

## Language Landscape: Supporting community-led language documentation<sup>1</sup>

Sandy Ritchie, Samantha Goodchild, and Ebany Dohle

*Language Landscape/SOAS*

Different groups have differing motivations for participating in language documentation projects. Linguists want to increase our knowledge of languages and linguistic theory, but constraints on their work may lead to issues with their documentation projects, including their representations of the languages they study. Native speakers participate to maintain and develop their language, and may choose to represent it in a way which showcases their culture and attitudes. In order to encourage more native speakers to take part in documentation projects, a simple integrated system is required which will enable them to record, annotate and publish recordings. Language Landscape, our web-based application, enables native speakers to publish their recordings, and Aikuma, a mobile application for documentation, enables them to record and orally translate recordings, in both cases with minimal cost and training required. Language Landscape benefits communities by allowing them to document their language as they see fit, as demonstrated by our outreach program, through which some London school children created their own projects to document their own languages and those spoken around them.

**1. INTRODUCTION.** Linguists and native speakers participate in language documentation projects for different reasons. “. . . [T]he linguists’ research aim is to contribute to our scientific knowledge of the world’s languages or to linguistic theory, while the local language workers’ aim is to do something for the maintenance and development of their language and culture” (Mosel 2006: 67). These differing motivations may lead to differences in the way that the two groups choose to represent languages. As Dobrin et al. (2007) opine, “. . . linguists’ efforts to preserve [endangered] languages seem to lead inexorably to their reduction [. . .] to indices, objects, and technical encodings. . .” (2007: 60). Native speakers may prefer to represent their language in a way which showcases the culture it encodes and expresses the community’s collective view of themselves and others in their society. So far, documentation projects initiated and led by native speakers outside of the academic sphere have remained largely hypothetical. Two reasons for this are lack of resources and lack of technical knowledge. Documentation projects often rely on costly and complex technology and are carried out by students and professionals who have undergone extensive training. These requirements are barriers to entry for native

<sup>1</sup> We would like to thank the students and staff of Bow School, east London for working with us through our outreach program, and Teresa Poeta for her comments on an earlier draft of this paper. All remaining errors are our own.

speakers who might otherwise be motivated to participate in and even initiate documentation projects. Removing these barriers is desirable as native speakers can then exercise greater control over the way their language and culture is represented in the public sphere. In order to achieve this, a system is required which enables people to create, annotate, and publish recordings of their language with minimal training and with the technology that is available to them. Given the availability of cheap internet-enabled mobile phones<sup>2</sup> and high speed mobile telecommunication networks (GIA 2014) in an increasing number of countries, it seems that a mobile application which incorporates all of the capabilities listed above would be the best format for such a system.

This paper sets out to show how Language Landscape (<http://languagelandscape.org/>), a web-based application developed by our team,<sup>3</sup> may meet some of the requirements of such a system. The website enables users to publish their recordings on a global map of languages. The paper also discusses a mobile application for documentation currently in development: Aikuma (<http://www.aikuma.org/>) is an Android application developed by researchers at the University of Melbourne which enables users to create recordings and provide time-aligned oral translations of them. Both applications may be viewed as attempts to engender greater participation in documentation projects by employing simple interfaces which require minimal training and cost to be able to use. Language Landscape offers simple forms which make publishing a recording or project online quick and easy, and collaborative features which enable people to work together on documentation projects. Aikuma employs an innovative “phone call” interaction style whereby the speaker listens to a recording as if they were on the phone and can interrupt it at any time to add a segment of oral translation simply by speaking into the phone (Hanke & Bird 2013).

Each of these applications goes some way to lowering the barriers to entry in language documentation projects. However, neither currently provides a complete integrated system for people equipped only with internet-connected mobile phones to create and publish a documentation project. This paper speculates on how the development of an accompanying mobile application for Language Landscape would bring our system closer to this ideal.

The paper also briefly discusses issues with current documentation practices and contrasts them with how documentation may be done using software like Language Landscape and Aikuma. It goes on to highlight the benefits to language communities and individuals offered by Language Landscape. An evaluation of a pilot outreach programme carried out in a London school is also included as a case study of the benefits of community-led documentation.

