

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-17

Lecture 17 : Different Approaches of Language Documentation

Welcome to the 17th lecture of the course, Tools and Technologies of Language Documentation. Today, I will talk about different approaches of language documentation. So I will start with history of language documentation, I will talk about different approaches to documentary linguistics, timeline of different types of field works and different types of field work depending upon the community which can be like field work on a language, field work for the language community, field work with the language speakers and field work by trained language speakers and about active participation of the community members. So, let us begin with the history of language documentation. The term 'language documentation' has been used historically in linguistics to refer to the creation of grammar, dictionaries and text collections of undescribed languages. So it was already there traditionally, where data was collected from the field and based on that grammars, dictionaries and other materials were created.

So, that was taken as language documentation. However, it emerged as a sub-field of linguistics during 1995, where there was a crisis being understood about the dying languages of the world. The world became more conscious about the endangered languages. So previously, language documentation was not focusing on the dying languages as such, but it wanted to create resources like dictionaries and grammars and other materials for the under-resourced or under-described languages.

But with this emerging crisis or with the fact that people understood that languages are dying at a fast rate, the focus was more on the endangered languages. So, we understand that language documentation of course, focuses on endangered languages because they are dying and we need to document them before they vanish, but it can also be done on different aspects of a healthy language as well. So, development of technology made the

whole process easier and also, we saw a shift from the traditional method where the experts were there and data was collected manually, manually means no digital equipments were there. So generally, what happened was researchers collecting data and they were writing it, transcribing it at that very moment sitting in the field and language experts had to repeat it several times and then the researchers they were writing it. So whole process was done in the field itself .

So of course, it was very very difficult. With technological aid, it has become little easier and also, we are talking about creation of metadata and archiving, storing and preservations, which are have become more accessible or easy because of the development of technology. So, there are different approaches to documentary linguistics, that is, Himmelman talks about language documentation as a process or he says that language documentation is concerned with the methods, the method in which the process is being done, tools and theoretical underpinnings of compiling a representative and lasting multipurpose record of a natural language or one of its varieties. So basically, he is talking about the process, the method, tools and theoretical underpinnings. So what we are seeing is that, he is talking about the whole process and what is involved in it, but we also see that it is only about the process.

While Woodbury talks about after the process thing as well. So, he focuses on outcomes of the documentary linguistics or what happens after documenting a language. So, whether things like primers, orthographies, dictionaries, grammars are created or not and he also specifically says that things for non academic audience. So, the target is non academic audience. So if there are outcomes, can they be used by the speech community or not.

So, not only about description or IPA transcription of the data, but also creation of some useful thing for the speech community. Because if the data is collected and stored or kept in just say, in the digital form and they are transcribed in IPA and then kept, it is of no use for the speech community. They might not understand what IPA or what will they do with the IPA transcription anyway. So, but if dictionaries are created, grammars, primers, orthographies, then that can actually make the language flourish; it can save the language from dying. So, Woodbury also talks about the outcomes of the documentation of a language, while Himmelmann talks mostly about the whole process of documenting a language, which again has different dimensions.

So, different types of field works that we have seen which have actually emerged over the period of time, what we see is field work on a language. So, what was it? It was the more traditional way where the researchers went to the field, they collected their data and came back. So, they were basically working on a language and trying to capture different

aspect of a given language. Field work for the language community. So now, you can see the difference when field work on a language is there, it is about the language only, how well we can collect data, how well we can describe it, analyze it and all.

But when we are talking about 'for' the language community that means, we are trying to give back something to the language community in the forms of primers, in the forms of dictionaries, grammars. So, what can we give back to the community? How can the process of data collection or documenting a language be useful for the speech community? So, it can be of use in the academics or for the academicians, but how is it going to help the speech community? So, that aspect was added. And then again with the time, we saw that the language speakers were involved, where we saw that they were not only passive part of the whole process, not like you are asking something and they are saying something; they were made active participants. So, there was an involvement of both the researcher and the language experts. So, it was not only about this sitting somewhere and just responding to what you are saying, they were also actively participating.

They were also saying that what do they need or why a dictionary should be created in their language or they could also guide you what to collect and what you can get from where. And then field work by trained language speakers, which is very recent, what we see here is that in many of the projects which is working on language documentations, they involve community members who are trained in field work. So, they are not just community members, they are trained community members and they are working in language documentation. So, what we are seeing is that since they are part of the community, they know what to collect, when to collect, how to collect, all these aspects they know; they know their community. And since they are trained, they know how to do it because only being the community member might not make them suitable for the purpose because they are not technically trained.

