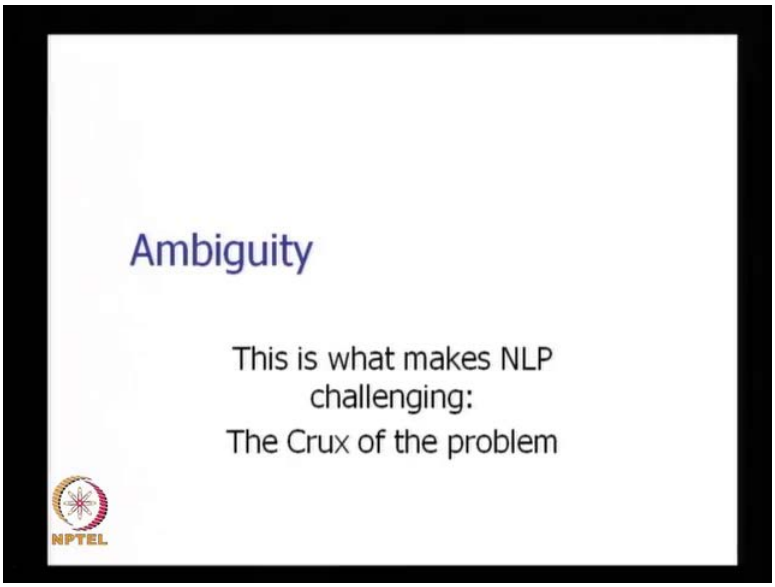
Natural Language Processing
Prof. Pushpak Bhattacharyya
Department of Computer Science and Engineering
Indian Institute of Technology, Bombay

Lecture - 3
Stages of NLP Continued

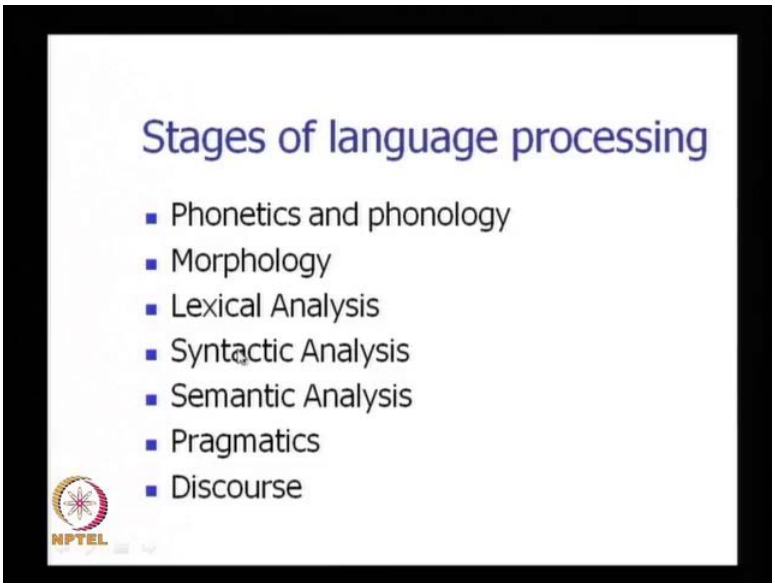
We continue our lecture on natural language processing. This is lecture number 3 and we continue with the stages of natural language processing.

(Refer Slide Time: 00:33)



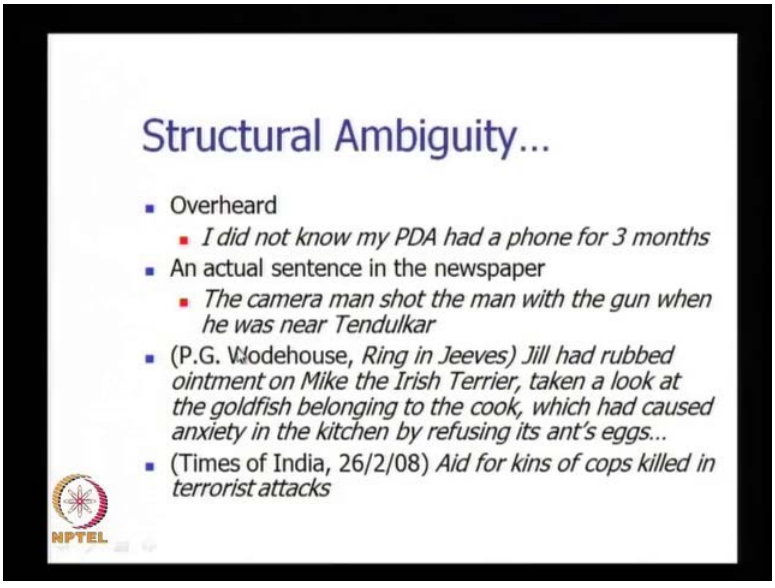
Looking at the slide again reemphasizing, the fact that ambiguity is the crux of the problem in natural language processing and ambiguity makes natural language processing the challenging job that it is.

(Refer Slide Time: 00:47)



We remind ourselves of the stages of natural languages processing: phonetics and phonology, morphology, lexical analysis, syntactic analysis, semantic analysis, pragmatics, discourse. In the last lecture, lecture number 2, we are discussing syntactic analysis.

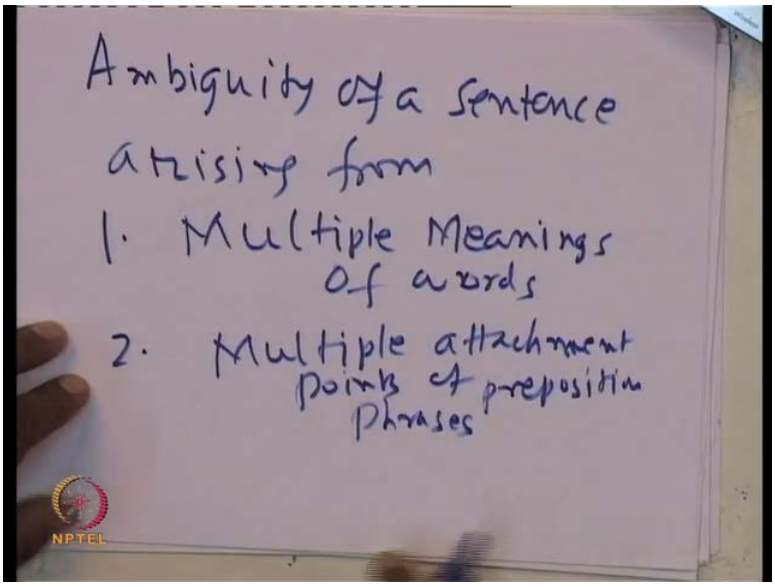
(Refer Slide Time: 01:06)



We showed this example of structural ambiguity. I did not know my P D A had a phone for 3 months, the camera man shot the man with the gun when he was near Tendulkar. This long sentence from P G Wodehouse and this particular caption from Times of India.

