

*Received 2 October 2022.*

*Accepted 13 January 2023.*

*Published January 2024.*

DOI: 10.1344/DIALECTOLOGIA.34.8

## MOBILE SURVEYS FOR DATA COLLECTION: A CASE STUDY OF KANAUJI

Anu PANDEY \*

Indian Institute of Technology Kanpur

anu.pandey45@gmail.com

### Abstract

This paper demonstrates how mobiles can be effectively adopted for language documentation fieldwork through case studies in Kanauji during COVID-19. Remote fieldwork conditions in India support data collection via mobiles, over laptops or computers, especially for low-resource languages like Kanauji. Telephonic surveys have been used extensively for corpora creation; however, their use in remote language data collection is new. The popular messenger WhatsApp can effectively be used for remote linguistic fieldwork in India. Its multimedia features and low network requirement are a boon. WhatsApp is used for asynchronous elicitation of conversational recordings in Kanauji. The dynamics and ethics of mobile fieldwork documentation differ from traditional fieldwork. However, if the challenges are addressed effectively, it can prove to be a viable model for language documentation.

**Keywords:** remote fieldwork, Kanauji, mobile interviews, COVID-19, conversational recordings

### ENQUESTES MÒBILS PER A LA RECOLLIDA DE DADES: UN ESTUDI DE CAS DE KANAUJI

#### Resum

Aquest article demostra, mitjançant estudis de casos que es van dur a terme a Kanauji durant la COVID-19, que els mòbils poden ser eficaços per obtenir documentació lingüística. Les condicions de treball de camp a l'Índia admeten la recollida de dades amb mòbils a través d'ordinadors portàtils, especialment per a idiomes amb pocs recursos com el Kanauji. Les enquestes telefòniques s'han utilitzat àmpliament per a la creació de corpus; tanmateix, la recollida de dades lingüístiques de manera remota és un procediment nou. WhatsApp també es pot fer servir de manera eficaç per al treball de camp lingüístic. Les seves funcions multimèdia i el baix requisit de xarxa són una avantatge. Així, WhatsApp s'ha utilitzat per obtenir de manera asíncrona enregistraments de conversa a Kanauji. La dinàmica i l'ètica de la documentació del treball de camp mòbil són diferents del treball de camp tradicional. Tanmateix, si s'aborden bé els reptes, pot resultar un bon model per obtenir documentació lingüística.

---

\* 124/ new civil lines, Vinayakpur, Kanpur, Pin- 208024, Uttar Pradesh, India.

© Author(s)



**Paraules clau:** treball de camp remot, Kanauji, entrevistes mòbils, COVID-19, enregistrament de converses

### ENCUESTAS MÓVILES PARA RECOLECCIÓN DE DATOS: UN ESTUDIO DE CASO DE KANAUJI

#### Resumen

Este artículo demuestra, mediante los estudios de casos que se llevaron a cabo en Kanauji durante la COVID-19, que los móviles pueden ser eficaces para obtener documentación lingüística. Las condiciones de trabajo de campamento en la India admiten la recolección de datos con móviles a través de ordenadores portátiles, especialmente para idiomas con pocos recursos como el Kanauji. Las encuestas telefónicas se han utilizado ampliamente para la creación de corpus; sin embargo, la recogida de datos lingüísticos de forma remota es un procedimiento nuevo. WhatsApp también se puede utilizar de forma eficaz para el trabajo de campo lingüístico. Sus funciones multimedia y el bajo requisito de red son una ventaja. Así, WhatsApp se ha utilizado para obtener de forma asíncrona grabaciones de conversación en Kanauji. La dinámica y la ética de la documentación del trabajo de campo móvil son distintas del trabajo de campo tradicional. Sin embargo, si se abordan bien los retos, puede resultar un buen modelo para obtener documentación lingüística.

**Palabras clave:** trabajo de campo remoto, Kanauji, entrevistas móviles, COVID-19, grabación de conversaciones

## 1. Introduction: alternative research methods during COVID-19

The coronavirus pandemic (COVID-19) has taken a significant toll on several areas of human life, including the domain of academic research. Government-enforced social distancing and lockdown measures have confined interviewers in their homes disrupting their ongoing projects. Isolation measures imposed by countries have also meant that in-person interactions had to be avoided to safeguard all parties from being infected by COVID-19.

Scientific research involving laboratory experimentation halted as authorities vacated institutions and moved education primarily online. Thus, interviewers from STEM fields suffered disruption to their ongoing research. The circumstances were equally debilitating for social science interviewers whose projects had fieldwork components. Since interviewers' plans for fieldwork were delayed or disrupted, they were left to evolve their research using newer methods. Some were also forced to modify their already collected data.

Language documentation fieldwork also saw changes in terms of new methods for remote fieldwork. Sneller (2022) features articles that give insights from several sociolinguistic research projects involving remote fieldwork. Grzech & Shaca (2022) and Leeman et al. (2020) discuss how Zoom as a videoconferencing software can be employed for remote fieldwork. This involves establishing a collaborative online relationship between the interviewer and the language consultant. Then, the consultant is given Zoom-mediated training regarding the transcription and translation of audio-visual recording in ELAN, which assists them in data collection. Rice (2021) has talked about how YouTube can be used as a platform for achieving collaborative remote transcription and translation workflow. The benefit of such a workflow is that no training in ELAN is required to capture audio-visual recordings. Hence, it saves time and effort in providing fieldwork training and can be used by even an untrained speaker.

