**Applied Linguistics.**
**Professor Rajesh Kumar.**
**Indian Institute Of Technology Madras.**
**Lecture -29.**
**Language and Computers.**

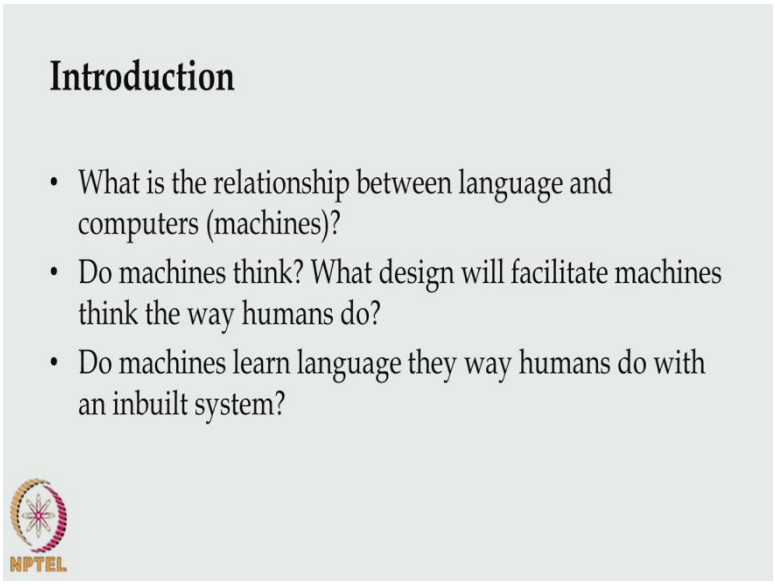
In our discussions in applied linguistics, we have reached to a stage where we are looking at applications of the study of language in somewhat greater details. We have looked at several aspects of social life where we see the applications of language in problem-solve and applications of language in making sense of the other domains of our lives in a significant way. Today, we are looking going to look at one more very significant domain, that is machines or computers. Where we see the applications of language in a very intricate and exciting ways. We need to look at couple of preliminary things to understand this this detail. Primarily we know that machines require a particular kind of language to interact with other machines or human interface.

There is a role of conversion of electrical signals and role of inputs in the machine and finally the one of the bigger goals in the domains of machines is to continue making machines more and more intelligent. We will look at the idea of an intelligent machine in some more details, however, in short it simply means machine that can process language to a greater complexity and yield simplicity for our interactions. The languages that are significant for machines to interact are called artificial languages. There are several names given to those languages. Then it comes down to the point that we are really talking about natural language and artificial language.

So, how does a machine handle natural language and what are the things required for machines to handle natural language is what goes in making machines more intelligent. We are going to look at some of the aspects both ways, where the study of language has already contributed a lot and also has potential to contribute to various domains of the study of machines and study of language through computers as well.

So, we have looked at some of these questions and yet we need to keep them in our minds all the time, what is the relationship between language and computers, is there any? Do machines think? Can they? What design will facilitate machines think the way humans do? Do machines learn language the way humans do with any innate system?

(Refer Slide Time: 4:03)




Answers to some of these questions are quite obvious, therefore we do not spend much time on these questions. However the totality of the answer requires attention, yet another larger goal is to develop a system, a machine which will think the way humans do, which will perform the way humans do. That is one of the ultimate goals of computer science in particular and engineering in general.

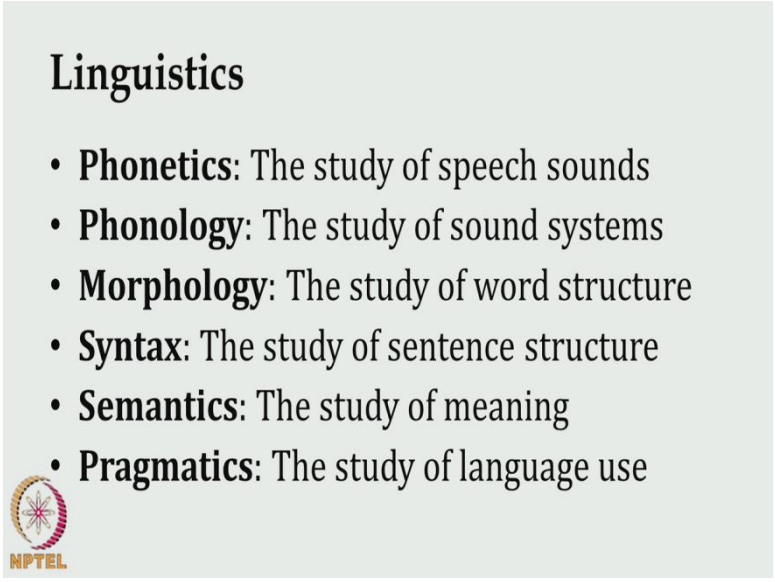
This is the area where the study of language has to contribute a lot because language is what fundamentally defines humans. Therefore, to make a machine work the way humans do, we need to make machines more compatible with natural language, more intelligent in processing natural languages, faster, intelligently and the way human mind it does.

