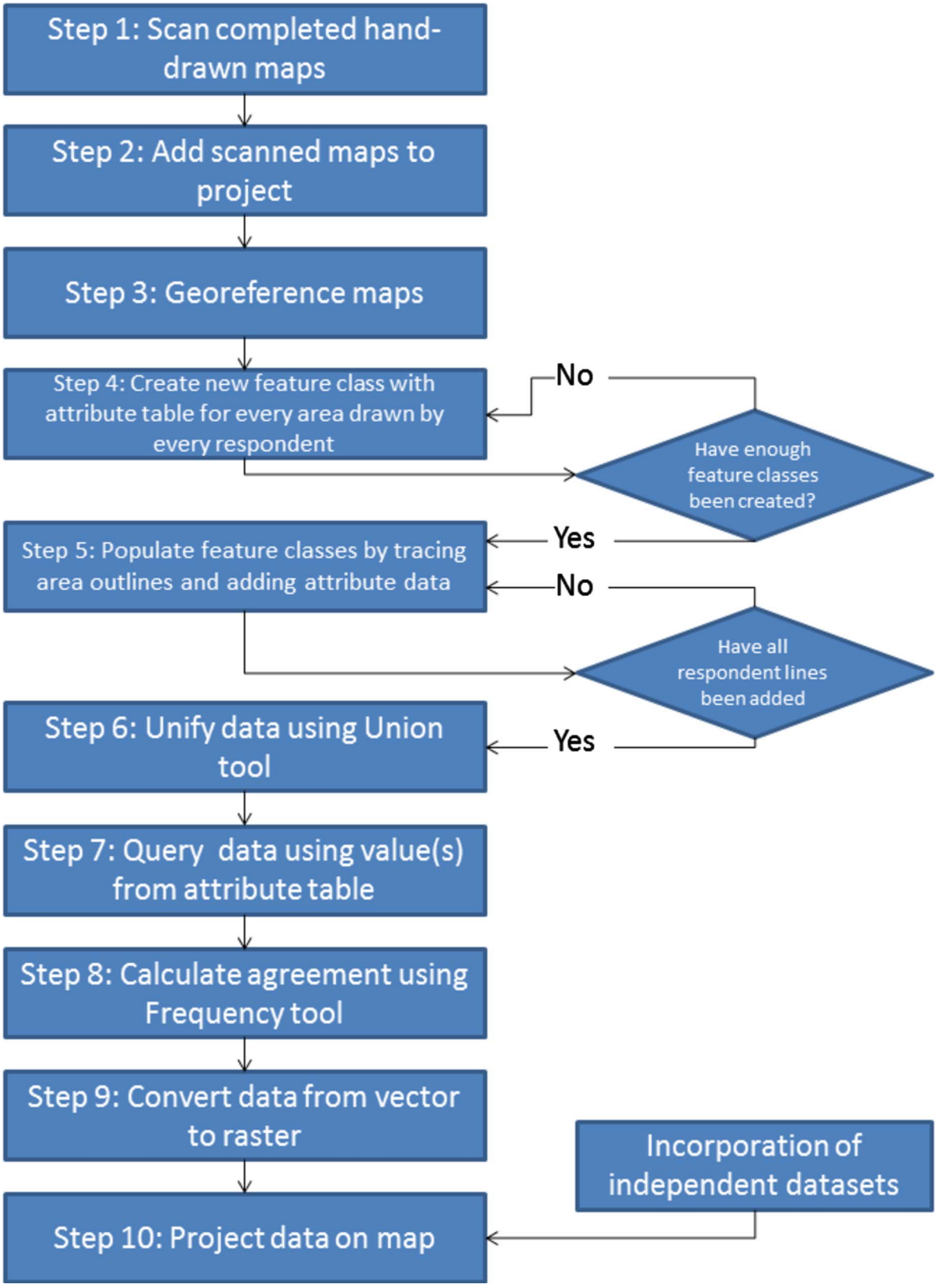
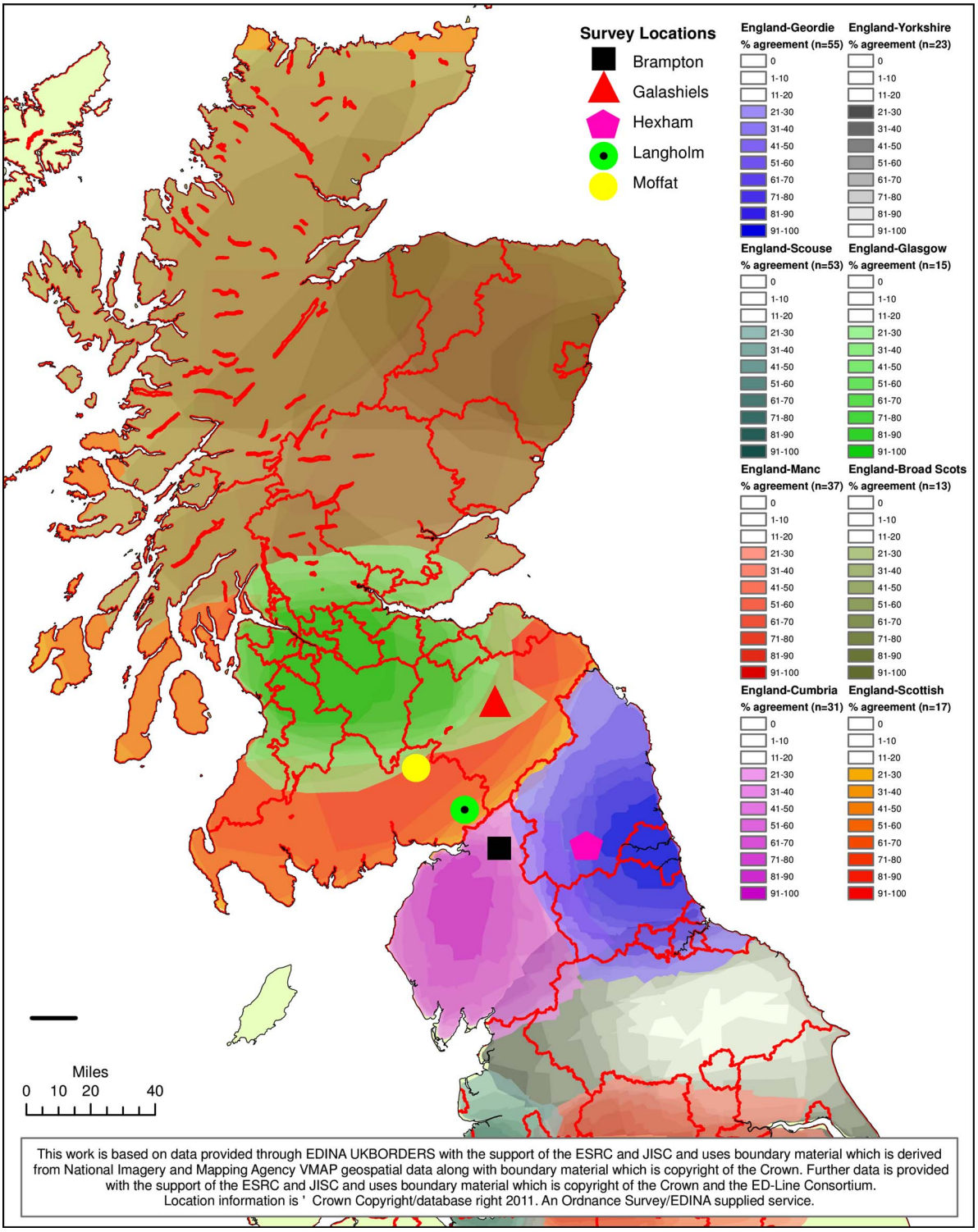


Figure 2. Workflow for processing draw-a-map data and projecting onto a map in ArcGIS.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

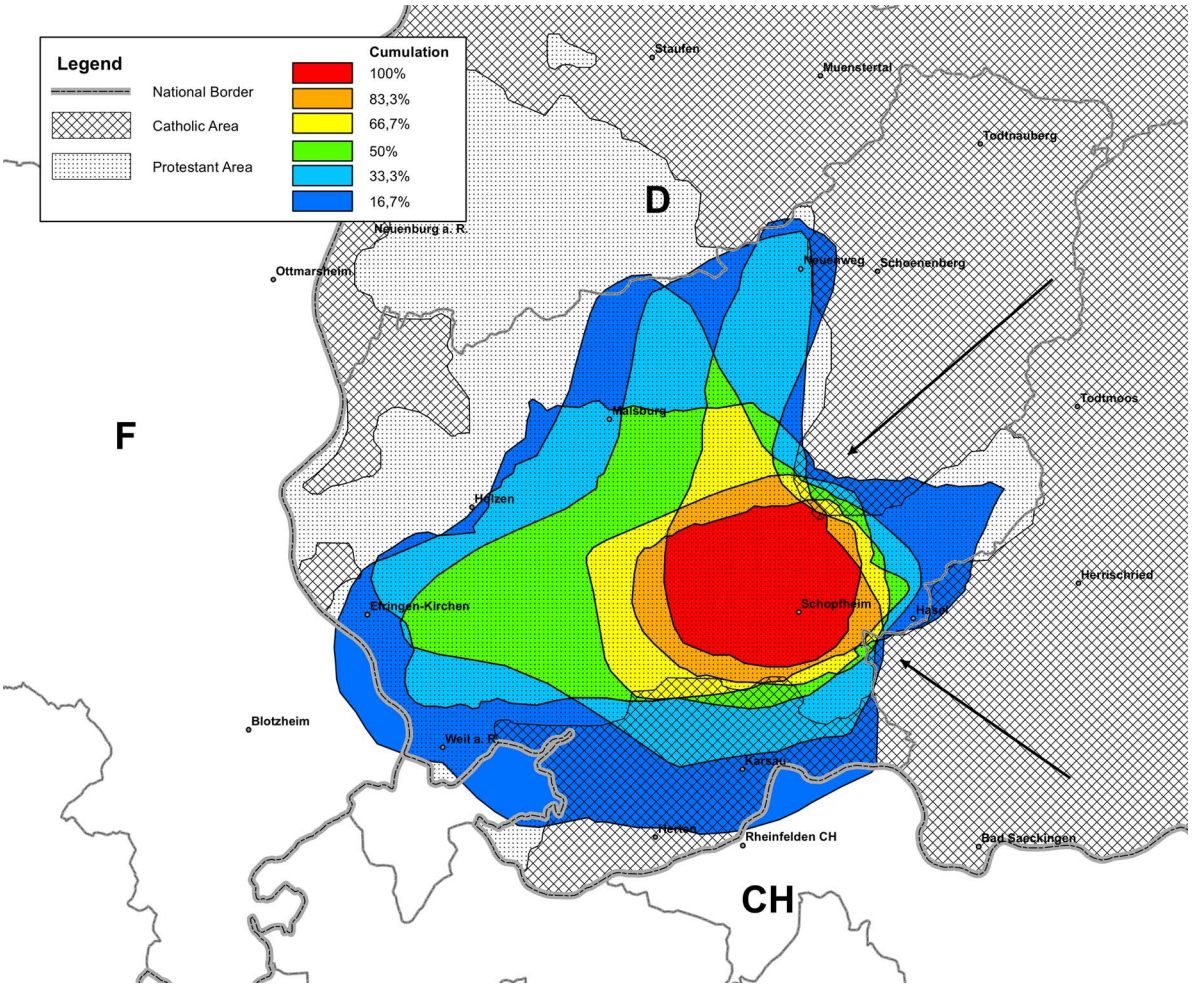


Map 15. English respondents perceptions of dialect areas and Northern England and Southern Scotland, with national and county boundaries superimposed.

used for isolating a group of respondents from a particular location) (Montgomery, 2007:95). The ability to query data entered into a GIS is, on the other hand, practically unlimited, dependent only on what an attribute table has been set up to contain (step 4 of the workflow in Figure 2).

