

Tools and Technologies of Language Documentation

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

Department of Humanities and Social Sciences

IIT Kharagpur

Week-02

Lecture-06

Lecture 06 : Assessing Language Vitality

Hello everyone. Welcome to Lecture 6 of Module 2, Various Attributes of Language Documentation and you are attending the SWAYAM NPTEL course, Tools and Technologies of Language Documentation. The title of today's lecture is Assessing Language Vitality. So if you remember I already told you what Language Vitality is, briefly, in the previous lecture. So today our target will be to talk about the assessment of Ethnolinguistic Vitality of a speech community. So I will be discussing the 9 factors which have been suggested by UNESCO, these factors are for the assessment of Ethnolinguistic Vitality.

And along with discussing the factors, I will also talk about our experience from the field regarding the assessment of the vitality. And finally, I will briefly mention the community's role in vitality assessment. So far we have already learnt the ideas that now you can see on the screen. Language shift and endangerment, we understood the basic ideas and we know that these can be triggered by some external causes.

For example, education related factors, economic, religious and cultural factors. And we also know that this trigger can be internal. External factors can trigger endangerment and internal factors are important too. For example, if the community members have negative attitudes towards one language speakers, there can be an internal trigger for language endangerment. So when we think about loss of linguistic as well as cultural knowledge, which I am sure you have been thinking about while attending the lectures, we have to think about a few factors which are responsible for language vitality.

And we know that awareness of language loss and diversity is needed. I briefly mentioned the idea of reversal of language shift which is important for revitalization. So today our target will be to discuss these 9 factors. I quickly read them out and then explain one by one with examples. The first one is intergenerational language

transmission, second one is the absolute number of speakers, then proportion of speakers with the total population, trends in existing language domains, response to new domains and media, materials for language education and literacy, institutional language attitude and policy, community members' attitudes towards their language and amount and quality of documentation.

So, all these factors are needed. If you are a community member and attending our course, I will ask you to pay attention to this. If you are not a community member, and working on a language which is not your own language, I will ask you to pay attention to these factors. It may have different significance for you, but it is important for both. So as I told you I will be talking about our experience from field while explaining all these factors, so I will also request you to note down your own examples while going through these factors.

So the first one is intergenerational language transmission and you briefly know about it, when I discussed about this GIDS scale by Fishman for example. I am sure you remember that we did talk about this intergenerational language transmission right. So here I am talking about these layers of marking an endangered language or the level of endangerment which you are already familiar with. You already know them through Dr. Lahiri's lectures.

So I am talking about them again so that you understand the layers in terms of the 9 factors. So here we are talking about these factors and also suggesting a score for them. So in this slide, I will be showing the score. I will not show the score for other slides, but I will ask you to do this calculation for yourself which will benefit you. So if a language is 'Safe' for example, spoken by all generations then you can put a score of 5.

If it is 'Unsafe'- Most, but not all children speak the language. You can find that only a few children of the community are using and others are not, but most children speak them, then you can put the score of 4. If it is 'Definitely Endangered', where language is not being learnt as mother tongue by children; maybe they learn a few words of the language, but they do not learn it as a mother tongue, you can put a score of 3. When we say it is 'Severely Endangered', where it is spoken only by the grandparents and older generations which is very common in Indian context. I will give you some examples.

You have to mark the score of 2 here. For 'Critically Endangered', where younger speakers are in the great grandparental generations; it means it is basically moving towards endangerment, then you have to put a score of 1. And if it is 'Extinct', there is no one who either speak the language or remembers the language, then you have to put a score of 0. So similarly for all factors please put the scores. So when I am talking about

intergenerational language transmission, I will give an example of Kurukh, a North Dravidian language which is mostly spoken in Jharkhand, but we happen to work on some migrant population who migrated to the tea garden areas of northern Bengal.

It is also spoken in Assam, but we worked in northern Bengal in this Kalchini Out Division area of a North Bengal tea garden. We found that Kurukh is not spoken and learnt by younger generation. So we can mark them or we can say that they are definitely endangered in the target area. At this point I should mention that in case of other places it can have a different status. May be in Jharkhand it has a different status which is basically the homeland of Kurukh speakers, but in our target area we found that it is definitely endangered.

If you are trying to chalk out the factors for intergenerational language transmission you have to look at these factors. We need to check the number of speakers for the languages on which we are working. So if it is a small speech community, they have a chance of merging with the neighbouring groups, specifically if the neighbouring groups are dominant speech communities. So I will give an example of Toto which is spoken in the India-Bhutan border. So the number of speakers are available through census data.