**2. ISSUES IN LANGUAGE DOCUMENTATION PRACTICES.** Current language documentation practices have led to the creation of a previously unparalleled range of materials relating to minority and endangered languages. The scope and detail of these materials is due in large part to the efforts of a small community of highly specialized researchers. These linguist-documenters have spent years collecting, annotating, and archiving bodies of material which often come close to Himmelmann’s “comprehensive record of the linguistic practices characteristic of a given speech community” (1998: 166). However, linguist-led documentation projects are not without their problems. There is often a large cultural gap

<sup>2</sup> <http://www.theguardian.com/technology/2014/jan/13/smartphone-explosion-2014-india-us-china-firefoxos-android> (10 December, 2015).

<sup>3</sup> Language Landscape team. Directors: Ebany Dohle, Samantha Goodchild, Karolina Grzech, Charlotte Hemmings, Teresa Poeta, Sandy Ritchie (all SOAS); Main Developer: Graham Ritchie (The University of Edinburgh); Designer: Nick Ritchie (BBC Radio).

between linguist and community, leading to problems with recording certain cultural and linguistic practices. Linguists are often working on tight budgets and timescales which limit the potential of their documentation project. Many linguists are also engaged in teaching and research activities which take them away from their fieldwork.

One solution to these problems is to carry out documentation work with the community. Grinevald advises “fieldwork with teaching, training, and mentoring native speakers for sustainable documentation projects” (2003: 60). Typically, the linguist-documenter trains community members in the techniques of language documentation both to collect data which would normally be inaccessible to the researcher and in the hope that they will continue the work of documenting the language beyond the scope of the funded research project. This is an admirable endeavor and there have been some success stories, for example Grinevald’s own work with the Rana community in Nicaragua and McGill’s work with the Cicipu community in Nigeria (McGill 2009).<sup>4</sup> However, for many projects, training of community members in the use of recording equipment and software packages is impracticable due to lack of time and resources. The requirement of basic computer literacy to complete tasks such as transcription and translation work is also sometimes an issue as some communities have had little or no experience with computers until the arrival of the researcher.

**3. THE ROLE OF TECHNOLOGY IN COMMUNITY-LED DOCUMENTATION.** In the face of these issues, truly community-led language documentation projects might seem like a distant prospect. However, with the increasing availability of cheap mobile technology, it is possible to envisage a future in which many more people own or have access to tools which could readily be harnessed for language documentation purposes. In a community with several internet-connected mobile phones installed with a dedicated language documentation application, making, annotating, and publishing recordings could be carried out by a much wider range of people. In such a situation, the possibilities for involvement in or even initiation of a documentation project by the community seem much more likely.

**3.1. CROWDSOURCING TRANSLATION: AIKUMA.** A mobile application which includes some of these features is already available: Aikuma uses an innovative user interface which enables users to create recordings and provide time-aligned oral translations of them.

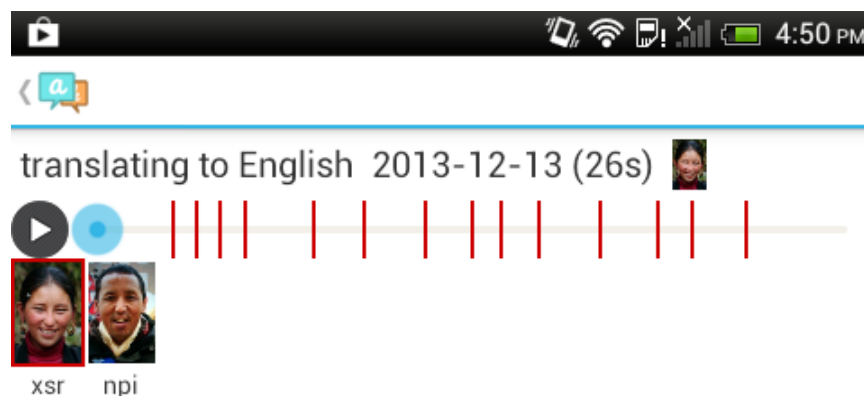
The developers’ motivation for creating the app is to remove the bottleneck in current typical language documentation workflows in creating transcriptions and translations of recordings (Hanke & Bird 2013). Trials of the app have been successful and the developers plan a series of improvements which will make the app even more practical for use in language documentation projects.<sup>5</sup>

Aikuma is an excellent example of how to harness the capabilities of low-end mobile phones<sup>6</sup> to broaden participation in a language documentation project. By removing the requirement of an established writing system and/or literacy, a much wider range of people become potential collaborators in a project. Even in communities which do have high rates of literacy, using Aikuma negates the need to provide extensive training in recording techniques and the use of computer software packages. This frees the linguist-documenter of a considerable commitment and allows them to involve a wider range of people in the documentation project.

