Basics of Language Science
Professor Rajesh Kumar
Humanities and Social Sciences
Indian Institute of Technology Madras, Chennai
Lecture 13
Phonotactic Rules and Word Formation

Professor: How do we produce sounds? What is the mechanism involved in that? And then there, what is the role of airflow, extra air, voicing that is vibration in the vocal cord and things like that? That was the idea of that discussion on sound. So now we will begin with words.

Student: When languages differ in the number of consonants that they have or the different sounds they have, what might be the reason that some languages have adopted more sounds as part of their languages while others have not? Can it be justified by reasons? Why do languages differ in the number of sounds?

Professor: Say it again, his question is why do languages differ in number of sounds? Why do some languages have more sounds, some have less sounds, some have certain specific sounds and others do not have. You were saying something?

Student: I did not understand. Every language has the same number of sounds.

Professor: No, not true, approximately the same. For example, you will not find a language with 20 and the other with 50. See that. So, they will have approximately similar numbers like 42, 46 or 48. No language will have 20 or 25 and then others will have 50 or 70. They do not have the same number, at the same time the differences are not too wide.

Student: There are some languages in Africa, which have more than 100 sounds. I cannot remember the name of the language, but they have unique sounds like, they also have that part of their language.

Professor: Right, right. They are called clicks.

Student: clicks, yes, the exclamation point. So, what might be the reason?

Professor: See, first of all, there are not 100. Okay. Second, some of those are called, some of those things are suprasegmental features. For example, in our languages, we do have something called nasalization. It is a part of many languages. But to give you an example from our languages, we have nasalization, which is different from nasals. For example, we

have a sound called ma, right, like, mama, ma. But then we have something else, which is called nasalization.

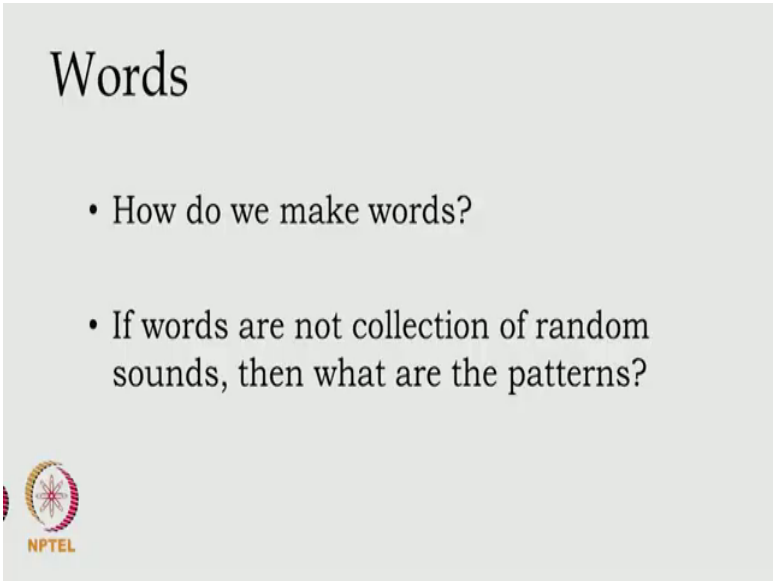
Now, what is nasalization, it is you said, it is something like, let us say when we say a word aankh, aankh. Aankh is the word for eye. It is a Hindi word for eye. Now, the first sound of this word is aa, it is not aakh, what is it aankh, the first vowel is nasalized. That moon and the dot thing is the marker in the writing system for nasalization. Now, you may have noticed, I do not have it on my schedule to talk about the written system, but you may have noticed in some words you just see a dot and in some you see moon and the dot, there is a difference between the two. I hope we get some time to talk about that.

However, we need to get to principles in more detail. Nonetheless, they represent two different things: the moon and the dot represent nasalization, which is a feature of a particular sound, the dot is simply representing another sound which is nasal. Now, how they do those things, I will show you some other time and if I find some space for that. The important thing with reference to your question is nasalization is something, which is called a suprasegmental feature.

