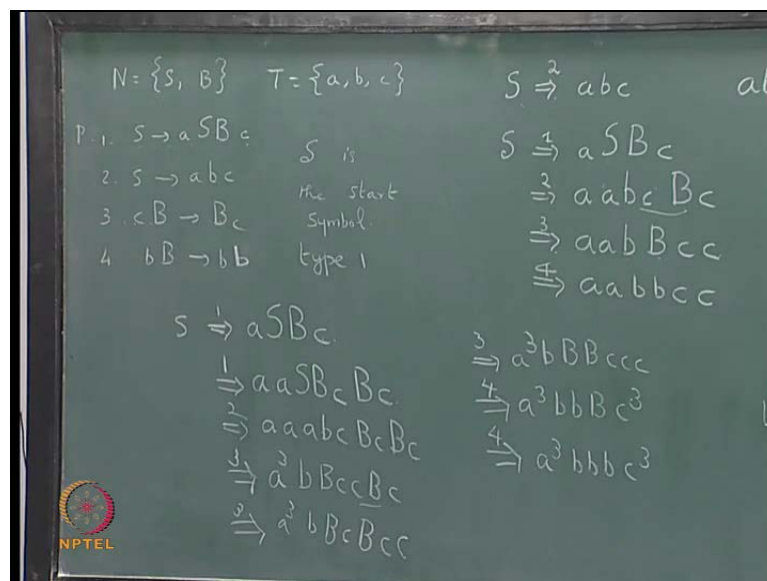


**Theory of Computation**  
**Prof.Kamala Krithivasan**  
**Department of Computer Science and Engineering**  
**Indian Institute of Technology, Madras**

**Lecture No. # 03**  
**Grammars and Languages Generated (Contd)**

(Refer Slide Time: 00:19)



So, in the last lecture we considered some examples of context free grammars, the grammars and the language generated by them. Today we shall consider some more examples. The first example we consider is this, there are two non terminals S and B, and three terminal symbols a, b, and c. And there are four rules, the rules are of the form S goes to aSBc, S goes to abc, then cB goes Bc, bB goes to bb, this is small b.

Now, you can see that this is not a type two grammar or a type three grammar, because on the left hand side, if you have a single non terminal it is type two, but here you find that on the left hand side you have a string cB and bB here. So, this does not belong to type two, but you can also see that the length of the right hand side is always greater than or equal to length of the left hand side. So, this is a type one grammar.

And let us see what are the strings, terminal strings generated in this grammar, the language generated by the grammar consists of all terminal strings derivable from the

start symbol. So, we will see what the terminal strings derivable? So, using rule two alone, from  $S$  we can derive  $abc$  using rule two alone. So,  $abc$  belongs to the language,  $abc$  belongs to  $L$  of  $G$ .

Now, starting from  $S$  if you use rule one you will get  $aSBc$ , now again for this  $S$  you can use rule two, then you will get  $a$  this  $S$  will be replaced by  $abc$ , this  $abc$  then this  $Bc$  remains as it is, this is not a terminal string. Now, you can apply the rule three  $cB$  goes to  $Bc$  to this  $cB$ . So, this  $cB$  becomes  $Bc$ , the idea of this rule is to bring all the  $c$ (s) to the right and the capital  $B$ 's to the left.

Now, this is applying rule three. Then applying rule four this  $bb$  can be, this capital  $B$  when there is a small  $b$  on the left side will be turned into a terminal  $b$ . So, rule four you apply and you get  $ab, aabbcc$ . This you can represent as  $a$  squared,  $b$  squared,  $c$  squared and this belongs to  $L$  of  $G$ . Now, let us consider one more example; I will start from here  $S$  use rule one you get  $aSBc$ , then again use rule one for this  $S$ , then you will get  $aaSBc$ , this  $S$  is replaced by this  $aSbc$ , but this  $Bc$  will remain as it is. Now, apply rule two for the  $S$ , then you will get  $aabc$ , this  $Bc$  will remain as it is, this  $Bc$  will also remain as it is.

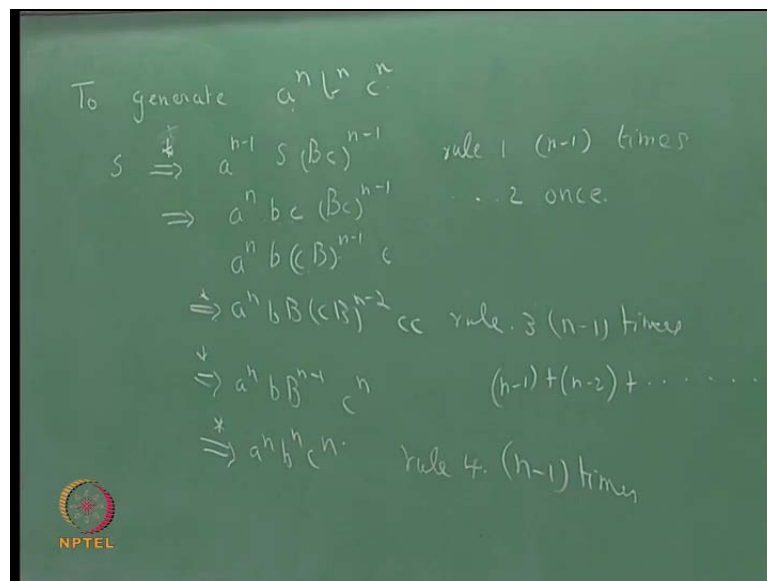
Now, use rule three repeatedly to bring the capital  $B$ 's to the left and shift the small  $c$ (s) to the right. So, you will have  $a^3c^2bd$ , if you apply the rule for this  $cB$  you will get  $b$ , this  $b$  is here then you will get  $Bc^2cBc$ . Again, you apply rule three  $a^3c^2bb^2c$ , if you apply to this  $cB$ ,  $b$  will be shifted to the left and  $c$  will be shifted to right, you will get  $a^3c^2bb^2cc$ .