But when they are technically trained, they know how to conduct the whole process, how to actually collect data. And as they are part of community members so, that also helps. So, this is a very recent approach and this actually has been proved to be very useful, where the community members are self-motivated or they are motivated enough to get training in linguistics and field work and then they start working for the language. So, what can we see is that the community members, they come forward, they are working in their own languages, which is actually very beneficial for the speech community as well. Because they know what should be documented first, what can be created for the community, they are well aware about it because they are the insiders and as outsiders, researchers cannot always decide that.

So, these are different types of methods which can also be mixed and which can also be adapted, depending on the speech community. So what we can see is that, the categorization of communities have been done based on the relation of land that we have seen, where communities which do not stay long in a particular area. So, we see there are certain communities who keep on travelling. So in India, we have Banjaras and then there are also various subgroups to that. So for those type of groups, who are continuously travelling, so they might be staying in a particular place for say 6 months and again, they are travelling to another place.

These type of groups we find across the world, where there is an urgency to collect the data, because as they keep on travelling, sometimes they lose their language, they go to some other place and then adopt another language. And in those cases, what is also not possible is to ask some of the community members to get trained in field linguistics. So, one can use traditional questionnaire method and quickly collect data because there is an urgency, there is no time, they will shift to another place, they will not wait for your work. So, these type of communities generally, what we can adapt is traditional method where researcher visits the community and quickly collect data. And then there are fixed location settlements, there are also various subdivisions to it.

I am not talking about that, but I have given a citation in the reference where you can read about it. So there, these community members, they stay in a particular place. In such communities also, you can see where some people are very educated, some communities are quite educated in the sense that they are literate. So for them, they can easily get trained in field linguistics and start collecting data or help in documenting their own language. Sometimes we see these type of motivations directly come from the speech communities.

There are certain members who want to get trained in the community in the field methods and start working on their own language, may be at times at the individual level, may be they are not associated with any project, but start working on their own language. So, that is possible or sometimes we see that there is no one who can be trained in this field, there is no motivation. People might not be much literate. So, in those cases of course, we cannot use the last method that I talked about, where the member is trained in field linguistics and start working. So there what we can do? We can actually involve the speech community, we can make them active participants and start documenting the language.

Over the time, they can get influenced and start themselves getting trained in field linguistics. So, that can be a thing. So depending on the community, different methods can be adapted. So depends on lots of thing, one major criteria is the land and their

relation to the land, whether they are traveling continuously or are settled groups. There are other relations like whether they are educated, literate, about their socio-economic background.

So, if people are quite busy earning their money, they would not have time to document a language right, livelihood is much more important. So looking at various other factors, different types of methods can be adapted among the four methods that I described. So, active participation of community members, what can be done? So, where we see that the members are not trained in field linguistics, but still they can be involved actively in the process of language documentation. So manuals quite automatically assign a passive role to the native speakers. What they say is that there are questionnaires, you ask the sentences and thus language expert will translate it into his or her language.

But that means, they are just sitting there passively, they are just reacting to what you are asking. But recent approach talks about their active participation, where you do not ask anything as such, but you can actually tell them about the process of language documentation, why it is important, how it can actually give lots of different types of outcomes which can be helpful for the community. And in those cases, the community members can actually suggest. They can say that "we need primers" or they can say that "we need picture books in our community". So those thing they can suggest, they can help you collecting data, they can tell you more about their language which they think is important.

So, it makes the whole elicitation process more like teaching learning process or more like both parties are quite actively involved. It is not like the researcher is active and the language expert is passive, both are actively or together they are working for documenting a language. And then, the cooperation is also more from the speech community. So this is a very recent approach and is widely adopted. So, it takes language documentation to be directed by this activity, rather than by a set of goals regarding accomplishments with respect to an objectified language.

So generally, what do we see in many projects that there are certain objectives, like you have to collect say 500 or 5000 words in the language, you have to collect say 10000 sentences. So, this is an objective which one has to follow, but what happens in this case we see that language is being made like a commodity where you have to go and collect this thing. But actually language documentation does not work in that manner. If one wants to really document the language, then one needs to look at the overall thing and it cannot be like target oriented, in the sense of words and sentences as such. But of course, sentences and words will be collected, but it is done in a more, not in that strict way, rather by involving the speech community members as well.

So what happens there, you cannot say that I have to collect like 10 words related to body parts. Or you can say that, but you cannot say that I have to collect "what do you call eye, what do you call nose, what do you call ear?". You will get all these terms, but in a different manner. You can ask your language experts to give you the terms for different body parts. You can ask your language experts to say that "Can you tell me some terms related to your body parts, some terms related to flowers and vegetables?" or "What are the most common vegetables that you eat generally?".