This has been in short a long-term goal of scientists. We have come a long way, there are miracles that machines can do in our modern times. We have seen, rather we have lived, at least some of us have lived the change in the paradigm. We have seen the difference making an impact in our lifetime. We did not have roles of machines, computers and computer-based programs in our lives the way we see these days.

That is 25 years ago, things were very different, in the last 25 years, all kinds of inbuilt, underlying advancements and the way we communicate, the technology that we use, the based on the point in making machines more intelligent. It is not an exaggeration, the fundamental point is machines are more intelligent in dealing with natural languages and therefore the perform more.

Of course there are advancements in the areas where they do not have to interact with natural language as well. So, how do scientists achieve that goal? And what is the role of language in that? We are going to touch upon these things very briefly in understanding how, what is the role of language in computer science and what is the role of computers in understanding more details about language.

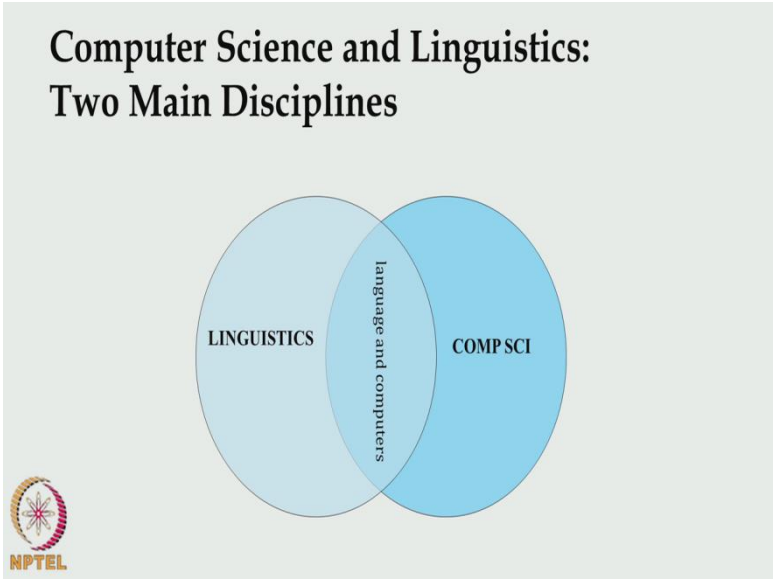
(Refer Slide Time: 7:26)




When we look at the discipline called linguistics where we study in a very systematic way several branches, several factors, several aspects of language like phonetics, phonology, morphology, syntax, semantics and pragmatics, to name just a few, there are more aspects, more branches of study which deal with other aspects of the study of language.

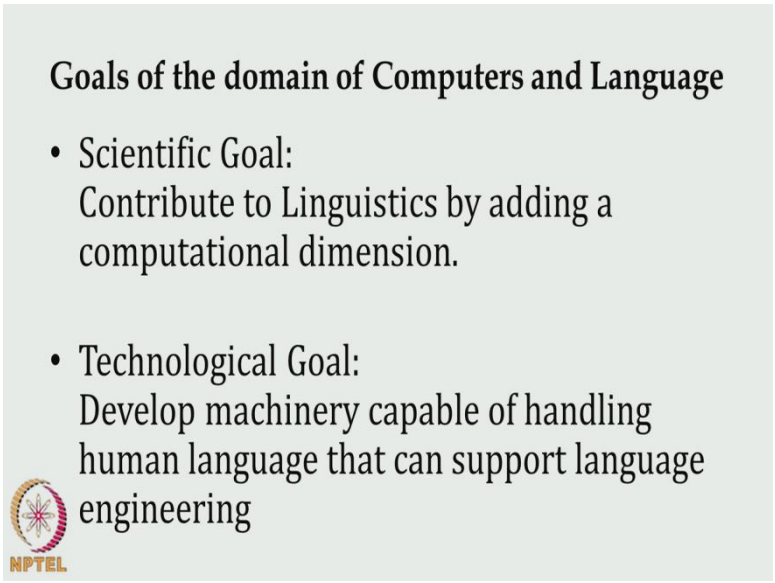
But primarily the study of the sound system, study of the sound production system, study of the word structure, sentence structure, how they contribute in making meaning and all of these put together, how do we use language in the real world. The use of language in the real world has lot of underlying algorithms inbuilt, however when we try to separate them, the research that has been in the domain of these branches of linguistics have uncovered a lot.

(Refer Slide Time: 8:55)




So, on the basis of these things, when we try to see interactions between 2 main disciplines when we talk about in the area intersection between language and computers, we look at 2 things, 2 larger domains, that is not to say other fields of study has not much to do with this intersection. However, linguistics and computer science have to come close. And the intersection of the 2 is where the role of language has to make an impact.