<sup>4</sup> See also <http://cicipu.org/> (10 December, 2015).

<sup>5</sup> See <http://www.aikuma.org/development.html> (10 December, 2015).

<sup>6</sup> The researchers used inexpensive Android mobile devices in their trials.

FIGURE 1: Time-aligned oral translation on Aikuma<sup>7</sup>

Aikuma demonstrates that technology can be an inclusive force in language documentation projects. However, the typical documentation workflow does not end at transcribing and translating recordings. Beyond that, recordings are tagged with metadata, annotated using dedicated software such as Toolbox/FLE<sub>x</sub> and ELAN and eventually deposited in an archive. Respecting access restrictions, the archive will then publish the collection online, making it accessible to anyone with an internet connection.

**3.2. CROWDSOURCING PUBLICATION: LANGUAGE LANDSCAPE.** Where Aikuma is designed for crowdsourcing of the early stages of the documentation workflow – that of making and translating recordings, the Language Landscape website is designed in part for crowdsourcing of the final stage – publishing recordings online. We will return to this point but before that we will introduce the project and describe some features of the website.

Language Landscape is a not-for-profit organization based in London, UK<sup>8</sup> which aims to raise awareness of language diversity. We run a website and outreach program to bring together language communities online, to help people to better understand the languages spoken around them, and to help to raise the profile of minority and endangered languages. Our website enables anyone with an internet connection to add language recordings to a world map. The recordings are mapped where they were made in order to reflect the geographical spread of languages. Other users can access the recordings by browsing or searching the map. In order to encourage more people to contribute to the project, we have initiated outreach work with several communities and completed a pilot educational program with a school in east London. The program provided students with practical training in recording techniques and helped them to learn about issues such as multilingualism and language endangerment, using the website as a starting point for discussion and activities (see Section 5).

The website is based around a series of maps which display where recordings added to the site were made.<sup>9</sup> Each recording is represented by a marker which users can click on to see information about the recording and play it in page. The homepage displays all the recordings currently available on the site.

<sup>7</sup> Image source: <http://www.aikuma.org/getting-started.html> (10 December, 2015)

<sup>8</sup> Language Landscape is a company limited by guarantee registered in England and Wales (8694275).

<sup>9</sup> We use the Google Maps API and an in page JavaScript media player to play the recordings. The website runs on a Python web application server with recordings and associated metadata is tracked in a relational database.