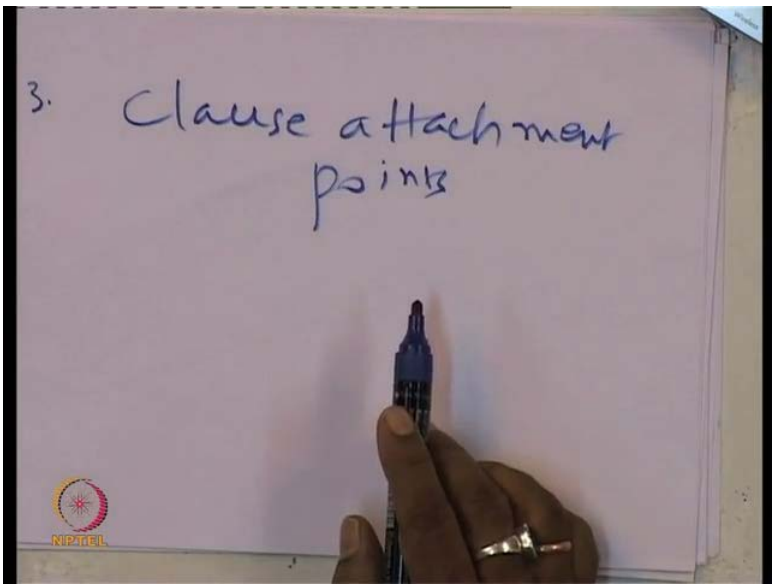
All of them have multiple meanings because of the ambiguity. Ambiguity arising from different sources.

(Refer Slide Time: 01:44)



I would like to repeat the sources of ambiguity by writing down, ambiguity of a sentence arising from multiple meanings of words, multiple attachment points of preposition phrases. These 2 are the most important ambiguity sources.

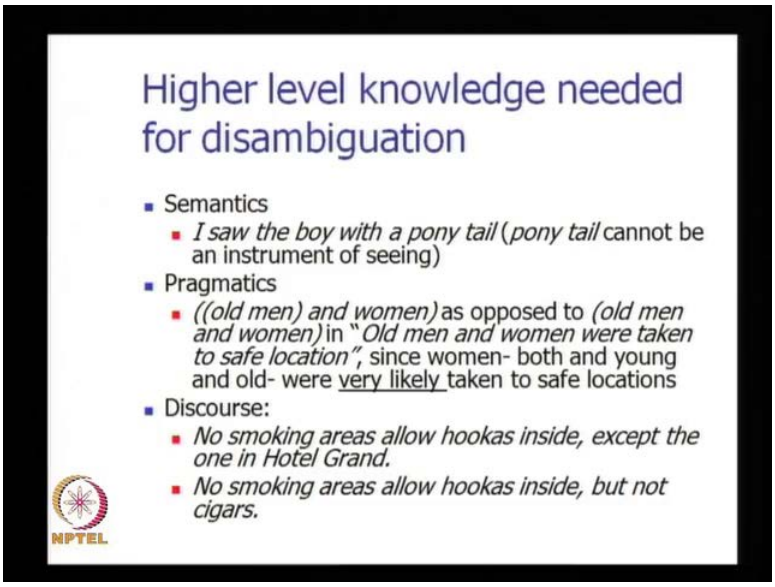
(Refer Slide Time: 02:36)



The third reason is Clause attachment points ok. So, multiple meanings of words, multiple attachment point of preposition phrases and multiple clause attachment points.

The interaction of these three produces ambiguity of sentences, different meanings of sentences ok. So, we proceed further.

(Refer Slide Time: 03:07)



In this slide we are saying that higher level knowledge is needed for disambiguation. A machine will automatically produce a multiple phrases depending on the: Attachment points and the Clausal points. The noun that takes the preposition phrase gets modified by the preposition phrase ok. So, this is a longer noun phrase where the head noun has gotten as a modifier in the form of the preposition phrase.

So, this is the preposition phrase ambiguity. The preposition phrase may be attached to the verb where, the whole phrase is like an adjunct for the verb. The other attachment ambiguity is that, which comes from a clause attachment points. The clauses can get attached to different points in the sentence ok. Now, when these different kinds of ambiguity arise, how is it possible that we still understand the meaning of sentence from the context and from the interaction of the sentence with many other sentences in the neighborhood? How does it happen? Many times it happens from, what is there in the slide. Higher level knowledge is the needed for disambiguation, which is where semantics comes in to being. I saw the boy with a pony tail.

Here, the machine, the parser will produce two phrases. In one case, the attachment will be shown with the boy, with pony tail is attached to the boy. This is a modifier for the boy. In the other case, with a pony tail will be part of the verb phrase. Saw the boy with a

pony tail. Here, the verb see has an object in the form of the boy and with a pony tail is an adjunct for the verb. So, just like I saw with a telescope, the boy. So, this para phrasing would be, I saw with a pony tail the boy. And, we immediately know that this particular sentence or this particular reading does not have any meaning. A pony tail carried out to be instrument of seeing. So, this is the word knowledge which is to be brought to be are. So, these possibilities excluded and now, we have a single phrase for the sentence. I saw the boy with a pony tail means, the boy has the pony tail.

Next example, is that we do disambiguation through pragmatics. We consider this sentence once again. Old men and women were taken to safe locations. Now, this particular sentences has two meanings. One meaning is that, both men and women were old, the other meaning is, only men were old.

Now, old men and women were taken to safe locations. Since, women both young and old, both young and old women were likely to be taken to safe locations. Our surmise or presumption would be that the word old qualifies only men because women both young and old will be taken to safe locations. So, imagine there is an attack on a region, on a country and men women everybody are taken to a safe locations. Young men of course, would go and fight the enemy. Old men will have to take into the safe locations.