Now, again to this  $cB$  you can apply rule three. So, you will get  $a^3c^2bb^2Bccc$ . Now, you see that all the terminal  $c$ (s) are shifted towards the right, all the  $a$ (s) are on the left, now you have 1  $b$  and capital  $B$ 's in the middle, the capital  $B$ 's have to be converted into small  $b$ (s) using rule four. So, apply rule four repeatedly you get  $a^3c^2bb^2B^3c^2$ , again you apply rule four  $a^3c^2bbb^3c^2$ .

So, the string generated is  $a^3c^2bb^2c^2$  belongs to  $L$  of  $G$ . Now, you can see the pattern what will be the language generated by the grammar, the language generated will be  $L$  of  $G$  will be  $a^n b^n c^n$ ,  $n$  greater than  $r$  equal to 1. So, this is a type one language or a context sensitive language, this cannot be generated by a type two grammar. In type two, what happens is when you replace by another non terminal, you

can generate equal number of a(s) and b(s),we considered an example,where you can generate a power n, b power n.But,you have to keep the same count for a,band c, for this something else is required, you have to shift the as seen this example, you have to shift the c(s) generate equal number of b(s) c(s) etcetera, but then shift the b(s) to the side and shift the c(s) and so on, and that require some context. So, this cannot be... Later on we give a formal proof that it cannot be generated by a type two grammar after learning some theorem.But,at present we will just make a note that this is the type one language which cannot be generated by any type two grammar.

(Refer Slide Time: 07:38)



Now to generate a power n b power n, how do you go about...To generate a power n b power n c power n,how do you go about doing this?First,from S rule one you have to apply repeatedly, I will put star.You have to use rule one repeatedly, when you use rule one S will be replaced by a S B c. So, one a will be generated on this side, one B c will be generated on this side.

So, n minus 1 times if you apply rule one,rule one n minus 1 times you will get a power n minus 1 S B c n minus 1, BcBc Bc like that.Then you apply rule two once, you will get a power n Bc Bc power n minus 1. So, this you can also write as a power n b cB to the power of n minus 1 right?Now, all this cB you have to shift the capital B to the left and small c to the right,for which you have to use rule number three there cB goes to Bc.How many times you have to apply this?

(( ))

There are  $n - 1$   $c$   $b$ 's,

(( )):

So, how many times you have to apply this? If you can apply the rule  $n - 1$  times to this, but then at the end one capital  $B$  will come here, but then you will again have  $n - 2$   $c$   $b$ 's. So, after applying say rule three  $n - 1$  times, what you will get is a power  $n$   $b$   $c$   $B$  power  $n - 2$   $c$ . Now, again to this  $n - 2$   $c$   $b$ 's you have to apply the rule again, that will end up with  $n - 3$   $c$   $B$  in the middle you have to apply again and so on.

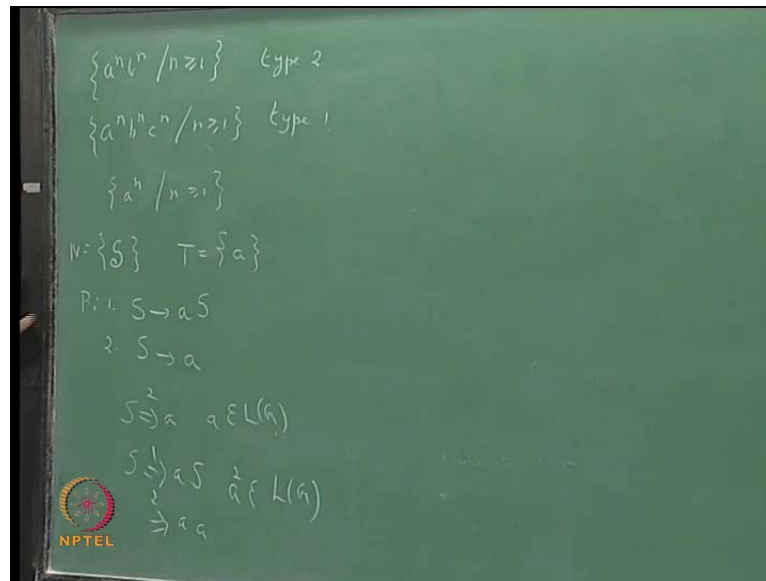
So, ultimately when you get a power  $n$   $b$ , this is again star  $b$  power  $n - 1$   $c$  power  $n$  when you get this, how many times you would have applied rule three? You would have applied it  $n - 1$  plus  $n - 2$  plus plus plus one times. What is this value?  $n - 1$  into  $n$  by  $2$ , so you would have applied rule three so many times for example, when  $n$  is equal to  $1$  you do not apply that rule at all, when  $n$  is equal to  $2$  you are applying  $n - 1$  into  $n$  by  $2$  is  $1$ , you are applying rule three only once.

