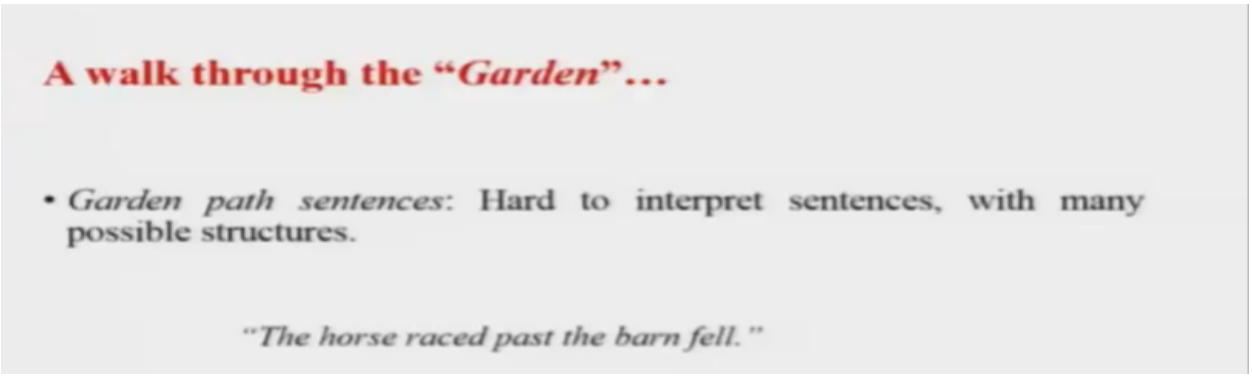
Lecture 22

Parsing Sentences - 1

Hello and welcome to the course introduction to the psychology of language. I am Dr. Ark Verma, from IIT Kanpur; we are as you know, in the fifth week of the course talking about sentence processing. In the last lecture we talked a little bit about, what are sentences? How do we work with sentences? What is parsing a little bit of that and today, we kind of try and look a little bit more you know, in a more detailed sense into, what parsing could be about. So, let us move ahead, now one of the kinds of sentences that you know, researchers have been really interested in, understanding, are having referred to as the garden path sentences.

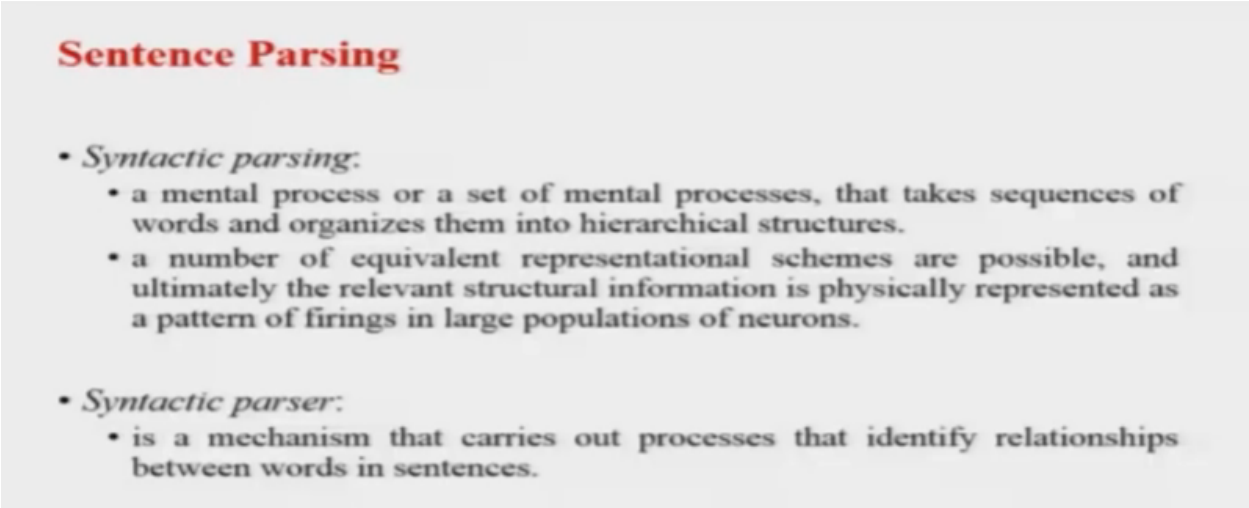
Refer Slide Time :(0: 53)



Now, garden path sentences are, typically as the name suggests, they are like paths in a garden you know, all kind of going round and round, mushy mashing, into each other and all of that. And why is this metaphor really taken, is because these sentences, are very hard to, make sense of, there are multiple structures, multiple paths that you can take to understanding each of these sentences most parts are kind of valid, parts and you don't know exactly, which is the correct path. Okay? And this is also, kind of presents a challenge, to the scientist, in order to understand, which is the exact part that the particular comprehended or listener is going to choose, what are the rules he is going to kind of employ, in choosing this path, how is this path going to lead to the correct interpretation of meaning. So, in that sense I hope you understand: that it is, kind of interesting for linguists or a psycho linguist for that matter to understand, how people are going to process, these garden path sentences. I am taking an example, again from Traxler, not one of the best examples, but, this is the example, here. So, the horserace past the bond fell, the horse raised past the barn fell or the horse raced past the bond fell. Okay? So, there can be many readings of this sentence, the horse raced past the barn fell, the horse race, horserace past the barn fell something like that. So, one of the things about, garden path sentences is: that you can kind of come up with multiple readings of these, most readings will be you know, grammatically acceptable, there will be many possible structures and you have to be able to choose, the correct

mathematical structure and interpret the correct single intended meaning, of the individual. Okay? This is why, garden path sentences are interesting. Now, how do we deal, with this obviously we have to do a little bit of parsing, but before we move, ahead with parsing, let us try and understanding, the understand that also.

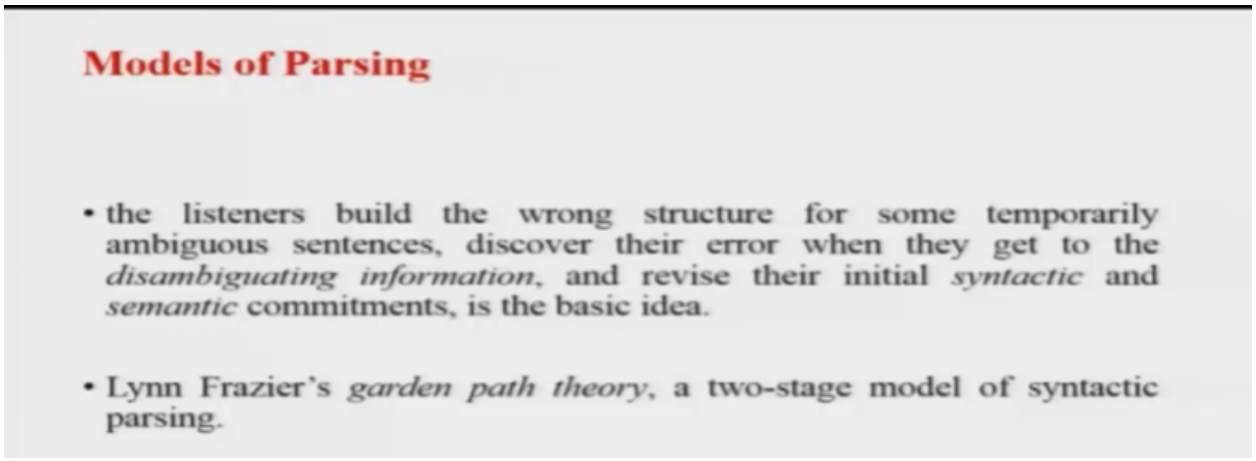
Refer Slide Time :(2: 52)



