

Full title:

A perceptual dialectological approach to linguistic variation and spatial analysis of Kurdish varieties

Main Author:

Eva Eppler, PhD, RCSLT, Mag. Phil

Reader/Associate Professor in Linguistics

Department of Media, Culture and Language

University of Roehampton | London | SW15 5SL

e.eppler@roehampton.ac.uk | www.roehampton.ac.uk

Tel: +44 (0) 20 8392 3791

Co-author:

Josef Benedikt, PhD, Mag.rer.nat.

Independent Scholar, Senior GIS Researcher

GeoLogic Dr. Benedikt

Roergasse 11/18

1090 Vienna, Austria

josef.benedikt@geologic.at | www.geologic.at

Short Title:

Perceptual Dialectology and GIS in Kurdish

Abstract:

This paper presents results of a first investigation into Kurdish linguistic varieties and their spatial distribution. Kurdish dialects are used across five nation states in the Middle East and only one, Sorani, has official status in one of them. The study employs the ‘draw-a-map task’ established in Perceptual Dialectology; the analysis is supported by Geographical Information Systems (GIS). The results show that, despite the geolinguistic and geopolitical situation, Kurdish respondents have good knowledge of the main varieties of their language (Kurmanji, Sorani and the related variety Zazaki) and where to localize them. Awareness of the more diverse Southern Kurdish varieties is less definitive. This indicates that the Kurdish language plays a role in identity formation, but also that smaller isolated varieties are not only endangered in terms of speakers, but also in terms of their representations in Kurds’ mental maps of the linguistic landscape they live in.

Acknowledgments:

This work was supported by a Santander and by Ede & Ravenscroft Research grant 2016.

The authors would like to thank Yüksel Gonul and Mehmet Duran for collecting field data, and

Christina Membrive Altés and Philipp Kalchbauer for helping to establish a GIS database.

We furthermore want to acknowledge that we are aware of a detailed small-scale perceptual dialectological investigation of northern Kurdish in Iran conducted by Asadpour (2011).

## 1. INTRODUCTION

This paper presents the results of the first perceptual dialectological analysis of linguistic variation in Kurdish supported by Geographical Information Systems (henceforth GIS). Kurdish is a bundle of closely related northwest Iranian varieties with approximately 30 million native speakers in southeast Turkey, southwest Armenia, northwest Iran, north Iraq, northeast Syria and the diaspora. Kurdish people's perception of linguistic variation will be investigated using the "draw-a-map" task, one of a range of methods from Perceptual Dialectology (henceforth PD, Preston, 1982, 1998, 2010). PD is the cross-disciplinary study of lay people's perception of linguistic variation. It aims to capture non-linguists' mental representations of linguistic variation associated with geographical space. The spatial features of the collected PD data are stored, processed, analysed and displayed in GIS.

Kurdish is under-researched both in linguistic terms and in terms of its spatial distribution, despite its geolinguistic importance. Its classification into varieties is disputed (see section 2 and Haig & Öpengin, 2014), and dialect maps of Kurdish overlap in some, but not in other areas (Belelli, 2016; Fattah, 2000; Hassanpour, 1992; Izady, 2014; Öpengin & Haig, 2014; Sheyholislami, 2015). Research is complicated by the fact that varieties of Kurdish are spoken in five different nation states (Turkey, Iraq, Iran, Syria and Armenia). Only one variety (Sorani) has official language status at the national level in Iraq, while another variety (Kurmanji) is an official minority language in Armenia. What is completely missing from the picture about Kurdish so far is: which of the varieties of their language Kurds are familiar with, and what they know about where these varieties are spoken. The current paper aims to fill this gap. The results can impact on various different areas, such as identity building, language planning, ethnolinguistic vitality and traditional dialectology.

The results of this study will provide information about Kurdish people's mental maps of the varieties of their language and where they are spoken. This information will contribute to answering the question of whether the Kurdish linguistic area also exists in its speakers' minds, and how it is

spatially distributed. The conditions under which the Kurdish speech community exists are very different from those of better-known languages for which PD research has been conducted. The study presented in this paper therefore constitutes an interesting test case for assessing how much of this kind of knowledge is available to a speech community that lacks almost any kind of institutionalized / state-sanctioned dissemination of knowledge regarding the varieties, the linguistic areas in which they are spoken and the speech community in question. This information can then be used to evaluate whether the Kurdish language can contribute to identity building.

As only one variety of Kurdish (Kurmanji) had been developed into a fully functioning medium for education, printing and broadcasting in Soviet Armenia in the early 1920s and 1930s, some of the numerically bigger varieties of Kurdish are currently undergoing language planning processes. The results of this study can also inform decisions on which varieties language planning efforts might want to focus on, e.g. those Kurds are most familiar with. Some Kurdish scholars and institutions (e.g. Institut kurde de Paris, Kurdish Academy of Language) are furthermore debating the question whether one variety of Kurdish should be developed into a trans-national standard, and if yes, which one. The findings of this study can also inform this process by showing whether the linguistic diversity of Kurdish seems to pose a problem for Kurds.

Two related varieties of Kurdish are on UNESCO's list of endangered languages (Gorani as "definitely endangered", and Zazaki as "vulnerable", Moseley 2010). Another contribution this study can make is to show whether this status is reflected in Kurds' mental representations of the varieties of their language, and which of the factors known to influence a variety's ethnolinguistic vitality may contribute to the vulnerable/endangered statuses. If some varieties are also weakly presented in Kurds' mental representations of their language, there is even more reason to support efforts to maintain these varieties.

Digital spatial representations of Kurdish people's mental maps of the varieties of their language and where they are spoken can furthermore be compared with maps produced through

traditional dialectological methods (Hassanpour, 1992; Izady, 1992; Öpengin & Haig, 2014). This can not only yield useful comparisons and challenge assumptions made in both research traditions (Butters, 1991), but also encourage traditional and perceptual dialectologists to look afresh at their results (Montgomery & Stoeckle, 2013). With these wide potential applications, the paper addresses the following research questions:

1. Which varieties of their language are Kurds familiar with?
2. Where do they think these varieties are spoken?
3. How do perceptual dialectology boundaries pattern with traditional dialect boundaries, i.e. bundles of isoglosses?
4. How do non-linguists' perceptions of linguistic variation pattern with non-linguistic facts, such as nationality, demographics and related statistics?

To address these questions, we have used the most established method for the study of non-linguists' conceptions of linguistic variation, the "draw-a-map" task (Preston, 1982) of PD. For the analysis we employ GIS because it provides a set of useful tools to collate, visualize and analyse spatial along with non-spatial data.

The paper is structured as follows. Section 2 introduces the object of study, the different varieties of Kurdish. Section 3 outlines methodological aspects of PD, motivates our choice of GIS as a primary analytical toolset, and describes the participants and original data. The results of Kurdish people's perception of linguistic variation in their homeland are presented in Section 4. Section 5 summarises our findings and addresses potential limitations of the study. Section 6 presents some ideas for future research and discusses some implications of our findings.

## **2. KURDISH**

The 'Kurdish language' is a continuum of related northwest Iranian dialects spoken across a large contiguous area in the Middle East (see Section 3 for more detail). Most of the Kurdish speaking areas belong to four different nation states: Turkey, Iraq, Iran and Syria. The most sizable

communities of Kurdish speakers live in Turkey (approximately<sup>1</sup> 10-15 million), Iraq (approximately 6-7 million), and Iran (over 3 million); smaller communities of Kurds live in Syria (approximately 2 million), Armenia and other Caucasus and Central Asia republics in addition to Lebanon. With a number of Kurdish speakers living outside the Kurdish homeland, the total number of Kurdish speakers is estimated at over 30 million.

Because the Kurds lack a state of their own, Kurdish is not a national language. Currently only Sorani has official language status in Iraq, and Kurmanji in the self-declared autonomous Jazira Canton in Syria. Bahdini, a northern variety related to Kurmanji, has some official status in Iran. Kurmanji is also a recognized minority language in Armenia. As a consequence of this, there is no unitary normative Kurdish standard. This, in turn, complicates the question of which varieties belong to Kurdish.

Chambers and Trudgill (1998:9) propose that, from a sociolinguistic point of view, certain varieties of dialect continuum x are dialects of x (while others are dialects of y), because of the relationship these dialects bear to their respective standard languages. The formulae that speakers of x dialects are speaking x, that they read and write in x, that any standardizing changes in their dialects will be towards x, and that they in general look to x as the standard language which naturally corresponds to their vernacular varieties (ibid.) cannot easily be adapted to Kurdish, because there is no unified writing system (different varieties of Kurdish are written in different scripts)<sup>2</sup>, changes in some varieties seem to be diffusing rather than standardizing (Belelli, 2016; Matras, 2016; Öpengin & Haig, 2014), and there is no unitary standard to look to<sup>3</sup>. A precise demarcation of the Kurdish language is thus not a simple matter from a sociolinguistic point of view.

It is also not a simple matter from a linguistic point of view. Based on lexical, phonological and morphosyntactic data, Kurdish can be considered a superordinate unit with three to five main

dialect groups (cf. Belelli, 2016; Haig & Öpengin, 2014; Hassanpour, 1992; MacKenzie, 1961; see Figure 1).

1. Northern Kurdish varieties are spoken in Turkey, Iran and Iraq. They are generally referred to as Kurmanji and Bahdini (in Iraq); both Kurmanji and Bahdini include a number of other regional dialects (Öpengin & Haig, 2014).
2. Central Kurdish varieties are spoken in a large area on both sides of the Iraqi-Iran border. Like all varieties of Kurdish, Central Kurdish is internally diversified. The main variety is Sorani; major sub-varieties include Mukri and Sineyê (mainly spoken in the Mahabad and Sanandaj regions of Iran, respectively) and Hewlêrî, Silêmanî and Germiyanî (in the Erbil, Suleymania and Kirkuk regions of Iraq).
3. Southern Kurdish varieties are also spoken on both sides of the Iraqi-Iran border, but mainly in Iran. Southern Kurdish is potentially the most diverse grouping<sup>4</sup> and includes the varieties Bijari in the Northeast, Kelhori, Kolyai, and Kermanshahi northeast of the Zagros mountain range, and Malekshahi/Feyli, Badrei and Kordali southeast of the Zagros range.

Zazaki and Gorani are considered as distinct from Kurdish by some researchers (e.g. Belelli, 2016, Haig & Matras, 2002; MacKenzie, 1961); other scholars (Fattah, 2000; Izady, 1992) propose a separate Kurdo-Caspian group for them. Hassanpour (1992) and Haig & Öpengin (2014:110) suggest avoiding sub-grouping Zazaki and Gorani (at least until positive evidence in favour of such a move is forthcoming). For this reason and a reason outlined later, we include Zazaki and Gorani as varieties of Kurdish in our study.

4. Zazaki, also referred to as Dimili, Kirdkî and Kirmanjkî, is only spoken in Turkey. It is usually categorized into Northern, Central and Southern Zazaki.
5. Gorani is predominantly spoken in the Kurdish areas of Iran and covers what is known as Hawram(an)i, including the dialects of Paveh and Halabja among others. Several varieties spoken in present day Iraq, e.g. Sarli and Shabak in the north and Bajalani in the south, are also frequently

grouped with Gorani (cf. Fattah, 2000:62-70; Mahmoudveysi & Bailey, 2013 for discussion of Gorani).