(Refer Slide Time: 9:28)

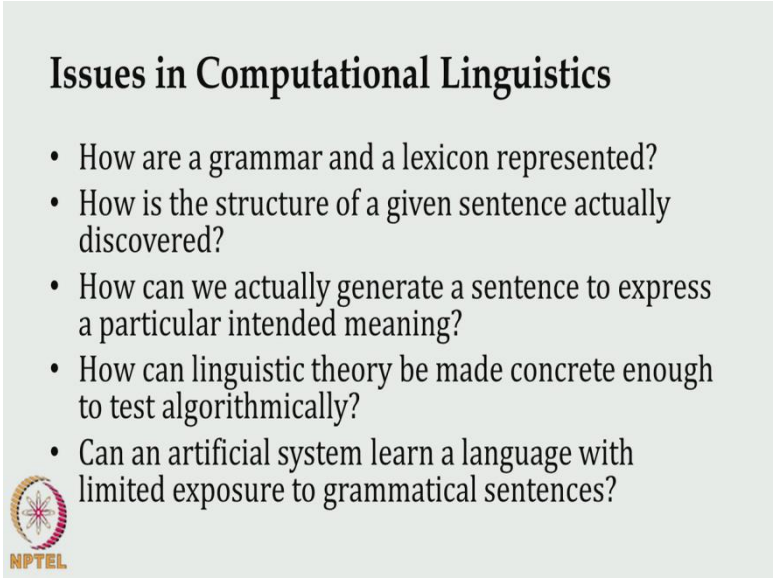



If we having looked at the fundamentals of linguistics and the study of language, when we try to understand the scientific goal and technological goal for the study of computers and the intersection between competence in language, when we try to look at the scientific goal, we

reach that it contributes to linguistics by adding a computational dimension. So, the use of machines in study of language is one of the scientific goals of computer science.

However, when we look at technological goal, it develops machinery capable of handling human language that can support language engineer. So, we have been discussing about the domains of language technology and language engineering so far, how machinery becomes capable of handling human language to support the kinds of things machines can do with human interface is the technological goal of this larger domain.

(Refer Slide Time: 10:37)




There are several issues that are interesting in computational linguistics. So, the intersection of linguistics and computer science is called computational linguistics. In that intersection area, there are several interesting questions that we need to keep in mind. Such, the questions are listed on your screen and think about these questions and you can come up with more than we are discussing here today.

How are grammar and lexicon represented? In short, how is grammar represented? What is the meaning of representation of a drummer, that is we are familiar with the grammar and the grammar of language in human mind when they are trying to look at application through aspects of such understanding, we want to see how is that grammar, how can that be represented in machines.

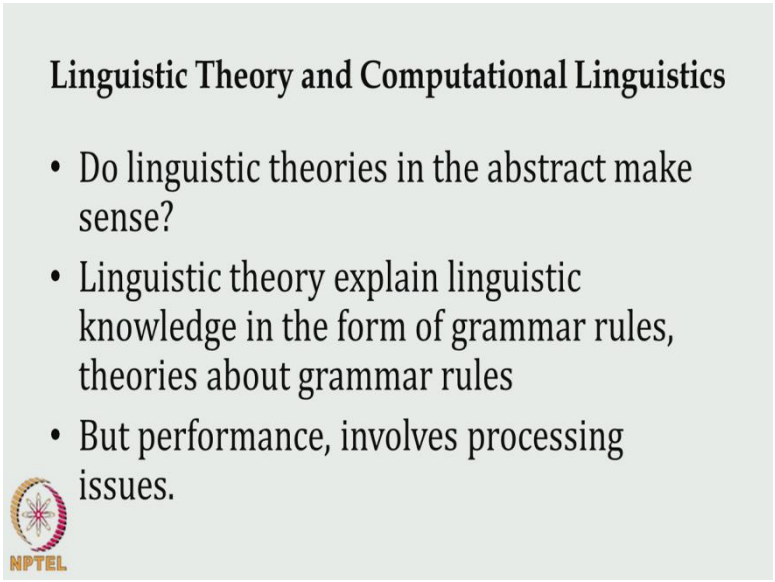
How can machines be made intelligent to represent that grammar? How is the structure of a given sentence actually discovered? How does machine deal with this structure? Or the

computational algorithms underlying the structure of a sentence, how does the machine deal with that? How can we actually generate a sentence to express particular intended meaning? So, this is about ambiguity.

Lot of sentences we speak of full of ambiguities, however human mind has no difficulty making the correct sense or the intended sense obvious from several available interpretations of a sentence. Can machines really do that? Can they perform, and if they cannot, how can we make them do so? How can linguistics theory be made concrete enough to to test to test algorithm? How can they be simulated?

How can they be designed in a way that it could be transmitted and translated to machines? Can an artificial system learn a language with all that and even if we end up doing all that, is it possible for an artificial system to learn a language with the limited exposure to grammatical sentences? These are much complex questions in the domain of computational linguistics. And in simpler way, when we interact and when they want to study the 2, that is machines and language together.

