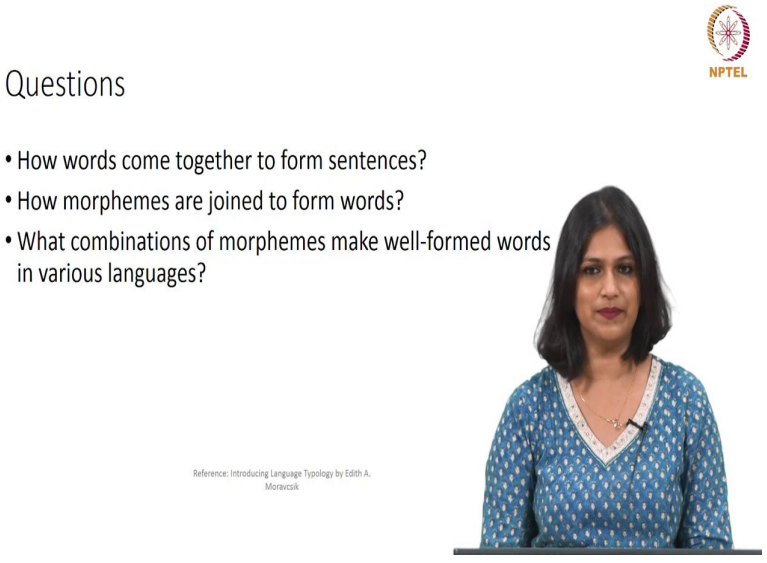
Appreciating Linguistics: A typological approach
Dr. Anindita Sahoo
Department of Humanities and Social Sciences
Indian Institute of Technology, Madras

Lecture - 21
Morphological Typology: A Cross Linguistic Study - Part 1

Hello everyone. Good morning. Welcome to a new session of my NPTEL course Appreciating Linguistics: A typological approach. It has been a while we have been talking about morphology and its introduction; what is a morpheme, how would you find out what is a free morpheme, what is a bound morpheme. And then, what is the connection of the types of words with the types of morphemes and with these basic informations about morphology as a part of linguistics, we will move to morphological typology now. I am going to talk about some empirical evidence from a crosslinguistic perspective.

(Refer Slide Time: 01:00)



So, without giving more details I will straightaway move to the discussions on morphological typology. If you remember it correctly, when we were talking about morphological typology, we were supposed to find out what are the different types of morphemes available and then what are the different combinations of morphemes available in various world's languages. I am going to give you the data that has been discussed in Moravcsik's book, but after that you

can find out the language that you speak, it falls in which category. What do you think, what sort of examples you have for this language?

To begin with, I have three very basic questions which eventually we are going to discuss through this unit or through this section. The first question that we need to figure out is how words come together to form sentences. This is not exactly in the domain of morphology rather it will be primarily in the syntax domain. But then to understand how words come together to form the sentences, we also have to understand how morphemes are joined to form words.

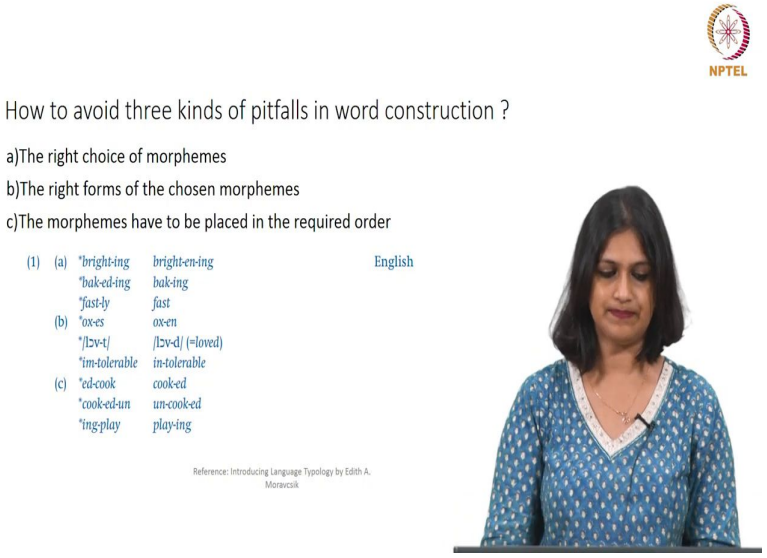
It is like one is a superset the other one is a subset. If words would be the supersets then its parts, the morphemes, are going to be the subsets of words. Before we understand how to frame a grammatically correct sentence or how to put the words together to find out a sentence which is acceptable in that sense, we need to find out first whether the words have certain organizational category or not. Do the words also have to follow the same parameter or the same pattern to be considered as words? So, once the words are considered as words, we will move to the arrangement of words to make it a different sentence.

