

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-24

Lecture 24 : Language documentation and the problem of standardization

Hello everyone! Welcome to the SWAYAM NPTEL course on Tools and Technologies of Language Documentation. And today, we will listen to Lecture 24 of Module 5. Module 5 is Other Dimensions of Language Documentation and the title of today's lecture is Language Documentation and the problem of Standardization. So, I hope the term 'standardization' does ring a bell. You are familiar with the term, but today, we will be going to a more critical level and try to explore the idea of standardization in more detail. So, we will start the class by defining 'standard language', then we will discuss different models of standardization, how it change with time and then I will introduce the idea of Koineization, which is a related concept.

And then, we will learn about how purism and standardization are related concepts. And finally, we will discuss some challenges in identifying the contemporary standardization processes. So I'll talk about different timeline, when exploring the idea of standardization and also try to give you an overview of the changes that took place in standardization studies. And in between, I'll be giving some examples of field and talk about some potential problems that we face while we collect data.

So if you ask this question, what is a standard language? Many of you are already familiar with this idea. If you are coming from a background of linguistics or if you are from allied discipline, you have attended some course in sociolinguistics, I am sure you are familiar with this concept. So here is a very basic definition. A language L conceived as an assemblage of dialects. So, when we consider a language, we always talk about some kind of an assemblage of dialects D1 to Dn.

But, when we are talking about several dialects inside one umbrella of language, there is

one clearly defined standard language. So, why this is different and why it is clearly defined and why we are terming it as standard? One criteria is that it is a dialect with a written norm. So, the standard will have some kind of written norm and I will begin with this definition of standard language. Simultaneously, I will point out to a problem that we often face at field. This is related to the belief that non-standard varieties of any language cannot be documented.

So, this is a problem that you also find conceptually. You think that apart from the standards, we should not talk about the other varieties, but while on the field, if you think that non-standards are not worth documenting, then there is a problem. And when non-standards are not written, often the speakers tell us that "please do not document these forms because these are not written". So, we have to be alert and we have to remember that these are some traditional norms that are driving us. And, if you are a community member, I am sure you are alert and you are aware of the factors that there are non standards which are not written, but we should not skip them.

We should document them as well. Stages of standardization as discussed by Haugen, which is a classical study. It talks about different stages: selection, codification, elaboration of function and acceptance. And, you can also, I will explore this little bit and if you remember, I also briefly talked about them previously. So, if we select a variety, then we codify it through say, dictionaries, grammars, etcetera.

Then, the function is elaborated in the society, those dictionaries, grammars, textbooks, as well, these are spread and then, the society or the speech community they accept that this is the standard form These four stages which have been given by Haugen is considered as a classical stage, a classical model for discussing language standardization, but I will also show you how these are challenged and how there are new models coming up. I will also mention Milroy at this point, who is saying that standardization and social access are related. If you think about contemporary times and talk about what kind of access is there for the speech community, you may need to revisit the ideas by Haugen. There are different ideas about standard which you should also know. You are already aware of language attitudes that we discussed in detail.

So, you can figure out that there are different kind of attitudes towards non-standard varieties. This is also related to the problem that I have just pointed out. If you are on the field and you figure out that there is some kind of negative attitude towards the non-standard varieties, you need to note it. Because this may affect the language use, there can be some kind of control in the natural language use. And Sallabank (2013) mentioned that people can have different feelings about their own language and the language of the others.

So, you remember the lessons of language attitude again. When we talked about the feeling of your own community and the adjacent communities, we need to take into account both the different types that occur in a speech community. Now I am introducing another very important idea, which has been given by Lippi-Green in 1997. This is related to a bias towards homogenized language. So, you can think about different layers of the bias.

It depends from one case to another case, but there is some kind of bias and it is imposed and maintained by dominant institution. You can figure out the role of dominant institution for your case and also, need to think that it acts as a model of written language and usually, is the spoken language of upper middle class. So, whatever homogenized varieties there inside your head, which is like an ideal standard, which is maintained by dominant institutions, we always think that this is the model for any kind of writing. We cannot write in non-standards. Again, there is this strong belief in different societies and we also find that the standard, the so called standard is spoken by the upper middle class.

So, if we revisit this model, we will figure out that from 2010, there is a departure from the traditional models. If we need to look at the departure, we will think that there is this idea of re-standardization and the application of the models of standardization to minority and minoritized language. So, this is very important for us because we are talking about different indigenous minority and minoritized languages. So, what kind of re-standardization is there in the scene? We need to analyze it in different layers. And, if you look at the postmodern world, you find that there are social changes and role of standards is weakening.

So, if you look at the role of standard in say 19th century, it is definitely very different in contemporary times. And, we see that the boundaries between standards and non standards are gradually becoming blurred. So, all these are challenges or rather, posing challenge against Haugen's model and that is why, it is equally important for us to understand Haugen's model for your situation, your kind of linguistic situation, where you are working and also, to look at it from the contemporary viewpoint. Neustupny talks about four historical types of language planning. I am dragging the term 'language planning' here purposefully, which is associated with standardization as a process, because you know language planning is basically done as a part of standardization.