The attribute table could contain information about basic biographical data of the type we might

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



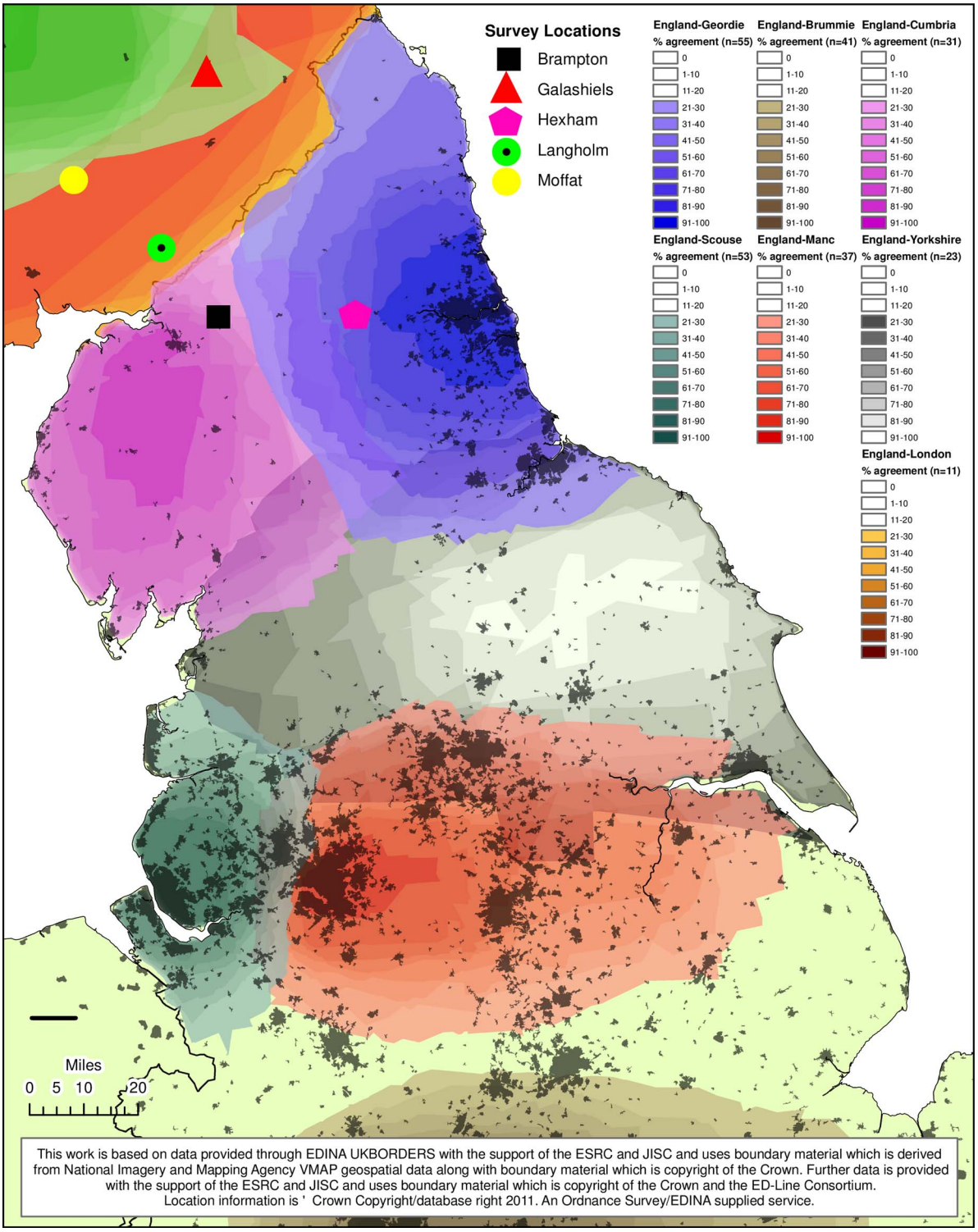
Map 16. Mental maps from Schopfheim respondents and traditional religious affiliation structuring.

expect of modern sociolinguistic approaches to speech communities (e.g., social variables such as sex, age, gender, social network score, etc.). As (perceptual) dialectologists are interested in spatiality in addition to these factors, other attributes might also be important, such as travel history or postcode (ZIP code) information relating to each respondent. We might also be interested in those dialect areas characterized as "rough," "posh," or "friendly" areas (or other labels of this sort). Details of all such variables can be added to the attribute table and then used to query the data. Map 18 shows the result of a query from Study 2 in which polygons drawn only by the male and female respondents are indicated.

Querying the datasets in a GIS need not only rely on information contained within the attribute table, and it is possible to use the geoprocessing tools, which we have previously discussed (e.g. for the calculation and display of unions, frequencies, and contours, etc.) to further interrogate processed data. In a similar fashion, GIS programs contain different kinds of measuring functions which allow calculations of distances, areas, and lengths (Gomarasca, 2009:500). Common questions that perceptual dialectologists may want to ask are: *How large is perceived area A in comparison to perceived area B? Which people draw the largest dialect areas?* (Cf. Map 18, where female respondents appear to have drawn larger areas than male respondents); *How big is the distance between a subjective dialect area and the national border?* Of course it is also possible to combine different types of dialect areas, for example "subjective" and "objective" dialect areas, and examine where they intersect and how much they overlap.

Although the primary function of PD research is to examine perceptions of dialect areas through aggregation of hand-drawn maps, in some contexts it can be interesting to determine where subjective *borders* are particularly stable (cf. Preston, 1986). Map 19 shows a summary of all dialect areas drawn by the respondents from Study 2. At first glance the image looks quite confusing, although it already gives an idea of where lines occur at a higher frequency.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

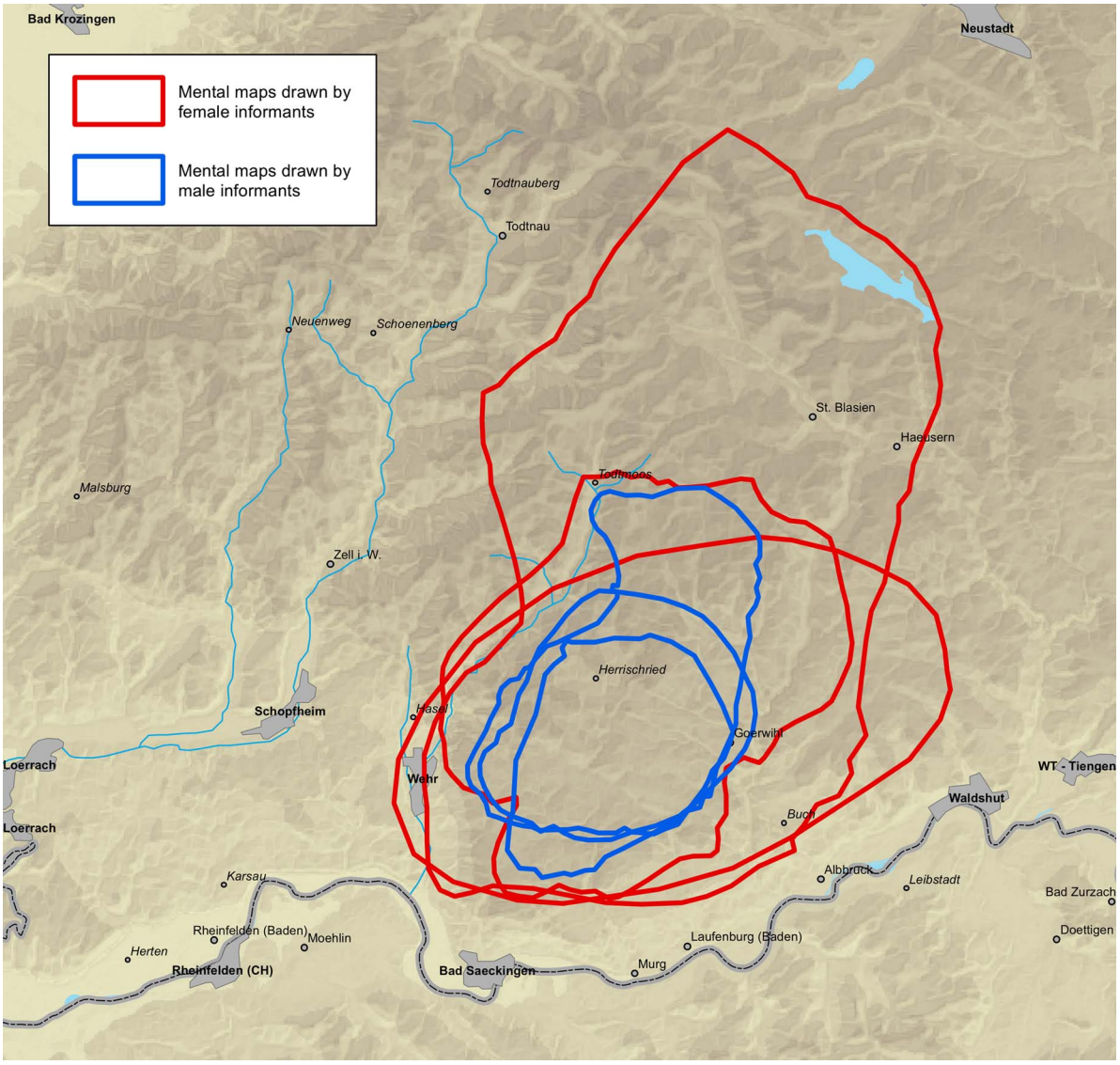


Map 17. English respondents perceptual areas, with urban areas superimposed.

