Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute, Madras

Lecture – 63
Introduction and Historical Approaches to Machine Translation

We are going to be starting a new session today on the topic of machine translation. I am sure you would have noticed that we have been progressing from the understanding of the corpus, then finding out the frequency of words and then we tried to find out the meaning of embedding. Try to understand the context of the words and then try to get the neighborhood words in terms of the; in terms of the word embedding.

And then later we try to formulate newer machines or newer machines in terms of neural networks to really do the same job. And then later we try to address from the words to the sequences during accrual neural networks. I am sure you have been noticing that progression right, from the words to the word embedding to the sequence of words and so on.

So, what is the next stage in this understanding is to really understand a sequence of one language and then translate it into another language. We have been considering only the language English for our purpose so far. Any language that is very similar to the English structure, can also be used in this scenario that we had discussed ok.

Now what we want to do is we want to find out if the mechanism that we have understood the technologies that, I have that we have understood, the fundamental that we have seen so far could it be used for machine translation. So, this is going to be an interesting topic that we want to take up today. So, what are the important things that we have seen so far right?

So, trying to understand the meaning of the words without even looking at the dictionary; trying to generate sentences and then try to figure out the probability of a sentence formed using the machine or the neural networks or using some probability models right. So, now, we are going to take it to the next level in terms of identifying a new model where we can feed in a text in one language and we want to get the translated version through that model.

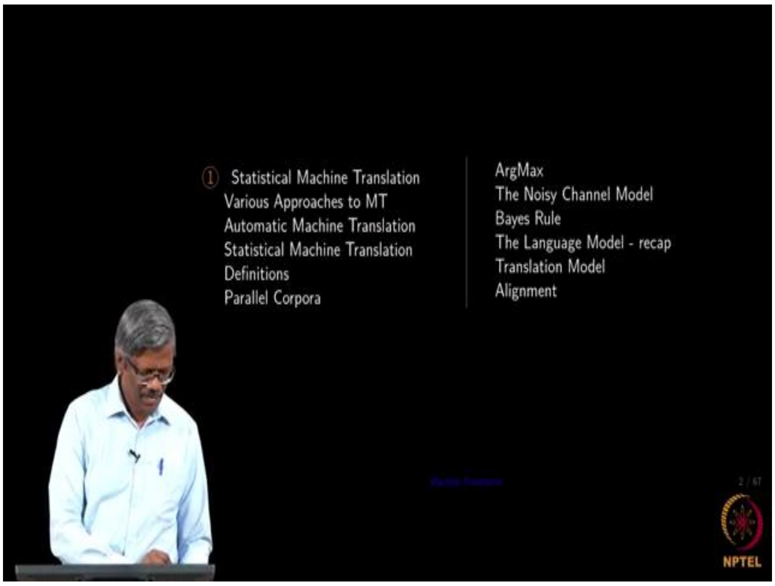
So, that is going to be a very interesting task and a very challenging task. And this is not very new people have been trying to do the machine translation for over 50 plus years. When they started the machine translation, the power of the machines was too small even you know some of the digital watches that we have today, they are more powerful than the computers that they were they had earlier.

So; that means, more and more processing can be done then more and more corpus can be processed. Parallel corpora can be constructed and processed and automatic ways of extracting words and translations are going to be a reality very soon. Even though you have seen that these techniques that we had seen so far have not been successful to the level of 100 percent.

There are some elements of error here and there we are still improving those, I am sure in about 20 plus years we will have very good models that would be able to do all the jobs with the 100 percent confidence and so on right ok. So, in order to understand how people have been doing machine translation over the last 50 plus years; we will first look at the basic element that they had looked at and how they progressed.

And then later we look at some statistical models or statistical machine translation aspects and later we will try to apply a neural model into and see how we can successfully translate sentences in one language to the other language alright ok.