Similarly, women will have to be taken to safe locations. So, the reading that we prefer is this first reading. Only the men are old and this kind of consideration is known as the pragmatics consideration ok. Here syntax is giving you as 2 possibilities. Semantics is also giving 2 possibilities. It is not excluding the isolating the possibility of women also being old ok. So, up to the level of semantics ambiguity remains. There are two different phrases with old ok, with old being a qualifier for both men and women.

Semantic also saying that both men and women can be old. Here comes pragmatics which says that both men and women being old and they been taken to safe location, is less probable. Then the fact that only the men are old and old men and women both are being taken to the safe location ok.

So, this is a purely it is known as the pragmatic consideration and here pragmatics is coming and disambiguate. The next example in the slide is the Discourse. Example where the other sentences are helping to disambiguating. So, we took the sentence no smoking areas allow hookas inside. We saw that this particular sentences has 2

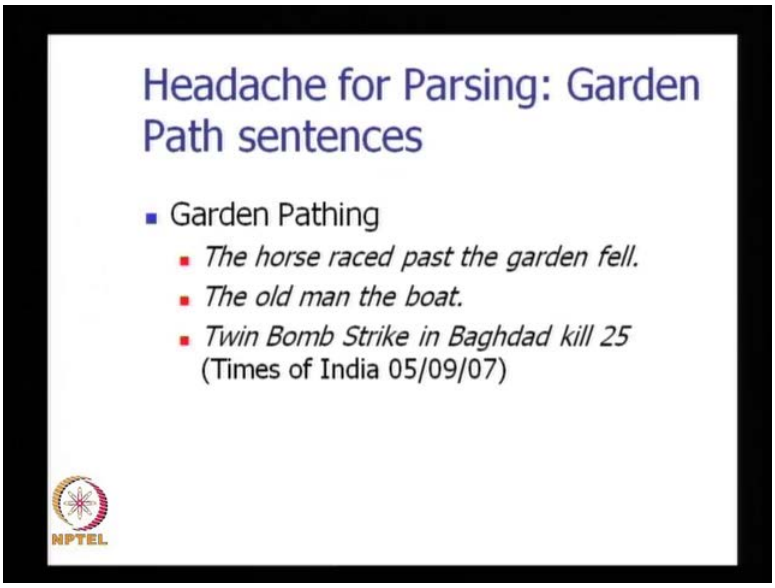
meanings. One meaning is that there are special designated areas called: No smoking areas. Those areas will not allow in cigarettes or cigar inside but they will allow hookas. Because, you can have this artificial flavored water kept inside the hooka and one can smoke that so called smoke that and enjoy the experience of hooka. So, no smoking areas they allow hookas, those artificial hookas.

The other meaning is that this knows the qualifier the whole sentence, all the remaining words in the sentence. So, this would mean that there is no smoking area. You will not be able to find any smoking area which allows hookas inside. So, they allows cigarettes and cigars but they do not allow hookahs. So, just this constrain with the earlier meaning we talked about. In this earlier meaning, the smoking areas a they did not allows cigars and cigarettes but they allowed hookas. The next meaning is saying that smoking areas are allowing cigars and cigarettes but they are not allowed hookahs, completely opposite meanings.

Now, let us see how a particular meaning is isolated from discourse, from the discourse. So, if you look at these sentence. No smoking areas allow hookas inside except the one in hotel grant, this is the discourse. The other sentence is: no smoking allows areas, allow hookas inside but not cigars. So, we have connectives here in the form of except and but and the rest of the text helps disambiguate. Let see how, no smoking areas allows hookas inside, this is that other reading. You will not find any smoking area which allows hookas insides except the one is hotel grant.

So, this except one in hotel grant that is helping to disambiguate the previous sentence. No smoking areas allow hookas inside. So, the meaning that is conveyed by this piece text is: you will not find any smoking area which allows hookas inside. The next sentence: no smoking areas allows hookas inside but not sugars not cigars. This is the meaning where, no smoking area is a special designated area where one cannot smoke. So, they allows they allow hookas but not the cigars. So, you can see how an additional piece of text coming after the sentence is helping to disambiguates. So, this is an example in this slide actually we show that even though, syntax can produce multiple phrases. In one case, semantics will exclude one possibility. Next case, pragmatics will exclude possibility and in a last case discourse, other piece of text will exclude possibilities.

(Refer Slide Time: 12:01)



Proceeding further, we now come to particular phenomenon called Garden Pathing. Garden Pathing is really a headache for parsing and there are special kinds of sentences called garden path sentences. Please understand what it means. Let us look at this first sentence here. The horse raced past the garden fell. Second sentence is, the old man the boat. Third sentence is, Twin Bomb Strike on Baghdad kill 25.

So, all this sentences have some interesting peculiarity sentence. Let us look at the first sentence. The horse raced past the garden fell, this particular sentence could have been over at garden, the horse race past the garden. What is this fell doing here? Even human being will get a mild shock. Lets us say he will get surprised after having processed all the words of to garden and then encountering fell.

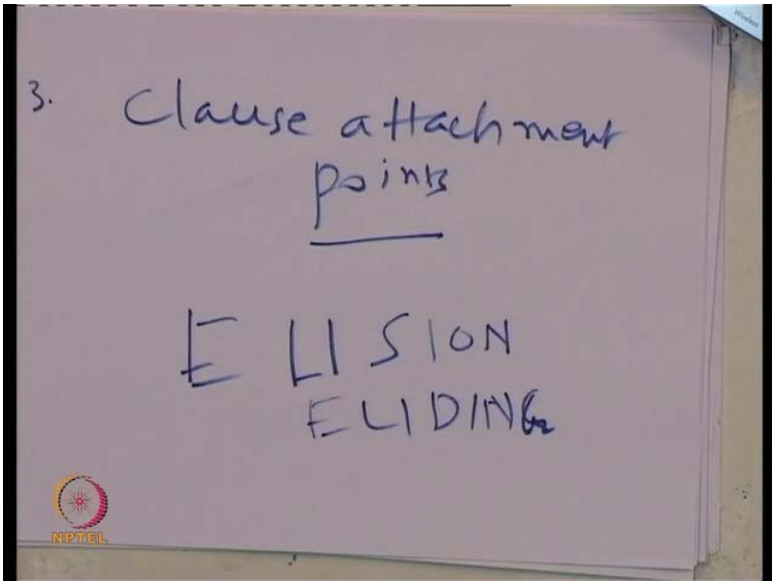
So, what you will think? You will think that, I have processed these sentences. I process this words in the sentence. I process the, I processed horse raced past the garden and then I encounter fell. The sentence could very well have been over at garden. Is there a mistake in sentence? Is this sentence grammatically wrong? The sentence is not grammatically wrong except that, after processing garden and thinking that the sentence is over here.