Like Belelli (2016), Hassanpour (1992), Izadi (1992), Fattah (2000) and Haig & Öpengin (2014) we do not include Laki<sup>5</sup>(another northwest Iranian variety), and Luri (a southwest Iranian variety) among the Kurdish.

The classification of Kurdish varieties which we use in our study can be summarized in the following map (Öpengin, 2013:16, see also Haig & Öpengin, 2014; Öpengin, 2016).

[insert Figure 1 approximately here]

Figure 1: Map of language varieties spoken by the Kurds (from Öpengin, 2013:16)

The lack of consensus in the literature when it comes to defining and classifying Kurdish in linguistic and sociolinguistic terms illustrates that geographic, socio-historical, ethnic and linguistic criteria yield different results when used in language classification. These different factors furthermore influence the way the varieties listed above are named/labelled by different people and different research disciplines. This poses a methodological challenge that we will return to in the next section.

That said, for the current study it is crucial that there is consensus in the literature regarding two things: despite the lack of a single standard or unifying writing system, and despite the Kurdish speaking areas spanning minimally four nation states (which pursue different language policies towards the varieties of Kurdish spoken in their state), there is consensus that Kurds are generally defined as people who claim Kurdish identity for themselves and that the majority of people thus defined also speak Kurdish (Haig & Matras, 2002:3). The Kurdish language is therefore an integral component of any conceptualisation of “Kurdishness” (Haig & Öpengin, 2014:99).

Second, despite the lack of consensus on whether Kurdish is a language (or a continuum of related dialects) Haig & Öpengin (2014:103) assert that we can ‘meaningfully investigate what speakers of the varieties concerned perceive about their own variety in relation to others. In this

case, there seems to be a relatively broad consensus among speakers of Sorani/Central Kurdish, and speakers of Kurmanji/Northern Kurdish that their respective varieties can be identified with a larger-order entity Kurdish/Kurdi. Similar perceptions may hold for speakers of Southern Kurdish (Fattah, 2000), and for some varieties of Gorani' (ibid.). We will approach this statement with the appropriate methodology, which is outlined in the next section.

### 3. METHODOLOGY

#### 3.1 Perceptual Dialectology

Perceptual dialectology is a sub-branch of one of three ways of looking at language.

[insert Figure 2 here]

Figure 2: Three approaches to language data (Preston, 1999:xxiii)

Figure 2 shows three approaches (a-c) to language data:

(a) presents investigations and classifications of actual language use, and the states and processes which govern it (a'). Branches of linguistics taking this approach to language data aim to account for linguistic variation; they include historical linguistics, traditional dialectology and linguistic geography.

(b) represents language attitude research, quantitative like the matched guise technique, or qualitative such as discourse data (Preston 2010).

(c) covers all investigations of non-linguists' views of linguistic concepts (such as language and dialect) and what lies behind such views (b' and c').

PD is the study of how non-linguists perceive variation in language, and how the varieties they are familiar with are spatially distributed; it is a sub-branch of (c). PD thus differs from traditional dialectology both in aims and methods. In traditional dialectology, linguists create geographic boundaries of dialect areas based on phonological, lexical and morphosyntactic features. In PD, non-linguists create mental representations of linguistic variation associated with

geographical space; the data are then analysed by linguists/geographers. As in traditional dialectology, PD data are collected from speakers of the varieties under investigation, but they differ in being spatial rather than speech data.

The main technique used for capturing perceptions of linguistic variation associated with geographical space is the “draw-a-map” task (Preston, 1982). It requires respondents to indicate on a map of the region being studied where they believe dialect areas to exist. The researchers (linguists/geographers) then digitize, geographically align and analyse the data to arrive at composite maps which give a generalized picture of participants’ perception of linguistic variation.

The “draw-a-map” task is the first of Preston’s (1982) five-point method for the study of PD. We only used the first method for two reasons: one outlined earlier, i.e. the absence of an accepted standard for Kurmanji and Sorani and the lack of any standard for southern Kurdish preclude the “correctness” task (in which participants are asked to rate how “correct” the varieties spoken in a region are), and a second methodological reason, i.e. we wanted to include participants from all over the Kurdish speaking area (which precludes other language regard tasks in which participants are asked to rate varieties in comparison to their own and thus require a fixed sampling location).

In its original conception (Preston & Howe, 1987) participants in a “draw-a-map” task are given minimally detailed maps of the region being studied. They are then asked to draw lines or “borders” around the locations where they believe different dialects to exist, and to label the demarcated areas (with folk-linguistic names). We modified the “draw-a-map” task for the reasons outlined below.

The first methodological modification we introduced to the “draw-a-map” task is to use a map which shows more detail than is customary in PD studies of, for example, the US (Evans, 2013; Niedzielski & Preston, 2003), or the UK and Germany (Montgomery & Stoeckle, 2013). The motivation behind this decision is twofold.

First, the geographical area in which varieties of Kurdish are spoken is large. Although we excluded discontinuous areas in the west of Turkey, to which many Kurdish people were moved in the 1970s, and northeast Iran, the Kurdish “homeland” in the eastern Taurus and northwestern Zagros mountain ranges spans approximately 190,000 square kilometres (roughly the size of Mexico). The area displayed on our questionnaire includes all contiguous regions in which the main traditional dialect maps of Kurdish (Fattah, 2000; Haig & Öpengin, 2014; Hassanpour, 1992; Izady, 2014) indicate varieties of Kurdish to be spoken and thus overlaps with the area shown in Figure 1.

Second, a pilot study showed that many of our participants, especially older ones, are not as accustomed to map reading as people educated in the West (in Section 3 we present two hand-drawn-maps, one of which illustrates this point).

We believe this modification of methods was not only necessary but justified because PD researchers must ensure that those completing the “draw-a-map” task have a grasp of the basic geography of the area in which research is undertaken (cf. Montgomery, 2012:646; Postlep, 2015), and that respondents’ geographical knowledge is consistent so that the spatial data they provide can be treated as accurate (Preston, 1993:335, cited in Montgomery & Stoeckle 2013:83).

The second methodological modification is potentially more controversial. Rather than asking our participants to label the dialect boundaries they drew, we asked them to draw lines/borders around the areas where the main varieties of Kurdish are spoken, alongside the option to add and label additional varieties they believe to belong to the Kurdish language family. The motivations behind this decision are as follows. The linguistic situation of Kurdish (see Section 2) is arguably more complex than that in regions where PD studies have been conducted to date (e.g. the Netherlands, the US, Japan, Spain, Korea etc.). By asking our participants to indicate the main varieties of Kurdish (rather than indicating any area in which Kurdish is spoken), we aimed at (a) assuring some level of consistency in the completion of the task, and (b) comparability of the results, i.e. participants’ knowledge of the main varieties of Kurdish and where they are spoken.

This approach furthermore provides an alternative solution to the challenge all (language) classification studies face, i.e. that of deciding how many categories to work with. The main difference between our approach and the standard method is that traditional labelling approaches face the classification problem after the data have been collected (the literature generally contains very little information on how this was handled by the different researchers); this paper uses pre-defined groups.

In addition, we asked participants to draw the dialect boundaries in four different colours: blue for Northern Kurdish/Kurmanji, red for Central Kurdish/Sorani, green for Southern Kurdish varieties and black for the related varieties Zazaki and Gorani. The lines for the latter two varieties could be easily differentiated because Zazaki and Gorani are spoken in very different regions. This complicated work in the field, because the fieldworkers always had to have access to four different coloured pens, but facilitated the task for the participants, because they could easily differentiate the varieties they had already drawn by their colour. As already stated, participants had the additional option of adding more varieties/finer distinctions and labelling them, as in the traditional method, to demonstrate more in-depth knowledge of the complex linguistic situation of Kurdish.

The problem of many of the varieties of Kurdish being known under several different (linguistic, ethnic/tribal or geographical) labels/names (with spelling/pronunciation variants), first mentioned in Section 3, was meliorated with the help of the fieldworkers. They verbally supplement the given labels with alternatives. A good example in point is Zazaki, which is also known as Zaza, Dimilî, Dimilkî, Kirmanj, Kirmankjkî, Kirdkî and Kirmanjkî. The Northern and Central varieties of Kurdish are less affected by this problem, as they are generally known as Kurmanji and Sorani (with only minor spelling and pronunciation variants). Especially for Southern Kurdish we cannot rule out that some participants are familiar with more varieties and where they are spoken than they indicated on the questionnaire; they may not have indicated them because they know them under

different names (Kelhori/Kalhuri, Kirmashani/Kermanshahi, Palawani, Feyli are other cover terms for Southern Kurdish). If anything, our results may therefore have a false negative bias.

The last modification to the standard task was that we did not have a limited number of sampling sites but recruited participants from all over the Kurdish speaking area in the Middle East. We believe that this provides a better general picture of Kurdish people's perception of dialect areas and where they are located than asking people from a limited number of locations (cf. Demirci & Kleiner, 1998; Montgomery, 2012). We could use this sampling strategy because we did not include the "Degree of difference" and "Correct and Pleasant" tasks from Preston's five-point method, which rely on fixed sampling locations. In addition to basic demographic information (age and gender), we also know the respondents' place of origin.

### **3.2 Geographic Information System (GIS)**

A GIS is commonly referred to as a database application designed to collect, manage, analyse and display spatially referenced information. In the context of PD, a GIS offers a variety of techniques for geoprocessing, spatial analysis and statistics as well as ways of visualizing collected data and creating maps (Burrough & McDonnell, 1998; Kemp, 2007). Traditional dialectological (production) studies see geographic space as a 'blank canvas' (Britain, 2009:144) onto which different linguistic features are assigned. In PD, on the other hand, geographical data, i.e. the hand drawn maps produced by the participants, present the primary object of study. In contrast with pre-GIS PD attempts which treated maps as graphics and thus did not make full use of spatial aspects of the data (Onishi & Long, 1997; Preston & Howe, 1987), more recent approaches to data in PD (e.g. Cramer, 2010; Evans, 2011, 2013; Montgomery, 2012; Montgomery & Beal 2011, Montgomery & Stoeckle, 2013) use GIS to deal with data gathered using "draw-a-map" tasks. Using GIS has the following main advantages in a PD context:

- GIS anchors people's perceptions of where they believe dialect areas to exist in the real world (georeferencing) and enables the merging of PD data with other kinds of spatially referenced data;
- the topological nature of data stored in spatial databases allows for the analysis of connectivity, relationships, overlays and geostatistical methods (de Smith, Goodchild & Longley, 2015);
- GIS provides tools for extensive quantitative and qualitative analysis to support decision making processes such as the calculation of agreement levels, areas, distances (e.g. proximity) and spatial frequencies (Maguire, Goodchild & Rhind, 1991); and
- GIS analysis takes the geographical space and the way in which people inhabit it (or not) into account. Space and spatiality can be used to help understand patterns of language variation and change (Britain, 2010; Preston 2010).

