Natural Language Processing

Prof. Pushpak Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Lecture - 39
Parsing Ambiguous Sentences; Probablistic Parsing

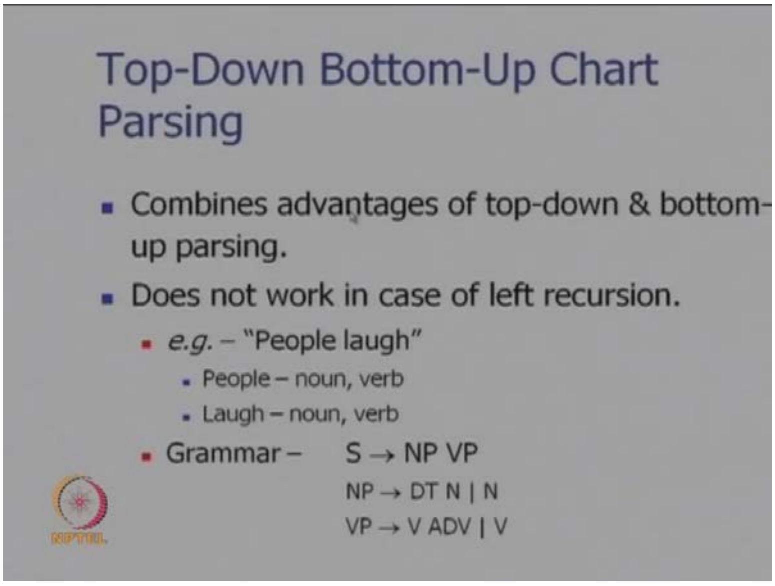
In this lecture we will complete deterministic parsing and start probabilistic parsing one question that met arise in our mind is why should be worry about probabilistic parsing? Why bring in probability? What is probabilistic about parsing? The answer to that question is that natural language sentences are ambiguous either very nature. There is ambiguity with respect 2 words can a multiple meanings, and only one meaning is applicable in a particular context generally. And there is also structural ambiguity where phrases, clauses can have multiple attachment points.

A very famous example is the example of I saw the boy with a telescope, with a telescope is a preposition phrase which can get attached to boy willing their by the boy has the telescope or it can get attached to see willing thereby that I had the telescope, I saw the boy with a telescope. Now, the point is that this attachment of the preposition phrase is a structural question, in which part of the tree does it exist what is its note. So, sentence as the some in substances that sentences can have multiple parses, so this is the ambiguity issue. And now we have to choose the best possible parse tree as the dictate of the context or the discourse ((Refer Time: 02:06)).

So, now the question is when we have multiple parses and we have to choose one on what basis do we chose a parse tree. So, here comes the notion of a scoring function you have to score the parse trees and choose the one which is the highest code, this code is often it terms of probability values. The probabilistic framework provides a systematic platform by which in this kind of codes can be computed. And probabilistic parsing takes its birth in this notion that ambiguous sentences are a reality, choosing amongst ambiguous sentences each also reality.

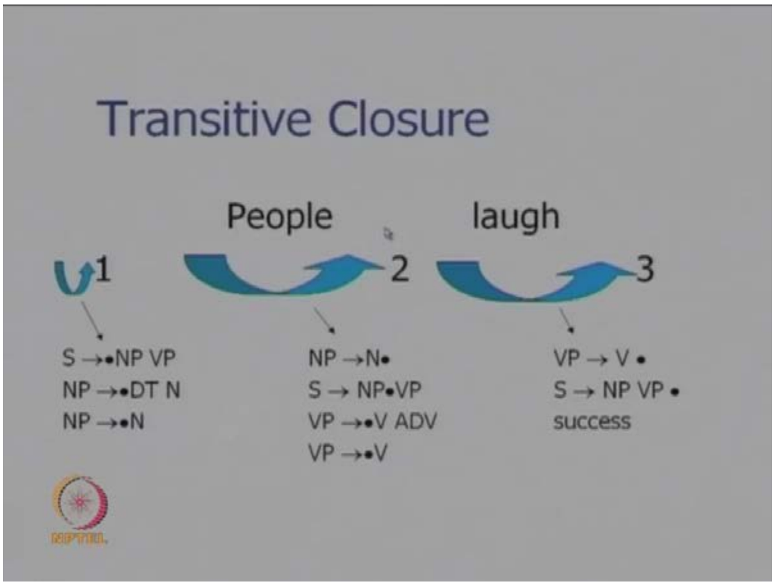
How to a make this choice is a matter of scoring those trees and probability provides a principal way of arriving at this course. Now, we take the first step of discussing how to parse an ambiguous sentence? The 3 algorithms we discussed namely top down bottom up and top down bottom up chart parsing can they produce multiple parse tree when structural ambiguity exists and how do they appear. So, parsing ambiguous sentences and probabilistic parsing is the topic.

(Refer Slide Time: 03:32)



So, just to recapitulates the top down bottom up parsing.

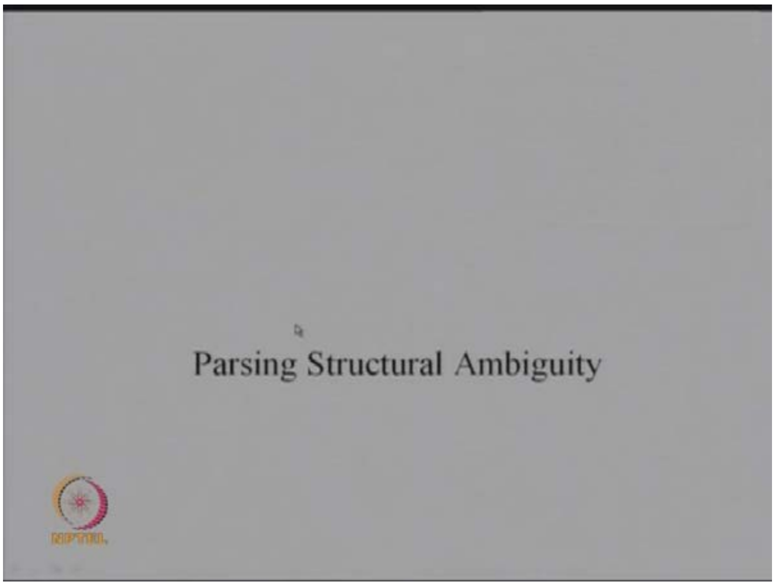
(Refer Slide Time: 03:35)



We say that when we have a sentence like people laugh loudly we first have the transitive closure of the non-terminal in front of dot. So, non-terminal expanded to D T N and N thereby exposing D T and N. So, since we have to look for a noun or a D T, we find the noun here in people and that finishes a noun phrase when the noun phrases over then V P gets exposed V P as verb requires finding about. Now, verb as in found in loudly so that adverb gets exposed and now a need to find the adverb loudly is that adverb.