For a more sophisticated insight, it is possible to calculate the line density of the subjective dialect borders using a GIS "Line Density" tool. The result is the raster map shown in Map 20 that displays the number of lines that occur within a certain research radius for each cell.

This technique gives a much clearer idea of where mental borders accumulate. There are certain correlations that are immediately apparent, most significantly the coincidence of mental and political borders.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



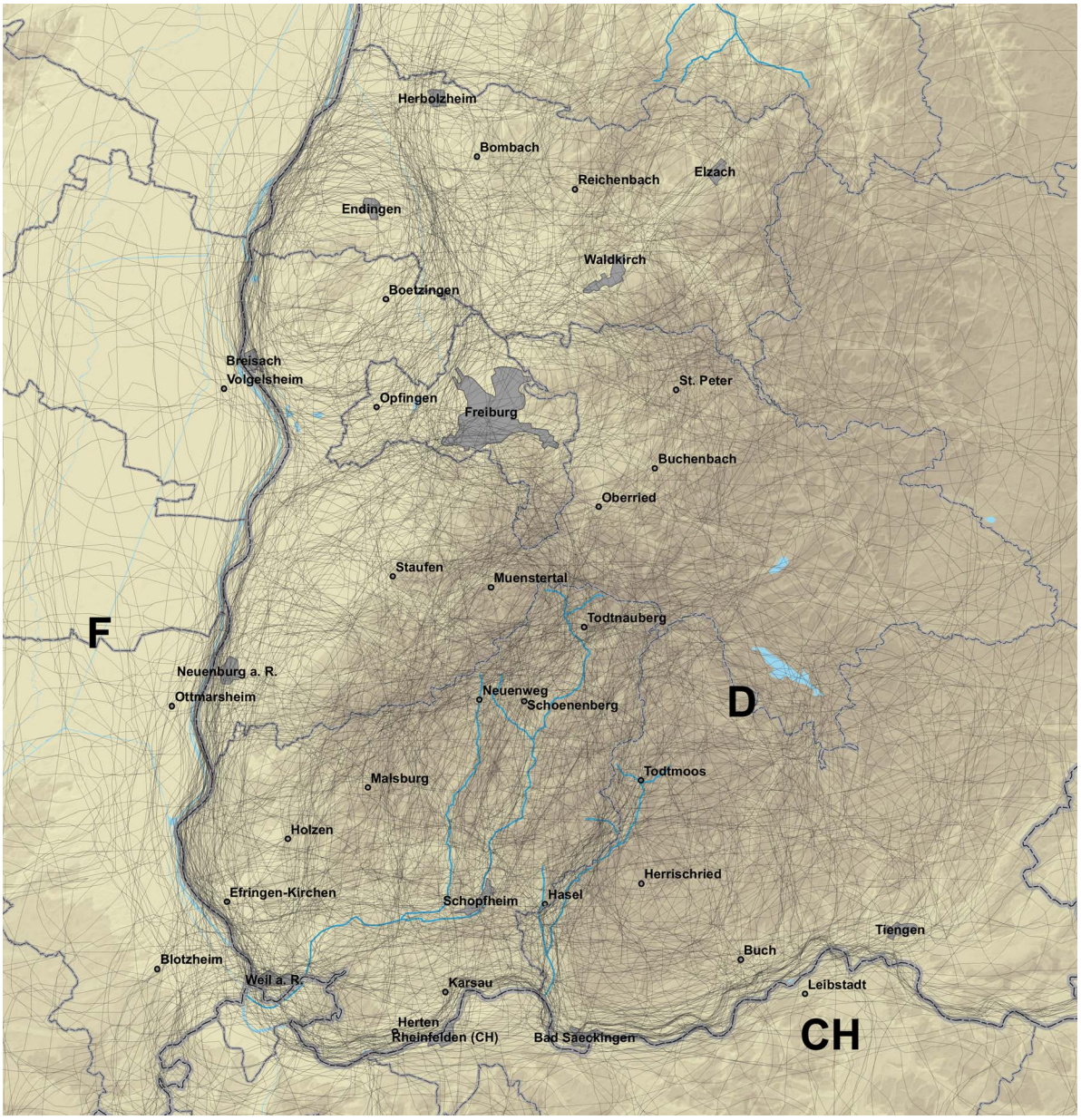
Map 18. Mental maps drawn by female and male respondents.

GIS tools also permit the customization of the display of aggregate data, something that the techniques used by Preston & Howe (1987) and Onishi & Long (1997) were not able to accomplish. In many cases it is useful to show percentages of agreement instead of absolute values (cf. Long, 1999; Montgomery, 2007). This can easily be achieved using raster datasets by using interval shading instead of continuous visualization scales (such as that seen in Map 14). Map 21 shows the use of interval shading.

Map 21 shows the hand drawn maps from the ninety-five respondents who drew the "Kaiserstuhl" dialect area. The interval size to display steps of 10 percent is therefore 9.5. Of course, PDQ permitted such a display of percentage agreement, as demonstrated in Map 2. However, what PDQ did not allow was the customization of the percentage display, for which there were fixed intervals (either 5 or 7 percentage boundaries). In addition, all of the data are shown on the composite map. There is no possibility of making some of the lower agreement level transparent, for example, in order to present the "best fit" data.

The approach that we describe here enables the user to control the amount of information presented in the aggregate map. Percentage agreement levels[16] can be customized, with low levels of agreement made transparent. Although it is possible for the user to customize the percentage agreement levels in the GIS program, most GIS software has in-built methods for class interval shading (Heywood et al., 2006:258–60). Such methods include the "equal interval" described in relation to the percentage agreement levels above, as well as other techniques including "nested means" and "natural breaks" (Heywood et al., 2006:259).

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 19. Summary of all dialect areas drawn by the respondents from Study 2.

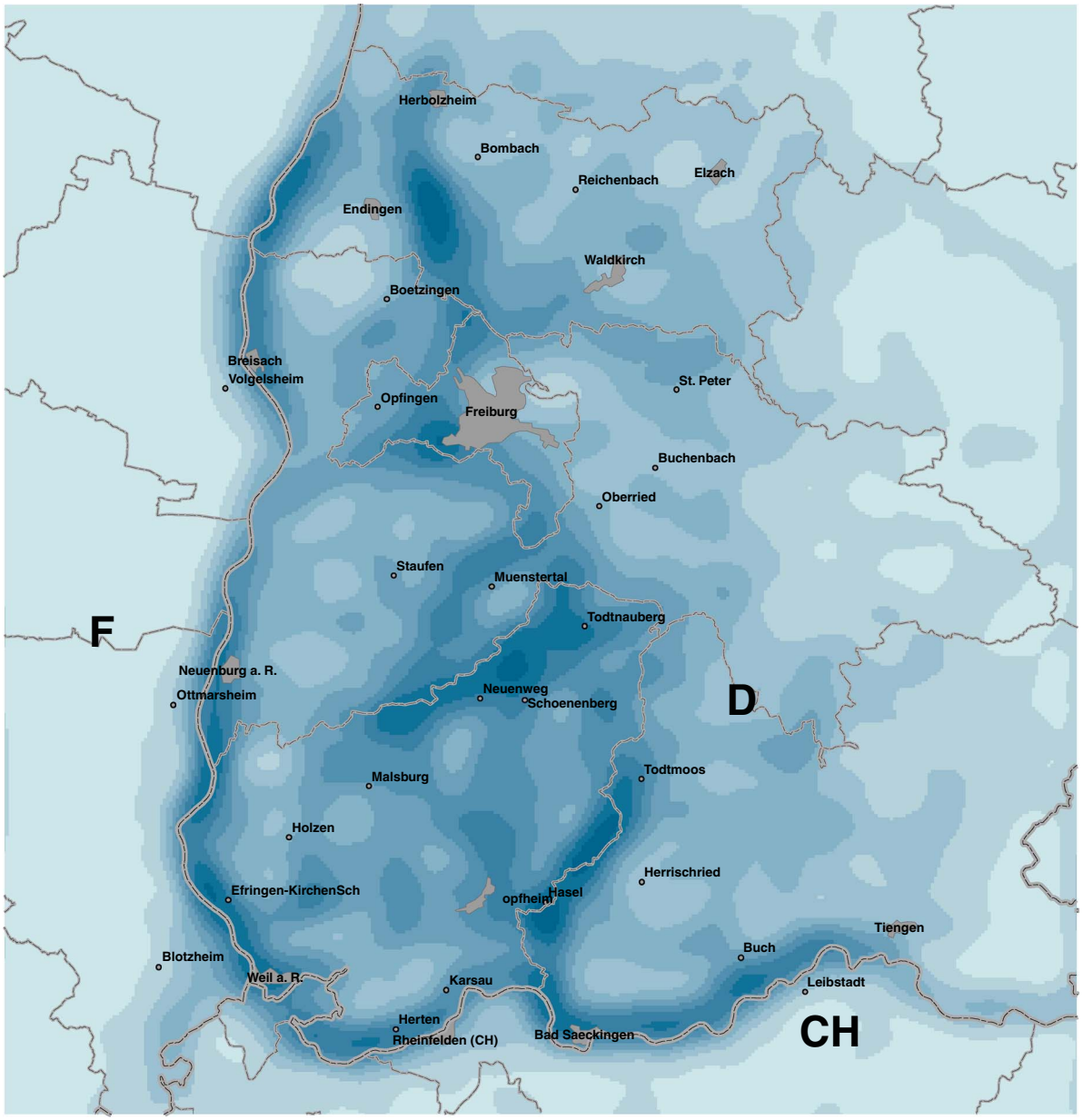
Solid blocks of color without percentage shading can also be created using the equal interval method in order to compare PD data with other raster datasets. Map 22 demonstrates this functionality, with all examples taken from data gathered as part of Study 1 indicating a Geordie [Newcastle upon Tyne] dialect area.