Now, syntactic parsing: as I was saying is basically understood as a particular mental process or a set of mental processes that takes the sequences of words, reading at you kind of hearing and organizes, them into hierarchical structure. So, it organizes them into structures, wherein like okay. This is, this particular phrase, this is that phrase, this is the head verb, this is the head noun and this kind of structuring will give you an idea of what, is the correct reading, of the sentence and what is the correct interpretation of the meaning of the sentence. Okay? If you kind of want to understand it in terms, of you know, how the brain will must be doing it or how you know, a neural network, must be doing this, so you can kind of, think of that, number of equivalent representational schemes can be come up with, you know all of these represented, so how do you represent this sentence? Option ABCDE, all options are possible. So, this can be kind of represented and they can be physically represented as, pattern of firing in the sense of neurons and whichever one kind of bins, whichever pattern of firings kind of bins, becomes the in accepted representation or accepted intended meaning of the individual. That’s basically how, this really works you don't really have to, think about the physical representation part for now, what you really need to understand is that, syntactic parsing is basically, a set of mental processes: that kind of evaluates, different representations, of sentence that are possible and organizes these words, into these

particular representations structures and helps us choose, the correct structure. Okay? Now, if there is a process there must be a processor, sort of so there is, a hypothetical device, called the, ‘Syntactic Parser’ and the syntactic parser can be thought of as a device or a mechanism: that carries out processes, in identifying relationships between words in sentences. So this, processor or parser basically, kind of starts you know, taking up words organizing them, on the basis of their relationships with each other and this organization broadly kind of helps you understand the whole sentence. Okay? So, this is this and now, if you have to kind of we can move ahead, what we have to actually do is, we have to kind of look at some models, of parsing.

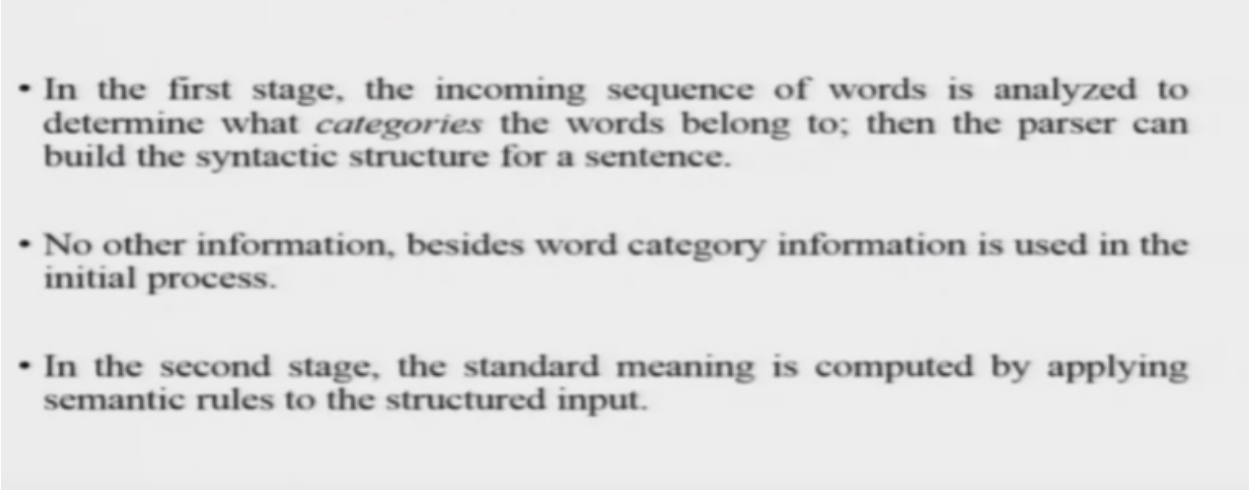
Refer Slide Time :(5: 16)



But, before we do that, there is, let me give you a brief idea of how this, enterprise really happens. So, the idea is that the listeners are continuously listening, for whatever is being said so, sentences they are kind of attempting to, organize them, into particular structures, sometimes, what can happen is? On at least in these grammatically, ambiguous sentences, the garden path or the globally ambiguous sentences, the listeners will build a wrong, structure for this for some temporarily ambiguous sentence, then discover that okay. This is erroneous, this is not the correct grammatically acceptable thing, when they get to the disambiguity information, which is say for example, I Susan was dressing the baby played on the floor. So, they will realize: that why? Susan was dressing the baby played on the floor will go that side. Okay? So, they'd revise their initial syntactic construction and syntactic and semantic commitments and then, kind of the latch onto the correct idea. So that is, broadly what we're planning to do. Okay? Lin phrase here, gave this very interesting theory, called you know his, his theory was called the, ‘Garden Path

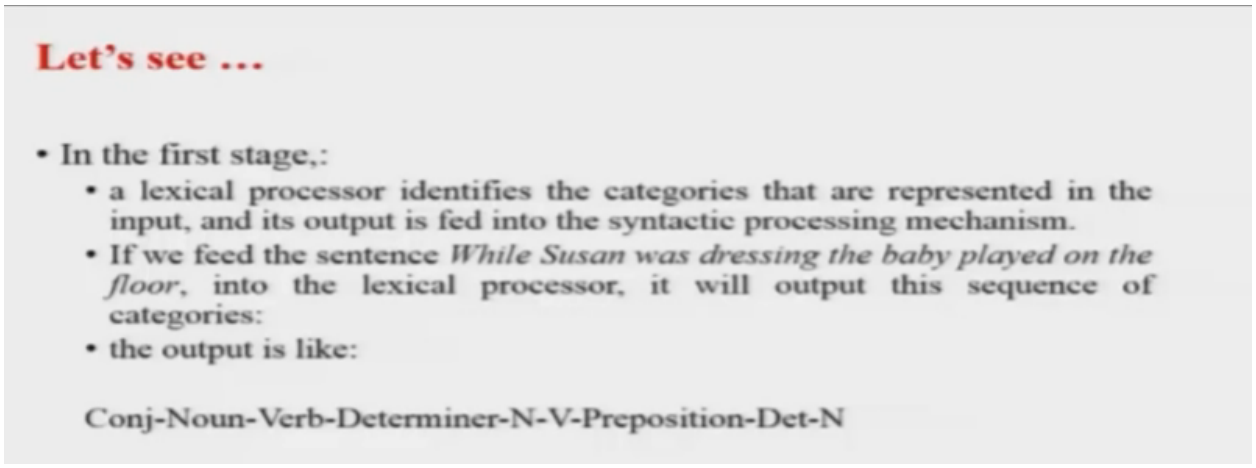
Theory’ which was basically, there to understand, how people process these garden path sentences and he said that, this is a two stage theory of syntactic parsing.

