elpublishing

Language Documentation and Description

ISSN 2756-1224

This article appears in: *Language Documentation and Description, vol 20*. Editor: Peter K. Austin

COVID-19 and documentary linguistics: Some ways forward

NICHOLAS WILLIAMS, W. D. L. SILVA, LAURA MCPHERSON, JEFF GOOD

Cite this article: Williams, Nicholas, W. D. L. Silva, Laura McPherson & Jeff Good. 2021. COVID-19 and documentary linguistics: Some ways forward. Language Documentation and Description 20, 359-377.

Link to this article: http://www.elpublishing.org/PID/241

This electronic version first published: December 2021

COVID-19 and documentary linguistics: Some ways forward

Nicholas Williams,[1] W. D. L. Silva,[2] Laura McPherson,[3] Jeff Good[4]
University of Potsdam,[1] *University of Arizona,*[2] *Dartmouth College,*[3] *University at Buffalo*[4]

Abstract

In the wake of widespread and ongoing travel restrictions that began in early 2020 due to the COVID-19 pandemic, many documentary linguists worldwide shifted to remote work methods in order to continue or, in some cases, begin new projects. This pandemic situation has prompted questions about both methodological and ethical considerations in doing remote fieldwork. In this paper, we discuss the pros and cons of working remotely and discuss ways of working remotely based on our experiences working on projects in West Africa, northwest Amazonia, and Indonesia. We argue that elements of remote fieldwork should become a permanent part of linguistic fieldwork, but that such methods need to be considered in the context of decolonizing language documentation and centering the community's needs and interests.

1. Introduction

The ongoing COVID-19 pandemic has had a dramatic effect on the work of documentary linguists worldwide. Widespread travel restrictions beginning in early 2020 made it impossible to continue with traditional on-site fieldwork, in which an outsider linguist visits a community to conduct collaborative research. Even as the situation has improved in some parts of the world, we face the difficult question of whether or not to start or continue projects in vulnerable communities that are suffering higher infection rates, have under-resourced healthcare infrastructure, and have been more negatively impacted by the economic fallout of the pandemic and ongoing social distancing measures. Despite the challenges, communities and linguists are seeking ways to continue working and building on the success of prior documentation and revitalization projects. In some cases, we are building on workflows that were developed prior to the pandemic. In other cases, we are finding ourselves inventing new models of working in the current environment.

Williams, Nicholas, W. D. L. Silva, Laura McPherson & Jeff Good. 2021. COVID-19 and documentary linguistics: Some ways forward. *Language Documentation and Description* 20, 359-377.

Drawing on our experiences working on projects in West Africa, northwest Amazonia, and Indonesia, we discuss ways of doing language research remotely and consider the implications of this way of working for language documentation and conservation. Although we have used some of the tools and software discussed below for being in contact with our collaborators, for some time, many of us only gave serious consideration to remote fieldwork when forced to by a global pandemic. while we ought to have been thinking outside the box in this way all along. We hope by sharing these methods we will spur further discussion of remote methods, which we expect (and hope) will continue to be used indefinitely even 'after' the pandemic is 'over'. We also emphasize the importance of using this opportunity to reflect on what we are trying to accomplish in our language work and how we might incorporate some of the beneficial aspects of remote fieldwork in the future.

Finally, although this paper addresses methods and techniques for remote linguistic fieldwork, we want to point out that it can be difficult to start a project remotely. Indeed, we believe the work described here, in most cases, has only been possible because of the rapport we have built with our collaborators through face-to-face relationships in previous years. While certain documentation activities are possible remotely, it is important that we ensure any remotely conducted activities are also beneficial to collaborators in the field and the ongoing relationships we are building. In other words, we must ensure that the changes further the goals of decolonising linguistics (Leonard 2018) and avoid perpetuating colonial dynamics in our work.

We begin in Section 2 with a discussion of the pros and cons of remote linguistic fieldwork learned since the beginning of the COVID-19 pandemic, and consider the ethical implications of this shift, particularly in the context of the ongoing movement to decolonise linguistics and academia more broadly. In Section 3, we focus on some practical tools and their uses for remote language work. We conclude in Section 4 with some suggestions for future directions in remote fieldwork.

2. Some pros and cons of remote language work

Doing the work of language documentation remotely has benefits but also brings certain challenges. Additionally, there are various practical and ethical implications of conducting this work primarily or entirely remotely. If we are to continue conducting language work remotely, it is important to consider these implications and to adopt only the most useful and ethical methods.

Prior to the pandemic, most documentation projects involved travel of one or more linguists to the community where the language is spoken. With the onset of widespread travel restrictions, this traditional model of fieldwork became impossible to continue. Those with projects just beginning or already

in progress needed to find ways to continue their projects while travel was not possible. The obvious solution was to develop ways to work remotely, that is, without traveling to the community.