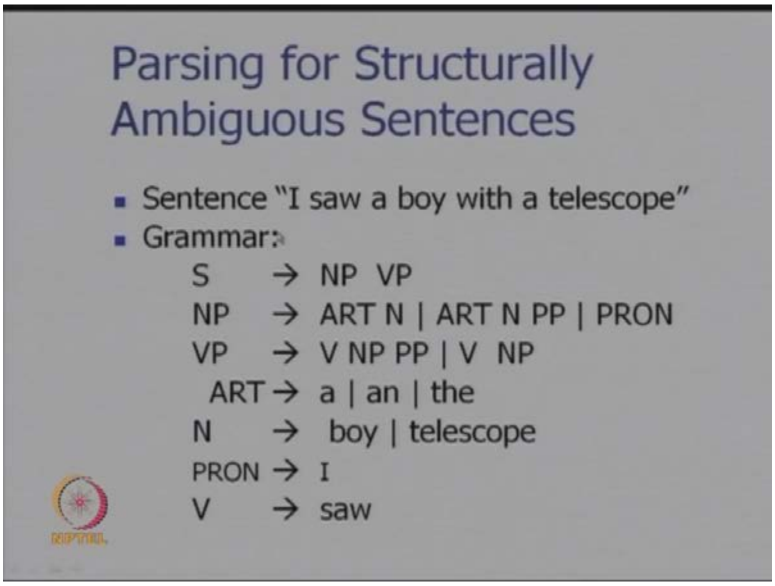
Now the verb phrase is over and the verb phrase for the sentence is also over. Now as the grammar goes since a V P is over here we could also finish the sentence through noun phrase and verb phrase at this step. And indeed this is the case because people laugh is also a value sentence, but we cannot signal the end of parsing or success of parsing. Because there is still a word left and this word next to be accounted for in the parsing process.

(Refer Slide Time: 05:02)



So, now we come to this point about parsing structural ambiguity.

(Refer Slide Time: 05:05)



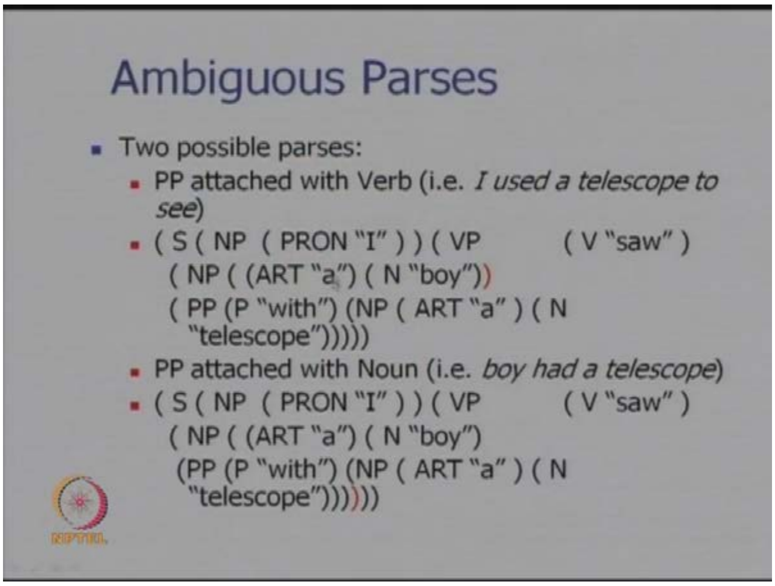
If the sentences I saw a boy with a telescope, the grammar is as goes to N P V P as usual N P can be article and noun or article noun and preposition phrase or simply a pronoun. Verb phrase also can be verb noun phrase and preposition phrase or verb and noun phrase. Article can be a, and the noun can be boy and telescope our vocabularies limited here only 2 nouns boy and telescope. Pronoun is I, verb is seen. Now, let us look carefully at these rules. So, these first the second role for example, says that we have a wide variation possible in constructive noun phrases.

They can be simply article noun combination or they can be article noun combination where the noun itself is modified by preposition phrase. Preposition phrase becomes a qualify for the noun or the noun phrase itself can be a pronoun. Similarly, we have variety possible on the verb phrase so we have verb and noun phrase and the preposition phrase. The preposition phrase can now they can now get attached to the verb as is shown in this combination.

And there by signaling the fact that the verb phrase has a preposition phrase embedded which is a consequent. When here be noun phrase is simply the depiction of the case where the verb is a transitive verb with taking an objective and so on. So, a let us see if you understand this particular point that these 2 rules namely article going to article noun and P P and verb phrase going to verb noun phrase and P P is the potential ground for multiple parse trees or structural ambiguity.

The preposition phrase can be attached to noun or preposition phrase can be attached to verb and the grammatical rules these 2 grammar rules capture that phenomenon. Now it must be clear to you as I explain this that ambiguity is unavoidable in natural language. And sentences have a different structures a particular sentence can have multiple structures and when these multiple structures obtain define grammatical rules come into play. And a close examination of the grammatical rules capturing the language or language fragment is sufficient to tell us that that is potential ground for ambiguity. And when multiple sentences multiple parse tree come out from a sentence, what we do? We have to choose 1 and we choose this 1 by a scoring mechanism through probability.

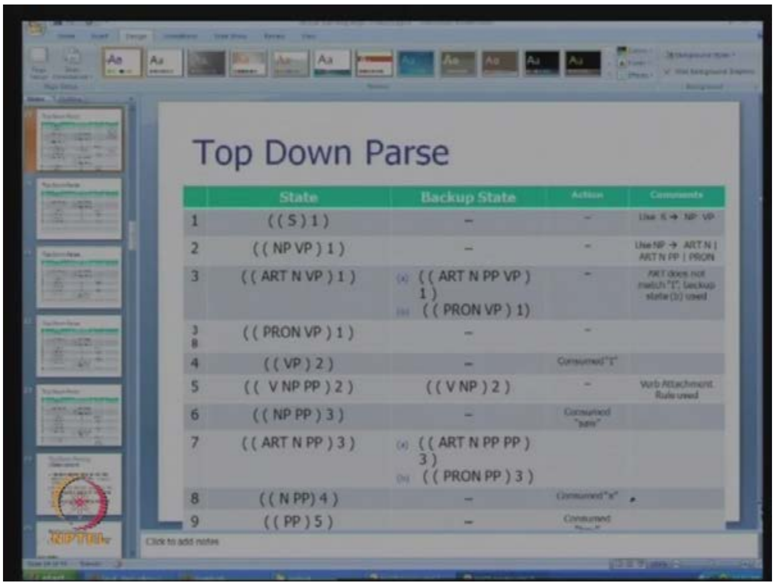
(Refer Slide Time: 08:18)



Looking at the slides 2 possible parses for the sentences I saw the boy with a telescope is at this P P is attached with verb that is I used a telescope to see. So, here is where to be seen in the form of a bracketed structure. Now, S is the starting symbol so here we see that there are this S is composed of a noun phrase which finishes here. And the verb phrase which is more complex or this particular verb phrase as the main verbs saw a boy with a telescope. Now, we has this question where is with a telescope with a telescope is a preposition phrase no doubt. But what does it attached to, we see that these preposition phrases is in the scope of this noun phrase.

The preposition phrase is in the scope of the verb phrase and therefore, this attaches to the verb and I have the telescope. In the other case, the preposition phrase is attached to the boy, because it is contained within the noun phrase not in the verb phrase. It is contain in the verb phrase, but more so within the noun phrase because its immediate parent is the noun phrase. So, this is the attachment to boy there are 2 ambiguous parses. By the way this kind of bracketed structure is another alternative representation for the parse tree where use to seeing the tree. The tree has the non-terminal than then expanded non-terminals than terminal so on for the same time think, and the express by creating bracketed structure. A particular enclosing of bracket finishes a phrase.

