

Tools and Technologies of Language Documentation

Prof. Bornini Lahiri and Prof. Dripta Piplai (Mondal)

Department of Humanities and Social Sciences

IIT Kharagpur

Week-04

Lecture-18

Lecture 18 : Challenges in the Indian field, Part 1

Good morning! Today is the 18th lecture of the course, Tools and Technologies of Language Documentation. Today, I will talk about challenges in the Indian field. So when we are collecting data, there are various challenges that we face in the Indian field. So what can be the challenges? The challenges I have basically divided into two groups, one is linguistic challenges and another is technical one. So what happens is that in linguistic challenges, you can talk about various challenges which are related to methods or methodologies, but I am not touching on them. I am talking about the logistic related issues.

So, how data collection or what data you are getting can be little challenging when you are going to the field and talking to specially endangered communities or endangered language communities. So those features, I will talk about. I will not talk about the limitations of various methods because that I have already done. For technical challenges, I can talk about various types of challenges which creates problems for digital recordings or due to these circumstances, we get lots of noise while documenting a language.

So, there can be locational disadvantage, environmental disadvantage, gathering of unrequired people, then there can be in certain communities, self-proclaimed leaders who come to talk to you or who can create problems at times. So, we will talk about all these in today's lecture. So when we talk about challenges, I am focusing more on Indian fields, Indian linguistic communities where we find different types of challenges. And most of these are actually gathered from my own experience or experience of my research scholars, who are working in various other languages. So one is problems related to what
to record.

Now, what to record here does not mean that you do not know whether to collect sentences or words. Of course, you are well prepared. You know that you are going to collect some sentences or some words, but then what happens is that specially in endangered languages, what you are eliciting can be lots of mixing or the speaker himself or herself may not be very sure about the knowledge. So, maybe the language informant is saying some word, but herself she does not know that whether it is the correct word or not. Because, there is a lot of influence of the dominant languages.

So, they may not be able to recall the original words or original structures of the sentences. And this is not only about linguistic information, but also about the cultural information. So because of the influence of the dominant languages, maybe the community has adopted the neighboring cultures. And then they cannot recall what actually they used to do, maybe how they used to perform various rituals during marriage or what were the rituals performed after death because there may be some mixing up or shift towards the dominant language. So maybe nowadays, they are following the patterns of the dominant language more and their own knowledge has been marginalized or is being not used actually, not practiced.

So there can be times when they can consult the older people and then they can recall something. So things like that can happen. Then there are technical challenges. So, I will talk more about both linguistic and technical challenges. In technical challenge, it is more about how you collect it, because there are so much of noise, so much of disturbance which cannot be avoided at times, then how to tactfully try to collect data or document the language.

So when I talked about linguistic challenges, I was talking about word collection or sentence collection structures of the sentences. So, at times for words we often see that the speakers might not be able to recall the words. So, they they cannot remember the exact word. Sometimes of course, they do not understand what the researcher is asking and sometimes they just cannot recall it. So, one speaker gave a word for one meaning while the other speaker can give another word with other meaning.

I will show you examples from Dhimal language. I will deal with Dhimal, as a case study and I will share my experiences of the field with you. So Dhimal, a little linguistic profile: Dhimal is a Tibeto Burman language, but surrounded with Indo-Aryan languages, like Bangla, Nepali, Rajbanshi, so these languages are there. So basically, Dhimal is spoken in Naxalbari area, it is in Darjeeling district of West Bengal. So, all Dhimal speakers are fluent multilinguals, they can speak Dhimal, Rajbanshi, Bengali, Hindi, Nepali and at times, some of them can also understand Santhali or Sadri.

So, we can see that there are lots of borrowing from these languages and also, the dominant languages are Indo-Aryan languages. So, there are lots of influence of these languages at the level of language and of course, at the level of culture. So now, they worship more like Bengali communities or Nepali communities rather than following their own or original forms, which we can find in the varieties of Dhimal, which is spoken in Nepal; Dhimal is also spoken in Nepal. So now what we see is that, when we ask about a word like "What do you call flower?" someone gave us the word 'bar'. And then another speaker said no 'bar' is the word which is used for seed.

