

This is an accepted version of an article which will be published in Language Documentation and Conservation: <http://nflrc.hawaii.edu/ldc/>

Downloaded from SOAS Research Online: <http://eprints.soas.ac.uk/21930/>

Map Making for Language Documentation and Description

Lauren Gawne
NTU, Singapore

Hiram Ring
NTU, Singapore

Abstract

This paper introduces readers to map making as part of language documentation. We discuss some of the benefits and ethical challenges in producing good maps, drawing on the GIS literature. We then describe current tools and practices that are useful when creating maps of linguistic data, particularly using locations of field sites to identify language areas/boundaries. We describe a basic workflow that uses CartoDB, before describing a more complex workflow involving Google Maps and TileMill. We also discuss presentation and archiving of mapping products. The majority of the tools identified are open source or free.

Key words: Map-making, GIS, language map, cartography

Map Making for Language Documentation and Description

1. Introduction

Linguists are often concerned with representing linguistic data so that others can easily see and understand what a linguist is attempting to describe. One graphic that can communicate the geographic reality of a speech community is the language map. People are used to processing geographical and spatial cues on a map, and maps can be extremely helpful to illustrate the geographical spread of one or many language varieties. However, language maps such as those found in high school atlases can give the erroneous impression of a clear relationship between nation-states and homogeneous national languages (Mackey 1988: 24-25), while even detailed language maps of some linguistically-diverse areas can obscure small languages (Dahl & Veselinova, 2005). The majority of the world's languages are under-documented, and thus our knowledge of where these languages are located is also lacking. Often the location of language-speaking populations is known only to the speakers themselves, as well as specialist researchers (Dahl & Veselinova, 2005). The ability to make maps has also been limited to those with access to the technical skills and capital, creating an imbalance in who can claim the right to be represented (Peeters 1992: 7).

All maps are not created equal. Some maps have entirely too much information, while others have too little. It can be difficult at times to know what a map is attempting to illustrate, and particularly with linguistic maps there are fine-grained distinctions that (if included) can cloud or confuse the graphic content. It should always be understood that geographical boundaries on language maps are not the 'hard' boundaries suggested by a line, but rather more variable, dependent on speaker mobility and interaction.

Maps are useful for indicating language (or dialect) locations and features in a clear, concise and graphical format. But as noted above, many maps relating to languages or dialects are out of date or based on incomplete information. This is particularly the case in remote or inaccessible areas where little is known about a particular language group or variety. This paper introduces readers to language map making. This includes some of the considerations necessary for making language maps, and the introduction of techniques and methods that have made the task of map creation much easier in the last few years (see Dahl & Veselinova 2005, who a decade ago observed a lack of an easily accessible tool for language mapping). These techniques are centered around making use of Global Positioning Service (GPS) data with free software such as Google Earth and TileMill and online services such as CartoDB, and converting the data into points, lines and polygons to overlay on a digital map of the world. We first discuss some considerations regarding map making (§2). We then identify ways of gathering GPS data (§3.1), how to convert the data into a usable format (§3.2), how to use it with various free mapping software (§3.3-3.5), and finally how to share your map, including exporting the map as a print-ready image (§3.6).

2. Considerations regarding map making

In this section we will discuss the nature of linguistic map making, and things that will need be addressed before starting the map making process. This includes both practical and ethical considerations. Even though language documentation often involves site-specific communities who have lived in an area for a long period of time, there has been a general lack of discussion regarding map making in the language documentation literature. Bowern (2008: 111) mentions the need to take a device on fieldwork to collect GPS data and, where possible, using existing high quality maps to mark features. The ability to generate your own high quality maps from the GPS data collected has simplified this process somewhat.

To demonstrate the researchers' own map making development, Figure 1 and Figure 2 are the maps used by Author1 in her completed PhD dissertation. This outline of Nepal was traced from an atlas, and then MS Paint was used to label the relevant language varieties. The villages in Figure 2 were identified in Google Earth using satellite imagery and the road traced in approximation. These maps offer limited contextual

information, and are limited in their reproducibility. We are unable to reproduce the map Author1 was using earlier, as it was an image with copyright restrictions, but it also contained a great deal of irrelevant information and was not very visually appealing. Figure 3 is a map that Author1 created using TileMill and GPS data she collected. Unlike Figure 1 it is easily re-purposed and replicated, and looks much more professional. We will demonstrate how to create a map like that in Figure 3 in §3.5 below.

<Figure 1>

<Figure 2>

<Figure 3>

2.1. Currently available maps

Linguists generally use one of a number of strategies for making language maps for presentation and publication. Some may use existing resources like Google Maps, dropping digital pins to locate languages, or may use image editing software to add that information to an existing map. While this serves the necessary functional purpose, it is limited. Firstly, the data is hard to re-purpose into another map, unlike digital mapping where GPS data points are created separate to the styled map. Secondly, every map is designed with a specific purpose in mind (c.f. MacEachren, 1995), so co-opting an existing map may mean including extraneous information. Thirdly, there may be limits to reproducing such maps in publications for reasons of copyright and visual clarity (i.e. graphic resolution). Finally, these maps are often not particularly aesthetically appealing. Speakers of endangered languages are often minorities, rendered invisible on official maps. Making targeted and professional quality maps for these languages can help increase their perceived status, both within the community and among the general public.

The most comprehensive language maps that currently exist are those created by the Summer Institute of Linguistics (SIL), which are updated periodically for editions of the Ethnologue (Lewis, Simons & Fenni 2013), based on the World Language Mapping System (WLMS <http://www.worldgeodatasets.com/language/>). This information is also used to provide the GPS information for ISO 639-3 labelling,¹ which is used by services such as Glottolog (<http://glottolog.org/>) and OLAC (<http://www.language-archives.org/REC/discourse.html>). These maps can be useful, especially the ISO 639-3 point data, however they also have some limitations. The first is that individual linguists and specialists would often prefer to have a map that illustrates different features than what these larger data sets offer. Linguists (and the larger linguistic community) would benefit from a way to create new maps based on their research and knowledge of a particular area, using GPS data to clarify and assist their efforts. These maps could then be posted online or be printed in publications. The second limitation is that these maps cost money to obtain, and are still licensed products of WLMS. Current publicly available software is free to use, and builds upon open access data. This means that maps created by researchers can be easily distributed online and in print without the licensing limitations that pre-made maps can have.

2.2. Ethical considerations

Language maps are problematic constructions. As static and selective representations of complex phenomena, they can act as both representations of power and tools of power (Luebbering 2013: 49). As Peeters (1992:7) notes, there is a certain authority given to a place or a people by putting them on a map. For endangered language communities this may be a good thing, giving them recognition and status that they may never previously have experienced, even if those maps are only used within the community and research publications. But mapping can also be used as a tool of division, or evidence in territorial claims. Therefore, linguists need to think very hard about the potentially exclusionary or over-expansive story their map might tell.