(Refer Slide Time: 10:43)

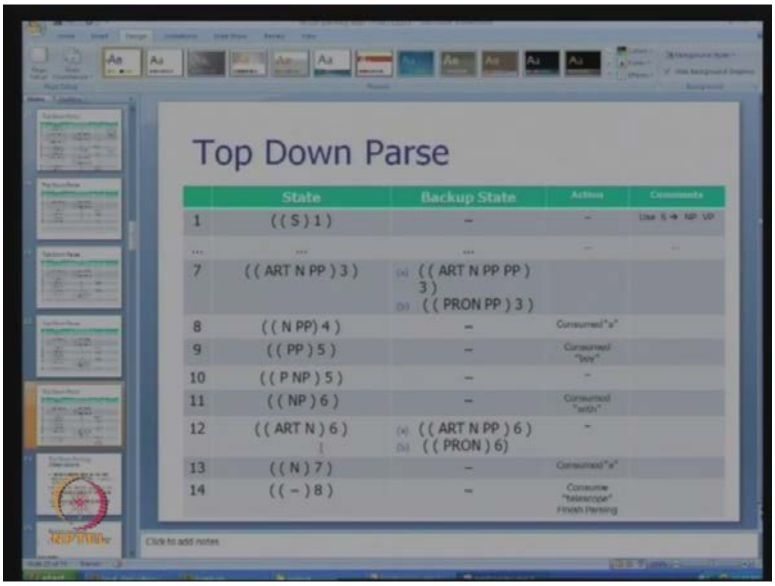


Now, we take this ambiguous sentence and round the top-down parsing algorithm. We first have the starting state input point the, that 1 no backup state S is expanded by N P and V P no backup state N P V P. We use there is no backup state because s can be expanded on the in 1 way. Now, we have to use N P goes to article noun or article noun preposition phrase or pronoun and we have to take up textually first role which is article and noun. And there are backup states which article noun preposition phrase and pronoun. So, the backup states are article noun preposition phrase verb phrase and pronoun verb phrase with input pointer at 1. So, these states are in the waiting in case these state fails the alternative states will be after the stack.

So this of course, fails because article does not match anything, we do not also bring the state A. Because again it is fronted by A R T which will not able to be match anything on the input data. Now, pronoun V P is brought in. So, I is consumed input pattern moves to

2. Now V P is expanded by V N P P P, V N P P P and V N P is the rule in vatic. Now the verb is consumed so that brings N P P P exposed on the stack. Now, N P has to be found out N P is expanded by article noun preposition phrase article noun preposition phrase N P is expanded by article and noun article noun preposition phrase is in the waiting. Similarly, pronoun is in the waiting, now the word a on the input is consumed by article that makes N N P P exist on the stack input pointer is move to position number 4. At this point noun is used to consume boy when boy is consumed. Then we find that P P and 5 exist on the stack P P and 5 exist on the stack by and the before that step boy was consumed now we have P and N P here.

(Refer Slide Time: 13:54)



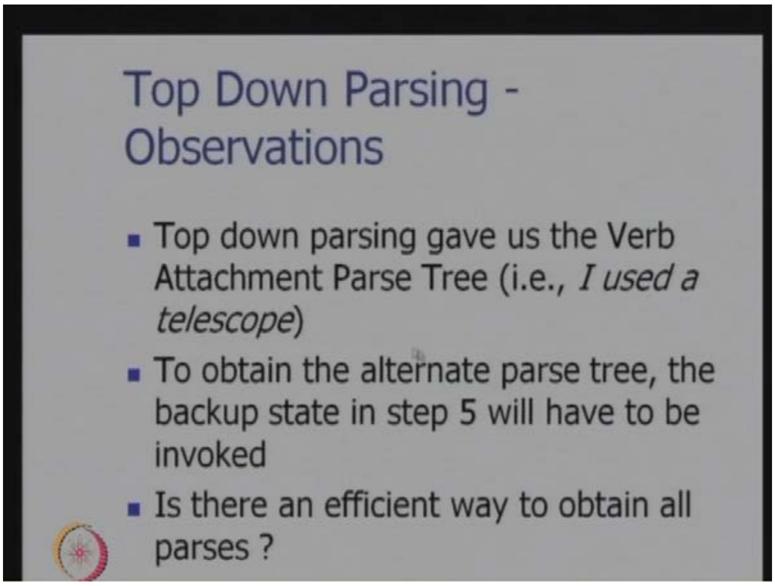
Now, continuing in this way where the first step as 1 and the sixth step had exposed N N P. Now, we have article N N P P and the input pointer is A 3 backup states are article N P P with input pointer A 3 pronoun P P at input pointer A 3 again we consumed A that brings P P that brings N P P with 4 at the input point of position. Now, P P is exposed P N P comes in N P comes in with is consumed similarly, all the N P is are developed. And then finally, telescope is consumed and there is no symbol on the stack the input data is over and that finishes parsing.

So, what is seen here is that we can multiple parses if we do not stop at a particular stage of processing where all the input symbols have been consumed and the start symbol also have been popped from the stack. So, this is a particular way of parsing which attaches telescope to the boy. Let us because in the sixth step we had V N P P P ,V N P and P P and N P P P K N P and P P K and V N P P P path was perceived this V N P P P path was perceived. And that finally, let to telescope being consumed at the final step and S also getting path from the stack.

So this is the path which is perceived with V N P P P. Now if we take the other path which is with the state waiting in the backup stack V N P and 2. And follow similar process of expansion will get the parse tree where the preposition phrase with the telescope is attached to noun. We can see as step 5 says verb attachment rule is used, and we are followed this path completely to get the parse tree finally. Now, this is the example this is the example where 1 of the parse trees has been produced 1 of the parse trees is produced. So, this shows how the stack can the stack base top down parsing can handle ambiguity by following 1 particular state on the stack up to completion.

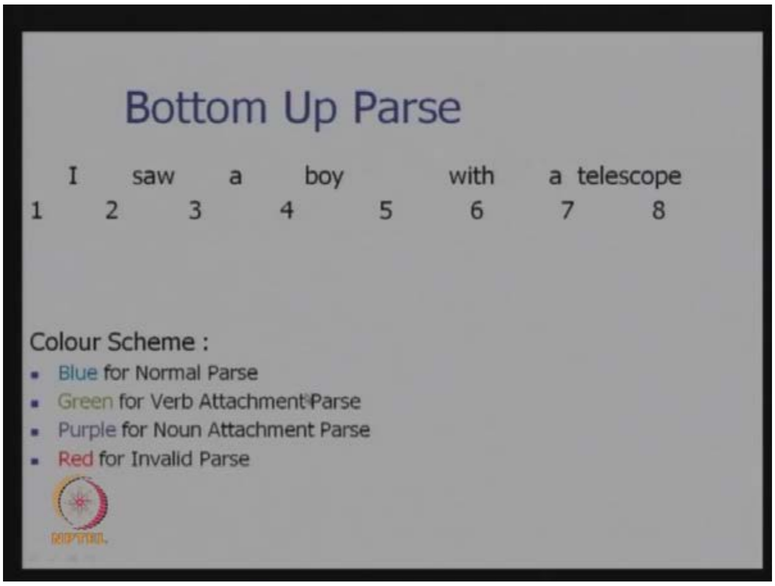
And then revisiting the backup states and seeing if the parsing can be completed from these backup states so that would produce the other parse trees. So, I suppose this particular point is clear about how ambiguity can be handled in parsing in top down parsing. The possibility of pursuing a different path with a different state always exists in top down parsing the stack contains that information finish the parse. Then revisit the stack see what kind of states exist obtain their at follow those to finally, finish the parsing process. And if there are multiple parses possible from that sentence they will come out by this procedure.