Refer Slide Time :(6: 28)



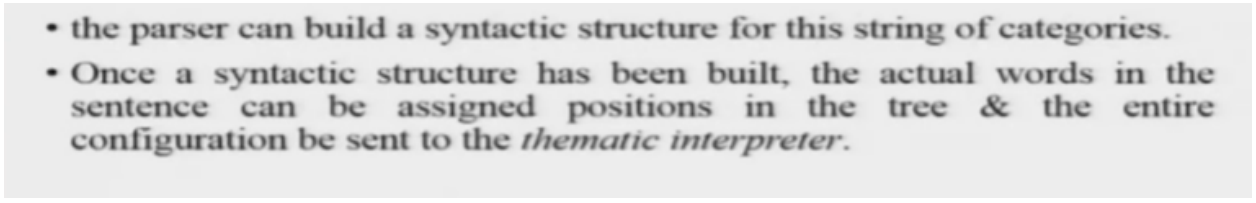
Let us look at the two stages. In the first stage, the incoming sequence of words is analyzed to determine what categories these words belong to. So, one of the very important information, in understanding where each of these words will basically you know, or how each of these words can be grouped, which are in coming from these sentences, is by understanding, what categories do they belong to is, this a noun, is a preposition, as an article, is it a verb, is it an adverb, all of that knowledge you need to have and that, knowledge will help guide you, it will help you, put, things that go together in one particular phrase or in one particular Clause. So, no other information, besides word category information, is used in the initial process. So, they're saying: that in the first stage of the garden path, processing theory or garden path parsing theory, basically what you will just get? Is you will get the information about word categories? So that is the important part. In the second stage, you have to kind of now go, go towards the meaning. Okay? You have to kind of generate a particular acceptable meaning and this; acceptable standard meaning has to be computed, by applying semantic rules, to the structured input. So, basically you kind of on the basis of this knowledge, about okay, this is the structure and these are the word categories, you kind of put some organization there and then you kind of evaluate, their plausibility in a semantic manner, is this making sense, is this not making sense, if this is making sense, in the process it's alright it, ends there, if it is not, then let's repeat the process. So, let's see what, what is it, what is done in the second yeah! So yes, let's look at this in a bit more detail.

Refer Slide Time :(8: 11)



So, the first stage, basically in a bit more detail is: that there is, this lexical processor and this lexically processor basically identifies the categories that are, represented in the input and its output is basically fed into the syntactic processing mechanism. So, the lexical processor is just looking at words adding, category information to it. So, you can look at the bottom of the slide, there is this kind of, output that you can expect. So, a conjunction noun verb determiner, noun verb, preposition determined noun. So, this could be say for example, you know, the output that would come out of reading this sentence, while Susan was dressing the baby played on the floor. So, while is a conjunction, Susan is a noun, was verb and was dressing is the noun and verb. So, this is basically so now, Susan was dressing is the verb, determiner is the, noun as baby, played is verb, preposition on and then I did mine at the and floor is another noun. So, this is basically how this mapping needs to be done, this is the kind of output: that we come out of the first stage that is the lexical processing stage.

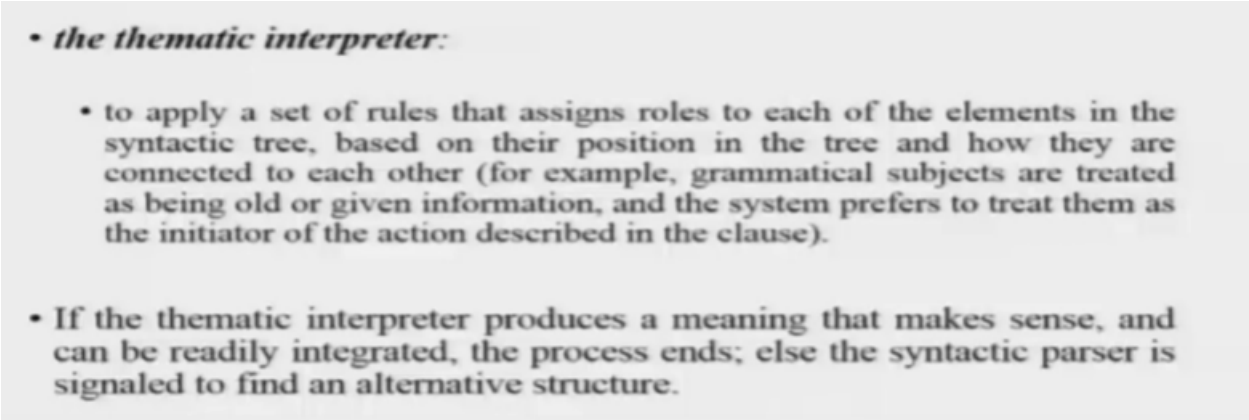
Refer Slide Time :(9: 15)



Then, what happens is that a parser kind of builds an initial syntactic structure, for this string of categories. Once a syntactic structure has been built and the actual words have been assigned in these positions, the entire configuration is sent to, what is called a thematic interpreter? So, this is where the second stage of