When  $n$  is equal to  $3$ , what is  $n - 1$  into  $n$  by  $2$ ?  $2$  into  $3$  by  $2$  that is  $3$ ,  $3$  times we are applying check, we are applying  $3$  times. So, when you want to generate a power  $n$ ,  $b$  power  $n$ ,  $c$  power  $n$ , first you have to apply rule one  $n - 1$  times to get a string of this form, then this is equivalent to this **and you have to** you get this, and then you apply rule two once to get the  $abc$ , to get rid of that is  $S$  you have to use rule two. Then when you get a string of this form, you have to bring all the capital  $b$ 's to the left and small  $c$ 's to the right. So, you have to apply the rule  $cB$  goes to  $Bc$  repeatedly. So, first time you can use  $n - 1$  times, then again  $n - 2$  times and so on, so you will be using that rule  $n - 1$  into  $n$  by  $2$  times. When you do that you get a string of this form then finally, rule four is used to convert all the capital  $b$ 's, which is a non terminal capital  $B$  is a non terminal, those  $b$ 's have to be converted into small  $b$ 's. So, how many times you will be using rule four? Rule four will be used  $n - 1$  time, so ultimately you get. So, you must realize that, if you have a  $c$  you cannot to the left of  $b$  you cannot convert the capital  $B$  into a small  $b$ .

Only when the  $b$  get shifted in the middle towards the  $b$ , the capital  $B$  gets converted into

a small b. So, you may sure that all the b(s) are shifted, if you write b goes to b in the middle, somehow the b(s) you can replace in the middle also that you do not want. So, you want to convert the capital b(s) into small b(s) only when they are shifted to the middle, that is why this rule is there. This rule makes sure that all the capital b(s) are shifted to the left, then you are converting them into small b's.

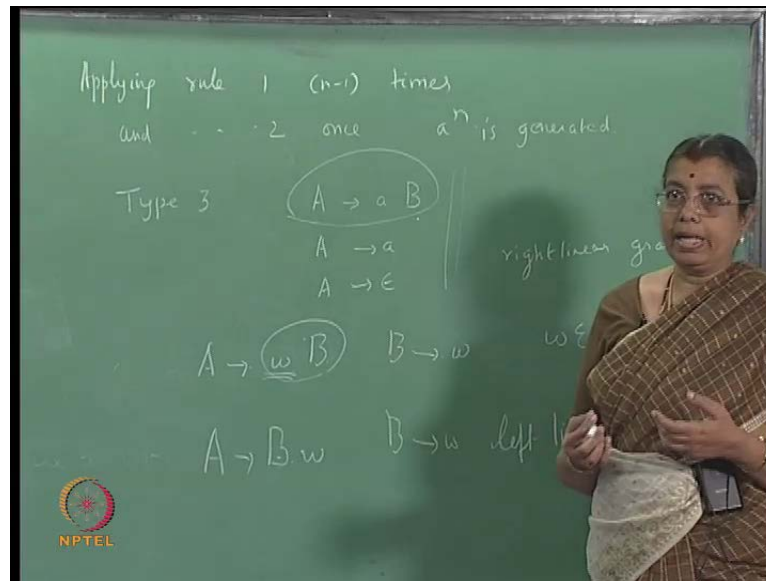
(Refer Slide Time: 14:59)



So, this is an very simple example of a context sensitive language or a type one language. Now, we have seen the grammar for a power n, b power n, we gave a grammar in the last lecture for this context; this is type two we gave. And a power n, b power n, c power n, today we have considered a type one grammar. I will consider one more a power n, n greater than or equal to 1. What grammar can you give for this?

Consider this grammar with S as a non terminal; only non terminal, and the only terminal is a and the rules are of the form, S goes to aS, S goes to a. There are only two rules; S goes to aS, S goes to a. So, with this we can see that applying the second rule alone you generate a, applying rule one and then applying rule two you get a a and so on, aa belongs to L.

(Refer Slide Time: 16:29)



Similarly, applying rule one  $n$  minus 1 times, **applying rule 1  $n$  minus 1 times** and rule 2 once you will get a power  $n$ . So, the language is this, **the language is this** what type of a grammar is this? This is regular grammar or a type three grammar. Actually, you will find that there are slight difference definitions given for type three. So, the way we defined this; type three we defined is the rules are of the form  $A \rightarrow a B$  something like this, left hand side non terminal right hand side as terminal symbol followed by a non terminal, or like this a non terminal going into a terminal, or you can also have something like  $A \rightarrow \epsilon$ . Sometimes for example, in this one instead of this if I have epsilon, what will be the language generated?