Now it comes on a sound, it is not a sound by itself, therefore several of such things may not be, I am not denying because I do not know which language you are talking about. I do not know the sound inventory of that language to deny anything outright, but some of them are suprasegmental features that are number one. Number two and those are the things which are responsible for when someone says accents. Accent most of the time is used as a derogatory word to mark the language low, it is said your language has an accent.

However, though, the technical term accent also means differences caused by suprasegmental features. So that is also another aspect of that. There is one more thing which we need to discuss before we move to words, because now we are moving from sounds to words.

(Refer Slide Time: 6:13)



When we look at words, we know that they are not just a random collection of sounds, that is well established. We may not know each and every rule, each and every pattern which is responsible for the strings of sounds that we get in a particular word or what are the strings that are allowed and what are the strings that are not allowed, we may not know all of them. However, we know that they are not random collections of sounds. They have an underlying pattern in it, we are going to see some patterns today, which is going to answer the question of how we make words.

But there is another element which is larger than sounds and smaller than words and this thing is called a syllable. Have you heard this word syllable? What does it mean to you, when you hear the word syllable?

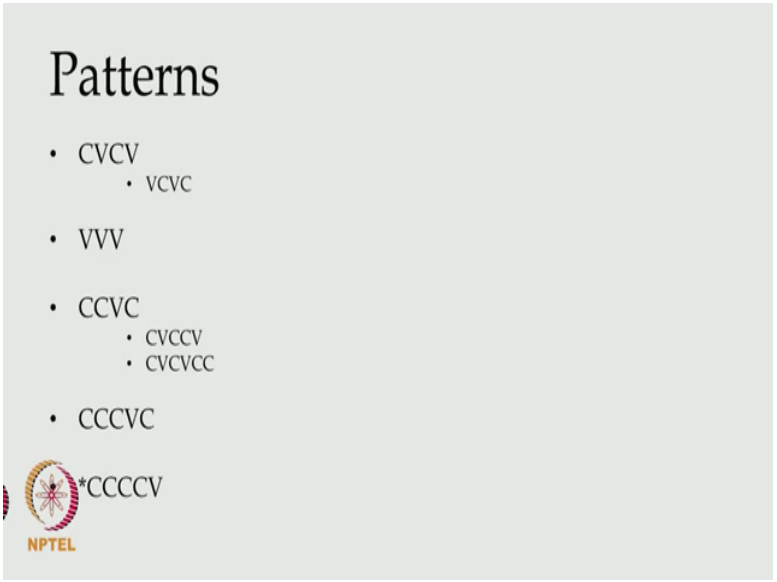
Student: Reference to a certain sound.

Professor: Reference to sounds. Actually, what it is, is the following. It is more than sound; it is larger than sounds and smaller than a word. For example, a word may have two to three syllables or maybe four syllables. However, one sound may also constitute one syllable. But the question of syllables comes in only when we are talking about words. If we cut the words in different parts, then we get several syllables. This process is called syllabification.

And we know that a word may have, may be divided into two syllables then they are called bi-syllabic or disyllabic words or a monosyllabic word if we cannot divide them into two or three then they are called monosyllabic words. So, please keep this thing in mind. There is

something called a syllable which is larger than sounds, but smaller than words. Syllables do not extend beyond the word boundaries; they stay within words. So now, let us very quickly look at these things, my idea is to get to sentences by tomorrow.

(Refer Slide Time: 8:56)



So, we have been looking at these things off and on, where these things simply mean consonants and vowels. Now, the first pattern that you see is CVCV, which tells you several things. Probably we have referred to these things while discussing other aspects, but once again let me say this thing. This, the first pattern simply tells you first and second, that the most common pattern for the formation of words is CVCV. There is nobody knows how many words we can make with this pattern, that simply means a lot, a lot of words.

There is no restriction on how many of them can be made. Also, it means we must have a vowel in a word. When we say most common pattern, we mean, if we have a proper alternation, or at a regular interval, consonant vowel, consonant vowel, then we probably get more, which also means that we can have a vowel consonant, vowel consonant, alternate alterations, which will mean the same thing.