It is also worth considering whether the community will want their location mapped and distributed. While some groups may approve of this activity, it should not be assumed that everyone wants to be “on the map”. It may be appropriate to only indicate

¹ <https://www.ethnologue.com/about/language-maps> (visited 6th of May 2015)

their location very generally within the larger nation or state where they reside, and then present a more detailed map that doesn't necessarily give traceable GPS information.

It is also worth considering whether all local geographic features are appropriate for mapping. Some communities may have sacred spaces or totem lands that they do not wish to have located. Including the community in any GPS co-ordinate plotting and sharing early map drafts with them can help ensure that they are comfortable with the amount of information you are building into a map.

2.3. What to put on the map

Choosing what to put on a map can change how it is read. In this paper, we assume that the main target to be mapped is the location of one or more language speaking communities. But there may be other geological or political features that need to be considered. You may choose to include national or local borders, to indicate the relationship between a language group and larger geo-political constructions. Including other language groups or urban centres may help contextualize the linguistic landscape.

Within the target group being mapped, you may choose to map them as a single unit, or break the groups down further. You will need to decide on what grounds you wish to make such distinctions, and what divisions that may indicate in a map. Some groups that you have good reason to consider as separate dialects or languages on linguistic grounds may have strong social or cultural reasons to be identified as a single group. Addressing these considerations may require separate maps and/or careful planning of colors and shading to clarify such distinctions.

Of course, the location of language groups may not be the only language-related mapping target. As Lubbering (2013:44) notes, there is a diversity of topics and variables, and their geographic arrangement, when it comes to languages. With the tools we demonstrate in this paper, it would also be possible to create maps that illustrate lexical variation between groups, density of language maintenance in particular areas or areal contact.

Each language context will present its own challenges and points of interest for the task of mapping. Any map is inherently controversial, as no one map will satisfy all users (Peeters 1992: 6) but there are some generalizable considerations to be made. For this paper we are assuming that the map will be the location of settlements in which the target variety is spoken, be they individual homesteads, small villages or large towns and cities. You will have to decide if you want to mark all of these as individual points, which may be preferable if there are a manageable number of tokens, or to indicate the language area with a single polygon (an enclosed shape made from a collection of points, see GPS section below for more detail). You may wish to represent each village as a separate point, but then have larger polygons representing the influence of different contact languages or related languages.

2.4. The challenges of borders in language mapping

There are challenges with assigning languages to space, and delimiting between language (Williams & Ambrose 1988). One particular challenge is how to represent the borders of different language groups. Unlike natural geological features, which are inherent to the landscape, and geopolitical borders, which are measured with great precision, the fluidity of language, and the nature of language contact can make representing the edge of a language group a particular challenge.

Mapping methods with one-language-per-area and clear borders between language groups are at odds with our understanding of the way languages can often co-exist in an environment (Luebbering 2013: 41). Dahl & Veselinova (2005) suggest that with modern mapping technology and small language speaking populations the best way to represent endangered language communities is the map at the level of individual settlements. This allows us to see the specific location that languages are linked to, rather than forcing ourselves to carve specific dominions for each language. Even this kind of boundary-marking may distract from the reality of language contact, bilingualism, trade or marriage movements in an area (Mackey 1988: 26; Williams & Ambrose 1988:110; Luebbering 2013: 44). These researchers also note that a focus on language boundaries distracts from the fact that there are many interesting features of the contact between language groups. Building your own map allows you to represent the areal reality in the best way possible, however dealing with borders will always be a particular challenge.

2.5. Maps as objects

The status of maps as objects has changed drastically in the last decade. Peeters (1992: 7) and Dahl & Veselinova (2005) note that static maps must, in their nature, exclude and overcrowd information. While all maps are selective snapshots of a moment in time, modern digital mapping offers some solutions to the problems inherent in static maps.

Digital maps allow collected data points to be represented in different ways, so that multiple maps can be made to demonstrate the complexity of a linguistic situation. Also, each map feature set is produced on its own layer of the map, so additional information can be easily toggled on/off. Digital maps can include a zooming feature, allowing for detailed village-by-village views on closer zoom and more generalized polygons at greater distance. These digital maps are obviously not suitable for print publications, and there are still the same concerns about what is included and excluded, however they allow for more complex language stories to be told.

2.6. Possibilities of GIS

In this paper we demonstrate how to make basic maps for locating the distribution of languages in space. This is not the only use of mapping that linguists may find useful. Geographic Information Systems (GIS) allow not only for the collection and representation of geographic data, but analysis as well. While exploration of the possibilities of GIS for language documentation projects is beyond the scope of this paper, we hope that in giving language documentation researchers a basic introduction to language mapping we can open the possibility of further collaboration with GIS researchers. Language mapping and geographical analysis has central to areas of socio-linguistic research including dialect studies (Trudgill 1983). Hoch & Hayes (2010) give a good overview of GIS/linguistics research and discuss potential analytical possibilities in the field.

If linguists have not made full use of the possibilities of geographic research, it is comforting to know that many geographers have long neglected the complexities of mapping language (Williams 1988: 1). This situation has changed, with the development of the field of geolinguistics (Mackey 1988; Williams 1988; Luebbering 2013), however there is still much possibility for developing geolinguistics within the domain of endangered languages. Hildebrandt & Hu (2013: 57) offer one example of such a collaboration, creating a multimedia mapping of languages and their varieties in the state of Manang in Nepal, which is helping to “visualize and re-think the complex settlement patterns and histories of Manang and the current socio-economic pressures on the vitality scenarios of the languages spoken there.” Linguists who understand the basic process of language mapping will be better equipped for possible collaboration with GIS researchers.

3. Making maps

In the following sections we will work through the process of making maps, from GPS data gathering and processing to using that data to build a map. It is possible to make many different maps, so it is important you have decided exactly what you want your map to represent before starting the process. We begin with an explanation of how to gather data using GPS (§3.1.), and how to make this data ready for mapmaking (§3.2.). We then introduce popular software for map creation (§3.3.), which vary in terms of the degree of flexibility they offer, and the amount of coding required to manipulate the final product. We introduce a simple workflow (§3.4.) using CartoDB, which allows you to add and manipulate data, and has many predefined map features. Following this we introduce a more complex workflow (§3.5.), involving editing data in Google Maps and styling the final map in TileMill. This option requires use of CartoDB, a simple markup language based on CSS and built specifically for map making. We conclude with a discussion of the distribution of electronic and static maps, attribution of copyright and archiving (§3.6.).

3.1. Gathering GPS data

GPS data is the primary means of locating positions on earth in the 21st century. Satellites orbiting the earth have allowed us to identify particular points in relation to a global spatial map since at least the 1980s (see Parkinson 1996; Guier & Weiffenbach 1997; Pellerin 2006). Only in the last 10 years, however, has this become commonplace for consumers, with dedicated GPS devices becoming accessible for hikers, trekkers, ocean and land navigation. In the last 5 years GPS has become an essential part of smart phones and is used by various apps to identify users in relation to nearby businesses and advertising, or for navigation in apps such as Google Maps, with GPS functionality generally being referred to as ‘Location Services’.