Now, we do not know whether 'bar' is the common word for both flower and seed, or it is only for flower or it is only for seed. So, there can be lots of confusion; we do not know how to actually record it, how to take note of that. So maybe at times, what we can do in such situations, we can cross-check it with lots of other informants, lots of other language experts and then we can conclude. So there are different options. If the language experts cannot remember or recall what actually is the word for flower or whether 'bar' is used for both the words or not, then we do not know what to do.

So sometimes, that might happen in the case of endangered language because they cannot recall and maybe they are recalling this word 'bar', and themselves, they are not very sure, whether it is used for flower or seed because in their day-to-day discourse, they might be using words from the dominant language, not the original word from their own language. Those type of cases we often see. We also see sometimes that they give words with may be the dominant language, but then if they are reminded of the original word, then they can use that as well. Like in this example, one of the Himalayan language experts, he gave me the sentence "Baba na lo" which means that someone is calling his father. So "Father, you come".

Now, 'baba' is the Bangla word and when the other Dhimal speakers, he corrects the original speaker then he says, "ok ok it is not baba, it is aba". So now, we get the word 'aba'. So, if no one is actually reminding them about the original word or if no one is there who can actually recall the original word then we might not get that. So secondly, it was often noticed that Dhimal speakers often mix words from Hindi or Bangla or even Nepali in their speech. So now, you can see in this example, the first one "I am going to work in the field", the word for the 'work'- 'ka:m' is being used.

'ka:m' is used in Bangla, 'ka:m' is used in Hindi. So, this word 'ka:m' for 'work' has been used. So now, we do not know whether it has been borrowed or whether the speakers cannot recall the exact word for 'work' or whether there is no word for work in demand. So, there is no concept of a word as such like 'work'. So there might be words for fishing, different types of work, but work in general, might not have a word.

That can also be a thing. So, those can be confusions for us where we do not know, whether what to do. Again in the second example, you can see the adverb 'slowly slowly' or 'aste aste' is being used, which is again a Bangla borrowed word. So when someone is saying "I will walk slowly", that can be used, the reduplicated form which is used for the adverb is the borrowed one from Bangla, it is not the original Dhimal one. So again, if the speaker or the researcher is not aware 'aste aste' is a Bangla word, might take it as a Dhimal word.

So, one needs to also cross-check with the linked language or the dominant language. So for me, it was easier because I am a Bangla speaker. So, I can identify that 'aste aste' is a Bangla word, but if the researcher does not know the linked language or the dominant language, then one might take it as an as a word of Dhimal; one might not know that it is actually borrowed from Bangla. So, there can be instances like that where we see so much of borrowing or so much of mixing up of two languages that at times, it becomes difficult to identify the original one or the borrowed one. Now, this sentence also tells us or creates confusion where we do not know whether the language uses this type of reduplication form for adverbs or not.

Because this whole structure is borrowed, 'aste aste' is a reduplicated adverb. So, we do not know whether these type of structures are found in this language or not. So for that, what we can do, we can of course, cross-check with the other similar examples and then we can know whether these type of reduplicated adverbs are being used in the language or we only see such usage when there is a borrowing from the dominant language. So, that way we need to cross-check and verify, we just cannot collect something and say that this is happening in the language. We need to also cross-check with the linked language, with the dominant language and also with the language within it, by comparing it with other examples.

Sometimes, the informant replace the Bengali word with the Dhimal words when he or she was corrected by someone. So, they might be using the Dhimal word or Bangla word, but they can change it, depending on their surroundings when people are suggesting them that use this word, then they can actually do it. But when the informant is not reminded of the original Dhimal word, then they might not say it, then they are sticking to the Bangla word. And this instance of Dhimal is not an isolated example. This is what is happening across India, you can find in different speech communities, where we are talking about endangered languages or languages which are actually not being spoken much.

So in those cases, what we are seeing that these type of tendencies are there, where people are actually forgetting different words and then, when you ask them all of a

sudden, it is difficult for them to recall. We can see a similar example from Mahasui language. Mahasui is a language which is spoken in Himachal Pradesh. So here, what you find is that respondent gives data, where you can see the first example, "hal lagata tha". So "lagata" is a Hindi word, which has been used, but when he is corrected, then by another language expert, then he said "hor lada deo tho".