We can also have simple words with two to three vowels. And again, the underlying thing is no word with only consonants in any language, these are the underlying things. The third thing when you see, what does this tell you when you see there are two consonants in the beginning. That is a cluster. It means when two of them are together, they are not necessarily, but they will form a cluster. Does anybody understand what a cluster means?

Student: Yes.

Professor: Cluster is not just the juxtaposition of two sounds. What does it mean? Somebody?

Student: A little part of the sound and the starting part of the second sound may be cut off.

Professor: Exactly.

Student: Might overlap.

Professor: No, no, no overlap. The part of the previous consonant this appears. It is important to know which part of that previous consonant, we are talking about a sound and again we are talking about a part of that sound that has to disappear in order to make a cluster. Do you see how, what a microscopic look it becomes when we look at words? So which part of that sound do we cut or get disappeared?

Student: Second one.

Professor: The vowel part, we know that every consonant comes with a vowel, which is called an inbuilt vowel, which is just one particular vowel, this vowel is a short a, it is shorter than a, and definitely way shorter than aa, aa is a completely different sound, we do not find that anywhere, in any consonant, when you see aa in a consonant, that is an additional sound. In a word like mama or kaka, k and a are two different sounds.

So, the sound like k comes with an inbuilt vowel a, which is a small vowel, it is called a shwa, but let us call it a vowel. In order to make it a cluster that goes away that vowel disappears and then we get the following consonant forming cluster with that one. And we have seen examples like school, station, school, scooter and many more. This cluster is possible in the beginning of a word, in the middle of a word also at the word boundary.

However, there are a couple of things that that work as constraints and that are important for us to keep in mind. Only the first sound is responsible for forming the cluster that is in a cluster, the first sound that is first consonant will lose its inbuilt vowel, the second one will not. If the second one loses its vowel, then it is going to form a cluster with the following one.

And this can happen at any position in a word, beginning, middle, and end with the break of a vowel. If there is no break, then they are going to form a cluster with three consonants. If we say there are some words which have a cluster of three consonants in the beginning of a

word. Okay, you get this question. If we say there are some words, which have three consonants, forming a cluster in the beginning of a word, what do we mean? What are the two consonants that are losing their vowels? First and second. Third one is not going to lose its vowel.

Now, before we see some more examples of the first, second, third, fourth one, let us look at the third one again. The difference between the first and second put together and the third is striking, which is, remember, what I told you about the first one, the total number of words that we can make with this possible pattern is a lot, that is infinite. We do not know how many, therefore we do not even count. However, the presence of a cluster in a word reduces the number of words way too low.

That is, the number of total words in any language with clusters is very few. This very few may not be a handful that you can count, but when we say very few, we mean compared to the first one, way too low. What is the meaning of this thing? If you look at it, if you can elaborate this thing a little bit more, what does it mean? Which one is more common? Which one is easier? We get an unlimited number of words with a pattern, definitely that pattern has to be an easier pattern. Only then we are getting so many of them.

Now, we are getting very few that is way to fewer with clusters, which simply means that is a heavy word. And for the vocal apparatus, for the word-formation process and generative mechanism, that is an expensive word, that is a tough, tough one. Therefore, we have very few of them. I am very well aware we did not completely answer your question of why there are differences in the number of words, but this will tell you something.

But one more thing I want to add to your question, not necessarily as an answer to this question is why do we have so many languages, which we do not have a clear answer of. It is just that we have so many of them. Similarly, it is an empirical fact that languages differ in terms of the number of words. Getting this thing.

Student: Sir, can we say that, because say a long time ago, because there was geographical isolation between different groups, civilizations, there were no civilizations if you go long back, but different groups because there was geographical isolation. So, one group, they developed their own languages, each group developed their own languages. And since they did not communicate earlier, they were not exposed to new sounds or different sounds which

can be produced, some groups explored their vocal apparatus more compared to other groups. Can we say that?

Professor: Probably yes, but you can say the same thing in a different way also, which is or similar things in a different way. See, we have talked about the language continuum even today. We have seen the continuum of Assamese, Bangla, Oriya and Telugu. Definitely Telugu has some sounds which Bangla does not have and Bangla has some sounds which Assamese does not have, or Oriya may not have, still they form a continuum. It is possible for us to see that continuum today, because we are aware of that geography.