In this paper GIS is used to manage the workflow from creating a digital database to analysing georeferenced data. Topological overlay techniques were used to aggregate maps and to visualize the results. Hand-drawn maps were traced and georeferenced by assigning real world coordinates to graphic attributes. All linear data indicating dialect areas were completed to gain areal features showing the specific dialect's extent. Additional data (age, gender and place of origin) were collected and entered into the database. The added value of spatially referenced maps lies in their aggregation. The so derived frequency maps provide a generalized picture of perceptions of dialect areas from groups of respondents. This has more explicative power than single images of mental maps produced by individual respondents (cf. Goodey, 1971; Lynch, 1960; Orleans, 1967). Grouped perceptions then show the perceived extent and placement of dialect areas, and can be used to ask further questions about the data. They can, for example, be used to directly relate perceptual data to other linguistic (e.g. traditional dialect maps) and non-linguistic (e.g.

demographic) data. With the exception of Fig. 5, which was produced in ESRI®ArcGIS® 10, all GIS operations were performed with Manifold®System 8.

### 3.3 Respondents and Data

The study is based on “draw-a-map” data from 186 respondents who are all lay people. They are from five countries: 70.4% are from Turkey (mostly from Tunceli, Muş and Mardin), 8.6% from Iraq, 5.9% from Iran, 2.7% from Syria and 0.5% from Azerbaijan; the place of origin of 11.8% of participants is unknown, i.e. they did not indicate their place of origin on the map.

[insert Figure 3 approximately here]

Figure 3: Place of origin of the 186 participants

140 participants are male, 39 are female (with 7 missing values). The age profile of the participants is as follows: 66.1% are 25-50 years old; 24.2% between 50 and 70; 5.9% are younger than 25; and 2.2% older than 70. Most participants now live in Western Europe. They emigrated between the 1980s and the time of data collection (2014/15). This may potentially have influenced the results as many of them are ethno-politically mobilized members of the Kurdish community with strong ties to the Kurdish movement in their respective countries of origin. They may thus have better knowledge about the varieties of their mother tongue than non-mobilized members of the community.

Fieldwork was conducted in late 2014 and early 2015 by two fieldworkers. They are both native speakers of Kurmanji from Turkey living in London. Both are graduates; one of them is a PhD student of linguistics. Participants completed the “draw-a-map” task and the demographic questionnaire under close supervision from the fieldworkers at various Kurdish Community Centres or cultural venues (concert halls, galleries, exhibition centres etc.) in Western Europe. All questionnaires were done with pen on paper, i.e. participants drew different coloured lines round the dialect areas they are familiar with, indicated their place of origin with a cross on the map, and completed the short demographic questionnaire (age and gender). Participants who wanted to be

informed about the results of the study were given the opportunity to provide us with their e-mail address.

The fieldworkers were instructed to explain the task and provide alternative labels for the varieties of Kurdish (see Section 2), but to - under no circumstances - assist the participants in the completion of the task. Participants were not allowed to use smartphones or access the Internet during the task. The fieldworkers reported that none of the participants was familiar with the concept of a dialect map. The “draw-a-map” task took between 3 and 20 minutes to complete. To illustrate this considerable difference and explain why we used a more detailed map for the “draw-a-map” task than customary in standard PD methodology (section 3.1), we present two maps: the left hand one (a) was drawn by a 18-25 year old male geography PhD student participant who took approximately 3 minutes to complete the task; the right hand map (b) was drawn by a 26-50 year old female respondent, who carefully navigated her way round the place names indicated on the map we provided, and took approximately 20 minutes to complete a version of the task.

[insert Figure 4 here]

Figure 4: Two examples of hand-drawn maps, (a) and (b)

Figure 4 illustrates that the ways people approach the “draw to map task” are manifold and so are the results of the graphic representation. GIS, however, uses only three topological representations of real world phenomena, namely points, lines, areas and additionally text and raster/image data (Robinson, 1998). Figure 4 illustrates typical challenges posed by “draw-a-map” data to GIS processing which are mainly due to the linear representations of areal features (Couclelis, 1996). Given the processes outlined in Section 3.2, all lines were topologically checked for “completeness” and original data were transformed into areas to produce the frequency maps discussed in Section 4 (Results). As for Figure 4 (b) all lines were considered part of areal features with the outmost blue and black lines defining the border lines. In order to (internally) keep track of the different quality levels of the original data for GIS purposes, and to develop a future research agenda on quality

management of PD data (see Section 6), we marked all hand drawn maps with numbers between 10 (perfect for GIS analysis) and 1 (extensive topological pre-processing necessary).

## 4. RESULTS

### 4.1 Familiarity with Kurdish Varieties

The first research question addressed the familiarity of participants with varieties of Kurdish, i.e. recognition levels. The questionnaire provided participants with the opportunity to indicate in separate colours the areas in which the three main varieties are spoken: blue for Northern Kurdish or Kurmanji, red for Central Kurdish or Sorani, and green for the Southern Kurdish varieties. Participants were asked to indicate the related varieties Zazaki and Gorani in black (see Section 3 for why the use of the same colour does not impact on the results). In addition, participants could indicate in any colour other varieties which they believe to belong to the Kurdish family, but were explicitly asked to label them.

In PD, the presence of lines drawn by participants to indicate specific dialect areas serves as an indicator for recognition levels.<sup>6</sup> Table 1 shows raw figures indicating how many respondents recognized an area, along with a percentage respondent recognition figure.

[insert Table 1 approximately here]

Table 1: Raw recognition figures and percentages of varieties of Kurdish

The results presented in Table 1 indicate large differences in recognition levels, with all respondents having indicated an area where they believe Kurmanji to be spoken. The—in terms of number of speakers—second biggest variety, Sorani, also achieves the second highest recognition level in our study. The related variety Zazaki, which is only spoken in Turkey, reaches the third highest recognition level. Southern Kurdish varieties are only indicated by slightly over half of our respondents (55%). The related variety of Gorani, which is not only “definitely endangered” according to the UNESCO’s Atlas of the World’s Languages in Danger, but also spoken in the far

southeastern corner of the Kurdish speaking area, is also the by far least recognized variety in our study. The numerically biggest varieties thus also have the highest recognition levels in our study.

Other factors that may have influenced recognition levels include: proximity; the varieties' status in the various nation states; the presence/absence of a "unified" label/name for the varieties (see Sections 2 and 3); and a methodological issue.

Bare proximity has long been identified as an important factor in previous PD work (e.g. Montgomery, 2012; Preston, 1998) and also seems to play a significant role in our study. Table 1 shows that the highest recognition levels are achieved by Kurmanji, Zazaki and Sorani, the three varieties spoken in areas closest to most of our participants' places of origin in the northwest of the Kurdish language area. The Southern Kurdish language area, by contrast, is furthest away from where most of our participants originate. So are the pockets where Hawrami, a variety of Gorani, is spoken. Tobler's (1970:236) first law of geography, 'everything is related to everything else, but near things are more related than distant things', may thus not only account for the low recognition levels of Southern Kurdish and Hawrami, but is likely to have contributed to the 100% recognition level of Kurmanji, the variety spoken the by the majority of respondents from Turkey.

Out of the varieties that were frequently drawn by our participants, only Sorani has official language status in Iraq (see Section 2); this is also widely known among Kurds from other regions and may thus have contributed to the high recognition level of Sorani. Given that we have no participants from Armenia, the minority language status of Kurmanji there cannot have influenced recognition levels. None of the other varieties of Kurdish have official status at national level in any of the countries where they are spoken.

The absence of a "unified" label/name for Southern Kurdish varieties is likely to have contributed to their low recognition level. Although we provided more examples for Southern Kurdish varieties in writing on the questionnaire than for Central and Northern Kurdish (and the fieldworkers supplemented more verbally), the absence of an umbrella term under which Southern

Kurdish dialects are known may have contributed to the low recognition rate of the Southern Kurdish language area. The same holds true for Gorani, for which Sheyholislami (2015:35) notes that '[t]here is no consensus as to what this group should be called'. The fact that Gorani is frequently associated with literary tradition rather than a spoken variety, may have further contributed to its very low recognition level.

Recognition levels may furthermore be influenced by a methodological issue. Respondents may know that a varieties exist, but do not indicate this knowledge because they cannot associate the variety with an area in which it is spoken.<sup>6</sup> The PD requirement to delineate dialect areas may have discouraged participants to indicate varieties they are aware of, but do not know where they are spoken. This methodological issue may have created a false negative effect in our results.

Other varieties of Kurdish our participants indicated and labelled on the questionnaire are the northern variety of Bahdini, the northern variety of Gorani (Shabak), and Leki. The results to research question 1 have thus provided us with a picture of recognition levels of varieties of Kurdish among our participants. The varieties which are spoken at or near the participants' place of origin (Kurmanji, Zazaki) and either have a fairly unified name (Kurmanji and Sorani) or official status in one of the Kurdish regions (Sorani) achieve the highest recognition levels. Numerically small varieties with many different names, which are spoken far away from the majority of participants' place of origin and have no official status in the nation state(s) they are spoken in, achieve the lowest recognition levels (all southern varieties and the related variety Gorani).

#### **4.2 Spatial Pattern of Kurdish Varieties**

Research question 2 investigates the location and extent of dialect areas of Kurdish. Participants were asked to represent their mental maps of where different varieties of Kurdish are spoken by indicating their perimeters in different coloured pens on a given map of the Kurdish homeland. In addition, they could delimit sub-varieties and label them. The lines drawn by all 186 respondents (note that not all respondents indicated all varieties, which led to the different recognition levels

discussed in the previous section) were then aggregated in GIS to yield two primary sets of visualisations of results: line density data and frequency maps. Figure 5 shows the line density results for Northern Kurdish.

[insert Figure 5 approximately here]

Figure 5: Kurmanji line density

The line density procedure calculates a grid representing the prevalence of lines within each raster cell. The more lines fall within a raster cell, the more intense the colour appears. The dark blue areas show a high level of line density and indicate that our respondents can delineate the area where Kurmanji is spoken with a high degree of certainty.

We expected high recognition levels and high accuracy in terms of dialect area delineation for Kurmanji as the majority of respondents are speakers of this variety from Turkey (bare/spatial proximity effect). The individual lines outside the dark blue high density area in Figure 5, however, also show that a number of respondents indicated a smaller or larger Kurmanji speaking area. The former show no discernible pattern; most of the latter included regions where other non-northern varieties of Kurdish are spoken. A clearly noticeable bundle of lines, for example, include the Sorani speaking areas; another group of respondents drew the entire Kurdish-speaking area. This may indicate that, for some respondents, Kurmanji represents Kurdish. A more detailed analysis of “typical” areal border line drawings using predefined groups of respondents may provide more insight into the spatial delineation of a linguistic variety and constitutes an area for future research (see **Error! Reference source not found.** 6). The fact that the southernmost high density bundle of lines clearly lies south of the Turkish border suggests a minimal effect of national borders on the line drawings, and shows that the majority of respondents are aware that Kurmanji is also spoken in the north of Syria and Iraq. A closer look at these lines on the individual maps reveals an interesting “cultural prominence” effect.

