

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-21

Lecture 21 : Issues related to Ethics

Hello! Welcome to the Lecture 21st of the course, Tools and Technologies of Language Documentation. Today, I will talk about issues related to ethics. So, ethics is a very important aspect of any work that you do and so, for language documentation as well. In some of the journals, if you want to publish your work, they need to see the ethical clearance that you have got from your institute or from your sources. So, that is important and also, ethical clearance should be received from the language experts. So, that is there and other than that also, there are certain ethical practices which we should follow in the field, and also when we come back and working with our data then also we should follow these ethics.

So some of the ethical issues related to the field, I will discuss. There are some points like informing the language expert about your work, taking permission, then not forcing too much the language expert to give you data, respecting their rules and others. I will also show you a sample ethical form, which we generally use in the field and we ask the language experts to put their signatures on them to show that we have the ethical clearance. And also, there are certain ethics related to the data annotation and when you are actually analyzing the data.

So, we will talk about that. So ethics of the field, very very important aspect. When you begin your work, you are in a speech community, you should inform the community members or your language experts what are you working on or what is the purpose of the work, what are you doing. So the language experts should be aware of the researcher's work. One should not think that maybe they will not understand, we should not describe; it is not so.

If you feel they will not understand the technicalities of your work, you have to explain them in a simpler language, so that they can understand. So, it is our responsibility to make language experts understand what are we working on. So, because they should know why they are contributing the data. It is we are taking their time, we are consuming their energy and time. So, they have the right to know what are they giving this data for.

So, we need to explain that to them. Inform the language expert about the use of data. So, how are you going to use the data? In most of these cases, the data is used for academic purpose, the data is used for various type of academic purpose. So, you can mention that if there is any financial profit involved then, of course, you should also mention that. And then according to the rules or according to the sponsors, they might also have some share to that financial profit, if you are gaining something out of that data.

So, sometimes the language experts might feel that the researchers are using the data for commercial purpose. So, maybe you are just collecting the data for academic purpose or just for documenting a language. So, you are not going to sell the data or earn money out of it, but then you should clarify that to the community otherwise, they might not know about it. So, that should be clarified. Permission is very important to be taken.

You cannot just record anything, you cannot use hidden cameras or hidden recorders; that is very unethical and that type of data collected cannot be used anywhere, because you do not have the permission. So, we need to take permissions of the language experts, when we are in the field. Before every session, we should have their permission to record something. So, if the individuals involved, they should give you if there is a group, so everyone present in the group should give you the permission. If there are certain pictures which you are going to click some objects which are there in their house, in their courtyard, then also you should have the permission to click pictures, otherwise it is unethical.

Permission from the guardians should be taken in case of children or babies. So if you want to take data from children then, you need to take permission from the child's parents or guardian. So for India, it is under 18 years old, you need to take permission from the guardians or parents. So, a researcher may feel like recording some session secretly. So, you might see something happening and all of a sudden, you feel like recording it because something is happening spontaneously.

There is no time to take permission. In those cases, what do you do? So if something is happening, if suddenly two people are like arguing and you want to collect that data, how do they argue, what is the language that they are using, the pragmatics related to it. So you want to record it, then you might record it and then afterwards, you should take

permission from the people who were involved in the whole discussion or discourse. If you cannot take permission or if they do not give you the permission, then you should delete that recording. That recording should not be kept with you.

So if by chance, because of some spontaneous activity, some performance you have to record and you do not have prior permission, in that case, if you record, you need to delete the recording. Otherwise, if you get permission, then that is fine. So in most of the cases, what you do is that, you take prior permission and then you carry on with your recording session. This is a sample ethical form. So generally, what we do is that we create an ethical form, when we go to the speech communities.

So, here you can see that there are certain criteria like the language expert writes his or her name, can write his or her father's or mother's name, age. So, little details about the language expert, so that he or she can be traced. And then about the language also. So the person can say that "I give my consent to provide information with regard to my language.

." So the name of the language can be there "and culture for academic purpose.." So it is clearly mentioned that, the data is collected for academic purpose for whatever project it is. "I understand that the session will be recorded and the recording may be stored, duplicated and distributed for the academic purpose of the project". So, all the details, how the data can be used, all those details are mentioned here in a very very short manner.