Williams et al. (2021: 361) elaborate on the benefits of doing remote fieldwork and argue why it should become a permanent part of linguistic fieldwork just like the traditional methodologies. Remote fieldwork is cheaper than traditional fieldwork as it can save the travel and lodging costs incurred when making a trip to a fieldwork site. This kind of fieldwork can be carried out by collaborating and coordinating with (trained) community members through various digital platforms like WhatsApp, Google Drive and Zoom. Unused research grants can be redirected to compensate the native speakers of the communities collaborating in the project. Building local networks and remote collaboration projects can enable more community participation with the linguists. Moreover, the collaborators (community speakers) can accommodate this work in their daily schedule and even work asynchronously (Williams et al. 2021: 362).

Most of the linguistic documentation studies that have been done using remote fieldwork in different parts of the world during COVID-19 focus on the requirement of (trained) community members. Training a community member in modern documentation methods is challenging and time-consuming. Secondly, one can effectively employ these methods of remote fieldwork only for those few endangered

community members that are technologically equipped and have good technical support. None of these studies captures the perspective of remote language documentation in low-resourced communities which might not have the monetary resources, technical support, or technical skills to engage in remote fieldwork through platforms like Zoom, Google Drive, and YouTube.

This study focuses on using mobiles and mobile-assisted technologies like WhatsApp for corpora creation via remote fieldwork in low-resourced language communities like the Kanauji language community. At the same time, it highlights the fact that remote fieldwork in India might work differently from how it works in other parts of the world.

## **2. Remote fieldwork in India**

Dash (2020: 108) states that India is home to more of the world's most endangered languages than anywhere else, with around 197 threatened languages according to the UNESCO Interactive Atlas of the World's Languages in Danger.<sup>1</sup> Dash (2020: 120) highlights that poor digital presence and interactiveness on social media are significant reasons behind the endangerment of many languages. He adds that the digital divide is one of the significant differences between safe and endangered languages in India. There is a chance of revitalizing endangered languages through increased internet and social media use. The Scheme for Protection and Preservation of Endangered Languages of India (SPELL), initiated by the Government of India (GOI), aims to preserve and save endangered languages through digital documentation and archiving.

Williams et al. (2021: 364-365) state that remote linguistic fieldwork requires some broad technical support:

(a) Tools assisting in communication with community members. These include limited access to technology, such as smartphones and laptops.

---

<sup>1</sup> <http://www.unesco.org/languages-atlas/> (Accessed 2022-4-20.)

(b) Tools to support the transfer of linguistic data—audio/video recordings, annotations, metadata, etc.

(c) Regular or sporadic access to high-speed internet connectivity and mobile network.

Williams et al. (2021: 371) vouch for the effectiveness of Zoom for remote training of community members in field methods and of Google Drive to transfer large databases like audio or video recordings. Nevertheless, they point out that such web-based remote learning software platforms require excellent and stable internet connections. These methods of remote fieldwork can be effectively employed for those few endangered communities with technically qualified members and good technical support.

Remote fieldwork is a challenge in India's rural and vulnerable tribal communities. Currently, 8.2 percent of the Indian population speaks tribal languages, most of which are endangered to some extent. These linguistically marginalized communities suffer from poverty, illiteracy, and state apathy (Dash 2020: 152). Lack of awareness and prestige in these minoritized language groups perpetuates a lack of community participation. Being poor and uneducated, most community members cannot afford laptops. They also do not have the technical skills to use digital platforms like Google Drive, Gmail, YouTube, or Zoom for linguistic documentation. Moreover, frequent electricity power cuts and the absence of high-speed internet connectivity add to the technical handicap.

Unlike web-based digital platforms, a messenger tool like WhatsApp holds much greater promise for remote fieldwork in India. There are several advantages to using WhatsApp to conduct remote fieldwork (Williams et al. 2021: 366-367):

(a) It is a widely used app, and people are comfortable using it for day-to-day communication.

(b) It works well in low-powered network environments, too.

(c) It has multimedia features like messaging, audio-video calling, and file transfer.

(d) It functions equally well for both synchronous and asynchronous communication.

Smartphones and WhatsApp usage are popular among India's rural and urban masses. The linguist can use WhatsApp to interact remotely with community members by exchanging texts or digital recordings. There is no need to train the community members to use the multimedia features of WhatsApp as they are quite familiar with the app. Such fieldwork approximates traditional in-person fieldwork with direct interaction between a linguist and the speaker of a target language. WhatsApp files are in OGG format, a compressed audio format. The audio quality is good for linguistic fieldwork but may not be sufficient for detailed phonetic analysis. WhatsApp files can later be converted by a linguist using audio-conversion programs like Audacity on the desktop version of WhatsApp (Williams et al. 2021: 367-368).

### **3. History of telephone use in corpora creation**

Telephonic surveys for data collection have long been used for making a corpus in various projects. The Switchboard corpus, dating back to 1993, is an open-access database that consists of 2,400 two-sided telephonic conversations involving 543 speakers from different parts of the United States of America. A computer-driven robot operator system managed this corpus. It used a variety of functions—giving the caller appropriate recorded prompts, selecting and dialing another person, and introducing the topic of discussion. These conversations between two speakers were recorded in separate channels. They were restricted so that no two speakers could make a call more than once and conversation on a specific topic happened only once. Later, these recorded conversations were transcribed and converted into speech files, leading to an open-access database (Godfrey & Holliman 1993).

According to Bernstein et al. (1994: 27-30), another significant corpus created using telephones is the Macrophone. This corpus consists of many American English speakers and was sponsored by the Linguistic Data Consortium (LDC). It forms a part of a bigger project, Polyphone, that has such telephonic datasets from many major

languages worldwide. A planned and structured method was used to build this corpus. Firstly, prompt sheets were sent to the respondents via mail. Each callee received a different kind of sheet that elicited reading and prompted spontaneous responses. Calls came in the next few days. The read responses and spontaneous response transcriptions were made using SRI software. The goal of these corpora was to assist in developing an automatic voice-interactive service, like those used for transportation, ticketing, and shopping applications, among others.