Smart phones provide a useful tool for the map-making field linguist. Smart phone use is increasing in developing countries (where many of the world’s unwritten and under-described languages are spoken), and carrying a smart phone will no longer make you look out of place. With location services enabled for photographs taken on a smart phone device, the GPS information is automatically recorded when the picture is taken. It is culturally acceptable and even expected by most local individuals that a foreigner would take pictures at every opportunity, and there are generally few restrictions on what can be photographed and where, although sensitivity should of course be exercised. There are many free GPS tracking apps that can store and display real-time GPS information. Thus, collecting GPS data that can be used in mapping does not have to be an additional task in the linguist’s workflow, nor interrupt other activities.

Location services are available on any modern mobile phone. On an iPhone, this is accessed by selecting ‘Settings > Privacy > Location Services’ and turning Location Services to ‘On’. Then location services for individual applications can be enabled. To have photos that you take on your device automatically store GPS coordinates, for example, select ‘Camera’ in under Location Services and turn on the location services for the built-in camera. Other phones may vary in how the location services are enabled for photos, but the process is similar – searching ‘How to enable location services for camera on [Name of Your Device]’ using Google should give step by step instructions for your particular smartphone.

With location services enabled for photos on your smartphone, GPS points are stored whenever you take a photo. Keep in mind that this information is not shared online publicly unless you upload the photo to a public site such as Flickr, Facebook, Instagram, or a personal website. Here your particular privacy settings will determine the photo’s accessibility and unless you have removed the GPS information from the photo, the GPS coordinates will be included, as they are embedded in the image file itself. Most people do not mind this, but it is worth being aware of, particularly if you or community members have privacy concerns.

Collecting GPS data with photos also helps make it easy to recall the exact location of a GPS point – in terms of mapping, you should plan ahead of time what points you want to capture and the metadata to include with each point. This is easier with a photo – you can take a picture of informants you are working with, a village sign, or a particularly recognizable geographical feature, landmark, or building. These visual elements can help you link individual photos with metadata that help to identify the particular layer you are building with the help of that point.

A dedicated GPS device can also be extremely useful. The benefits of a dedicated device are that they generally have longer battery lives, are more robust and can manipulate data points on the go. However, they are not as commonplace and as such may attract more attention than comparable use of a smart phone. They also may require additional software to transfer the data to your computer when mapping. However, for those who are doing a lot of walking or who want to plot boundaries or roads manually, these may be the better investment. Many models, particularly those for bushwalkers, are quite robust and may survive fieldwork better than a modern mobile phone. They can also be used to demonstrate to people the task of collecting data for mapping, therefore starting the community discussion of what should be mapped.

Unlike for smart phones, GPS trackers may allow you to mark specific locations, but may simply auto-generate a numeric value for each of the recorded GPS points. Ensure you test your GPS device before taking it to the field. You may need to record your own metadata for the names of the enumerated points so that you can remember what they refer to when you return to your mapping data at a later date.

It should be noted that GPS data exists primarily as points. Vectors are a collection of points that when connected form lines. Polygons are vectors in which the

first and last points are linked to form a shape that can be filled with a color of varying opacity. Keeping this in mind can help the linguist decide what kind of GPS data (and how much of it, i.e. how frequently points should be taken) is necessary to help identify the points and boundaries for creating a map to display the kind of information that the linguist is interested in. For example, creating a map of villages where each village is represented as a single point will require only one GPS data point per village, while representing clusters of groups as polygons will require a lot more GPS data points to be taken in order to produce an accurate map.

It is also possible to take a more armchair approach to aggregating GPS data. Services like Google Earth (discussed below) now offer high quality satellite images of many places on the earth's surface. Remote hillsides may not have the same detail as a major urban centre, but once basic GPS locations have been established it may be possible to mark out other villages, landmarks or features based on the satellite photography and familiarity with local landmarks rather than using a GPS device in situ.

3.2. Uploading, editing, cleaning GPS data

Removing GPS data from a device will vary from product to product, but many will have support software, or online documentation for how to do this. Whatever the file type the GPS program works with, you should find a workflow where the final product is a KML (or potentially KMZ) file. KML is an acronym for 'keyhole markup language', a specific arrangement of XML data that many of the mapping programs uses to store GPS coordinates and other information. If you open a KML file in a text editor you will see the XML tags being used to organize the data. A KMZ file is a zipped file that contains a KML and additional supporting files. This zipped file is often smaller than a KML because of the compression, and many programs that can read KML also support KMZ.

When managing mapping data you should think in terms of layers. 'Layers' refer to demarcations that you intend to build in to your map. These are collections of similar information that you will color or shade the same, in order to contrast with other layers. Layers are the primary way that map data is organized and identified in most mapping software. Maps may have a single layer that identifies oceans, another that identifies land masses, and others that identify geographical features such as rivers and mountains, or those that identify political boundaries such as country, state, and local borders. Geographical layers may be shaded or colored to give the impression of elevation, while political boundaries are often lines of different width or type.

Thinking in terms of layers will help you identify and separate the kind of GPS information you want to represent on your map. As mentioned in §3.1, this will likely be a series of points to begin with. However, you may need several different sets of points – a set of village points for one language variety, a set of village points for another variety, a line or periodic points to identify potential trade routes, or a set of points to identify marketplaces or cultural/historical points of interest. With careful metadata assisted by photos that you take along the way, these GPS coordinates become a rich source of information to support future maps and mapping projects. If you take points without photographs remember to make note of the location of each point (for example the village name, or the route name).

One of the most useful tools for uploading, editing and cleaning GPS data is Google Earth, a free cross-platform program (www.google.com/earth/). Google Earth will accept data from a wide range of formats, which can then be manipulated, and allows KML export. Many GPS devices will allow direct upload to Google Earth. Google Earth allows you to move data points, create data points and export this for use in other programs. It is possible to take a screen capture of Google Earth to use as a display map, however it is not a clear or elegant way to represent geodata and therefore we only use it as a first step for editing and manipulating. For example, Figure 4 is a screenshot of the Google Earth representation of the Yolmo villages in Figure 2. As you can see, the pushpins are not very attractive, and while the satellite images of the terrain look fine here, may be too much information for other mapping purposes.

<Figure 4>

If you have taken pictures with embedded GPS data, you can download or transfer your pictures to your computer, ensure that they contain GPS data, and use a program such as Geotag (www.geotag.sourceforge.net/) or Photo KML

(<http://www.visualtravelguide.com/Photo-kml.html>), both free, to convert your images into sets of GPS points that you can open in Google Earth or another map editor. Geotag is a free, Java-based program for getting GPS data (in EXIF format) from images and exporting the GPS information into several formats, including KML. To use it, simply download the Java program from the SourceForge site, make sure you have the latest version of Java, and run it. To add data select ‘File > Add image...’ or ‘File > Add images from directory...’ To export data, select the files that you wish to group together, right click, and select an export option such as ‘Google Earth > Export selected images’. Then select a location on your computer to store the data and type a name for the KML file.