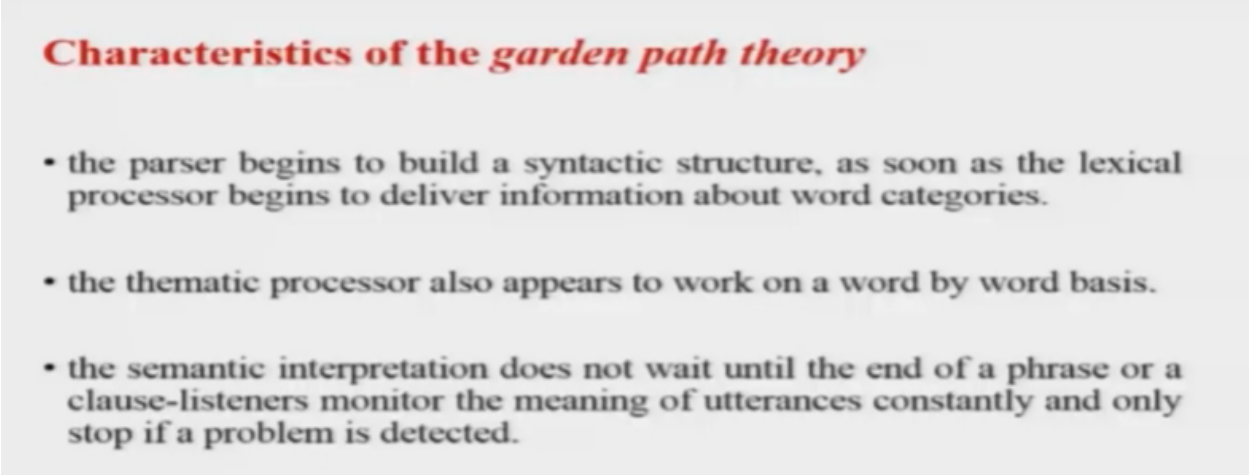
meaning, making has to start. Now, the semantic interpreter, this thematic interpreter sorry, the thematic interpreter basically, what it needs to do is?

Refer Slide Time :(9: 42)



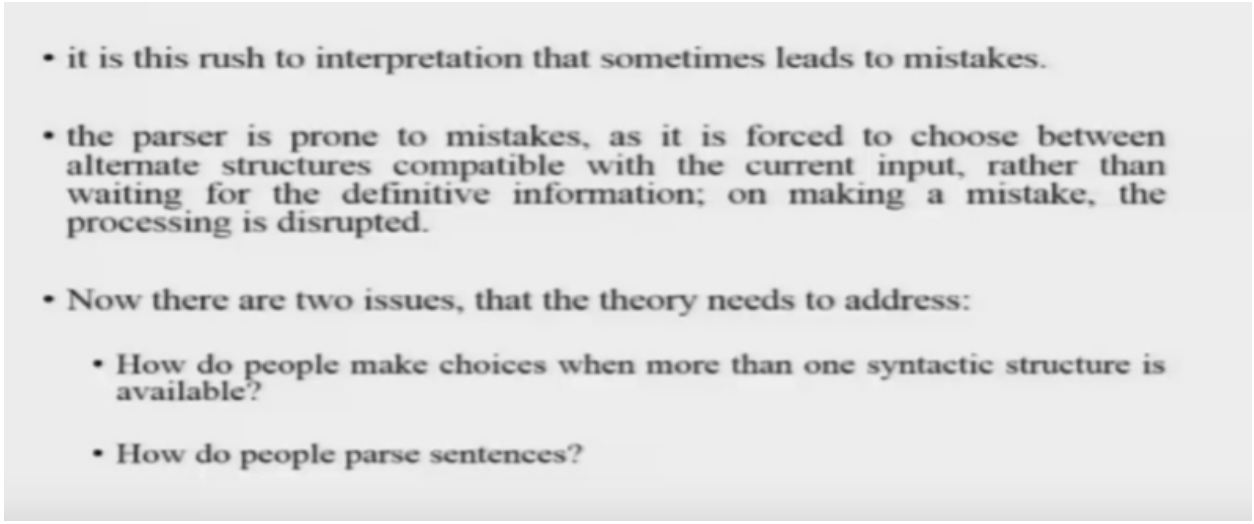
It needs to apply a set of rules and that assigns you know, roles to each of the elements in the syntactic tree. So, you saw that syntactic tree earlier, now you have this you know, conjunction, noun, verb and you have to kind of you know, make this tree and assigned role to them where will, each of these elements go in that tree. Based on their position in the tree and how they are connected to each other, all of that has to be factored in. For example, grammatical subjects, are treated as being old or given information say for example, if a sentence starts with a particular grammatical, subject that's a given: that is considered as the initiator of the action. Okay? Now, if the thematic interpreter can produce, a meaning that makes sense, as I was saying, in the first iteration, the thematic interpreter kind of finds that okay, this organization is all right and this already is making sense, then the process shall end, else the process will kind of go back and revise again. And this process will keep on iterating, again and again until you come up, with a thematically, semantically and syntactically you know, consistent reading of this interest, so that is basically the idea. Now, let us look at the characteristics of the garden path here in some more detail, now the parser here, begins to build a syntactic structure,

Refer Slide Time :(10: 55)



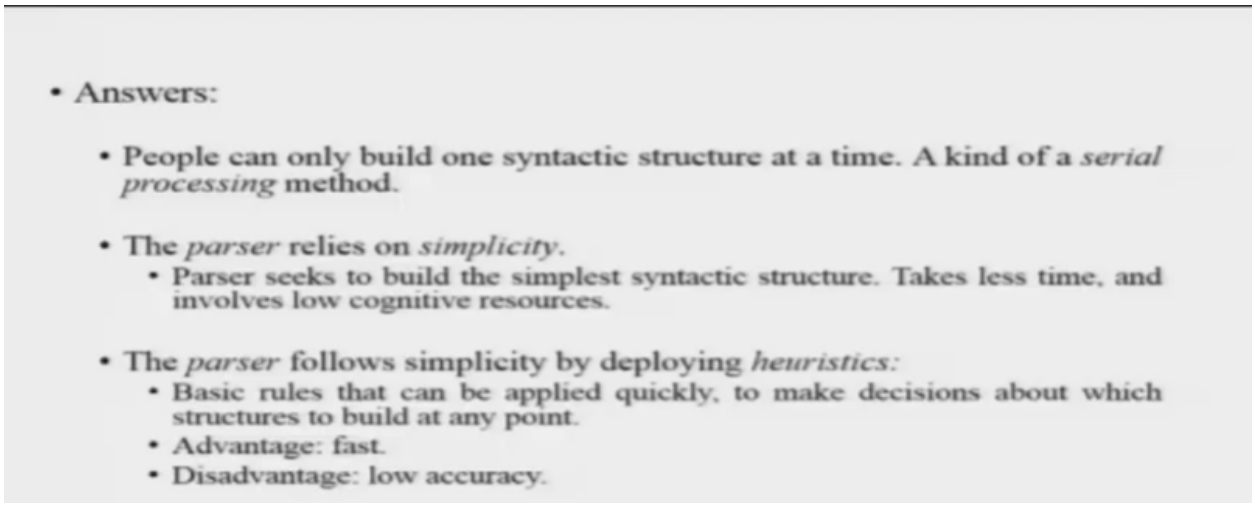
as soon as the lexical processor begins, to deliver information about word categories. So, as soon as the information starts coming about word categories look at the first one is noun, in second is verb, third is determine as etc, etc. The structure also, starts being built, almost on the fly, almost in parallel. Okay? The thematic processor also, appears to work, on a word, by word by so, the thematic interpreter is kind of also going, as in when each of the words are coming, the thematic process is attaching or integrating that to a particular structure that could make meaning. Now, the semantics in depression also, of the sentence you know, you're reading of whether the structure is going to make sense or not, also does not wait till the end of the phrase or clause and basically, what is happening is that, the listeners are monitoring the meaning of the utterances, constantly and in an evolving, fashion. Okay? And they only stop okay, this way, if I'm going it makes a wrong sense, let me revise my reading of this, this way so for example, you know, a lot of times if you are in a conversation, sometimes you miss say, a word and somebody kind of or miss hear, somebody miss hears the word, they kind of ask okay, what did you say? In order to just revise, so this process of understanding, is happening constant you know, is happening constantly and it's happening in a continuous, manner as well. So, as I am saying, your understanding, if you get confused, will kind of pause or remind back, in a normal conversation, we might as well ask the other speaker saying okay, what are you saying exactly? Because the what based on whatever I heard, my meaning interpretation goes in a particular direction, but say for example, I don't really think that this is what you might be meaning. So, I'm requesting, you to do it again, to say this again so that I can kind of start on with a new structure, more often than not you don't really need to ask, you just kind enough you know, can revise your interpretation, on the fly without really anybody knowing about it as well. This is basically, what is happening in parsing.

