

Tools and Technologies of Language Documentation

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

Department of Humanities and Social Sciences

IIT Kharagpur

Week-01

Lecture-01

Lecture 1: Introduction

Welcome to the first lecture of Tools and Technologies of Language Documentation. To know what is language documentation and why it is important, we need to know what is language endangerment. So, in today's lecture, I will cover three concepts mainly about language endangerment, how are the languages distributed across the world and what are the causes of language endangerment. because when we know about language endangerment, then only we can understand why language documentation is important and what are the techniques or tools to do that. So, let us see what is language endangerment. Like any species language can also get endangered.

So, language endangerment is like any other thing or species where it is predicted that a particular language can die or go away from the earth no one can use it. So, it will become endangered. Many languages are falling out of use and being replaced by the other languages. So, when we say that a language dies or a language has become endangered that means, that the speakers have shifted to another language that does not mean that the speakers have also died.

So, basically what we see is that sometimes speakers of a particular language shift to other languages and they stop using their own language and that is the time when that particular language dies. So, like the use of Hindi and Bangla instead of Khorta or Mahali. So, you can see in Mahali communities, at times the younger generation they prefer using Bangla compared to Mahali And if this trend goes on then after few generation there will be no one left to speak Mahali and then we can say that Mahali has become a dead language. So, David Crystal says "a language is said to be dead when no one speaks it anymore". So, there are lots of definitions about language death or language endangerment.

So, you can also explore UNESCO's Atlas which shows the list of endangered languages. There are around 196 Indian languages which are endangered. And I hope that you will go

through this link and you will explore the other endangered languages of India. And you can also see that some of the languages are marked as severely endangered, while others are marked as critically endangered. So, I will like you to explore little bit about it, because in my next class, I will talk about these levels and why some languages are marked as severely endangered while others are marked as critically endangered.

So, I will talk about these stages in my next class, until then you can always explore these terms. So, it is also predicted that 50 to 90 percent of the world's languages will disappear in the next century. So, you can see that at a very fast rate languages are dying. And that is why 2019 was declared as the International Year of Indigenous Languages by UN. And UN also declared International Decade of Indigenous Languages beginning in 2022.

And now you know that language endangerment is a big issue, it is very important and that is why UN is also talking about it. And this awareness has spread across the world, people want to save languages and one of the steps of doing that is by doing documentation, because documentation also helps in revitalizing a language. But then there are various process involved in it which you will know through this course. So, as I mentioned languages become endangered when speaker shift to another languages. So, speakers of many smaller less dominant languages stop using their own language and shift to the dominant language of the locality.

So, parents may start discouraging their children in speaking their own native language, because they might think that it is useless, there is no educational material available, there can be some negative attitude related to the language. So, we will also talk about language attitude. Professor Dripta will talk about that in her next class. So, as a consequence what happens is that no speakers are left who speak their own native language and they shift to the dominant language because of various reasons. So, a language may become dormant or extinct.

Perhaps only in recording or written records we can find them, if the language is well documented we can only find the documents of the language available, but no speakers. But in that case, there is a hope that we can revitalize the language or reintroduce the language among the community using the documents. Language death is very serious issue as language death is related to various other things because language is not only a means of communication, it is also related to culture, it is also related to one's identity. So, when a language is lost, the identity is also shaken and that is why language is not only limited as a means of communication. It is like a vehicle which carries the information of culture, collective memory, values, knowledge system from one generation to other.

And when this language is lost all this collective information, memory everything gets lost

along with the language. Moreover, it has been seen that biological diversity is related with linguistic diversity. So, when we see that biodiversity is decreasing we also see linguistic diversity is decreasing and vice versa. So, if linguistic diversity is affected it also affects biodiversity. So, they also are very closely related and studies have shown that.

So, it is estimated that number of languages in the world is around 7000, but the ratio is not 1 to 1 right and that is why there is a problem. So, 96 percent of the world's population only speaks 4 percent of world's languages. Now, you can know why so many languages are endangered. But then we can try and save these languages from dying. So, we need to know what are the causes of language endangerment? Why does a language die? So, there are basic four main categories in which we can categorize the causes of language endangerment.