In this context, one important question arises immediately: is it actually possible to conduct the same language 'work' remotely that was previously done through in person fieldwork? Certainly it is possible to coordinate work with (trained) community members, so that projects may continue by making video and audio recordings, transcribing and annotating those recordings, and preparing the resulting files and associated metadata. But is the 'work' of documentation projects limited to the production of archivable annotated video and audio recordings? Arguably, the 'work' includes a range of other activities in the community that bear the intangible benefit of raising the language's profile and promoting maintenance and revitalization simply by being present for the explicit purpose of language work. Nevertheless, we focus here primarily on the practical and ethical considerations for conducting the 'real' work of producing documentary materials remotely.

First, while it is possible to work remotely, one possible issue we must remain sensitive to is that in a time of pandemic, communities may have more pressing issues to attend to than language documentation research. Families may be grappling with illness, loss of employment, or school closures that make focus on research difficult or impossible.

On the other hand, collaborators who have lost employment or are confined to their homes may appreciate work more than usual, both for the income it generates and for a distraction from the drudgery of confinement. Unable to travel to either the field or to conferences, researchers may have research funds sitting unused, which could be put to good use if redirected to collaborators and communities.

One obvious benefit of doing language work remotely is that it is much cheaper than traditional fieldwork. In many research grants, international travel and living costs are one of the biggest budget categories. If even one or two planned trips during the grant period could instead be achieved remotely, whether in a pandemic situation or not, it could make the funding go much further (i.e., that funding could be redirected to participants, whether in the field or students at home or host universities) while at the same time reducing the carbon footprint of our fieldwork.

A second potential benefit of remote research is that the physical absence of the principal investigator (PI) could also serve to decenter the researcher in the project. If work can only take place when the PI is able to be in the field, this leaves any scientific progress, monetary compensation, and documentation progress subject to the personal availability of one (often outside) individual rather than responsive to the community's needs and availability. It may also be that remote fieldwork offers the opportunity to change the nature of the work in such a way as to decenter the outside researcher's primary goals and focus instead towards centering Indigenous

perspectives and activities in the work of language documentation and maintenance. For instance, it is possible for the attention of the work to shift and more clearly focus on the speakers and interaction in the community itself, and away from the community's interaction *with the researcher.*

In the past, the typical fieldwork model involved travel to a location to record materials and later transport these materials 'home' to be organised and eventually deposited in an archive, which may or may not be accessible to speakers and the community. Given that travel has been heavily restricted during the pandemic, any recordings must be made by community members, who can immediately do as they wish with them without concerns of access or having to coordinate with an external researcher. This shift places more agency in the hands of speakers and community members to negotiate with external researchers what will be recorded and what will happen to the recordings. Ultimately, this model of research could serve to flip the script from 'linguists' working with 'language consultants' to the language community working with a 'linguist consultant', more on their own terms. Grzech (2020) has suggested the same approach, based on her experience proposing a new project remotely to a community she had not previously be in contact with.

Many (ideally most) projects include community training and capacity building as part of their research activities. With remote fieldwork, collaborators do not need to wait until the following fieldwork season to continue their training. This is convenient to all, as participants can accommodate other activities in their daily schedule, especially with asynchronous work. The use of commonly used platforms like *WhatsApp* and *Google Drive* (see Section 3) means that the skills and proficiencies gained in the course of language documentation are also highly applicable in other contexts.

Finally, the implications of remote fieldwork extend beyond the current pandemic into both the past and the future. Many regions with high levels of linguistic diversity also face unrest that can disrupt pre-existing field projects and make it difficult to commit to long-term projects going into the future, due to uncertainty about whether travel to the field site will be safe. Having a workflow for remote fieldwork in place allows the project to keep moving, and for collaborators to keep being compensated, even if physical travel is interrupted by instability. Developing local advising networks and remote collaboration also permits larger cooperative projects that will be necessary to effectively document all the under-documented languages in places with high levels of linguistic diversity. This has been the approach taken in Good's collaborative *Key Pluridisciplinary Advances on African Multilingualism – Cameroon* (KPAAM-CAM) project,[1] which, for a number of years, has developed

[1] https://kpaam-cam.org/ (accessed 2021-12-16). KPAAM-CAM research has been supported by NSF grants BCS-1360763 and BCS-1761639.

a model for conducting interdisciplinary language documentation focusing on a rural region of West Africa.