Perceptual geographers (e.g. Gould & White, 1986) as well as linguists (e.g. Britain, 2010) have pointed out that proximity is not static, because speakers do not inhabit a homogeneous information space. Proximity can be influenced by other factors, such as physical geography (mountain ranges, rivers) and cultural prominence (Montgomery & Beal, 2011). This is, distant places may be “brought closer” and become more prominent through, for example, media exposure. Stuart-Smith (2011:3) proposes ‘that the broadcast media will have an impact on metalinguistic awareness of linguistic varieties and variation [...] and the ideologies surrounding them’. Such forms of cultural prominence may have influenced recognition levels in our study. More interestingly, however, we have concrete evidence that media exposure seems to have influenced our participants’ perceptions of the extent of dialect areas.

As already stated, data collection took place in late 2014 and early 2015, during the battle over Kobane (<http://www.bbc.co.uk/news/world-middle-east-29688108>), North Syria, and IS assaults on the Yazidi community in Sinjar, Iraq (<http://www.bbc.co.uk/news/world-middle-east-31962755>). 16<sup>7</sup> of our participants who completed the questionnaire during this time “extended” (see Figure 6 for an example) their Kurmanji line south into Syria, and 29<sup>8</sup> southeast into Sinjar/Shingal, north Iraq, after having originally drawn lines closer to the Turkish border.

[insert Figure 6 approximately here]

Figure 6: Questionnaire 29 by a male respondent aged 25-50

It seems likely that these “extensions” into very specific areas were related to the intense media presence these Kurdish-speaking areas had at the time. We therefore believe that the post-hoc additions of dialect areas into zones that had high media presence at the time of data collection present first concrete evidence of increased “cultural prominence” (Montgomery, 2012:640) or metalinguistic knowledge via the media.

Another feature of the line density for Kurmanji presented in Figure 5 requires an explanation. Although the northernmost high density line bundle lies close to the edge of the map on which the

participants were asked to draw their mental maps of Kurmanji, it clearly falls within. The westernmost PD line bundle, on the other hand, is right on the edge of the map that was provided to the participants, despite the map having a similar “safety” margin to the westernmost traditional dialect boundary of Kurmanji as it has to the north (on Haig & Öpengin’s map – see Figure 1 – the northern dialect boundary of Kurmanji runs along 39.9° N, while our map stops at N 40°10'; on Haig & Öpengin’s map the western dialect boundary of Kurmanji runs through Gaziantep 37.3781° E; our map stops at E 37° 7'E). Fieldnotes suggest the following reason for this anomaly: several participants complained that we had excluded areas in the west of Turkey from our map, to where many Kurmanji speakers were dislocated in the 1970s (Gunes, 2012:39). Drawing the westernmost dialect boundary of Kurmanji right at the edge of our map may thus have been a form of protest against the exclusion of these Kurmanji speaking areas.

To make full use of the areal features of the collected data, one of the core GIS tools, a topological overlay, is performed on all maps. The frequency maps resulting from this operation indicate the amount of agreement by respondents on the spatial presence of a dialect. Figure 7 presents the frequency maps for Kurmanji and Zazaki, overlain because the Zazaki speaking areas form islands within the Kurmanji speech zone. For mapping purposes, all data have been visualised using ‘natural breaks’, a data classification method which groups the project data and sets class breaks where there is a jump in values, so groups having similar values are placed in the same class.

[insert Figure 7 approximately here]

Figure 7: Frequency maps for Kurmanji and Zazaki

Figure 7 shows a large area in dark blue. This indicates the high level of overlap in participants’ mental maps of the Kurmanji dialect area. Figure 7 thus demonstrates—even better than the line density map (Figure 5)—that the participants agree to a large extent on the core Kurmanji speaking areas. The dark blue core area for Kurmanji furthermore does not have the “western map edge

effect” discussed in relation to the line density map, and, as we will see in the next section, largely coincides with the western border of traditional dialect maps of Kurmanji.

The frequency maps for Zazaki are even more interesting than the one for Kurmanji. They reveal that our participants are aware of a) Zazaki being spoken within the Kurmanji dialect areas and b) Zazaki speaking areas clustering round three centres (Northern Zazaki round Dersim / Tunceli, Central Zazaki round Bingöl, and Southern Zazaki round Diyarbakir and Siverek). This gives rise to the “island effect” of 3-4 high density areas separated by low density areas. This reflects the actual situation of Zazaki much better than the uniformly grey area for Zazaki on the traditional dialect maps by e.g. Hassanpour (1992) and Haig & Öpengin (2014) (see Section 2). This finding furthermore goes some way towards explaining why Zazaki is classified as “vulnerable” on UNESCO’s List of Endangered Languages. Studies of ethnolinguistic vitality<sup>9</sup> have demonstrated that, in addition to intergenerational language transmission, absolute number of speakers (approximately 2 million in the case of Zazaki), proportion of speakers within the total population (2/78.5 million), and language attitudes and policies, being spoken in discontinuous / non-adjoining geographical areas has a detrimental effect on ethnolinguistic vitality (e.g. Edwards, 1992, 2006; Ehala, 2005; Östman & Mattfolk, 2011).

In Figure 8 below we present the areas in which our participants believe Sorani and Southern Kurdish varieties to be spoken next to each other in order to highlight the unexpected results we got for areas in Iran in general, and in the West Azerbaijan and Kordestan provinces in particular.

[insert Figure 8 approximately here]

Figure 8: a) Frequency map for Central Kurdish b) Frequency map for Southern Kurdish

The frequency map for Sorani shows a high level of agreement, which indicates that the vast majority of the participants (122/186) have good knowledge of the core Sorani speaking area. The recognition levels (see research question 1) for the varieties discussed so far are high (Kurmanji 100%, Sorani 92% and Zazaki 87%) and we have therefore not drawn particular attention to the fact

that not all maps show all varieties. The results for Southern Kurdish in particular need to be looked at in the context of the fact that only 55% (103/186) of participants indicated Southern Kurdish areas. The highest agreement value these 103 hand drawn maps for Southern Kurdish achieve is 45, and the dark green area in Figure 8b marks the region in which more than 31/103 and up to 45/103 of our participants believe Southern Kurdish varieties such as Kelhuri, Feyli and Kirmashani to be spoken (when Laki was indicated by the participants, it was always labelled separately, suggesting that the participants do not see Laki as belonging to the Southern Kurdish group). This shows that our participants were considerably less confident in identifying the extent and placement of the Southern Kurdish dialect areas than they were for Central and Northern Kurdish varieties. As already indicated, this may be due to distance from the participants' place of origin, the considerable linguistic diversity of Southern Kurdish, and the lack of an umbrella term for Southern Kurdish varieties.

The distinctly odd result emerging from the “draw-a-map” task is that the mental maps for Central and—to an even greater extent—Southern Kurdish of a considerable number of participants included an area roughly coinciding with the West Azerbaijan province in Iran. That is, up to 52/171 participants think that Sorani is spoken in northwest Iran, and more than 31/103 participants believe that Southern Kurdish varieties stretch far north into Iran and into an area in which—with the exception of Bijari—no southern variety of Kurdish is spoken (Anonby et al., 2016). Figure 9 shows different spatial patterns of the Southern Kurdish varieties in Iran by respondents from Turkey, Iraq and Iran. The comparison presented in Figure 9 thus indicates that our Turkish and Iraqi participants have only vague knowledge of the linguistic situation in Iran.

[insert Figure 9 approximately here]

Figure 9: National splitting (place of origin: a) Turkey, b) Iraq, c) Iran) of Southern Kurdish frequency maps

These results did not come as a big surprise to linguistic experts on Iran. One of them notes:

‘I was not surprised to discover that your “participants’ perceptions of the varieties of Kurdish spoken in Iran went pretty haywire”, as you put it [at the 3<sup>rd</sup> International Conference on Kurdish Linguistics, Amsterdam 2016]. The fact of the matter is that overall there is very little information about the Kurds in Iran in general let alone linguistic information. For example, since the early 1990s in particular there has been an exponential growth in research about and publication on the Kurds from Iraq and Turkey. This has also been similar with media coverage in both the West and the Middle East (e.g., in Persian, Arabic, and Turkish). In all these respects, interest in Iranian Kurds has been extremely limited and any reliable data and information has been meagre. This was also the case with the Kurds from Syria but this has changed notably since 2012.’ (Sheyholislami, e-mail 30 August, 2016)

The last sentence furthermore supports our interpretation of the “extensions” to participants’ Kurmanji lines into North Syria and North Iraq presented earlier in this section as a “cultural/media prominence” effect.

Due to the low recognition rate for the definitely endangered related variety Gorani (3.2% or 6/186), the results for Gorani were not incorporated into frequency maps. The six participants who indicated Gorani on their hand-drawn map, however, showed the placement and extent of where it is spoken with high accuracy.

### **4.3 PD Patterns Compared to Isoglosses?**

Research question 3 asks how the results to research question 2, i.e. PD dialect boundaries, pattern with traditional dialect boundaries, i.e. bundles of isoglosses. It has been suggested that PD research can play a role in looking afresh at the results of production studies (Montgomery & Stoeckle, 2013) and that it could even challenge assumptions made from such studies (Butters, 1991). This can only be done with aggregated data such as that presented in the previous section. Figure 10 presents the summary results of our PD study with all four main varieties of Kurdish indicated, overlain with the most widely used traditional dialect map, i.e. Figure 1 (Haig & Öpengin, 2014).

[insert Figure 10 approximately here]

Figure 10: Kurdish varieties: traditional vs. perceptual dialectology

The results to research question 3 show that the participants in our PD study have very good knowledge of the extent and placement of the core areas where Northern and Central Kurdish and the related variety Zazaki are spoken. For Northern Kurdish, the trend observed in the previous section, i.e. that the participants extended the core Kurmanji speaking area slightly to the west, is also noticeable here. The perceptual Kurmanji speaking area shows a slight national border effect in the east, but more than half of our respondents (84-109/186) drew the eastern border of Kurmanji in accordance with the traditional dialect boundary for this variety. This lends further support to the interpretation of the line density and frequency maps of Kurmanji which have already indicated that our participants have solid knowledge of the extent and placement of Kurmanji.

In comparison with the traditional dialect map, the perceptual core area of the Sorani speech zone is also shifted to the northwest, towards the Iraqi-Turkish border. Like the new *Atlas of the Languages of Iran* (Anonby et al., 2016), but unlike the traditional dialect maps by Hassanpour (1992) and Haig & Öpengin (2014), some (27-52/171) of our participants extend the Sorani speaking area to the very south of Kordestan province in Iran. The perceptual Sorani speaking area, however, does not stretch anywhere near as far east as suggested by Hassanpour (1992) and Haig & Öpengin (2014), let alone southeast, as suggested by the *Atlas of the Languages of Iran* (Anonby et al., 2016). These findings should therefore encourage traditional and perceptual dialectologists to look afresh at their results, as suggested by Montgomery & Stoeckle (2013). Our participants' mental representation of the Sorani speech area furthermore clearly includes the Mukri speech zone. This tentatively suggests that our participants "correctly" classify Mukri as part of the Central Kurdish dialect group.

Our participants also show good knowledge of the Kelhuri and Kermashani areas of Southern Kurdish varieties. The biggest discrepancy between our PD maps of Southern Kurdish and all

traditional dialect maps, however, is that our participants extend the Southern Kurdish area far into the west (Iraq) and northwest (Iran). As suggested in the previous section, this can currently only be put down to the limited knowledge Kurds have about their fellow Kurds in Iran, and clearly invites further investigation.