(Refer Slide Time: 13:25)




Do linguistic theory in abstract makes sense to machines? It does make sense in the sense of, sorry, in the domain when we study human mind, it has contributed a lot. Insights that come from study of language has contributed a lot in our understanding of how our mind functions. How it is structured, how the study of the underlying structure of language has contributed in building, we have explored the underlying structure of given language and that has contributed a lot in construction of linguistic theory. And the 2 together has helped us

understand the functioning of human mind to great extent. And that does that have potential to deal with machines as well, that is what this question means.

Linguistic theory explains linguistic knowledge in the form of grammatical rules. It is theory about grammatical rules, but performance part involves processing issues. We are again talking about the distinction between I language and E language. Can I language be simulated is the question. And the E, the areas of E language where processing issues are involved, can they be simulated as well?

Because dealing with the areas of E language, that is performance part, that is the use of language in real-world and also that deals with lot of issues that come from our understanding of pragmatics. And without machines being in a position to be dealing with such issues, we do not see machines doing much. So, there are some of the areas in which we have applications of these things. That is where an intelligent machine can help us perform fast.

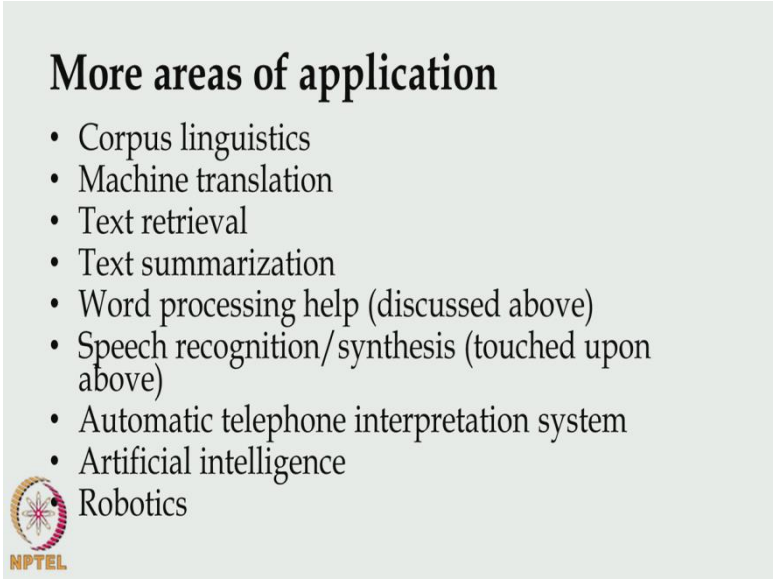
(Refer Slide Time: 15:33)


The areas like language teaching or some of the medical areas like speech therapy for aphasic and others with difficulties. Now, if you put the 2 together, that is how machines help us in dealing with speech therapy, particularly for aphasics and also people with other difficulties, can teaching be made faster? There are a lot of studies in this domain and it has yielded positive results in this area as well.

So, the one way to start is to look at computer assisted language learning and also the areas of disabilities where the applications of machine-based learning programs has yielded a lot in

people with difficulties learning language. There are certain computer-related applications in the domains like computational linguistics that we are talking about, natural language processing, language engineering and computer simulations of human language processors.

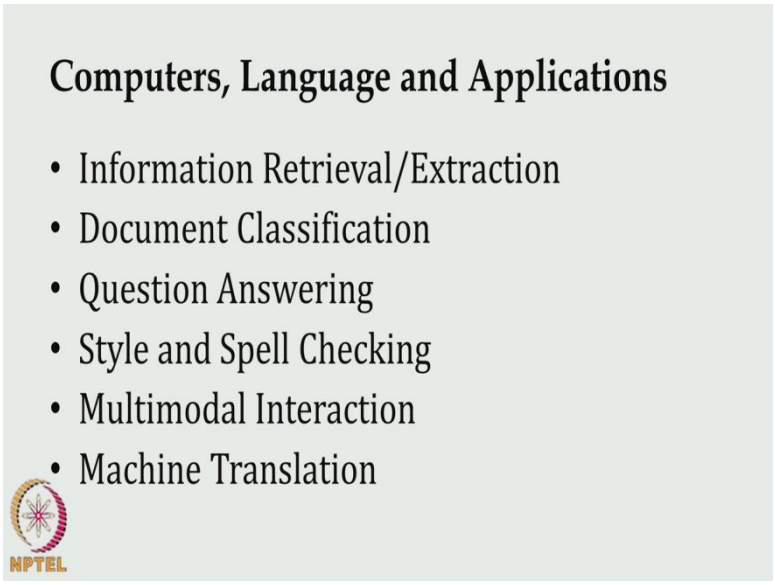
(Refer Slide Time: 16:57)




There are overlaps between these areas, however they are distinct from one another and each one of them, we see the application of the things that we are talking about. So, there are some more areas where we see the applications of this aspect of machine where it requires language and insights from the study of language are corpus linguistics, machine translations, text retrieval, texts summarisation, total put together and more are parts of data mining, speech recognition, automatic phoning interpretation system, artificial intelligence and Robotics, just to name a few.