(Refer Slide Time: 18:38)



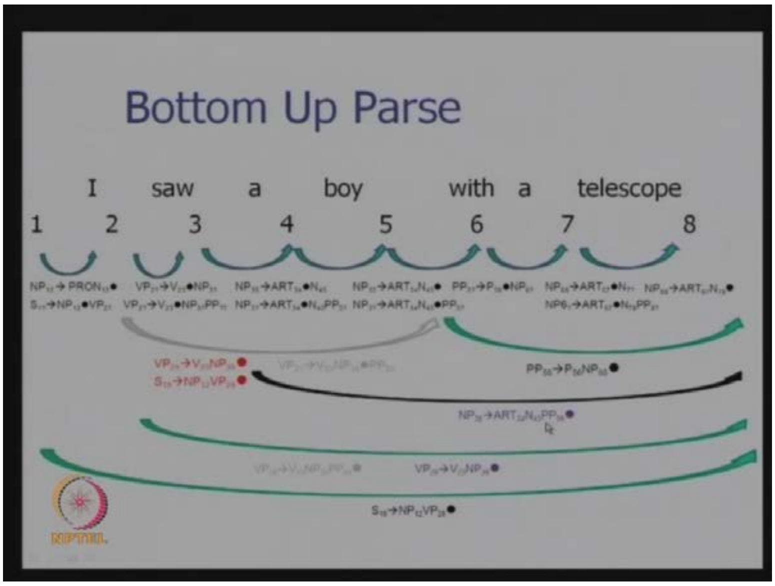
Now, we move to some observations which is that top-down parsing gave us the verb attachment parse tree. I used a telescope to obtain the alternate parse tree, the backup state in step 5 will have to be invoked is there an efficient way to obtain all the parses that is an interesting question to think about.

(Refer Slide Time: 18:59)



Now, how does bottom up parsing deal with ambiguous sentence and does it have the capability to produce multiple parse trees? This is the question we now parse you with the same example I saw a boy with a telescope note the between 1 positions 1 before I 2 between I and saw 3 between saw and I and so on. And finally 8, which should have been after telescope to indicate the end of sentence. So, since they are all ambiguous situations and multiple parse trees are possible. We follow this color convention blue for normal parse, green for verb attachment parse, and purple for noun attachment parse, red for invalid parse.

(Refer Slide Time: 20:00)



So, when we have I saw a boy with a telescope the first stage produces noun phrase and the pronoun. So, first we see that between 1 and 2 way have resolved the pronoun and N P is resolved a pronoun the dot moves were pronoun. In S where the dot as gone over N P now V P as to be found out the suffixes have to be noted N P 1 2 and front 1 2 says that between 1 and 2 we found a pronoun and a noun phrase S 1 question says that there is a sentence beginning at 1 ending at where we do not know N P 1 2 is find where found a phrase between 1 and 2, V P begins at 2 ends at where we do not know.

In the next item set, we have resolved the verb with saw and thereby V P has completed at the work of finding A V. Now, at N P has to found starting from position number 3. Now ambiguity situations corpus 2 rules are applicable in this situation. So, we can also have the verb phrase with we resolved and N P and P P to be found when we go to the next word which is A. And article has been found thereby the noun phrase has completed the work of article. And noun has to be found or a noun phrase as completed the work of noun and the noun and preposition phrase has to be found; now boy is a resolved.

So, a noun has been found between 4 and 5 that resolve a noun phrase between 3 and 5 a boy. And it also resolves partially another noun phrase which is article and noun, but we have to find a preposition phrase after that. Now, what we sees that since we have found a noun phrase with article and noun we can we can complete this verb phrase here. So, that is why the noun phrase the dot loops over this noun phrase, and a noun phrase has been found between 3 and 5, 3 and 5 is a boy and thereby of verb phrase is resolved between 2 and 5 saw a boy. And thereby a sentence is resolved between 1 and 5 I saw a boy and it is indeed true. Because I saw a boy is a complete sentence statement or a complete sentence actually within the whole sentence. But at the same time since the noun phrase is has been found we can move the dot over.

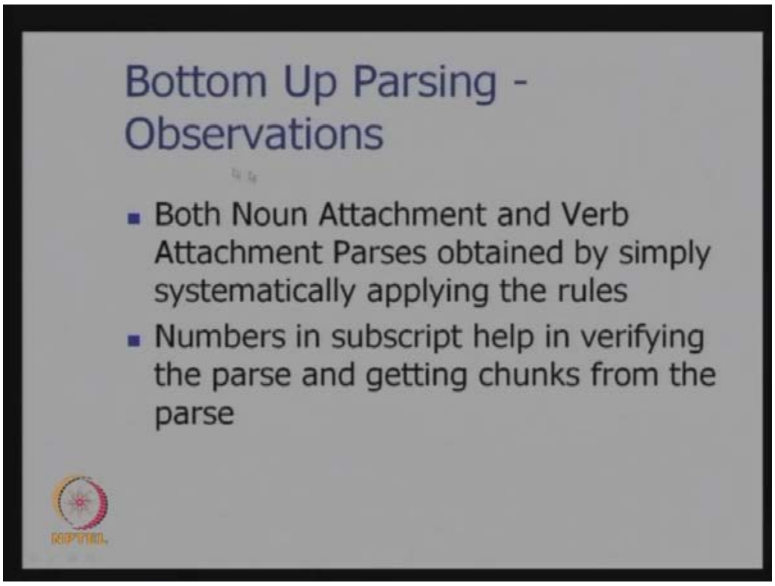
The noun phrase in this production group and now it exposes P P so we have to found V N P which is saw a boy. Now, we have to find a preposition phrase now here though a sentence has been found, we cannot finish the paths because there are still words remaining in the input sentence. Proceeding further we have now found a preposition with so a preposition between 5 and 6 is found P 5 6 we are waiting to find a noun phrase to complete the preposition phrase starting at position 5. And these we have found partially, because now to a found an a and an article between 6 and 7 is found so that now once a noun to be found at position 7.

And a noun phrase with article and noun or a noun phrase with article and noun and preposition phrase is waiting to be completed. At this point we find telescope which is point 10 this noun phrase from 6 to 8 a telescope is completely done a is completely done S. And that will finish the preposition phrase with a telescope between the position 5 and 8 with a telescope is a complete preposition phrase. So, that will complete this part verb phrase here this or other this noun phrase here a boy with a telescope 3 to 8 and when this is done this verb phrase between 2 and 8 will be finished.

But it will finish in 2 ways; one in a verb between 2 and 3 and a noun phrase between 3 and 8 or a verb phrase which is composed verb between 2 and 3 noun phrases between 3 and 5 a boy and a preposition phrase from 5 to 8. Now, the sentence gets over with noun phrase from 1 to 2 and verb phrase from 2 to 8 how about the verb phrase from 2 to 8 is completed in 2 different ways either by P P attachment to be verb or by P P attachment to the noun a boy this particular situation. So, this shows that bottom of parsing also can handle ambiguity by maintaining within its item set or the chart all the applicable rules.