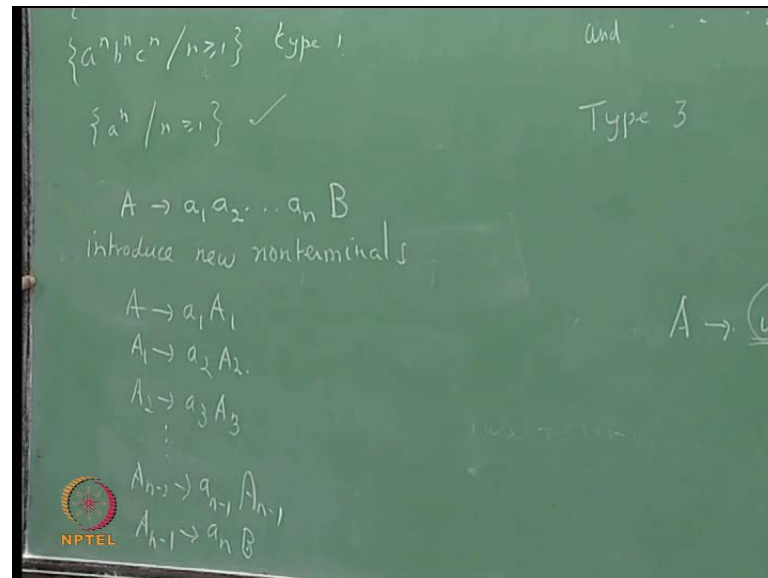
The empty string

The empty string also will be generated. So, you will get the language as a power  $n$ ,  $n$  greater than or equal to 0. Sometimes, the definition of type three is given like this, on the left hand side you have a non terminal, on the right hand side a terminal string  $(())$  by a non terminal or a non terminal going into a terminal string. So, right hand side you can have a terminal string it is a string;  $w$  is a string,  $w$  belongs to  $t^*$   **$w$  belongs to  $t^*$**  and  $B$  goes to  $w$ , that is on the right hand side you can have a terminal string followed by a non terminal or just a terminal string, it can be empty string also.

Now, the point to be noted is, this capital  $B$  which is a non terminal occurs there is only one non terminal on the right hand side, and that is the last symbol. This definition is also

considered as right linear grammar. **rightlinear grammar** Even though you write like this, these two definitions are equivalent, what is the reason? Suppose, I have a rule of this form,

(Refer Slide Time: 19:59)

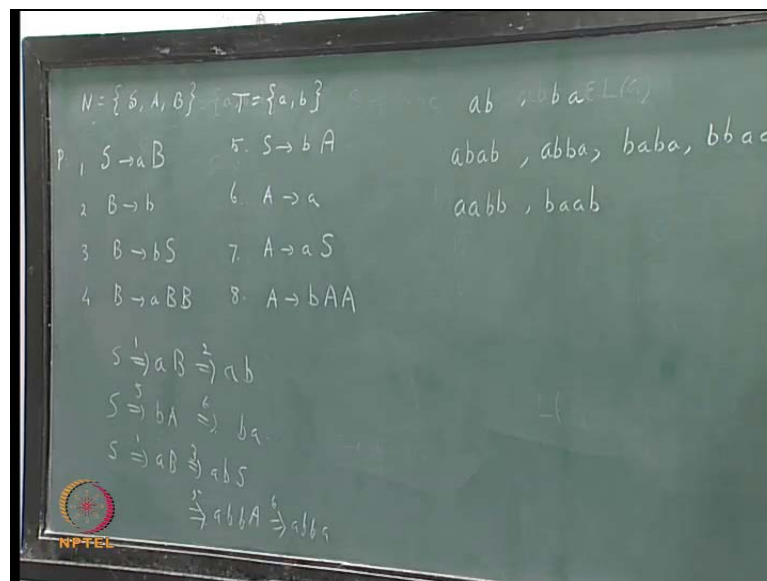


I have a rule of the form say  $a_1 a_2 \dots a_n B$ , this is a string instead of a single symbol, all are terminal symbols,  $a_1 a_2 \dots a_n$  are all terminal symbols. If I have a rule like this, I can split this into several rules of this type. How is that possible? I can introduce new nonterminals, **introduce new non terminals**; A going into say  $a_1 A_1$ ,  $A_1$  goes to  $a_2 A_2$ ,  $A_2$  goes to  $a_3 A_3$ , like that up to  $A_{n-2}$  goes to  $a_{n-1} A_{n-1}$ , then  $A_{n-1}$  goes to  $a_n B$ . Now, each one of this rule is of that form, left hand side you have a single non terminal, on the right hand side you have a single terminal followed by a single non terminal.

