

Tools and Technologies of Language Documentation
Prof. Bornini Lahiri and Prof. Dripta Piplai (Mondal)

Department of Humanities and Social Sciences

IIT Kharagpur

Week-05

Lecture-25

Lecture 25 : Script & Language

Welcome to the 25th lecture of the course, Tools and Technologies of Language Documentation. Today I will talk about script, what is the importance of script and other issues related to it. So today, the topics which I will cover are the advantages of having a script because we know there are lots of languages, which do not have scripts and then, there are languages which have scripts. And when the languages, they do not have scripts, then we talk about developing scripts for those languages. So, why do we talk about that? What is the advantage of having scripts in a language? I will also talk about various terms which are related to scripts like exographia, digraphia and bigraphia. When we talk about development of a script, then we see that there are various issues related to it.

There are issues which comes from within the community, there are issues which are related to linguistics and beyond that as well. So, we will talk about all those things and what are the requirements for successful development of scripts, what do we need for that. And when we talk about language endangerment, I would also like to mention that there are also cases where scripts gets endangered and it is also very common; it is not like very unique. In India itself, we can see lots of scripts which have got lost.

Scripts also gets endangered and sometimes, they are revived and sometimes, they are actually totally lost. So, lack of writing system becomes a reason for marginalization of a language at times. So often, we see that script is related with prestige. When a language has script, then people take it as a language, otherwise people may not consider it as a language itself. In common discourse also, you will find people saying that that language is not a language; it is just a dialect.

It is just a variety, just because the language does not have its script, but it is just a myth because all languages are languages whether they have script or not; that does not matter. But in common concept, it is always taken as, if the language does not have script, that means the language is not a standard language as such. And when we see script, we know that script actually gives political voice to the language. If the community members want to express their political ideologies, they can actually do it using scripts. It becomes easier if the language has its own script.

Socially marginalized or stigmatized, that we often see social marginalization of a language, which does not have its own script. So, little access to other domains. So, if a language does not have its scripts of course, there will be no books in that language written, but of course, there can be books written in that language using other scripts, but when we see borrowing of the scripts, then there are certain issues. So, we will talk about that. There are some disadvantages of not having scripts for a language and that is why, it has been argued in literature, that much of the success of a language documentation depends on casting these records in an orthography that appeals to the speech community.

That means, the orthography should be their own orthography, their writing way. So, it should be the speech community's way of writing. That means, script for the language because script adds prestige to the language. As a matter of fact, it is accepted that documentation has to be accessible to the speech community. That means, in the form of books, journals or other forms the development and implementation of a practical orthography for the speech community is necessary.

It is very important to have a script, if you are thinking about documenting a language and bringing outputs to the community, whether we are talking about grammar, we are talking about dictionaries or other pedagogical materials. If there is a standard script for the language, then it always becomes more convenient, It also reaches more members of the community, if the community has accepted a script as a standard version of it. So, it is important for language documentation outputs also, that a language should have its own script. Now there are other advantages related to having a script a greater chance of literacy. So, if there are scripts for a language then that means, people can actually read and write in their own language.

So, they have the provision of writing and reading it related to literacy. So, script is of course, related to literacy and also related to education and other opportunities. So, when there are scripts then of course, more books can be published in that language and then these books can reach more people because now the language has its own script. So, for example, if x language has its own script then books will be written in that script, but if it is using a borrowed script, then it will vary from place to place. For example, Mundari

uses Bangla script in West Bengal, Odia in Orissa, Devanagari in Jharkhand.

So now, the materials need to be either translated into all these scripts. Or when a book is published in Mundari written in Bangla script, it will not reach the Mundari people of Odisha or Jharkhand. Similarly, if a book published in Mundari written in Devanagari script, then it will not reach the Mundari people of West Bengal and Odisha. So, what we see is that there is restriction of use of the books, but if there is a standard script of Mundari, which actually the Mundari people are working on; they have developed a script for themselves. When this script becomes popular and the community everywhere, whether they are from Jharkhand or West Bengal or Orissa, that does not matter; everywhere they have accepted the script, then one book publishes and everyone can actually read it. There is no need of translating it into different scripts.

There is no need of actually restricting people, rather people can use the same book across the states, across the places. So, that actually unites the people and also adds to the scope of their usage of the language. Increased access. As I mentioned, it gives more access to the learning of additional languages. So, what happens is that when a book, when they have a Mundari script for example, and then there are books written for how to learn Bangla or how to learn English and then there are Mundari sentences written in Mundari script.