Refer Slide Time :(12: 52)



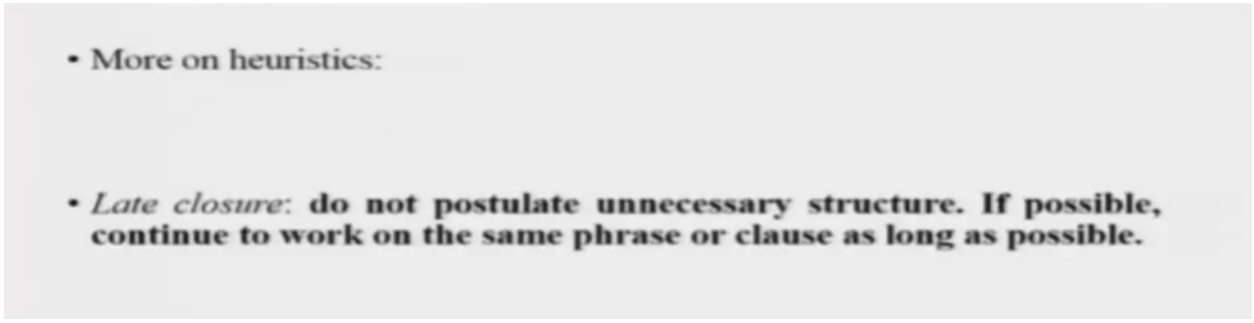
Now, this rush to interpret you know, somebody might ask why are you in such a rush of interpreting and why can you not do it, at the end of the sentence. Obviously that's, probably the correct way to do, but, then probably the process will not remain as fast, so the cost comes you know, the speed comes at a particular cost. So, there is this rush of interpretation: that obviously sometimes leads, to interpreting mistakes, the parser as, as all of us know, is prone to mistakes, as it is forced to choose between, alternate structures compatible with the correct input. So, whatever concurrent input is coming, many structures can be compatible with this kind of input and the pastor has to choose the correct, obviously the person will kind of make mistakes some places. Okay? So, rather than waiting for the definitive information or making a mistake, the part of the process you know, the processing is disrupted whenever a mistake is made. Now, there are two issues that in, in here you are parsing especially, the garden bar theory needs to address: is how do people make choices? When more than one syntactic structure is possible, question one and the second question is, how do people pass sentence in the first place? How do they come up with sentences, with these structures in the first place? Now let, let me answer this to some detail, now people usually, can build only one structure at a time, this is what the garden path theory is saying? So, the garden path theory is a serial processing method and it says that, people will be able to build only one syntactic structure at the time, if that fails, they will come back remake another structure. If that will fail, they'll come back remake our structure. So, this is a serial processing model one.

Refer Slide Time :(14: 30)



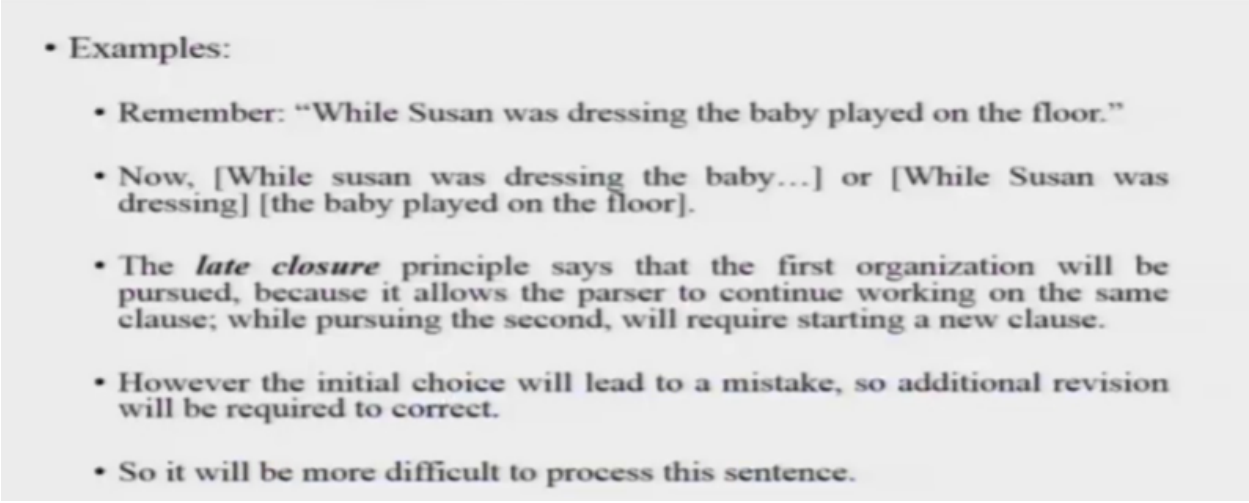
The parser relies on a simplistic idea, tree relies on simplicity and what it basically is seeking is? It's seeking to develop, the simplest syntactic structure. The simplest meaning, the most plausible meaning is the one that a parser will go with, it would take less time and it involves lower cognitive resources. So, that you can come back and redo all of the job again, if the mistake is that, more often not you probably not make, so many mistakes and the process will remain fast and efficient. Now, the parser could do this, by following some heuristics, shortcuts that is and these heuristics basically are the basis, are the basic rules: that can be applied very, quickly and these rules can be applied to make decisions, about which structures, to build, at any point in time. The advantage of this is, this keeps the process, fast the disadvantage is obviously there is a cost attached to this.

