

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-19

Lecture 19 : Challenges in the Indian field, Part2

Hello everyone! Welcome to the lecture of Module 4 which is Methods, Ethics and Challenges in the Field. And the title of today's lecture is Challenges in the Indian Field. So today, we will discuss how to address the heterogeneity issues in relation to multilingualism in Indian context. And we will also discuss the idea of fuzzy boundaries and pressure of pure languages. When we conduct field work, how do we come across these issues? We will consider 'language choice' as a major factor while collecting data in Indian context. We will also talk about the status and social hierarchy issues in terms of languages and language change.

Finally, we will also see whether linguistic homogeneity can be a problem in field as well. So, I will begin the class with two quotes, which are very important for our theme today. The first quote is from Anvita Abbi's Manual of Linguistic Fieldwork, where she says, "The unique nature of linguistic heterogeneity coexisting with the typological homogeneity among various languages and language families demands a different approach to fieldwork". So, this is regarding the Indian context.

So if you, at one hand, look at the heterogeneity issues that I will also explain in today's class, what kind of heterogeneities can be challenging for us, when we conduct field work in Indian context. And also, if typological homogeneity can be a problem as well. If we find there are some common features across different language families, that can be a problem for us as well. The second quote is from Prabodh Pandit's work and which is a very well quoted line, "India is a sociolinguistic giant". So, what do we mean by 'giant' here? So, we will talk about the sociolinguistic features which are very important for us to understand, in order to collect data.

So, there are certain theoretical tools which are super useful for us before we start collecting data. So we have to understand the sociolinguistic factors, before we go to the field and also while collecting data, we can try to go through them again, in order to clarify our own understanding. So, when we talk about the linguistic heterogeneity of India, we can quickly look at the Census data, which is the 1961 census. It mentioned that there are more than 1600 languages in India and you can understand the complexities, which is very different from many western countries. So, there are certain languages which are contributing to the multilingual nature of the country's linguistic behavior.

But here, we can see that there are many languages and we will also see that, there are some factors which are adding some more complexities to it. So, we know that language contact and convergence, these are important sociolinguistic factors, which are also important in typological literature. Some of you already know about it, specially those who are conducting advanced research. So, I am saying that we need to pay attention to both contact and convergence issues. I already mentioned in some of my previous lectures that, what do we understand by language contact and we need to pay attention to this while collecting data at the field.

So, Annamalai talked about functional multilingualism. So, I am asking you to stop here and think about this term and think seriously, what do we mean by functional multilingualism? So, while you will be collecting data, I will talk about it today that it can be confusing for us. I can give some examples and I can tell you that functional multilingualism can act as a barrier for us if we do not understand the kind of functions that different languages play in a multilingual society. So, I have a few questions here which are of course, taken from Annamalai's work that I have quoted, but I will ask you to think about these questions before going to the field. The first is, how is the nature of language maintenance among multilingual speech communities? I hope you remember the idea of maintenance, we talked about it previously and we also looked at the difference between language maintenance and shift, if you remember.

So, how is the language maintenance in the target speech community, where you are going for your data collection, needs to be understood before you arrive at your field. You need some data, maybe some secondary work, maybe some piloting data, but this is super useful for you to understand the nature of maintenance. It is also important for you to figure out what kind of code avoidance is there among the multilingual communities? So, I hope you remember the term 'code avoidance'. I talked about 'code switching', and 'code avoidance' is something where a particular speech community is not using certain codes. So, you can refer to it as some kind of a shift, but you know there is some kind of avoidance because of stereotypes and attitude issues, if you remember.

And the third question that I am asking you to think at this point is how is linguistic purity related to multilingual speakers? So, the idea of pure language, which is again associated with stereotypes and attitudes can compel the multilingual speakers to not use their ancestral language or avoid certain kind of code, which is termed as non standard or impure by the other communities, who are in the same neighborhood. At this point, I will also introduce this idea of fuzzy boundaries. So, if you look at Khubchandani's work, you will know more about it. So, what do we mean by fuzzy boundaries? You do not have a clear-cut boundary between one language and the adjacent language, the language next to your target speech community. So, you do not actually get a particular village or a particular hamlet where you move to a different language.