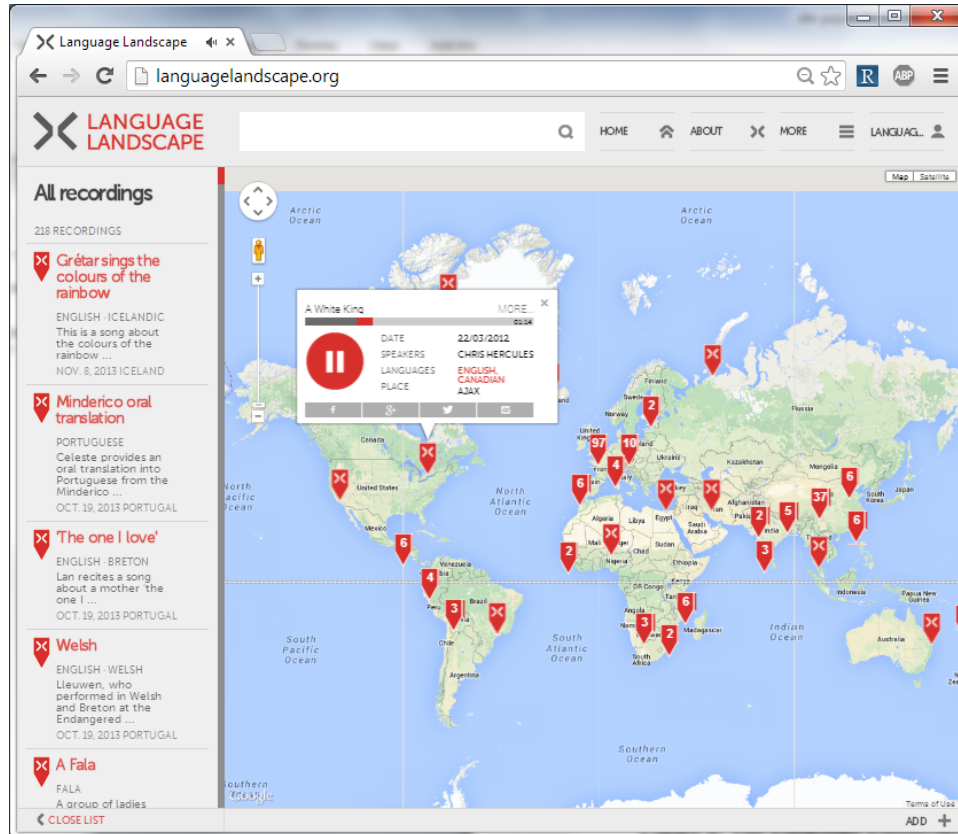


FIGURE 2: The Language Landscape homepage with an in page recording being played

As well as this primary map displaying all of the recordings currently available, the website also features potentially limitless “submaps” which show subsets of the recordings. For example, each language represented on the site has its own submap which displays only the recordings in which someone is speaking that language (see Figure 3). These pages also offer other information about the language, including a description and its position in the genealogical tree.<sup>10</sup>

If a user is looking for something specific, they can also define their own submap by using the Advanced Search functions (<http://languagelandscape.org/search/recordings/>). These allow users to conduct complex queries on the database with the results displayed as a custom submap (see Figure 4).

For reasons of space we have only outlined a few of the website’s features here. For more information and help on using the features of the site, see our Help pages (<http://languagelandscape.org/help/>).

Now that the website has been explained in more detail, we can return to the point of how it may help to encourage greater participation in the publication of language documentation materials. The central idea driving the design and implementation of the website was

<sup>10</sup> Genealogical trees are primarily taken from Glottolog – <http://glottolog.org/glottolog> (10 December 2015) – as well as other sources.

