

Tools and Technologies of Language Documentation

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

Department of Humanities and Social Sciences

IIT Kharagpur

Week-08

Lecture-37

Lecture 37 : Case Studies of Language Documentation from India

Hello everyone. Welcome to the Swayam NPTEL course, Tools and Technologies of Language Documentation. This is Dripta and today, we will be listening to Lecture 37 of Module 8, which is Case Studies. The title of today's lecture is Case Studies of Language Documentation from India. Today, I will be talking about different works on endangered languages from India, which have been done by different people and organizations. Initially, I will talk about the government initiatives from the government of India, then, I will introduce some research on endangered language by various academic institutions.

Then, I will also briefly talk about some individual linguists work on endangered languages. Then, I will talk about work from organizations outside India who worked on Indian languages. And finally, I will share some of the works that we have done. The first work that I want to initiate is the SPPEL initiative, which is the Scheme for Protection and Preservation of Endangered Languages, which is done by Central Institute of Indian Languages, Government of India.

And you can see the website here, you can go through it and explore more by yourself, but I will briefly talk about the kind of resources that you will be finding on the website. You will get some resources on documentation, for example dictionaries, you will get some bibliographies which will be useful for your work. There are some sketch grammars and apart from that you will also get the details of various programs on endangered languages which are offered by the SPPEL program. So, there are some highlights of this SPPEL project that I would like to share with you. If you go through the website, you will find that there are lists of endangered languages based on different zones.

So, if you are planning to select a language from a particular zone of your choice, you

can go through the list and select languages which are available in the particular zone. You will also find the details of some events on endangered languages of India, conference, workshop or any other training program for example, the SPPEL website will give you the details. Apart from that, the website also includes some video documentation of different festivals of indigenous communities in India and some other ethno-linguistic data. So, you can also access them for your work. Now let me talk about some projects which were funded by the University Grants Commission.

Some of these projects are ongoing and some of them are finished. So, I will initially talk about the project from the Indira Gandhi National Tribal University at Amarkantak, Madhya Pradesh. You will find this website and this is the list of the languages that this particular university selected. You can locate the languages based on different states of India. You can see that this particular university covered endangered and lesser known languages from Madhya Pradesh, Gujarat, Telangana, Andhra Pradesh mostly and some languages of Maharashtra as well.

So, if you are planning to work on languages or planning to select some languages for your work, this website will be useful for you. Tezpur University Assam also had a Centre for Endangered Languages and you can explore their website for resources and languages and it will be useful for you if you are planning to work on languages of North East. And if you can look at the map, you will figure out that there are languages from Assam, Manipur and Nagaland which they have chosen for their work. And you can also look at the location of these linguistic communities based on the map. So, this is another resource which will be useful for you.

So, if you want to explore more, for example, if you are planning to learn more about the sketch grammars, dictionaries or similar works done by other people, bibliographies, etc, these things can be accessed from the Tezpur University website. Next is the Center for Endangered Languages or CFEL from Vishwavarti University, West Bengal. So, they have some resources in their website too which you can explore. You can find the repository of Sikkim University Sikkim. And if you go through the website, you will see that the list of languages are mentioned in the first page of their website.

Sikkim University worked on languages from Sikkim and northern part of Bengal, including Bhujal, Gurung, Magar, Rai, Rokdong and Sherpa. So, you will also find some resources on these languages on their website. And there is a community home page which you can access. So if you look at the map which has been shared by the team of Sikkim University, you can look at the location of these languages. So these are mostly located in as I told you northern part of Bengal and Sikkim.

So you can also go to the detail of villages and the locations where these languages are spoken. If you are planning to work on these languages, this map is going to be useful for you. Next I will talk about the work by Jadavpur University website. So, there is a link that I have given where you can download a questionnaire which has been developed for conducting field work on endangered and indigenous languages in India. And this is done based on Indian languages and there are some language specific, family specific information as well.