So now, you can see how these structure of the sentence actually changed. So, when you are working on the syntax of the language, then also it is important to know what is the original form, because with the borrowing, it can also affect the syntactic structure of the language. It is not only about borrowing a word and using it, it is not only limited to the lexical level, it affects the morphosyntax of a language. So here again, you can see how in these examples, you can see how again, morpho-syntax has been affected with the borrowing thing. So here, in the 4a example, you see that "na esho lo", "esho" is a Bangla word.

So, basically you "come come". So, if you are calling someone "come near", so they are using the reduplicated form and strangely one word is Bangla, another word is Dhimal. So, 'lo' is the Dhimal word for 'come', while 'esho' is the Bangla word for 'come'. So it actually semantically getting reduplicated, but structurally it is not getting reduplicated. And similarly, in the another example, they are saying "na centa lo" which means actually 'near'. So, 'centa' is 'near', so, 'come near'.

Now, we have to look at these data and parallel data and see whether these structures are common, like 4a where reduplication, semantic reduplication is happening with the meanings being reduplicated with different words, one being borrowed and another, the original word. Or is it ungrammatical in the language because in many of the languages, it is not possible; it is actually wrong. So you can try speaking that in your own language. Even in Hindi if I want to do that, that would not sound very good. So, might be 4a is actually ungrammatical sentence, but the speaker or the language expert gave us that data, that sentence.

So, it becomes again important to know whether this data is actually correct, actually usable within the community or not. Because in the case of borrowing also, of course, borrowing are accepted by the linguistic groups and they use it in the certain communities, but then there are certain restrictions. So every borrowed word cannot be used in all type of structures. So there can be certain restrictions, like a 4a example can be ungrammatical, it might not be a grammatical sentence in the language. So, it is very important for us to cross-check and know and at times, this cross-checking can be very difficult as the language experts, they do not know which form is actually correct or which is ungrammatical.

Now, talking about technical challenges, I will talk more about the logistic issues. So till now, I was talking about the language aspect of it, how when we are collecting we can get little confused with the variations of data which can be grammatical, which can be ungrammatical or things which are actually practiced, which are not practiced, but just we are getting, just the language expert is creating something because he or she cannot recall the original structure. So, there can be issues related to that. When we are talking about technical challenges of course, we know that when we are documenting a language, we are using audio-video recordings along with of course, notes and photographs. So among other things, the quality of the recording is of utmost importance for long-term storage and hence, needs explicit attention; that is what is mentioned by Himmelman 2006.

And if you remember in my first class, I said that documentation is also about preserving it for a long period of time. So, the recording, quality of the recording should be good, very good. And of course, various scholars including Himmelmann have described the ideal environment for data collection, what is the ideal environment where you can collect data. So, one should avoid the noisy places and noise can be created by lots of things like humans, animals, birds and also blowing of the wind can create noise. So, noise can be created by lots of different things.

So, you might wonder, but if we are recording in a room where there is no noise, but tube lights are there, those can also create sound waves which can be recorded. So those type of noise should also be avoided. However, the real field scenario presents many many challenges when we are thinking about this ideal field, where there will be no noise and we can record a good quality of recording. We may not get that in the real Indian fields. Because when we are sitting in a village, collecting data, then there can be different types of things happening which can affect our recording.

So it is very difficult at times to control these situation, how things are happening and we need to adjust to the situation. So that, we can at least get some recordings some part of the languages well documented. So, what type of advantages or disadvantages that we see? Advantages, we can actually take out from these disadvantages which I will talk about. So, there can be locational disadvantage, there can be environmental disadvantage, gathering of unrequired people, self-proclaimed leaders of the community. So what are these? So when I talk about locational disadvantage in many of the houses, actually in most of the Indian villages, houses, what do we see is that they have house and outside house, there are courtyards made where people sit generally, visitors or guests are asked to sit there.

So inside the house, people generally do not enter. So, guests are not allowed quickly to enter inside the house, they are welcomed outside the house. So in the pictures, you can see how there are courtyards being made outside the house. And you can also see a person carrying a chair from inside, actually he was carrying the chair for us to sit. So they will bring you chairs and tables and they will make you sit there and this is actually common across houses, across villages which we have found.