I think you remember when I talked about the importance of demographic data that you can collect from different sources. So since this is a small speech community, it is possible to track the actual number of speakers through census data and also other data sources. So when we worked on this language we could get some idea about the actual number of speakers, but there is multilinguality in the speech community. The community members use Nepali, Bangla and Rajbanshi which are more dominant languages. So there is a strong chance of shift, but I should also mention through this study, we found that some of the community members are initiating the reversal of language shift.

I hope you remember what does it mean the 'reversal'. So community members are trying to use their language, the younger members are trying to use their language. So it is also affecting the number of speakers in the speech community. So at this point I have a question for you. If you are a community member and attending this course, I am sure there are some people who belong to different speech communities which are receiving some kind of threat.

So it is very important for you to understand the population. So can you comment on the initiatives of the reversal of language shift among your community members? Do you really find that the younger members of your community are basically trying to revive their language by using them? So this is an important question for you. But I should also

introduce the third point suggested by UNESCO, by linking with it with the second point because when you talk about the number of speakers there is a need to link it with the proportion with the total population. So if you have a small population of a language and if you also consider the total population of that area, it will be easier for you to figure out whether there is a threat for this language, whether there is some kind of shift which has been initiated already. So I will go back to these layers of threat; the ways in which we usually talk about endangered languages.

So if it is safe, we see that all the members speak the language. It is unsafe we see nearly all speak the language and if it is definitely endangered, we see a majority do not speak the language. They are shifting to other languages, but you have to take into account. Severely endangered- so a minority speak the language. If it is critically endangered, very few speak the language and if it is extinct, none speak the language.

So it is very important for you to check the total population and the proportion of speakers. So we worked on a speech community who speak Kurmali, in Purulia district of West Bengal. We found that majority speak the language including children. So it was our finding. So in this manner, you need to figure out what is happening in your case.

So now, I will talk about the trends in existing language domains. What do we mean by existing language domains? I am sure you remember what do we mean by language domain. I already talked about different domains in previous classes. So if we find that in different discourses and for different purposes the language is used, we can say that it is universal use. There is multilingual parity if we see that one or more dominant languages are there rather than the language of the group.

So, more options are available for shift and more options are there in terms of the multilingual nature of the speech community and we often see that in official domains more than one language will be used, like government or education. Then there is something called dwindling domain, where non-dominant languages lost ground. So non-dominant languages will not be used. In limited or formal domain we have languages which are used in only highly formal domains, for example, rituals or festivals. Otherwise you do not find the language.

It is there only for the highly formal domains. And highly limited domains: very restricted during special occasion, not all festivals, not all occasions, but only for special occasions. So I will give an example of Rajbanshi which is a language spoken in northern part of Bengal. The Rajbanshi community have Nepali and Bangla medium government schools available for the children which is a case of multilingual parity. So it is not only one dominant language which is pressurizing the community, but more than one

languages which are used for official purpose.

So similarly you can try to understand the status of your language. New domains and media, UNESCO talked about it, but the situation is changing so fast that we need to understand the role of new domains and media which is affecting the language use. So we often see that communities expand their language in new domains. So it is not the governmental language policies which are controlling language use, but it is also the members who have access to new domains who are basically trying to expand the language use. So we see that some languages are also being used in schools, maybe one Friday class or summer school in government schools; and private schools, you will also find there are some initiatives for local language classes.

But in new media, specifically broadcast media and internet, we find the existence of many languages which are otherwise threatened. We find that the presence of several endangered languages are there through internet. We can find Instagram reels handles, Facebook groups, YouTube channels where the young community members are trying to revitalize their language. And if you are a community member again, I think you can identify with what I am trying to say. So our experience of working with the Mundari community in West Bengal-Odisha border, we found that there are Mundari songs which are not only initiated from West Bengal or Odisha, but from Jharkhand.

So there are very popular channels of Mundari songs through YouTube which are basically helping the young members to learn their language. Some of them are learning their language in order to understand the songs. Next we will come to materials for language education and literacy and when I am using the word 'literacy' of course, it is related to the orthography. So, there are possibilities that one speech community is traditionally using some kind of established orthography. They may have a literary tradition which is either oral or written.

A community may have dictionary, some kind of texts, grammar, literature and use in everyday media. So we find that there are speech communities who have written materials which exist for literacy, may be printed materials. Here I am mostly talking about printed materials, which are widely used in schools and writing is also used for administrative works. At times we find that written materials exist only for some members of community, which is very important for you to understand if you are a community member. So we find that if you just tick the check box that there is literacy in the language, but do not count the number of people who are basically using it, then there is a problem.

So at times new orthographies are created, new pedagogic materials are created, but it is

restricted to only a few members, then there is a problem. So you have to talk about the materials and their users both, in order to understand the vitality. So we find that literacy materials are available, but not part of the school curriculum. Literacy materials are produced, but these are routed only through home-based literacy domains, but not for formal school curriculum; you need to check this point for yourself. And at times we know that or there are orthographies created for different communities, but only a few community members know them.