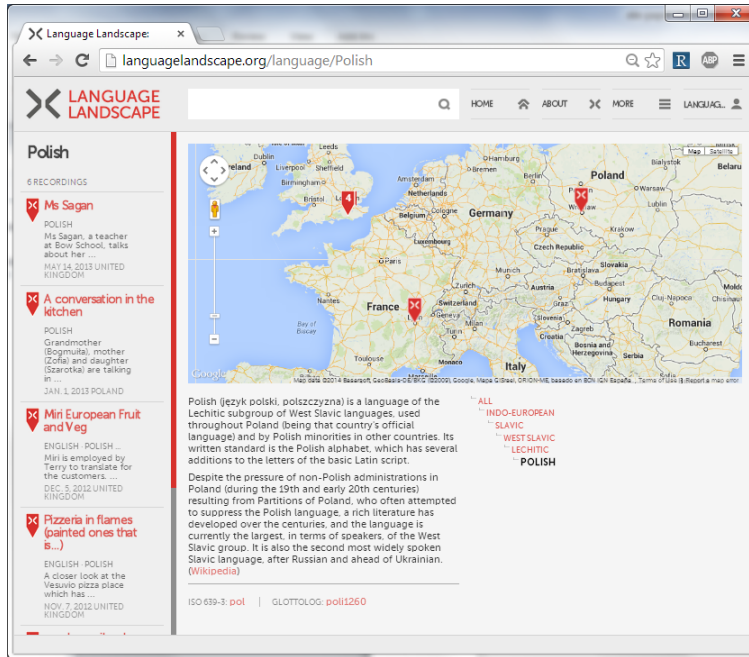


FIGURE 3: A Language Landscape submap: recordings of Polish

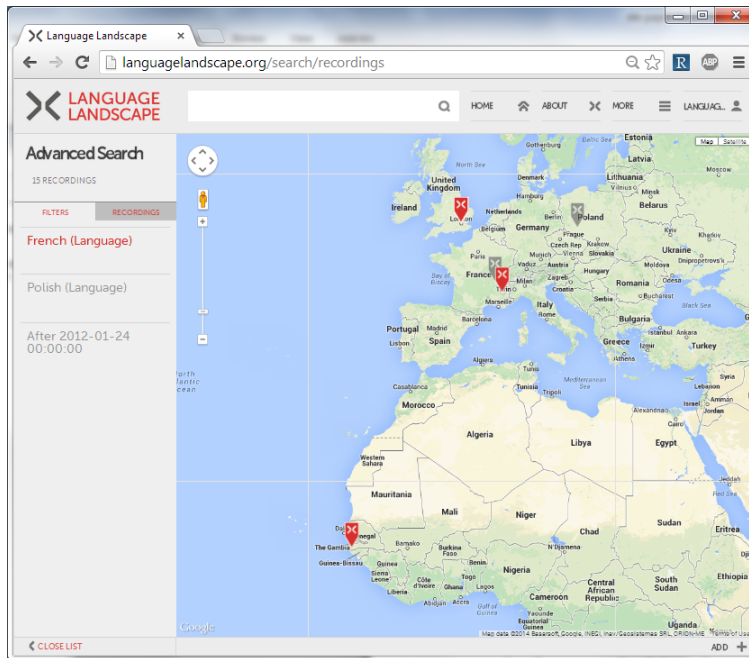


FIGURE 4: A custom submap: Recordings of either French or Polish made after 24 January 2012. The filters on the left show which recordings match which search term.

to make this process quick and easy for anyone with an internet connection and a Language Landscape account.<sup>11</sup> To add a recording, contributors upload it to the website and fill out a simple form.<sup>12</sup> Contributors can also map existing content from YouTube on Language Landscape by pasting the URL in the relevant field instead of uploading a file. Once the recording has been added to the database, it is flagged for moderation and published after being checked by a moderator.

The screenshot shows a web browser window with the URL `languagelandscape.org/add/recording/`. The page header includes the Language Landscape logo and navigation links: HOME, ABOUT, MORE, and LANGUAGE. The main content area is titled "Add Recording" and contains the following fields:

- TITLE:** Scots song
- AUDIO FILE:** Choose File | 60-Scots-sample.mp3
- YOUTUBE URL:** (empty)
- PHOTO:** Choose File | No file chosen
- TIME:** 2014-01-14 00:00
- PLACE:** Glasgow, United Kingdom
- Map:** A Google Map of Glasgow, United Kingdom, with a red location pin.
- SPEAKERS:** Type some text to search in this autocomplete (with a red plus button)
- LANGUAGES:** Type some text to search in this autocomplete (with a red plus button)

FIGURE 5: Language Landscape Add Recording form

If the contributor agrees to it, the recording's metadata can also be edited by other users. This allows people to improve on existing recordings by adding transcriptions, translations, and other information which the original contributor may not have provided.

As well as adding individual recordings to the map, contributors can also create their own "projects" on the website. A Language Landscape project is a user-defined submap featuring a collection of recordings grouped together around a particular theme. Projects are designed to be collaborative: several users can work together on a project by contributing recordings and metadata and personalizing their project page with our customization tools. The projects each have their own page with options to add a description and video.

<sup>11</sup> You can sign up for an account by registering your details here: <http://languagelandscape.org/accounts/login/>.

<sup>12</sup> The full list of metadata categories in this form is available here: <http://languagelandscape.org/help/add-recording/>.

Users can also customize the style of their project map using the Google Styled Maps Wizard<sup>13</sup> and add a custom marker by uploading their own marker image. The form for adding a project is also designed to be quick and easy to fill in.