A more recent monolingual spoken corpus<sup>2</sup> has been created and made available on CLARIN infrastructure online. It contains transcriptions of spontaneous and planned speech, such as dialogues, news, or elicited narratives recorded using telephones in several significant languages—Dutch, German, French, and Nepali, to name a few. Since it is freely available online, it is a good resource for a range of linguistic research.

Finally, an interdisciplinary project called the Talkback project, initiated by Carnegie Mellon University has a spoken corpus comprising hundreds of contributions and collaborations. It consists of telephonic interactions, speech narratives, and audio conversations in several major languages. These were recorded and transcribed later, simultaneously being made available online in a chat format. The corpus has been used in various domains of linguistic research—classroom discourse, aphasia, child language, etc. (MacWhinney 2021).

Unlike the Talkback project and the spoken corpus at CLARIN, featuring several major languages, most other corpora have been created only in English. No telephonic corpus exists in understudied languages and endangered languages. Records of such languages can be found in archives or in publications designed explicitly for endangered languages using other language descriptions and documentation methods. Telephonic surveys have not been used for corpus creation in endangered languages and can be used for language data collection. In the sections below, I will elucidate more on this new remote fieldwork method.

---

<sup>2</sup> <https://www.Clarin.eu/resource-families/spoken-corpora#Introduction> (Accessed August 2021-8-19.)

#### 4. Pros and cons of telephone as modes of research

Traditionally, telephones were used for the limited purpose of business and interpersonal communication. However, Hopper (1992) predicted that research interviews and surveys could be done via telephone. He claimed that telephones would potentially have a broader sociological impact on coming generations, bridging space and time. Telephonic interviewing has been quite popular in healthcare research for several decades. Robinson & McCartan (2011) discussed the advantages of telephonic interviews compared with face-to-face interviews: high response rate, opportunities to correct misunderstandings, etc. They also discussed the added benefits of such interviews: lower cost (time, effort, and money), low tendency to fabricate responses, and the neutralization of the observer's paradox, where the effects are minimal in telephonic interviews.

Telephonic surveys/interviews offer a greater sense of personal security and safety among the interviewer and the interviewees, which justifies the efficacy of such interviews in nursing and healthcare research (Marcus & Crane 1986). We can extend this applied research via telephone to other circumstances, such as the current pandemic Covid-19, as healthcare safety foregrounds our research methods. Hence, telephonic research can serve as a novel way to conduct surveys, perform interviews in scientific and social science fieldwork or to collect data. Generally, it is appropriate to support such interviews with preliminary phrases asking for informed consent and explaining the study's objectives and expectations of the consultant. Scheduling of time and date should also be done beforehand with the consultants.

One disadvantage of a telephonic survey is engaging a consultant in the interview for longer, thus causing fatigue to either party. Such a burden is profound, especially in the elderly (Herzog & Rodgers 1988, Worth & Tiernery 1993). The power imbalance in a telephonic interview may create anxiety in interviewees; the interviewer can normalize this by letting the interviewee set the interview pace and not asking over-complex questions (Hopper 1992, Worth & Tiernery 1993).

Nowadays, telephones have been replaced by mobiles, a quintessential device in our lives. Nevertheless, they carry the same function of interpersonal communication



over a distance. We cannot deny the need for mobile. Technological advancements have made the mobile a device that is easy to carry everywhere and that offers several advanced applications. Hence, people have shifted to mobile phones for distance communication, replacing the earlier trend of telephones. Just like telephones, it is crucial for mobiles to have a good network connection to achieve clarity in exchanging information during the interview.

## **5. Mobile interviews for language documentation fieldwork**

### *5.1 Ethics of mobile interviewing*

Mobile interviews and surveys can be quite effectively used to collect data without going into the field. Several changes must be made to the approach and style of interviewing for data collection via mobiles compared to face-to-face interviews.

First, the interviewer is required to gather a list of mobile contacts of different consultants. This can be done with the help of people of the community members whose contacts we already have. They can use their social circle to get us more mobile contacts. This list of consultants should be based on a set number of people of different ages, sex, gender, class, etc. Secondly, it is necessary to prepare certain essential documents that should be shared with the consultants or consultants before the interview. The documents should include the following:

- a) Informed consent
- b) Details of interview
- c) Interview questionnaire

An informed consent sheet that outlines the objectives and scope of the research. At the same time, asking for the consultant's consent affirms the consultant's anonymity and confidentiality. The means of paying consultants should also be outlined in this sheet. Instead of paying cash to consultants, Google Pay or Paytm can be used to remunerate the consultants. The interview sheet should include details of

the date and time of the interview and check whether it is agreeable to the consultant. The date and time of the interview should be negotiable and finalized, keeping in mind the convenience of both interviewer and interviewee. The interview questionnaire will comprise questions that will be asked during the telephonic interview. It should also elicit the consultant's basic personal information and sociolinguistic background. All the documents mentioned above should be prepared in advance and sent to the consultant via WhatsApp. Cold calling (contacting strangers) has been extensively used for commercial purposes in the UK, raising concerns regarding the intuitiveness of the telephonic interview method (Smith 2005). It is an unethical way of infringing upon someone's privacy and should be avoided at all costs. All the above guidelines will ensure that the process is ethical.