I hope you understand that. So, there are four historical types of language planning, which you can classify as premodern, early modern, modern and postmodern. So, I am giving examples of premodern and early modern here. So, if you think about the academies that were there in the four different languages like the French Academy of

17th century, you can explore more on the nature of the academy, if you want to know more about premodern kind of language planning. And early modern is a classical example of Europe, where for industrialization, we find that there is some kind of secular modern ideology of nationalism.

If you are familiar with works of Peter Burke for example, they are talking about this ideology in detail. And in 19th century, we observe different national movements of modern nations, like Norwegian, Finnish and like you can also look at these; the importance or significance of these movements are also reaching India at that point of time, because we were under a colonial control. So, the European notion of standardization were very much applicable for India. So, at this point, if I look at the language plan, a quote from Haugen again, regarding language planning in Norway, where he primarily worked. "The activity of preparing a normative orthography, grammar and dictionary for the guidance of writers and speakers in a non-homogeneous speech community".

So, this is a quote from Haugen. I hope you can understand how this quote is related to the four divisions or four stages of standardization that Haugen mentions. So, it is related to some kind of norm and orthography, grammar, dictionary, all these different aspects of codification. So, I am also mentioning Garvin (1959) here who talked about two properties of standard languages which is flexible stability and intellectualization. So, you can figure out that this is directly related to the definition given by Haugen or stages given by Haugen, but we need to remember these because these are basic skeleton of the standardization process.

One very important work was Le Page and Tabouret-Keller in 1985, where there is a difference between unconscious and more conscious standardization were made. You can figure out what is occurring for your target speech community and whether it is unconscious standardization, whether they are part of unconscious standardization process or it is a conscious choice for them to adopt a local standard. Individuals mold their verbal behavior as they want to be identified with a particular group and at times, they do not want to be identified with another group. So, it is changing from time to time. So, this is very important and that is why we need to figure out, we need to learn from Le Page and Tabouret-Keller, which is slightly different from Haugen, but it is another milestone in language standardization research.

At this point, I have a note and you should keep in mind while collecting data from the field that in Indian context, we already talked about projection of multiple identities. So, we find that one speaker is switching to different standards. if they have access to it. So, this is exactly what Le Page and Tabouret-Keller has figured out. At times it is

unconscious, but at times, it is a very conscious choice.

As a researcher, it can posit certain challenges. So, you have to be conscious about what kind of process is happening. Another very important notion in Indian context is Koineization, which is very common for certain societies in specifically multilingual societies. If you look at Ferguson's quote or Ferguson's definition, you can find that it is a reduction of dialect differences by leveling and simplification. If you find there are several varieties in one area, you will find that there is some kind of leveling which is taking place, some kind of simplification which is taking place.

So, this is referred to as Koineization, which is very different from a Pidginization process. I hope you are familiar with the process of pidginization and creolization. So, in case of Koineization, it gradually changes. It is a gradual kind of simplification which takes place. We also find that in Koineization, some stigmatized forms are common while the simplification process is taking place and also something called variety shifting is observable.

You find that one person has access to multiple varieties, they are constantly shifting again to relate it with the projection of multiple identities based on situations. So, I will give one example from our field here. So, when we were collecting data from the multilingual patches of West Bengal-Jharkhand-Odisha border. It is a border of three different Indian states and it is a huge multilingual melting pot zone. We were finding that people were constantly switching from different varieties and it is not the case that these are the standard varieties all the time.

So, you have to be alert that whether they are switching to standards or they are switching to non-standard and what kind of simplification process is acting for your case. So, you have to be alert about the nature of your data. You cannot label something as standard or non-standard very quickly. You have to observe this entire process.

Another related term is purism. So, I want to discuss it in terms of the standardization process at this point. And if you look at what Thomas 1991 says, purism is associated with language standardization, language contact and language variation. So, if you look at different stages of language standardization as Thomas argues, you find several challenges related to purism. So, at the stage of codification, you find different set of challenges, at the stage of elaboration of function, if you follow Haugen's classification again, you find a different set of challenges. You need to figure out what kind of challenge you are facing or what kind of challenges are probable challenges for your field.

For example, you can think about growth of a single autonomous standard. At times, we find that there are multiple standards. In Indian context, it is also there and at times, we find that there is a single autonomous standard. So, what is the case for your field? So, there were some recent developments in the field, which I should mention. I already talked about the issue of multilinguality, which is questioning the standard model of standardization, but we need to figure out that many such models are created based on western countries.

So, whether these models are valid for non-western multilingual countries, these are also being asked in more recent times. Another challenge is to understand the standardization history of minority and minoritized languages. So, what kind of standardization process is taking place? So, I will also say a few more points for this. And we have seen that Haugen's model was not useful everywhere.

So, we find double-edged standardizations. And in Indian context, we cannot generalize the entire picture with Haugen's model only. Lane, for example, mentioned that for minoritized languages, standardization efforts are very different from national language standardization, that we have seen for pre-modern cases in Europe. So, there are some layers which we need to understand and also, it is very important because contemporary standardization is documented, accessible and visible unlike the 19th century cases. So, we do not have enough evidence for the standardization.