That a GIS divides datasets into layers means that it is very easy to change the order in which layers appear in a map projection. This is especially true when the impact of various extra-linguistic (or linguistic) factors on subjective dialect perception is considered (cf. section 4.0). It is also possible to modify the transparency of layers in the GIS in order to examine the possible effects of other factors more clearly. In Map 23, roads, places, and political borders have been placed on top of the hand-drawn maps, and transparency has been used. In this way multiple possible influences, such as the political border between Germany and France, or topography, become more apparent

4.2. Combining aggregates of individual areas on the same map

Preston (1999a:326) pioneered the approach which saw the combination of aggregate data for individual dialect areas on the same map, resulting in maps similar to that

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



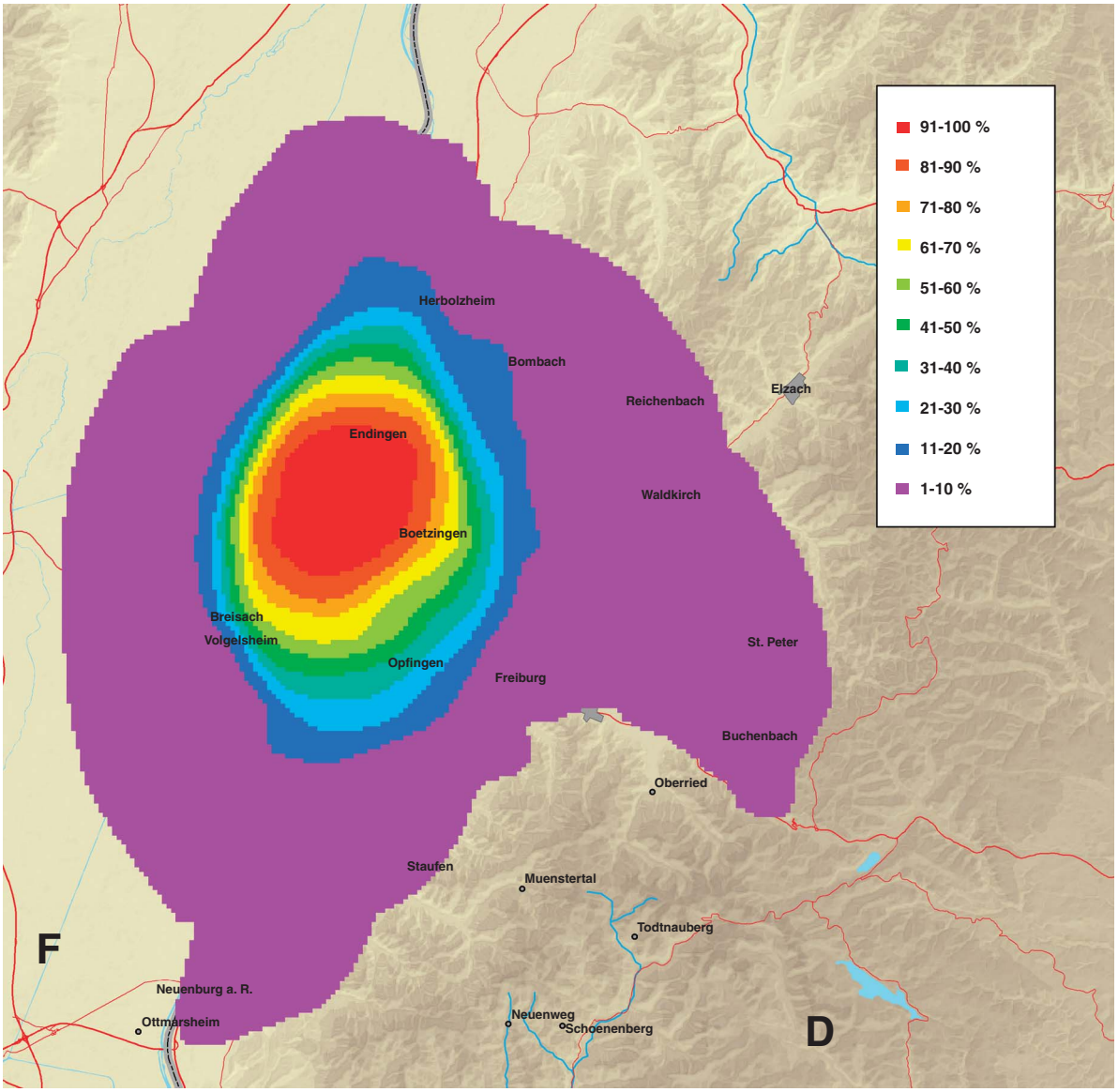
Map 20. Line density of mental dialect borders.

shown in Map 24. This approach has generally been used to display results from large-scale dialect studies, although its utility is also clear for small-scale research projects.

Such composite maps are helpful as they can be compared with other maps indicating boundaries arising from production-based studies (see Montgomery, 2007:242). They also give a useful overview of the perception of dialectal variation in a particular country (or area of a country). Hitherto, however, they have not been straightforward to create. PDQ does not easily allow the creation of such maps. Instead, in order to compile such a map the researcher must trace around the edge of an agreement level for each of the aggregate dialect areas. Each of these lines is then placed back onto a map and labeled manually. This is a relatively laborious process, and it introduces another level of potential error into the data. This is not the largest issue with the technique, but the loss of the agreement data for each of the areas is a more substantial problem. This means that for each area, the map reader is left with outline data only and as such has no idea where the perceptual "cores" of each area are to be found, nor where the lowest levels of agreement can be seen.

The GIS method we advocate here removes the need to undertake an additional stage of data processing.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 21. Rasterised version of hand-drawn maps displaying percentages.

Instead the GIS can work with all of the aggregate areas together in one map. Map 25 shows the type of map that can be achieved using this method.

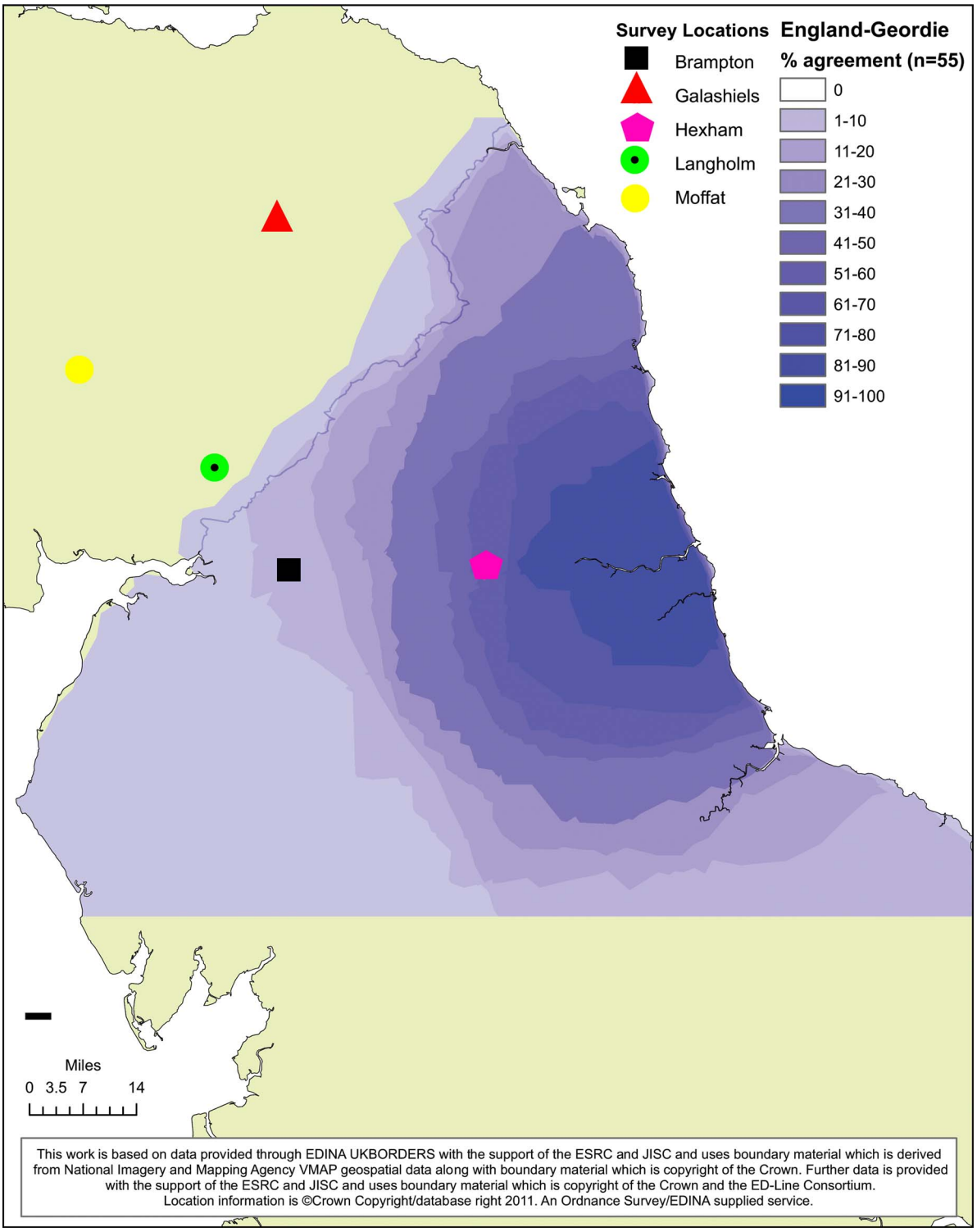
The resulting composite map loses none of the agreement data, while also permitting the display of overlapping dialect areas. In addition, the raster data generated using the method described here can be displayed alongside data held in a vector model, such as roads, political boundaries and other linear data.

5. Summary: the benefits of the use of GIS for PD study

The ability to offer improved visualization quality, to customize aggregate data, to combine individual areas on the same map, and to perform calculations and statistical analyses are all steps forward in the processing and aggregation of PD data. The use of GIS improves the quality of visualization tools available to researchers. This is a persuasive reason for us to move toward the wholesale adoption of the technology, although the way in which a GIS can work with data presents an even more appealing proposition. Thus, the ability to use the functionality of GIS technology to make PD data more comparable with that from elsewhere, as well as to subject them to all kinds of geoprocessing makes the case for using GIS very strong.