Several things can ease the creation of these layers from GPS photo data. The first is to organize your photos into folders. Create folders for each layer that you want to build into your map, and copy or move the photos that are associated with that layer into the proper folder. Then open each folder individually within Geotag and save each group of images in the Google Earth format. Here you will notice that the KML file links to the location on your hard drive where your image is stored, and that all the images in a layer are grouped together under the heading ‘Geotag’. This is information that you can edit in Google Earth. You can change the names/titles/headings, delete the images, and even adjust the location of each of the points. To export from Google Earth the easiest process is to right-click on the point, or folder of points you wish to export, select ‘email’ and email them to yourself. They will then be in an email as a downloadable KMZ attachment.

An alternative to using KML files is to create a list of longitude and latitude values for points you want to map and save them as a CSV (comma separated values) file in a program like Excel. Each layer should have its own CSV file. You may already have these values, or have taken them from Google Maps or another mapping platform. An example for locations in a CSV that will be familiar to many readers is given in Table 1. The X column is for longitude and the Y column is for latitude.

<Table 1>

CSV tables can only be created for point data, not for line or polygon data. If you have a larger number of points it may be more time-consuming to generate than working with Google Earth.

3.3. Map making using existing software

Below we identify some of the useful mapping software that is freely available. Some of these require an internet connection, while others can be downloaded for use offline. These programs also vary in terms of user friendliness, i.e.: visual simplicity, variety of menu options, format of data, intuitive nature of interface, and access to underlying code. The software that we review briefly here are, in order of complexity for the user from simplest to most technical; Google Earth, CartoDB, TileMill and QGIS.

Google Earth, already introduced in §3.2 above, is freely downloadable software for visualizing the earth as a globe and for zooming in to specific points or geographical features. It incorporates GIS data and allows users to add and share points and descriptions, and is particularly useful for identifying places you have been on the globe through geographical features. Users can import, draw and export points, lines and polygons with export to the KML and KMZ formats. For those who have an internet connection but do not want to install Google Earth, Google also offer My Maps as part of Google Maps (www.google.com/mymaps). This has fewer features than Google Earth but also allows for the creation of points or lines in layers that can then be exported to KML format.

CartoDB is an online map-making tool, primarily for digital maps, although they do also have a static map export option. CartoDB requires users to register an account, with the free user accounts limited in their storage size. Maps made in the free accounts are also currently limited to only four layers, which will constrain the number of features you can represent on a single map. All maps made with the free version are publicly viewable, so may not be appropriate for mapping sensitive data. Paid accounts offer more storage and private maps. CartoDB allows you to upload existing data sets, for example those created in Google Earth, but also allow you to create, or modify, data sets within the program, which is not possible in map design programs like TileMill. The data created or modified in CartoDB can also be exported again. CartoDB offers design

interfaces that offer less flexibility than fully featured software but allow for the quick creation of maps with no knowledge required of scripting languages.

TileMill is specifically designed with the non-specialist cartographer in mind, and focuses on allowing users to create visually pleasing maps. Unlike CartoDB it is not possible to edit datapoints within the program, they must be edited in another program such as Google Earth. Built mainly to create maps used in web browsers, a coding windows allows users to fine-tune adjustments to layers using a special kind of JavaScript called CartoCSS. While it is not necessary to fully learn this code, users may find that learning the basics will make their maps more visually interesting and understandable. Tutorials are available online, some of which are simpler and more accessible than others.

TileMill also has a dedicated server which hosts GIS map files and importing from websites or your computer is quite easy. With a focus on creating interactive web-based maps, the export options and user-friendliness are not ideal for users who are primarily interested in creating print-ready maps; exporting may require some finessing. As another caveat, TileMill is no longer in active development, with the developers having moved to development of Mapbox Studio. TileMill is currently easier to use than MapBox Studio, offers better static map export, and we believe it is still preferable for the kind of language mapping work linguists are likely to undertake. One limitation is that TileMill currently does not work on Mac OSX 10.10 (Yosemite), which may make it inaccessible for some readers. TileMill servers can be created, where the user logs into the program through a web browser, which avoids the limitation for Mac users. It may be possible to talk to your University IT services about setting this up

QGIS is a full-featured professional graphical GIS editor. This program is freely available and runs on all major operating systems. It imports from nearly any mapping format and allows the user to edit and create GIS layers of points, lines, and polygons in visual layout similar to PhotoShop. However, the extreme level of detail and array of menus and options outstrip most basic map making needs. This is an excellent and powerful tool but is much more powerful than the typical linguistic map maker requires. We do not include QGIS in our workflows, as we find that the less powerful programs are still more than sufficient for the needs of language map makers. If you are interested in building more complex GIS projects, learning QGIS may be worth the time, but starting with the programs we discuss will offer good basic training in mapping nomenclature and design.

There are many different ways to create maps, and so we describe two approaches that we have found useful. The first is a “simple” workflow using CartoDB that allows the user to create a basic map easily and also enables further expansion and collaboration, but which requires an internet connection. The second is a more complex workflow that makes use of several different programs, including Google Earth and TileMill, importing data between them and making adjustments along the way.

3.4. Basic workflow: CartoDB

CartoDB allows users to either import existing data in a range of formats, including KML and CSV, or to create and edit data sets within the program. You will need to create a user account before you begin, and all of your data will be stored online, so if you create or modify data in CartoDB remember to download the data occasionally for backup and archiving. CartoDB allows you to create data sets, with each representing one layer of information on a map, and then combine multiple data sets into maps. The top left of the menu allows you to toggle between your data sets and the maps you’ve created - multiple maps can draw on the same data sets allowing you to create different visualisations of the same data very quickly.

When you create a new map CartoDB gives you the option of creating a blank map, or using one or more of your data sets. Data sets can be imported, including KML and CSV, or you can browse the data sets available on the website, although it is likely their data sets will not match your needs. Each data set layer is represented in its own tab on the right of the map visualization. You may use CartoCSS, which we discuss further in the workflow below (§3.5), by clicking on the CSS icon, although this is not necessary unless you have very specific design ideas. Users who prefer to work with a graphic interface should select the paintbrush icon, which brings up the design wizards; remember that you will need to change between tabs to select the correct layer you wish to manipulate. You can move the map around and zoom to achieve the scale of map you are interested in. CartoDB comes preloaded with a number of basemaps, including simple

geo-political maps, satellite data, and styled maps, you can change between them in the ‘change basemaps’ menu on the bottom left of the map. It is possible to import other backgrounds, but it is easier to select one of the options already loaded. CartoDB basemaps cannot be edited. You may find that there are labels or features that you would prefer were not shown. We suggest you attempt to find one of the available basemaps that best meets your needs, but if this is insufficient you may want to consider using TileMill, which allows for complete control over all design elements.