We, will have to back track, to account for the next word which is coming. We will have to back track many many words behind and we will have to come back and stop at raced. The sentence can paraphrased as, the horse which was raced past the garden fell, then

there is no problem. So, if we now begin to analysis the sentence, let us not look at the slide any more. We begin to analysis the sentence, the horse raced past the gardens fell. Here, the other paraphrase of the sentence is: the horse which was raced past the garden fell.

Now, which was raced past the garden, this is the clause. This is a complete sentence in itself except that it is a relative, it has a relative pronoun which was raced past the garden fell. So, the course sentence is the horse fell subject and predict, subject and verb. The horse fell, which horse? The horse which was raced past the garden. So, raced past the garden is the clause, is modifier for horse. Now, English has this, has this peculiarity. That in the past tense and under certain conditions it can have what is called an Elision and eliding construct.

(Refer Slide Time: 15:43)



Let me, write it down. It can have an Elision or Eliding. Eliding means cutting out or dropping out. So, the sentence had a relative pronoun, the horse which was raced past the garden fell. This which was can be dropped. The condition is that in the past tense a relative clause, for a noun. In a past tense, relative clause for a noun can drop the relative pronoun and the verb, the auxiliary verb if that noun is an object and the verb is in past tense.

There are the 2 conditions: the relative pronoun along with the auxiliary can be allied or dropped if the noun is an the object and the tense of the verb is past. So, here also you

can see, the horse was raced pass the garden. So, the horse is the object, somebody raced the horse. So, horse is the object and the activity which is racing, it is in the past tense and therefore, which was can be dropped. Therefore, the horse raced past the garden fell is the completely grammatical sentence. Except that, the parsing process will move on go up to garden. Think that the sentence is finished and encounter for fell and it would do the back tracking to discover that race past the garden is actually a modifier for horse with elision of relative pronoun and the option.

So, this is an example of very interesting phenomenon which is challenge for phrasing. The phenomenon is that, the sentence seems together over. But, there is an additional textual metal matter, which demands that the whole phrasing process lead to do its works. Go back to particular point, which point it is? How to find out that particular point is accomplish problem. And therefore, these sentences are a big headache for the phrasing process. These sentences are known as garden phrase sentences presumably from this example. This particular example, which mentions the word garden. Another theory is that these sentences seem to lead you, lead the leader of the sentence or the lead the parser of the sentence, along a garden path from where the phrasing machinery will have to back track ok. This is the meaning.

So, garden pathing is a very important challenge for all parsing algorithms. Whenever we design new passing algorithms those passing algorithms are tense against garden pass sentences to find out, how efficient there processing is? The next sentence is also garden pathing phenomenon. If you look at this slide, the old man the boat. This is the interesting sentence because here the garden pathing is on the phrase. The old man and the leader expectation is that, this whole thing is noun phrase, the old man. And then, he or she is in for a surprise because the old man the boat. The whole sentence seem to be without the verb. And therefore, an ungrammatical sentence, the verbs have to been disappear. Again, we have to do back tracking and we have to consider other possibilities ok.

So, let us consider the sentence once again. The old man the boat. The old man can be noun phrase. What are the other possibilities? The man, the word man can be up. One meaning of man is, man as a noun. The other meaning of man is, man ((Refer Time: 20:15)) stair. We can manage ship and this would mean that we stair ship. The old man the boat would mean old persons ok old persons. Now, again peculiarity of English is

that one can use the adjectives as nouns. The old means, the old people. The old man the boat, means the old people, many word or stair the boat.

There is the meaning of the sentence. Now, in this case what will happen is that, we will think that the sentence is ungrammatical. Because there is no verb here, having come here. If you look at the slide the old man the boat having come here, after man we expect the verb. We do not find the verb here. So, we back track and see, what is the alternate possibilities. The other possibilities is that the word man itself is a verb. In which case the old will be the noun phase, men is the verb and the boat is the objective. So, subject the old man, the verb boat and the boat object, subject verb object everything perfectly.

So, this is the grammatical and this is found out by back tracking from before the point. So, this back tracking goes back and says that manage the verb. So, this is a garden pathing phenomenon just like before. Except that in this case the garden pathing is happening, because of the multiple part of speech of man. (Refer Slide Time: 12:01) Finally, the last sentences twin bomb strike in Baghdad kill 25. This is not a garden pathing sentence parsing but, this happens in the newspaper context.

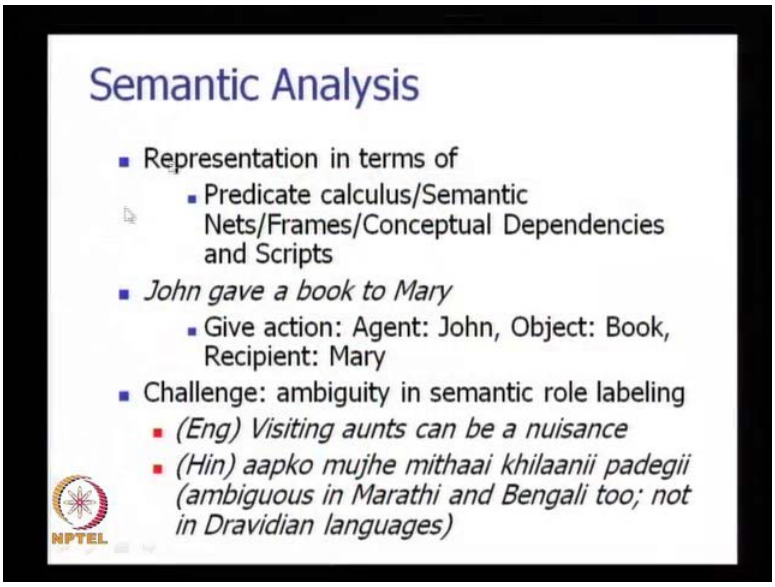
In the newspaper one is used to seeing headlines where the verbs are dropped. Headlines dropped verbs so, twin bombs strike in Baghdad that would finish the news item. Twin bomb strike in Baghdad, then we see kill 25. We know that this is not a normal newspaper heading. Twin bomb strike in Baghdad, instead the heading is complete sentence. Twin bomb strike in Baghdad till 25 and the normal procedure of processing is sentence proceeds.