We hope to have demonstrated above that the use of GIS for processing PD data can result in a good many benefits. Although the processing techniques can be labor intensive and time consuming, they are no more

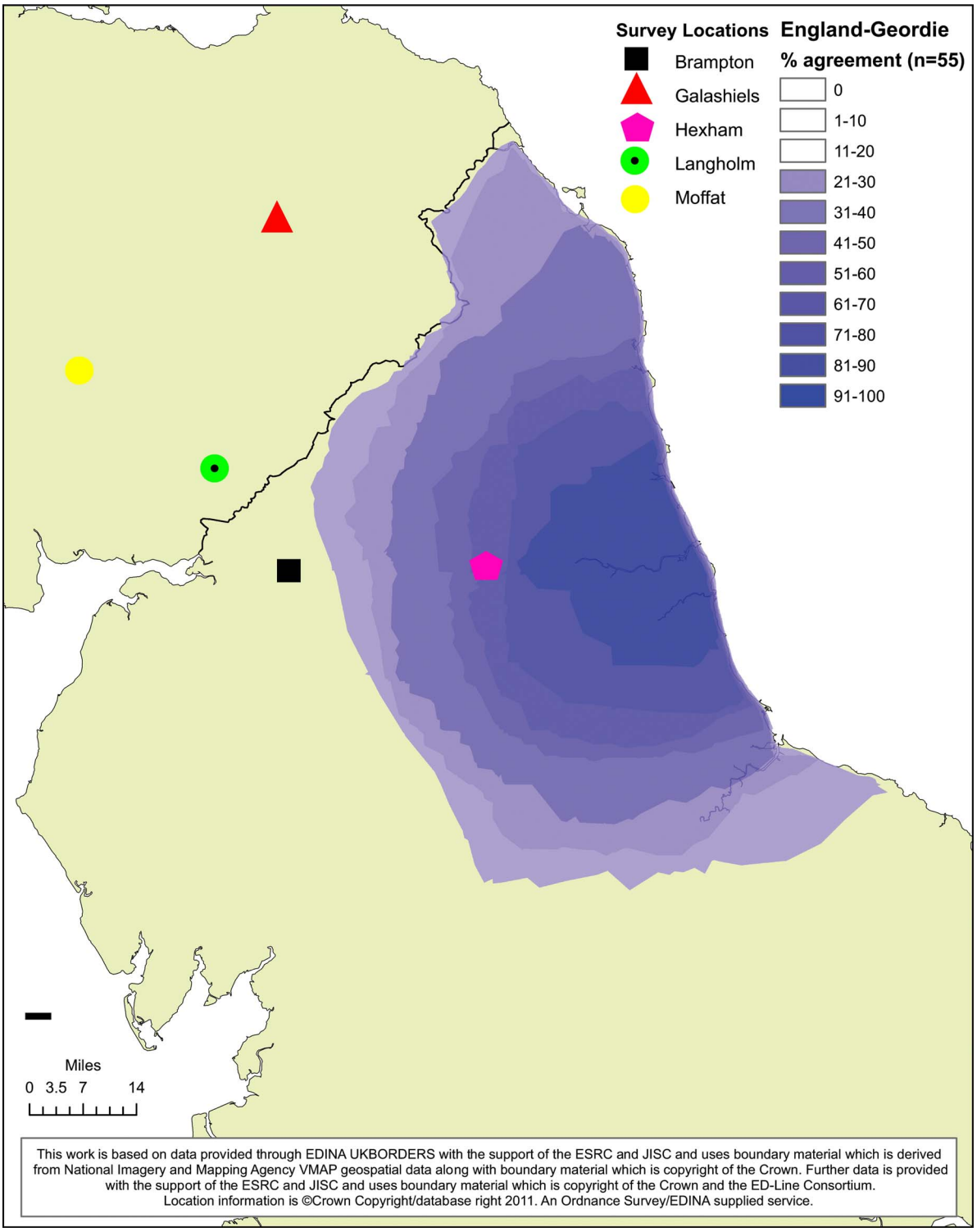
https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 22a. Differences in map display as a result of customizing aggregate data display.

so than the alternatives that have been used in the past (such as Onishi & Long, 1997). The time and effort spent processing data in a GIS is also not to be seen as an end in itself, as we have mentioned above. The ability to display PD data in a more readily accessible and visually more appealing manner is not the main benefit of the approach we outline in this article. Instead, the huge possibilities of working with PD data in a truly "spatially sensitive" (Britain, 2009:144) fashion should open up the use of this technology to

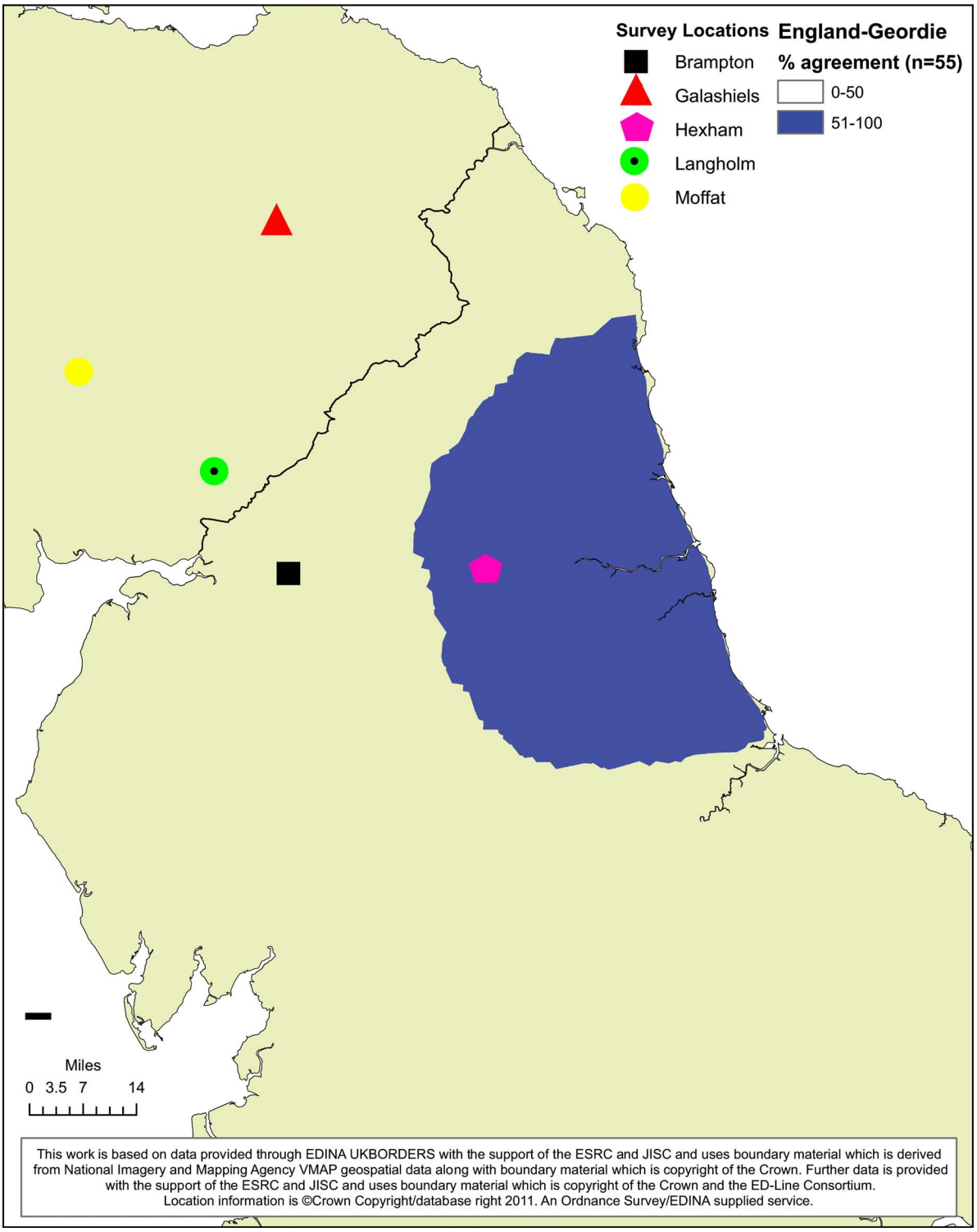
https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 22b. Continued.

others in the fields of dialectology and sociolinguistics. We urge that GIS be seen as an exciting new tool that can be used to integrate and interrogate data. In this way we echo van Hout (in Nerbonne et al., 2008), who has stated that this type of approach "opens up new vistas for doing research" by giving us "opportunities to open up, combine and integrate various rich data sources (e.g. historical, geographical, social, political, linguistic), again and again" (van Hout in Nerbonne et al., 2008:25).

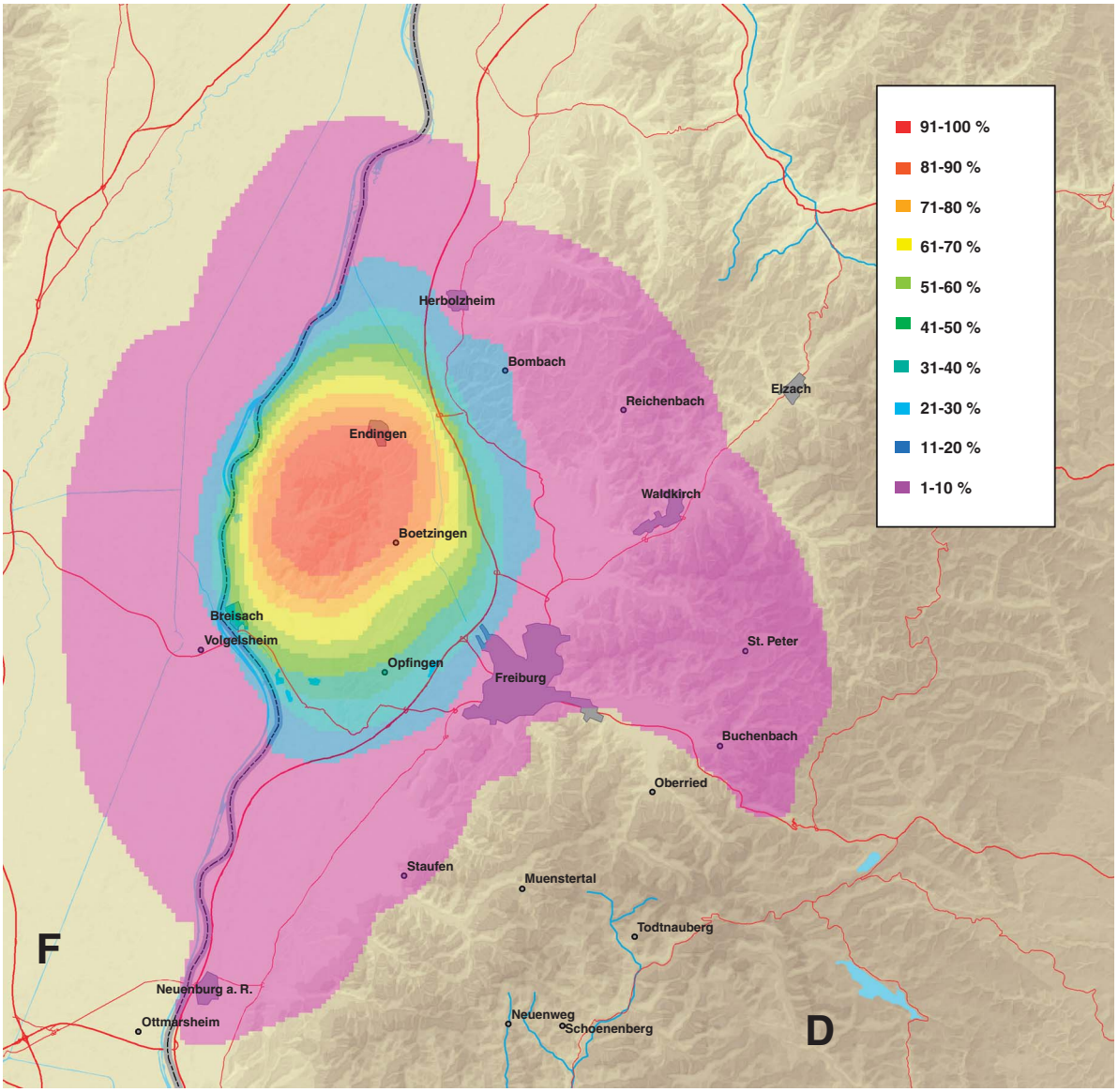
https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 22c. Continued.