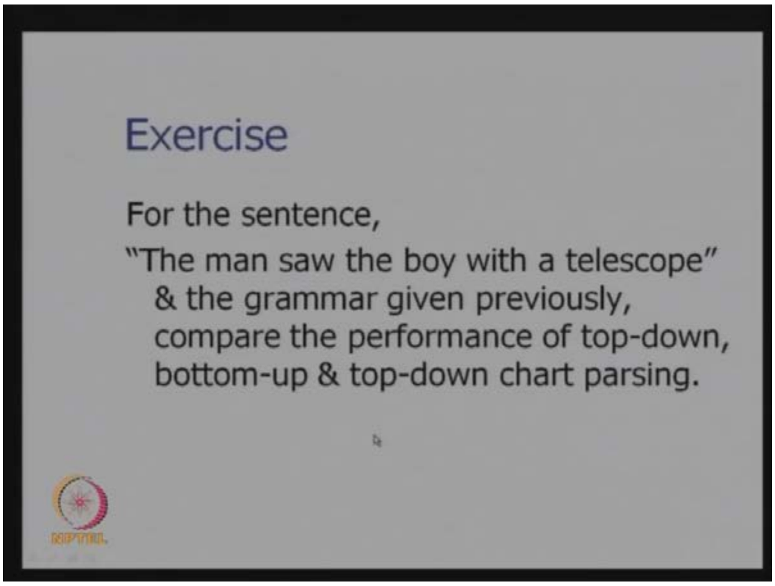
So, these bigger arts are indicators of larger parses and in larger phrases and in each phrase we maintain the indications of ambiguity. For example, this phrase just to possible structures verb noun phrase preposition phrase or verb. And noun phrase where the noun phrase itself is a noun and preposition phrase preposition phrase getting attached to the noun.

(Refer Slide Time: 27:44)



Now, we move forward and there is some observations in bottom up parsing both noun attachment and verb attachment parses obtained by simply systematically applying the rules. So, this we have seen in the past processing itself there are the indications of the rules which are applicable. Numbers in subscript help in verifying the parse and getting the chunks from the parse.

(Refer Slide Time: 28:11)



So, I suggest this exercise to you the man saw the boy with a telescope is the input sentence and grammar is given previously. We first compare the performance of top down and bottom up parsing on this ambiguous sentence. Comparison means were discuss what is which strategy maintains more cleanly the ambiguity indications. How is the ambiguity indications maintained in top down parsing? This is maintained very definitely and cleanly by backup states on the stack.

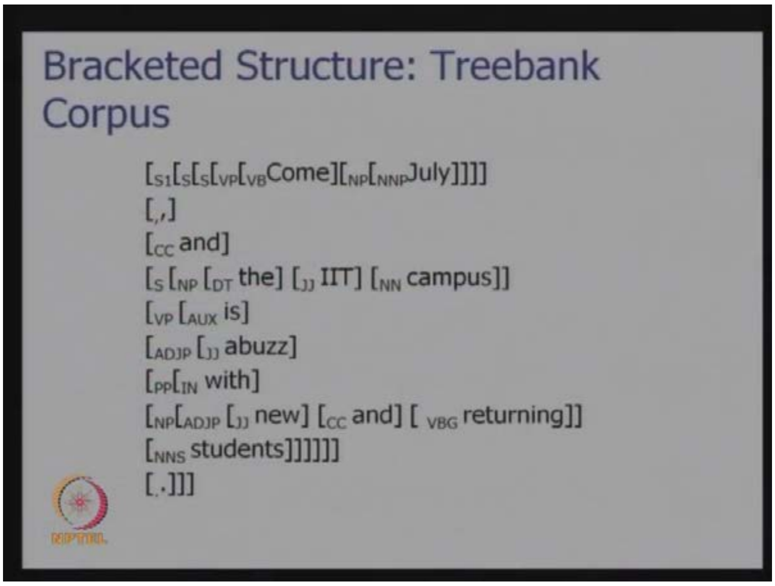
How is ambiguity indications maintained in the bottom up parsing? It is maintained by multiple items in the item sets or the charts on each art between 2 words. So, this structure which maintains the possible parse tree completion paths the potential trees which can come out of the grammatical rules. So, both of them maintain the ambiguity indications in their own way and after that the processing can reveal the ambiguity.

Now, in top-down parsing, a revisiting of the stack for all the backup states is necessary to obtain the ambiguities where as in bottom up parsing the whole ambiguity situation is where top both the trees and all the trees. Because of ambiguity which can come from the sentence and maintained in the charts so an amount of rereading of the charts is required to obtain the trees. But one will not have to expand the rules again or resolved the data or the words with the non-terminals to obtain the parse trees. In the top-down parsing re generation of trees are required.

So, one can say that in top down parse the development of the tree has to be done the development of the tree other than one obtain has to be done. In bottom up parsing the trees are already developed they there in the charts they only have to be retreat. In top down bottom up parsing it will be an interesting question to investigate. How ambiguity is handled and multiple parses are produced?

So with these we close our discussion on deterministic parsing we have seen top down, bottom up and top down bottom up chart parsing we have seen also ambiguous sentences are handle all this is find. Now, an important deterministic algorithm here is the C Y K algorithm. C Y K algorithm is a dynamic program being waste parsing algorithm which finds its usage in probabilistic parsing also. This we did not do here it will be done along with probabilistic parsing we begin probabilistic parsing now.

(Refer Slide Time: 32:02)



So, here is a very useful resource called the bracketed structure and these bracketed structures are the tree is actually for the sentences. And a collection of such bracketed structures found what is call the tree bank corpus if you look at this sentence come July and the I I T campus is abuzz with new and returning students. This is a complex sentence, because it has a clause embedment come July and the I I T campus is abuzz with new and returning students. Now the I I T campus is a buzz with new and returning students is a complete parse tree and the bracketing shows the inherent structure. So, let us go over the slowly the is a determiner it is resolved this way and the bracketing shows

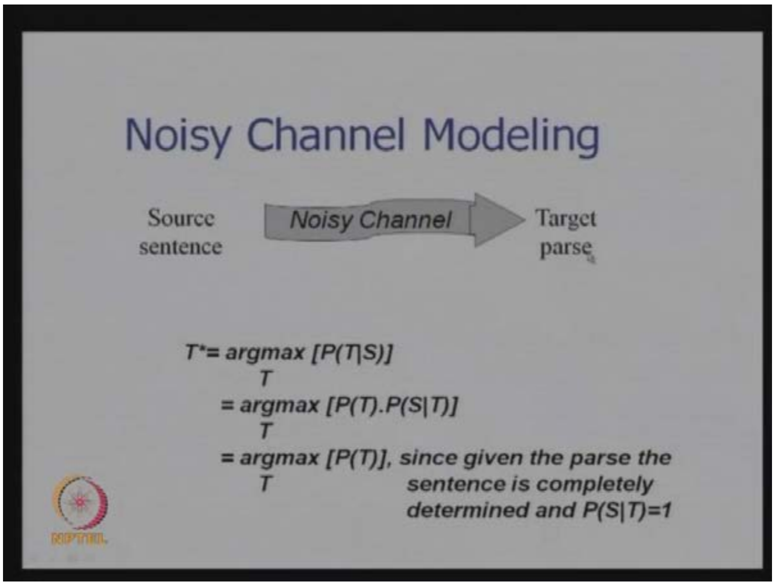
D T goes to the I I T has an adjectival role here. So, it is an adjective for the campus which is now. The I I T campus can be combined together to form a noun phrase which is shown here N P.

Now, a when N P is over it only requires a verb phrase V P to complete the sentence. Let us see is an auxiliary, there is an adjective phrase abuzz with new and returning students. So, abuzz with new and returning students is a complete adjective phrase. So, these adjective phrases along with the auxiliary verb which comes before the adjective phrase complete the verb phrase and verb phrase and noun phrase together completes sentence. Let us take a look into the structure of the adjective phrase itself adjective phrase is already an adjective has the head which is abuzz. And abuzz is taking preposition phrase after this which as with as the preposition the tag in is used for preposition in the tree bank corpus. Now, this adjective new and returning is a complex adjective phrase with 2 adjectives new and returning linked with at returning is a Gina dial verb as indicated by V B G.