So, the problem here is the following, what I am trying to say is this. That the third sentence is not garden pathing sentence. Typical garden pathing sentence is not of that kind. What is happening is that, we are in particular frame of mind when we are reading a newspaper. So, we finish processing at twin bomb strike in Baghdad. We finish processing here, twin bomb strike in Baghdad, finish processing in here. After finishing we encounter motive and then we revise our opinion about the sentence.

We, revise the processing a situation and then proceed with the other possibilities. So, this is the crux of matter in a garden pathing. Something is finished, some processing is finished. There is more material to be processed and therefore, backtrack and begin reprocessing with alternate possibilities. That is the whole crux of matter in garden

pathing. First sentence is the actual class garden path sentence, the second sentence and the third sentence are situation specific or part specific.

(Refer Slide Time: 24:17)



Proceed further, we now come to the next stage. What we have done so far is the processing of structure in the sentence. Namely: syntactic processing or parsing. We now, move on to much problem. The deeper problem of semantics, which is a much more complex task. In fact, in natural language processing a lot of progresses happen on syntactic analysis and parsing. There extremely sophisticated and very good passing algorithm. But, natural language processing to go, have to go have to really cover long distance. Before making inroads into the 5 points of semantic processing. So, this slide say all this things. Semantic analysis, the analysis semantic of sentences produces the knowledge representation of the sentence in a form of one of the schemes. The representation knowledge can in terms of predicate calculus or semantic nets or frames or conceptual dependencies and scripts.

All of these are classical, extremely well known knowledge representation schemes. Gives predicates calculus is a branch of logic, a very classical field of knowledge representation which is fundamental to any kind of enforcing work. Semantic net is concerned with representation of knowledge in the form of graphs. Where, we have notes at arks capturing relationship between concepts.

So, semantics needs, these are semantics graphs. Frames are structured knowledge representation schemes in the form of slots and figure. Where, you have a table like structure with different slots and there fillers. Conceptual dependencies are representation of knowledge in the form of lenitive constructed and scripts captured typical situations.

For example, going to lecture would mean, coming out from the hostel, coming out from home carrying one's pen, paper, and geometry box etcetera. Walking the road or taking the vehicle and reaching the class, listening to a lecture taking notes, writing an examination, all these are routine activities connected with at beginning lecture. So such things are called scripts. We will cover knowledge presentation in some amount to detail eventually.

So, semantic analysis is concerned with representation of a sentence in terms of one or more. Sometime it is more of this knowledge representation schemes. So, I take an example here, john gave a book to Mary. Here there is the give action, the giving action taking place. Who is giving? John is giving. So, john is the agent that is shown here, agent. What is he giving? He is giving a book. That is the object. To whom is he giving the book? Mary. So Mary is the recipient. Therefore, this give action has 3 entities especial entities, without which the give action is not complete.

Give action, requires an agent namely john here. Requires a object namely the book here and it requires the very beneficiary of the action namely Mary here, the person who is receiving the object of giving. So, this is an important illustration, in the sense that it shows what is obtained as a result of knowledge extraction from a sentence. John gave a book to Mary is a sentence. From here, the structure that we have obtained is give as the main verb having a agent as john, having book as object and having the recipient as Mary ok.

So, this then finishes the story about john giving a book to Mary. Now, this kind of semantic extraction is very, very crucial to natural language processing. Given a sentence if we do not understand, what does semantics of this whole sentence is, then any further processing is impossible. So, that kind of processing happen by means of what is called semantics roles. So, semantics roles capture the relationship of nouns present in a sentence, with the main verb of a sentence, the main action of the sentence. So, john

gave a book to Mary. Here the action is, give action. The nouns are: john, book and Mary and semantics roles are: agent, object and recipient respectively.

So, these things: agent, object, and recipient, these are known as semantic roles. They capture the relationship of the noun present in the sentence with the verb of the sentence. Now, when we capture the semantics of the sentence, are ambiguously, preciously, correctly then the semantics roles have become very clear for the sentence. So, to obtain the meaning of a sentence the semantics roles have to identify without any mistake. I make this point to show that, if the semantic roles is not correctly identified, there it can lead to distortion of meaning and there are kinds of ambiguity which arises from the ambiguity of semantics roles.

Look at this sentence here in the slide. Ambiguity in semantic roles labeling, we have an English sentence here. Visiting aunts can be a nuisance, it is a very well-known classical sentence in a natural language processing which illustrates the ambiguity of semantics roles. What can one make out from the sentence here? Visiting aunts can be a nuisance, one meaning is aunts who are visiting, aunts who are coming to see us, can be nuisance. In this case, the action is visit, who is performing the action? Aunts. So, aunts are the agents of visiting. The other reading of the sentence is visiting aunts can be a nuisance, where aunts are objects of visiting. So, I am called upon to visit my aunt and I am not happy about this.

So, in this case the visitor is I, the agent of visiting is I and the object of the visiting is aunt. So, if you contrast this, with the previous sense of the sentence there the visitor was aunt. So, aunt is the agent, in a next meaning aunt is the object. So, this particular sentence has left the ambiguity of semantic role leveling unreserved. From the sentence one cannot make out, what is the semantic role of aunt? Is aunt the agent of visiting or is aunt the object of visiting? Again here I insist that, you translate the sentence in your own mother tongue and you will see that unless you commit a particular semantics roles, the sentence cannot be translated on ambiguity. You have to leave both the meanings open. So, one for example, if we take Hindi.