According to Glogowska et al. (2010: 19), prior knowledge of the interviewer's expectations will aid in a smooth mobile interview, reducing the consultant's anxiety. Before the primary interview, a pre-call should be made to check the consultant's agreement with the proposed interview guidelines. Consequently, they can prepare themselves in advance for the interview session. Some drawbacks of this method of mobile interviewing should be kept in mind during the call. Firstly, it has been claimed that respondents are likely to be distracted and unable to maintain their concentration and energy in longer interviews (Glogowska et al. 2010). It would therefore be better to keep to shorter sessions with breaks in between.

## *5.2 Dynamics of mobile interviewing in data collection*

Before the mobile interview, it is crucial to charge mobiles in advance lest the call is interrupted due to low battery. Secondly, the interviewer should be punctual with the interviewee scheduled call date and time. The language to be used in the interview should be that which is familiar to the consultant. When talking to a bilingual consultant, the language common to both parties should be used for the interview. In other cases, the consultant can use English if they are comfortable with the language.

Each interview should start with some icebreakers by the interviewer to make the consultant comfortable. Specific general greetings can be used for this, as in 'How

are you?’ ‘How is it going?’ etc. The interviewer should keep track of the consultant’s mood and focus during the interview. If the consultant is getting bored or tired, the interviewer can stray from the interview topic and make some light talk. In other cases, where the consultant feels overwhelmed by a long session, it is best to postpone and carry out the rest of the interview later. Every interview should end with gratitude and a token of appreciation towards the consultant.

## 6. Mobile interview fieldwork: a case study of Kanauji

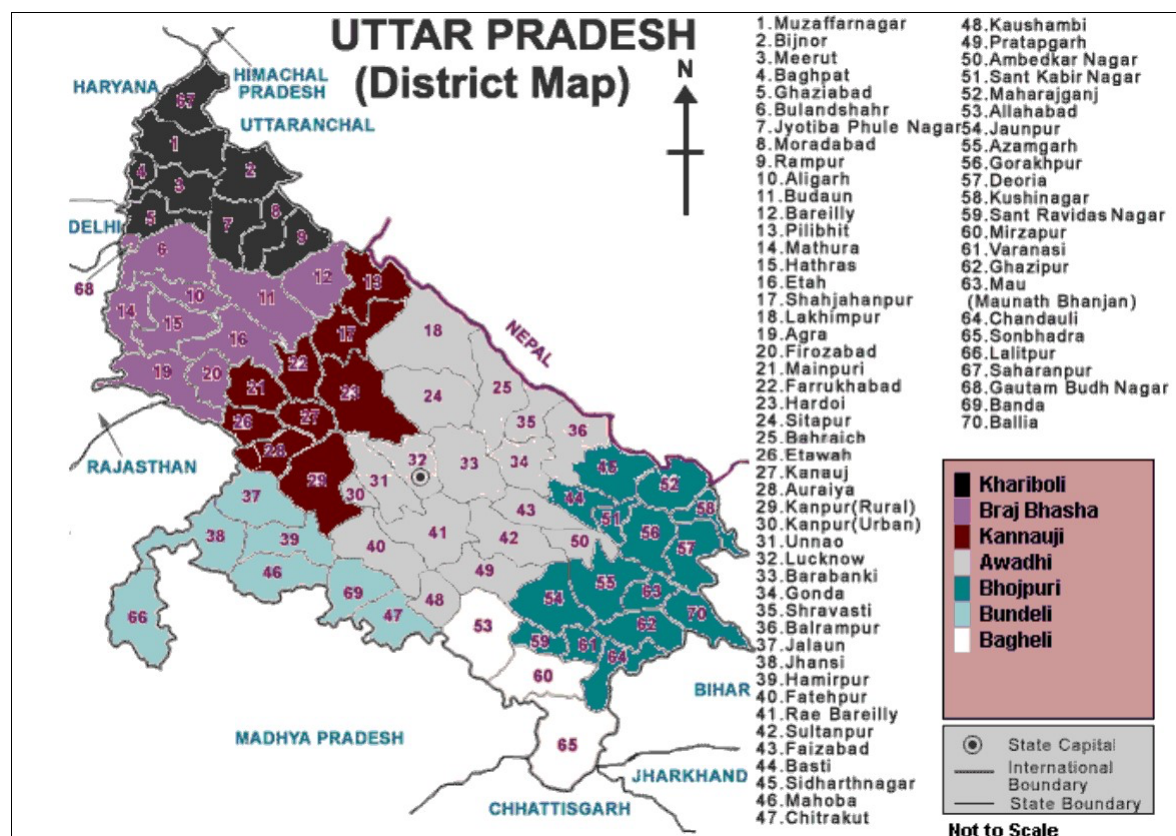


Figure 1. Map of Kanauji-speaking regions in Uttar Pradesh (Source: Languages of Uttar Pradesh, Wikipedia (Accessed 20 December 2021))

Kanauji is a dialect of Western Hindi spoken by around 9.5 million people, but it remains a sparsely documented variety. It is spoken in the Kanpur, Farrukhabad,

Etawah, Shahjahanpur, Pilibhit, Mainpuri, and Auraiya districts of Uttar Pradesh and is classified as endangered (Eberhard et al. 2021).

Mobile numbers of consultants belonging to different Kanauji-speaking regions, like Kanpur, Kanauj, Farukhabad, Mainpuri, Itawa, Auraiya, were collected by asking known native speakers. Overall, a list of 50 consultants was prepared, some of which are listed in Table 1 below.