But it actually is an adjectival rule that is why it can combine with an adjective through a conjunction C C and produces the adjective. Now, this adjective along with adjective phrase along with student which is a plural noun forms a noun phrase with and the noun phrase that is preposition noun phrase forms a preposition phrase which when combine with adjective phrase which when combined with an adjective produces the adjective phrase. And this adjective phrase combined with auxiliary verb produces the verb phrase. So, auxiliary verb followed by a noun phrase is a followed by an adjective phrase or adjective is a very classic case of a verb phrase. So, weather is hot the day is cold john is well my sister is studious these are examples of cases where auxiliary. And predicative adjective together forms a verb phrase it s a very common construction in English language.

So, this bracketed structure is nothing but at the tree parse tree itself return in a linear fashion. Now, we are discussing probabilistic parsing in probabilistic parsing what we are interested in is the probability of parse tree in probabilistic parsing. The entity of interest is the probability of the parse tree and as in other branches of statistical natural language processing.

(Refer Slide Time: 36:45)

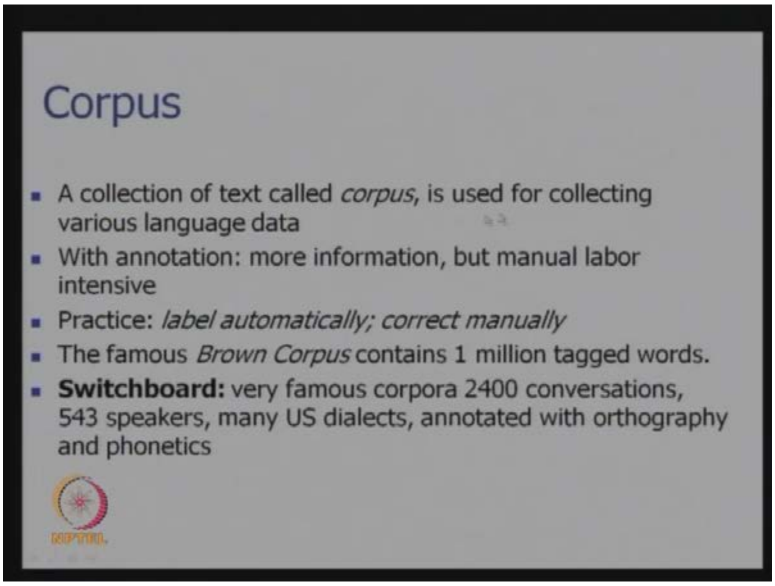


We make use of what is call the noisy channel model which is shown in the slide. The source sentence passes through a noisy channel and produces the target parse. The production of multiple parse trees from the source sentence can be looked upon as injecting of noisy as the sentence comes over the pipe line and transforming to the parse tree. So, T star is the desired parse tree the best possible parse tree how is the best possible parse tree found by maximizing the probability of the tree given the sentence.

And that is why this argmax computation T star is nothing but argmax T P T S we applied baits theorem here. And P T S P T given as is turned into P T into P S given T so likely hood of the sentence given the tree and the probability of the tree. Now, a very interesting thing happens the likely of hood of the sentence is given the tree seizes to be uncertain. It is no more uncertain given a parse tree the sentence is uniquely determinate. We simply have to traverse the leaves left to right and the sentence is completely determined.

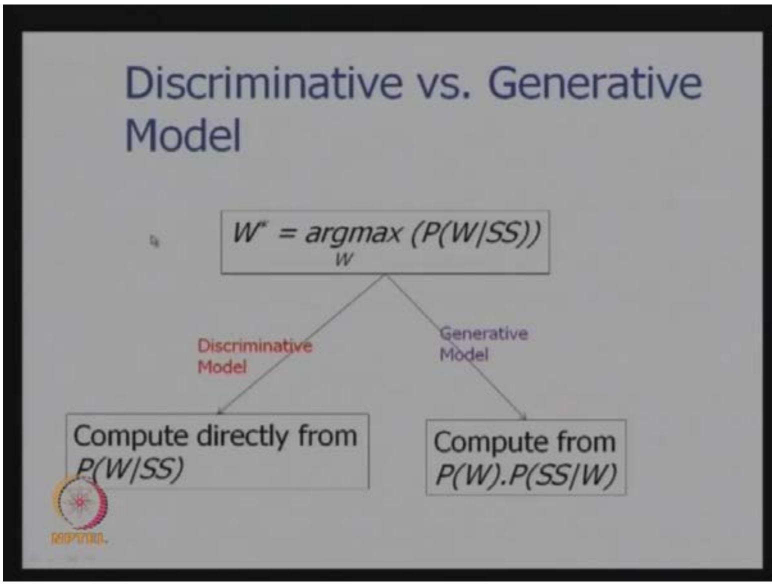
So, given the tree there is no uncertainty with respect to the sentence that is why P of S given T is 1. So, this produces an interesting expression T star the best possible parse tree is nothing but the tree with the highest probability which is intuitive so get the sentence get all it parse trees and tree which has the maximum probability is the best possible tree. Now of course, this has raise the ball in another court so far we were interested in finding an expression for the best possible parse tree that you found in terms of probability value. And the expression now demands at the probability of the tree we calculated what is the meaning of probability of a tree? We proceed to discuss this question.

(Refer Slide Time: 39:08)



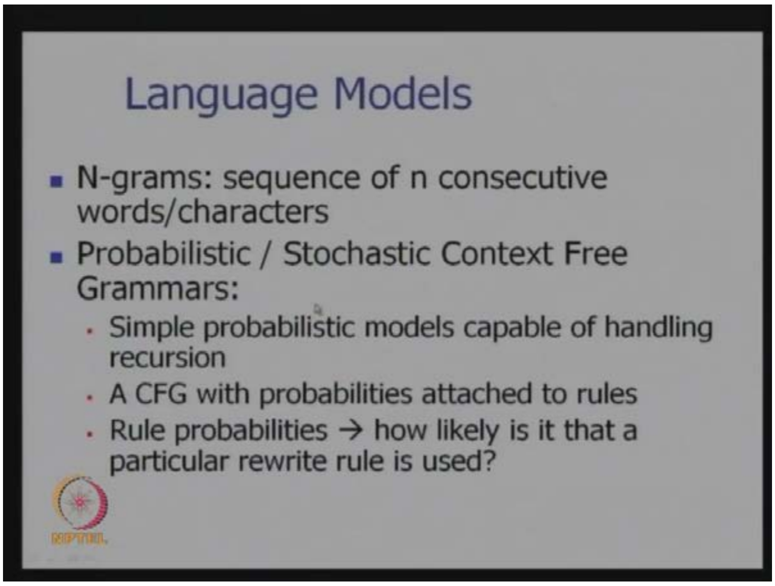
So, here is a side discussion a collection of text is called corpus it is useful for collecting various language data with annotation more information comes in. But the manual labor is intensive practice that is followed in the field typical is label automatically. But correct manually the famous brown corpus contains 1 million tagged words. A switchboard corpus is also very famous it contains 2400 conversations, 543 speakers many U S dialects annotated with orthography and phonetics.