((Refer Time: 33:15)) in this case I am visiting the aunt, so ((Refer Time: 33:21)) the other meaning is visiting aunts, Aunts were visiting us. So, in this case it is ((Refer Time: 33:38)). Depending on that, the semantic role of aunt is changing and this is making the

sentence. English sentence ambiguous, the Hindi sentence or for that matter any Indian language sentence. I believe will not be ambiguous or in other words, the sentence when translated will have to commit to semantic role disambiguation. We take an example just below it which is an Hindi example ((Refer Time: 34:19))

This particular sentence is ambiguous again because of semantic role. What is the action here? The action here is ((Refer Time: 34:37)) ok, or to give. So, here what is happening is that a giving action is taking place or feeding action is taking place, which is the action of ((Refer Time: 34:53)).

Now, there is no ambiguity with respect to object or what is being given effect. It is very clear it is ((Refer Time: 35:04)) Sweets. The problem comes in, who is giving sweets to whom? ((Refer Time: 35:12)) You have to look at to me now because I have to perform that action. ((Refer Time: 35:12)) I will eat the sweets. So, in this case you are giving me, the sweets. So, I am the beneficiary of given action, I am the recipient the object is clear the object is ((Refer Time: 35:41))

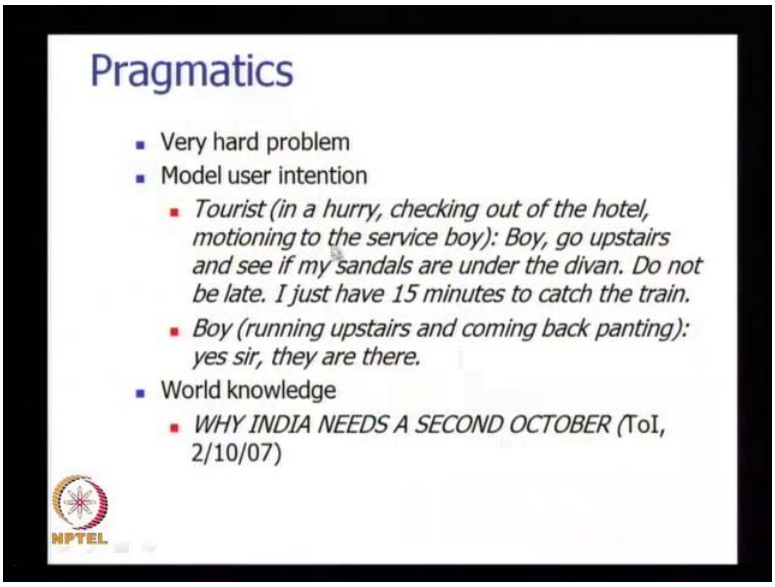
There is ambiguity with respect to agent and beneficiary. When I am the beneficiary ((Refer Time: 35:51)) I get the sweets ((Refer Time: 35:54)) you have the give sweets to me. The other reading is ((Refer Time: 36:01)) I will give sweets to you. The other reading was you will give sweets to me. So, there is the semantic role reversal between me and you, in terms of agent and beneficiary. So, in the European languages which are close to Hindi like Marathi and Bengali they retained this ambiguity. In Bengali we have to say ((Refer Time: 36:35)).

So, this sentence is ambiguous, because it does not specify who is giving sweets to whom. This sentence is ambiguous in Marathi also but, it is not ambiguous in Dravidian languages. Where you have to produce the sentence after resolving semantic roles because that will decide the case marker and others suffix information on the nouns. The semantic role will have to disambiguated, just like visiting aunts can be nuisance is an ambiguous sentence in a English. But, when we take into Indian language sentence, we have to commit to the semantics role and the ambiguity has to be resolved.

Proceeding further, we come to pragmatics. So, what we saw was semantic role labeling which is processing of semantics. When you come to pragmatics, we are concern with how a sentence is processed by a user. When speakers utters the sentence, when listener

listen to that sentence and information giver and information recipient. How they look at the sentence?

(Refer Slide Time: 38:01)



So, this are the very very hard problem know in natural language processing. We look at the transparency here. And, see an example of pragmatics, being important. Pragmatics is concerned with modeling using intention. So, see here I have a piece of conversation. There is tourist who is in a hurry, the tourist is checking out of the hotel and he is motioning to the service boy.

The service boy, boy go upstairs and see if my sandals are under the divan. Do not be late, I just have 15 minutes to catch the train. The boy running upstairs and coming back panting. Yes sir, they are there. So, the boy is answering the tourist question appropriately, there is no problem about that. So, he is saying that the sandal is under the divan but, the tourist intention was to get that sandal and he was already late for the train and therefore, this sandal had to be brought to him.

But, the sentence, his sentence only specified to the boy that he should go and see if the sandal is under the divan. There was no specific instruction with respect to the boy, getting the sandal and giving to him. So, that was the crux of the problem, human beings are no problem with this kind of situation. They are extremely good at dealing with it. The pragmatics of the sentence, so when a sentence is uttered, we understand the intention behind that sentence.

We, also understand who is the recipient of the sentence? Now, in this tourist boy conversation, it was actually an instruction for the boy to bring the sandal. Though, the sentence did not said in explicit terms, the intention was that. And therefore, the stage of pragmatics, for the natural language processing is concern with modeling the user intention which is very hard problem to illustrate. What I mean by this? Let me give you another example, many times we sit on the dining tables and we point to the neighbor, is that water? You ask the neighbors is that water point to jug and say is that water? The intention actually is, for you to obtain the jug of water. You like to drink some water.