You are introducing new non terminals  $A_1 A_2 A_3 \dots A_{n-1}$ . So, the application of a single rule of this form is equivalent to applying a sequence of rules like this. If you apply this first, then this, then this, then this and so on, the effect will be the same as applying this rule. So, this rule in essence you can split in this form, so even though you considered this definition or this definition there are equivalent, whichever way you consider it is immaterial. Now, as for the present we will keep the definition like this, does not make much of difference if you consider like this also, or equivalently you can also have in this form, this is also will generate the same type of language  $Bw$ ,  $B$  goes to  $w$ .

The only non terminal on the right hand side, it can occur at the last symbol or it can occur as a first symbol, such a grammar is known as a left linear grammar. **left linear grammar** Again, this also generates the same type three language, but you must not mix up this and this. In the same grammar, either you should have all the rules left linear or all the rules linear, if you have one left linear and one right linear that no longer it can generate higher class. So, these grammars generate what are known as regular sets, language generated is called a regular set. The regular set is also any regular set is accepted by a final state automaton, the equivalent machine characterization is FSA.

(Refer Slide Time: 24:16)



So, let us consider one more example. The non terminals there are three non terminals now, another grammar we will consider; this is of type two. The terminal symbols are a and b. There are eight productions; the production consists of S is the start symbol, and the production symbols, production rules are like this; S goes to a B, B goes to b, B goes to b S, B goes to a B B. S goes to b A, A goes to a, A goes to a S, A goes to b A A. S is the start symbol, there are eight rules; two terminal symbols a and b, and three non terminals S capital a capital B. Try to generate a few strings in the language, what are the strings generated? Let us try to generate some strings S goes to a B rule one, then rule two you will get a b, S goes to b A rule five, rule six you can generate a B you can generate b A. Then you can also generate S goes to a B rule one, rule three if you use a b S, then rule five a b b A, rule six a b b a. You can verify that length two ab can be generated, ba can be generated, length four ab ab can be generated, ab bacan be generated dba ba can be



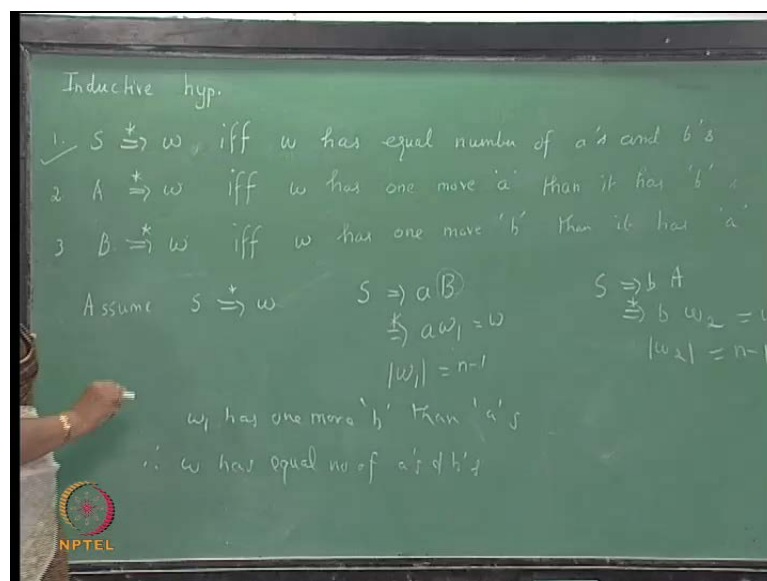
generated bb aa can be generated, aa bb can be generated, aa bb bb aa ab baba ab, ba ab also can be generated, baba ab ab.

In general, you will find that you cannot derive strings of odd length, you will be able to derive only strings of even length from S, and also that will have equal number of, any string generated in this grammar will have equal number of a(s) and b(s). It is not very obvious, see looking at the rules that is not very obvious, is not it? So, we will have to prove that **right we have to prove.**

So in general, what your purpose will be given a new language, you want to design a new language; programming language and construct the compiler means, first step will be to generate the grammar. You are framing the grammar, but then you must show that this grammar generates all syntactically correct programmes in the programming language you are designing, and all syntactically correct programmes can be generated by the grammar, that is the grammar generates all syntactically correct programmes in the language, and only those not anything else.

So, that way you have to prove, then only you can proceed for constructing the compiler. So, this is for example, here we know that the language generated has any string which has got equal number of a(s) and b(s), that is not very obvious from the grammar, isn't it?

(Refer Slide Time: 29:11)



So, you have to prove that, how do you prove? We will prove by induction, so the

inductive hypothesis is this. If a string is derivable from  $S$  in this grammar, then or if and only if  $w$  has equal number of  $a(s)$  and  $b(s)$ , but this is the one we want to prove, but in order to prove that we need to prove something more also.

String is derivable from the non terminal  $a$ , if and only if  $w$  has one more  $a$  than it has  $b(s)$ . Then another step would be a string is derivable from  $B$ ; a terminal string,  $w$  is a terminal string. If and only if  $w$  has one more  $b$ , than it has  $a(s)$ , so this is the inductive hypothesis. So, taking the basis class induction is on the length of the string, you must note that there are two things, if every statement has if and only if, so you have to prove in both directions.