(Refer Slide Time: 39:42)



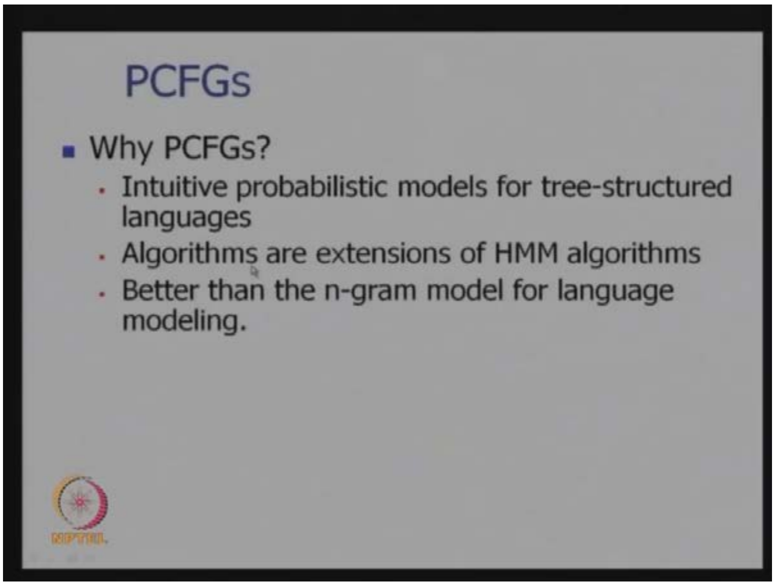
The next question which is also a really an issue on the side, but important nonetheless is the question of discriminative verses generative model are W star cab be found from probability of W given S S, where S S is some quantity W is another quantity the relationship between W and S S is that there is a noisy channel and 1 is the input other is the output. W star can be found directly from P W given S S which call the discriminative model we compute directly from P W is given S S other computing model is the generative model we compute from P W into P S S given W. Now, the debate between discriminative and generative model is not yet over the N I P community comes up with different arguments of why one approach better than the other or why one approach worse than the other so this debate is ongoing.

(Refer Slide Time: 40:52)



Language model; we have discussed already before there are N grams there are probabilistic stochastic context free grammars.

(Refer Slide Time: 40:58)

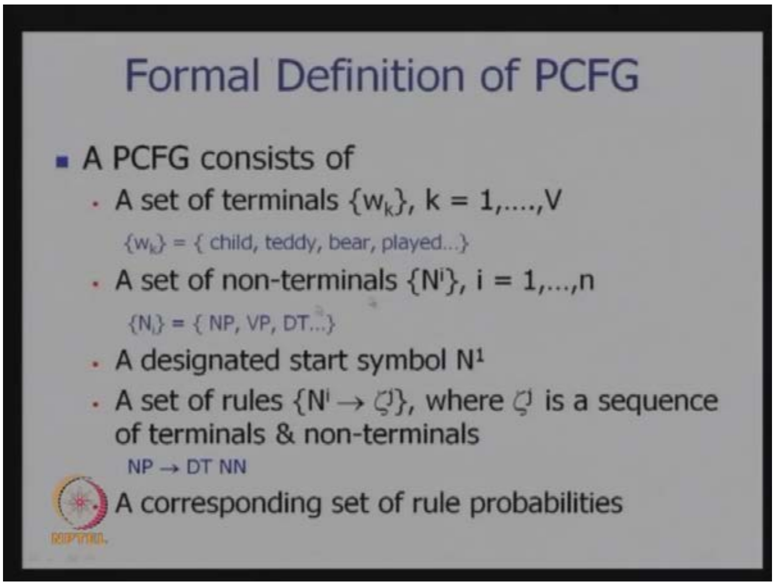


Now, context free grammars probabilistic context free grammars are very power full mechanism for capturing language regularities. Simple probabilistic models are capable of handling recursion P C F G with probabilities attached to the rules, rule probabilities imply how likely is it that a particular rewrite rule is used, we elaborate these ideas as we go ahead.

Now, why probabilistic context free grammar these are intuitive probabilistic models for tree structure languages all natural languages tree structure. Sentences and so on. Algorithms are extensions of H M M algorithms that is a very important powerful point hidden mark of model has been understood very lucidly in the probabilistic community, signal price same community, electrical engineering community. And now statically N N P community and powerful algorithms for finding the probability of the observation sequence.

Finding the probability of the automaton, learning the parameter values for the automaton these have been investigated thoroughly and interesting result have been obtained. Now, there is interesting correspondence very interested correspondence between H M M probabilistic context free grammars. We will show that algorithms in H M M translate directly to algorithms probabilistic context free grammar. So, these are the powerful ideas which can be used to produce the parse trees and their probabilistic calculations. Now, the P C F G is, are better than N gram model of language modeling. We will justify why?

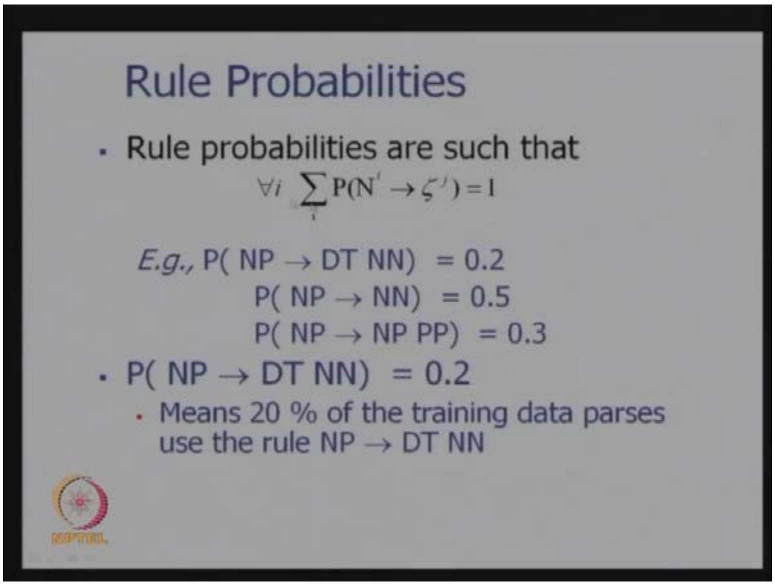
(Refer Slide Time: 42:45)



Now, first we give a formal definition of P C F G probabilistic context free grammar. A P C F G consists of a set of terminals W K K equal to 1 to V. For example, we a if our vocabulary is child, teddy, bear, played then this is the set of W K a set of non terminals N I, I equal to 1 to N N P V P D T etcetera the non terminals. A designator star symbol N

1, which is the special status in the parsing process a set of rules N I going to zeta j where zeta is a sequence of non terminals. And terminals for example, N P goes to D T N N a corresponding set of rule probabilities. So, P C F G consists of terminals, non terminals, star symbol, rules and probabilities. So, P C F G is nothing but C F G augmented with probabilities on the rules.