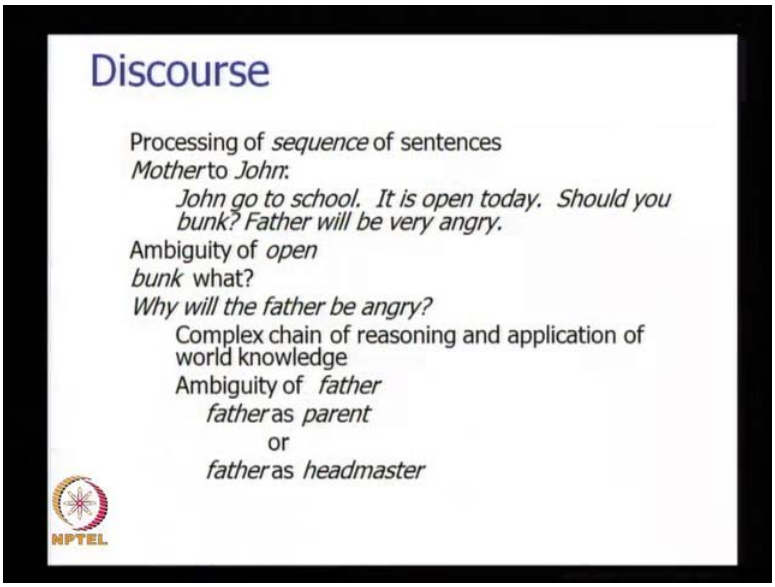
So, when you ask your neighbor, is that water? This is not actually the question, it is actually a request in the form of a question. The request is, please past me the jug of water and if the neighbors just says yes or no, is that water? Yes, it is water and then does nothing about it. Then there is the pragmatics failure. There is no: syntax, semantics or lexical processing in failure. There is pragmatic failure, we the neighbor the dining table has not understood the intention behind the question. Notice that, the same sentence is alright in a case chemistry lab situation. So, it is possible that an examiner comes to a student, performing a practical examination and points to water and says, is that water? The examiner preassembly does not have any intention of drinking the water but, he is examination the student with respect to what that particular compound is.

So, in a chemistry lab situation pragmatics again is ensuring that this question is actually question. In the dining tables situation, this question was not question, it was an actually a request. So, pragmatics is extremely situation specific. There is some kind of pragmatics playing a role in this sentence. Why India needs a 2nd October? Times of India 2nd October 2007. This particular sentence will not be understand very well not be understood very easily about by a non-Indian. That a person reading this sentence, we will have to understand the significance of 2nd October, which is the birth anniversary of Mahatma Gandhi. 2nd October has a special significant for many Indian. This is the birth anniversary of Mahatma Gandhi.

We celebrate 2nd October with different kinds of a: poojas, bhajan and songs and so on. So, why India needs a 2nd October will be wrongly interrupted or completely not understood by a person who does not understand the meaning of 2nd October, the special significant of 2nd October. So, this again is a pragmatics consideration, it shows the importance of pragmatics where a special word knowledge or situation knowledge is

helping the user, speaker, listener to understand a sentence. So, this is the scope of pragmatics. Pragmatics is a very hard problem mainly because, it has to do user model it has to know user preference, likes, dislikes. Pragmatics also to know situation specific constructs and their significance.

(Refer Slide Time: 44:23)



Now, we move on to the last stage of processing presumably. The very very difficult stages once again, like pragmatics. This is the stage of discourse processing, discourse processing is concerned with processing of sequence of sentences. So, far we have been discussing only one sentence. A sentence which is demarcated by 2 full stops on 2 sides. In this case now, in this course we are concerned with sequence of sentences, I take here a piece of conversation. Mother to john, john go to school. It is open today. Should you bunk? Father will be very angry.

So, this 4 pieces of text are uttered by a mother and the listener is a boy called john. Now, when we look at these 4 sentences, one of which is interrogative. We cannot but, we astonished is at the ease with which we process this sentences. Because there are many, many complex tasks involved here. The 1st challenge to process this sentence is, is the ambiguity of open. Open is very, very policy able verbs. A verb with many many different meanings and in this case we are concerned with that particular meaning of open, which says the school is working, school is open today. It is not at the school doors

and windows are open. The school is working. It is the open to means, the school is working today.

So, there is ambiguity of open and we have resolve this ambiguity by taking the particular meaning of open, which is working. The next challenging problem here is ellipses, we mentioned before that in garden path sentences are. The difficulty of processing comes, one kind of difficulty processing comes because of elision. The relative pronoun and the auxiliary verb has dropped. And therefore, there is a difficulty in parsing.

Now, when we consider this sentences here, the school is open today and should you bunk? Is the next sentence, which the mother utters or john should you bunk. The question is, how does john know what is the object of the bunking. Bunk is a verb, should you bunk what so, should you bunk the school naturally. The school as a piece of text is coming from one of the previous sentences, which sentence? The sentence is the first sentence. John go to school, it is open today. It is a pronoun which refer to the nouns here, this kind of pronoun the noun referencing is called an of area and big branch of natural language processing is concern with analysis referencing. How do you correctly buying, a pronoun with a particular noun.

The next problem that we are try to resolve here, is the problem of ellipses, should you bunk? Should you bunk the school and the piece of text. The school came from the previous sentence, not the immediately preceding sentence. But, the sentence before that you can presumably see that this kind of ellipse handling may require obtaining textual material. Textual material from a vary distance sentence may be 5 or 6 sentence away. You have to pick up the textual material matter from that. Therefore, ellipse is handling is a difficult problem. This is a challenge.

Now, we come to the last sentence, father will be very angry. Question is, why will the father be angry and it is imagine how or mind processes this sentence. There is complex chain of reasoning an application of word knowledge here. The father is angry because the son is disobedient, john is disobedient or he is angry because he is apprehensive about, john not going to the school and forming the bad habit and there by entering the possibility of a league future.

All this complex chain of reasoning comes to the father minds and he becomes angry. Again, we can see that we have applied world knowledge, namely the fact that discipline is important in our life. We have to attend the school regularly, we should have good habits. This is world knowledge and we are also resolving the ambiguity of father. This ambiguity is very interesting ambiguity. One ambiguity point is that the father is, john's father.