So, for induction bases we can note that no string of length zero epsilon, empty string cannot be derived from any non terminal here,  $a$  or  $S$ . When the length is one from  $a$  you can derive  $a$ , it has one  $a$  and zero  $b(s)$ , so the induction hypothesis is satisfied. From  $b$  you can derive  $B$  which has got one more  $b$  than it has  $a(s)$ , so the induction hypothesis is satisfied. From  $S$  you have seen that we can derive  $a$  or  $b$ , that is a smallest string you can derive from  $S$ , and that has one  $a$  and one  $b$ , so it has got equal number of  $a(s)$  and  $b$ 's.

So, the bases class for  $n$  is equal to 1,  $n$  is equal to 2 you have checked, so the basis class holds. Then the induction class, we have to prove the induction portion, we will just prove this first statement alone, the second and the third can be proved in a similar manner, just a matter of details only. So, let us take the first one only, we will assume that the inductive hypothesis is true up to  $n$  minus 1. So, assume up to  $n$  minus 1, **upto minus 1** this  $n$  is the length of the string, **length of the string** induction is on the length of the string.

Assume up to  $n$  minus 1, that assumption is all the three you are assuming up to  $n$  minus 1, prove for  $n$ . So, if you have a string  $w$  belongs to  $t^*$  and the length of  $w$  is  $n$ , you have to prove that  $S$  derives  $w$  if and only if  $w$  has equal number of  $a(s)$  and  $b(s)$ . So, there are two parts, if  $S$  derives  $w$  then  $w$  has equal number of  $a(s)$  and  $b(s)$ . And the other way round,  $w$  has equal number of  $a(s)$  and  $b(s)$  then it is possible to derive  $w$  from  $S$ , you have to prove two ways on two directions.

If it is possible to derive  $w$  from  $S$  then  $w$  has equal number of  $a(s)$  and  $b(s)$ , and if you have a string  $w$  which is having equal number of  $a(s)$  and  $b(s)$ , it should be possible to derive that from  $w$ , both ways you have to prove. Let us see how to go about this? First, assume from  $S$   $w$  is derivable. So, what you do is, from  $S$  if you start from  $S$  you

have to use rule number 1 or 5.

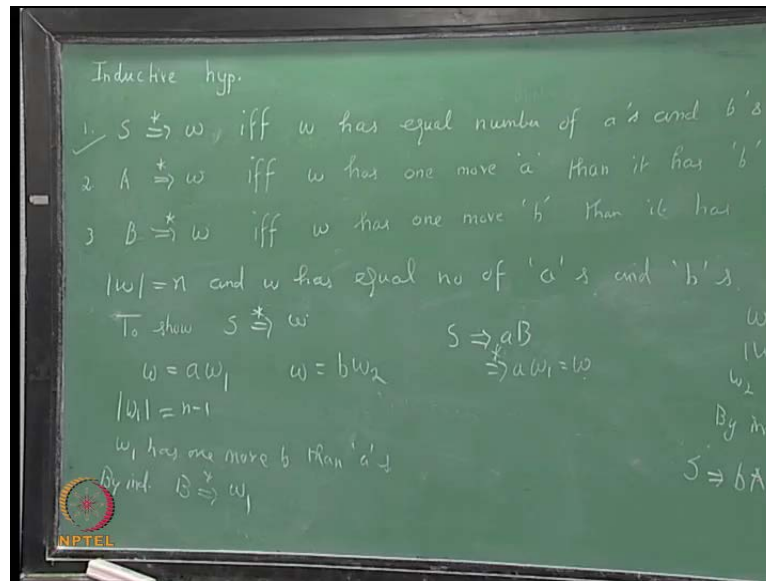
If we use rule number one the first step will be like this, if we use rule number five the first step will be like this. Then from  $b$  you are deriving a string  $w_1$  in many steps, and you are getting a string  $a w_1$  which is equal to  $w$ . So,  $w_1$  is derived from  $b$ , and what you can say about the length of  $w_1$ ? Length of  $w$  is  $n$ , so what can  $a$  be taken away, so what can you say about the length of  $w_1$ ?  $n - 1$ .

And the induction hypothesis holds for all strings up to  $n - 1$ , that is strong induction. All strings up to  $n - 1$  it holds. So, if  $w_1 \in S_n$  length of  $w_1$  is  $n - 1$  and it is derivable from  $b$ , what can you say about this? A string  $w$  is derivable from  $b$ , if and only if it has one more  $b$  than it has  $a$ 's.

So,  $w_1$  has one more than  $a$ 's. So, along with this  $a$  what can you say about  $w$ ?  $w$  will have equal number of  $a$ 's therefore,  $w$  has equal number of  $a$ 's and  $b$ 's. This one if you take  $S$  goes to  $b$   $A$  the first role is like this, then from  $a$  you are deriving a string  $w_2$ , what can you say about the length of  $w_2$ ? This is equal to  $w$ ; length of  $w$  is  $n$ , so length of  $w_2$  is  $n - 1$  and  $w_2$  is derived from  $A$ .