Now, as you said, long time ago, and that is possible, it is just a speculation. Long time ago, when people started moving, that movement was probably permanent. We are also familiar with what we know as Big Bang Theory. That was way too big. And I do not want to go that far, we do not know, we do not have evidence of that time if we had language or not, or if it was there in what shape and all we do not know. But when people started moving, they probably never met with one another again.

So, they had one way of communication, again, they separated and then moved to some other place, then they had one, some other way of communication. So, we see a common ground. But when people separated, and they were communicating among themselves, they came up with another few sounds, which were specific only to them. And then as they moved, probably some of them were retained, some of them were lost. We do not know how many sounds may have been lost even in the languages that we have today, because sound change is one of the big processes of language change. Sound change is one of the big indicators.

Remember, I was trying to show you the distinction between three kinds of sa, dental sa, that is dental fricative, palatal sh and then there was a retroflex sh, and I told you that retroflex sh, some languages have lost that one and what we have, we only have palatal sh. And even between palatal sh and dental sa as some languages have one and some languages have others. Which means any in language A we may not have both palatal sh and retroflex sh, in language B we may not have palatal dental sa and retroflex sh.

And this loss is recent, therefore, we can even put our fingers on them. Some of the losses may not have been recorded. So, we do not know what has happened to sound. However, we do know sounds have played a great role in language change. And therefore, some languages in the present form, when we see them have a particular set of sounds, some languages have

another set of sounds, nobody says that they are going to stay that way. They are going to change too. However, what is predictable is the following and which will probably not change if all the languages will keep sharing sounds.

There will not be a language which does not share sound with another one, that will probably not change. I am very well aware of this does not completely answer that question. That is probably, that question does not have an answer. You can think, you can read, one can speculate, one can write something else, but can never be a complete answer to a question like this. Coming back to our discussing first and second patterns they give us a lot of words.

The moment we start with clusters, we start getting, we start seeing reduced numbers of words, which simply means that clusters are heavy. Now, the evidence of a cluster being really very heavy is coming from the fourth pattern. Can you think of some words with the fourth pattern where we have three sounds, three consonants forming a cluster to make a word?

Student: String.

Professor: Loud.

Student: String.

Professor: String. So, what are the sounds here? I suggest please keep writing these words, string and underline the sounds that are part of the cluster. You will see magic in this thing. The sounds are sa, ta and ra. True, this is a cluster, it is string, sa, ta and ra. More?

Student: Stray.

Professor: Stray. So, what are the sounds here sa, ta and ra. All right, more.

Student: Spray.

Professor: Spray. And sounds here are sa, pa and ra. More?

Student: Screw.

Professor: Screw. Sounds here are sa, ca and ra. I can stop but for your fun part, can you tell me more?

Student: Sclerosis.

Professor: So, the sound sa, ca, ra and la. More?

Student: Prawn.

Professor: Sorry?

Student: Prawn.

Professor: Prawn. Sounds are?

Student: ((???)) (25:40) two.

Professor: we are talking about three, so you are right. In a word like prawn, we have pa and ra. Probably you thought of this example because you were looking at the writing system where you see a w or something right. But it has only two sounds pa and ra. We are talking about three.

Student: Throw.

Professor: Throw. The sounds are?

Student: Tha, ra.

Professor: Tha, ra and after that you get a vowel. So that is not three, there is only two. We are talking about three. Can you be more imaginative? Just give me a word. Give me a word. Sorry.

Student: Stroke.

Professor: Stroke. All right. So, the sounds are sa, ta and ra. Now, while you are thinking, and I am still waiting for a few more examples, let me say the following. We are talking about some of these words from English, clearly. You are not coming up with words from our languages. What we are going to see as generalization applies to our languages also. It is just that you are not giving those sounds, those words. So, generalizations derived from English words are applicable to all languages.

So, do you see anything common in what you have seen, what you have given so far? Anything common in the examples that you have given so far?

Student: Start from the front and then go back.

Professor: That is very nice, true. But look at the sounds. What is the first sound in all of them?