(Refer Slide Time: 04:49)



1 Statistical Machine Translation
Various Approaches to MT
Automatic Machine Translation
Statistical Machine Translation
Definitions
Parallel Corpora
ArgMax
The Noisy Channel Model
Bayes Rule
The Language Model - recap
Translation Model
Alignment

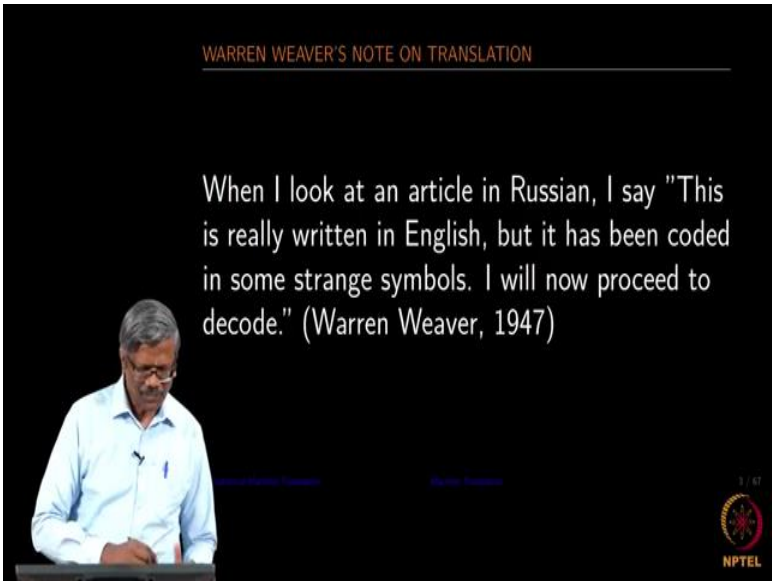
2 / 67

So, let me take you through the content that I am going to talking about today. So, some of the element that I am going to be talking about is on these statistical machine translation part ok. So, we will first look at various approaches followed by people to machine translation. And then we will look at how we can automatically translate one sentence to the other.

Then later we look at some parallel corpora that are really required for us to do this statistical machine translation ok. So, what we mean by parallel corpora and then we try to apply the model into a noisy channel model those of you who have studied in information theory you would know this very well. We look at the base role which is very fundamental to machine translation.

And then we will just have some recap on the language model because it is very important for you to know, the sentence that you have created really is something that we can use you know using some probability models. And then we will see what actually is a translation model; we also look at something called alignments and then in the latter part of the lecture we will look at the neural models for doing all these alright.

(Refer Slide Time: 06:20)



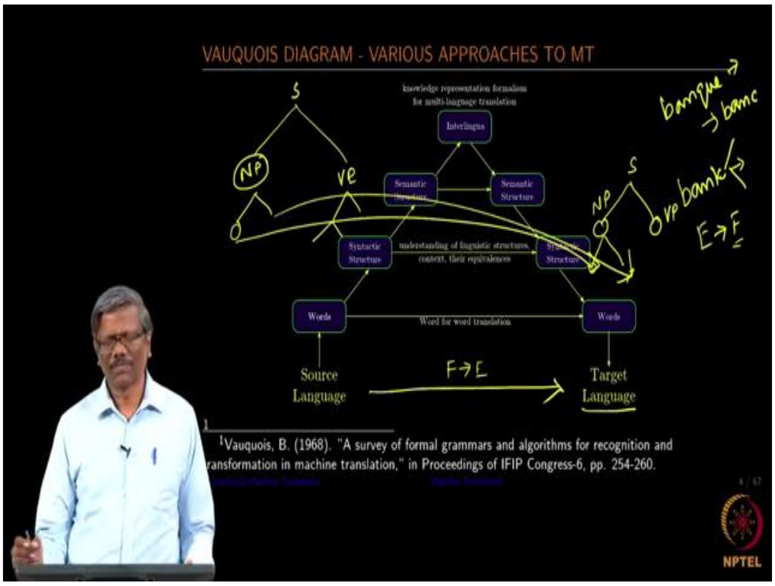
WARREN WEAVER'S NOTE ON TRANSLATION

When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode." (Warren Weaver, 1947)

So, this is something that you should be aware of ok. So, this person Warren Weaver in 1947 mentioned that I think its the beginning of the cold war. He said when looks at an article in Russian this is really written in English, but it has been encoded in some strange symbols I will now proceed to decode. So, this is how he looked at the translation from Russian to English ok.