Now this form is in English. In many places, people will not understand this form. So in those cases, what do we do? We translate it into the original language or the language which they can understand. If they cannot read it, we read it to them and make them understand that what we actually want to say. In certain communities, you can see that, may be certain elderly people, they do not speak the language you speak, they only speak their native language.

Then you get a translator and this translator, actually translate the whole thing to that person and then only that person can give you consent of recording. So, the person can sign or can put his or her thumbprint. That is there and so basically, when you have this form filled, that means you have the ethical form ready with you, you have their permission. So, this is very very important. Now, in many cases, what do we see is that of course, it is very useful to get this type of hard copy ethical form, but in certain cases, you can also record the whole session.

Sometimes both can be done together. So may be, you are reading this form to the

person, you are making him understand what this form tells and then, you are taking a verbal consent as well as you are asking the person to sign here and the whole session is being video recorded. So that way, you are safe; you have all the consent ready with you, in the form of this form and also, in the form of your recording. So, that is a good practice to do. So, you have the ethical form ready and everything, but still there have to be some limitations, like you cannot say the person that "You have to sit and just keep on talking".

So, if the person feels like he or she is tired, he or she wants to leave, then you cannot force them to just sit there and just because they have given you the consent, that does not mean that they need to sit there and keep talking to you. So that is very important; you can't force the language expert. Sometimes they are tired, sometimes they have their own jobs, other works to look at. So one can't force them to just sit and talk to them. Again, you cannot ask them questions which can be embarrassing for them.

So, if certain questions are making them feel embarrassed or if they do not like to talk about certain aspects, then they should not be forced to do that. Their rules, any speech community's rules should be respected. So, rules and customs vary from community to community. One may find certain customs or rules to be funny or weird at times, even immoral many times. But the researcher is not allowed to speak about those things, about those feelings, rather the researcher should respect those customs and cultures and practices.

So you might find it amusing, but then still you should respect it, rather than commenting on that. So, that is very very important. We cannot be judgmental towards the practices of any community. So, every community has its own practices and for an outsider, it can be amusing at times, but still we are not allowed to show disrespect towards that and that, I believe is not true for the field; that is true everywhere actually. So, we need to also limit our time.

So, it is important to make the language experts feel comfortable in every way. So ideally, it is said that the sessions should not be more than 2 hours. So, maximum it can be for 2 hours. So, generally 1 hour is actually enough, because in 1 hour, people feel tired.

So 2 hours. It can be extended to 2 hours with a little break. So, maybe after an hour, the person have some snacks or something and then again, return back to the session. So, there can be like 2 hours session. So, if it goes beyond that, then it can be very very tiring and we should not do that. So sometimes researchers, they do not have much time.

They are in the field for a limited period of time and within that time, they have to

collect lots of data or we feel like collecting more and more data, but still we should not actually burden our language experts with our targets, rather we should limit the time. We should not actually force them to just sit with us. They should feel comfortable in every possible way, physically, emotionally, every possible way language experts should feel comfortable. Then only they will be able to talk to us, discuss with us. At the same time, it is unethical to make the language experts exhausted.

So, we cannot just keep on asking something. Sometimes they might feel bored, then of course, we should shift the topic or we can talk something other than collecting data. So, that can be a thing. At certain times, we actually disturb their routine.

So, that should not be the part. We should not try to disturb their routine or stop them from doing their daily course. That can actually irritate them or can be harmful for them. So, we should not do like that. Again, when we are actually sitting with the language experts, we are consuming their time and energy. In those cases sometimes, we also need to pay them something, not always in cash or in the form of money but at any other type, in cash or in kind.

So, we can pay them something or sometimes you can gift them something, which can which they need or which can make them happy. So, if you are collecting data from children, you can also give them something, which will make them happy: books or something. So that is a good practice actually. If a daily wage worker is giving his or her time to you, is sitting with you, then of course you should remunerate.

The person should not be at loss. So data collection at the expense of language expert is unethical. So, if the person is not getting paid for the session and also not getting paid for his or her daily work, then the person is at loss and that is very very unethical. Some language experts, sometimes are young and they wish to have a certificate or something. So they will ask for that. It is our experience as well, where we have seen in certain villages, the young language experts, they help us a lot, they take us to different families and introduce us and help us in collecting data.