This project is team-based and involves cooperation between researchers at Good's home institution as well as a number of local universities. In particular, local linguistics students have been supported under the direct supervision of local academic staff, with additional advising by out-of-country researchers. There were a number of reasons why this model was adopted, including:

(a) the presence of a strong local linguistic research community with a number of members specifically interested in language documentation, which provided the foundation for cooperative research of this kind;

(b) the communities of focus lacked individuals who were in a position to do language documentation directly, which meant that research collaborations were more sensibly focused on the local scholarly community rather than with community members themselves (with other, more community-oriented collaborative endeavors developed in parallel to compensate for this);

(c) the fact that this part of the world has hundreds of underdocumented languages, which means that local capacity building is required if documentary work is to scale up beyond what can be conducted by outside researchers;

(d) a large number of local scholars can be involved in a project structured this way on more or less the same research budget as would support only a few outside scholars due to lower travel and salary costs.

This model has the further advantage that the research can continue even in cases where the outside researcher cannot spend significant time at the research site, not only due to unanticipated events (such as a global pandemic or civil war) but also due to more mundane ones such as professional or family obligations.[2]

While we have outlined above a number of benefits of remote fieldwork, it must also be acknowledged that in some settings members of speaker communities may positively value the physical presence of an outside

[2] An unexpected benefit of this model in the context of the current pandemic is that in-person research has been able to resume in-country during periods when local travel restrictions have been relaxed, even when international travel remains effectively impossible.

researcher as a way of validating the importance of their languages (see, e.g., Dobrin 2008). Moreover, the dehumanised approach to the collection and exchange of data associated with remote fieldwork has unfortunate parallels with processes of extractive industries that enter areas where local communities are socioeconomically marginalised, and they are forced to export commodities which are transformed into more valuable products elsewhere (Austin (2013: 14) calls this the 'plantation project' model). We do not have a specific solution to this problem beyond emphasising that conducting remote fieldwork ethically requires the outside linguist to pay careful attention to how mutually beneficial relationships can be developed, even at a distance. We also note here that our own efforts at remote fieldwork took place mostly after we had established in-person relationships with a speaker community. Starting a project remotely would clearly be associated with additional challenges, and would likely raise distinctive ethical issues.

In sum, while we should at all times be critical and self-reflective in our research, always keeping in mind not only 'do no harm' but also 'do some good', we believe that on the whole, the possibility of remote fieldwork bears significant benefits over the traditional in-person only model, while also opening up avenues for decentering the outside researcher and decolonising the language documentation process. We argue that all documentary projects should include plans for remote work as either an essential component of the project or, at the very least, as a contingency plan in the case of instability and travel limitations.

3. Tool support for remote language documentation and linguistic fieldwork

In recent decades, digital communication and data transfer tools have enabled much more extensive kinds of remote fieldwork than were previously possible. Successful remote fieldwork requires addressing a number of distinct problems of communication and interaction, and these must take into account the kinds of technologies that both the researcher and community members have access to. In this discussion, we assume a model where the outside researcher has regular access to high-speed internet and more-or-less to up-to-date technology while local scholars and community members have, at best, sporadic access to high-speed internet connections, and limited access to technology such as smartphones or, perhaps, laptop computers. We believe that this is a typical situation for remote fieldwork in a documentary context. Moreover, since this represents the most difficult scenario where remote fieldwork can even be considered, it is the one that needs the most detailed consideration. It contrasts, for instance, with the work reported by Leemann et al. (2020), which describes remote linguistic data collection in Switzerland, using methods that assume access to much greater resources than is the case in most of the world.

Remote fieldwork requires two broad kinds of technical support: (a) tools to facilitate coordination and general communication with community members; and (b) tools to support transfer of different kinds of data, including audio and/or video recordings, annotations, and metadata. Different kinds of data may themselves require the use of different kinds of tools for transfer, given that services that work well for relatively small files (such as those created by Elan annotation software[3]) will not necessarily function well for large files (such as a high resolution video recording). An additional dimension of consideration is whether a tool is optimized for supporting synchronous or asynchronous communication, or a mix of both kinds.

One general lesson we have learned, which is shared by other projects, is that, where possible, there are great advantages to working with widely-supported technologies that are already familiar and accessible to all team members. We have found, in particular, that *WhatsApp* and *Google Drive* are very effective for information sharing, with *WhatsApp* being the default communication tool, and *Google Drive* being the best option for sharing large files. These services operate very well with poor and sporadic internet connections, and we can build on the investment their commercial developers have made to ensure worldwide connectivity to their services. Moreover, these kinds of tools are much more likely to also be known and used by community members than specialized linguistic applications, thereby facilitating community engagement. The approach we describe below therefore contrasts with attempts such as Hanke (2017) to build fieldwork-specific tools for collaborative language documentation, though such tools clearly have a place in long-term considerations of how to support remote fieldwork. It is valuable to explore a variety of models of technological support required for remote fieldwork.