So, what he says is the text that is really written or that you see as Russian is written in English when it comes through some channels right, it becomes Russian and then I want to estimate what was originally written and that is what he wants to decode. So, that is what he said and then after this flurry of activities started in the natural language processing aspects of machine translation.

(Refer Slide Time: 07:30)



VAUQUOIS DIAGRAM - VARIOUS APPROACHES TO MT

[1] Vauquois, B. (1968). "A survey of formal grammars and algorithms for recognition and ransformation in machine translation," in Proceedings of IFIP Congress-6, pp. 254-260.

4 / 67

So, this diagram called Vauquois diagram discusses various approaches to machine translation in one shot ok. Rather it gives you the various approaches people have followed to machine translation in one simple diagram. If you look at this diagram on the left side you see this source language on the right side you get the target language right.

So, how do we really take the source language and then translate it to the target language? So, one way to look at is we have a sentence let us say in English and I want that to be translated into French or vice versa right. So, we will call the source language let us say as French and then the target language is English. So, in this case we are going to be translating from French to English right.

So, what are the ways that we can do it to if you do not know the language the target language what you do? You get the dictionary the English or French to English dictionary. Look at the source language sentence take one word at a time and then find out what is the actual meaning of that in English and then write that word. And then take the next word and then write the next word in the target language and so on right.

So, once you converted all the source language words into or the target language you have a bag of words ok. I am using the term bag of words because this structure of the language on the left side could be very different from the structure of the language on the right side. So, once you have translated rather converted all the source language words into target language word what you do you just rearrange them so that it makes sense ok.

So, how do you know that you really translated the real meaning of the source language into this it is not possible to really know if you do not know the source language? So, you need to go to somebody who understands both languages and then ask them whether the translated version is right or wrong. Let us assume that he is the person who has a gold standard who understands both languages and translates from source to target very well.

So, you take this sentence to him and then that person validates saying that no this word is not to be translated in this fashion it is different for example if I use. So, in case of English if you have the word bank ok, you fit up here. So, for the sake of understanding this I am just assuming that if we are going to be doing the translation from English to French.

The bank we have seen earlier that it means different things based on the context right so, this is polysemy right. The same word meaning different things it could be river bank or it could be the financial institution or it could be a verb you know I bank on you to do certain things ok. So, what do you really how do you really translate banks into the respective word in French ok?

So, there could be different words associated with the polysemy of English. So, we should know the context and then based on the context we can probably. Say I am just making it off it could be a ban n k, which will mean the river bank or the pardon me I do not know the French I am just making it up ok, this could be the financial institution this could be the river bank.

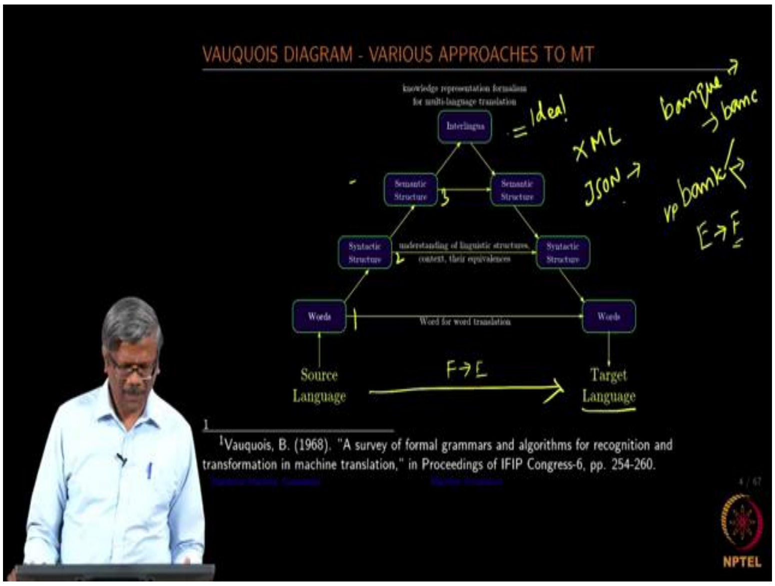
So, the person who understands the language would really translate the word, and then based on the context he will use the appropriate word in the translated sentence ok. So, is that really possible to convert all the languages in this fashion word by word without knowing the syntax structure? It is not possible right. So, that is why we go to the person who we assume as a person holding the gold standard of translation ok. So, it is almost impossible for you to do the translation without understanding the syntactic structure of both languages ok. So, what is the next level that people have done the next level is, they have taken to the syntax structure level for example, you can construct a syntax tree ok. So, you can construct something like this a sentence could be your noun phase verb phase and then you can still do all this right.