Student: Sa.

Professor: Sa. Do you see this? It cannot be a coincidence. The moment you want a word with three clusters, you cannot have one beginning with any other sound. Any sound other than sa. And this is no big-time discovery or anything, it is just pretty obvious, we speak these words every day, we learn these things, the way I am telling you, it is just that we are, this is the process that is called reinventing. We are not inventing anything new. It is called reinventing or making something a little bit more obvious.

It is not that these words you did not know, these are your examples. It is just that we did not pay attention to this thing.

Student: Sir, did the words come first and then we found out that a bunch of rules fitted them or will the words form according to the rules.

Professor: This is like chicken and the egg story.

Student: Yes.

Professor: Yes, so a good question, though, you are saying did we have first sounds and then we made words?

Student: Did we have words and then we figured out the way we form.

Professor: Definitely.

Student: Was naturally following a bunch of rules or did we make the words based on a bunch of rules that we created?

Professor: No, we have words. Okay. Let me put it the following way. These things that I am trying to show you when somebody came up with these rules definitely, they looked at the huge data set. And then they came up with this pattern. It must have been very exciting at the time when someone figured out that you cannot have a word without a vowel. Sounds are very generic and as a matter of general knowledge today, but must have been very exciting for them.

Student: Why is it that we cannot have words with different sounds, like instead of starting with the sa, is it like we cannot have it or it is difficult to pronounce and people could not do it?

Professor: That answer that I can give you, but let me talk about the question that you raised before. It is a very interesting question. So, these patterns were made explicit, definitely, on the basis of a large data set, where the claim is not this fun, the claim is there is an underlying pattern of word-formation. Now, these patterns are part of principles and parameters, which are part of universal grammar, and which are part of the language acquisition device that we are born with.

We trigger them with these examples and therefore we speak the way we do. See this thing. So, it is not that we have a set of sounds in one compartment, and then rules in the other. And rules apply to sounds, and then we get words. So, all together and definitely the generative process is that we need to trigger those rules to get these things. That is the answer I can give you for that. Now, let us look at more and then you will see why they all begin only with sa. What is the second sound in these clusters?

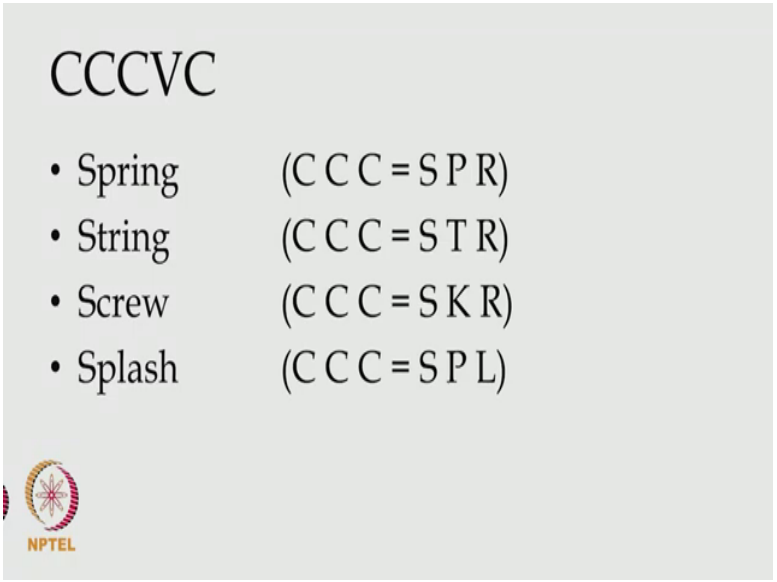
Student: Ka, pa.

Professor: Second sound is

Student: Ta, pa.

Professor: They are either pa or ta or ka.

(Refer Slide Time: 32:01)



You can come up with 100 words, but they all are going to have sounds like the second, one can only be pa, ta, ka, one of the three. Now, what is common in one of the three, we are going to look at that in a moment. And the third one can only be either ra or la. Now, again, the underlying idea is, this constraint is very strict, the stricter the constraint, the heavier the word. By heavy, you understand what we mean by heavy, the heavier the word for the apparatus, therefore, fewer in number.