One thing that should be noted about CartoDB is that different basemaps are licensed from different organisations and will have different usage limitations. We recommend using basemaps from Open Street Map, like in Figure 5. The OpenStreetMaps cartography is licensed under the [Creative Commons Attribution-ShareAlike 2.0](#) license (CC BY-SA). This means you can copy, distribute, transmit and adapt the data as long as OpenStreetMaps is given attribution. Maps provided by Nokia in CartoDB only allow for personal non-commercial distribution, which may not be appropriate for publication requirements.

New points can be added in the map view by selecting the ‘add feature’ button at the bottom of the list of icons for each tab. If the data set is a CSV then it will only allow you to add points, as that is all that a CSV can store, but if it is a KML you will be able to add lines and polygons as well. You may chose to create all your data this way, although CartoDB maps do not have the same density of references, or clear satellite images that Google Earth has, so you may wish to create your data there first, and then import it to CartoDB.

With the design wizards the size, color and outline of the data in each layer can be modified. Labels can also be added by selecting the value with the label you want to add. For example, with the US cities in Table 1 I would select the ‘Name’ column. The font color and style can also be modified. You can export the map for the creation of a static product by selecting ‘export image’ near the top left of the map. You can select the pixel width and height, with the export then calculated from the current zoom level and centering of the map. This export does not have the flexibility of TileMill export, and may require you to zoom in and then chose a large pixel area to ensure print quality. The export is in Portable Network Graphics (PNG) format. An export of the map created with the data from Table 1 is presented in Figure 5 below.

<Figure 5>

If the map is to be displayed digitally, the user can also add ‘infowindows’, which is a small amount of text that can be read when a datapoint is selected. This feature may be useful if you have plans to use a map online, but will not be available in print. This text would be an additional column in your CSV or KLM file, and can be added in the data view - it may be a small description of a village, a lexical item you want to compare across different populations, or whatever you like. You can also add a text legend to explain different colors used on the map by selecting ‘legends’ on one of the relevant tabs. This will only appear on the online map, not on the static export. The digital map can be embedded in another website, or distributed via a link. You can use the ‘options’ down at the bottom to change how people interact with the digital version, by toggling on or off a tile, the zoom, sharing options, and other features. The digital version of the map above, including legend and infowindows when you click on any of the cities, can be found at https://lgawne.cartodb.com/viz/5346bba0-02b7-11e5-86da-0e24745e8b53/public_map

Below is a version of the map of Nepal from Figures 2-3 made with CartoDB.

<Figure 6>

A dynamic map using the same data can be found at https://lgawne.cartodb.com/viz/94f75f72-a200-11e4-90b6-0e0c41326911/public_map. The online version uses a Nokia background map, which limits its print distribution. The basemap is a satellite image, which would be too detailed for print, but works satisfactorily for a digital map. Zooming in on the Lamjung Yolmo community reveals the distribution of the villages at zoom levels greater than 10. This means that instead of Figures 1 and 2 being separate maps, in a digital map they become different levels of the same map. The map can also be enriched over time. Mapping data for Kagate villages has been added more recently, and data for the other communities can be added when it is acquired at a later date.

It is also possible in CartoDB to make cluster maps, density maps and choropleth maps, which all serve different functions for representing information. CartoDB has extensive online support, including videos to demonstrate the features available. It is still in active development, so some features may change from this description.

3.5. Complex workflow: Google Maps and TileMill

This workflow describes a method of using My Maps in Google Maps or Google Earth and exporting data to TileMill. The first step is to get your data into Google Maps or Google Earth. This can be done simply by adding points or polygons to a map in My Maps or to a globe in Google Earth. If you have a user account with Google (which can be created easily and is also free), you have the option of saving the map so that it can be accessed from anywhere. Adding points or polygons to the Google map is a relatively simple matter of moving the map until you are viewing the area to which you want to add point or polygon data, and then using the tools in the toolbar.

Above in §3.2 we mentioned some options for getting GPS data from your photos tagged with location data by your smartphone. Unfortunately, Google Maps and Google Earth do not currently support dragging in or importing images with GPS location information, and other third-party applications must be used to get this data into a format that Google will recognize. Geotag and PhotoKML, for example, are free, Java-based program for getting GPS data (in EXIF format) from images and exporting the GPS information into several formats, one of which is KML.

Once your data is in Google Maps or Google Earth, it is a simple matter to export, again in KML format. This file can be imported to TileMill, and will precisely locate your points on the basic map of the world that TileMill provides. The KML file will be converted into a table that you can view within the TileMill interface and export for editing.

TileMill allows you to make adjustments to layers through the use of CartoCSS, a scripting language that can configure color, line width, and opacity, among other things. In order to use these features, some knowledge of the commands and options is necessary. Fortunately, there are a number of tutorials on CartoCSS (i.e. <https://www.mapbox.com/tilemill/docs/manual/cartocss/>) available online for the interested mapmaker.

When opening TileMill, you may first want to adjust your settings in the tab on the lower left (with a gear icon). In this window you can choose which folder your project files will be saved in. TileMill organizes maps based on projects, so under the ‘projects’ tab you will want to select ‘New Project’ to begin building your map. After allowing you to name your project, and with the ‘Default data’ button left checked, selecting the ‘Add’ button will create a basemap of the world with continents demarcated from oceans and countries outlined lightly in blue on the continents. Other layers can be added to this map by selecting the ‘Layers’ tab in the lower left hand sidebar and selecting the ‘Add layer’ button. Selecting ‘Browse’ will allow you to navigate to the ‘Mapbox’ server (the middle button on the menu that comes up) where free GIS data points and layers are stored. If you have map data on your computer (such as those created via Google Maps or Google Earth), navigate to the folder where those files are located, and select the file you wish to add. TileMill supports most map-related files and formats.

Each time you add a layer, the ‘Layers’ tab updates, as does the CartoCSS code on the right side of the screen. This code allows you to make global changes to how a particular layer is represented (styled) in terms of color, opacity, line width, etc. If you have multiple sets of data within a single table (think Excel table or CSV), you can make changes to a single part of the data set according to the names in a particular column. If you do not change the style, your data may simply blend in with the existing layers. You will need to learn specific commands for changing those properties (style elements) that you wish to be displayed differently. Colors are coded by a series of numbers and letters in hexadecimal, for example, which may require an online tool such as ColorHexa (<http://www.colorhexa.com>) for you to pick exactly the color you want for a particular element.

You will also notice that each time you exit and reload, TileMill may default to the large global view when you open your map again. Change this by opening your project and clicking on the button with the image of a wrench in the top right hand corner of the screen next to the ‘Export’ button. These are your individual project settings,

allowing you to shift-select an area of the screen to be the default view, as well as right-click to place a point for the view to center on. You may also need to play around with the zoom level and other parameters before saving your settings, to make sure that TileMill actually saves the settings and defaults to the level you want.