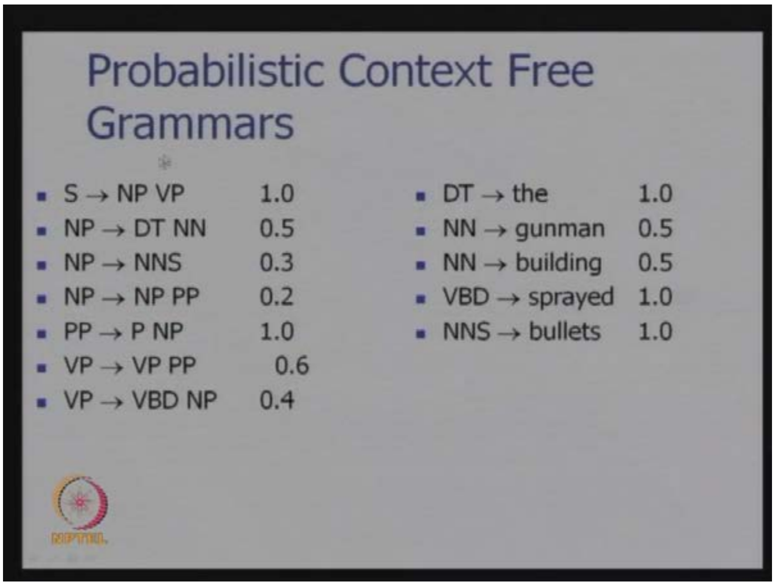
(Refer Slide Time: 43:43)



Now rule probabilities are such that sigma for all I sigma over I of P N I going to zeta j is is equal to 1, what does it mean if a non terminal expands into multiple sequences? Then the some of the probabilities of this different production situation must be equal to 1. So, for example, probability of N P going to D T N N is 0.2, probability of N P going to N N is 0.5, N P going to N P P P is 0.3 and we can see that all of these sum up to 1. Now, what is the meaning of this expression P N P going to D T N N equal to 0.2? This means that 20 percent of the training data parses use the rule N P going to D T N N. So, when we are parsing the corpora 20 percent of the times, we find that noun phrase consist of D T and N N noun phrase consists of an article and a noun. So, that is why the probability value is 0.2.

And you can possibly see that one can easily compute this probability by tracking, keeping track of the count of how many times this N P going D T N N rule was invoked? As, opposed to other instances of N P replaced by other sequences of terminals and non terminals. So, how many times have we applied N P going to D T N N as opposed to N P going to N P P P or N P going to adjective phrase N P or N P going to simply noun.

(Refer Slide Time: 45:35)



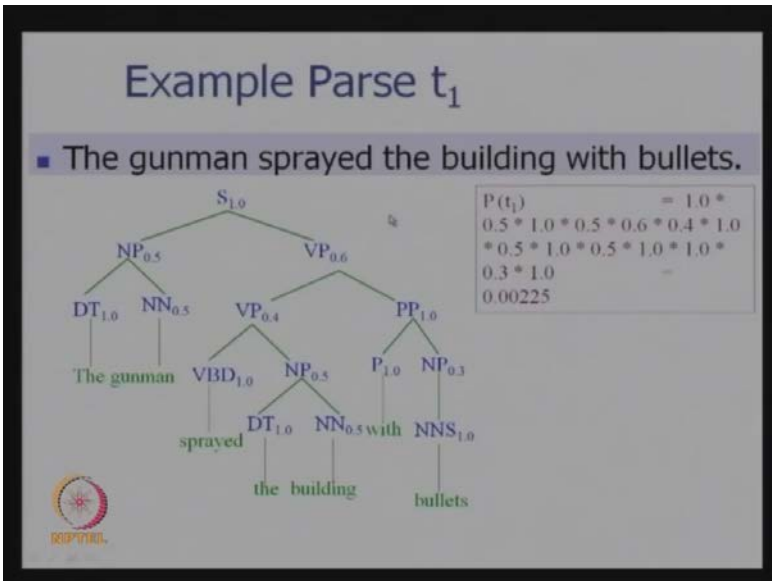
Now, here is an example of a probabilistic context free grammar rules capturing the language a fragment of English language S goes to N P V P probability is 1. 00. There is no doubt that S can be expanded only this way for declarative sentences if there are interrogative sentences then a V P will begin the sentence. So, N P goes to S goes to N P V P is 1.0 we are dealing with only declarative sentences. N P goes to D T N N as the probability 0.5 N P goes to N N is 0.3 that means N P consists of a plural noun that happens 30 percent of the times in the corpus data.

N P goes to N P and P P that means a noun phrase consist of noun phrase followed by preposition phrase that happens in 20 percent of the times. Preposition phrase is always preposition and non phrase that is why probability 1.0. Verb phrase can be verb phrase in preposition phrase 60 percent of the time verb phrase can be a verb in past tense and a noun phrase 40 percent of the time determiner is the 1.0 we are assuming there is no other article or there is no other determinant like that.

And this there is single noun which is gunman no there are 2 nouns gunman and building each with probability 0.5 that means the distribution of gunman and building in the corpora as noun is half V B D. It is a past tense verb we have only 1 past tense verb sprayed the probability is 1.0 because this is the only occurrence of the past tense verb plural noun is bullets again 1.0. So, this is the fragment of English language on the right hand side we placed what are called lexical rules.

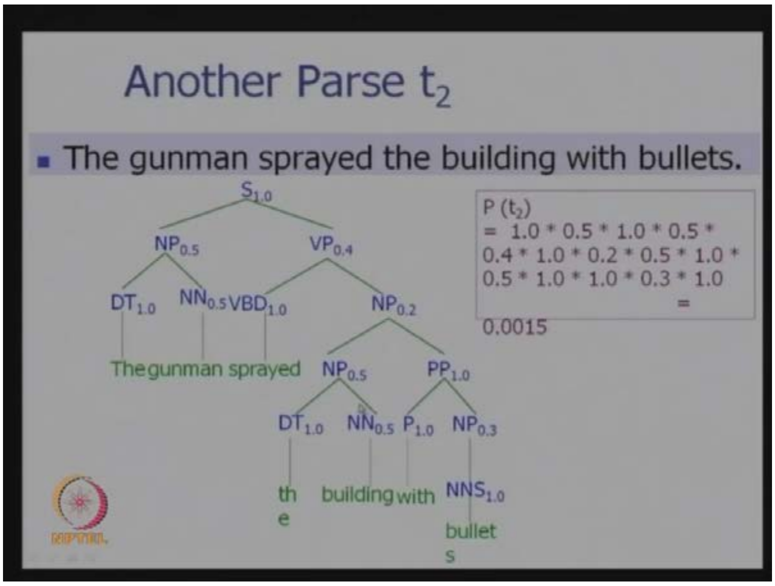
These are all rules which go to a terminal or a vocabulary of the language whereas, these are syntactical rules they capture the grammar of the sentence. For example, how is a noun phrase formed D T N N, N N S, N P V P? How is a verb phrase formed? Verb phrase in preposition phrase or past tense verb and noun phrase. So, this is capturing the syntactical rules this is capturing lexical rules.

(Refer Slide Time: 48:07)



Now the gunman sprayed the building with bullets if this sentence is taken then we can have 1 parse tree which is shown here. This is a parse tree starting with sentence s consisting noun phrase and verb phrase. In the noun phrase we have determiner noun the gunman the verb phrase itself can be a verb phrase and preposition phrase. Here the preposition phrase attached to the verb that with bullets is attached to spray. And the verb phrase is consisting of a past tense verb V B D sprayed, and a noun phrase D T N N the building. So, spray the building is the verb phrase and the building is to object for the, for spray with bullets is the preposition phrase attached to the verb so this is one parse tree.

(Refer Slide Time: 49:04)



The other parse tree is where the preposition phrase gets attached to the building as is shown here. So, N P contains the N P P P the whole thing forms a constituent of the verb phrase. So, there are two possible parse trees and they have different probability values as we will see in the next lecture. And will choose the tree which has the highest probability value. So, we compute some interesting probabilities and proceed with the probabilistic parsing in the next lecture.