The Gorani/Hawrami speech area is shown as a light grey blob around Paveh on the map in Figure 1. As already mentioned, those six participants who did indicate the Gorani/Hawrami area did so with high accuracy. This is not surprising as the Hawrami speaking area is compact, geographically distinct and includes the well-known dialects of Paveh and Halabja.

We would furthermore like to highlight that the picture the PD map presents for Zazaki is more differentiated than that suggested by traditional dialect maps (Fattah, 2000; Haig & Öpengin, 2014; Hassanpour, 1992; Izady, 2014; Öpengin, 2013, 16). Even the most nuanced of these maps (Le Monde Diplomatique, 2007) still shows the Zazaki speech zone as a more or less continuous area. The PD map, on the other hand, highlights the vulnerability of Zazaki caused by, among other factors, the loss of geographical connectedness between speakers of Zazaki due to the speech zone breaking up into small language islands (around Tunceli, Bingöl, Elazığ, Diyarbakır and Siverek) in an otherwise Kurmanji and Turkish speaking area.

#### **4.4 PD of Kurdish and demographic factors**

Research question 4 investigates how non-linguists' perceptions of linguistic variation pattern with non-linguistic facts, such as demographic, national, religious, ethnic, economic, cultural and political areas/boundaries. For this research question, GIS provide extensive means to establish and analyse spatially referenced databases and allow for advanced information retrieval using spatial queries (de Smith, Goodchild & Longley, 2015). At this point we address two factors only: how participants' perception of linguistic variation in Kurdish patterns with their country of origin and age. We select these two variables because, in future research, we would like to investigate in more depth how place/country of origin (e.g. Kurds being educated in four different educational systems)

affects their awareness of the varieties of their language and their spatial distribution, and how this awareness changes with age/over time (Anselin, Sybri & Kho, 2006). The results presented below are thus only an outlook into future research.

Figure 11 summarizes participants' recognition levels of the main varieties of Kurdish by country of origin.

[insert Figure 11 approximately here]

Figure 11: Participants' recognition levels of the main varieties of Kurdish by country of origin

The results presented in Figure 11 require a cautious interpretation because the number of respondents from each country varies considerably (Turkey n=131, Iraq n=16, Iran n=11, Syria n=5, Azerbaijan n=1, unknown n=22). This said, the line of enquiry looks promising because the results presented in the bar charts reveal interesting differences. Respondents from Syria and Iraq show the most balanced recognition levels of the main and related varieties of Kurdish. Two factors are likely to have influenced this result: the participants from both Syria and Iraq come from the more central parts of the Kurdish dialect area and are thus in closer proximity to all varieties of Kurdish than, for example, participants from northwest Turkey or southeast Iran. Historically Kurds in Iraq have furthermore enjoyed the more liberal cultural and language policy. Kurdish, for example, became the official language of Iraqi Kurdistan in 1992. As all Iraqi participants are over 25 years of age, all of them will have benefitted from the increased cultural prominence of Kurdish since the early 1990s. Proximity also seems to be the main factor behind the results from respondents from Turkey and Iran: the former show a significant drop in recognition level for Southern Kurdish, while the latter are already less familiar with some of the (southern) varieties spoken in their own country, and even less so with the related variety of Zazaki, which is only spoken in Turkey.

Age already emerged as a relevant factor from the national comparison, and yields even more interesting results when looked at in more detail. Recognition levels for Northern and Central

Kurdish and Zazaki are consistently high among respondents from Turkey, while the recognition level for Southern Kurdish drops to 11% (22/186) among the 18-25 age group and to zero among the old age group. As we have a good number of respondents from Turkey (131/186), these results are relevant. Distance is likely to be the most prominent explanatory factor behind the unfamiliarity of the old Turkish participants with the varieties of Kurdish that are spoken furthest away from their place of origin. The 18-25 age group should be less affected by bare proximity (due to globalisation and the World Wide Web). The results from the young age group may be due to the low cultural prominence of Kurdish in Turkey and thus particularly worrying in terms of the role the Kurdish language can play in identity formation in this country. The number of participants from Iraq, Iran and Syria are not high enough to attribute much significance to the further splitting of recognition levels of varieties by age groups. The general trend among Iraqi and Iranian participants, however, is so similar that the combined results are worth commenting on: Iraqi and Iranian participants over the age of 70 are equally familiar with all main varieties of Kurdish; recognition levels for Southern Kurdish and especially Zazaki already drop among the 50-70 age group and even more so among the 25-50 age group. These results may give rise to concern, as a) Zazaki is already considered vulnerable and b) non-linguists' perceptions have been proposed as a bellweather for linguistic changes affecting linguistic diversity (Montgomery & Stoeckle, 2013:66).

The studies published in Cramer & Montgomery (2016) suggest that a comparison between aggregate PD data, as presented in this study, and natural speech data from urban areas, as collected by Belelli (2016), is appropriate and can indicate regional dialect levelling.<sup>10</sup>For Southern Kurdish Belelli (2016:15) notes: 'Dialect blending is particularly common in major urban centres (e.g. Kermānshāh, Ilām, Qasr-e Shirin), where speakers of many different dialects live side by side'. The young (< 25) participants from Syria show a particularly balanced recognition level. Whether this is due to the increased cultural prominence of Kurdish in the Jazira Canton, where Kurdish became an official language at the regional level in 2014, remains to be seen.

## 5. SUMMARY

This paper presents the first investigation into Kurdish people's awareness of the varieties of their language and where they are spoken. It is based on a modified version of the "draw-a-map" task from Perceptual Dialectology. The data, maps drawn by 186 Kurdish participants indicating the varieties of their language and their localisation, are spatially referenced, analysed and displayed with GIS.

The results show that, despite the varieties under investigation being used across five nation states and only one (Sorani) having official status at the national level in Iraq, participants have good knowledge of the main varieties of their language (Kurmanji, Sorani) and where they are spoken. Participants from Turkey also identified the Zazaki speaking areas with high accuracy. Post-hoc extensions to 45 drawn lines demarcating the Kurmanji speaking area into Syria and Iraq, furthermore, appear to present first concrete evidence for the "cultural prominence" effect on metalinguistic awareness of linguistic varieties proposed by Montgomery & Beal (2011). It seems plausible that these "extensions" into highly specific areas are related to the intense media presence Kobane (North Syria) and Sinjar/Shingal (Iraq) had during the time of data collection (autumn 2014 to spring 2015). Awareness of the more diverse Southern Kurdish varieties is less definitive and depends on the participants' region of origin. Participants' awareness of the linguistic situation of Kurdish in Iran emerged as sketchy.

The results for the two main varieties, Kurmanji and Sorani, furthermore reveal a good match between the perceptual dialectology boundaries as indicated by our participants with bundles of isoglosses established by traditional dialectological methods. The aggregated frequency maps for the related variety Zazaki exposed an "island effect" of 3-4 high density areas separated by low density areas, which presents a more differentiated picture for this variety than the uniformly grey area for Zazaki on the traditional dialect maps. In combination with research findings from ethnolinguistic vitality research, this finding moreover supports one explanation why this variety is

classified as “vulnerable” on UNESCO’s List of Endangered Languages: being spoken in discontinuous/non-adjointing geographical areas has a detrimental effect on ethnolinguistic vitality.

One potential limitation of the study is that our participants live in the diaspora. On-line versions of the “draw-a-map” task that are currently being developed will facilitate data collection even in remote areas such as those where Kurdish is spoken. Perceptual dialectologists working with Preston’s five-point method may also find fault with our modified methodology. We believe that the modifications are appropriate for a first investigation into Kurdish peoples’ awareness of where the varieties of their language are spoken; in follow-up investigations we would potentially opt for participants labelling the varieties they indicate with the folk-linguistic names they know them under. Including “language regard” (Preston 2010) tasks from PD constitutes another interesting potential avenue for future research.

## **6. OUTLOOK AND CONCLUSION**

The main difference between traditional dialectology and perceptual dialectology lies in the fact that PD data show genuinely geographic attributes, whereas speech production data are projected onto geographic locations. In David Britain’s (2009:144) words, traditional dialectological studies treat geographic space as a ‘blank canvas’ onto which different linguistic features are assigned. In PD studies, spatially referenced lines delineating linguistic varieties form the basis of the investigation and dialect areas emerge from aggregating these geospatial boundaries into frequency maps. The resulting maps open up alternative ways of looking at the relevance of boundaries and mental representations of dialect areas.

GIS in general are considered to precisely represent points, areas and lines on earth. Uncertainties (due to vague as well as ambiguous data), however, remain inherent when using linguistic data as representation of spatial knowledge (Benedikt, Reinberg & Riedl, 2002, 2004; Kratochwil & Benedikt, 2005). These uncertainties can be addressed with weighting algorithms; we intend to experiment with spatial weights using ordered weighted averaging operators (see *ibid.*) in

ensuing research. The comparison of our participants' perceptions of linguistic variation with non-linguistic/demographic data suggests this GIS supported line of inquiry as promising for future research (Zazah & Desachy, 1993). Integrating PD data and spatially referenced socio-demographic attributes provides a clear profile of the demographic groups among which the mental representations of linguistic landscapes are most/least endangered.

Aside from technological challenges, the main findings of this study can potentially contribute to and impact on Kurdish language and identity research, language policy and planning, and language maintenance and shift.

As outlined in Section 2, only Sorani has official language status in Iraq and Kurmanji is a recognized minority language in Armenia. In all other nation states in which varieties of Kurdish are spoken, Kurdish has either been banned from public life, education and the media for the majority of the past 90 years (Turkey), or “enjoyed” ‘restricted and controlled tolerance’ (Iran, see Sheyholislami, 2015:41). Given the lack of almost any kind of institutionalized / state-sanctioned support and dissemination of knowledge regarding the Kurdish varieties in the country in which most Kurds and most of our respondents live(d), recognition levels for and placement of the main varieties Kurmanji and Sorani (as well as the related variety Zazaki among Turkish respondents) as emerged from this study seem high. We interpret this in a similar direction as Haig & Öpengin (2014), i.e. that the Kurdish language is an integral component of any conceptualisation of “Kurdishness” for the speech community and important for identity building.

Multilingualism<sup>11</sup>, the linguistic diversity of Kurdish, as well as the limited political influence Kurdish politicians have in Turkey, Iran and Syria, pose considerable challenges for language policy and planning. Most countries varieties of Kurdish are spoken in pursue a “one nation, one language” policy.<sup>12</sup> Language planning efforts are also fraught with difficulties at all stages (status and corpus planning), because different Kurdish groups nominate different varieties for corpus planning (a case in point being the struggle between Sorani, Badini and Hawrami speaking groups

in Iraq since the mid-1990s), and several—sometimes competing—groups of scholars work on corpus planning. What emerges from the current study, much more than from traditional dialectological studies, is a) that both Kurmanji and Sorani have strong mental representation among our participants, and b) that there is considerable overlap in dialect areas. These findings, applied to language policy and planning, seem to suggest as the most fruitful approach one which recognizes the linguistic diversity of Kurdish by accepting more than one standard (cf. Hassanpour, Sheyholislami & Skutnabb-Kangas 2012). Examples like Norway demonstrate that having two linguistically related standards is practicable for a considerable period of time; examples like Switzerland show that a coherent identity can be formed even with four official languages.