We structure our discussion below with respect to specific tools and their capabilities as a means of ensuring that the discussion is concrete enough for those engaged in language documentation to make specific plans for their own projects. However, will keep general considerations in mind during the discussion as well. We focus on three particular widely-used tools: *WhatsApp*, *Google Drive*, and *Zoom*. Each of these can be seen as standing in for a class of similar messenger, data storage and transfer, and video communication tools, respectively. However, as will be clear from the discussion below, these particular tools can support multiple functions beyond their core capabilities.

[3] See https://archive.mpi.nl/tla/elan (accessed 2021-12-16)

At the end of this section, we contrast the possibilities of using general purpose tools like these with the potential capabilities of a custom-made tool designed specifically for research purposes.

3.1 WhatsApp and the expanding capabilities of messenger apps

WhatsApp is one of the world's most widely-used messaging apps and has been successfully used for conducting remote fieldwork by the authors, as well as other fieldworkers. Its popularity is due to several factors:

(a) it is already in wide use and free (other than network and smartphone costs). This means that community members are often already comfortable with it for day-to-day communication;

(b) it works well in low-powered network environments;

(c) its core function is messaging, however it also effectively incorporates other kinds of functionality, such as voice calls and basic file transfer;

(d) it supports voice messaging very well, making it accessible to individuals with limited literacy;

(e) it functions equally well for both synchronous and asynchronous communication.

Taken together, this means that *WhatsApp* is not only an effective basic communication tool to interact with individual collaborators and to manage work across teams, but it can also support certain kinds of data transfer for language documentation. As mobile internet infrastructure spreads to even remote areas, we are now able to maintain relationships via *WhatsApp* that previously would have been impossible. For example, in the midst of the pandemic, Williams has been able to restart work with the Kula community in Alor, Indonesia, who now have electricity and mobile phone services in their village. One useful feature lacking in *WhatsApp*, but available in other messenger tools, such as *WeChat*, is a payments facility, which can streamline the process through which consultants and collaborators are compensated.

A messenger app such as *WhatsApp* can be used to conduct virtual fieldwork sessions with individual speakers where its multimedia features can be especially valuable. The linguist can send text, audio, or video messages, and receive responses back synchronously or asynchronously, as conditions allow. This way of working most closely approximates

traditional in-person fieldwork, which typically involves a linguist/language-learner interacting with a speaker of the target language. While the nature of this interaction varies according to the goals of the community members and linguists, and the nature of the project, remote interaction via *WhatsApp* is well suited to nearly any form of such one-on-one interaction between linguist and community member. If the community has some degree of training in language documentation, *WhatsApp* can also be used to help guide them in collecting data from other community members, and to monitor their data collection process (see Griscom 2020 for relevant discussion).

In Silva's work with Mũtẽã (also known as Karapanã) speakers in Colombia, *WhatsApp* was initially used primarily to facilitate elicitation and development of materials for teaching the language. However, over time another use of *WhatsApp* emerged. One active collaborator, Jhon Vargas, decided to create a *WhatsApp* group to serve as a virtual platform for speakers and learners to engage with the language and the documentation being done remotely. This *WhatsApp* group generates its own instances of language use (which could also be documented, subject to participants' consent), while also allowing wider and immediate dissemination of materials for language lessons developed during remote sessions. This highlights the value of exploiting tools already in wide community use when conducting remote fieldwork, since they facilitate the expansion of documentary work by community members with other members of the community, without the need for intervention (e.g., through file conversion) by the outside researcher.

Another way of using messaging software like *WhatsApp* is to facilitate oral transcriptions and translations of previously-recorded texts, when working with speakers remotely. McPherson describes this process for documention of songs in Seenku.[4] A Seenku speaker living in Vienna is able to go through already recorded videos piece-by-piece, playing the video, pausing it, then repeating back slowly what is said via *WhatsApp* audio. He then sends these audio messages to McPherson, who can use them to create written transcriptions. This technique of remote oral transcription builds on the *Basic Oral Language Documentation* (BOLD) transcription method (Reiman 2010). While somewhat slow and cumbersome, it is a straightforward way to continue transcription work while the linguist is out of the field. It can be easily applied to new projects and settings.

[4] Research supported by the NSF Documenting Endangered Languages grant BCS-1664335.

WhatsApp can also be used to transfer relatively small files between team members. For example, when collaborating with Kotiria and Wa'ikhana communities in northwest Brazil, Williams has trained Indigenous research team members to transcribe video-recorded conversations in Elan. The resulting annotation (.eaf) files can be transferred via *WhatsApp* through a desktop application to the linguist partners on the project. Given the small size of .eaf files and the ease of *WhatsApp* file transfer, this is a much simpler and more effective means of sharing information than email, which often requires a good internet connection to simply open an inbox and draft a message.