So, those type of bilingual or multilingual pedagogical materials can be made available for the community members. Opportunity for cultural expression and wider communication. If a material is written in Mundari using Mundari script, it might be about anything, any cultural aspect, any political agenda, whatever it is, if it is written in Mundari script, which is used across Mundari community irrespective of their geographical area, then this same material can reach everybody. The same message, whole community can know about it, whole community can know about the message that is being spread or printed. Increased cultural and linguistic self-confidence.

When a language has its own script, we know that a community member feels proud about it, they can associate the script with the language. And then they know that their language is not inferior to languages which already have a script. So, when they are borrowing scripts for example, from Bangla, from Devanagari, from other languages then somewhere, there is that feeling of inferiority, that the language does not have its own script and they need to borrow it from some other language. So, another language has its own script. So, there is this gap.

So, when the language itself has its own script then there is no gap. There is that self-confidence for the language and also, you can say there is prestige added to it.

Appreciation by others of the unique richness of the language. So, more expression can be there if the script is there. So, expression in the sense more books can be published, people can write about their language in their using their own script and that way, they get their identity, their identity is well established through their script.

The option to use the language in electronic media. So, if script is developed for a language, then of course, there can be unique codes being created and then it can also be used in electronic media. So, what you see is that when a certain language is written in its own script, then other linguistic communities can look at it, that it has its own identity. The identity is established in a better manner and when identity is established, that automatically adds to the prestige of the language. So, the existence of written code is seen as an essential prerequisite for many activities in favor of maintenance and revitalization of a language like creating dictionary, curriculum development, teaching learning material development.

So, for all these, scripts becomes mandatory actually. Script adds to the prestige of the language, which I am repeating, a language without script is often considered to be just a variety or a dialect. People do not consider it as a language. Linguistically, you might say that no, it is not. So, if you are student or a scholar of linguistic, you will know that every language is a language irrespective of the fact whether the language has a script or not.

But in common people's way or even people talk in their day-to-day discourse, they consider them as varieties and dialects, not as language, when these languages do not have scripts. You will often see people saying that "oh that language is not a language, it is just a 'boli'." A boli, we often hear, specially in northern India, where people say it is just a boli, it is not a language. And why do they say it? Because the language does not have its own script. So, maybe everything is there, there is oral literature in the language of course, there is grammar.

That is why people can communicate, there is a huge set of lexicon and everything, but still the language is not considered a language just because it does not have its own script. So, writing without a script can also be there. As I gave you the example of Mundari, when they do not have their script, which is established, then often, the community members, they borrow scripts from the other languages and those scripts are used. So, it is called exographia. It is when a script is borrowed from other language and that script is used to write the language.

We can see Bangla is used for writing Dhimal. Sometimes we can even see more than one script is borrowed. So, Mahasui is written in Devanagari script, Kurmali is written both Devanagari and Bangla script. So, people who stay in Jharkhand, they write in

Devanagari script while people who stay in West Bengal, they write in Bangla script. We have even seen books which are like written in Kurmali and half of it is written in Devanagari script and half of it is written in Bangla script.

So now, you see one who knows Devanagari script, can read half of the book, but cannot read the other half if he or she does not know Bangla script. So, that becomes a problem. Mundari, as I mentioned, is written in three scripts Devanagari, Odia and Bangla. So, when there are borrowing of the scripts, then at times, there can be divisions or restrictions for the others, others who does not know that script may not be able to reach to the content. So, exographia is very widespread in endangered and minor languages, because these minor or smaller languages they do not have their own scripts.

They often borrow scripts from other languages and in that script, the language is expressed. In those cases, we also see a problem that is, sometimes for example, in Bangla there are of course, certain letters, but in Mundari when we are talking about Mundari, there are certain sounds for which we do not find representations in Bangla script. So, either in those cases, when they are using Bangla script they can create some symbol or they use some diacritics or sometimes in the longer process, when exography is practiced then what do we see is that particular sounds are not at all represented in the writing system and that can actually influence the speech of the younger generation and that particular sound can be lost. So, there are problems. So, when there are borrowing of scripts, certain times we see lots of sounds are not represented through the other script.

Because when we are talking about Mundari, it is Austroasiatic language, when we are talking about Dhimal, it is a Tibeto-Burman language, but these languages are using a script of Indo-Aryan language, which is Bangla. So, it is very difficult to find representations of all the sounds in Bangla script and that is why, it becomes very difficult. So, either some non-standard ways are adapted or there can be no representation of certain sounds. So, it is often the case that, an official language occupies formal writing context and a regional lingua franca is used for writing in semi-formal and informal context. So, what happens that for these languages and many many other languages like this some script is borrowed and that is mainly used.