One is natural catastrophes, war and genocide, overt repression, cultural, political or economic dominance. I will talk about each of these points. So, when we talk about natural catastrophes, you know that there can be disease like COVID, pandemics, there can be other type of natural calamities which can affect a community. like we had Tsunami. So, what happens is that when these type of disasters happen then at times the whole community is washed away or it vanishes within few years and that can lead to death of that particular linguistic community and the language.

So, these can happen and we have seen in history that there are many communities which died out because of these natural catastrophes like earthquake, tsunami or very few speakers were left for a language. Then there can be a cause like war and genocide. So, when Europeans were settling down in Brazil they were actually killing Brazilian indigenous people to settle to get their land. So, when they were actually hunting those indigenous people and when they were hunting them or when they were killing them, then the whole community was killed and the language was also killed along with the community. And so, at that time lots of languages died along with the communities.

So, that is again that has been seen in history. Overt repression is again there where we see that a particular speech community is forced to not use its own language or is forced to use the dominant language often in the name of unity or assimilation. Why it is done? Because it is believed in some places that using one language across the nation might be more convenient and economic than maintaining linguistic diversity and that is why communities which speak languages which are not spoken by the dominant languages or communities which speak smaller languages, they are not allowed to use their language in the public domain and slowly they are encouraged to shift to the dominant language. In some places we have also seen that these people who use smaller languages in public domain are even punished. So, they are scared to use their own language and in that way

slowly the language dies out.

So, in Native America, it was seen at one point of time during 19th century, 1880s, there was a policy where it was said that native Americans cannot use their own language, they are forced to use English. And in some documents it was also found that these local languages or native languages were compared or said to be barbaric. And it was said that they need to be trained, they need to be cultured and they need to be forced to use English only because these barbaric languages or barbaric dialects cannot be allowed to use. So, in history we see lots of such instances where languages were forced to die, languages were actually killed. So, that is overt repression.

The most common cause that we see is cultural, political or economic dominance. Here there is no group which is forcing the other group to not use their language, but there is that dominance. How does that happen? You can see economic dominance by the other language or cultural dominance. For example, in rural areas, when people do not find job or their traditional job do not give them proper money they are forced to shift to the urban places. They are forced to shift to the cities and towns and when they migrate, slowly they forget their own language because they have to learn the dominant language of the cities.

And slowly, this is a very slow process when they are forced to migrate because of occupation, when they are forced to migrate because for the educational reason to get better education, along with living their own village they also leave their language behind and this is very common. This has been seen. For example, Mahali community is also known as basket making community. They are a smaller community and they used to make bamboo baskets. But now we know people generally buy plastic baskets, they do not buy bamboo baskets anymore.

So, now most of the Mahali people they are forced to migrate to bigger cities in search of better opportunities. We also know that Mahali children are now going to schools and colleges in Kolkata or other bigger cities. So, when we see that most of the people from the village are migrating to the bigger places, bigger towns and cities, they will of course, slowly forget their own language, they are forced somewhere to use the dominant language. So, at this point I will also like to make it clear that using the dominant language cannot be the only cause for language endangerment because our brains can very easily handle multilingualism. At the same time, we can use more than one language that is fine, but the problem is when we do not speak our native language.

When the Mahali speaker is only using Bangla or Hindi, but not Mahali. But if parallelly both are used Mahali and Bangla, Mahali, Hindi, Bangla, English everything then then that is fine and that is also what we see in many places across the globe and more so in our own

country because our country is multilingual. So, in rural places we see this type of economic pressure. Tourism is also promoted in some of the places, but tourism is promoted to just you know earn like they just showcase their thing, they just showcase their culture, their performances to earn money out of it, which is also known as folklorized or folklorization which means tourists are attracted showing these indigenous performances, but in reality people who perform, they might not know their own language.