When you have added data and made style adjustments to your layers, TileMill offers an excellent export function. Selecting the ‘Export’ button in the top right of the screen reveals a drop-down menu with options for exporting as PNG, PDF, SVG, and XML. Selecting one of these options opens a new window which defaults to the selection you configured in your project’s individual settings window. Here you can select new bounds and adjust the actual pixel size of the image to be exported. Some issues to be aware of when exporting are that the export depends largely on your individual project settings. ‘Zoom level’, ‘Scale factor’, and ‘Metatile size’ will effect how the image and relevant bits such as text will render in your image, and these features interact with the image’s pixel size - increasing the pixel dimensions will result in a higher-quality image but will render text as smaller, for example. This may require multiple exports to finesse your image into a format that works for your purposes.

You may want to add more data to your map than you currently have. Fortunately, adding another layer in TileMill is quite easy, but unfortunately editing such data is not. To add points, lines, or polygons, you will need to create them in another program and import them. To edit this kind of data, you will need to open the underlying layers in another program (such as Google Earth), edit them, and re-export them to a format that TileMill can read and import as another layer.

The map of Yolmo point data in Figure 3 was generated in TileMill. Note that the user had a lot more control over which design elements were included in comparison to the CartoDB map. Some features, including the India and Tibet labels and the scale, were added after the image was exported from TileMill.

3.6. Sharing your maps

Having created beautiful maps it’s understandable that you will also want to share them. In this section we discuss the logistics of distributing static and digital maps, as well as considerations for archiving your data.

Both workflows we described offer options to export maps to static images. These images can be used in traditional publication formats, like journal articles and presentations. Remember to always ensure that you correctly attribute the mapping data, whether it be licensed, or distributable like OpenStreetMap data is. Remember to also properly acknowledge the software used to create your language maps to encourage other linguists, researchers and community members to also create their own maps.

Although we have primarily focused on the creation of static maps for display and publication, both workflows also offer the option to share digital maps. Unlike static maps, digital maps allow users to explore, access additional information, and easily disseminate content. When creating your maps in CartoDB and TileMill it is worth considering how you might share your work online as well. With just a small amount more work you may be able to create zoom-able maps, or maps with additional labels or content for a wider range of audiences. An interactive map on your research webpage can help people engage with language documentation as a practice situated in real space. Interactive maps can also be incrementally updated and embellished as you gain more data and understanding of an area.

The formats that mapping data is stored in make archiving very easy. The KML format is based on XML, which is the same underlying structure as data from programs like ELAN, and CSV tables are also good for archiving. Remember to ensure that all points are properly labelled within the file, and sufficient metadata is included to explain what the map data pertains to. Even if you do not create a final map, the data may be archived for future use. High quality JPEG or PDF versions of your final maps may also be appropriate to archive. Your archivist will be able to inform you of any requirements specific to individual archiving platforms. Remember to also keep backups for your own records. CartoDB is a cloud based platform, so can act as an online backup of your data, but you should download your map whenever it is modified just in case.

As a final consideration, you may wish to make your mapping data more widely available, especially if it is a large areal survey, or historical data. For example, Claire Bowern uses her university website to host KML data of polygons of Australian languages and a map of Pama-Nyungan Languages that she created as part of a project on

Australian prehistory (<http://pamanyungan.sites.yale.edu/language-resources>). This data is available for other scholars to use as long as they abide by the conditions set. Alternatively, you could contribute your data towards the OpenStreetMaps initiative. This is particularly useful in countries without much mapping of basic infrastructure. If you have included lines for roads, or points for towns and villages, you can add these. As OpenStreetMaps provide much of the basic infrastructure for programs like TileMill and CartoDB it is a nice way to contribute to open and accessible data.

Conclusion

Mapping is an extremely useful way of making your data accessible and useful to a larger audience, and this paper only exposes the tip of the iceberg when it comes to making maps. Unfortunately the reader will likely have a lot of questions, and in this case the internet, online tutorials and user forums are your friend. As with any graphic, a map can condense and simplify a situation and display context in a way that a paragraph of text cannot. It can also allow researchers to show correlations between data sets that previously were very difficult to visualize abstractly. Although necessary simplifications and perceived correlations can sometimes be problematic, having such tools in our hands as linguists is an incredible opportunity to communicate specialized knowledge in a clear and succinct way. Along with this opportunity comes the responsibility of presenting the data in an ethical manner and ensuring that the communities we work with are respected. We hope that this brief introduction will help other linguistic researchers to add to the world's knowledge about languages and the communities that speak them through map-making.

References

- Author1 (2013a)
- Author1 (2013b)
- Bowern, Claire. 2008. *Linguistic fieldwork: a practical guide*. Basingstoke; New York: Palgrave Macmillan.
- Chamberlain, Brad. 2015. Watersheds and language mapping. Paper presented at the 25th Annual Meeting of the South East Asian Linguistics Society (SEALS). Payap University, Chiang Mai, Thailand, May 27-29.
- Dahl, Östen, & Ljuba Veselinova. 2005. Language map server. *Proceedings of 2005 ESRI User Conference*, San Diego, California, July 25-29.
- Guier, William H. & George C. Weiffenbach. 1997. *Genesis of Satellite Navigation*. John Hopkins APL Technical Digest 19(1). 178-181.
- Hildebrandt, Kristine A., & Shunfu Hu. 2013. Multimedia Mapping on the Internet and Language Documentation: New directions in interdisciplinarity. *Polymath: An Interdisciplinary Arts and Sciences Journal*, 3(3). 51-61.
- Hoch, Shawn, & James J. Hayes. 2010. Geolinguistics: The incorporation of geographic information systems and science. *The Geographical Bulletin* 51. 23-36.
- Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig. 2013. *Ethnologue: Languages of the World*, Seventeenth edn. Dallas, Texas: SIL International. Retrieved from <http://www.ethnologue.com>
- Luebbering, Candice R. 2013. Displaying the geography of language: The cartography of language maps. *The Linguistics Journal* 7(1). 37-69.
- MacEachren, A. M. 1995. *How maps work: Representation, visualization, and design*. New York: The Guilford Press.
- Mackey, W.F. 1988. Geolinguistics: Its scope and principles. In C.H. Williams (ed.), *Language in geographic context*, 20-46. Philadelphia: Multilingual Matters.
- Parkinson, Bradford W. 1996. Introduction and heritage of NAVSTAR, the Global Positioning System. In Bradford W. Parkinson, James J. Spilker, Penina Axelrad, & Per Enge (eds.), *Global Positioning System theory and applications*, Vol. 1, 3-28. Washington, D.C.: American Institute of Aeronautics and Astronautics. <http://site.ebrary.com/id/10516767>.
- Peeters, Y.J.D. 1992. The political importance of the visualisation of language contact. *Discussion Papers in Geolinguistics* 19-21. 6-8.
- Pellerin, Cheryl. 2006. United States Updates Global Positioning System Technology: New GPS satellite ushers in a range of future improvements. Url (accessed 5

- June, 2015):
<http://iipdigital.usembassy.gov/st/english/article/2006/02/20060203125928lcnirellep0.5061609.html#ixzz3cANu1lpE>
- Sloetjes, Han & Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008) 816–20.
- Trudgill, Peter. 1983. *On dialect: Social and geographical perspectives*. Oxford: Blackwell.
- Williams, C.H. 1988. An introduction to Geolinguistics. In C.H. Williams (ed.), *Language in geographic context*, 1-19. Philadelphia: Multilingual Matters.
- Williams, C.H., & Ambrose, J.E. 1988. On measuring language border areas. In C.H. Williams (ed.), *Language in geographic context*, 93-135. Philadelphia: Multilingual Matters.
- Williams, C.H., & Ambrose, J.E. 1992. Geolinguistic developments and cartographic problems. *Discussion papers in geolinguistics* 11-32.