Refer Slide Time :(15: 24)



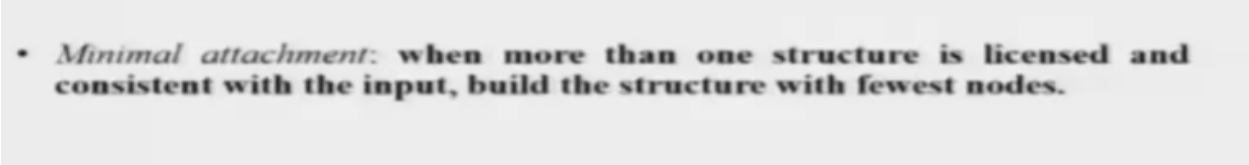
Now, let us look at some of the heuristics that have been suggested, under the garden path theory, of Lin parser. One of the first heuristics that I can talk about, is the heuristic of late closure: what is this heuristic of late closure? It says that the parser does not need to postulate any, unnecessary structure, if possible, it will continue on to work, on the same phrase or clause, as long as possible. So, while Susan was dressing the baby played on the floor, probably happened in this sense, because you're following this heuristic, while Susan was, dressing the baby, everything is being attached to the same, you don't really go on you know, you don't really create another, structure unless absolutely necessary. So, while Susan was playing the baby, was dressing the baby played on the floor. So, the baby that, it kind of remains there and played on the floor kind of becomes the next structure. This is how, late closure operates, this is basically what I was going to say here, so the late closure principle says: that the first organization will be pursued,

Refer Slide Time :(16: 25)



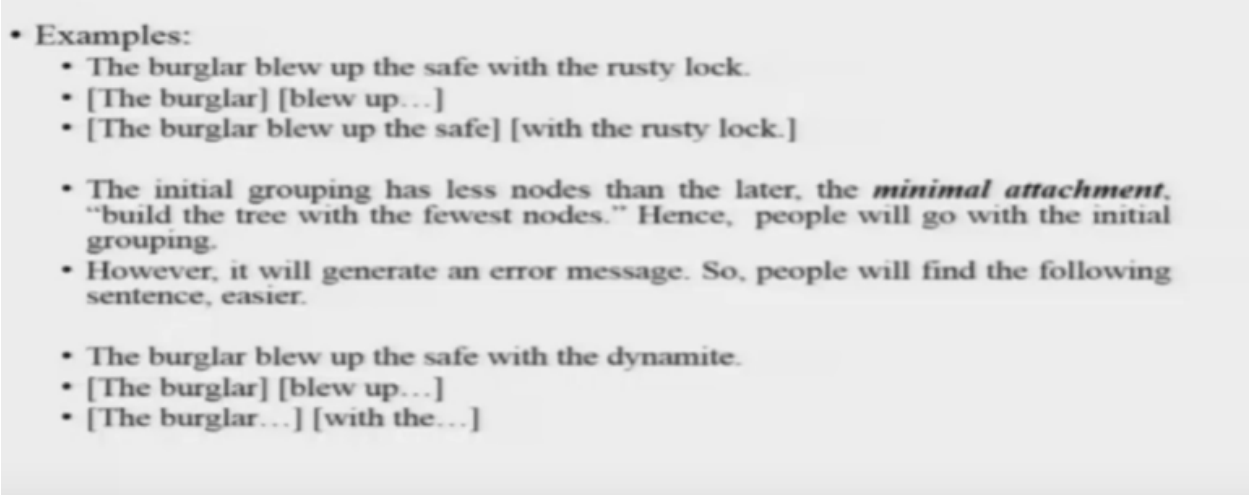
because it allows the parser to continue working on the same Clause, for a longer time. Okay? So, while Susan was dressing the baby played on the floor. However, you later realize that, you know dash played on the floor is not really correct and so, you will kind of you know, then move on to the second, but that will have an in the second iteration, if you are following the principle of late closure. So, this is, this is one, another heuristic: that we can take up is the heuristic of the minimal attachment, you do not, attach

Refer Slide Time :(16: 56)



so, when more structure, one more than one structure is licensed and consistent with the input, build the structure with the fewest nodes. So, in your phrase structure tree that you're developing, you should have least number of nodes, the more roads you have, the more complicated the structure is and the idea that, you know, the system seems to follow, is that of minimal attachment. Let us, look how that rolls out.

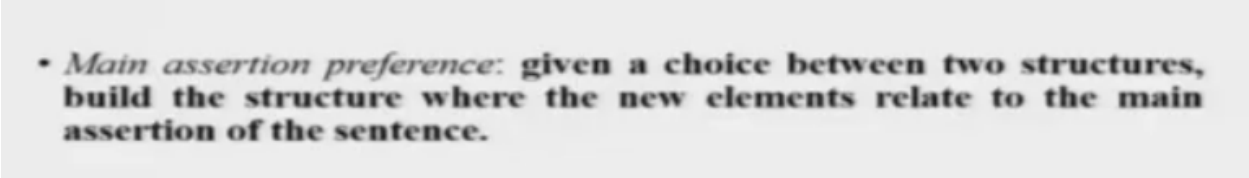
Refer Slide Time :(17: 22)



So, there is this example, the burglar blew up the safe with the rusty lock. Now, there can be many interpretations, the burglar can come up here, blew up the safe with the rusty lock can be the second one. The third interpretation could be the burglar blew up the safe and then with the rusty lock: that could be another one. Now, the initial grouping, the second one the burglar and blew up the safe it has only two nodes and less than the later, minimal attachment says, "build of the tree with the fewest nodes. So people here, by reading this sentence, we'll go up with the initial you know, grouping. However, it will generate an error message, why?