In those cases, they wanted certificates from us that they worked as language experts, they helped us in documenting a language. And in those cases, we should provide them with a certificate, if that is possible. Because that way, it also motivates them and also, that is a way to show gratitude towards them, that of course their help actually helped us to improve our work. So, that way also it helps. So this work, if you are providing them with a certificate, with a document, which they need, if that is legal, we can provide them with that.

That can actually work as an encouragement for them. So next time also, they will be eager to help and at the same time, those certificates specially in Indian scenario, can be helpful for their career as well or for taking admission to certain courses or something. We know that in Indian scenario, certificates are very important. Those type of certificates can actually help them.

So, we can try to provide them with those. Now, these were basically ethics related to the field, what I was talking about. These ethics we should follow in the field. Now, when we come back from the field with the data we have collected, then also there are some ethical practices which we should follow. Firstly, we should always acknowledge our language experts. If we have collected data from our language experts, we should try to acknowledge that.

It is important to acknowledge the language experts and the community from where the data has been collected. Details of the community members should not only be logged in the metadata, but also be revealed in the research works and other publications in the form of acknowledgement. But if they have helped you more with the data, then also you can make them co-authors. In certain papers, we do see that community members become co-author with the researcher. So basically, we should also involve them in the outputs as well.

It is not like just getting the data, coming back and just saving their name in the metadata book. It is also about acknowledging them and including them in the whole process. So, that is very very important. Another very important ethical aspect is not misusing the data or manipulating the data and that at times, people do. So, for the convenience of analysis or for some other aspect, people do manipulate data which is very very unethical.

So, that should not be done. Secondly, the data should not be misused. If there are certain video clips which you have got, you should actually use it for the purpose you have brought it and if you feel it should not be made public, then it should not be made public. So, there can be certain things which language experts tell you, just to tell you; they do not want it to be public. In certain cases language experts, they do not want their names to be revealed.

So, in those cases we should respect their choice. We should not make their names public, we should not try to actually publicize some videos or something, which they do not want. So, we should not do things like that. So, it is very unethical also to sell video clips or certain type of data, which you have collected for academic purpose or using them for any type of non academic purpose. For gaining popularity or for any other purpose also, one should not use those clips.

But at times, we have seen people do that. They upload video clips which they have brought from the field. They upload it in different types of social media pages, YouTube and all, which is again very unethical. If you do not have the permission for that, you cannot do it. So, those types of things should not be done.

And as I mentioned, you cannot manipulate your data. Secondly, you should not do things which can actually hurt the community. Sometimes we have seen people write things, which actually hurt the community. So, of course, we know that we collect data and one of the methods of collecting information is also observation method. So in the field, we observe lots of things.

But how do we interpret it, that depends on us. So, there can be again various types of interpretation, but if we are writing something which can actually hurt the community members, then we should not write that. Because the community members feel bad because they have given their time, their energy and then you are actually writing something which is hurting them. We have seen some research papers or some articles, where people do that and in most of the cases, community members they know about it. Somehow, they come to know about it because of course, they have been mentioned so they will know about it. And when they know about it, of course they are hurt and they are sad.

What happens is that, when in future, other researchers go to the same community, the community members do not want to interact with the researchers, because they have lost their trust on the researchers or academicians. So, one should not do things like that, which actually makes the task for future researchers difficult and at the same time, also hurts the sentiment of the community members. And this point, giving back is very very important, which is also actually taken as the part of language documentation, that is, giving back something to the community, not in the form of money or some goods. It's in the form of grammar, dictionary, primer, any thing, any app-related thing, so that the community members can actually use it. If you are collecting linguistic data from the community, then you should give something related to the language which the community members can use.

And why this is important? Because mostly, we are dealing with the smaller languages, with the languages which are less resourced. So, in those cases, if you are creating resource for the language, that is the best thing that you can do for the speech community. That way, the language gets a scope to be used. That way, you are actually helping the language to be preserved. So, it is very very important that we should not just collect data and keep it somewhere or just publish our academic papers in several journals and all,

rather we should also try to create something out of it and give back to the community.

So, if you are creating dictionary and you are keeping only IPA fonts in it and there is no local script being used, then the community members might not be able to use the dictionary. So, it will not be useful for them. Rather what we can do? We can of course, include IPA, but we can be more inclusive and also, at the local language, may be the script that they use, and other local languages which they understand. So if we create a dictionary like that, that will be helpful for the community members as well, or if we can create simple picture books or pocket dictionaries. So, basic smaller things can also actually help creating language resource.