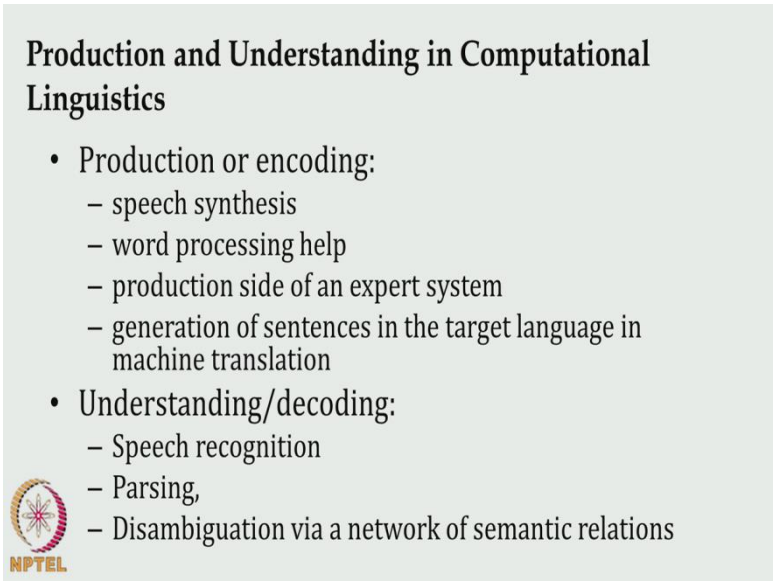
The more we see, the more we find the presence of intelligent machines, the more the areas we can come up with where we see applications of these things. Therefore the domain of the applications of intelligent machines is unlimited and it more and more areas can be explored if it is more intelligent and in order to do so we need a solid foundation in dealing with language.

(Refer Slide Time: 18:02)




Information retrieval or extraction document classification, question and selling, style and spellchecking, multimodal interactions, machine translations, all are the areas where we see applications of language. Let us get to the point where we see a role of production and understanding income potential linguistics that is in are not in dealing with intelligent machines.

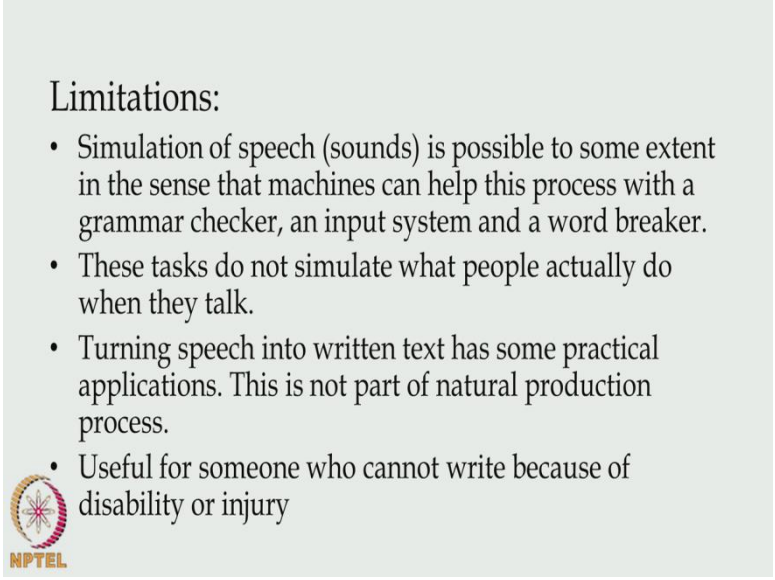
(Refer Slide Time: 18:30)




So, the technical terms in computer science for production is encoding. It requires speech synthesis, word processing help or it can do these things. Production side of an expert system and generation of sentences in the target language in machine translation. Whereas

understanding part or the decoding part has role in speech recognition, parsing, disambiguation that is disambiguities and it has to do with the network of semantic relations. So, these are the 2 parts of a machine which are also important for us to understand the production or encoding or understanding and decoding.

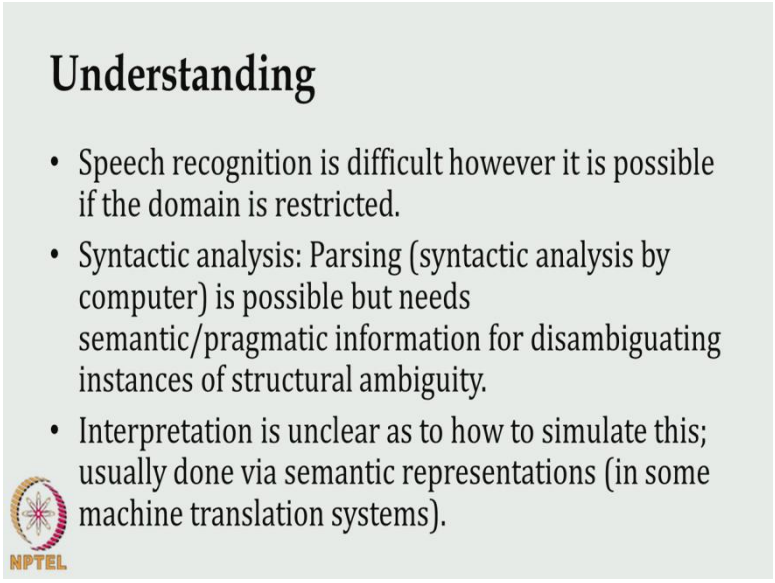
(Refer Slide Time: 20:00)




There are certain limitations of applications as well where simulation of a speech, that the speech sound is possible to some extent in the sense that machines can help this process with the grammar checker as input system and a word breaker. These tasks do not simulate what people actually do when they talk. That that is a very complex area to simulate and that is one of the roadblocks here.