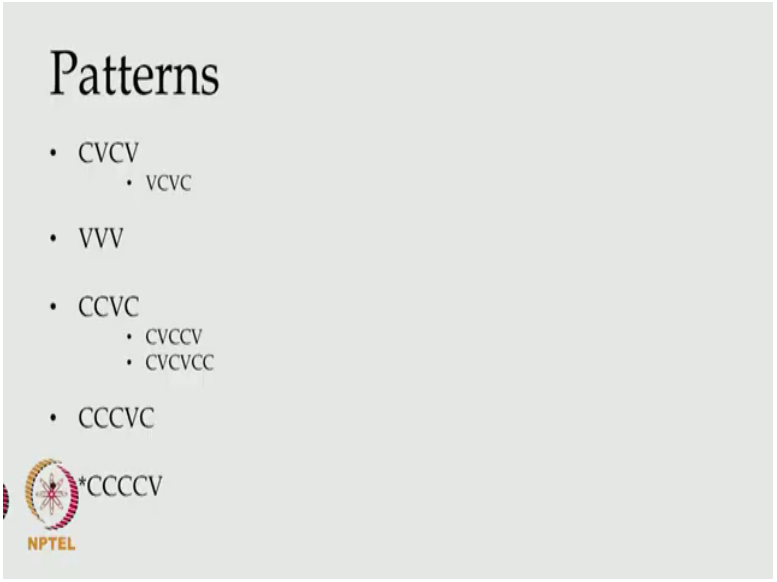
Now, these words with three consonants, if you compare them with the words with clusters of two consonants are again way too few in number in any language, they could be a handful of them. And these things being handful, I am not concluding on the basis of you not giving more examples. We do not come up with more examples, have nothing to say that we have established well before. They are very few in number, you can count them. I mean, you will not be too wrong, if you say English may not have, let us say more than 100 or 150 words like them.

By 150 I am not trying to put a number on that it could be 200. But, or even for that matter, it is 500, what is the big deal about that, which is not true, definitely 500 is not true. But the point is, if you can come up with 5,000 with one pattern, and then the third pattern, which just makes another string a little bit tighter, and gets reduced to 10%. And if you look at the previous one, which was millions, and you added one more constraint on that and got reduced to, let us say, 5 lakhs, or for that matter, 1 lakh.

That is too heavy a constraint. Just put one small constraint, it gives you, it becomes so heavy that it has a dramatic change in total number of words. That is the constraint we are looking

at. I am coming to this in a moment. To conclude this thing. Yeah, I am coming to that in a moment. I am coming to that slide in a moment.

(Refer Slide Time: 35:01)



Again, see the last one. Four consonants as a cluster. Do you see star in the beginning of that? For your information this star means not possible. Four consonants as part of the cluster, not possible. And this answer could be simple for you, why?

Student: Extremely heavy.

Professor: Extremely heavy, not possible, will become too heavy for processing, therefore not possible. And this is not possible in any language, no language gives you four clusters of consonants. Now, for you and in the interest of time, I should add in the beginning of the word. There are few words in English, where, at the end of the word, a cluster of four consonants is possible.

I think the word sixth has probably four clusters, or there are a couple of more, again, not very significant thing to discuss. However, the point is, even if at the end, even if it is possible at the end, it is possible only for a few words, not more than five, definitely. And five, I am being generous in the beginning of a word, definitely not possible in any language. Now, the point for us to take from here is to see the weight of these constraints.

And see the patterns involved in the formation of some of the words. These are called phonotactic rules. By no means I am giving you all the possible patterns and all the possible rules of word formation, I am just giving you a flavor of these things. Yes, sir?

Student: You told everyone has to start with s, if it has a cluster of three.

Professor: If it has a cluster of three consonants, the first consonant must be sa, not s.

Student: Suppose if there is a word called fcron, Fcron.

Professor: Fcron, how is that pronounced?

Student: Fcron.

Professor: So, the first one is sa.

Student: No, F.

Professor: No, but the way I hear you saying is Fcron. I am not familiar with that word. What I am asking you is you tell me what is the first sound.