We will answer the first question a little later. But before that we will try to figure out how we are going to arrange the morphemes together to form words; that is what our concern is or our intention is. And, as a part of this question the second question which is the most important one to address in this section. Let us go step by step.

First, we are trying to understand how the words are to be arranged in a sentence. But before that, we have to find out how the morphemes are arranged from words. Then we have to find out which combination of morphemes are able to make well formed words in different languages. I am not going to discuss all the languages in the world for sure; whatever data that I have in hand, I am going to talk about it and eventually I will leave it up to you to figure out which category your language belongs to.

The combination of morphemes is more important. After that, we will see how the morphemes are joined to form words and then we will move on to how the words are joined to make sentences.

(Refer Slide Time: 04:15)



The right kind of game starts from here. How to find out which way the morphemes are to be arranged? And if you look at this slide, you will see that there is a word pitfall. There are three different kinds of pitfalls in the word construction. Whenever you are trying to create or when you are trying to find out whether a particular combination is a word or not, you generally encounter three primary challenges. And what are the three primary challenges here? The first challenge is the right choice of morphemes. You should be able to understand whether the morpheme X has anything to do in this word or not.

If not, then let us drop it out, or if needed, we will add something else; that is the first problem or the first challenge that you encounter. The second challenge, the right form of the chosen morpheme. You have chosen the morpheme X; X might have different forms. X might have X-ed, X might have X-ing, it might have X-s, it might have X-es, taking into account the English lexicon. So, which form is more important here, that is also equally a challenging task. And the third one when you are choosing many morphemes you have to find out what is the required order for it.

So, if it has to go in xyz format, then yxz is going to be wrong, or zyx is going to be wrong; it must be in the xyz format. I hope I made it clear. So, three points: the selection of right morpheme; second, the selection of the right form of the chosen morpheme and third, the sequence of the morphemes which you have already chosen; it has to be done in the required

order. If these three pitfalls can be challenged or if these three pitfalls can be identified and worked on then you are going to get the right kind of words.

Let us look at the English examples given over here. Refer to 1 a and all these examples on the left. They are wrong because they are marked with an asterisk. The right side are all right because they are in the right form, right morpheme and in the right order. So, three things: right selection of the morpheme, right form of the selected morpheme and right order of the selected morpheme. If these three things are there then obviously, you are going to get the right kind of morphological order.

So, now look at 1 a. What is the first one in 1 a? Why is it wrong? Can you think about it? When I say brightening – brightening the left side of the left form of the word that we have written here, so this one is wrong. If this one is wrong, what is right? The right side is right. Right is not always right, but in this case, right side is right. So, brightening is the correct form; brighting is the incorrect one. Similarly, bakeding is the wrong form, but baking is the right form; fastly is the wrong form, but fast is the right form.

Now, taking into consideration these three examples brighting, bakeding and fastly; of the three kinds of pitfalls that we have recognized or that we have listed, can you tell me these three come under which category? Take half a minute and then think about it. The three categories that I have listed here, right choice of morpheme, right form of the chosen morpheme and right order of the chosen morpheme.

Which one is actually violated here? If you look at the data carefully, you will see that it is actually the first form which has been violated. Why has the first firm been violated? Let us look at each of the data carefully. When it is brighting, you have not chosen the right form like you have not chosen the right morpheme. So, the choice is wrong. So, what is wrong here? The mistake is that another morpheme which is extremely essential which is -en, you have missed it like you forgot to add it over there. Considering the choice of morpheme is not wrong. Brighting is not really a good word. So, -en is missing here.