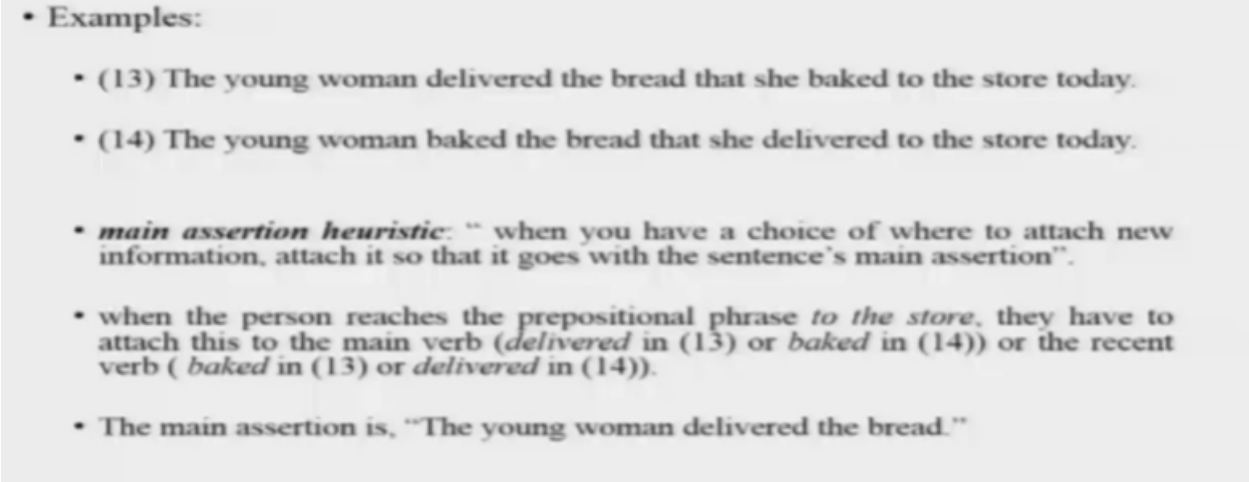
Because blew up, the safe with the rusty lock kind of candle okay, the rusty lock is not really a dynamite. So, you cannot blow up anything, with that rusty lock. So that kind of leads to incorrect interpretation, so people will need to revise it, when they revise it, then they will find the following sentence easier. So, the burglar so, if you have to really go with the minimal attachment, this sentence is the more plausible one, the burglars blew up the safe with the dynamite and then, if you kind of blew up with the burglar and then blew up the safe with the dynamite kind of leads to two kind of structures, lesson nodes. This is what will you know, give you the correct interpretation, because this is what kind of comes out with minimal attachment. Okay? With the, with the initial example that wouldn't have worked. So, moving on third heuristic that people might be following,

Refer Slide Time :(18: 45)



is that given a choice, between two structures, we can basically, build the structure, where the new elements relate to the main assertion of the sentence. So, the idea is given a choice between two structures we have two structures possible, we have to build a structure, where the new elements relate to the main clause or main idea, of the sentence, main assertion of the sentence. Let us look at, how that works out.

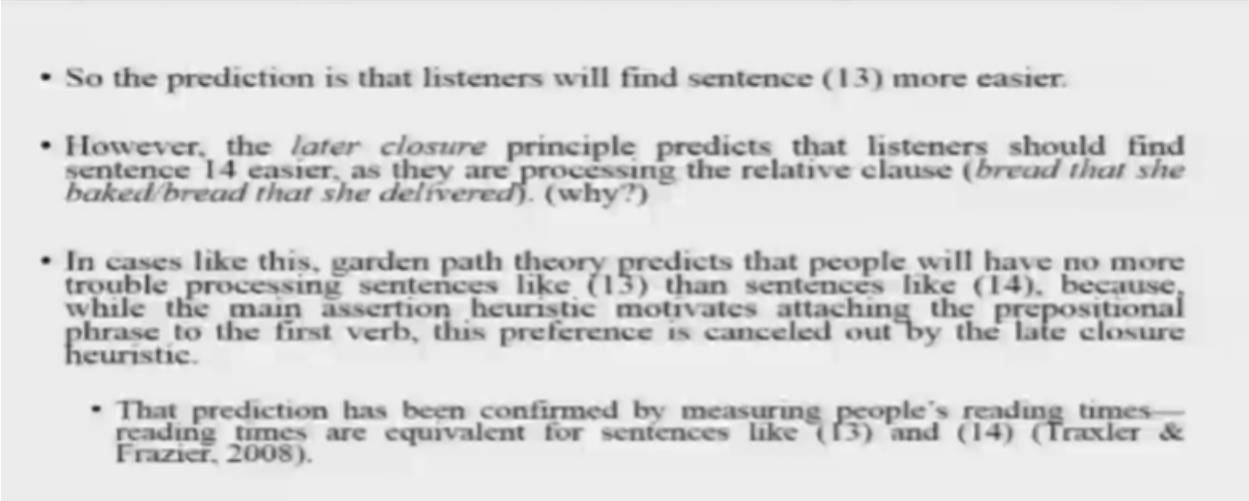
Refer Slide Time :(19: 11)



There is these examples, saying this is 13 and 14. The young woman delivered the bread she baked to the store today. Another example is, the young woman baked the bread that she delivered to the store today. Again once again, reading me sorry, yeah! The young woman delivered the bread that she begged to the store today, the young women baked the bread that she delivered to the store. These are two different sentences. Now, let's try and understand the main assertion thing, the main assertion is when you have main assertion heuristic says, when you have a choice of where to attach the new information, so in this indents, what is the new information? In the first one baked to the store, to the store today and the other one is delivered to the store today. So that is new information. Okay? So, how do you have to do it? You have to kind of attach the new information, to the main assertion, what is the main assertion in the two sentences? The first sentence, the main assertion is delivered, is in the second sentence, sentence 14, the main assertion is baked. So, the new information, delivered to the store today, can be attached to the women baked or to this totally can be attached to the main assertion delivered. Let us see, how that will work out. Now, being a little bit more closely, I was giving you all the options, when the person reaches the proper prepositional phrase, to this store, they have to attach, this to the main verb, delivered in 13, baked in 14. Okay? Or the recent verb baked in 13 and delivered in 14. So, there are two kinds of verbs there, the main assertion however is, the young woman has delivered the bread. Okay? Now, does the prediction is that, listeners will find reading you know, sentence 13 more easier. Let us read that again. The young women delivered, the bread to the store today. Okay? That she baked this there, the young woman delivered the bread: that she baked to the store today. So, to the store has to be attached to this one and then you can kind of get the correct reading, out of it. Okay? If you kind of do this attachment, in the fourteenth sentence, it'll create a problem, see how will that create a problem? There young woman baked, the bread to the store today, is going to create a problem. So, baked the bread that she delivered to the store today, in this one, the to the store today needs to be attached

to the more recent work. So, if you following the main assertion, you ristick the fourteen one will be more difficult, because you'll have to kind of first attach to the store to do Baked, you will realize that is not working, you will come back and attach to the recent verb. In the first sentence however it's very easy, to attach to the store today, to deliver and that is the correct easier, faster interpretation. Okay?

Refer Slide Time :(21: 52)



So, this is if you plainly go just by the main assertion heuristic, but obviously you know the other use six as well. So, the later closure, lead closure principle predicts: that listener should find sentence 14 easier, because they're, as they're processing the relative loss bread that she baked or bread that she delivered. Okay?

Refer Slide Time :(22:12)

- Examples:
  - (13) The young woman delivered the bread that she baked to the store today.
  - (14) The young woman baked the bread that she delivered to the store today.
  - ***main assertion heuristic***: " when you have a choice of where to attach new information, attach it so that it goes with the sentence's main assertion".
  - when the person reaches the prepositional phrase *to the store*, they have to attach this to the main verb (*delivered* in (13) or *baked* in (14)) or the recent verb (*baked* in (13) or *delivered* in (14)).
  - The main assertion is, "The young woman delivered the bread."