Appendix

Below is the CSS code used to make the interactive Yolmo map, which can be found here: https://lgawne.cartodb.com/viz/94f75f72-a200-11e4-90b6-0e0c41326911/public_map. #villages refers to Yolmo villages and #villages2 refers to non-Yolmo villages.

```
#languagelocations [zoom<11]{
  marker-fill-opacity: 0.9;
  marker-line-color: #000000;
  marker-line-width: 2;
  marker-line-opacity: 1;
  marker-placement: point;
  marker-type: ellipse;
  marker-width: 20;
  marker-fill: #D6301D;
  marker-allow-overlap: true;
}

#languagelocations::labels [zoom<11]{
  text-name: [language];
  text-face-name: 'Lato Regular';
  text-size: 15;
  text-label-position-tolerance: 0;
  text-fill: #FFFFFF;
  text-halo-fill: #000000;
  text-halo-radius: 2;
  text-dy: -20;
  text-allow-overlap: true;
  text-placement: point;
  text-placement-type: dummy;
}

#villages[zoom>10]{
  marker-fill-opacity: 0.9;
  marker-line-color: #FFF;
  marker-line-width: 2;
  marker-line-opacity: 1;
  marker-placement: point;
  marker-type: ellipse;
  [zoom>10] {marker-width:15}
  marker-fill: #D6301D;
  marker-allow-overlap: true;
}

#villages::labels [zoom>10]{
  text-name: [name];
  text-face-name: 'Lato Regular';
  text-size: 14;
  text-label-position-tolerance: 0;
  text-fill: #000;
  text-halo-fill: #FFFFFF;
  text-halo-radius: 2;
  text-dy: -15;
  text-allow-overlap: true;
  text-placement: point;
  text-placement-type: dummy;
}

#villages2[zoom>10]{
```

```
marker-fill-opacity: 0.9;
marker-line-color: #FFF;
marker-line-width: 1.5;
marker-line-opacity: 1;
marker-placement: point;
marker-type: ellipse;
marker-width: 12;
marker-fill: #3E7BB6;
marker-allow-overlap: true;
}

#villages2::labels [zoom>10]{
  text-name: [name];
  text-face-name: 'Lato Regular';
  text-size: 14;
  text-label-position-tolerance: 0;
  text-fill: #000;
  text-halo-fill: #FFF;
  text-halo-radius: 1.5;
  text-dy: -12;
  text-allow-overlap: true;
  text-placement: point;
  text-placement-type: dummy;
}

#road[zoom>10]{
  line-color: #FFCC00;
  line-width: 2;
  line-opacity: 0.7;
}
```

Figures and Tables

Figure 1 Map of Nepal with Yolmo villages used in Author1 (2013a):

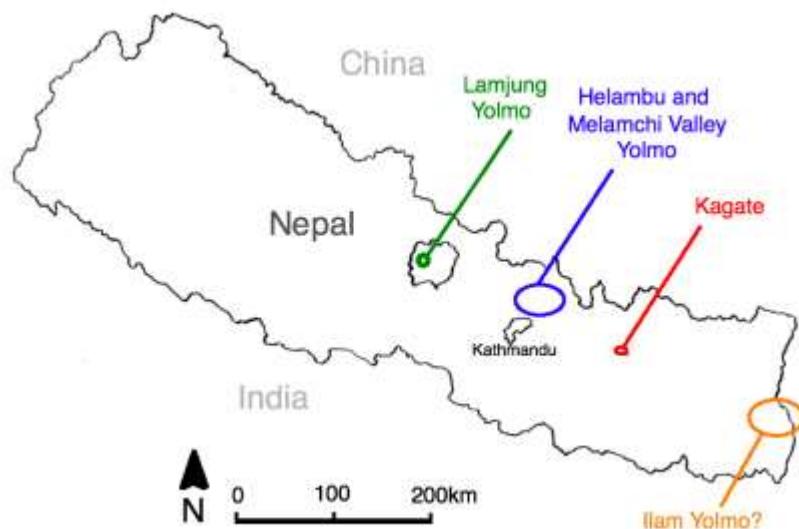


Figure 2 Map of Yolmo villages in Lamjung used in Author1 (2013a) :

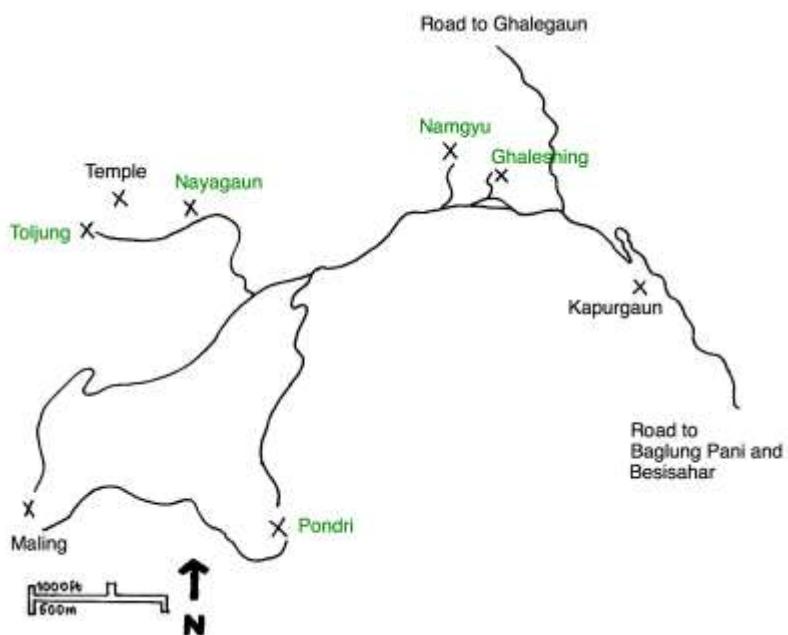


Figure 3 Map of Nepal made in TileMill used in Author1 (2013b) :