So, when we see also this type of borrowing then the community members are also somehow forced to learn the script of other language. So, if Dhimal community is writing Dhimal in Bangla script, then somehow, the younger generation or the community members they are forced to learn Bangla script. So, maybe, they can speak Bangla, but somehow, there is also indirect force of learning Bangla how to write and read Bangla, because Bangla script is used in writing them all. So, that also becomes an indirect pressure on the community members. Then there are practices like bigraphia and

multigraphia.

So, these terms are actually taken from bilingualism and multilingualism. So, what we see is that more than one script in representing one language. And in fact, in whatever language you use, you can also think about it. Nowadays, we actually do bigraphy a lot. So, I speak Bangla, I write Bangla in Bangla and also I write in Roman script.

So, due to social media, we often use Roman script to write our own native language. So, that way that is bigraphy. We are using more than one script to write our own language. Sometimes like code-mixing, we also mix our codes. So, if we are writing something on social media, maybe we will write one phrase in Bangla script and then, I will mix it and write something in Roman script, but I am writing Bangla and then again, I will write in Bangla script.

So, there can be this type of mixing of codes in writing as well, which we do while speaking, like code-mixing can also happen in writing. So, Manipuri is written in Bangla and is also written in Manipuri script though nowadays, the community is shifting towards Manipuri script, which has been actually standardized. So at times, we also see that scripts are developed for the languages and those become standardized and the community slowly actually shifts towards the standardized script and they learn it and start writing in only that script. So, we see for Manipuri, it is like a transition period where we see the older people they still know Bangla and they prefer writing in Bangla. They are writing Manipuri, but in Bangla script, but slowly people are actually shifting towards the Manipuri script and Manipuri script is used in the formal context for writing Manipuri.

Digraphia denotes multigraphic writing traditions in context where one tradition is dominant than the other. So, you see it is something like Diglossia, where you see one writing system is taken as more as a better way of writing than the other one. So, there are two ways of writing or two scripts in the same language for the same language, but one script is more dominant than the other. So, now when we are talking about development of script. As you know, development of script is important for the smaller and minor languages, because that can help in revitalization of a language; that can help in maintenance of a language.

So, when we talk about development of a script, there are certain things when which we keep in mind. The successful creation of an orthography involves the consideration of historical, religious, cultural, identity related and practical factors in addition to linguistic ones. So, there are actually lots of things which are involved when we talk about development of script, because it is not only about linguistic things, it is not only about a

linguistic relation, it is also about political, historic, all these factors because the community members, they should agree to a particular style of writing and then depending on that, script is developed. So, like for example, Kurmali as I said is written in Devanagari, Bangla and Odia scripts and if the community now decides to adapt one script or to develop one scripts, then they have to actually discuss. All these representatives have needs to be discussed and they cannot choose one.

So, if one community says that let us adapt Devanagari script with some minor changes into it, other two users like those who are using Bangla or Odia script to write Kurmali might not accept that because for them, it will be a burden to learn Devanagari script; they will think that those who are using Devanagari scripts are actually in a privileged position because they already know Devanagari script. So, they will not need to learn more, but for Bangla and Odia writers, they will need to learn another script that is Devanagari script. So, they might not agree to that and same is the case if Bangla is chosen or Odia is chosen. So, it becomes a real problem and that is why, it needs to be thought about and discuss how to actually develop script; there are lots of issues involved into it. So, in Saurashtra language again, Saurashtra is written in Tamil, Telugu and Devanagari and all these scripts are very different.

Now, if we have to choose one script which one it can be, but Saurashtra language has actually designed its own script where we find different letters are there. So at times, what we see that among all these scripts, which are used to represent the language, one of the script is chosen like for example, for Kurmali maybe Devanagari is chosen, if it is accepted by everyone in the community or sometimes, like Saurashtra a new script is being developed for writing the language. So, scripts are not identical with orthographics and spellings. The standardized versions of scripts for specific languages or varieties are different. So, we cannot always say that scripts are in one to one relation with the sounds.

So, grapheme and phoneme may not match, but it is ideal to keep them match while developing a script. Like in English, we see there are 24 letters, but there are around 44 sounds. So, there is no one-to-one match. So similarly, when writing Hindi versus Kurmali speaking, there can be difference. So, there can be a common word in both Hindi and Kurmali, but the spellings can be different in both the languages.

So, when we are talking about script, we also need to think about the rules, how to write spellings. There are certain rules which are followed while developing a script, because that needs to be followed across the writings. So in every writing, we see that there are certain rules in every script related to spelling, related to various combinations of consonant sounds. So, what are the rules those are also finalized. It is often assumed that the writing systems of modern orthographics will be alphabetic, but that is not true.