So, that can be done. Or one can actually be part of their lifestyle, one can be with them, one can see, observe what they are doing and then ask what is this happening, what are these things which you consume and totally become active in the process and make the community members active as well. So, they also try to say that "Okay we have a festival, you can join us, you can also document and know more about it". So whole process becomes a joint venture. So there is more cooperation as well. The advantage is that it represents a way for speaker and linguist to collaborate on creation of the documentation.

They might even suggest that "See, body parts are not getting lost very quickly, we use the terms for body parts, but what we are forgetting is our folklores". So you should first think about collecting the folklores. So maybe they can direct you. So those things can be done or they can say that "Kinship terms are being lost. So, I think those terms should be collected before you collect body parts".

So, it is not like you create the objectives, rather they can direct you towards that, help in collecting discourse practices and also restricted information. So, if both the process, if the language expert becomes an active member for the process of language documentation as the person is part of the community, she can help you collecting the discourse material like the discourse practices, how discourse is practiced, different pragmatics of different discourse. So, all these can be collected through that person. It is also a more inclusive way of documenting a language, where both the community members and researchers both are included. So it is not like you are just getting data from the language experts and not involving them in the whole process.

It is like they are also part of the process, they are also guiding you, they are also telling what needs to be documented before one thing or they are also talking about what are the outcomes that their community needs. So, there can be all these things where the community members can participate. So the whole process becomes more inclusive. It also helps in remote data collection. So think about remote data collection, which I already talked about.

When you cannot visit the field, you can contact the community members who are trained or who might not be trained, and you can train them little bit and they can actually collect data, they can actually do the whole language documentation thing. So it also helps in that way. So for collecting words generally, what is done is that traditional field work manuals recommend on compiling word list where you see there are 100 word list or 500 word list and bilingual method is used for collecting data for translation and all. But in these cases generally, what is done is that "Tell me the names of five fruits, tell me the names of edible animals", those types of questions are asked rather than the translation methods. So, these type of things can actually make the speech community choose.

So they might not say about particular fruits, but say about some more common fruits which they use. So rather than asking the native speaker to translate, you can actually ask them to teach you what can be the fruits which are edible or what are the animals which you can eat. So things like that. So it makes the whole process more useful. So, I would like to conclude by saying the whole process of language documentation is changing with time and lots of new dimensions are being added to it.

People are talking about how to document a language, not about what to document, but also how we should approach the field or how should we approach the speech community, because when we are talking about language documentation, our main aim is also to preserve the language. It is not only about recording and huge amount of data and keeping it, preserving it somewhere. It is also about how we can actually make the language speakers speak their own language, how we can actually help them maintaining their language. So, one way of doing it is that like giving them outputs, creating some materials for them like dictionaries, grammars and of course, various types of apps, which can help them. So if we have gaming apps, pedagogical apps, these can motivate them using their own language.

Secondly, when we make them active participant of the whole process, they also feel more involved and in many cases, it was seen that when language documentation process was started, then actually the people started speaking more about their language, which actually helped in reviving the language. So, they were trying to recall their language, they were trying to use their language more, because they were participating actively in the whole process of language documentation. So in both the ways, when we are providing them with outcomes and when we are making this an active process for both the community members and the researchers, then we are trying to help the language sustain. So, the field of documentary linguistics not only looks at the process of data collection and recording of linguistic and metalinguistic information, but also concerns

with the outputs and the methods involved in the process. So, it is not only about the data linguistic and metalinguistic data, but also about the process, how it is done and also about the outcomes, what can be the outcomes.

And when we are talking about outcomes, then we also channelize our data collection process in such a way so that, we can have outcomes. So, if we are thinking about creating a basic picture book for a speech community, we will focus on that area and we will start our process, keeping that target in our mind. So that, we can have that outcome at the end of the process. So that way also, we adapt methods. So of course, we will take lots of pictures and we will collect terms for that, so that at the end of the process, we can give that picture book to the community.

So depending on various socio-cultural aspects of the communities, different approaches have been suggested. So there are different approaches of language documentation or data collection, and those can be adopted, depending on the various aspects of the speech community. One can also adopt mixed approaches. So it is not like one-to-one; in this community, this process will work or in this community, this approach will work.

It can be mixed; it is not always very strict. So one can begin with say translation method, but slowly, one actually become friends with the community members and then, she or he starts actually involving the speakers in the whole process and then, it can become a very inclusive process. So it depends how one begins and then, how one proceeds towards it. So it is not like one-to-one match. I hope you enjoyed this lecture today and I also hope that you can also think about documenting maybe, certain aspects of certain languages, may be a language which you speak or may be a language which you think someone is speaking, but is not used much in the community. So, you can think about documenting maybe, one particular aspect like kinship terms, body parts or certain syntactic structures you think about.

So, you can also think about documenting some aspect of a language. Please go through these readings. Thank you!