There is always a lacuna. but for contemporary standardization, you understand that the entire process is somehow documented. So you can clearly figure out that the stages can overlap and it can also differ from the western models. So we also find that for minority languages there are negotiations, debates, contestation and also appropriation. So, here I will also point out to another problem that we often find at field. So, how to assess the standardization venture for your target language? This is not very easy.

If the standardization process is relatively new, you have certain documentation, but there can be multiple layers to it. Another problem is a major shift in the agents of standardization, which is of course a postmodern challenge. Earlier, the agents were bodies, the academies, as I have mentioned for French. The academies were solely responsible for creating the codified materials as a part of standardization, but now, there are local agents and also bottom up policy makers. So, if the community members are becoming a part of the standardization process, you may have to think it differently.

You have to understand what kind of community members are being involved in the standardization process, which is relatively new. and how social actors support, alter, resist and reject the processes of standardization. May be the community members are

accepting the new initiatives of standardization, but it is also true that, in some cases the community members do not accept, they reject the standardization initiatives. At this point, you should note that the role of dominant speech communities can be understood. If a minority or minoritized speech community is involved in a process of standardizing their own language, there is a possibility that dominant speech communities are either resisting or rejecting.

It may result in a very different kind of language use. So, you need to figure out what is happening for your field. I will also talk about the idea of standard by mere usage, which is a key concept of communities of practice. Now, you understand that there are discourse communities. If people have never met, even then, they create a kind of solidarity and they act as a group of agents for standardization process.

So, what is happening in those cases? So at this point, if you consider Eckert and McConnell-Ginet, they are saying that "People who come together and mutual engagement is an endeavor". This is applicable for the modern or the rather postmodern initiatives of standardization. So, people from speakers from different points, they are coming together. and their target is to standardize a language. So, you can relate it with the concept of communities of practice and certain points for your field.

So, this is very important in contemporary Indian context, because we find digital solidarities. So, people creating social media groups And this is also creating a huge corpus and getting some kind of consciousness, spreading some kind of consciousness among the community members. There are Sunday schools which I have mentioned earlier, Sunday schools for teaching learning of the minority or minoritized languages, which is again creating some kind of solidarity. And the Sunday school networks are not located in one place.

So, these are in different places. So, which one to use? How to standardize the use of one particular variety. So at this point, I will also tell you some potential problems that we often face. And you may also ask some questions. You may consider the questions as silly, but these are not silly and very serious rather. So, if you ask, if you are supposed to document the standard language only, the answer is No.

We need to document all the varieties available. So, if you are visiting a field and if you figure out that there is some kind of written documents in one variety, it does not mean that you have to figure out that particular written variety and document it. You have to document all the different varieties available, all the varieties that are available, used by people in different geographical areas, used by different occupational groups, different age, gender and social class. So, the target of documentation is to get the entire picture,

not document on the basis of standard variety only. So, there are certain problems that we face on the field. At times, when our informants try to use the standard versions of their language.

So, if you find that one of your informant has two or three different varieties of his or her language and if you are asking them to translate something in their language, they will tend to translate in the so called standard variety which is now being written. So, you have to be aware of this fact that there can be different varieties. If you are doing a pilot study, you will be getting some idea of the other varieties which are existing in the same belt. So, your task is to figure out whatever different varieties are accessible to the speakers and document all possible varieties that you get. So, at times these speakers think that their home varieties have a lower status.

So, they should not use the variety for documentation purpose, but it is your call to explain or assert that it is equally important. I am listing some more problems which are often confronted by the field workers, which are related to the problem of standardization. So, if the community has several varieties, that I have already mentioned and users of less prestigious varieties avoid them, you have to tell them to use it. If they are consciously avoiding it, may be, you can ask another speaker who is comfortable with the speech data. There is another problem that you also face that, if a new standardization process is being initiated among the community members, if there are existence of some newly published codified text like grammars, dictionaries, etcetera, the speakers of other varieties may be asked to use the target variety for standardization.

So, may be some agents will tell these users of other so called non-standard varieties to use the written variety. So, that can be dangerous and that can also be important for language endangerment. So, we all think about certain varieties which can also die. For these cases, the non-standards are never documented. If you emphasize on the standards only for one language, then you will never be able to document the non-standards.

And also, if the community members are deliberately shifting to the standard, then the non-standard varieties will die. So, I will come to the conclusion of today's lecture and also try to give a recap of whatever we have discussed in today's class. So initially, we talked about the definition of standardization and models of standardization, keeping in mind Haugen's idea. So, I primarily talked about Haugen's idea. We also talked about different stages, then we discussed the idea of Koien and how purism, the idea of purism is associated with standardization.

We mentioned that validity of western models and for non-western multilingual countries is not very useful because there can be gaps and if you remember, the

difference between conscious and unconscious choice. And finally, we talked about some potential problems that we have identified while conducting field work. All these problems are important for Indian context. So, these are the key references and I hope you enjoyed this lecture. Thank you!