On the other hand, in the second example bakeding, you do not need -ed. So, -ed is the wrong choice. Why is -ed the wrong choice? It does not have to do anything in this word. From the word bake, you can simply do baking. So, bakeding is bad. And, in the third one, you also

have a morpheme which is not required here; the -ly morpheme. So, fast itself is enough. When you are writing fastly, that means you have not chosen the right morpheme. So, the first pitfall listed you must have the right choice of morphemes. That has been compromised or that has been violated in the set of data that is given in 1 a.

Now, let us move to 1 b. Considering we have written a, b, c pitfalls. I assume that b is going to be correlated with data number b. The point number b and data number b they should come together. So, when it is oxes, it is actually the right form is oxen. So, -es and -en, what is the problem here? -Es is also a plural morpheme, -en is also a plural morpheme at some instances. But, in this particular case, when you are trying to work on the word called ox, the right morpheme is not -es.

So, you have chosen the plural morpheme fine, absolutely fine. No problem about it. Ox is a singular noun and by adding a plural marker you will make it plural, that is absolutely fine. But the concern here is that the form of the plural morpheme that you have chosen is wrong. Instead of -es you have to put -en, then the word is going to be correct. Then the second one is actually written in the phonetic transcription. So, the phonemes are listed here. If you say loved, then that is not the right morpheme.

So, sleep the past tense could be slept; keep the past tense could be kept, but love the past -t morpheme is not the past morpheme here. If you look at the past tense markers of English, sometimes it is -ed sometimes, it is -et or simple -t basically the /t/ phoneme. But in this case, when the word is love, it does not have the /t/ phoneme associated with it. It will always have the d thing. So, instead of lovet, it has to be loved – [d].

So, the right form of the chosen morpheme is not correct here. The chosen morpheme is the past tense marker, absolutely fine; love is a verb and it has been marked with -ed which is the past tense marker, that is fine. But, instead of -ed, if you are choosing the other past tense marker which is /t/, then that is going to make the word wrong.

Similarly, the third one im-tolerable and you have to compare it with in-tolerable. So, both im- and in- are the opposite markers. Think about two other examples; patient, what should be the opposite of patient? Impatient. Possible, what is the opposite of possible? Impossible.

So, im- as a morpheme it is generally the opposite marker. But in this case, when you are talking about tolerable as a word, it does not necessarily take the im- marker, it has to be in-.

So, both im- and in- are opposite markers. Patient – impatient; then you have tolerable – intolerable. In both the cases, both im- and in- are are opposite markers, but in case of tolerable, you cannot consider im- and in case of patient you cannot consider in-. So, inpatient is wrong similarly imtolerable is wrong. So, when you are trying to choose the right form of the opposite marker, you have to see what is the word to begin with what is the root word and accordingly the selection should be done or the choice should be made on the basis of the word that you have in hand. If the word is tolerable, it is going to be intolerable; if the word is patient, it is going to be impatient.

Now, come to the third category. So, these two things we have understood. You have to choose the right morpheme; then the second one, even if the morpheme is right, you have to see which form is used. Right form of the chosen morpheme should be used. Then the third one is the morphemes have to be placed in the required order. So, look at this the first 1 c; 1 c the first example ed-cook. So, ed-cook, -ed is also the past tense marker and when you say cook and cooked, cooked should be in past tense if you begin it with a verb, but -ed should come after cook not before. If you put it before cook then the morphological order is wrong.

So, the required order is actually post verbal. If the verb is cook, morpheme is going to come as in the post verbal position. Now, look at the other example given in the same data set 1 c cooked-un. So, in this case, cooked is not a verb rather cooked is treated as an adjective because when you say uncooked that is fine. I have some uncooked rice. If I can cook it, I am going to eat it. So, if I can say I have some uncooked rice, what is the required order of the morphemes? So, first you have to have un- which is the prefix, then you have to have cook, then you have to have -ed.