3.2 Messenger tool technical considerations

Certain technical aspects of messenger tools like *WhatsApp* bear mentioning. First, we have found that working with *WhatsApp* voice messages has several benefits:

(a) avoiding problems of limited internet connectivity, which may result in dropped calls;

(b) potential for asynchronous work;

(c) researchers (or consultants) can listen to messages multiple times;

(d) messages are able to be downloaded and archived; and

(e) in terms of confidentiality, all messages are automatically encrypted by the application.

WhatsApp records in 'opus', or OGG, format. This format is playable in most available audio programs (e.g. *Windows Media Player*, *VLC Media Player*, *Apple QuickTime Player*, etc.). Although OGG is a compressed audio format, the quality is generally better than MP3. In our experience, *WhatsApp* message audio quality is generally sufficient for linguistic fieldwork, especially if the consultant is in quiet surroundings, although it may not be good enough for detailed phonetic analysis, especially of consonant phenomena. While we do not have direct experience, using external microphones specifically designed for smartphones might improve recordings, if consultants can be given access and trained in their use.

On the desktop or web version of the app, OGG files can be downloaded and easily converted to WAV format (e.g. for archiving) using *Audacity*, or other audio conversion programs. Files are automatically saved with a label containing: (a) the service name; (b) the date; and (c) the time the message was received, as illustrated in Figure 1.

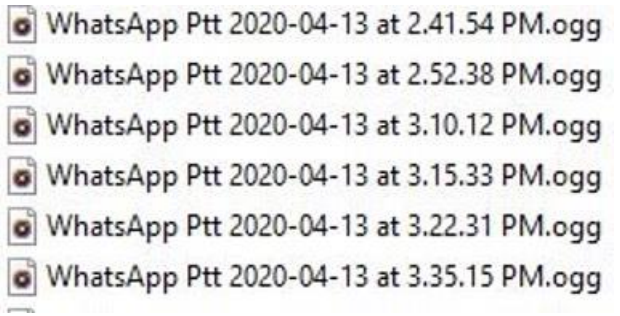

Figure 1. Screenshot of filenames download from WhatsApp

Such file names make it easy to recover later the order in which messages were received. They can either be saved and archived individually, or concatenated into a single file for the work session, using *Praat* or *Audacity*. Another benefit of using the desktop or web version of the app during the elicitation session is that it is easier to type notes using the computer keyboard than the phone keyboard. If only audio messages are being exchanged, this is less an issue. The desktop, web, and mobile versions of the app synchronise with one another, so material gathered using the mobile version can also be accessed through the desktop app.

WhatsApp can also record video in MP4 format. We have less experience using *WhatsApp* video for fieldwork, though we suspect that the audio quality would suffer due to the necessary physical distance between the phone's microphone and the person speaking (Zhang et al. 2021).[5]

While *WhatsApp* is in very wide use, it is not a good tool for every part of the world, notably China, where *WeChat* dominates and can be used in a similar way. We are not familiar with *WeChat*, and would encourage those who do use it for fieldwork to discuss its strengths and weaknesses, and to compare the file formats that it produces with those of *WhatsApp*.

3.3 Cloud-based file transfer and tools such as Google Drive

Transferring digital video and audio recordings, or even large databases, can be challenging in places with access to limited or sporadic internet. Various cloud-based services are available, but for fieldwork contexts we have found *Google Drive* to be most effective. While file uploads may still take some

[5] See Sanker et al. (forthcoming) for an evaluation of audio signal fidelity in remote recording setups.

time, the synchronization process can effectively make use of times when internet connections may be stronger (e.g., at night) to help expedite file transfer. Once the files reach Google's cloud services, they can then be accessed quite quickly from locations with good internet connections.

Depending on the needs of a particular project, once the files are uploaded to *Google Drive*, they can be directly downloaded, or copied within the web-based *Google Drive* application between different folders. One difficulty in sharing files in this way, which is easier to manage with non-remote fieldwork, is ensuring that consistent file-naming conventions are followed, and that appropriate metadata is collected. This needs to be checked after the files can be accessed rather than at the time of their creation; establishing their provenance requires extra steps, since the person receiving the files was not present during the recording. As with non-remote fieldwork, however, taking care of these issues promptly can alleviate issues that can arise when the details of a recording are forgotten over time. Developing project-internal standards for recording metadata using tools that the entire team is comfortable with helps make metadata collection and curation more robust.