So, that is also a responsibility. And not only creating or publishing these things, but also bringing them to the community. So that the community members can actually know about these. They should know that these things are available, they can use it. So if I am creating an app, then I should also take that app to the community members. So that they can use, they can give their reviews and that will also help me improve my app and at the same time, they will know that with the data, I am doing something which is actually helping the community.

So, it is very very important to give back something to the community and taking it to them actually. So, that they can use it. So, I will like to conclude by saying there are two types of ethical issues related to language documentation. One, which we follow in the field and another, which we follow when we come back from the field. So, it is not only enough to be ethical with your work, but it is very very important nowadays, to keep a record of the ethics that you are following.

So, this record you can keep in the form of hard copy. But nowadays, it is more preferred to have a digital recording of it. So, to have a digital copy. So, if you are asking them something, taking their permission, if everything is done through the video recorder or audio recorder, then that is a best practice. So, you can have the hard copy as well as the recorded form and both, you can use actually where, whatever is required you can use it. For publishing your work in academic journals also, sometimes they ask for these type of permission letters.

So, you can use that for that. So, I would suggest if you are planning to work on certain aspects of language or if you are planning to work or document a language and you are planning to visit the field, then prepare a permission letter or ethical form. At the beginning actually, you should take the permission and then proceed your work. So that your work has that ethical clearance and then, you can actually publish it anywhere, talk about it anywhere because you have the permission and you do not need to actually get

scared of anything. So in both the cases, the researchers need to clarify the requirements, aims and objectives of the whole process. So, either you are recording it or taking hard copy or doing both, in the both the cases, what you do is that you actually tell the language experts everything about the data collection thing.

Why you are doing it? What are your aims and objectives? What are the outcomes that you are thinking about? So, you should tell everything to the language expert. Everything should be clarified. While transcribing or analyzing the data, one needs to be careful and should not change the data at any cost. So, certain ethical practices which we follow in the field and certain ethical practices, we follow after the field. So, when we are back and analyzing our data, transcribing our data or translating our data, then also we should follow these practices.

We can never change the data for the sake of convenience of analysis. So, there might be some complex sentence, we cannot actually analyze it and we cut it something and make it smaller and try to analyze it. No, that cannot be done, that is actually unethical and wrong. Changes as you find it offensive or immoral. So, there might be some narrations, some sentences which you do not like, which you feel that the speaker should not have said like that. This sounds very immoral or you might find that this sounds very offensive, one should not talk like that.

So in those cases also, you have no right to actually change it or manipulate it. Let it be there because you have not said it, the speaker has and again, from community to community what is offensive, what is taboo actually varies. So, the thing might not be offensive in that community. So, always keep the data. You should not actually change it because of your judgment.

Creating false data to increase the volume of data. Sometimes people actually think that "Ok I have these many sentences and in similar pattern, I can create more sentences". Learning the language is a different thing. You can always try to learn the language, you can then try to speak it, create sentences and actually ask the speakers whether you are saying it right or not, that is different. You can always learn the language and try to use it, but you cannot actually create false data to just increase the volume of your data.

That is again, unethical. Observation, as I mentioned, is one of the methods, but based on that you cannot just pass any judgment. So, you cannot say that I have observed this and that is why I am saying this, this is how I interpret it. Of course, you can do it in certain aspect, but you cannot do it by saying something, by publishing a statement, which can hurt or harm the community. So, we should not say things, which can actually hurt the community or anyway harm the community. Maybe you have seen certain practices and

then you talk about it openly and the community gets harm.

So, that should not be the thing. There should be acknowledgement, respect and involvement of the community members in the whole process of data collection and the outputs which are created. So, if you actually become more inclusive, if you actually involve the community members in the whole process, automatically you become more ethical. So if they also become part of the whole process, if the community members are there to guide us, to help us and we involve them in the project, then that way, we automatically become more ethical and we become more inclusive. In those cases, we, of course, we acknowledge them, we respect them and also involve them in the whole process.

So, it is not like as I mentioned, I and them. It is more like we together, the community members, researchers, together, we work for a language, together, we work to create resources for a language, together, we work to create outcomes like dictionaries, primers for a language. So that is it. I hope you enjoyed today's class. These are the references. Thank you!