So every house has this veranda where they actually ask their guests to sit, they are served tea, everything is done in this place which is outside the house. And this place is actually open, as you can again see in this picture where data is being collected. So in these places, what you see is that the place is open from every side, there is no wall and when you are collecting data in an open area of course, there will be lots of noise. There will be noise of the birds, animals and again, there will be noise of the wind itself. In this picture, you can see that the person who is holding the recorder is also, there is a wind shield used.

So that the wind, noise of wind is not recorded in the recorder. And then in some cases, we do find sometimes a roof on the veranda. In some places it is not even that, it is totally open and there we are asked to sit and record the data, where it is impossible to make the place silent because there will be noise lots of noise. So mostly, every house have their pets and they move freely. So it is not like they are cage, they move freely in village houses.

You will easily find pigs, cow, hen, duck, every type of pets are there available and of course, dogs are there. So you do not know, when they will start barking or making noise. So all of a sudden, they can start barking or they can start making noise and you do not know how to stop them. They are not within anyone's control. So in those cases, your recording is hampered and that can happen for a long period of time; you are actually sitting for half an hour, there are lots of noise and you cannot record anything because with a dog barking in the background, you will not get anything recorded.

The human speech will not be recorded properly. So that can be a thing and there can be lots of things which can actually divert your attention. I will like to share one of my experience where I was collecting data in Totopara and I was sitting and I was collecting data, and then there comes a piglet, baby pig and it starts chewing my sandal and then I got scared and I shouted. So what is happening and then the language experts, they said "that is ok, it is just a baby pig, do not worry, it is fine". So, things like that can also happen where you get distracted and then you actually disturb your own recording, but then we have to adjust to those experiences and we have to carry on with our work. And of course, we cannot even hurt the sentiments of the language experts.

We cannot say like "why is it happening or I do not like this". You cannot say things like that which can actually hurt the community because they are helping you with their data and they are there for you. So, you should always consider that. So there can be lots of noise and sometimes, you cannot actually ask the language experts to keep on repeating something. So maybe if you were recording something and suddenly, a dog barked and your recording was not good, you can ask twice maybe, but if you keep on doing that for every part of the recording, the language expert may feel exhausted and irritated at times because he or she might not like repeating it several times and repeating also wastes your time.

So if you get a narration and if you ask the whole narration to be repeated, one thing is that you will not get the same narration again because every time the person is narrating something, there will be certain changes in the narration. Secondly, it will be a waste of time, every time you are asking or half an hour someone spoke something and if you want to repeat that thing again, that is waste of time and at the same time, it is also wasting of your resources, like if you are carrying batteries for your recorder, batteries can also get exhausted. You might not have proper charge in your recorder. So these things can also happen because in the fields, you might not easily get your devices charged, there can be problems of electricity. So, you have to carry your backup and then those backups can also get exhausted.

So repeating things might not be possible in every cases, for few sentences or few words it can be applicable, but for narrations, you will not get the same narration; that is the most important thing. Rural houses have wells outside these courtyard. So people work around the wells, they will do lots of things like fetching water, washing utensils, clothes, everything. People also cook outside their house.

So in the picture, you can see that a lady is cooking. So of course, there will be noise and you cannot ask them not to cook and you do not know when to go because every time you are going, in most of the time, they are cooking. In the morning time, they will cook. So that time actually during cooking, she can talk to you. Other times maybe, she will be out for some other work.

In the evening, there can be certain other types of problem. So now, those are very tricky area, how to actually collect. In certain cases, in the other picture, you can see me holding a camera where this person is from Mahali village. He is making bamboo and there in the picture also, you can see far away, a girl is also sitting with a recorder and collecting some data. So in those cases, what you can do is that the person is working, you can actually document the whole process and keep on asking what is this, what are you doing,

what do you call this process and then, several terms related to the process. So whole community you can actually document in that way, but then of course, there can be lots of disturbances, lots of disturbing noise around of the daily household work.