At times we find orthographies where many community members know them and at times there is no orthography. So I will give an example here and before that, I will also tell you that it is a very tricky issue in Indian context because we are I mean there is multiscryptality, multilinguality and it is a complex question in terms of Indian background. So we find that for several communities, literacy materials are restricted to informal schools, and only few speakers; orthographies are not known. Again I will give an example from Mundari, where there are Mundari Sunday schools for different age groups in West Bengal-Jharkhand-Odisha network, which is very useful for language revitalization and many young community members are also learning their language through these Sunday schools. But this is not available for formal schooling.

So there is a restriction. So, if these are being introduced in formal schooling things will be different because there will be one round of acceptance from a different layer. And again as I told you that if one script is made by a community member and only a few members know the script; it is neither used in school nor taught to the children through informal initiatives, only some adults know this and use them for literary works, then there is a problem. So I have a question for you at this point and I am asking you to identify your challenge. If you are a community member and if you are bothered about the script issue and literacy issue, you have to pause and think about all these factors. Now we will think about the institutional language attitude and policies.

Through institutional policy, we will come back to it even in future when we will talk about language attitude, but here I will tell you that at times there are equal support for different languages. There are countries which value all the languages as assets, but there are differentiated supports as well, where non-dominant languages are explicitly protected, but there are differential treatments. May be on pen and paper all the languages have equal status, but in reality there is some kind of differential treatment. So if you are a community member, I hope you can talk about it better than me and because you know, we can go as an outsider and certain issues which are important, but as a community members you can talk about the differentiated support in a more detailed way. At times we find there is passive assimilation where dominant groups are indifferent of the minority languages spoken.

At times the dominant groups do not have any idea about the minority languages which are existing in their own area. So this is also some kind of indifference and this indifference affects the use of language by the speech community who are non-dominant. We can label it as passive assimilation. There is also active assimilation where we find that government encourages the speakers to abandon their language. So maybe there is a pressure from the government to use one dominant language only and as a result, the smaller languages are abandoned.

Forced assimilation is something where there are directives- materials or published on behalf of the government and directives say that you are not supposed to use the smaller languages for formal domains like school or administration. And finally, there is prohibition where minority languages are actually prohibited from all the domains. So it is important for you to evaluate the attitude and policies of the institutions in your area whether they are supporting your language or the language on which you are working on and it affects the use of language and it can also initiate shift. So I will give an example of the Rabha speakers where we worked in the Assam-Bengal border and we found that dominant speech communities like Bangla, Nepali and Rajbanshi, all of them are putting a pressure on the Rabha community members. Though the language is retained in Assam, but in West Bengal these small villages where Rabha is spoken, they are shifting to mostly Rajbanshi and Bangla and in some cases, Nepali.

So, this is something that we need to consider while talking about the vitality. Now I will talk about the community members' attitude. So earlier, we were talking about the institutional attitude and policies. Now I will talk about the community members' attitude towards their language. So there are two possibilities which are very common in Indian context.

One is that speakers are ashamed of using their own language specifically in public domain. So I have seen children in classroom are ashamed to use their own language if I ask them to speak their language. And as a result, they are explicitly trying to shift to a different language, but the opposite is also true.

We can find that there is.. the community members think that it is essential to protect the language because it is a part of their identity. So, as a result they will use it even in public domain and they will not be ashamed of their language. So I will give an example of Rajbanshi again from Jalpaiguri district of North Bengal, where it is a very interesting case because I found that the children from Rajbanshi community were earlier ashamed to speak the language in public domain. This is before the creation of a language academy which is a governmental support and after the creation of the academy, we find

that there is a growing consciousness about community members. And we find that children are willing to speak in Rajbanshi in school in the same domain where they did not use the language, their own language earlier.

Now the children also view their language as a part of their identity. So it is very important for you to consider whether the external factors or internal factors are important for the vitality. Now I will come to the point where I will talk about the amount and quality of documentation. So you have been learning about documentation and different aspects of documentation. I will tell you that documentation directly plays a role here when we are listing different factors for assessing language vitality.

So we say that following UNESCO, the type and quality of existing documentation of the language needs to be assessed. So any documentation will not be ok. Again it is not about ticking a check box, that there is some kind of documentation available for a language. The language has a grammar, the language has four dictionaries, five grammars and audio and then the language has some resources.

So it we should receive some status. It is about assessing both the quality and amount as I said. So there are some layers that have been suggested by UNESCO. So we can say that there is a layer called superlative, where you find that one language has number of grammars, dictionary; there is a flow of text. There are more people who are writing in their own language, maybe by using different scripts. Scripts does not matter here and there is some amount of audio-video recording for the community.