Once recordings have been made and uploaded to shared folders in *Google Drive*, team members can then access and work with them in various ways. One possible workflow is being implemented by Williams and Dr. Tasnim Lubis for their project documenting Leukon in Simeulue, Indonesia. Here, Lubis has uploaded video recordings to *Google Drive* along with an *Excel* spreadsheet containing metadata. Williams is then able to access and download these files to create .imdi metadata files and bundles (sets of audio, video, transcription, and metadata files) for archiving. Williams then does initial processing, including segmentation in *Elan* and making preparations for written or oral transcription. New versions of these bundles can then be re-uploaded by Williams and downloaded by Lubis for further processing with speakers in Simeulue. While these processes can be cumbersome and time-consuming, especially with slower internet speeds prevalent in many areas, *Google Drive* allows file synchronisation over time. Additionally, as conditions have improved locally and health protocols are in place, Lubis has been able to continue making recordings. Through this hybrid approach, Lubis and Williams have been able to continue their work and achieve the goals originally set out for their project.

3.4 Video conferencing software and remote training

Widespread use of video conferencing software such as *Zoom* has transformed remote work across many domains. Some projects are using *Zoom* to record elicitation sessions, while others are conducting entire field methods courses remotely via *Zoom*. Since mid-2020, Williams has been running remote

workshops on Elan for language documentation in Indonesia. While there are challenges for remote areas due to limited connectivity, *Zoom* and similar software makes it possible to greatly expand the reach of training and participation in language documentation. In cases where consultants have access to audio and video recording devices but could benefit from remote training or supervision, *Zoom* can also play a useful facilitation role, as experienced by Good, and also described in Leeman et al. (2020). In our experience, *Zoom* is also more stable and less prone to interruptions than other video conferencing applications like *Skype*, making it more appropriate for weaker internet connections. The ability of *Zoom* to easily record a session is also useful where participants may want to access a recording later, e.g. for review or training purposes.

Because synchronous video conferencing requires a strong internet connection, at present we do not see it as a core tool for remote data collection. However, given that community members may be able to temporarily travel to locations where they do have access to high-speed internet, tools like *Zoom* seem especially well suited to intensive training sessions. These can be coordinated through messenger tools like *WhatsApp* and have files transferred to community members via tools like *Google Drive*. However, to reach the widest possible audience, it would clearly be preferable to develop techniques for remote training in language documentation which do not rely heavily on synchronous video conferencing. Unfortunately, commonly used web-based remote learning software platforms have similar needs for stable internet connections as video conferencing tools, and they are not well-designed for multilingual support (see, e.g., Libbrecht et al. 2019).

3.5 What would a successful research app look like?

In addition to these general tools, linguists have developed specific software for remote linguistics work. Good and colleagues have spent a number of years developing a smartphone-based data collection tool for Android. Superficially, this resembles *Aikuma* (Hanke 2017), which is also a smartphone app designed to support documentary research. However, it differs in that it is intended to be used by researchers trained to some degree in documentary techniques rather than by community members. After a successful pilot trial involving field-based data collection, synchronization of data to a server, and transformation of that data for automated archiving into the University at Buffalo Institutional Repository,[6] the app is now being

[6] https://ubir.buffalo.edu/ (accessed 2021-12-16)

redeveloped drawing on lessons from the pilot. The initial version of the app made use of the FAIMS Mobile platform (Ballsun-Stanton et al. 2018), which was originally designed for the collection of archaeological data. Archaeologists often work in settings with limited internet connectivity, and this app allowed for structured data collection using a smartphone, with data synchronisation taking place when reliable internet access was available. The data collection needs for archaeology could effectively make use of a simpler user interface than is possible for language documentation, which is why we are working on a reimplementation with more extensive interface capabilities.

This approach is based on the following observations:

(a) a smartphone is, in effect, a remote sensor and, therefore, can be used to collect a wide variety of kinds of data;

(b) smartphones can automatically record crucial metadata (e.g, time and location), thereby facilitating metadata collection; and

(c) a smartphone app can be used to enforce good data collection workflows.

With respect to point (i), while a smartphone cannot replace a high-quality audio recorder, they make acceptable recordings where sound quality is not crucial (e.g., sociolinguistic interviews conducted in a majority language), and using external microphones with them does enable good quality audio recording. With respect to point (iii), the tool is based on a 'questionnaire' where a researcher defines a series of 'questions' (broadly defined) that guide data collection. By ordering the questions properly and providing instructions in the questionnaire, good workflows can be supported.