So, turning a speech into written text is definitely one of the practical areas of applications and this does not become the part of the natural production process, yet we see the applications in this area as well. It is useful for someone who cannot write because of certain types of disability and injury. So, just imagine, it is going to revolutionise the voice life and functioning of people with certain kind of speech disabilities.

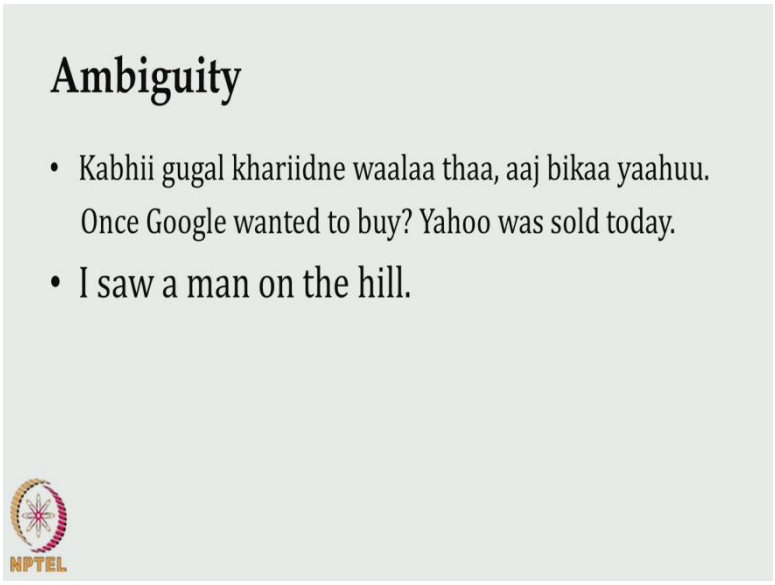
(Refer Slide Time: 20:39)




If speech to text converter is available and can be made more user-friendly with the simulation of applications of language to a greater extent. The speech recognition is difficult, however it is possible if the domain is restricted. Now, how to remove these stations in speech recognition will again revolutionise the world. It has potential to revolutionise the world.

Syntactic analysis by computers and for a machine to understand the chance of a sentence is possible but it needs semantic and pragmatic information for dealing with ambiguities. And these are encoded in a structural part of language. We will look at an example of this. So the interpretation is clear as to how to simulate this, usually than by a semantic representation in some machine translation system.

(Refer Slide Time: 21:29)

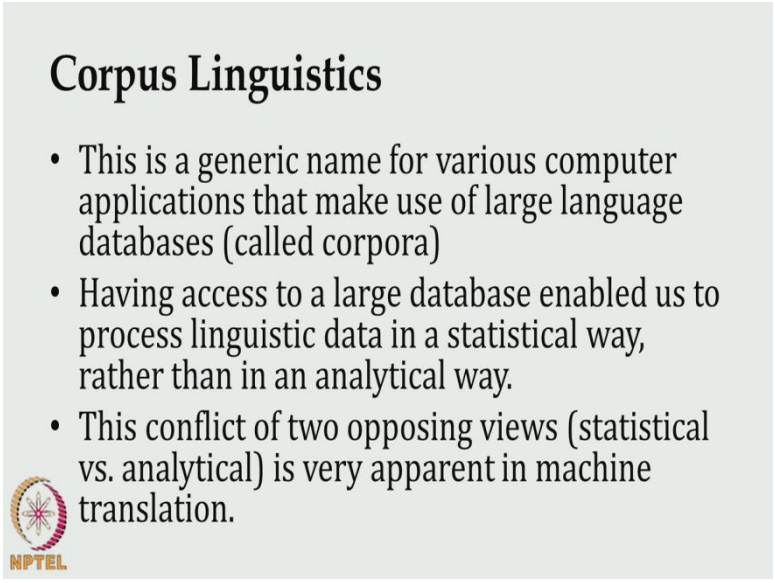



So, we were just looking at the area of ambiguity. Take an example of these 2 sentences, one from Hindi and the other from English. The Hindi sentence is an example that comes from the headline, in fact it is today's headline of a Hindi daily, very famous Hindi daily, Navbharat Times. One of the headlines of this from the front page says Kabhii Gugal kharidne waala tha aaj yahoo bika.