One of the strongest findings of this study is that the varieties of Kurdish on the United Nations' list of endangered languages, Gornai and Zazaki, are also weakly represented in some of our participants' mental maps; Gorani for all participants, Zazaki particularly for the central age groups from Iran and Iraq. If we subscribe to the notion that non-linguists' perceptions of varieties can be a harbinger for linguistic changes affecting diversity, and take seriously the ethnolinguistic tenet that 'to the extent a particular language is the main focus of a group's identity, ethnolinguistic vitality is an important indicator of possible language shift or maintenance' (Giles et al., 1977:307), all possible attempts should be made to maintain the linguistic diversity of Kurdish.

**REFERENCES**

- Anonby, Erik et al. (eds.). 2016. *Atlas of the Languages of Iran*. Ottawa: Geomatics and Cartographic Research Centre. <http://iranatlas.net> (retrieved 22 January, 2016.).
- Anselin, Luc, Ibnu Syabri & Youngihn Kho. 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38. 5–22.
- Asadpour, Hiwa. 2011. *The Computer Developed Linguistic Atlas of Azerbaijan-e Qarbi: Notes on Typological-Perceptual approach in Geolinguistics*. Tehran: Islamic Azad University dissertation.
- Benedikt, Josef, Sebastian Reinberg & Leopold Riedl. 2004. Vague Geographical Knowledge Management - A Flow-Chart Based Application to Spatial Information Analysis. In Rita de Caluwe & Guy de Tré & Gloria Bordogna, (eds.) *Spatio-Temporal Databases: Flexible Querying and Reasoning*. 314–329. Berlin: Springer.
- Benedikt, Josef, Sebastian Reinberg & Leopold Riedl. 2002. A GIS application to enhance cell based information modelling. *International Journal of Information Sciences* 142. 151–159.
- Belelli, Sara. 2016. *A Study on Language and Folklore in the City of Harsin (Kermānshāh Province, West Iran): Sketch Grammar with Texts and Lexicon*. Naples: University of Naples dissertation.
- Britain, David. 2009. Language and space: The variationist approach. In Peter 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 Alfred Lameli, R. Kehrein & S. Rabanus (eds.), *Language and space, an international handbook of linguistic variation*, 2. 69–97. Berlin: Mouton de Gruyter.
- Burrough, Peter A. & Rachael A. McDonnell. 1998. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.

- Butters, Ronald R. 1991. Review of Dennis Preston, perceptual dialectology. *Language in Society* 20. 294–299.
- Chambers, J.K. & Peter Trudgill. 1998. *Dialectology*. Cambridge: Cambridge University Press.
- Couclelis, Helen. 1996. Towards an operational typology of geographic entities with ill-defined boundaries. In Peter A. Burrough & Andrew Frank (eds.), *Geographic Objects with Indeterminate Boundaries*. 45-57. Boca Raton: CRC Press.
- Cramer, Jennifer. 2010. *The effect of borders on the linguistic production and perception of regional identity in Louisville, Kentucky*. Urbana, IL: University of Illinois at Urbana-Champaign dissertation.
- Cramer, Jennifer & Chris Montgomery (eds.). 2016. *Cityscapes and perceptual dialectology*. Berlin: Mouton de Gruyter.
- Demirci, Mahide & Brian Kleiner 1998. The Perception of Turkish Dialects. In Dennis Preston (ed.), *Handbook of Perceptual Dialectology* 3. 263–281. Amsterdam: John Benjamins.
- Edwards, John. 1992. Sociopolitical aspects of language maintenance and loss: towards a typology of minority language situations. In Willem Fase, Koen Jaspaert & Sjaak Kroon (eds.), *Maintenance and Loss of Minority Languages*, 37-54. Amsterdam: John Benjamins.
- Edwards, John. 2006. Players and power in minority-group settings. *Journal of Multilingual and Multicultural Development* 27(1). 4-21.
- Ehala, Martin. 2005. The role of MTE in language maintenance and developing multiple identities. In Siegfried Kiefer & Kari Sallamaa (eds.), *European identities in mother tongue education*, 36-50. Linz: Trauner Verlag.
- Evans, Betsy. 2011. Seattletonian to Faux Hick: Mapping perceptions of English in WA. *American Speech* 86(4). 383-413.
- Evans, Betsy E. 2013. “Everybody sounds the same”: Otherwise overlooked ideology in perceptual dialectology. *American Speech* 88(1). 63-80.

- Fattah, Ismasil K. 2000. *Les Dialectes Kurdes Méridionaux: étude linguistique et dialectologique*. Louvain: Peeters.
- Giles, Howard, Richard Y. Bourhis & Donald Taylor. 1977. Towards a theory of language in ethnic group relations. In Howard Giles (ed.), *Language, Ethnicity and Intergroup Relations*, 307-348. London: Academic Press.
- 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.
- Gunes, Cengiz. 2012. *The Kurdish National Movement in Turkey: From Protest to Resistance*. London: Routledge.
- Haig, Geoffrey & Yaron Matras. 2002. Kurdish linguistics: a brief overview. *Sprachtypologie und Universalienforschung / Language Typology and Universals*, 55(1). 3–14.
- Haig, Geoffrey & Ergin Öpengin. 2014. Introduction to Special Issue Kurdish: A critical research overview. *Kurdish Studies* 2(2). 99–122.
- Hassanpour, Amir. 1992. *Nationalism and language in Kurdistan, 1918-1985*. San Francisco: Mellen Research University Press.
- Hassanpour, Amir, Jaffer Sheyholislami & Tove Skutnabb-Kangas. 2012. Introduction. Kurdish: Linguicide, Resistance and Hope. *International Journal of the Sociology of Language* 2012 (217). 118.
- Izady, Martin. 1992. *The Kurds: A Concise Handbook*. London: Taylor & Francis.
- Izady, Martin. 2014. Languages of the Middle East.  
[http://gulf2000.columbia.edu/images/maps/Mid\\_East\\_Linguistic\\_lg.png](http://gulf2000.columbia.edu/images/maps/Mid_East_Linguistic_lg.png) (22 January, 2017.)
- Kemp, Karen K. (ed.). 2007. *Encyclopedia of Geographic Information Science*. Thousand Oaks: Sage Publications Inc.

- Kerswill, Paul. 2002. Models of linguistic change and diffusion: new evidence from dialect levelling in British English. *Reading Working Papers in Linguistics* 6. 187-216.
- Kratochwil, Susanne & Josef Benedikt. 2005. Talking Space – A Social & Fuzzy Logical GIS Perspective On Modelling Spatial Dynamics. In Frederick E. Petry, Vincent B. Robinson & Maria A. Cobb (eds.), *Fuzzy Modeling with Spatial Information for Geographic Problems*, 159-185. Berlin: Springer.
- Le Monde Diplomatique. 2007. Geographic distribution of Kurdish and Zaza–Goran. <https://ethnicgeography.files.wordpress.com/2013/09/000654a.jpg> (22 January, 2017).
- Lynch, Kevin. 1960. *The image of the city*. Cambridge, MA: MIT Press.
- MacKenzie, David N. 1961. *Kurdish Dialect Studies*. Oxford: Oxford University Press.
- Maguire, David. J., Michael F. Goodchild, & David W. Rhind (eds.). 1991. *Geographical Information Systems: principles and applications - volume 1*. London: Longman
- Mahmoudveysi, Parvin & Denise Bailey. 2013. *The Gorani Language of Zarda, a Village of West Iran. Texts, Grammar and Lexicon*. Wiesbaden: Reichert.
- Matras, Yaron. 2016. “The Dialects of Kurdish” Project. <http://kurdish.humanities.manchester.ac.uk>. (22 January, 2016.)
- Montgomery, Chris & Joan C. Beal. 2011. Perceptual dialectology. In Warren Maguire & April McMahon (eds.), *Analysing variation in English*, 121-148. Cambridge: Cambridge University Press.
- Montgomery, Chris & Philipp Stoeckle. 2013. Geographic information systems and perceptual dialectology: a method for processing draw-a-map data. *Journal of Linguistic Geography* 1. 52–85.
- Montgomery, Chris. 2012. The effect of proximity in perceptual dialectology. *Journal of Sociolinguistics* 16(5). 638–668.

- Moseley, Christopher (ed.). 2010. *Atlas of the World's Languages in Danger*, 3rd edn. Paris, UNESCO Publishing. <http://www.unesco.org/culture/en/endangeredlanguages/atlas>. (22 January, 2017.)
- Niedzielski, Nancy A. & Dennis R. Preston. 2003. *Folk Linguistics*. New York: Mouton.
- Onishi, Isao & Daniel Long. 1997. Perceptual dialectology quantifier (PDQ) for Windows. <http://nihongo.hum.tmu.ac.jp/~long/maps/perceptmaps.htm>. (22 January, 2017.)
- Öpengin, Ergin. 2013. *Clitic/affix interactions: a corpus-based study of the person marking in the Mukri variety of Central Kurdish*. Paris, Université Sorbonne Nouvelle & Bamberg: Universität Bamberg dissertation.
- Öpengin, Ergin. 2016. *The Mukri Variety of Central Kurdish: Grammar, Texts, Lexicon*. Wiesbaden: Reichert.
- Öpengin, Ergin & Geoffrey Haig. 2014. Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish Studies* 2(2). 143–176.
- 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.
- Östman, Jan-Ola & Leila Mattfolk. 2011. Ideologies of standardisation: Finland Swedish and Swedish-language Finland. In Tore Kristiansen & Nikolas Coupland (eds.), *Standard Languages and Language Standards in a Changing Europe*, 75-83. Oslo: Novus.
- Postlep, Sebastian. 2015. *Wahrnehmung von Sprache und der kommunikative Raum – Möglichkeiten und Grenzen einer raumbasierten Auswertung von Perzeptionsdaten*. Paper presented at the Workshop zur Perzeptionslinguistik, Universität Wien, 24. Oktober 2015.
- Preston, Dennis R. (ed.). 1998. *Handbook of Perceptual Dialectology*. Amsterdam: John Benjamins.

- 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. 2010. Perceptual Dialectology in the 21st Century. In: Christina Ada Anders, Markus Hundt & Alexander Lasch (eds): “*Perceptual Dialectology*”. *Neue Wege der Dialektologie*. Berlin/New York: de Gruyter, 1-29.
- Preston, Dennis R. & George M. Howe. 1987. Computerized studies of mental dialect maps. In Keith M. Denning, Sharon Inkelas, Faye McNair-Knox, John 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.
- Robinson, Vincent B. 1988. Some Implications of Fuzzy Set Theory Applied To Geographic Databases. *Computers, Environment & Urban Systems* 12(2). 88-97.
- Sheyholislami, Jaffer. 2015. The language varieties of the Kurds. In Wolfgang Taucher, Mathias Vogl & Peter Webinger (eds.), *The Kurds. History – Religion – Language – Politics*, 30-52. Vienna: Austrian Federal Ministry of the Interior.
- Smith, Michael J. de, Michael F. Goodchild & Paul A. Longley. 2015. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools - Fifth Edition*. <http://www.spatialanalysisonline.com>. Winchelsea: The Winchelsea Press.
- Stuart-Smith, Jane. 2011. The view from the couch: Changing perspectives on the role of television in changing language ideologies and use. In Tore Kristiansen & Nikolas Coupland (eds.), *Standard Languages and Language Standards in a Changing Europe*, 223–239. Oslo: Novus.
- Tobler, Waldo. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46. 234–240.