They just know the performance. And this I can say from my own experience. I have seen a group which were performing in a particular language. They were dancing to a particular indigenous song, but when I asked them the meaning of the songs word by word, they were unable to say that because they did not know it. So, just they are performing for the sake of performance. And in these cases what do we see that a particular culture or a particular part of a culture is just used like a commodity.

It is commodified for economic gain of course, but then that does not mean that it is helping in using the language or revitalizing the language. So, cultural power or cultural dominance of the neighboring language can at times force people to shift to other languages. When we see that there are smaller communities like Mahali, Dhimal which are surrounded by the dominant languages like Bengali, Nepali, then there can be a chance that these smaller communities will have the tendency to shift to the dominant language. Because more facilities are available in these dominant languages which are not there in the smaller languages.

There can be political reason. So, if education policies which ignore or exclude local languages or they do not recognize the local languages or in extreme cases if the local or the smaller languages are banned from using in the public places or in educational domain, if in certain domains they are banned to be used, in those cases also the speaker will try to shift to the major language. There can be historical reason like colonization, boundary dispute, the rise of one group and their language variety to political and cultural dominance. So, we have seen some of the historical reasons as well. So, we know that there is a group in West Bengal, I will not name the group, but there is a linguistic group which was termed as a criminal tribe. So, now, the group which speaks a particular language is termed like this of course, by the colonizers, then they will not feel like using their language because the moment they will use their language, people will notice their identity, people will know their identity and then people will also know that these are the communities which were labeled as criminal tribes by the colonizer.

So, there is a negative attitude related to that and that brings me to the next point which is attitudinal. So, what is the attitude related to the speech community? Whether the speech community thinks highly about its language or whether it feels bad about its language,

negative attitudes is related to its language. So, depending on the attitude also we know whether the language can be endangered or not. So, sometimes when there are historical reasons that can bring negative attitude to a particular speech community. The speaker of a particular language might feel bad about his or her own language and in that case the person will not like to identify himself as part of that particular speech community.

So, there can be these causes which can lead a language to endangerment. Now, I gave you the example of Dhimal. Dhimal is a language which is spoken in Darjeeling district of West Bengal. So, they basically worship nature, they worship trees, rocks, hills and now they are surrounded by communities like Nepali speakers, Bangla speakers. These communities they go to temple, they worship idols and all.

So Dhimal community see that the neighboring communities are going to temples they are worshipping idols and all and this creates a pressure- indirect pressure due to which now we see that there are temples in the Dhimal village which was not there earlier. And dhimal word for the priest is 'dhami'. So, 'dhami' is not only a person who is priest, 'dhami' is a person who can heal, who can give medicines, who can also worship. So, all this combination is dhami. Now, this word 'dhami' is being replaced by the word 'pujari' which is a common term used in many of the Indo-Aryan languages.

So, why Pujari, because with the establishment of a temple in Dhimal village, they need a priest to worship the idols of the temple. And that is why this person is called Pujari because the neighboring communities call him Pujari. And then we can see that there is a shift from Dhami to Pujari. And now people know who is Dhami or they still maintain the tradition of Dhami. But if we think about future generations after two or three generation may be then they will not know what is dhami or who is a person who is called dhami.

So, this whole knowledge will be lost, this word will be lost and also the knowledge which dhami had. Dhami knew how to cure some of the local disease. So, that knowledge might also be lost along with this whole word. So, when we look at one particular word, it is not only one word; actually a world view is related to that word and that is why it is important to talk about language endangerment or language documentation.

Fishman talks about domains of language use. So, what are the domains? Domains basically mean participants with whom I can use my language, what are the topics in which I can talk in that particular language and where, what is the location where we can talk in the language. As you know that in many of the academic places in our country we need to use only English. Like in many schools, children are punished if they do not use English in the classroom. So, that means, it is a domain in the classroom domain with the teacher and the classmates the children are expected to use only English in English medium

schools. So, similarly if we look at the domains then we can know about the status of a language, whether the language is endangered or not.