Region	Kanpur	Kanauj	Farukhabad	Itawa	Auraiya
Gender	Male	Male	Female	Female	Female
Age	19	26	32	45	39
Name	Sanjay Sharma	Alok Trivedi	Rani Katiyar	Kalpana Dwivedi	Shivani Gupta

Table 1. Sample data pertaining to Informants from different Kanauji-speaking regions

Documents were sent to the consultants before the primary interview, comprising informed consent, details of the interview, and the primary questionnaire. These documents were sent to the consultant well before the primary interview to allow them to understand the nature of the research and their role in it. The documents were in both Hindi and English. As Hindi is the second language of most Kanauji speakers, they can understand the instructions in Hindi. Some Kanauji speakers are educated enough to know English. They can read and understand the English version of the document.

The consultants read and were informed by the documents. However, they preferred to consent and discuss the interview timings over WhatsApp chats with the interviewer. They did not favor filling in the form and sending it back as a Word file via WhatsApp as that is a more time-consuming and cumbersome task for them. Many consultants had queries regarding the research purpose and mode, which were resolved. The consultants raised various questions in the discussion on WhatsApp to get clarity after reading the documents sent to them. All of their questions regarding the interview method were answered. A few consultants called and asked questions to clarify the nature of the research: What was expected of them? How would they be

able to help in the research? Details of the informed consent and the interview were explained in straightforward terms to clear their doubts. After confirmation of the date and time slot for the interview over WhatsApp chat, the next step was to proceed with the primary interview.

The time constraint is a drawback in most mobile interviews, as it is pretty challenging to conduct long-duration mobile call surveys without causing a feeling of fatigue and boredom in either party. Hence, the interview sessions were divided into two or more slots of 30-45 minutes each. The burden of talking continuously on the telephone was significantly reduced by dividing the interview session into two slots. The interview questions were also divided depending on the duration of the interview slot. Each consultant was willing to devote at least 30 minutes at a time to answer the questions in a focused manner.

The interviews began with informal greetings and an explanation to the consultant of what to expect from the questionnaire. Each primary interview began with a brief explanation of the translation method, with an example of Hindi to Kanauji translation. A questionnaire consisting of 45 items in standard Hindi was put to each consultant, and sentences in Kanauji were elicited over the mobile voice calls. The preferred duration for each interview was 30 minutes to an hour. However, depending on first-hand experience with data collection, each interview lasted for around 40 to 45 minutes. The voice calls were very clear for the most part except for occasional disturbances in the background. These disturbances were generally due to interruptions by family members or people talking in the background, or occasional noise, etc. However, such disturbances were brief and caused little disruption to the interview process. There were challenges with poor network connectivity in the case of some consultants. Sometimes, these distractions seemed also to affect the quality of the data. However, the quality of data was unaffected by the duration of the interviews.

Beginning the primary interviews with a brief overview of the task and informal icebreakers like 'How are you?' or 'How is everything going?' enabled the interviewer to put the consultant at ease about their role in the interview. It is also the

responsibility of the interviewer to keep the consultant motivated throughout the interview session, to make it more lively and less dull. The interviewer had to explain again and give representative examples whenever there were discrepancies in the interviewee's understanding during the interview. The primary mode of data elicitation was the translation method.

Sometimes, the interviewer proposed to shift a couple of interviews to the next decided date and time slot as the consultant was busy. The consultant gratefully acknowledged this. The interview ended with gratitude from the interviewer, thanking the consultant, who was then remunerated for their time by sending money via Paytm or Google Pay.

### *6.1 Conversational audio recordings via mobile*

Griscom (2020) states that digitally recorded remote elicitation can only be done when a pre-established relationship exists between the linguist and community members. A computer with an internet connection or mobile device is needed to give specialized training to the data collector in elicitation from the speaker. Asynchronous remote elicitation happens in two stages. Firstly, the facilitator/interviewer communicates what is to be recorded through a message or live communication to the data collector; then the data collector makes a post-hoc recording of the speaker. This kind of data collection can be conducted using a laptop or mobile phone even when the internet connection is unstable and intermittent.

Asynchronous remote data collection was employed for our study. The linguist communicated directly with the native speakers and requested that they make audio recordings using mobile recording apps. The speakers could make asynchronous recordings in their own free time. Unlike synchronous elicitation, a high-speed internet connection is not required for asynchronous elicitation.

François (2019: 156) proposes a new type of questionnaire called the conversational questionnaire. The native speaker reads the conversations in a contact language and then renders the translation in their native language. François claims that this method is an efficient way of obtaining natural and well-formed utterances

from the speaker. Such conversations are anchored in a fictitious but realistic context, which avoids ambiguities and misunderstandings. This enables the linguist to focus on particular linguistic features and can be used in any stage of linguist fieldwork.

For the present study, the interviewer used Hindi as the contact language to prepare the conversational scripts, similar to the conversational questionnaire by François (2019). These scripts were shared with the consultants via WhatsApp chats. The consultants were asked to prepare audio recordings in the target language (Kanauji) following the scripts given to them in the contact language, standard Hindi. The task was to “[t]ranslate a written narrative from the contact language to the target language” (Chelliah & De Reuse 2011: 427). Later, the consultants used the sound recorder application on their android mobiles to make audio recordings of guided scripts in Kanauji. After that, these recordings were shared back with the interviewer via WhatsApp. The recordings contained first-hand data on regional varieties of Kanauji. An overview of the phonology, morphology, and syntax of Kanauji could be obtained from the recordings.