The current working model for the functionality of this app is schematised in Figure 2.[7] A 'manager' develops a 'questionnaire' using the app, which then becomes available to a data collector. The data collector creates documentation sessions, which, in turn, are loaded to a server from which the data can be downloaded and used in different ways. The reimplementation has not yet developed the server synchronisation functionality, though this was in place during the pilot. The successful pilot of this tool suggests that it should be possible to use smartphones to support relatively complex data collection workflows in remote settings, while also using them as tools for transferring

[7] The Android robot image in Figure 2 is modified from work created and shared by Google and used according to terms described in the Creative Commons 3.0 Attribution License (https://creativecommons.org/licenses/by/3.0/).

data from the field to a central server for analysis and archiving. A key feature of the design is flexibility in terms of the kinds of questions that can be included in a questionnaire, and the kinds of data that can be collected as answers. For instance, a question could be associated with an informed consent script, or be a request for the translation of a specific concept (e.g., to collect a wordlist), or be designed to gather information about a consultant. Answers can be in the form of text, audio or video recordings, or photographs, using the full range of a smartphone's data collection capabilities. In principle, this should allow an app of this kind to be used to support a wide range of projects.

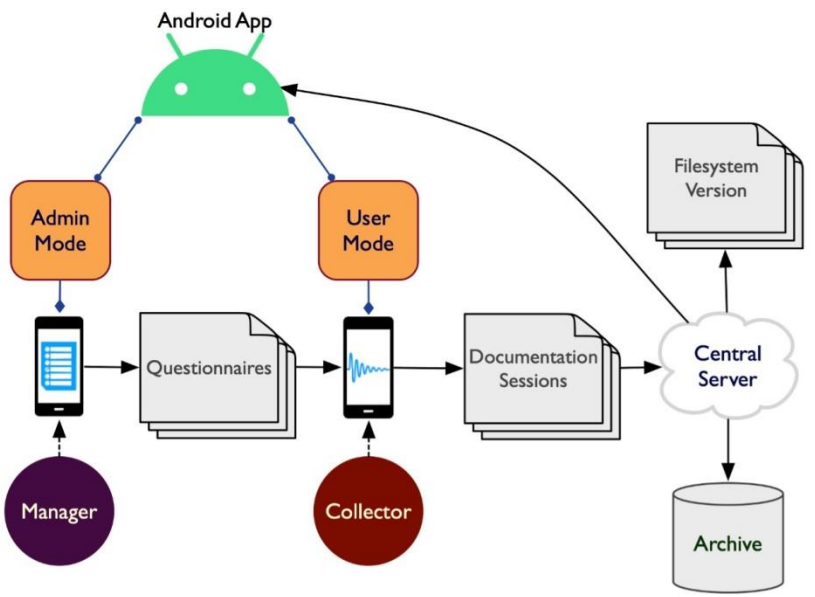

Figure 2: Schematized workflow for remote data collection app

Figure 3 shows what the app looks like in practice. Here we have used the app to create a basic (Swadesh) wordlist. The first screenshot is used to define and create metadata for a recording session. The second screenshot presents the interface to select items in a wordlist questionnaire. The last two screenshots show the interface for recording a particular word.

Figure 3: Illustration of workflow using the app under development

A purpose-built tool such as this has some clear advantages over general purpose tools insofar as it can produce data that is structured in ways that facilitate further annotation, analysis, and archiving. It is particularly well suited for contexts where local linguists with some training in language documentation (including students and trained community members) are in a position to do fieldwork in their home countries but may lack the technology and training needed to produce well-structured documentary corpora on their own. A disadvantage of such tools is that they cannot be readily employed by community members, and making the resources created using them available to community members would require extra steps. Future research on remote fieldwork can hopefully allow documentary linguists to better understand how to balance the potential of specifically-designed apps against the wider accessibility associated with using general purpose tools.

4. Future directions for community-based language work, during COVID-19 and beyond

As the COVID-19 pandemic continues, it is imperative that we embrace new ways of working remotely, and capitalise on them, rather than waiting for a return to pre-pandemic conditions. While we do not anticipate the complete elimination of travel and in-person fieldwork research, future documentation projects will benefit from taking a hybrid approach that involves a mix of both in-person and remote collaboration methods.

Furthermore, we must ensure that we are not using the available tools simply to continue 'collecting data' from speakers to serve the goals of academic linguistics, and to store materials in archives which are largely

inaccessible to the speaker communities. This forced hiatus from in-person fieldwork, which has distanced outside linguists from the communities they work with, has opened up an opportunity for linguists and communities to re-evaluate their needs, and establish more collaborative working relationships. Linguists could use this opportunity to reflect on and evaluate their research methods, while also engaging in dialogue with communities in which they work to rethink project designs and goals.