The processes we have detailed above mean that the datasets created within the GIS are useable in a widely supported format, permitting further use of them by other interested parties. The use of georeferenced datasets in other areas of geolinguistics (Lameli et al., 2010) means that similarly referenced datasets from PD research can be used in conjunction with these data in order to further query data we already know well. In addition to this, the processing techniques we outline here mean that we can move beyond the static

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 23. Transparent rasterised version of hand-drawn maps.

representation of perceptions of dialect areas, and instead use the tools present within GIS programs to perform sophisticated analyses on the data. This was always the aim of Long, who adapted parts of the PDQ program to do just this (see Long, 1997), and continuing along this path should make the use of GIS essential for accessing some of the hitherto "hidden" aspects of PD data.

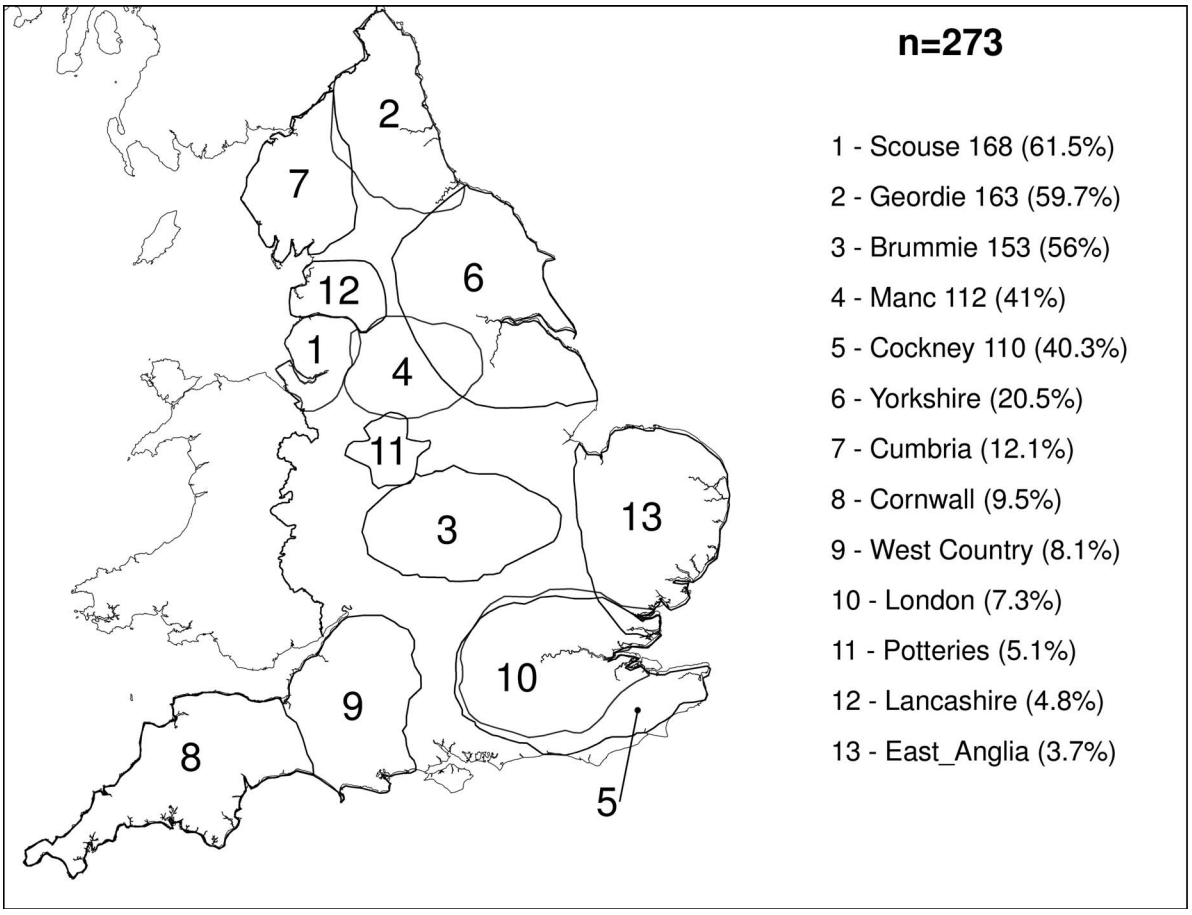
Other visualization possibilities should also not be neglected, and it seems that the ability to produce 3D animation in order to further explore PD data (see Animation 1) might have its uses. The production of change-over-time animations, which the use of GIS can facilitate, is also of clear benefit to sociolinguists and dialect geographers, as well as those involved in PD study, who wish to examine such phenomena in their data.

5.1. Possibilities of GIS for general linguistic study

Having demonstrated some of the advantages of GIS for PD research, we do not think that this is all that can be said about this technology. Although the possibilities offered by GIS may be essential for processing and analyzing hand-drawn map data, there are also many benefits for other types of linguistic research. Many of the questions and research referring to the relationship between language and space (cf. Auer & Schmidt, 2010; Lameli et al., 2010) could profit from the opportunities outlined in this paper.

Among their observations concerning the digitization of language mapping, Kehrein et al. state that mappings of linguistic data often are "subject to all kinds of limitations" (2010:xvii); that is, large parts of

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



Map 24. Composite perceptual map of England (Montgomery, 2007:237).

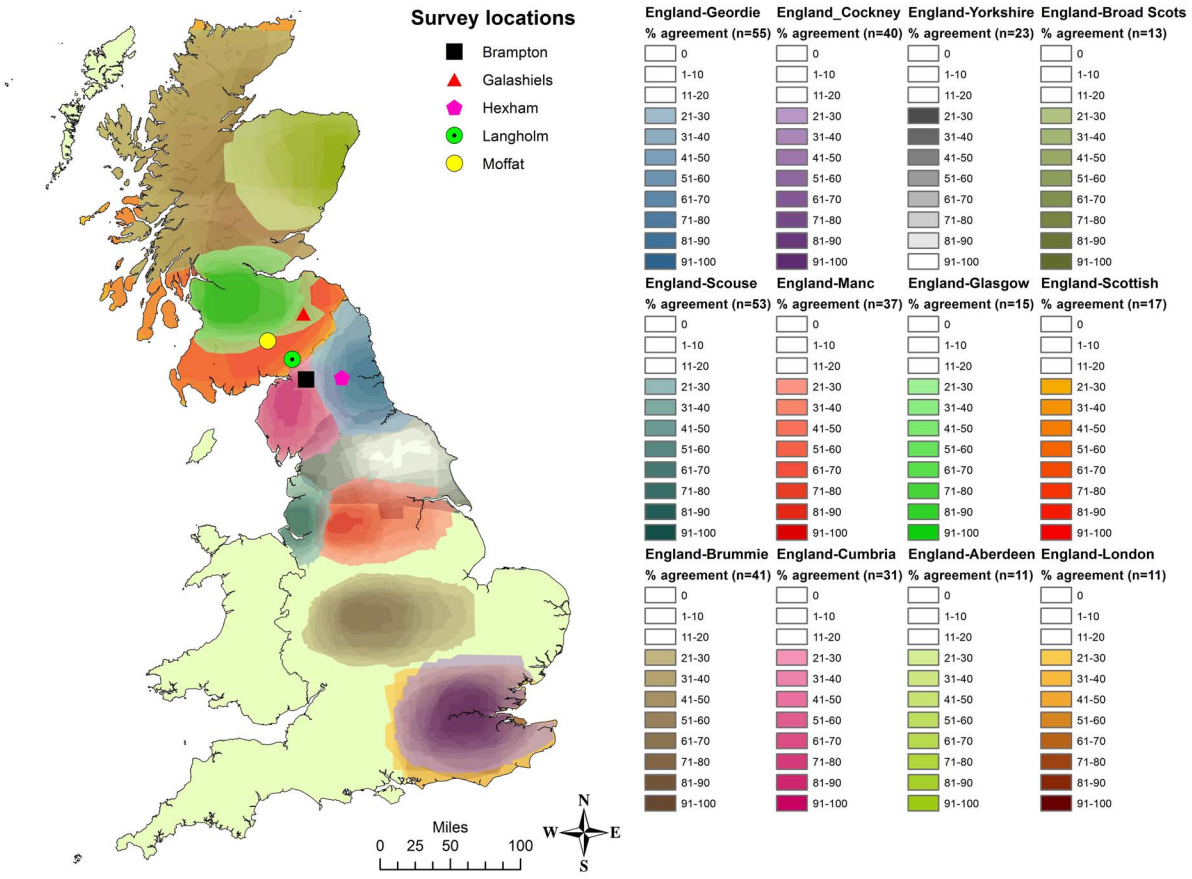
the data are not displayed and thus not accessible for other linguists. The use of GIS could contribute to overcoming this lack of information, since the outcomes of linguistic studies could be presented as datasets (cf. section 5.3) rather than just as images. Even more important seems to be another aspect which Kehrein et al. observe: "Linguistic maps are often difficult to compare because they all use their own (idiosyncratic) symbolization, map projection, scale, etc." (2010:xvii). In a GIS, all of these factors can be handled freely, which would enhance the comparability of different data.

5.2. Use of the technology: future directions

This article has focussed on PD data and the benefits of working with it in a GIS. However, we do not wish to claim that this is the only area of sociolinguistic investigation that can benefit from the use of the technology. Scholars working in neighboring disciplines, such as those who deal with questions about language and space, can also benefit greatly from the use of GIS. Georeferenced data is all that is needed for such scholars to start using the technology, and all that is required for this is the collection of postcode/ZIP code data. Once such data is captured, results of these studies can be worked within a GIS. In order for data from linguistic investigations (PD or otherwise) to be truly useful for those in other fields, gathering accurate metadata is essential. Metadata documents how spatial information has been captured and stored, which is important, since when data is captured and stored in digital form it is seldom questioned by later users (which means that metadata must be accurate and fulsome in order that later users do not compare "apples and oranges"). Accurate and complete metadata is therefore vital if linguists wish to add to the body of spatial data.