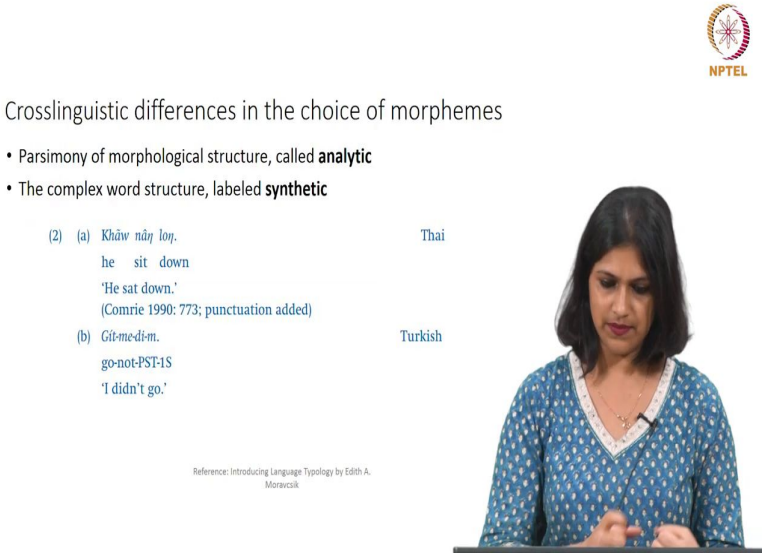
So, uncooked is the right order; but if you write cook-ed-un, that is going to be wrong. So here, what is the issue? You have chosen the right morphemes. You have also chosen the right form, that is absolutely no problem with the first condition and with the second condition. So, which condition is violated? The third one. Even if you have the right

morphemes and right forms have been used, you still have to take care of the place of the morphemes or the required order of the morphemes.

And, on similar lines, we have the last data in 1 c that is ing-play. So, -ing is the continuous marker or you can say progressive marker, or imperfective marker depending on the language that you speak. So, when you say ing-play, that is wrong. But you can say play-ing, that is right. The children are playing in the field or my child is playing in the school. Something like that. So, here also the same thing, -ing is the right morpheme. -Ing is also the right form of the chosen morpheme. So, what is wrong? The problem here is the third; third issue or the third pitfall that we encounter.

Instead of writing -ing before play, we should actually write -ing after play, so that the morphological order is going to be considered correctly. So, these are the three primary issues that we encounter when we are going to understand morphological typology. With these three pitfalls that we generally encounter in most of the words in any of the languages in the world. This is not exclusively related to English. These problems: the right selection of right morpheme, selection of right form and selection of right order. All these three things are a crosslinguistic set of features. So, you will find it in any given language.

(Refer Slide Time: 17:41)



With this we will move to the next set of data that will give us an example of two different types of morpheme or two different morphological types. So, you have to find out here when you are talking about your own language or when you are trying to understand how your own language works or for that matter any language that you want to study, as a linguist or otherwise, whether primarily they would be in the analytic category or in the synthetic category. So, when I focus on morphological typology these are the two very essential things I must focus on.

When I am trying to discuss this remember the languages like English and the languages like Chinese or languages like an Indo-Aryan language like Hindi or Dravidian language like let us say Telugu, they have different kinds of morphological patterns. These morphological patterns are broadly divided into two types, either they would fall under the category of analytic pattern or synthetic pattern. So, this is the first typological word that I am introducing in this discussion.

However, it is not that clear to identify whether a particular language is always analytic or always synthetic rather we will put it in a spectrum. There are certain languages which have more analytic features, there are certain languages which have more synthetic features. Let us look at the data that Moravcsik is talking about in the book. So here, if you look at the example given in 2 that is a Thai example and then the second example which is given this is the Turkish example.

So, in Thai, look at the word the gloss has been given here the first khaw nan lon whatever look at the first line of data. If you see, the corresponding English words are he, sit and down and there are three different words which are related to three different words in the sentence. So, he has a different word or he has been denoted with a different word, sit has been denoted with a different word and down has been denoted with a different word.

On the other hand, if you look at the Turkish data given over here, there is only one word and you see git-me-di-m. So, git, me, di, and m there are four different morphemes in a bigger word that is a complex word for sure because this is a polymorphic word. There are many morphemes given in just one word. So, what sort of difference do you see when you compare Thai data with Turkish? Let us have a look at it carefully. Turkish one word and it has been

loaded with information; there is a verb, there is the negation, there is a tense marker which is the past tense and there is also the number marking and then the person marking.