## **7. Challenges faced during the interview sessions**

Interruptions during the mobile interview call occurred due to network connectivity issues. One of the consultants residing in the interior Singhpur rural region in the Kannauj district did not have good network coverage and the network broke down several times during the interview session with the consultant. This problem was tackled by asking the consultant to move to a good network connectivity area. In some cases, interruptions were caused by the consultant’s family members, who might come to talk to him during the ongoing session. In such circumstances, it was crucial for the interviewer to be patient and to wait to reconnect with the consultant. Often, there was a lot of background noise, making it difficult to hear the consultant’s responses. Also, the consultant’s voice was unclear due to network problems. In these instances, a request was made for translations to be repeated.

A common observation was a consultant experiencing fatigue and boredom with a long, tedious interview. In such cases, it was essential to keep the consultant motivated by initiating short informal talks regarding his village, lifestyle, occupation, etc. It would give the interviewer and the consultant a necessary break from the ongoing interview and relax them. After five minutes, they could continue the questionnaire session with renewed vigour. Secondly, it was important for the interviewer to remind the interviewee of the time required for the session. This had the advantage that it was less likely to bore the consultant as he would know that he must only concentrate for a certain length of time. It is preferable to fix a 30-45-minute time slot beforehand, as was done with interview sheets so that the consultant can free that time for this event. Still, the duration can vary, and one should be open to renegotiating this time depending on both parties' comfort.

One practical problem faced during mobile interviewing was that the consultant and interviewer felt tired of holding the mobile, so switching to speaker mode was feasible. Speaker mode had further benefits as one could get grammatical judgments from interviewees' family members at the same time.

## **8. Conclusion**

Various digital platforms like Google Drive, Zoom, and WhatsApp have been used worldwide as alternative methodologies for remote linguistic fieldwork during the pandemic. These platforms are mostly accessed via laptops/computers and need high-speed internet connectivity with good power connections. However, many rural, under-resourced, or isolated communities do not have good internet connections and electricity supply, so mobile fieldwork comes to the rescue.

The technical infrastructure and socioeconomic conditions of the low-resourced linguistic communities in India also support data collection via mobiles over other social media platforms. Successful remote mobile fieldwork in Kanauji demonstrates the usefulness of this method. Mobile-assisted technologies like WhatsApp also function effectively for asynchronous data collection, as exemplified through audio



recordings in Kanauji. One can also use WhatsApp as a tool for remote fieldwork for synchronous data collection in response to non-verbal stimuli like pictures, audio-video recordings, etc. This will be ideal for communities with high-speed internet, such as urban tribal and endangered communities in Indian cities. In conclusion, mobile fieldwork documentation shows real promise for remote fieldwork in India and other parts of the world and should be integrated into the linguistic fieldwork toolkit.

## References

- BERNSTEIN, J., K. TAUSSIG, J. GODFREY (1994) "Macrophone: An American English telephone speech corpus", in *Proceedings of the Workshop on Human Language Technology (HLT 94)*, Australia: IEEE, 27-30.
- CHELLIAH, Shobhana L. & William J. DE REUSE (2011) *Handbook of descriptive linguistic fieldwork*, Dordrecht: Springer.
- DASH, Rajendra Kumar (2020) "Revitalizing endangered languages in India: Can public-private partnership (PPP) work?", in *Proceedings of 2nd International Conference on Social Sciences in the 21st century (ICS 21)*, United Kingdom: Oxford University Press, 99-137.
- EBERHARD, David M., Gary F. SIMONS & Charles D. FENNIG (ed.) (2021<sup>24</sup>) *Ethnologue: Languages of the world*, Dallas, TX: SIL International. <<http://www.ethnologue.com>> (Accessed 2021-3-12.)
- FRANCOIS, Alexandre (2019) "A proposal for conversational questionnaires", in Lahaussais, Aimée & Marine Vuillermet (eds.), *Methodological tools for linguistic description and typology* Language Documentation & Conservation Special Publication, 16, Honolulu: University of Hawaii Press, 155-196.
- GLOGOWSKA, M., P. YOUNG & L. LOCKYER (2010) "Propriety, process, and purpose: Considerations of the use of the telephone interview method in an educational research study", *Higher Education*, 62(1), 17-26. <[doi:10.1007/s10734-010-9362-2](https://doi.org/10.1007/s10734-010-9362-2)>
- GODFREY, John J. & Edward HOLLIMAN (1993) *Switchboard-1 Release 2 LDC97S62*, Philadelphia, PA: Linguistic Data Consortium (DVD).
- GRISCOM, Richard T. (2020) *Remote linguistic elicitation methods*, <<https://elararchive.org/blog/2020/06/25/remote-linguistic-elicitation-methods/>>