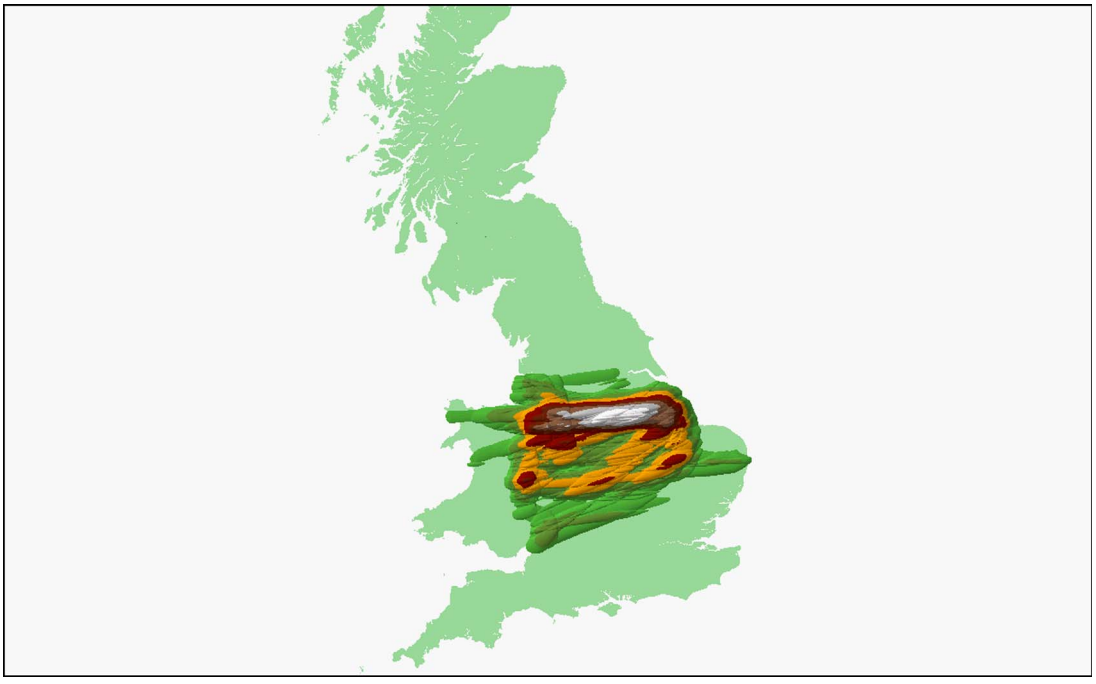
In PD, however, the use of this technology is not only helpful but instead it seems vital. Not only does it improve the quality of visualization of data, but it also permits spatial analyses of linguistic data that would not be possible with other types of computer software. Besides the gains that could be made in PD research, more extensive use of GIS by a greater number of linguists would lead to a good deal of progress in many respects. Comparable to other databases [such as

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press



This work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown and the ED-Line Consortium.
Location information is ©Crown Copyright/database right 2011.An Ordnance Survey/EDINA supplied service.

Map 25. Composite perceptual map of Great Britain, showing aggregated dialect areas drawn by English respondents.



Animation 1. Fly-over of a 3D representation of North-South dividing lines drawn by respondents from English locations in Study 1.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

the "Archiv für Gesprochenes Deutsch" [Archive for spoken German] (Institut für Deutsche Sprache, 2011), the Digital Wenker atlas (Lameli et al., 2010), and the Linguistic Atlas Projects web pages (Kretzschmar, 2005)], data and outcomes from studies in PD could be available for other linguists. As we have argued, they could also be compared with and merged very easily with other datasets, be they linguistic or nonlinguistic. Moreover, like any other kinds of statistical data published on the web (e.g. population density, demographic factors, education, etc.) linguistic data could make up databases available for other linguists, but also accessible for the interested public (cf. Lameli et al., 2010; Evans, 2011).

As GIS is used in many fields, it is subject to constant development and improvement. More users dealing with linguistic topics would promote academic exchange and lead to more ideas, more forums, and more progress in answering questions related to language and space. Kehrein et al. predict that the connection between linguistics and GIS "will be of increasing importance in the coming years" (2010:xviii). We hope to have established some of the most important uses of GIS in PD and delivered some of the decisive arguments for the use of GIS.

Notes

[1] Trace-and-overlay techniques can be useful for "quick and dirty" analyses, and should not be dismissed out of hand as they can be instructive as to the general patterning of perceptual areas. In such a technique, lines are compiled using an overhead transparency onto which can be traced all instances of a particular dialect area. The same can be done by scanning maps and manually overlaying them in a graphics program. Producing very detailed composite maps using this type of technique is, however, almost impossible, as is working with data from more than a limited sample (around thirty respondents). Therefore, a trace-and-overlay technique should only be used for small-scale or preliminary studies, or where the aim is to find broad general patterns from a limited cohort.

[2] The research in Study 1 was funded by the Economic and Social Research Council, Grant number PTA-026-27-1956.

[3] The survey was part of a larger project called "Regional Dialects in the Alemannic Border Triangle" (together with Sandra Hansen). The investigation aims at analysing dialectal variation from both linguistic and folk perspectives and to combine the outcomes of the two approaches.

[4] The question of how different types of information (such as cities, rivers, administrative boundaries, etc.) may influence the outcome of the draw-a-map task is addressed by Lameli et al. (2008) and shall not be discussed here. In this case, the decision to include the city location dots was made to ensure that respondents' geographical knowledge was consistent and the spatial data they provided could be treated as accurate (cf. Preston, 1993:335). Further details relating to this methodological decision can be found in Montgomery (2007).

[5] This question was included in order to increase the likelihood of respondents completing the draw-a-map task. As Montgomery (2007:71–73) has discussed elsewhere, the inclusion of such dots increases the number of respondents willing to draw lines on the map, as it reduces the chance of getting the geography of the country "wrong."

[6] A question relating to the "north-south" divide was included as it is an important concept in the United Kingdom (although it is perhaps of most importance in England). Barely a month goes by without media outlets reporting on the existence of the divide (or its "widening" or "shrinking") (e.g. Wachman, 2011). In this sense, the concept is convenient shorthand for a complex situation. Although often thought of as a modern or recent concept, Jewell has stated that it is "literally, as old as the hills" (Jewell, 1994:28). The preoccupation with a countrywide "divide" is perhaps not as surprising as one might think, as implicit or explicit contrasts have been shown to be important in creating a sense of "social self" (Cohen, 1985:115). Despite this, the divide is not an official boundary and, as such, there is a great deal of disagreement about where the dividing line falls (Montgomery, 2007:1–4). This question was included for the reason that the north-south divide is: (a) consistently mentioned, (b) a persistent concept, (c) potentially important for a sense of "social self," and (d) undefined.

[7] All interviews were attended by at least one of the researchers, which made it possible to resolve confusions concerning the task immediately.

[8] There are various other pieces of GIS software, such as MapInfo (MapInfo Corporation, 2011). Some GIS platforms have a free license [such as Quantum GIS (QGIS, 2011) and GRASS GIS (GRASS Development Team, 2011)].

[9] General introductions to GIS can be found in Gomarasca (2009) or Wise (2002). Moreover, there are individual information sites and tutorials for different GIS software providers (such as ESRI, 2011a; GRASS Development Team, 2011; QGIS, 2011).

[10] It should be noted that three control points are the minimum required to georeference an image. More can be used in order to improve accuracy and best practice dictates that four or more control points should be used in digitizing or georeferencing. Ideally numerous control points spread out within the area of interest should be identified using discrete (unambiguous) locations (such as borders).

[11] It is worth noting here that the red coloring of the area is totally at random and that the visualization, as will be shown in section 4.2, can be performed at will.

[12] This is especially true if many nodes—as in our case, where large numbers of polygons are combined in one dataset – are digitized. Simple vector datasets containing few nodes, however, require less memory than comparable raster datasets.

[13] If following this process, make sure to use the frequency count given by the use of the "Frequency" tool as the value field for the raster.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

[14] It has to be mentioned here that all modifications should be carried out thoughtfully. While smoothing the sharp edges makes the map "cleaner" looking, this technique can also distort the data.

[15] It should be noted that this is not the only way of producing aggregate maps in GIS. For example, it is also possible to convert each single hand-drawn map into a raster dataset and then calculate the sum of all datasets. Since with this method data queries are much more laborious (step 7/8), we follow the scheme presented here.

[16] Known as "choroplethic" values in the GIS literature, but termed "percentages" here because of the familiarity of this concept in sociolinguistics.

References

Anders, Christina A. 2010. *Wahrnehmungsdialektologie: Das Obersächsische im Alltagsverständnis von Laien*. Berlin: De Gruyter Mouton.

Auer, Peter & Jürgen E. Schmidt. 2010. *Language and space: An international handbook of linguistic variation, vol. 1: Theories and methods*. Berlin: De Gruyter Mouton.

Britain, David. 2002. Space and spatial diffusion. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (eds), *The handbook of language variation and change*, 603–637. Oxford: Blackwell.

Britain, David. 2009. Language and space: The variationist approach. In P. Auer & J. E. Schmidt (eds), *Language and space, an international handbook of linguistic variation*, 142–162. Berlin: Mouton de Gruyter.

Britain, David. 2010. Conceptualisations of geographic space in linguistics. In A. Lameli, R. Kehrein & S. Rabanus (eds), *Language and space, an international handbook of linguistic variation, vol. 2: Language mapping*, 69–97. Berlin: Mouton de Gruyter.

Bucholtz, Mary, Nancy Bermudez, Victor Fung, Lisa Edwards & Rosalva Vargas. 2007. Hella nor Cal or totally so Cal? The perceptual dialectology of California. *Journal of English Linguistics* 35: 325–352.

Bucholtz, Mary, Nancy Bermudez, Victor Fung, Rosalva Vargas & Lisa Edwards. 2008. The normative North and the stigmatized South: Ideology and methodology in the perceptual dialectology of California. *Journal of English Linguistics* 36: 62–87.

Bundesamt für Kartographie und Geodäsie. 2011. GeoDatenZentrum. Bundesamt für Kartographie und Geodäsie. http://www.bkg.bund.de/ (6 November, 2011).

Butters, Ronald R. 1991. Dennis Preston, perceptual dialectology. *Language in Society* 20: 294–299.

Cohen, Anthony. 1985. *The symbolic construction of community*. London: Routledge.

Cukor-Avila, Patricia, Lisa Jeon, Patrícia C. Rector, Chetan Tiwari & Zak Shelton. 2012. Texas – It's like a whole nuther country: Mapping Texans' perceptions of dialect variation in the Lone Star state. In *Proceedings from the Twentieth Annual Symposium about Language and Society*, Austin, TX: Texas Linguistics Forum.

Edina. 2011. Digimap collections. http://edina.ac.uk/digimap/ (28 September, 2011).

ESRI. 2011a. ArcGIS Resource Center. http://resources.arcgis.com/content/web-based-help (10 November, 2011).