It means once Google wanted to buy, Yahoo was told today. Now, if you talk to Hindi speakers about the 1st part of the sentence, Kabhii Google kharidne waala tha, it is not clear whether Google was going to buy Yahoo or Yahoo was going to buy Google. That ambiguity is underlying in this Hindi sentence. Just to give a clarification note here, the 1st Hindi sentence we have tried to put it in little bit of transliteration to make it readable for non-Hindi speakers and then the translation is given below.

So, that part of ambiguity is clear, when we say things like I saw a man on the Hill in English, it has 2 interpretations as well, whether I was on the hill where I saw the man or I was somewhere else and I saw the man on the hill. This ambiguity is encoded in structure and how human mind performs parsing, that is how it recognises junks can be captured with the generative structure. Can machines be made that intelligent where it can do the structural parsing which will help disambiguate these embedded underlying interpretations?

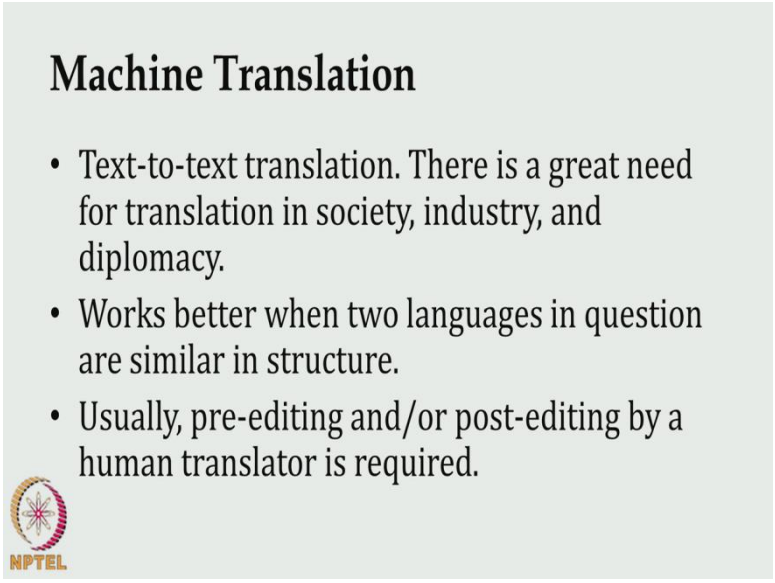
(Refer Slide Time: 23:38)




There are some of the other areas related areas where we see the applications of these interaction between language and computers and we are going to be talking about some such areas like corpus linguistics. What happens in corpus linguistics? This is a generic name for various computer applications that make use of larger language database called corpora. So, when we have access to corpora and such access enables us to process linguistic data in statistical way, rather than in analytical way.

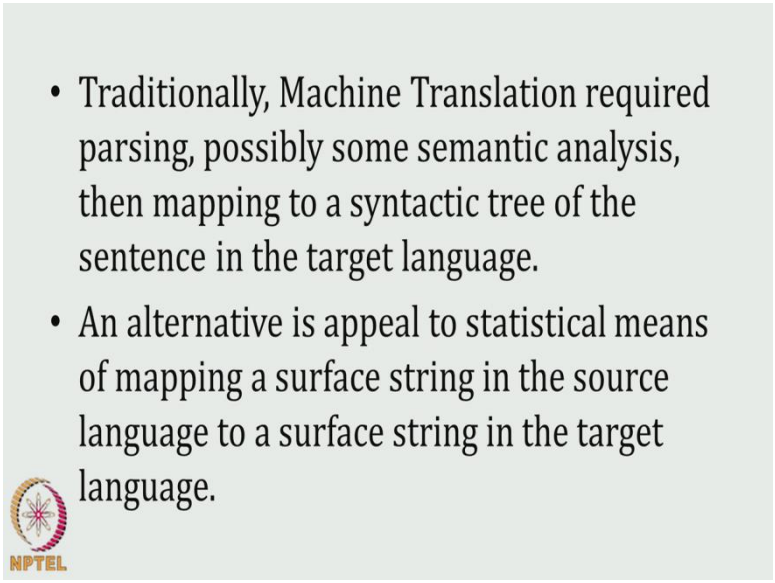
Now is the analysis of linguistic data from corpora possible? This conflict of 2 opposing views, that a statistical versus analytical is very apparent in machine translations and lot of time becomes roadblocks. The way we can deal with a larger database statistically if we can deal with that analytically, that will be a paradigms shift in field of computational linguistics which is largely applied linguistics.

(Refer Slide Time: 24:54)




When we do translations, that is machine translation in terms of text to text, there is a greater need for translation means it has emerged from the requirement of translation in the society, industry and diplomacy. It works better when 2 languages are similar in sentence structure. Usually pre-editing or post-editing of a human translator is required but can we remove human interface in text to text translation at least? Speech to text conversion and then translation will add more dimensions to the domain of machine translations.