Maybe you are moving to a different community, hamlet, but there is some kind of transition. If you remember I talked about transition very briefly, there is some kind of transition which is there and it is very important for you to understand it. Languages are not very watertight compartments where there will be no interaction between them. So, it is very important for us to understand. And we also try to understand in case of Indian context, that what kind of changes that we observe for administrative borders.

So, I already gave example of Bihar-Bengal border or Jharkhand-Bengal border earlier, where we find that there were multiple changes of borders in the same area. As a result, the official language of one area altered several times. So, we will find that the choice of the speakers will alter too. So, these are important for us to understand before we go to the field. So, we usually do some background studies on the linguistic landscape and try to understand the choice of language based on boundaries as well.

So at times, we have found that in Indian context, there are some new provinces which were created from the older provinces or you can say states, older provinces new states are created. So, the example of Chhattisgarh and Telangana which are comparatively new. So Chhattisgarh, you will see it is a state with several communities, while Telangana, you do not get that kind of a variety that you get in Chhattisgarh, but creation of new provinces will also come up with new set of language policies. And new set of language policies which definitely affect the language use, the everyday language use of the speakers. So, you have to be aware of these factors because boundaries affect the language choice.

Along with fuzzy boundaries and change of borders, I should also mention that 'continuum' is an important notion. So I am sure, some of you are familiar with this idea of 'dialect continuum'. If you have attended another course that have been offered by IIT Kharagpur for linguistics, we have discussed continuum in detail. So, continuum is something where you have a chain, where you have the transition that I was talking

about. You have a chain which is slowly diffusing some kind of linguistic features, whatever it is.

So, you find that there is some kind of continuum through state boundaries and it adds to several complexities due to multiple social linguistic factors. So, the factors that you are listing for one area or one community can be different for another community in a different geographical area. So, there are of course, both historical and geographical reasons. So when we talk about this continuum, we look at the existence of one speech community in terms of one space, we will try to look at the complexities related to sociolinguistic factors as well. And we also need to pay attention to the fact that the boundaries are crossed freely, it is not that this boundaries are not crossed.

So the state boundaries inside India, we find that people cross them for various social purposes, may be to access education, health or occupational needs. So these boundaries are crossed and as a result, you have two different states adjacent states with different official languages, where speakers are continuously moving from one side to another for various social purposes and it affects the language use. We talked about the idea of pure language. So, I will come back to it because we need to understand it and emphasize in terms of Indian context. So, if you think about the identity issues that we discussed previously, we find that at times there are loss of identity as the communities move towards purism.

So, it is again a kind of language shift, which may lead to endangerment. I am sure, you are now aware of this fact. So, we need to understand the kind of purism that is very much there in the community's mind. So, it is a collective notion. You have to understand it with all the subtle layers that exist.

So, of course, there is subjective discourse rationality and we have to understand these layers that we find, in order to access pure language, the so called pure language, I must say, and language shift due to the pressure of pure languages. It is such a case that the direction of shift is not very easy to find. You know that language shift is taking place because of the target use in some kind of pure language, whatever it is, but you do not understand the direction very easily. Maybe you have some kind of heterogeneity even there. So, you see that the pattern of shift is not uniform everywhere.

We already discussed about different kinds of language shift and how do we understand different aspects of it. So we, at times, find that the nature of shift is different, even in the same area. So, we need to understand it. So, when we are collecting any kind of data, we do some piloting for the communities, we find that the shift is different for different social layers. So, these are the stratification issues and I will come back to it in today's

class,

again.

So, we find we have difficulties in determining the nature of stratification, we collect some kind of data related to language stratification. If you are familiar with Labovian literature, I am sure you can understand what I am saying, what kind of social stratification we find through language use. And we also find that the dominant communities have separate codifications. So, we need to understand the language use of the dominant community, even if you are looking at some minority speech community in an area, otherwise the picture will not be complete. And there is culture-coded purism by elites.