- GRZECH, K. & Selenia T. SHACA (2022) "Building trust on Zoom: A workflow for language documentation via videoconferencing software", *Language Documentation & Conservation*, 16, 79-97. <<http://hdl.handle.net/10125/74673A>>
- HERZOG, A. Regula & Willard L. RODGERS (1998) "Interviewing older adults: Mode comparison using data from a face-to-face survey and a telephone survey", *Public Opinion Quarterly*, 52(1), 84-99. <<https://doi.org/10.1086/269083>>
- HOPPER, Robert (1992) *Telephone conversation*, Bloomington, IN: Indiana University Press.
- LEEMAN, Adrian, Péter JESZENSZKY, Carina STEINER & Melanie STUDERUS (2020) "Linguistic fieldwork in pandemic: Supervised data collection combining smartphone recordings and video conferencing", *Linguistics Vanguard*, 6(s3), 1-16. <<https://doi.org/10.1515/lingvan-2020-0061>>
- MACWHINNEY, Brian (2021) "Talkbank: spoken interfaces and dialogue processing", <<https://www.lti.cs.cmu.edu/projects/spoken-interfaces-and-dialogue-processing/talkbank>>
- MARCUS, Alfred C. & Lori A. CRANE (1986) "Telephone surveys in public health research", *Medical Care*, 24(2), 97-112. <<https://www.jstor.org/stable/3764591>>
- RICE, Alexander (2021) "Using YouTube as the primary transcription and translation platform for remote corpus work", *Language Documentation & Conservation*, 15, 514-550.
- ROBINSON, Colin & Kieran MCCARTAN (2011<sup>4</sup>) *Real-world research*, Oxford: Wiley Blackwell.
- SMITH, Elizabeth M. (2005) "Telephone interviewing in healthcare research: A summary of the evidence", *Nurse Interviewer*, 12(3), 32-41. <<https://doi.org/10.1046/j.1365-2648.1993.18071077.x>>
- SNELLER, Betsy (2022) "COVID-era sociolinguistics: Introduction to the special issue", *Linguistics Vanguard*, 8(s3), 303-306. <<https://doi.org/10.1515/lingvan-2021-0138>>
- WILLIAMS, N., W. D. L. SILVA, L. MCPHERSON & J. GOOD (2021) "COVID-19 and documentary linguistics: Some ways forward", *Language Documentation and Description*, 20, 359-377. <<http://www.elpublishing.org/PID/241>>
- WORTH, Allison & Alison J. TIERNERY (1993) "Conducting research interviews with elderly people by telephone", *Journal of Advanced Nursing*, 18(7), 1077-1084. <<https://doi.org/10.1046/j.1365-2648.1993.18071077.x>>

## APPENDIX

### Informed Consent

I volunteer to participate in a research project conducted by Anu Pandey from the Indian Institute of Technology, Kanpur, Uttar Pradesh.

As a part of this project, the researcher will document Kanauji in as many genres and domains of the language as possible. She will study the phenomena of interrogatives/questions in Kanauji through her data collection. She will document this data through questionnaires via mobile interviews. I understand that the project is designed to gather information about a research topic of a Ph.D. student on campus. I will be one of approximately 15 people being interviewed for this research. My participation in this project is voluntary. I understand that I will be paid a sum of One Hundred rupees per day for my participation. I may withdraw and discontinue participation at any time without penalty. If I decline to participate or withdraw from the study, no one in my village/city will be told.

If I feel uncomfortable in any way during the interview session, I have the right to decline to answer any question or to end the interview. Participation involves being interviewed by a researcher from the Indian Institute of Technology, Kanpur. The interview may last approximately 45 minutes to 1 hour. The researcher may take notes or make recordings. If I want, I may become an anonymous consultant for the project. The data stored by the research will not contain my identity. But she may use a direct quote from my recordings without revealing my identity.

I have read and understood the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study. I have been given a copy of this consent form.

---

My Signature

---

Date

---

My Printed Name

---

Signature of the Investigator

**Details of the interview**

I \_\_\_\_\_ am willing to participate in mobile interview sessions with Anu Pandey related to her Ph.D. research project.

I have received the documents pertaining to the interview from her and agreed upon the same.

I am available on the following dates and times for interviewing -

Date 1:

Time slots:

Date 2:

Time slots:

The finalized date and timeslots will be confirmed with her via WhatsApp. I have been given a copy of the details of the interview sheet.

\_\_\_\_\_  
My Signature

\_\_\_\_\_  
Date

\_\_\_\_\_  
My Printed Name

\_\_\_\_\_  
Signature of the Investigator

## SAMPLE INTERVIEW QUESTIONNAIRE

### Personal information and Language Identification

1. Name of the respondent:
2. Sex:
3. Age:
4. Education:
5. Occupation:
6. City and district:
7. languages spoken:
8. languages written:

### Language Identification

12. Does your language has a written form?
  - a. If yes, which script is used? The indigenous script, Roman Script, Devanagari, or any other, etc.
13. Is there any book published in your mother tongue?
  - a. If yes, provide details.
14. Do you have a language/cultural organization such as a literary committee?
  - a. If yes, provide details.
4. What are the other languages that you know?

### Interrogative structures (to be Translated)

Simple future

- (1) kən pərhegə je kɪt̪ab?  
Who read-FUT.3MSG this book.FSG  
'Who will read this book?'

Simple present

- (2) kən so rəhə hɛ?  
Who sleep live-PFV.3MSG be-PRS.3SG  
'Who is sleeping?'

Simple past

- (3) kən pər<sup>h</sup>t̪a t̪ha je kɪt̪ab?

Who read-HAB.3MSG be.PST.3MSG this book  
 'Who used to read this book?'

(4) *sītā-ko kīsne kīṭab dī?*  
 Sita-Acc who-ERG book give-PFV.3FSG  
 'Who gave a book to Sita?'

(5) *ṭum roz kīṭna pani pīṭe ho?*  
 You every day howmuch.3MSG water drink-PFV.2SG be-PFV.2SG  
 'How much water do you drink every day?'

(6) *vo kəhā pər beṭha hoga?*  
 3.SG where on sit.PFV.3MSG happen-FUT.3MSG  
 'Where will he be sitting?'

(7) *ṭum ghər kəb aoge?*  
 You home when come.FUT.3MSG  
 'When will you come home?'

(8) *ṭum kəl bənarəs jaoge ja nahi?*  
 You tomorrow Benaras go-FUT.3MSG or not  
 'Will you go to Benaras tomorrow or not?'

(9) *rohən kīs-ki saikl čəla rəha he?*  
 Rohan whose cycle drive live-PFV.3MSG be-PRS.3SG  
 'Whose cycle is Rohan driving?'