Student: Fa.

Professor: So, the word is Fcron.

Student: Fcron.

Professor: So, how do we say this word, tell me.

Student: Fcron.

Professor: You cannot say this thing.

Student: He is saying fa without eliminating the shwa sound.

Professor: Facran, something like that, you cannot do that. Is there a word like that?

Student: No, I am just asking you if there is a word like this.

Professor: No. You are artificially creating one where the problem is, like someone said in the sound of fa, the shwa is not deleted, the sound a is not deleted. Therefore, you are having difficulty saying this. I am having difficulty saying that. And imagine we drop that. We will not be able to say that. I invite you to say that word. So, you are right on the paper we can do anything, but that is not going to be a word.

And definitely that word is not going to be available in the inventory of words to be assigned meaning. See the point. See, the formation of words is one part and how such words, if they

are legitimate words based on these patterns get meaning is a completely different process. And that also I think I have told you that the relationship between a word and its meaning is arbitrarily matched, and is arbitrarily put together. Why this thing is called a pointer has got no rationale behind it.

We could have called it aeroplane, it would still mean the same thing. I mean, if in the beginning, if everybody called it aeroplane we will call it an aeroplane, would not make any difference. It sounds ridiculous today, because the moment we say aeroplane, we have something else in our mind, because we know the word aeroplane is assigned for something else. Get this thing. So that is a different process. Coming to this, you can do it on paper, we will not be able to say that.

In order to be able to say a word, this is what has been derived. This is not a prescriptive rule that you must say sa, this is if you happen to say three clusters in a word, the first one must be sa, or first one appears to be only sa, let us put it that way. That is the point.

Student: What about words is like accreditation, where it is like a, k, r put together?

Professor: So, tell me what is the word accreditation?

Student: First vowel.

Professor: So that is a vowel, accreditation.

Student: when we say CCVC, that means, does it mean only to five letter words or any kind of?

Professor: Again, it takes a long time to get out of this thing. When we say five letter words, we are not talking about letters, we are talking about sounds.

Student: Does this arrangement generally mean only for words with five sounds or word that start with these five sounds?

Professor: No. First of all, we are talking about these words, this cluster of three in the beginning of a word that is the thing that I am trying to show you. But I am saying something more than that also. What I am saying is, when we look at a word, we are talking about, looking at sounds involved in the word, not the letters involved in the word.

Student: My question is, I mean, those rules apply only to five sounds words or?

Professor: No, it could be any number of sounds in a word.

Student: So, but this type of pattern, that means starting of a word, you can generalize it to starting of a word that can make a single word also. For example, if you say string, it is stringing, the same rules apply to springing also.

Professor: Sure, definitely true. I heard your question as a different one. And that is important to clarify. When you say words like accreditation, what are the sounds involved in that? The first sound is a vowel so, that is out. Second one is ka and ra. So that is a cluster of only two sounds, the following sound becomes a vowel.

Student: What about words like Kripa?

Professor: Kripa. So, what is that? Tell me.

Student: Ka, ra, pa.

Professor: Before you say pa, there is an intervention of a vowel. Before you say the word third one ka, ra and pa before there is a vowel. See that, true. So, you must get three of them. If you have a break, then you can have any, because we are not talking about words with two clusters, words with clusters of two sounds. If we are talking about two sounds in a cluster, then we can get any two sounds. We are talking about a cluster of three sounds in that three, you must have first one sa.

Please think about more words, I invite you, this is not the only 50 minutes where we can talk about this. Think about more words, but just keep in mind that we are talking about sounds, not letters. So, the first one must be sa, second one can only be either pa, ta or ka. Yes sir.

Student: You said that the second sound can only be sa, ka or pa.

Professor: The second sound can only be pa, ta or ka.

Student: But it could also be ta, right? It may not be possible in English but it is possible in Hindi.

Professor: Yeah. See, this is why we discussed both retroflex and dental. And then we have also discussed something called alveolar, the English ta. So, they are all from the same region, therefore I am saying ta. So, in our language, it could be either ta which is complete

dental, or in other languages, it could be an alveolar ta, softer, in a word like let us say string. It is not a retroflex ta, it is ta, therefore I am putting it that way.