So it is case of superlative. We can mark it as good if the language has at least one good grammar. Grammar means I am talking about descriptive grammar here. I am not talking about prescriptive grammar. There are some texts and also audio-video materials which are accessible for others. May be it is accessible for you as a researcher or accessible for you as a community member.

We can say that it is fair if there are adequate number of dictionaries, some grammar and audio-video. We can say it is fragmentary, if there are some sporadic grammatical sketches and a few word lists. You have already learnt about some word lists. So, if maybe there is some kind of word list made by someone, some community member or school teacher.

So it is available for you, but it is fragmentary, sporadic. And we can say that it is inadequate if there there is only a few grammatical sketches, word list, fragmentary text and no audio-video. So there is no audio-video. As a researcher, you may create some audio-video and that can add to the documentation part. So here we are talking about the

importance of documentation.

This is a point that you must note. So I will give an example of Koda, an Austro-Asiatic language. We worked on Koda from West Medinipur district of West Bengal and we found that the language has fragmentary word list and almost no audio-video. We also tried to figure out whether we can find something from the public domain, but we found that there is almost nothing. So we discussed with the community members and we also figured out that there is a strong need of documentation for the language, so that the language can be revitalized.

So now, I will come to the assessment of all the factors. I am sure you noted down the factors, all the 9 factors that I have discussed and initially I told you about the scores. So you can put the score for all the factors that I have mentioned, following the different layers that you already know by now. And I will also share the link of UNESCO document towards the end as a reference. You can go to the website and figure out how the scoring has been done.

Now, there is a question of self assessment by the community members. So if you are a community member, so I will ask you to assess the situation of your community. So it has to be neutral and it has to follow all the parameters that I have mentioned as suggested by UNESCO. And if you figure out the nature of or if you can assess the vitality for your own language, the next step will be to figure out the need for support. What kind of support is needed? Do you need support of funding agencies in order to revitalize your language? Do you need support from academic perspectives for creating say teaching-learning material? Do you need a community-academics collaboration for something else, may be to develop language technology for your language? So it is important for you to note down the nature of support that you may need in order to revitalize your language. And I hope by then you will all already figure out the level of endangerment for your language and accordingly, you will figure out how to go for the vitality assessment.

So here I will give an example of Mundari speakers where we worked for a language technology initiative for low resource languages through MSRI and we collaborated with Mundari speakers. We are still working with some of them. We are trying to understand that if we can build some corpus which will be helpful for language technology in future, then it will be easier for the community members to have some technology in their own language and it will also be easier for the next generation to access technology through their own language. It will be helpful for the academicians to understand and analyze the nature of Mundari for academic purpose as well.

So it will serve different purposes and it is very important for us to note. So I have been talking about internal evaluation, but what about external evaluation. So you can say that if you are not a community member, you are going to the community to collect data and as a part of your research, you are trying to collect this vitality data for doing assessment of vitality, overall assessment of vitality for the language, whatever evaluation you are doing it can be used as a policy tool. So what do I mean by policy tool? If you have a detailed account of vitality survey of the language, the target language, then you will have a an overall picture of the situation whether there is a chance of shift, whether the members are maintaining the language, whether there is lack of material in particular domain, etcetera. So if you have all these details, all this corpus of language vitality assessment can be used as a policy tool may be for educational policy.

So I will be explaining educational policy in next few classes. So I will also tell you that external evaluation can be official; may be one institution is willing to conduct this vitality survey for official purpose. It can be voluntary, may be one academician is trying to do a work; may be you want to work on your PhD and/or masters' thesis and want to understand the vitality assessment of a community. So, this is voluntary work. So you have to think about all these areas: revitalization, literacy development and documentation. These three concepts are super useful for us throughout this course and we will come back to these concepts again and again, but here I will tell you regarding vitality assessment, you need to focus on these three factors, very important for all of us.

And if you are individual language specialist, if you are a local speaker of a community, if you are a part of an NGO who is working on a language or literacy, if you belong to a governmental or an other institutional organization, I am sure you have certain targets in your mind. So I will request you to think about your targets and then try to understand the overall assessment of vitality whether it is meeting your targets or not and then come to the generalization. Because without the complete picture of vitality assessment we cannot go for evaluation. So we talked about both internal and external evaluation here.

So I will come to the conclusion part here. I will again tell you that we have discussed 9 factors for assessing language vitality in today's class. I hope you will remember them and we also talked about how the community members can assess it. So it will be an assessment from inside. But external evaluation and collaboration is also important because external evaluators help or collaborate for revitalization support.

These are the key references. So this is the UNESCO document you can go through, and Thank you!