Documentary linguistics will benefit from a shift to remote and hybrid models of collaborative work in several ways. Perhaps most obviously, remote methods enable teams to maintain communication, and to make progress on data collection and annotation throughout the year, even while collaborators are not in the same location. A reduction in frequency of travel will also reduce costs and environmental impacts of our projects. Available funds will be able to support more projects and more participants for longer periods of time. Additionally, the new possibilities for remote work expand access and participation. As we write in late 2021, field methods courses are being conducted entirely online through *Zoom*, making it possible to work with speakers of languages from almost any part of the world. Documentation methods training workshops can now reach participants from disparate locations. We are only beginning to see the outcomes of such expansions in training and collaboration. Still much needed is the development of better digital training materials, including in languages other than English, and the creation of training materials that are accessible to people with limited access to high-speed networks. In the past, we have relied perhaps too heavily on in-person training events, which are difficult and costly to access for many students and community members, especially from lower-income areas or countries whose citizens have difficulty obtaining the appropriate travel visas. The development of robust digital training materials that can be accessed from anywhere could have a significant impact on efforts to expand participation in language documentation.

While the COVID-19 pandemic has put a sudden stop to in-person fieldwork and the previously common practice of summer research trips, it has not stopped efforts to document and revitalise languages around the world. On the contrary, the impact of the pandemic on vulnerable populations has spurred some to redouble efforts to document severely endangered languages with mostly older and at risk speakers. At the same time, the shift to remote work has enabled new forms of collaboration that will ultimately benefit documentation and revitalisation.

Acknowledgements

Research reported on in this paper was funded by NSF grants BCS-1360763 and BCS-1761639 (Good), NSF grant BCS-1664335 (McPherson), NSF grant BCS-1664348 (Williams and Kristine Stenzel), NSF grant BCS-1159510 (Williams), NSF grant BCS-1837852 (Silva), and ELDP grants IGS0181 (Williams) and SG0581 (Williams and Tasnim Lubis).

References

Austin, Peter K. 2013. Language documentation and meta-documentation. In Mari Jones & Sarah Ogilvie (eds.) *Keeping Languages Alive: Documentation, Pedagogy and Revitalization*, 3–15. Cambridge: Cambridge University Press.

Ballsun-Stanton, Brian, Shawn Ross, Steve Cassidy, Nathan Reid & Jens Klump. 2018. FAIMS Mobile 3.0: A next-generation field data collection platform. *EGU General Assembly Conference Abstracts*. 4163.

Dobrin, Lise M. 2008. From linguistic elicitation to eliciting the linguist: Lessons in community empowerment from Melanesia. *Language* 84(2), 300–324.

Griscom, Richard T. Remote linguistic elicitation methods. Endangered Languages Archive Blog. https://elararchive.org/blog/2020/06/25/remote-linguistic-elicitation-methods/ (accessed 2021-12-16).

Grzech, Karolina. 2020. Remote fieldwork in post-pandemic times. *Language Landscape Blog.* http://languagelandscape.org/blog/remote-fieldwork-in-post-pandemic-times/ (accessed 2021-12-16)

Hanke, Florian R. 2017. Computer Supported Collaborative Language Documentation. PhD thesis, University of Melbourne.

Leemann, Adrian, Péter Jeszenszky, Carina Steiner, Melanie Studerus & Jan Messerli. 2020. Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard* 6(s3). https://doi.org/10.1515/lingvan-2020-0061.

Leonard, Wesley Y. 2018. Reflections on (de)colonialism in language documentation. In Bradley McDonnell, Andrea L. Berez-Kroeker & Gary Holton (eds.), *Reflections on language documentation 20 years after Himmelmann 1998*, 55–65. Language Documentation & Conservation Special Publication no. 15.

Libbrecht, Paul, Stefan Dreisiebner, Björn Buchal & Anna Polzer. 2019. Creating multilingual MOOC content for information literacy: A workflow. In *Proceedings of the conference on learning information literacy across the globe.* https://informationliteracy.eu/conference/assets/papers/LILG-2019_Libbrecht-et-al_Creating_ILO_MOOC.pdf.

Reiman, D. Will. 2010. Basic oral language documentation. *Language Documentation & Conservation* 4, 254–268.

Sanker, Chelsea, Sarah Babinski, Roslyn Burns, Marisha Evans, Juhyae Kim, Slater Smith, Natalie Weber & Claire Bowern. 2021. (Don't) try this at home! The effects of recording devices and software on phonetic analysis. Yale University, MS. To appear in *Language*.

Zhang, Cong, Kathleen Jepson, Georg Lohfink & Amalia Arvaniti. 2021. Comparing acoustic analyses of speech data collected remotely. *The Journal of the Acoustical Society of America* 149(6), 3910–3916.