## SAMPLE RECORDINGS

## Recording 1: čhət puɟa

Rohan: rahʊl čəʈ puʃa ker ɖusra nam ka he?

Rahul: puja ker d̥usra nam səssəti puja ija čət he. i kartik juḳl pəḳj ka mənai jaṭi he

Rohan: jo t̪johar k̪itt̪e d̪in t̪ək mənao ʃat̪ h̪e ɔr ɪnma ka hoʃ h̪e?

Rahul: je t̪ɔhar ʧar d̪ɪn t̪ək mənava ʃat̪ hɛ. ɛ m̪ã ɡəŋɡa m̪ã nəhava əʊr vrət̪ rəkha ʃat̪ hɛ. vrət̪ ke t̪aim panɪ nai pɪja ʃat̪ hɛ.

Rohan: je kəb mənao ʃaʈ hɛ?

Rahul: jo dīvalī ke baad chāṭe dīn mānaṭ jāt he

Rohan: c̣attis gḥanta ka vṛaṭ je mā paṇi nahi pi ṣaḳaṭ ḥē?

Rahul: hã bɪkʊl, əɪsɛ hɪ hoɽ hɛ

Rohan: ka koi lokgit hota hai?

Rahul: čhət puʃa ke bəhʊt̪ ləkɪt̪ hoʊt̪ hɛ ʃo gaje ʃaʈ̪ hɛ

Rohan: jo t̪johar kəhã-kəhã mənao jat̪ hɛ?

Rahul: Čhaṭ puṣa bihar əvər uttər-prədes ke hindu logən ka pərv hē. ekhē əlavə jo nepal mā bhi mənava jət hē.

Rohan: i puja ma ka astha hoti he?

Rahul: logən ki məntə he ki čəť puja kəre se ghər mā həsi-khufi əvər  
səkh jənti avəti he.

Rohan: Is čar din ki puja mā̃ ka-kāise hot̃ he?

Rahul: bhaiduj ke dusre din-se chāt-puja ki suruat hoti he. pahile din  
sēdha-namak ghi aur kaddu ka prasād banakār khāt hē.

Rohan: vr̥əṭ̪ kəb rəkho jət̪ hɐ?

Rahul: d̥usre d̥in se vr̥ət̥ ki ʃur̥s̥at̥ hot̥i h̥e. Bina kh̥ae pure d̥in r̥əha j̥at̥ h̥e. Jam č̥ar b̥əje t̥ək ək̥he baḍ pr̥as̥ad̥ bin j̥at̥ h̥e fir č̥h̥ət̥ ke baḍ bin j̥at̥ h̥e.

Rohan: čəthe dɪn surəj ka dʊdʰ čədhəo jət hɛ? je ka ədɦɪa bolət hɛ̌, je mā lokgɪt gaje jət hɛ̌ əʊr ghərən mā puɾa kin jətɪ hɛ̌.

Rohan: tisre din ka hot hε?

Rahul: તિસ્રે ઢિન પુજા કિ તૃતીયારી હોતી હૈ. મેહરિજા ગ્હાત પૅર જાતી હૈ. નૅઢી-કે કિનારે પુજા કા સમાન લૅકે વૅહૅ ગિત ગાતી હૈ ઓર પુજા કૅરતી હૈ. કુવા વ્રૅત કા ઇધવા હોત હૈ.

Rohan: ekṛa ka hot̪ hɛ?

Rahul: əɪ mā phəl, narɪjəl, məsala, ghi hoʈ he ʈəʊn puʈa mā čəɖhavət hẽ.

This audio recording is about two Kanauji speakers discussing the ritual of Chahht pooja. Rohan wants to know information about Chahht pooja from Rahul. He asks various questions to Rahul regarding Chahht pooja. Rahul tells Rohan that Chahht pooja is also known as Sassati pooja and is celebrated in the festival of Kartik Shukla paksha. Further, Rohan asks what happens during this festival and how is this celebrated. When do people keep fast? Rahul tells Rohan that it is a 4-day festival. There are different rituals that are carried out each day. Rahul tells Rohan the details of the festivities on each day. On the second day, a fast is carried out. The offering given to the gods is called 'Ekua.'

## Recording 2: Devvuthana Ekadashi

Ankit: akanṣa i devvuthana ekadashi ke din ka hoti he?

Akansha: ankṣi evo tujhar lakṣmi mata ke jad mā manava jat he. i din tṛṣi aur  
shaligram ki jad karai jat he.

Ankit: i puja kaise kin jati he?

Akansha: i puja mā ghar saf kin jat he. okhe bad vishnu bhagvan ke pav  
āgan mā banaje jat he.

Ankit: okhe bad aur ka hoti he?

Akansha: phal, mitṭhai, ganna adi chadhava jat he aur diya jalava jat he

Ankit: kab je puja kin jati he?

Akansha: vṛṭ mā ghar ke log bhagvan ki puja karṭe he. okhe bad  
subere sankh bajakar jagao jat he.

Ankit: jo bahut majedar tujhar he. aur ka hoti he i tujhar mā?

Akansha: as mana jat he ki vishnu ka janam tṛṣi ma avṛṭ he.

This recording is a conversation between two native Kanauji speakers, Ankit and Akansha who are discussing the ritual of "Devvuthana Ekadashi". Ankit asks a series of questions to Akansha- What is Devvuthana Ekadashi? To this, Akansha replies that it is a festival that is celebrated in winter. On this day, the tulsi tree is wed to the lord shiva i.e the shaligram. Ankit asks what happens after this. How is the worship done? Akansha replies to that by saying the home is cleaned for worship. After that, the feet of lord Vishnu are made there. After that, she explains the entire festivities of that day to Ankit.