How do we know that? If we can use a language in most of the domains where we represent our self then that means, the language is quite healthy, but if the domains are very narrow that means, the language can get endangered. Like, if I take the example of Kurmali, Kurmali speakers use Kurmali within the family. They can use it in the local environment like with the shopkeeper, local people. They can talk about different topics like health issues, politics. They can talk in home, in the shop, in localities and all.

So, they can use it. But in the classroom, in most of the schools, in that locality, it is difficult to say that whether they can use Kurmali or not. Similarly, when we take the example of Hindi, we know that Hindi can be used in colleges, in the universities one can talk in Hindi with a doctor. That means, in professional domain also it can be used Hindi one can talk in Hindi with a lawyer, with a doctor, in different domains of education, health issues within the hospital, within the university, but can the person do the same with Kurmali? Can the person also use Kurmali in professional domain? So, that we need to explore and that is, Kurmali is just an example we can see that with various other languages, whether the language can be used in different domains. You can even explore this with your own native language or a language which is surrounding you. So, you can take any language, may be a language which is around you or a language which you speak or your friend speaks and then you can think about the domains where the language can be used and where the language is actually restricted from being used.

So, now, you can see these two diagrams or figures. Here you can see that Mahali, again I will come back to Mahali, Mahali can be used in home. But if it is compared with Bangla you know that it can be used for entertainment there are lots of things available for entertainment like books, journals, YouTube channels and all these things social media everything is available where you can see representation of Bangla. Educational materials are also more or less available in Bangla. In office Bangla is widely used in West Bengal.

So, in all these domains Bangla can be used. But when we talk about Mahali can it be used in schools? Do we find educational materials in Mahali? Do we find proper language technology available in Mahali? We do not. Secondly, can we watch movies made in Mahali? No. That means, source of entertainment like newspapers, journals those are not available, educational materials are not available and of course, in government or any private offices also, Mahali is not very well represented. So, this shows that of course, Bangla can be used in much more wider domains compared to Mahali.

Mahali has a narrower domain. Now if you look at these circles and if I ask you to assume

that the yellow circles are representing for example, Bangla language Bengali language right. So, these are Bangla and then you see Mahali. So, this bigger circle is Mahali. So, maybe if you imagine this bigger circle to be a person who is a Mahali speaker. You will see that this person will have to shift to Bangla for education, will have to shift to Bangla for entertainment.

will have to shift to Bangla for healthcare facilities if that person goes to hospitals in the cities and all. If this person has to write an application then again he or she will have to use any other official language like Bangla, Hindi, English. So, because in many of the offices they will not understand Mahali applications. And also the government advertisements or private advertisement anything, the hoardings and all the it is very difficult to find in Mahali. But in this region where Mahali is spoken easily Bangla representation can be found.

Now this person is exposed more to Bangla than Mahali. So, automatically there will be a tendency of shift from Mahali to Bangla which can be witnessed. So, if the person is not very loyal towards his or her language, if the person is thinking that ok I have to use my language in my home otherwise there is a tendency that the language might shift from Mahali to Bangla. because there is continuous exposure of the other language and parallely there is no facilities available in Mahali like as I mentioned no means of entertainment or education or other things available. So, that will force indirectly the Mahali speaker to use the other dominant language of the area. And when the person will face these difficulties it might be a chance that he or she will not like her children to learn the language Mahali rather will ask the children to learn Bangla or will not encourage the children to learn Mahali rather will ask them to learn Bangla or other languages and this is a case which we can often see .

So, basically now I will conclude and I will say that what you learnt from this lecture is that languages when it is predicted that a particular language is likely to become extinct in the near future it is termed as the endangered language. It happens because the speakers of a language slowly stop speaking their own language and shift to another language which often carries more power and prestige. There are four main causes for language endangerment. among which cultural, political, economic dominance of other language is the most common cause. When a particular language dominates over other language then the dominant language is used in most of the domains.

While the domains of the use of the subservient language shrinks often leading to endangerment of that particular language and that is why the language becomes endangered. So, in the next lecture we will know about the various stages of language

endangerment and how we can know whether a language is endangered or not. So, I hope you enjoyed this lecture. Thank you.