So, by the induction hypothesis what can you say? A string is derivable from  $A$ , if and only if it has one more  $A$  than it has  $b$ 's, so  $w_2$  has one more than  $b$ 's. So, counting along with this  $b$   $w$  has equal number of  $a$ 's and  $b$ 's. So, this is one portion, what we have proved is this, if a string is derivable from  $S$  then it has equal number of  $a$ 's and  $b$ 's. The other way around we have to prove, if you have a string which has got equal number of  $a$ 's and  $b$ 's, it should be possible to derive that from  $S$ , that way also you have to prove.

(Refer Slide Time: 39:33)



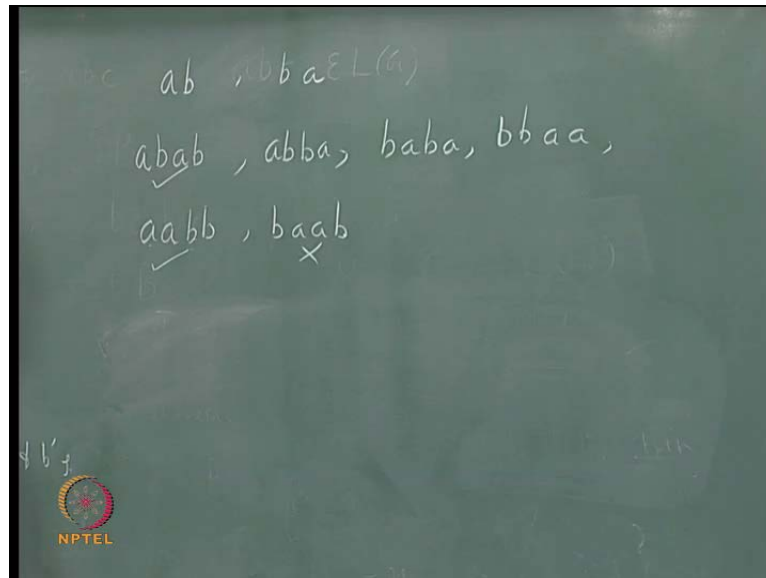
So, how do you go about proving that? Assume, that length of  $w$  is  $n$ , and  $w$  has equal number of  $a$ 's and  $b$ 's. Now, you have to show it is possible to derive  $w$  from  $S$ , but to show that it is possible to derive  $w$  from  $S$ . Now, there are two possibilities;  $w$  is a string of  $a$ 's and  $b$ 's there are two possibilities; the first symbol can be  $a$  or the first symbol can be  $b$ . So, it is possible that it is  $aw_1$  or you can write  $w$  as  $bw_2$ , and what can you say about  $w_1$ ? Length of  $w_1$  will be  $n$  minus  $1$ , and  $w_1$  has  $w$  has equal number of  $a$ 's and  $b$ 's,  $1$   $a$  is taken out. So,  $w_1$  has one more  $b$  than  $a$ 's. So, by the induction hypothesis it should be possible to derive  $w_1$  from  $B$ . So, by inductive hypothesis it should be possible to derive  $w_1$  from  $B$ , but we have a rule **we have a rule**  $S$  goes to  $aB$ , we have a first rule is this. So, apply this rule then from  $B$  you derive  $w_1$ , that is from  $S$  you can derive  $w$ . So, if  $w$  has equal number of  $a$ 's and  $b$ 's and assuming that it begins with the  $a$ , it is possible to derive  $w$  from  $S$ , that is what we have proved now.

Now, second case take  $w$  is equal to  $bw_2$ , so what can you say about the length of  $w_2$ ? The length of  $w_2$  is  $n$  minus  $1$ , and  $w_2$  will have one more  $a$  than it has  $b$ 's.  **$w_2$  has one more than  $b$ 's.** So, by the induction hypothesis what do you get? It is possible to derive  $w$  from  $A$  if it has got one more  $a$  than it has  $b$ 's. So, by induction hypothesis **by induction hypothesis** it should be possible to derive  $w_2$  from  $A$ .

So, but we know that there is a rule of the form  $S$  goes to  $bA$  rule number five,  $S$  goes to  $bA$  is there. Apply this first, then from this you can derive  $w_2$ , but  $bw_2$  is nothing but

w. So, it is possible to derive w from S. So, the language generated consists of equal number of a(s), any string having equal number of a(s) and b(s) will be generated by this grammar. What is the difference between this and the Dyck set which we considered in the last class. Dyck set is the well formed strings of parenthesis,