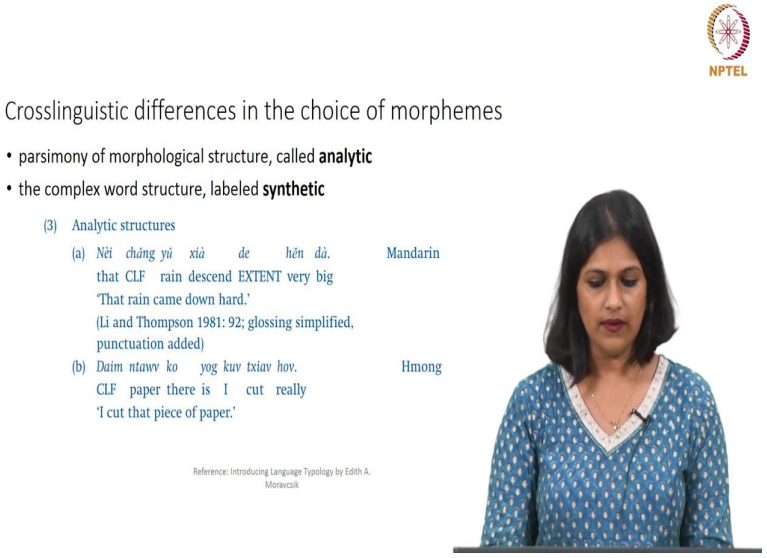
So, what is the person marking? It is first person. What is the number? Singular. What is the tense? Past tense. And also we have the negation, then we have the main verb go. All these informations are available or all these informations are loaded in one word that is git-me-di-m. I would apologize for my pronunciation if there has been any problem because I am not a speaker of either Thai or Turkish, but I will just refer to the data given over here. So, that is the reason we see a difference, one word with many informations; on the other hand, in the Thai example you have each word has only one information carrying.

So, khaw is a word, it means only he. It does not have any other information associated with it; nan is a word it means sit that is the verb, it does not have any other information; lon means that is down, it could be considered as a particle or it can be considered as a preposition it is up to the language speakers. So, when we compare, we see that Turkish as a language from the given data Turkish words very lightly are mostly loaded with informations at one go whereas in Thai, that is just different. So, Thai would be analytic type and Turkish would be synthetic type. I hope I made it clear.

Now you think about your own language. How does it work in the language or languages that you speak? If in your language each part like each unit denotes a different word then it will be an analytic language and if in your language one word has been loaded with multiple domains of information then it will be considered as a synthetic language. English is generally considered as an analytic language, but that does not mean that it never has loaded information based morphemes. However, we can claim safely that it has more analytic features.

Similarly, Turkish which seems to be a synthetic language, might also have one word, one morpheme, with one set of information. So, let us remember like this. Analytic words: one morpheme, one word, only one information set. Synthetic words: many morphemes, many informations, one word. So, the complexity of the word increases in the synthetic languages and the simplicity of the words is found in the analytic languages.

(Refer Slide Time: 23:50)



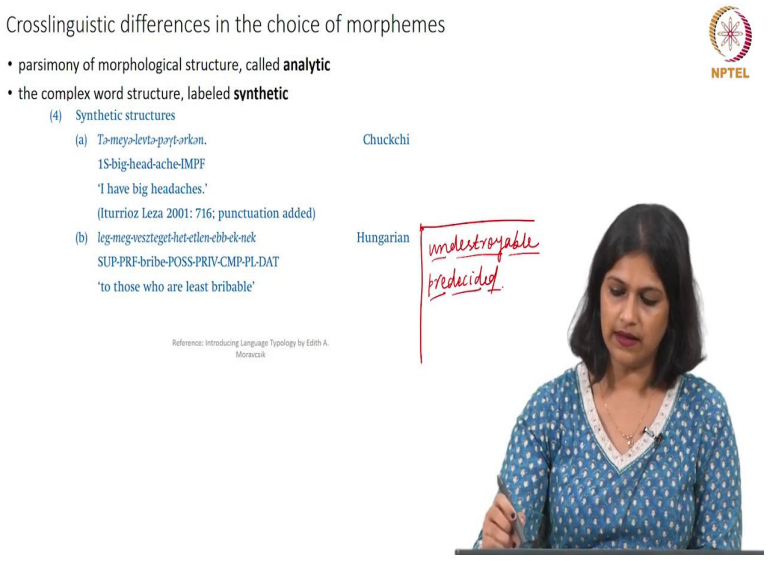
Let us check some more data. And, in this case, there are two different languages we have here. In most of the cases, Chinese languages or basically Sino-Tibetan or Tibeto-Burman languages they are considered to be analytic in nature. Why they are analytic? Because most of their words are of single morpheme. They are monomorphemic words. Remember this word, listen to me carefully when you say monomorphemic, that means, you have only one morpheme in the word and when you have polymorphic, that means, you have many morphemes in any given word. So now, let us see some more data related to analytic languages.