ESRI. 2011b. What is GIS? GIS.com. http://www.gis.com/content/what-gis (12 October, 2011).

Evans, Betsy E. 2011. Seattle to Spokane: Mapping English in Washington state. http://depts.washington.edu/folkling/. (12 October, 2011).

Garrett, Peter. 2010. *Attitudes to language*. Cambridge: Cambridge University Press.

Geofabrik. 2011. Geofabrik. www.geofabrik.de (2 July, 2012).

Gomarasca, Marion A. 2009. *Basics of geomatics*. London and New York: Springer Dordrecht Heidelberg.

Goodey, Brian. 1971. *City scene: An exploration into the image of central Birmingham as seen by area residents*. Birmingham: Centre for Urban and Regional Studies, University of Birmingham.

Gould, Peter & Rodney White. 1986. *Mental maps*, 2nd edn. Boston: Allen & Unwin.

GRASS Development Team. 2011. GRASS GIS. http://grass.osgeo.org/ (12 October, 2011).

Grootaers, Willem A. 1999. The discussion surrounding the subjective boundaries of dialects. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: 115–129. Amsterdam: John Benjamins.

Heywood, Ian, Sarah Cornelius & Steve Carver. 2006. *An introduction to geographical information systems*, 3rd edn. Harlow, UK: Pearson Prentice Hall.

HM Government. 2011. Opening up government. http://data.gov.uk/ (28 September, 2011).

Institut für Deutsche Sprache. 2011. Archiv für Gesprochenes Deutsch. http://agd.ids-mannheim.de/ (6 November, 2011).

Jeon, Lisa & Patricia Cukor-Avila. 2012. Urbanicity and language variation and change: Mapping dialect perceptions in and of Seoul. *Sociolinguistic Symposium 19*, Berlin, Germany.

Jewell, Helen M. 1994. *The north-south divide: The origins of northern consciousness in England*. Manchester: Manchester University Press.

Kehrein, Roland, Lameli Alfred & Rabanus Stefan. 2011. Introduction, xi–xvii. In A. Lameli, R. Kehrein & S. Rabanus (eds), *Language and Space – An International Handbook of Linguistic Variation*, Volume 2: Language Mapping. Berlin: Mouton de Gruyter.

Kerswill, Paul. 2003. Dialect levelling and geographical diffusion in British English. In D. Britain & J. Cheshire (eds), *Social dialectology: In honour of Peter Trudgill*, 223–243. Amsterdam: John Benjamins.

Kinchen, Rosie. 2011. Howay! Youths adopt hip accents. *The Sunday Times*, p. 11.

Kirk, John M. & William A. Kretzschmar. 1992. Interactive linguistic mapping of dialect features. *Literary and Linguistic Computing* 7: 168–175.

Kretzschmar, William A. 1999. Preface. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: xvii–xviii. Amsterdam: John Benjamins.

Kretzschmar, William A. 2005. Linguistic atlas projects. http://us.english.uga.edu/cgi-bin/lapsite.fcgi/ (10August, 2011).

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English: Phonetics, phonology, and sound change: A multimedia reference tool*. Berlin: Walter de Gruyter.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

Lameli, Alfred, Tanja Giessler, Roland Kehrein, Alexandra Lenz, Karl-Heinz Müller, Jost Nickel, Christoph Purschke & Stefan Rabanus. 2010. DiWA. Digital Wenker Atlas. http://www.diwa.info/ (10 August, 2011).

Lameli, Alfred, Roland Kehrein & Stefan Rabanus. 2010. *Language and space, an international handbook of linguistic variation, vol. 2: Language mapping*. Berlin: De Gruyter Mouton.

Lameli, Alfred, Christoph Purschke & Roland Kehrein. 2008. Stimulus und Kognition. Zur Aktivierung mentaler Raumbilder. *Linguistik Online* 35(3): 55–86.

Llamas, Carmen. 2000. Middlesbrough English: Convergent and divergent trends in a "part of Britain with no identity". *Leeds Working Papers in Linguistics and Phonetics* 8: 123–148.

Long, Daniel. 1997. The perception of "standard" as the speech variety of a specific region: Computer-produced maps of perceptual dialect regions. In Thomas, A. R. (ed.), *Issues and methods in dialectology*, 256–270. Bangor: University of Wales Press.

Long, Daniel. 1999. Geographical perception of Japanese dialect regions. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: 177–198. Amsterdam: John Benjamins.

Lynch, Kevin. 1960. *The image of the city*. Cambridge, MA: MIT Press.

MapInfo Corporation. 2011. MapInfo. http://www.mapinfo.com/ (12 October, 2011).

Mase, Yoshio. 1999. Dialect consciousness and dialect divisions: Examples in the Nagano-Gifu boundary region. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: 71–99. Amsterdam: Benjamins.

McKinnie, Meghan & Jennifer Dailey-O'Cain. 2002. A perceptual dialectology of Anglophone Canada from the perspective of young Albertans and Ontarians. In D. Long & D. R. Preston (eds), *Handbook of perceptual dialectology* 2: 277–294. Amsterdam: John Benjamins.

Montgomery, Chris. 2007. *Northern English dialects: A perceptual approach*. Sheffield: University of Sheffield dissertation.

Montgomery, Chris. 2011. Perceptual dialectology, bespoke methods and the future for data processing. *8th United Kingdom Language Variation and Change Conference* (UKLVC8), Edge Hill University.

Montgomery, Chris. 2012. The effect of proximity in perceptual dialectology. *Journal of Sociolinguistics* 16(5): 638–668.

Nash Parker, Robert & Emily K. Asencio. 2008. *GIS and spatial analysis for the social sciences: Coding, mapping and modelling*. London: Routledge.

Nerbonne, John, Paul Heggarty, Roeland van Hout & David Robey. 2008. Panel discussion on computing and the humanities. In J. Nerbonne, C. Gooskens, S. Kürschner & R. van Bezooijen (eds), *International journal of humanities and arts computing, special issue on computing and language variation*. 2: 243–259. Edinburgh: Edinburgh University Press.

Onishi, Isao & Daniel Long. 1997. Perceptual dialectology quantifier (PDQ) for Windows. http://nihongo.hum.tmu.ac.jp/~long/maps/perceptmaps.htm (2 July, 2012).

Ordnance Survey. 2011. OS open data. http://www.ordnancesurvey.co.uk/oswebsite/products/os-opendata.html (28 September, 2011).

Orleans, Peter. 1967. Differential cognition of urban residents: Effects of social scale on mapping. In J. G. Truxal (ed.), *Science, engineering, and the city*, 103–117. Washington, DC: National Academy of Sciences.

Preston, Dennis R. 1982. Perceptual dialectology: Mental maps of United States dialects from a Hawaiian perspective. *Hawaii Working Papers in Linguistics* 14: 5–49.

Preston, Dennis R. 1986. Five visions of America. *Language in Society* 15: 221–240.

Preston, Dennis R. 1988. Change in the perception of language varieties. In J. Fisiak (ed.), *Historical dialectology: Regional and social*, 475–504. Berlin: Mouton de Gruyter.

Preston, Dennis R. 1993. Folk dialectology. In D. R. Preston (ed.), *American dialect research*, 333–377. Amsterdam: John Benjamins.

Preston, Dennis R. 1999a. A language attitude approach to the perception of regional variety. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: 359–375. Amsterdam: John Benjamins.

Preston, Dennis R. 1999b. Introduction. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: xxiii–xxxix. Amsterdam: John Benjamins.

Preston, Dennis R. 2010. Perceptual dialectology in the 21st century. In C. A Anders, M. Hundt & A. Lasch (eds), *Perceptual dialectology. Neue Wege der Dialektologie (Linguistik – Impulse & Tendenzen)*, 1–30. Berlin: Mouton de Gruyter.

Preston, Dennis R & George M Howe. 1987. Computerized studies of mental dialect maps. In K. M. Denning, S. Inkelas, F. McNair-Knox & J. Rickford (eds), *Variation in language: NWAV-XV at Stanford* (*Proceedings of the Fifteenth Annual Conference on New Ways of Analyzing Variation*). 361–378. Stanford, CA: Department of Linguistics, Stanford University.

Purschke, Christoph. 2011. Regional linguistic knowledge and perception: On the conceptualization of Hessian. *Dialectologia* 2, Special Issue, 91–118.

QGIS. 2011. Quantum GIS: Welcome to the Quantum GIS project. http://www.qgis.org/ (12 October, 2011).

Saurer, Helmut & Franz-Joseph Behr. 1997. *Geographische Informationssysteme*. Eine Einführung, Darmstadt: Wissenschaftliche Buchgesellschaft.

Sibata, Takesi. 1999. Consciousness of dialect boundaries. In D. R. Preston (ed.), *Handbook of perceptual dialectology* 1: 39–63. Amsterdam: John Benjamins.

Stoeckle, Philipp. 2010. Subjektive Dialektgrenzen im Alemannischen Dreiländereck. In C. A. Anders, M. Hundt & A. Lasch (eds), *Perceptual dialectology – Neue Wege der Dialektologie*, 291–315. Berlin: Mouton de Gruyter.

Stoeckle, Philipp. 2011. The constitution of subjective dialect areas: Towards a hierarchisation of lay classification strategies. *6th International Conference on Language Variation in Europe*, Freiburg, Germany.

Stoeckle, Philipp. 2012. The folk-linguistic construction of local dialect areas – linguistic and extra-linguistic factors. In S. Hansen, C. Schwarz, P. Stoeckle & T. Streck (eds), *Dialectological and folk dialectological concepts of space. Current methods and perspectives in sociolinguistic research on dialect change*, 142–163, 271–275. Berlin: Mouton de Gruyter.

The Economist. 2011. England's regional accents: Geordie's still alreet. http://www.economist.com/node/18775029 (7 September, 2011).

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press

Trudgill, Peter. 1999. *The dialects of England*, 2nd edn. Oxford: Blackwell.
U.K. Data Archive. 2011. Census.ac.uk. http://www.census.ac.uk/ (28 September, 2011).
Wachman, Richard. 2011. North-South divide widens as public sector cuts hit businesses. *The Guardian*. http://www.guardian.co.uk/business/2011/oct/17/north-south-divide-widens-cuts (2 July, 2012).
Weijnen, Antonius A. 1946. De Grenzen Tussen de Oost-Noordbrabantse Dialecten Onderin. In A. A. Weijnen, M. Renders & J. van Guineken (eds), *Oost-Noordbrabantse Dialectproblernen*, 1–15. Bijdragen en Mededelingen der Dialecrencommissie van de Koninkjke Nederlandse Akademie van Wetenschappen te Amsterdam VII. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.
Wise, Stephen. 2002. *GIS basics*. London: Taylor & Francis.

https://doi.org/10.1017/jlg.2013.4 Published online by Cambridge University Press