If you again lies, read this sentence the young woman baked the bread that she delivered to the swords, for later it says you kind of, keep it, on you kind of keep adding whatever you can't do the current phase. So, the young woman baked the bread that she delivered, to the store. So, you kind of till delivered you can go on, in one particular clause: that is how the late closure thing goes and in that sense reading 14 is easier.

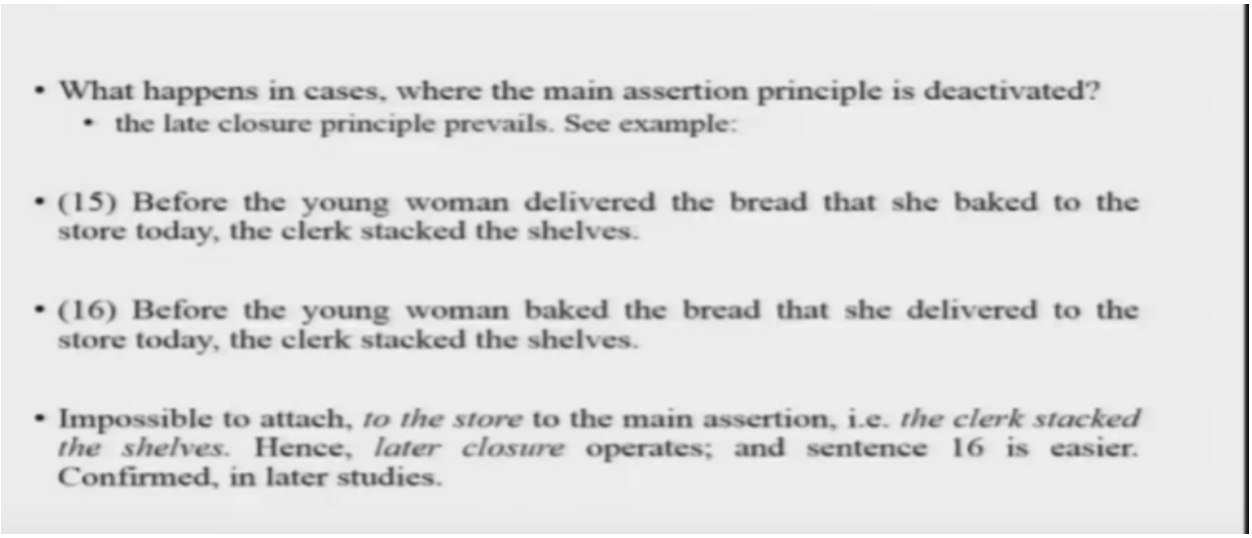
Refer Slide Time :(22:39)

- So the prediction is that listeners will find sentence (13) more easier.
- However, the *later closure* principle predicts that listeners should find sentence 14 easier, as they are processing the relative clause (*bread that she baked/bread that she delivered*). (why?)
- In cases like this, garden path theory predicts that people will have no more trouble processing sentences like (13) than sentences like (14), because, while the main assertion heuristic motivates attaching the prepositional phrase to the first verb, this preference is canceled out by the late closure heuristic.
  - That prediction has been confirmed by measuring people's reading times—reading times are equivalent for sentences like (13) and (14) (Traxler & Frazier, 2008).

So, in cases like this, this is a classic example, of a garden path theory, in cases like this the garden path theory, predicts that people will have no more trouble processing 13 and 14, because my main assertion 13 is easier, by late closure 14 is easier and both of them and the effects kind of seem to get canceled out by each

other. So that is, that is precisely, what is happening? And this prediction has been confirmed, by measuring people's reading times and what they found, is that reading times, are equivalent, for sentences 13 and 14. Okay? So, if one was easier, another was difficult, then 13 would be faster or 14 would be faster, depending on what you're just accusing but obviously the theory says that you are aware of both, kinds of heuristics and because you are aware of both kinds of heuristics, both sentences are equally, easy or difficult to read.

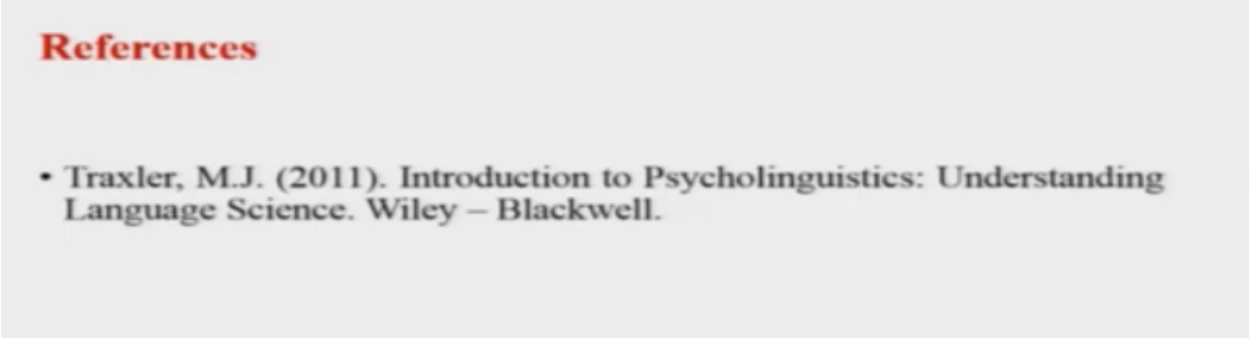
Refer Slide Time :(23:40)



Now, what happens in a case, where the main assertion principle is deactivated and obviously the late closure principle is going to prevail. Okay? Let's see another example, 15 and 16 sentences before the young women delivered the bread that she baked to the store today, the clerk's tagged the shelves. Before the 16, is before the young women baked the bread that she delivered to the store today, the clerk's tagged the shelves. So, you kind of it's basically, two different sentences and you have to kind of make sense of which one is easier to read. Now, if you see, the you know, if you go by main assertion, to the store, I cannot attach the clerk's tag the shelves, to the main assumption there, to the store the clerk started the shelves to the show cannot be attached. So, later closure principle operates and hence, reading 16 would be easier and just revising this. The clerk's tagged the shelves is the new information, you have to attach it to, the young women delivered: that is not really going to work out, but, the clerk started the shelves, the young woman baked the bread that is going to be easier. So, for example, if you are just reading by late closure, you can read 16 as, before the young women break the, bake the bread that she delivered, to the store today, the clerk's tagged this. So, this kind of leads to a longer read, longer clause, but makes the reading easier. If the main

assertion is not there, if their principal is not operating and obviously you will find sentence, a reading sentence 16 easier,

Refer Slide Time :(25: 05)



by the principle of late closer. That is all for today, I hope, I am not really confusing you a lot, this is a slightly dense chapter and you have to kind of you know, do a little bit of reading, then trying to understand. Obviously you can ask any number of questions you want, on the forum itself. Thank you.