So, if you are collecting data from a community, often it is not very easy to figure out the culture-coded purism. So, at times, we do not collect data from the elites, but it is also very important to get the idea of the elites, in order to understand the community's shift, the direction of shift and the nature of shift. So at this point, I can mention the term 'elite bilingualism' or 'elite multilingualism' and there are certain problem areas that I should mention. So, if you are collecting data from the elite representatives of the community, then you will figure out that you can see a trap. So, what kind of trap? Because if you are looking at the data of only the elite community members, you may not get the pattern of mass use, you may not get the pattern of maintenance that is still existing in the community, because the elites will tend to shift to some extent.

It is not everywhere, but it is very common in Indian context. So, the choice for the elites, we find that it is sometimes different from the non-elites and we need to, at times, if you take data only from the elites, you miss those points. You think that you know that is the default languages of the community, which is not in most of the cases. So, we have to collect data from different layers of the society, otherwise we would not get the real picture; it will be a partial picture, which is not fair. So, we talk about formal and informal bilingualism or multilingualism.

So, what do I mean by formal bilingualism? So, it is basically what you acquire through educational institutions. So, I will elaborate it later, but you acquire some language which is official, which has a "pure status" in the society. So, you acquire it as a part of formal education. But the social stratification may need something else, may be the prestigious language is something else and there are two prestige languages in one area. So, we discussed about these issues previously, if you remember.

And at this point, I will also mention Pandharipande's term "functional load", which is extremely important for Indian multilingualism. So, how do we understand it? We find that in different domains, people use different languages. So, we are already familiar with

this idea, if you remember. So, functional load is something that we need to figure out when we think about multilingualism. So, we are talking so much about multilingualism because this is quite complex in Indian context and if you are not from India and collecting data in India, you have to understand all these theoretical tools and all these layers, otherwise it is going to be difficult.

If you are a community member, you understand what kind of multilingualism is there in your community. There are different names of it. I am not going into it, but it is important for you to understand the types. So, language choice as I have already mentioned, it depends, and we have to focus on it. So, if you are using a say, Translation method, you are asking your informant to give some data.

So, you have to constantly be aware of the choice, because if you are getting data in something, which is not their choice all the time, there can be a problem. For collecting vitality data, if you already have a set of vitality data with you and then going to collect some data through translation or you are like, may be one of your team members has already collected vitality data and shared with you, it will be good because you will have some idea of the choice, otherwise it can be all mixed up. So, we talk about context, we talk about formality, if it is a very formal domain because if you are an outsider, there is a chance that the community member will try to maintain the formality, not use the non-formal codes, but at times, if you can also observe two community members talking to each other and make some notes, it will be useful. So, it changes according to context as well. So, ideas of stereotypes will also compel them to avoid some codes.

As I have just mentioned today, you have to be aware and the kind of code avoidance, which is common among the speech community needs to be noted. All of us know that multilingualism is a norm in India. It is not an exception. It is the default phenomena in case of Indian speech communities, wherever they are, in whichever location they are, whichever condition they are. So, there are no monolingual speakers in Indian context.

We do not find any monolingual speakers. Even if you are staying in a cosmopolitan location and mostly using English, even then, you have some kind of use of some Indian languages. So, it is very old and it is default. So, we have to understand why we are often saying that it is a norm and not an exception. It is not an exception. So, there are variations in all data in Indian context.

So, the problem is how do we take account of the variation data. So, maybe your target is not to concentrate on the variation data, but even then, whatever data you are collecting for documentation purpose, you need to take into account that there is some kind of variation. So you can find that there is a difference from demographically multilingual,

but functionally monolingual countries. You can have some countries where you have different demography, but functionally, it is only one language which is being used everywhere.

So, it is very different. Indian context is very different and that is why, we are emphasizing on all these theoretical tools, that you need to understand. There are some problems when you collect data. So, if you are collecting data from only one age group and or gender or only one location or social class, there can be a problem. Maybe you will miss out something, maybe you will not be aware of the potential variation, maybe some amount of useful data is being missed because you are not looking at the other layers. And also, if you have some preconceived idea about the nature of data, it can be good if you have some background reading, but it can also be problematic if you think that this is a stereotyped variety.