Zazah, El-hadi & Jacky J. Desachy.1993. Symbolic and Numeric Data Analysis in a Geographic Information System: A Fuzzy Neural Network Approach. In Erich P. Klement & Wolfgang Slany (eds.), *FLAI93, Conference Proceedings*. Vienna.

**ENDNOTES**

1 Only rough estimates are possible because no linguistic (or ethnic) census data on the Kurdish language and people have been available from Turkey since 1965.

2 Roman is used for Kurmanji in Turkey and Syria; Cyrillic was used in Soviet Armenia; and modified Perso-Arabic is used for Sorani in Iran and Iraq.

3 We are not arguing for a unified standard. There are plenty of examples worldwide which show that having more than one standard need not pose a problem, e.g. Nynorsk and Bokmål in Norway.

4 ‘Our knowledge of the S.Kd. [i.e. Southern Kurdish] dialects is still incomplete, but it is sufficient to show that they differ almost as much one from the other as they do from their northern kin’ (MacKenzie, 1961:79). Belelli’s (2016) recent presentation of a morpho-syntactic classification of Southern Kurdish was entitled ‘Southern Kurdish Dialectology: where to begin?’.

5 Belelli (2016) also considers the core Laki dialects as separate group from Southern Kurdish. The classification of Laki as part of the Kurdish continuum or not also affects how far southeast the Kurdish speaking language area stretches. There is a considerable discrepancy between the Hassanpour (1992) and Öpengin (2013) map we use, and Fattah’s (2000) and Belleli’s (2016) maps.

6 This issue also arises in PD studies which ask participants to indicate areas where people speak differently and then label their line drawing. In this version of the “draw-a-map” task respondents have the option of drawing an area without labelling it; the researchers analyzing the data then need to decide what to do with these unlabelled lines (see Evans 2013). Montgomery (personal communication) does not count unlabelled areas in his analysis.

7 Questionnaires 9, 13, 22, 29, 30, 64, 68, 70, 71, 73, 94, 97, 104, 115, 148, 149.

8 Questionnaires 1, 8, 12, 13, 23, 24, 28, 29, 30, 39, 50, 62, 64, 68, 70, 71, 85, 93, 94, 97, 99, 104, 107, 112, 120, 148, 149, 156 and 168.

9 The concept of ethnolinguistic vitality (V) denotes the property ‘which makes a group likely to behave as a distinctive collective within an intergroup setting’ (Giles et al., 1977:307). The higher V is, the better the chances are for the maintenance of this group over time; the weaker V is, the more likely it is to cease to exist through assimilation. To the extent a particular language is the main focus of a group’s identity, V is an important indicator of possible language shift or maintenance.

10 The levelling of differences among what was at first a conglomeration of varieties, often leading to a new variety (Kerswill, 2002:680-689).

11 Most mother-tongue speakers of Kurdish younger than 50 years of age also tend to speak, read and write the official language of their country of origin (Turkish in Turkey, Persian in Iran, Arabic in Iraq and Syria).

12 With outright prohibition of Kurdish (e.g. Turkey 1925 - 1992; Iran 1925 - 1941; Syria 1965 - 2011), to various degrees of tolerance (e.g. Iran 1942 - present; Turkey 1992 - present; Syria mid-1930s - early 1960s), and official recognition (e.g. Iraq 1930s - present [local level] and 2005 – present [national level]); see Sheyholislami, 2015).

**TABLE (columns/rows are not to be altered)**

<b>Variety</b>	<b>Raw recognition figures</b>	<b>Percentage</b>
Northern Kurdish / Kurmanji	186	100 %
Central Kurdish / Sorani	171	92 %
Southern Kurdish (Kelhori, Feyli, Kirmashani etc.)	103	55 %
Zazaki	161	87 %
Gorani	6	3.2 %