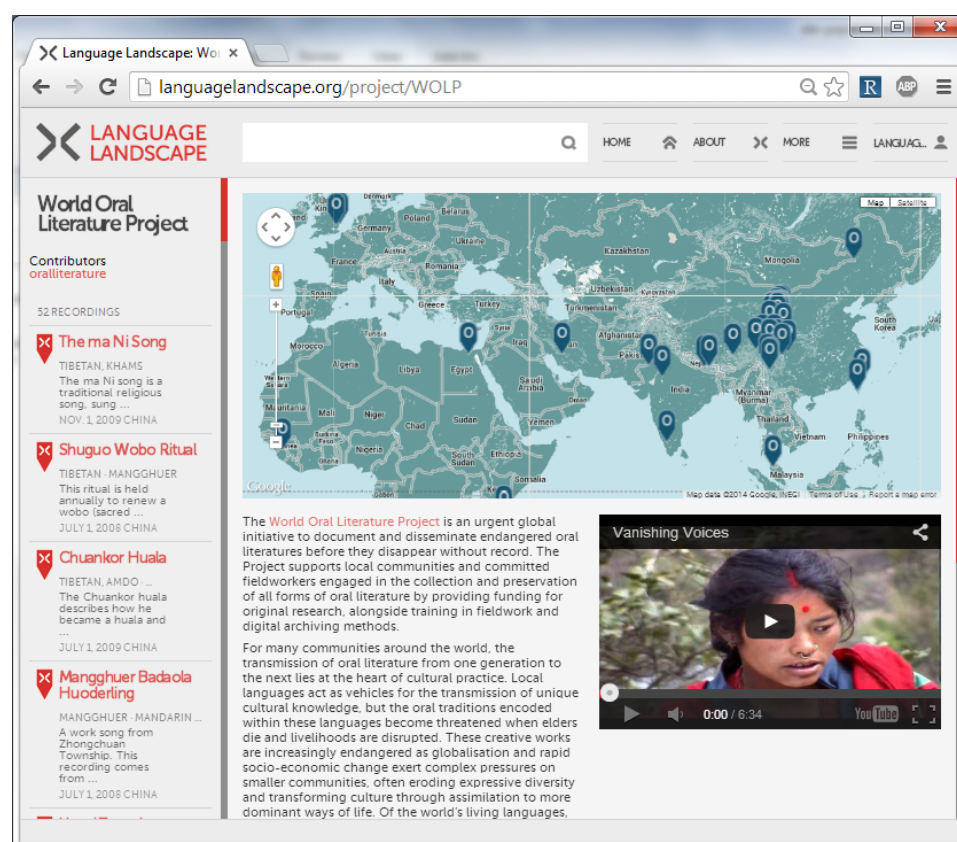


FIGURE 6: A customized Language Landscape project page (<http://languagelandscape.org/project/WOLP>)

We believe that by offering these intuitive content creation features for free online, we will help to encourage more people to get involved in documentation projects. The system would also be greatly improved by the addition of an accompanying mobile application à la Aikuma. Such an app would integrate the ability to make a recording using a phone with the tagging and publication tools we have already developed. This would make the entire process of making, tagging, and publishing recordings online even simpler and more streamlined. Given the global trend towards smart phones as the primary communication tool, development of such an application is an important next step for the project.<sup>14</sup>

This section described how software like Aikuma and Language Landscape can encourage greater participation in documentation projects by simplifying the processes in-

<sup>13</sup> <http://gmaps-samples-v3.googlecode.com/svn/trunk/styledmaps/wizard/index.html> (10 December, 2015).

<sup>14</sup> If you are interested in supporting the development of Language Landscape, please get in touch with us: [admin@languagelandscape.org](mailto:admin@languagelandscape.org)



volved. The next section will discuss why using Language Landscape in particular may benefit communities and individuals.

**4. BENEFITS TO COMMUNITIES AND INDIVIDUALS.** The primary advantage of Language Landscape is that it enables speakers to represent their languages and cultures online as they see fit. Representing one's own languages offers several possible benefits to communities and individuals, in particular to those who speak lesser-known and endangered language varieties.

**4.1. NAMING THE LANGUAGE.** Firstly, it allows people to decide on their own name for their language variety. A linguist who is not a native speaker of the language he/she studies may not be aware of subtle dialectal differences and may represent many varieties as one and the same language. Native speakers make these distinctions intuitively, and the freedom to name their variety as they wish allows them to represent these differences. Other speakers who also speak the same variety can then contribute to the new submap, giving the variety an identity and a "homepage" online. This also extends to smaller and perhaps less prestigious varieties of larger languages. Some varieties of British English, to take an example close to the authors' home, are popularly thought of as aberrations of the "standard" variety or indications of low educational attainment (e.g. Watts & Trudgill 2002). Enabling speakers to showcase their unique variety by putting it on the map may lead to greater respect and understanding of the diversity present within large languages as well.

Allowing user-defined language names is potentially problematic as in some cases these names will not match formal terminology used in linguistics and related disciplines. We intend to work around this by retaining the user-defined name for a variety but also adding alternative names in the metadata, so for example colloquial names for varieties such as Jafaican will also be tagged with more formal designators like Multicultural London English. This will allow for the use of standard terminology while retaining the ability for native speakers to name their variety as they see fit.

**4.2. GENRES AND REGISTERS.** As with language varieties, native speakers are also best placed to document the diversity of registers and genres present within a speech community. Young people in particular are often at the cutting edge of new forms of expression as well as being au fait with the speech of their parents' and grandparents' generations. Encouraging young people to record these genres and collaborate on innovative projects to document and map them will help them to appreciate the expressive power of their own language in comparison with dominant global cultural forms.