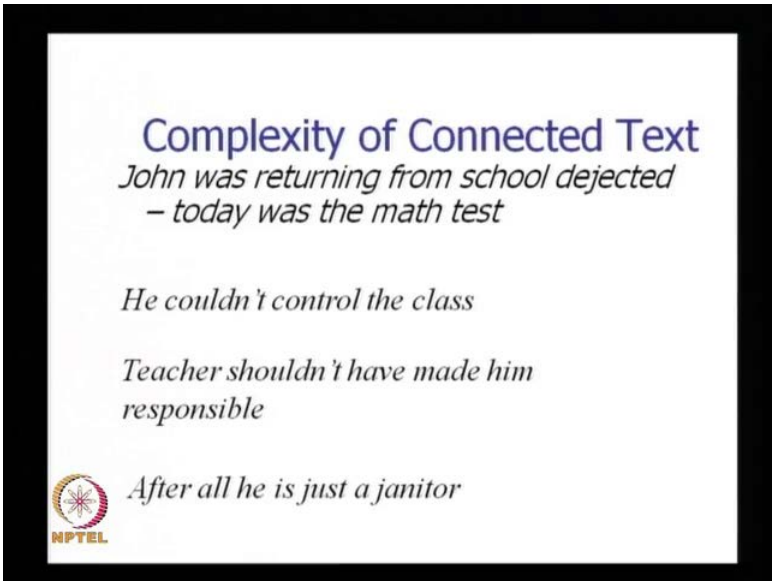
When we say father will be angry, it could be somebody else father also, jacks father. But, form the context it is clear that the father that is being refer to actual john's father. The other ambiguity consideration is that, father itself is ambiguous, it can make in either the principle of the school or parent. So, john's father would be john's parent and the school's father is the principle of the school and we also notice that somebody else father will not be angry, which is why we can sort of sorting that the father being refer to here is john father.

It could also be, without any problem it could also be the principle of the school. Father will be very angry that possibilities also remains. So, the mind will operate with two hypothesis, one is john's father and the other is school principle, the school's father. And since, the mother is saying to this john there is some kind of proximity consideration, mother preassembly is referring to john's father. He is the person, he is closet that ends and therefore, it possible means johns father. But, we cannot rule out the possibility of the principle of the school, being angry.

So, in somebody what have I illustrate through this text? You can listen to me and not look at the slide. This means that when we process connected sentences, this course sentences we keep the knowledge of previous sentence in mind. We also try to predict of what sentence is coming in the future and the complex interaction of all this. Finally, produces the meaning in our mind and on the way, we solve many problems. The problem of ambiguity of words, multiple means of words. We solve the problem of ellipses dropping of text or text which is not mentioned. We solve the problem of all ambiguity of word, lexical ambiguity, worlds have been multiple meanings. Then complex is an emphasis to finally arrive it the meaning.

So, all this goes in our mind and it is indeed and a remarkable fact that, we can process discourse at all. We can process in a number of sentences together by solving all this little difficulties on the way.

(Refer Slide Time: 52:53)



Moving further, here is the very interesting example, look at this slide, the complicity of connected text. This was pointed to me by student of mine from an actual web example. Look at this sentence here, the john was returning from school dejected, today was the math test. Anybody, looking at the sentence when asked, what do you think about john? Who is john? So, the reader or the hearer of the sentence would, in all probability say that john is student in the school, john was returning from school dejected. Today was the math test. He probably could not do well in the math test.

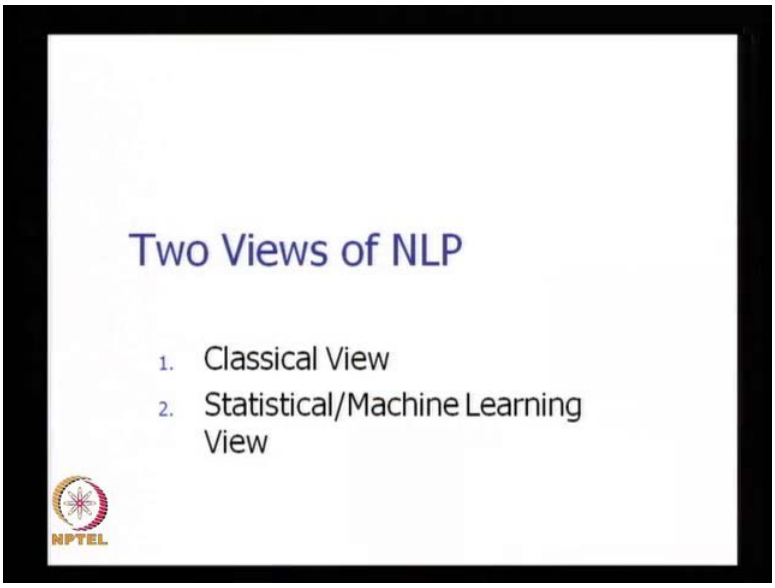
Next sentence, he could not control the class. So, after seeing the first sentence our hypothesis about john was, john is a student of the school. When we encounter this sentence now, he could not control the class and therefore, he was dejected, our hypothesis. The previous hypothesis, actually does not we are up to this new evidence. Who control the class? The teacher control the class and seems johns is returning dejected from the school and he could not control the class or hypothesis about john is slightly to be teacher. John is a teacher. So, the previous sentence showed so, previous sentence sort up let to the surmise the john is student. The next, sentence is saying john is possibly a teacher.

Next sentence in the slide, teacher should not have made him responsible. See here, we have come back to the student hypothesis about john, teacher should not have made him responsible. Whom does a teacher made responsible? It is a special student class who called the monitor, the head boy of the class. Now, we are back to john being a student, special kind of student namely monitor on the head boy. Finally, we encounter this sentence: after all he is just a janitor.

So, this says that john was returning school, from the school dejected. Today was the math test, make john student. He could not control the class, makes johns teacher. Teacher should not made him responsible, makes john a monitor. And then, he finally when we encounter the sentence, after all he is just a janitor, that over throw all hypotheses about john neither teacher nor student nor monitor. He is a janitor who cleans the class rooms, swipes the floors and so on. So, this are very, very instructive example which shows that, in natural language processing for every new piece of data. We will the hypothesis form so far in our mind against the sentences which are arriving.

So, the first sentence made john student, second sentence made john a teacher, third sentence made john monitor, special student and the forth sentence made john a janitor, the person who cleans the class. So, this shows that when we process a set of connected sentences it becomes a very complex problem, we not only solve world ambiguity. We not only solve anaphoras by finding a pronoun to noun, we not only solve ellipses where there are unmentioned text. We also form, we also solve the problem of forming hypothesis and discarding them on the phase of new evidence or data.

(Refer Slide Time: 57:27)



Just look at the last transparency and with that we will close the class. This will be our next topic of discussion where we say that natural language processing has been attempted from two different directions. One is the classical approach to natural language processing, which make the huge of knowledge and rules. And, second approach is statistical machine learning approach to natural language processing, which is the current approached to natural language processing. With that we will close and we will discuss this topic in a next class.