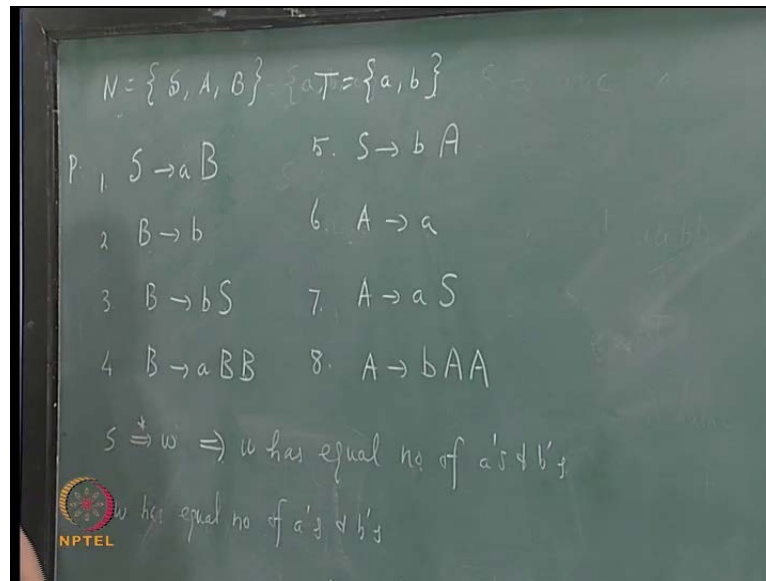
(Refer Slide Time: 44:06)



so for example, if you look at a as a left parenthesis, b as the right parenthesis, this will belong to the language, this will belong to the language, but this is this will not, because this will be right left left right like that, which is not a well formed string. So, there is a difference, here any string will have equal number of a(s) and b(s).

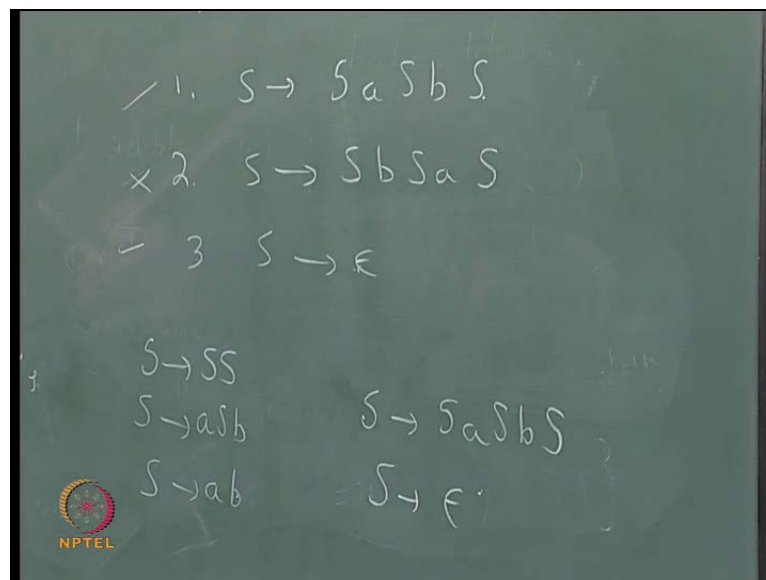
Dyck set also, any string will have equal number of a(s) and b(s), but there is some more restriction, what is that restriction? Left should come first I mean; the well formed string parenthesis means, left parenthesis should come before the matching right parenthesis. Now, one more point to note is, can this grammar generate an empty string? This grammar cannot generate the empty string.

(Refer Slide Time: 45:10)



The smallest string generated by the grammar is of length two a borb a.aalone can derive a, bcan derive b, but from S the smallest string derivable is of length two a B (())

(Refer Slide Time: 45:32)



Now, I will write another grammar, look at the rules. There is only one non terminal S, terminals a and b, the rules there are three rules S goes to S a S b S, S goes to S b S a S, S goes to epsilon, a set of reductions consists of these three. Here the empty string can be generated, the empty string can be generated here, but you can see that whenever i apply rule, 1 a and 1 b will be generated.

Whenever, I apply rule two and 1a will be generated, but ultimately any string generated will have equal number of a(s) and b(s). In between some of the s's we can make it go to epsilon. So, it is very obvious that any string generated by the grammar will have equal number of a(s) and b(s), but you have to prove the converse also. Any string having equal number of a(s) and b(s) will be generated by this grammar, that is not as easy, that is not very trivial, one direction it is very easy to see. Any string derivable in this grammar will have equal number of a(s) and b(s), but if the language consists of all strings having equal number of a(s) and b(s) you must also prove that, any string having equal number of a(s) and b(s) will be generated by this grammar, you can try this that is not very trivial step, you have to prove that.

Now, the dyck set the grammar which we gave was this,  $S \rightarrow S S$ ,  $S \rightarrow a S b$ ,  $S \rightarrow \epsilon$ . Here again note that epsilon cannot be generated, the empty string cannot be generated. An equivalent grammar which will also generate the empty string is this, that is in this grammar leave out this have these two rules alone. This again will generate only well formed strings of parenthesis, but in this case epsilon is generated, in this case epsilon is not generated. Similarly, these two generate the same language apart from epsilon, here epsilon will be generated, epsilon will not be generated in the other one. So, these are some of the examples which we consider for type two and type three, and type one also we have considered one more examples. In the next lecture we will see some more example and also about ambiguity. By the way, you can try to prove that the other converse of it, that is any string having equal number of a(s) and b, a(s) and b(s) will be generated by this grammar.