Other than that, we also see trees make sound, there can be sounds of the birds. So, if you think the lady is cooking in the morning, maybe in the evening I will go to collect data, but in the evening birds makes lots of noise. So, there will be lots of noise in the evening and that is actually impossible to collect data. In most Mahasui households, people get free at night, but in night, there is no electricity and then, video recording is not possible. So, how will you actually video record without lights because they do not have proper electricity connection.

Weather changes can be a challenge. So if suddenly, there is rainfall of course, there will be sounds of rain; how are you going to collect data? So before going to field, we should also cross-check with the weather, but then sometimes, it is not even predictable and we do not know; there is lots of lightning happening, thunder sounds are there, rainfall is there, then of course, in those situations, one cannot collect data. Also, we see that in Indian villages, people are very well connected, they actually behave as a bigger family. It is not like different families staying together, it is like all are related to each other and when there is an outsider, actually everyone tries to gather to that person. For at least first two three days, whole village will be there gathering; they want to see what is happening, they want to know what is happening. So, gathering of these lots of people actually makes lots of noise and then there can be children also.

So, mothers are carrying their children. In the picture above, you can see there are two children. When we are there to record the mother, the children are of course, there. And for children again, they are unpredictable; we do not know when they will shout or cry or talk. So, there can be lots of noise due to that. And in those cases, you cannot even say that do not bring the children because if you are recording the mother of course, the children will come if they are there.

People also start talking among themselves. So, they are watching you and then they are talking among themselves that what is going on this and that. So, these type of noises can also be part of your documentation, which we can try to avoid, but at times, it is not possible to avoid. You might not ask them not to come there or leave the place. So in the recording, in the left side also, you can see that there are few people standing.

So, that always happen. Now you see in certain cases, that there are certain members of the community who might be little educated and they at times want to do good for their community and when you are an outsider, they take you as a threat and they want to

actually save their community from you because they are unaware about you; they do not know much about you. In those cases, they can be judgmental about you and they might not allow you to collect data or document their language. So sometimes, they have a very genuine issue, like they think that you can be a threat. So, once we were asking some language experts to sign in an ethical form where we ask them that we can use the data for academic purpose and this person comes and he says, "No, no, do not sign.

They will take all our land and home and everything". So, there is that fear. So, that can be a threat for us in documenting a language; that can be a problem for us. Then sometimes there is some political or some different types of agenda or propagation made by these people. They want you to record what they want, you cannot record your own thing, but they will say you or dictate you. So, those things become very difficult then, because that might not be your interest, that might not be what you want to collect. But then they will dictate you or direct you and they will only allow you when you listen to them.

So, there can be issues like that, but then we have to handle them with care and respect and sometimes making them understand what is our real motive, our work is more academic or how we want to actually create or produce something for the language. So, if you try to make them understand, sometimes we might succeed, actually in most of the cases we will succeed, we succeeded. So, that is what we cannot start arguing with them, but we can make them understand the whole situation and that might help. So in conclusion, I will like to say that there are can be various issues when we are documenting a language, when we are actually in the field. So, there can be some issues which are unpredictable, which I did not talk about, which I did not face, but you might face.

So, there can be certain things, but we should be mentally and physically prepared for those type of challenges. There can be confusion regarding linguistic data which we have collected, but we should always try to cross-check it before making any generalization, we should always try to cross-check from different sources or from some other language experts. So, challenges related to technical aspects will also be there, there can be lots of noise, there can be certain disturbances, disturbing people or disturbing elements, but we should try to manage them in a very calm manner so that we can actually properly do our own work. In most of the cases, if we try to make them understand, what we are here for, then people do understand and then they also help us. So, every challenge should be tackled with care; one should not start arguing or starting a fight with the person because there can be circumstances where it can happen.

So, the language experts can be very judgmental about you and more so, if you are a

lady. So, these things can happen, but one needs to tackle them with care and then try to go back to one's own work. What is the aim? Aim is documenting the language and trying to do that. And in the process, it is good to involve most of the community members. You can explain them your work and you can take their help, then they will be very helpful to you.

So, there are lots of challenges in documenting a language, but these are worth taking. These are actually very enriching experiences and also part of the job. So, we should not be scared of them, rather we should face them strongly.

So, these are some of the references. Please go through them. I hope you enjoyed today's lecture. Thank you!