**4.3. CODE-SWITCHING, CODE-MIXING AND MULTILINGUALISM.** For many communities, using several languages in the course of their daily lives is the norm (see e.g. Lüpke & Storch 2013 for an overview of the situation in Africa). A single language map will not be sufficient to represent these communities' linguistic practices and repertoires. Offering multilingual communities the ability to tag their recordings with multiple languages and represent their practices as project submaps rather than language submaps will enable them to document the complex and dynamic nature of their linguistic situations.

**4.4. COLLABORATION.** The ability to collaborate is key to encouraging more people to get involved in a project. Collaboration allows communities to represent their language more democratically: the work of making and translating recordings is distributed amongst many people, and everyone can have a say in which recordings get selected for publication. A particular advantage of using an online platform is that it also enables communities who are separated geographically but share linguistic and cultural ties to collaborate remotely. This will also help to maintain links or establish new connections between diaspora communities and the community back home.

We will now turn to the benefits of our outreach work, focusing on a case study of an outreach project with a group of mainly Sylheti-speaking school children living in east London.

**5. LANGUAGE LANDSCAPE OUTREACH.** In 2013 Language Landscape carried out a pilot outreach project with Bow School, in East London, with a view to hosting regular outreach activities with various communities in the city. In this section we will look at our aims in our outreach programs, including increasing participation, before focusing on the Bow School case study.

**5.1. INCREASING PARTICIPATION.** Language Landscape can be used to document any variety of language anywhere in the world. In order to encourage greater use of the website, we feel it important to focus our efforts on individuals and communities who are most likely to become active users. We carried out our pilot study with a school in London (see below section 5.3) and will be working in the future with other minority and endangered language communities. We hope to encourage further uptake of our outreach projects which aim to increase community- and individual-led documentation efforts through providing basic training on using readily available technologies and a background to linguistic issues. By promoting our website and outreach program, we hope to engage with a greater number of speakers and communities, fostering greater interactivity among users of the website and creating an enthusiasm for undertaking community-led documentation projects.

**5.2. GENERAL AIMS.** The outreach projects carried out by Language Landscape are flexible in nature and can be adapted to suit the individual needs of different communities.<sup>15</sup> There are, however, core themes which form the basis of our outreach projects. As all current Language Landscape directors and volunteers have been trained in documentary linguistics, we are able to offer classes and workshops in the basics of sound recording and editing, both using professional equipment, and also how to achieve good results using the technology available to most people. In addition to providing a solid grounding in technology training, the outreach projects also give a broad overview of linguistic issues to present a better understanding of the multilingual world which we live in. Some of the areas which we cover in the outreach project are: an introduction to the discipline of linguistics, why people study language(s), minority and endangered languages, and the benefits of multilingualism. The outreach projects provide speakers and communities with the skills and knowledge required to create their own projects to document their language varieties in whichever way they wish, as demonstrated in our pilot outreach program with Bow School below.

---

<sup>15</sup> If you are interested in running an outreach project with Language Landscape, please contact our outreach team at [outreach@languagelandscape.org](mailto:outreach@languagelandscape.org)

**5.3. CASE STUDY: BOW SCHOOL.** Language Landscape carried out a pilot outreach project with Bow School in east London in spring 2013. The pupils created their own projects which are available to view on the website. The outreach project consisted of an eight-week series of workshops which culminated in a visit to SOAS, University of London, where the pupils took part in a debate on endangered languages and presented their projects to students, staff, and the general public.

Bow School is set in an area of East London, the Borough of Tower Hamlets, where over 70% of schoolchildren have English as a second language.<sup>16</sup> In this particular school, the outreach team worked with a class of pupils between the ages of 12–13, where the majority of children also speak Bengali and/or Sylheti. Throughout the eight weeks the team devised a number of interactive activities to introduce the children to topics such as minority languages, multilingualism, and endangered languages. They were also taught the basics of sound recording and were able to try it out on semi-professional equipment provided by the outreach team. The main focus, however, was on applying the recording techniques for use on their own mobile/smart phones<sup>17</sup> to carry out their own projects. The pupils were encouraged to come up with their own topics for their projects and many of the topics chosen were inspired by issues covered in the workshops and by the children's own experiences. For example, one group surveyed teachers in their school and recorded them speaking various languages, whilst another group chose to focus their project on languages and scripts used for writing, interviewing teachers who can write in the Bengali script. As is evident, when free to choose their own topics, the pupils created projects which were interesting to themselves, projects that reflect their community.