This is the reason why we discussed the differences between, do people remember here the difference between ta as a dental, ta as a retroflex and then the intermediate alveolar one, which are the sounds that we get in English, very nice. Now what is the common between these three sounds pa, ta and ka. It is not all of them are from the front of the vocal apparatus, ka is velum. What is common between all three of them is they are all stop sounds.

Stop sounds mean total closure and then release of air pa, ta, ka. The common manner of articulation in all three of them is they are all stop sounds. And the common between ra and la is these two sounds are called liquid sounds. Now, liquid does not mean much, liquid is not something like physical liquid, water or oil. However, it is something of that sort in nature. I do not have much to say about them at this moment. But I can simply give you one generic example that you may find some people who interchange these two sounds.

When they would want to say ra, they would end up saying la or the other way around. Some of such things are also attributed to some kind of a speech disorder. And some of these things are very commonly known as metathesis or something like that. But the point is, whether it is a speech disorder or not, is not our job to hint at. Our thing is even in this speech disorder, the alternation is between only these two sounds. Some people, can you think of some examples, I have found instead of saying a rail I have found some people saying lail.

And where they are supposed to say la, for example, laal instead of la, they put ra. Now, this may not give you the answer. But this is just an application of what I am saying, which is these two sounds or liquids, therefore this interchange. Now, the third one of this cluster must only be either one of the two. Therefore, you get sounds like spring, string, screw, splash, and many other words that you gave.

Student: So, it is the case only one has to start at the beginning.

Professor: Most of them have been seen in the beginning of it, yes. Beginning of a word. It is difficult to get a cluster of three sounds somewhere else. However, that is not denied.

Student: Somewhere else is quite common.

Professor: Somewhere else, it is quite common.

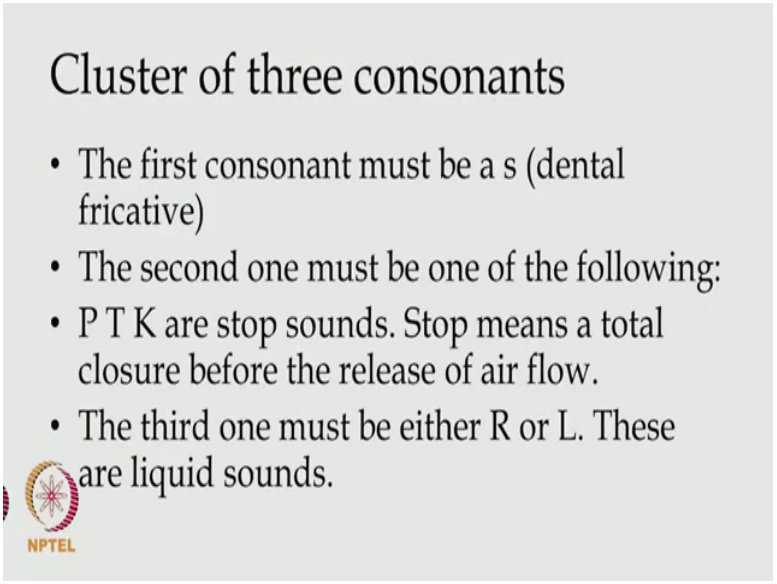
Student: At the end it is common.

Professor: Come up with some of the words. Probably there also it will be sa, pa.

Student: Pangs.

Professor: Again, you are talking about written things, written letters. If you say those sounds and pay attention to them, there will be a break with a vowel. Yes, and if at all that is that counts, that is one consonant, which is na, nasal, velar nasal. All right. That is all that is what I had to say to tell you today.

(Refer Slide Time: 49:15)



There are a couple of other things about words, which I do want to bring to your attention, because this is the, so far, we have seen some patterns. Now we are going to see some additional rules applicable for the formation of words. And then we go to sentences right away to see more rules of why and how sentences are not simply clusters of words. They will make more sense when we look at those things through the intermediate rules that are applicable in formation of words. And then a better understanding of sounds is definitely going to help us understand what we see a little later. All right, we stop here.