Table 1: Raw recognition figures

FIGURES

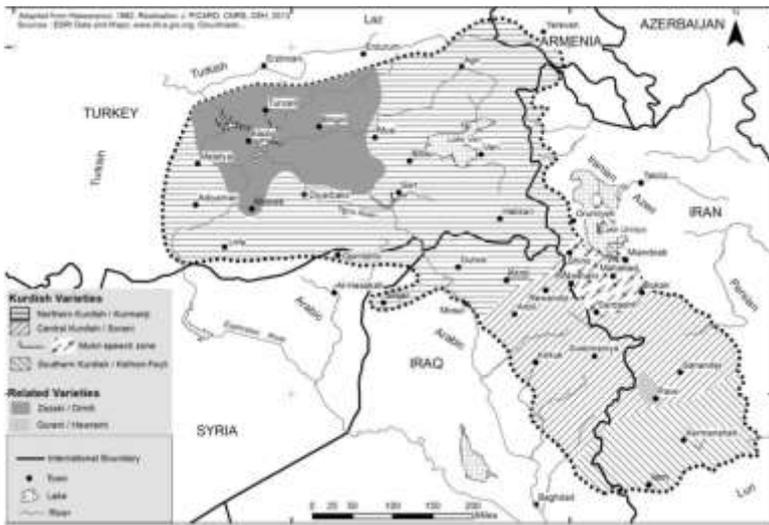


Figure 1 Haig\_Oepengin 2014

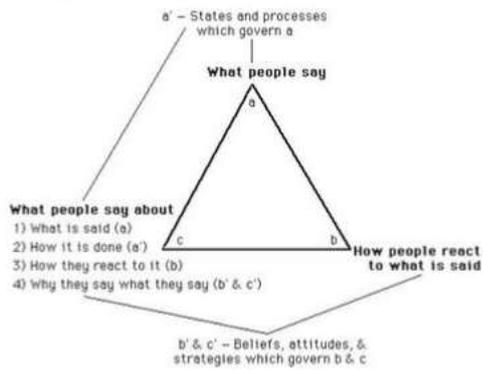


Figure 2 Preston 1999

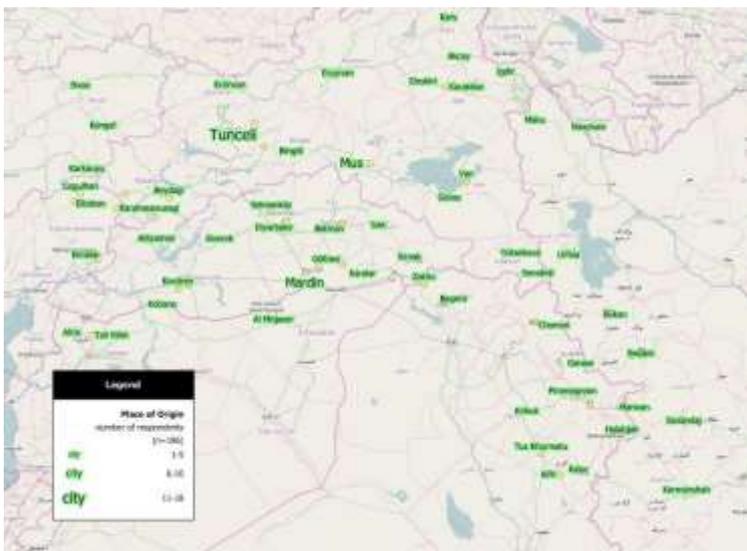


Figure 3 Place of Origin



Figure 4a male\_Q131

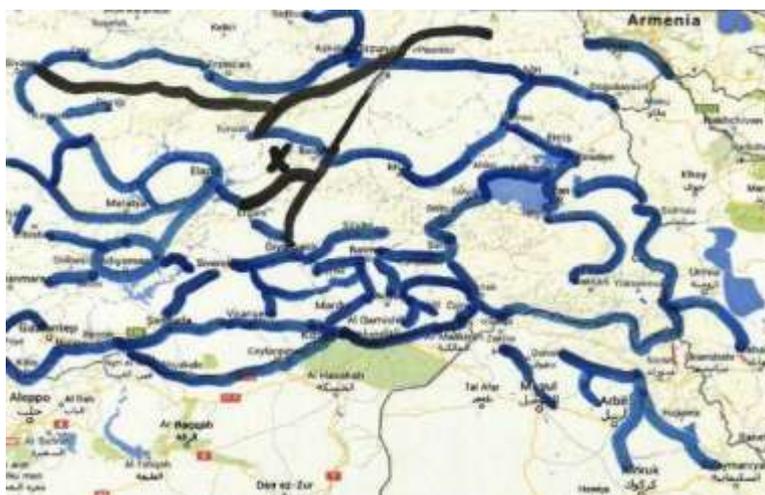


Figure 4b female\_Q150

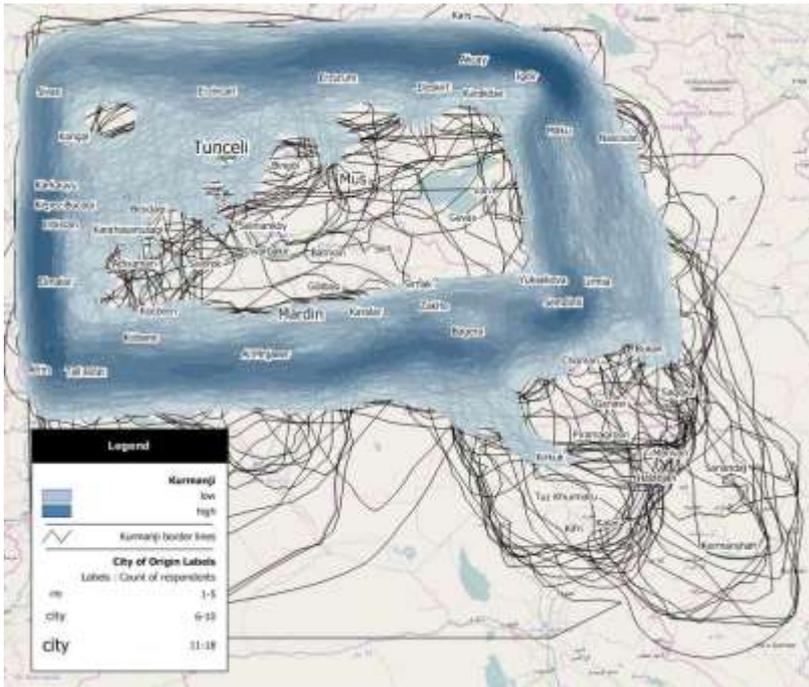


Figure 5 Kurmanji line density



Figure 6 Q029 extract

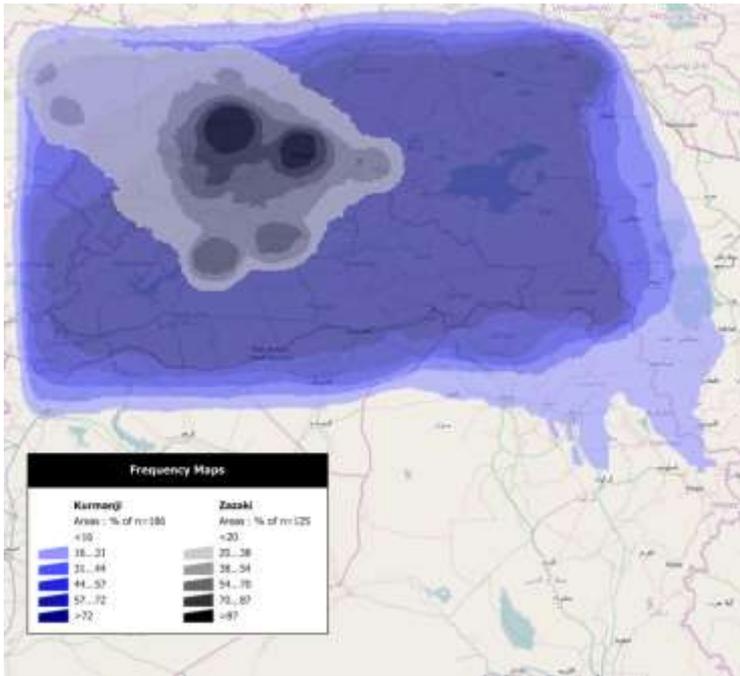


Figure 7 Frequency Kurmanji, Zazaki

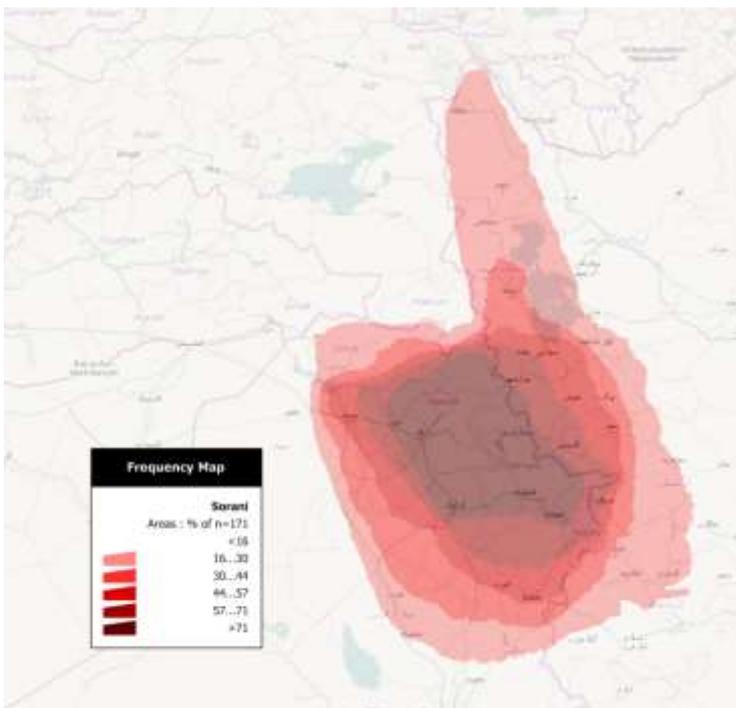


Figure 8a Sorani

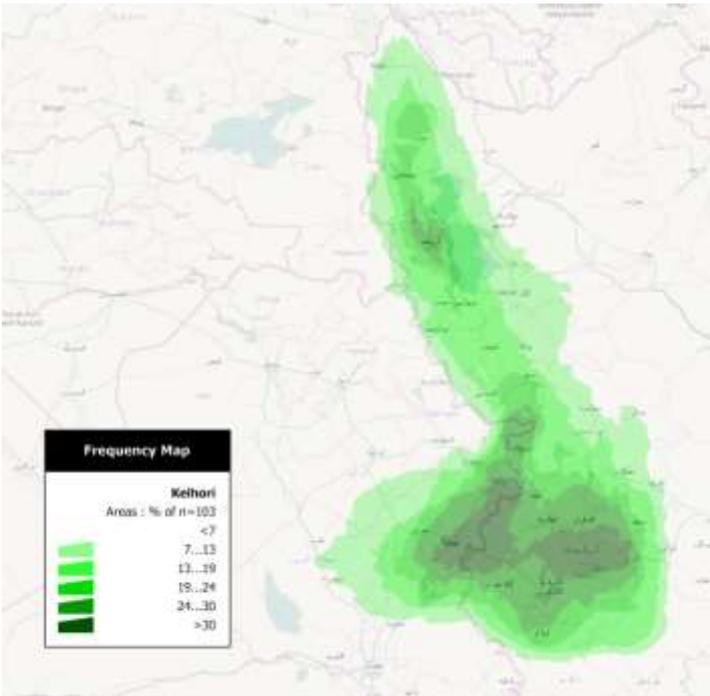


Figure 8b Kelhori

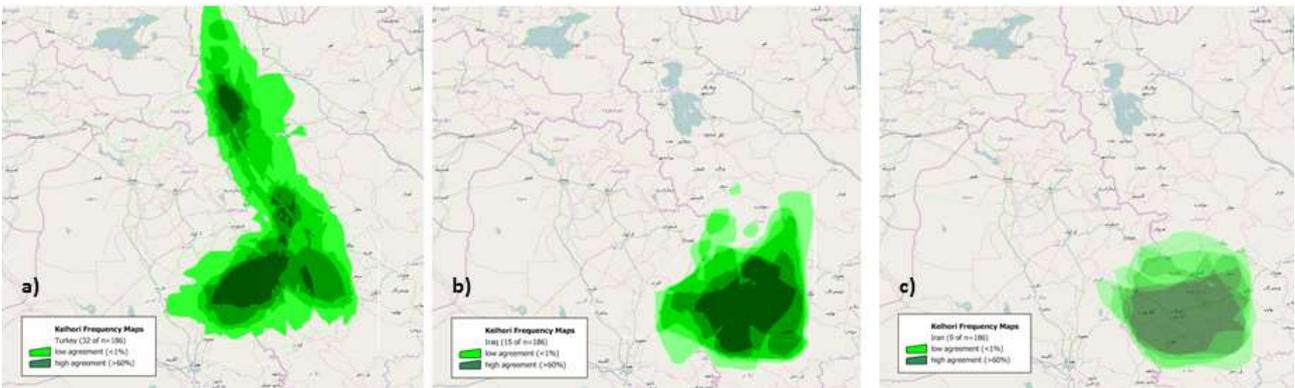


Figure 9 national splitting Kelhori

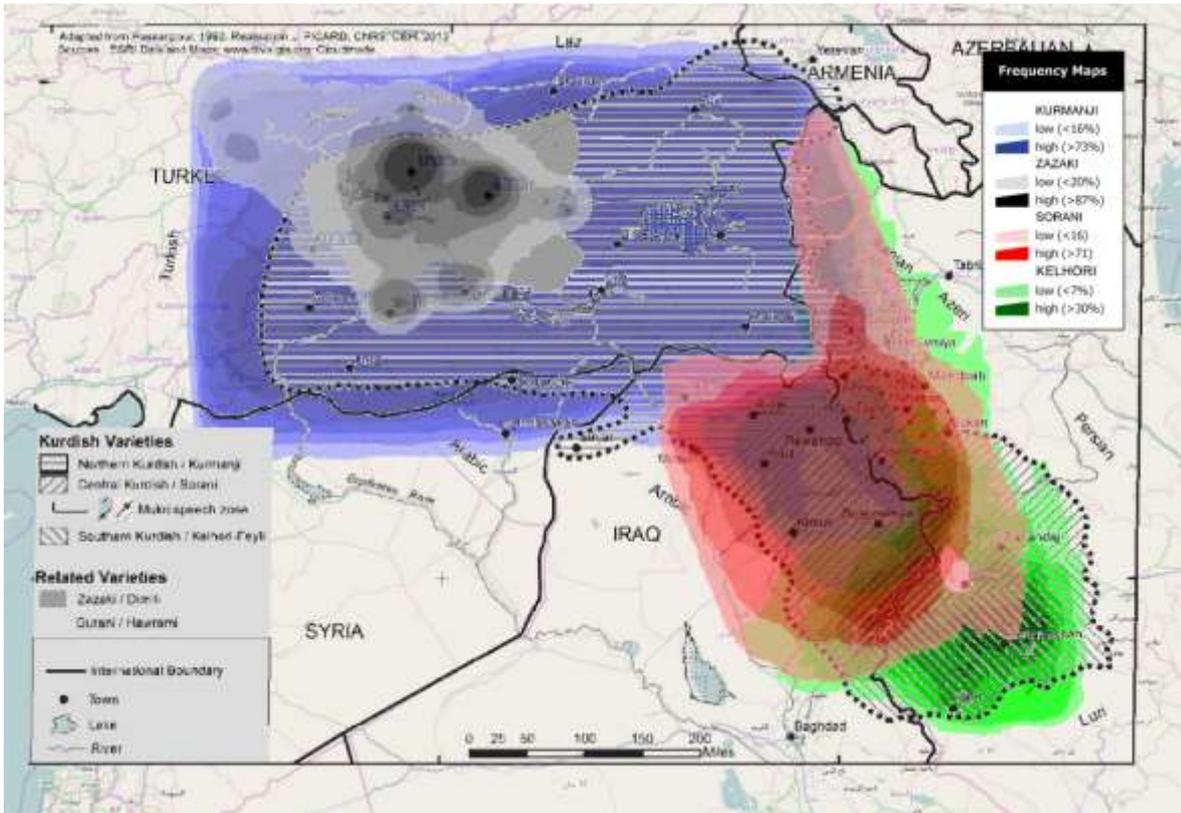


Figure 10 Kurdish varieties Öpengin

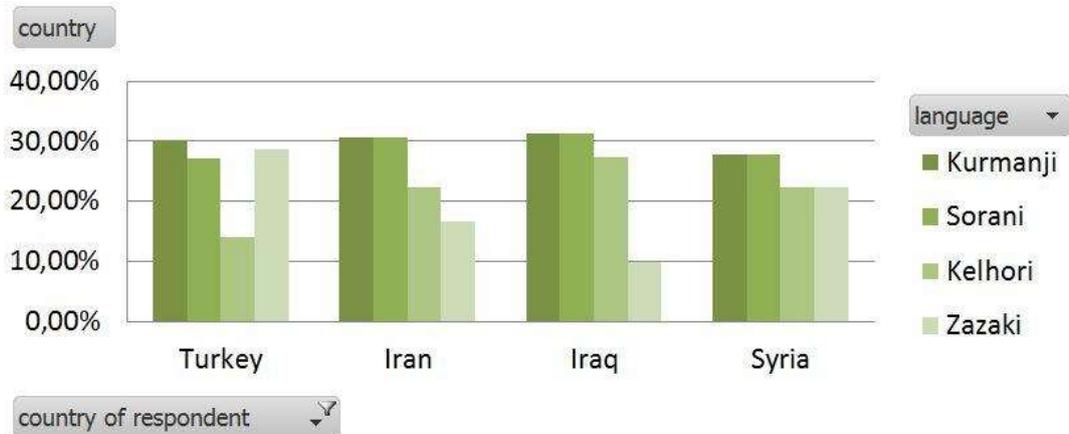


Figure 11 recognition levels