The recordings and projects created by the pupils in the case study could certainly be of interest to linguists and researchers investigating multilingual repertoires, certain types of speech genres and adolescent speech. Above all, though, the recordings are of definite interest to the speakers themselves, and by extension, their communities. We believe that this project shows some of the hallmarks of a community-led documentation project, which forms the basis of Language Landscape's model of documentation, which may provide an alternative or accompaniment to linguist-led documentation projects.

**6. CONCLUSION.** As demonstrated by the case study, simple, intuitive tools such as Language Landscape do enable communities and individuals to collaborate on documentation projects which represent their language and culture in the way they see fit. The publication features offered will also be greatly improved in future by the addition of a mobile application which enables users to add recordings directly from their mobile phone. The website offers many benefits to native speakers of smaller languages, not least the opportunity to take part in a global project which represents smaller language varieties on an even footing with larger languages. The website may also be a driving force for the establishment of new online communities, "bring[ing] together speakers and learners who are scattered over great distances" (Holton 2011: 371).

In addition to our online platform, the outreach projects provide an opportunity for some communities to receive face-to-face training in linguistics and technical skills. Although these are not required to create an interesting project on the website, our outreach work allows us to engage with communities directly, and we benefit from this by incorporating their suggestions and ideas into the website and outreach program. Gener-

<sup>16</sup> [http://www.towerhamlets.gov.uk/lgsi/351-400/367\\_census\\_information/2011\\_census.aspx](http://www.towerhamlets.gov.uk/lgsi/351-400/367_census_information/2011_census.aspx) (30 January, 2014).

<sup>17</sup> Although not all pupils owned a mobile/smart phone, enough children owned them so that when the class was divided into groups of four, each group had access to at least one phone.

ally, our aim is to raise awareness of language diversity and the many issues surrounding it by making language documentation projects open to a wider and more diverse range of people, thereby supplementing current language documentation practices with community- and individual-led documentation projects.

#### REFERENCES

- Dobrin, Lise, Peter K. Austin & David Nathan. 2007. Dying to Be Counted: The Commodification of Endangered Languages in Documentary Linguistics. In Peter K. Austin, Oliver Bond & David Nathan (eds.), *Proceedings of Conference on Language Documentation and Linguistic Theory*, 59–68. London: SOAS.
- Global Industry Analysts Inc. 2014. 2G, 3G & 4G Cellular Network Subscriptions. USA. <http://www.strategyr.com/TrendReport.asp?code=146042> (10 December, 2015).
- Grinevald, Colette. 2003. Speakers and Documentation of Endangered Languages. *Language Documentation and Description* 1. 52–72.
- Hanke, Florian R. & Steven Bird. 2013. Large-Scale Text Collection for Unwritten Languages. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 1134–38. Nagoya, Japan. <http://aclweb.org/anthology/I/I13/I13-1161.pdf> (10 December, 2015).
- Himmelman, Nikolaus. 1998. Documentary and Descriptive Linguistics. *Linguistics* 36. 161–95.
- Holton, Gary. 2011. The Role of Information Technology in Supporting Minority and Endangered Languages. In Peter K. Austin & Julia Sallabank (eds.), *The Cambridge Handbook of Endangered Languages*, 371–400. Cambridge: Cambridge University Press.
- Lüpke, Friederike & Anne Storch. 2013. *Repertoires and Choices in African Languages*. Berlin & Boston: De Gruyter.
- McGill, Stuart. 2009. Gender and Person Agreement in Cicipu Discourse. School of Oriental and African Studies, University of London.
- Mosel, Ulrike. 2006. Fieldwork and Community Language Work. In Jost Gippert, Nikolaus Himmelmann & Ulrike Mosel (eds.), *Essentials of Language Documentation*, 67–85. Berlin, New York: Mouton de Gruyter.
- Watts, Richard J. & Peter Trudgill. 2002. *Alternative Histories of English*. London & New York: Routledge.

Sandy Ritchie  
sandyfritchie@gmail.com

Samantha Goodchild  
299225@soas.ac.uk

Ebany Dohle  
ebdohle@googlemail.com