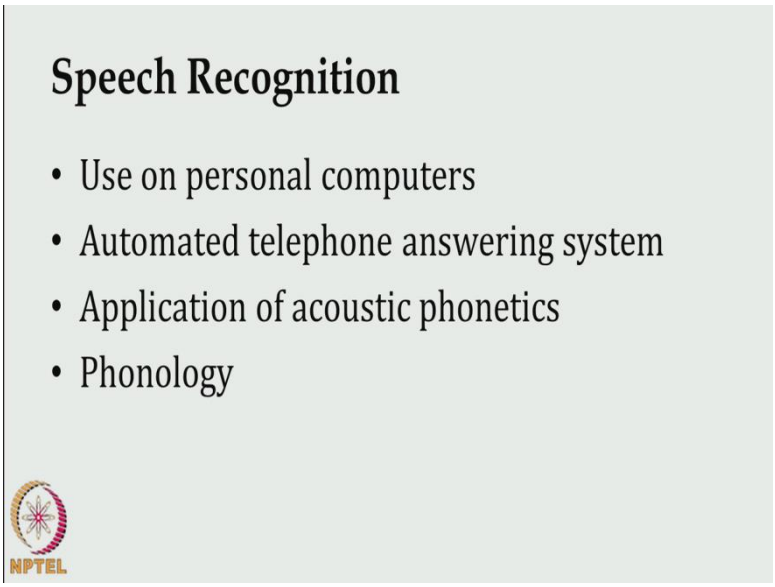
(Refer Slide Time: 25:44)




What traditionally it has required was possibly some syntactic analysis or some semantic analysis, then mapping to our syntactic tree, the kinds of discussions that we were doing a

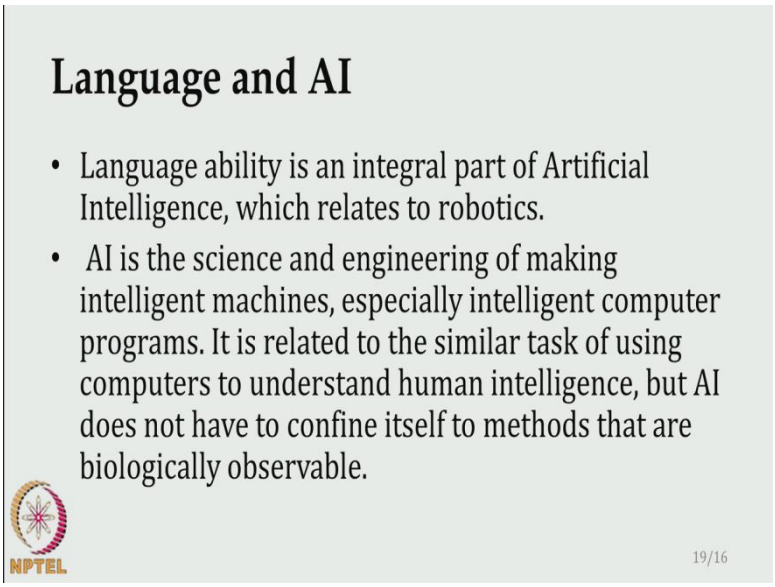
few minutes ago, of a sentence in the target language. But an alternative is to appeal to statistical means of mapping surface string in the source language to a surface string in the target language and this is where it requires more modifications.

(Refer Slide Time: 26:20)




Speech recognition by a machine is yet another area where we see applications of language where we use this on personal computers, automated telephone answering systems, application of acoustic phonetics and understanding of phonology becomes very significant in this domain of speech recognition.

(Refer Slide Time: 26:39)



What is, there is another area of artificial intelligence and an interaction between language and artificial intelligence and an interaction between language and artificial intelligence in the sense that we were discussing comparison and contrast between natural language and artificial language is yet another domain when we see these applicational aspects of these stuffs. Language ability is an integral part of artificial intelligence which is, this area to great extent is, has also to contribute to robotics where machines not only function in terms of internal processing but also does some physical activity as well.

So, what is artificial intelligence? In short it is known as AI and it is science and engineering of making intelligent machines, special intelligent computer programs, it is related to similar task of using computers to understand human intelligence. To match machines to human intelligence is the goal of artificial intelligence.

But AI does not have to confine itself to the method that are biologically observable alone, rather it has been functioning so far on the basis of mathematical simulation. Simulation of biologically observable stuff will be yet another challenge and probably can have revolutionary steps in this. So, with this we have seen some of the areas or the domains where we see the impact of the study of language and computers together.

The study of language and computers has resulted into emergence of several new disciplines of a study. It has revolutionised the way we live now and the way we do things in our society. And as we have underlined and tried to understand a larger domain of scientific community is to come up with a machine that that can function the way humans do. Once again to underline the significance of language and the study of language that has emerged from systematicity coming from the domain of linguistics has to contribute a lot in this domain.

I invite you to think about other related computation issues and come up with solutions, some of the areas where you see the use of language in more significant ways. Just thinking about those domains and also take you to some extent will take you forward in understanding the notion of language and computers in the larger domains of applied linguistics. Thank you.