This is something which is like I am going to get this kind of data X or Y. So, that can be a problem. So, we have to be open all the time. Whatever we get is data for us and we have to collect more to add to our corpus. And at times, we also expect homogeneous codes, it is a problematic statement.

I mean if you think that we are always expecting homogeneous codes, the codes are always heterogeneous. It will have mixture with other languages because hybridity is also default. So, you have to understand that what is being collected by you and when you will gloss your data, you will automatically understand the nature. So, if we talk about the research tools, you already know about some tools which are useful for us. So, we use some structured questionnaires, but I hope you also remember Role Play right? So, you can collect some data through role play, where you ask the community members to act like someone and you can get rid of the formality questions, because for role play, you are actually getting the real-life-situation-kind of data.

And we also talked about picture books, and you get the local options and perspectives through the picture book, when you collect data from them. We get data on the nature of language attitudes among the community members. And when we think about it in Indian context, we of course, think about variation, but we need to pay attention to the fact that there are both interspeaker and intraspeaker variation. So, what kind of variation is there in your case, you need to be aware of that. So, the status of language is important because when you are looking at the language.

If it is a minority language, then there is a chance, as I already told you, that they try to move towards some kind of, so called, pure or official language. So, some of you already know that there are 22 scheduled languages, which are recognized by the Constitution of

India, but we find that around 95 percent of India's population speak the scheduled language at least, in certain domains. So, there are some problems. So, if you are collecting data on a scheduled language, you would not face this problem because you get some amount of data which is needed for your purpose, but if you are not working on a scheduled language, there is a problem that you will definitely receive some kind of data in the official languages, which is there with your data that you are collecting in the target language. So, there will be some kind of mixture, code switching, etcetera.

So, you have to understand the pattern. And how do we get data from domain specific-use, is another related question. For domain-specific language data, we rely on observation, at times both participatory and non participatory. And we also get some kind of help from role play, but you are also collecting domain-specific vocabulary, but the way of use, may be for syntax you need to think about it in a more subtle way, keeping the socio-linguistic factors in mind. There are social hierarchies when we talk about languages. So, there are different kind of hierarchies in rural and urban setup.

I am not talking about suburban setups, you can think about them. So, in rural areas we had stratifications based on caste mostly. In urban areas, we had stratifications based on education. So, why am I saying 'had'? Because it is changing and we have found that, it is changing very fast in the globalized world. So at times, you can say that rural and urban stratification, they are overlapping.

So, I mostly work in rural areas and I find that people who have migrated there from the cities, they are affecting the language use of the rural population. So, that is something which is becoming very common in more recent days and we have to pay attention to these. Are people trying to imitate someone? Is it the natural language choice for the community members? So, these questions are relevant for us. So at times, we pay attention to them and we actually find that there are changes of language practices in the globalized world. We find that there are educated urban, who are migrating to rural areas for job.

For example, school teachers in the areas where I work, they migrate from cities to small villages all over West Bengal. So, this is something that I always notice. And the ideal situation that we had in mind, that lower caste rural population will not move, it is also changing because for jobs, they are also moving to big cities. So, the hierarchies, the older hierarchies that we had in Indian societies are changing very fast and as a result, the linguistic behavior is changing too.

So, there are some new challenges. If you are familiar with Blommaert's work, who is talking about super diversity. It is something which is a new challenge for post

globalization world, which is basically a never-before situation, where the traditional migration patterns, as I have mentioned, have changed. So, if you are collecting data where people have enough access to new media, people have enough access to new media in most of the Indian villages due to the advancement of technology. So, you can figure out that the traditional social stratifications recorded may not work. So, if you, for example, are looking at some published data from 1950s, you may not find the same kind of stratification in today's world in the same village.