So, as we were also involved in this project, you can find some of the learnings that we have shared with you already from our field work. So, you can download the resource from the link and you can also find some other works done by Jadavpur University like there are dictionaries for example, the dictionary of Kurmali which was a trilingual dictionary. And as a part of this endangered language research team, this university conducting language awareness campaigns with children as well from different languages which will be useful for you as well. So, these are the languages that were covered by Jadavpur University. All these are languages from West Bengal though they are spoken in other states as well.

So, Kurmali, Toto, Koda, Mahali and Lodha, you will get some resources. If you are planning to look at or searching for resources on these languages, you can get some information from the website. So, I have not included all the universities who did this work or who are still doing work on endangered languages from Indian languages. So, you can for example, go to the websites of Hyderabad Central University for work on Sora, Central University of Karnataka, Berhampur University, Odisha for some more languages spoken in Odisha, Central University of Jharkhand for some Austroasiatic languages and more. Next I will talk about some individual works which are seminal works for endangered language in Indian context.

So, you can go to this website of Professor Anvita Abbi, this is andamanis.org and you can get resources on Great Andamanese, Nicobarese and you can also get some more information of other languages in the section called other. So, you can get some resources, you can get news on the field work because it was a challenging field work and also overall idea of the endangered language research based on the linguists work. I have listed some more works which are important in this regard. So, for example, Shobhana Chelya's work, Reasons for Endangerment, Why Language Documentation Matters is an important work which was released in 2021.

You can go through it and learn more about Indian context. You can go through Uday Narayan Singh's work on a grammar of endangerment in this edited volume and learn about overall endangerment issues in Indian context. And there are also some specific

works on different languages in Indian context. I have given example of the work on Nihali and Korku by Shailendra Mohan in this journal which you can check from the link. So I did not mention all the works, I have just mentioned a few works which are very important for you to note.

And if you are looking for other works on endangered languages done by linguists in India, I think all these resources that I have suggested can be explored. There are some works on Indian languages which have been done by people who are not from India. So, such a resource is the Living Tongue Institute resource. So, I have given some details of this Living Tongue Institute research. So, for example, you can find this living dictionaries on different languages.

And if you can look at the map, you will find that there are some languages from India as well. For example, Living Tongue Institute has some repository on Mundari. So, you can look at this work which will be useful for you I suppose. There are some more works on Indian endangered or lesser known languages done by people who are not from India, but working extensively on Indian languages. So, I have given four examples which you can explore.

The works by Alexander Coupe are mostly on Tibeto-Burman languages. So, if you are planning to work on Tibeto-Burman languages, you can go through this website. Works of Dr. John Peterson, it is mostly from the languages of Jharkhand and Chattisgarh, both Dravidian and Austroasiatic. So, if you are planning to work in this area or explore either Dravidian or Austroasiatic languages, you can go to this website.

This website from Mark Post is again related to Tibeto-Bowman languages. So, if you are planning to work on Tibeto-Bowman, go through it. And finally, the website of George van Driem, which is again of Tibeto-Burman will be a very useful link for you, where you can get the repositories, sketch grammars, dictionaries and resources of other similar works from the websites. So, again I will tell you that this is not the end of the story. There are many other works by people who are working on Indian languages or they have worked on Indian languages previously.

I have given you some examples, so you can explore on your own. I will talk about the research that we have done at IIT Kharagpur and I am sure that we have both Professor Lahiri and me, we have given examples of several languages in our lectures. So, you are partially familiar with the names of the languages, but I will tell you that some brief account of the projects that we had or some of them, we are still having. How we are approaching these endangered or lesser known languages through different types of work. So, the first example that I have already given, but I will mention again is the work on

Mundari.

So, we finished two projects on Mundari. We tried to develop some language technology as we collaborated with Microsoft Research Labs India for this work. and our target was to develop some language technology. For the first project we focused on West Bengal and for the second project we focused on Jharkhand. So, we tried to develop some corpus in Mundari which will also be useful for future works if someone is planning to work on Mundari.