So, here if you look at Mandarin data, you will see that that train came down hard. So, each of the words here has its English counterpart and then they are primarily monomorphemic nei that is that, chang that is yu that is xia de hen da all of them. And, each unit in the first sentence it has only one morpheme and also it corresponds to only one set of information or only one information in that sense. Similar is the case with another language like Hmong. So, in that case daim so, that is the if you look at it. So, here also you see each of the morphemes has only one set of information and the meaning is I cut that piece of paper.

So, no complexity; you do not see this hyphen or dot things. If you see hyphen-hyphen in one given word; that means, many morphemes are associated with it or many morphemes together they create or they form a complex word. Here you do not really see the complexity

of the words. These are simple words, only one morpheme and they tell or they have only one information associated with them. So, Mandarin and Hmong these are the two languages listed here which are analytic in nature.

(Refer Slide Time: 26:05)



Now, let us see how to identify the synthetic languages. The synthetic structures, I just mentioned it, it must have many informations or many morphemes at one go. Let us look at the Chuckchi data given over here. I have big headaches. So, in this construction I have big headaches, the entire sentence which English requires at least four words 1 2 3 4. All of these informations are loaded in just one word in a language that is Chuckchi. So, how many morphemes? Let us count 1 2 3 4 5. So, five morphemes are attached to each other to create one word in Chukchi.

And, in Hungarian also, we see a similar kind of function happening. The example given in 4 b to those who are least bribable. So, this phrase, this is not really a sentence to those who are at least bribable, count the number of morphemes that it has; 1 2 3 4 5 6 7 8. So, this is a language which has a huge complex word that consists of eight morphemes.

So, one word having multiple morphemes obviously, is loaded with multiple sets of informations. There is a superlative adjective marker, perfective marker, bribe which is the noun here, then the possessive marker, then it is the pronoun and then there is a

complementizer, there in the plural marker and there is the dative marker. So, all these informations are loaded in just one word. You can think about how just one word can be so synthetic in nature. This is going to be heavily loaded with examples. When you look at Chuckchi, and when you look at Hungarian, you get an idea how synthetic languages should look like.

Now, let us think about English which is primarily considered to be an analytic language. Why this is considered to be an analytic language? Most of the words in English are monomorphemic or the complexity of the morphemes is not that huge in English. However, we have the words like let us say undestroyable. I do not know, I hope this kind of construction is correct; something which you cannot destroy.

So, in this case, when you say undestroyable, in that case so, un- is one morpheme, destroy is one morpheme, -able is another morpheme. Or you can say let us say predecided, predecided as a word. So, you have pre-, then you have decide, then you have -ed. These are also words which are heavily loaded in a construction like English. Generally English is put under the category of analytic words, but that does not mean that it would never have synthetic morphemes, not really.

So, we will consider the analyticity and syntheticity as a spectrum. Certain languages are purely analytic mainly the languages like Sino-Tibetan, Tibeto-Burman or like languages spoken in the far east countries. That would be mainly analytic. Then the languages which are in the extreme end of the spectrum which are synthetic, primarily the languages like Hungarian then if you consider India as a linguistic area, we have all the Dravidian languages. These are extremely synthetic in nature. Their morphemes are heavily loaded with informations.

So, let us not consider it as two types, but these types can also form a spectrum. In certain languages, the analyticity is higher, certain languages this syntheticity is higher. However, there are also languages which are purely analytic and there are languages which are purely synthetic. So, this is one type of morphological typology you need to understand. Now, my concern for you would be try to find out which category does your language belong to?

Whether that is purely synthetic or purely an analytic or that will be somewhere in the middle in the spectrum let us find it out.

So, that is an assignment for you. Think about it and then you might encounter questions.

Keywords: morpheme, morphological typology, right morpheme, right word form, right order of morphemes, analytic category, synthetic category