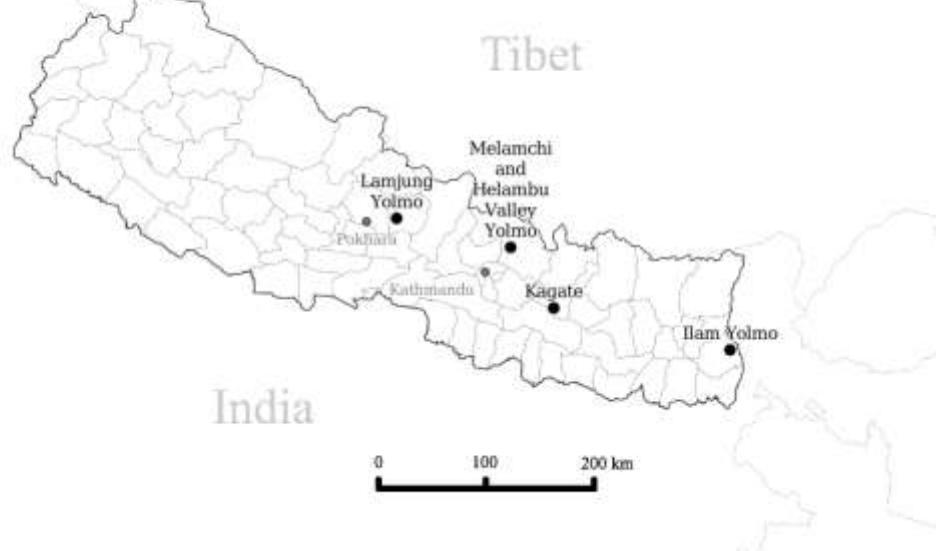


Figure 4 - The Google Earth representation of Yolmo village GPS data in Figure 2:

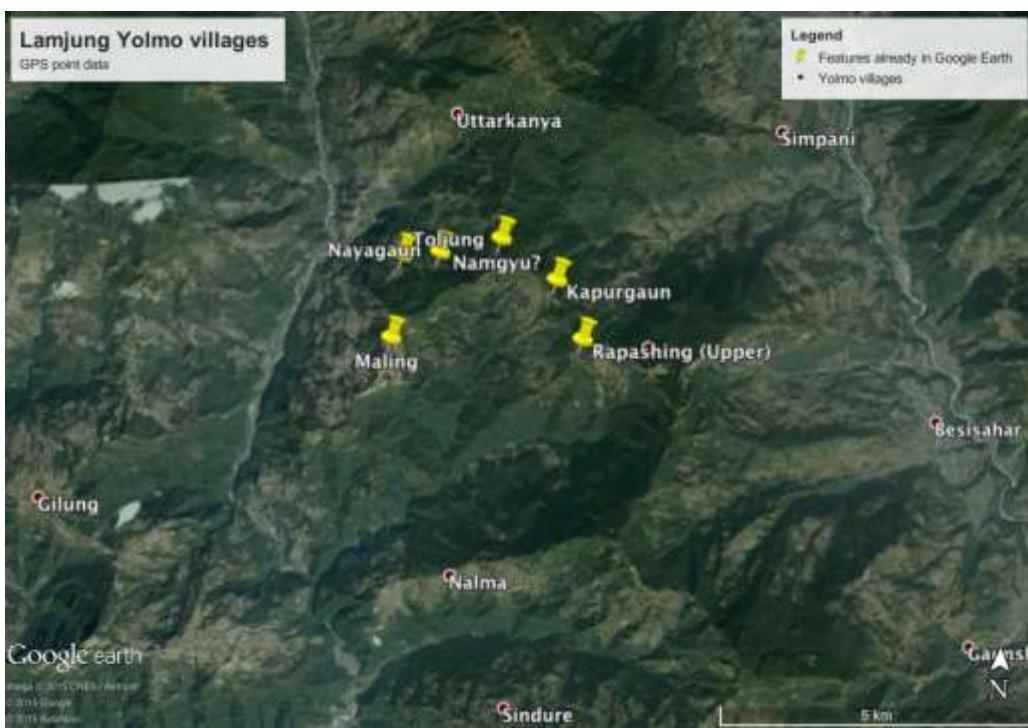


Figure 5. Map made using the CSV in Table 1 - basemap © [OpenStreetMap](#) contributors, design © [CartoDB](#), rest © Author1



Figure 6. Map of Nepal with Yolmo language areas © [OpenStreetMap](#) contributors © [CartoDB](#) © Author1

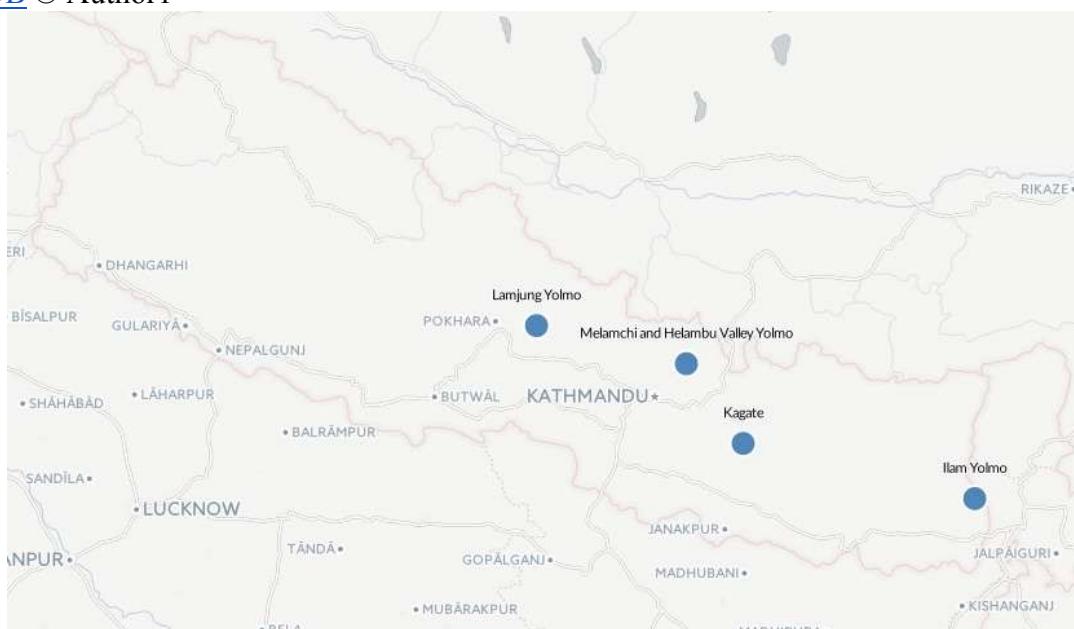


Table 1: A set of longitude and latitude values for six US cities

Name	X	Y
Anchorage	-150.02	61.17
Honolulu	-157.93	21.35
Los Angeles	-118.4	33.93
New York	-73.98	40.77
Seattle	-122.3	47.45
Houston	-95.35	29.97

Author contact details:

Lauren Gawne

lg21@soas.ac.uk

Note Lauren has left NTU and can now be contacted at her SOAS address. She can also be contacted at her personal email address lauren.gawne@gmail.com

Hiram Ring

hiram1@e.ntu.edu.sg

[note: This is Hiram's student email address and may stop working in the near future, he can be contacted at his personal email address ringrecords@gmail.com]