There was of course a community connect. So, we talked to the community members, tried to develop a network. and we talk together. For example, computer scientists, linguists, community members together regularly tried to have meetings and built a network to get some idea of language technology. I will also tell you that we made an app for language learning specifically for children as a part of our projects.

So, we designed one very basic app with vocabulary items from selected semantic domains and we tried to make it bilingual or even trilingual in some cases, tried to experiment with audio instructions as well. And the basic target was to help the children from the Mundari community to learn the language when they are losing it. We are also working on a Tibeto Burman language called Toto. We have given example of Toto for various kind of areas that we discussed in this course. So, this is the project that we are having and we are also trying to develop some kind of language technology for selected Tibeto-Burman languages and Toto is a part of it.

There are other languages as well. So, we are trying to develop a corpus and analyze the possibility of developing some language technology. Again we have some community connect because some community members were a part of this project. So, they have contributed to some data of their own language and then we together tried to develop some technology interventions. The next language is another Tibetobaman language which is Dhimal. So, this is being done by this project called Intergenetic Study of Cases of Languages of West Bengal by Professor Lahiri.

And this project includes study of analysis of case systems of different languages. Dhimal is one of the languages that she has chosen. This is an endangered Tibet-Burman language, which is spoken in West Bengal, and there is also some language technology interface for this project. So, this is the project that I am having which is on different hybrid languages, again it is funded by IIT Kharagpur. And the target is to create data bank on various hybrid languages which are in difficult position, which are in vulnerable position and children are losing them.

And the target for this project is to analyze the data after collecting the data create some data bank and also develop some pedagogical material. So, as I told you in detail the importance of creating pedagogical material if you remember. So, we said that creating a repository and developing some TLM or teaching learning material can also help a language to survive. So, this is exactly the task that I am trying to do along with the team to develop some pedagogical material to revive these languages which are gradually becoming non-existent.

So, and I am working in West Bengal. So, coming to the conclusion part of this talk, I will tell you briefly that we discussed about different kind of resources that can help us to get some idea on the endangered languages of India. You can definitely get resources of endangered languages of the world from different resources like UNESCO, Ethnologue for example, but in Indian context, you can also explore different kinds of resources to get some idea on Indian languages. So, the first kind of resources that I discussed was the Government initiative to conduct research and documentation to protect and preserve endangered languages in India. This is the SPPEL project that I have discussed and also I tried to give some overview of the resources that are available on the SPPEL website. For example, dictionaries, videos, sketch grammars, other resources, etcetera.

And, you also have some detail of the languages which they have chosen. Then, I talked about the documentation initiatives by different academic institutions, how the University Grants Commission funded different universities in India to conduct research on endangered languages. And I gave example of from several universities not all, but I tried to give you some idea of the kind of resources that are available through the university websites. For example, in some cases you get statewide names of endangered languages which can be helpful for you. You can get maps of the location of these endangered languages created by some universities.

Some universities have online repositories like dictionaries, sketch grammars, academic papers and you also have something like questionnaires for use in Indian context. So, you can access all of them for according to your work. After that I talked about individual research that have been done by Indian linguists on different languages and again, I gave some examples which are very important and quite well known, but you can again go through other works on the endangered languages that I have not mentioned. Then, I talked about the different works on endangered languages. These works are talking about Indian endangered languages, but they have been done from outside India either by academicians, academic institutes or by some non-government institutes.

So, you can access their data, you can access some media resources that they have shared and also community based interventions, videos, talking dictionaries and all these

resources are available for public use. And towards the end, I talked about the work that we have done or we are still doing at IIT Kharagpur. So, we mostly work on languages of West Bengal. So, we gave examples of languages like Mondari, Toto, Dhimal, hybrid languages. So, all these are considered as languages under our projects, and it will give you some idea of the kind of work that we do.

So, I hope all these resources will be helpful for you and you can search for more resources for your work and if you are planning to search for a language for your work, these resources will be helpful. Hope you enjoyed this lecture. These are these references, selected reference. Thank you.