And we need to pay attention to new migrations, what kind of migrations are taking place, are there people migrating for jobs in different places and may be they are connected to the family members through social network, which is affecting the language use of the community members, etcetera. And the influence of new media, as I have told you, will of course, contribute to language change. So, access to television for example, will affect to language change. So now, if we talk about language change and documentation, I am sure when we are discussing about all these factors in Indian context, you can already figure out that we are entering the domain of language change, which is extremely useful in the domain of social linguistics, as well as language documentation. So we know that language changes has accelerated because of new media, that I have just mentioned and migration, the changes in migration pattern.

So, do we consider the changes while collecting data? Do we have different age group-based data, that I have already mentioned? So at times, you find that younger children and older children of a same school have different kind of data, girls and boys have different kind of data because girls do not have access to technology in most of the Indian villages. So, you have a major change which needs to be figured out. And at times, we also considered that the changes of Indian societies altered the choices of codes, practices and code avoidances in domain-specific data use. So, all these things are important and we have briefly mentioned them. So, if I talk about typological homogeneity now, if you remember the quote by Abbi, 2001 that I have mentioned in the very beginning, we also consider India as a linguistic area.

We talk about the heterogeneity in terms of multilingualism, but we also talk about some kind of homogeneity because of the areal features. So, India is a linguistic area and we have different language families which have similar features, be it phonological, morphological and syntactic. So, there are of course, family specific features, for example, tones in Tibeto-Burman languages. So, we will talk more about it in next few classes.

So, family specific features are often overlooked if we try to generalize. So, we know that there are some typological features which are common. So at times, we do not pay

attention to the language-specific features. So, this is very important in Indian context. You have to do some research before collecting the data on a particular language family.

You know that there are tones in most of the Tibeto-Burman languages. So, you have to train your ear for clarifying and understanding the tones, but you know when we talk about the homogeneity based on arial features, it can act as a problem. So maybe, we will not pay attention to the specific features. So, background research will help you to understand this. I will talk about some common problems because it is important for us to understand in Indian context. So, we find that there are some link languages which are common in highly multilingual zones.

So, these link languages are used in marketplaces for cross community interaction and I already gave examples of Sadri, if you remember. So, there is a common problem that we find in Indian context is.. the link languages are often mistaken as the language of the minority communities by some researchers.

So, community members have a plus point here. If you are a community member and attending this course, I am sure you are aware of the different layers. You have the link languages at some domains and some layers and there are domain-specific use and because of the stereotypes, the communities shift to the link languages, as I have already discussed with you. So, outsiders may not figure out this. So, I will give an example of collecting data from Kurukh and Mundari speech communities and what I have found that at times, outsiders have a difficulty to understand between Mundari, Kurukh and Sadri. If you have not done enough homework regarding the features of Mundari and Kurukh, you may just transcribe Sadri and say this is Mundari and this is Kurukh.

So, this is a major problem which many people have actually done. Another common problem is you know to generalize and label it as a dialect of a pure or standard language. So, may be you are collecting data on the linked language and you think that this is the language of a community, you do not have data on the actual language use, instead you have collected some link language data. So, you figure out that there is some kind of structural similarity in the standard language. So, you conclude that this is a dialect of some standard language. So, we will talk about standardization in future and I will come to this point again, but at this point, it is very important for you to note because this is a common mistake in Indian context.

So in the conclusion, we will talk about or we will figure out some points that we have already discussed. I hope this will be useful for you while discussing the challenges in Indian context. First of all, we have functional multilingualism as a major challenge, because we do not understand the functional load, we do not have a clear-cut

understanding of functional load. Fuzzy boundaries can be difficult for us; we do not understand the use of the area where the use of language is being diffused, in which direction it is being diffused and what kind of choice is there, in terms of this diffusion. And the pressure of the so called pure languages will compel the speech communities to shift to a different language and some kind of hierarchies will be there and which can be unfamiliar to the outsider-researcher.

So if you are not familiar with the hierarchy, the linguistic hierarchy and the pressure of pure languages, it can be a problem in Indian context. And homogeneity can also be a problem for language documentation if we generalize, if we not look at the language-specific or family-specific features and do not do enough homework beforehand. So these are the common problems that we need to address. I am sure you can remember more problems, please note down those problems in your notebook along with the problems that we have discussed. I hope both will be useful for you. So, these are the key references. Thank you!