For many languages, we do see that the pattern of writing is not alphabetic. There can be different types of writing system, pictorial representation can be there. So, those types of scripts are also developed sometimes for some of the indigenous languages. So, what are the requirements for developing a successful script? One is to see that there is one-to-one match between the phoneme and the grapheme. That means, all the sounds which are there in the language should be represented through a symbol in the language.

It should not be like there are certain sounds which cannot be represented through a grapheme in the language. In the ideal situation, we see that there is one-to-one match. So, there should be certain symbols for all the sounds which are there in the language. So, that is a linguistic aspect to it.

Selection of writing system and set of graphemes. So, how the language will be written? It is not only about the letters or alphabets or the symbols or the graphemes, it is also about how will it be written left to right, right to left. what will be the spelling system or the rules for the spellings, what will be the rules for the geminations. So, there are all these things which are considered while developing a script. Establishment of rules specifying the relationship between sounds and grapheme. So, how are they actually represented? Like in many languages for example, in Devanagari there are symbols which are used to mark the vowel sounds.

So, with consonant sounds. So, those type of symbols, what will be used when a consonant and vowel is mixed together. Production of dictionary, listing spellings and materials for learning and later independent reading is required. So when these rules are created, word boundaries, punctuation, every type of rule related to writing is developed and then, that is being recorded, written somewhere. Mostly what we see we see collection of dictionaries where the words are written in that script and all these rules related to these writing system is also recorded in that dictionary. And this dictionary is actually used to teach the script and the rules related to it to the community members.

So initially, it takes time. Slowly community members learn the process of writing, how to write and learns the rules related to the writing and then, when it spreads in the community, slowly people start practicing it. And then we see more books and journals being written and printed in the script, then we can see that the script is being well established within the community. So, some languages are tonal languages. How to actually express the tones in the language in the writing system, that is also important. So now, we cannot say that, no, tones cannot be represented in the writing system.

So, if we see Vietnamese, then we can see that there are 6 distinctive tones 5 of which

are normally represented in writing. So, it is not like they are in not including the tones when they are writing something they are actually also representing the tones. So, writing system should be more inclusive. It is about writing all the things which we are actually speaking and the way we are speaking it. So we also mark the intonations, we have various type of punctuation marks to mark them.

So, those are also part of the script development thing. This is a process chart which is a guide given by UNESCO for developing writing system or script for the unwritten languages. They have mentioned the factors which one should keep in mind and then, it is followed by process and the outcome. Factors which influence development of a script are social relations within the group and outside the group. How people are actually reacting to this script development, what are their ideas, what are their decision, that is very important because when a script is finalized, then the community needs to accept it.

So, if the community takes part in developing the script. then it is always good. So, it cannot be done from outside. It cannot be like some language sit together, they develop a script and then, they give it to the community. The community might not accept it, might not take it.

And then there are political context. So, we see that there can be lot of government policies or other political groups which have their own interest and they can force to the community to have certain script. So, we can see that because of certain political reasons, some groups can be forced with certain type of scripts. So, there can be things like that. While developing a script, cultural heritage is also taken into account. So, that the oral literature, local knowledge, all of these things can be represented through the script.

Sometimes, even the alphabets or the symbols used to symbolize these sounds are based on these cultural histories. The signs, the symbols are actually outcome of those oral literature and part of those things. And a very important point where the linguist can play a vital role, that is looking at the language structure, looking at the sound system and the grammar of the language and based on that, a script can be developed. Looking at the sound system, the phonemic chart, the grammar, the punctuation things, all are incorporated while developing a script. So there, linguist can be active and other things are more dependent to the community members.

So, these are the factors which is followed by the process where we see that there are dialogues and negotiation within the group and also outside the group. There are negotiations for the political parties and all. And when a particular community is recognized, then its culture, its oral history, oral literature also needs to be compiled in the script, needs to be reflected through the script and that has to become from the

community. And then of course, the linguist's job is to analyze and look at the language and then help with the creation of the script. So, all of these process together can help in designing a script and then when a script is designed then it is given or it is accepted by the community and slowly it takes time to spread over the community and that way, we see scripts are developed for the unwritten languages.

So, we can see that as I said, some of the scripts at times also gets endangered and then they fade out. So, you can see Kangri script was earlier used to write Mahasui for official documents, but nowadays, we do not see use of Kangri script for writing Mahasui language. Similarly, in 18th century, the Manipuri script was replaced with Bangla script, but again in 2006, the Manipuri script was introduced and we see that nowadays, Manipuri script is used for writing Manipuri in all the formal context. So, we see that sometimes scripts fade out, sometimes they are revitalized and again, they appear and sometimes, they are also totally lost and new scripts are actually developed for the languages. So, these are the references and I hope you enjoyed today's class. Thank you!