So, if I am able to construct my sentence in this fashion and then assuming that the language that I am going to be translating the source into the target also has a similar structure. If I have a similar structure then I can take each one of those that are in these slots or nodes into the respective node on the right side so here like this. Suppose if this target language also has the same structure so, I can take it like this ok.

So, is there a difficulty with this I think there are some difficulties with this because sometimes the words are swapped; you know before or after depending on the language every language is very different in their own way. So, we cannot expect that to be very similar to what we have seen that in English ok. So, what you do is you construct a very similar structure syntax structure.

Or the tree which need not be the same as what we have constructed for the source language and then try to map that in. Still there would be a problem because we do not understand the semantics of that right. So, what you do you instead of doing this you go into the semantic word try to understand the meaning of the sentence, and then based on the meaning of the sentence do the translation.

(Refer Slide Time: 14:39)



VAUQUOIS DIAGRAM - VARIOUS APPROACHES TO MT

[1] Vauquois, B. (1968). "A survey of formal grammars and algorithms for recognition and transformation in machine translation," in Proceedings of IFIP Congress-6, pp. 254-260.

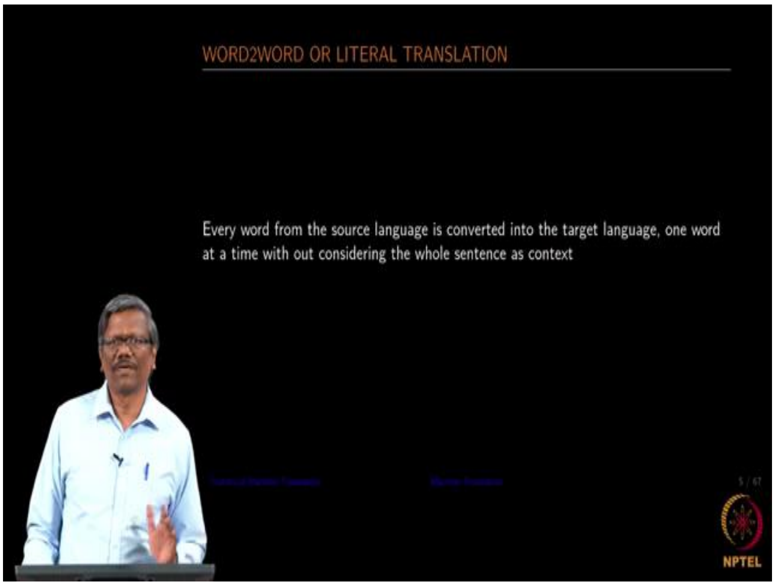
For this again you require a good understanding of both languages correct then if you want to do any of this right. So, whatever you have listed in 1, 2, and 3 we can only do translation from one language to the other. So, there are difficulties people have been attempting this and they have not been very successful in 1, 2 or 3 so far. So, its a very difficult task its a very intellectual task you know well right.

Then, what they also attempted was to try to create an interlingua. Interlingua is something that represents the knowledge of the language its a very in neutral model where you can represent a language. And when you get a source language you can translate that into that interlingua model. If I create this interlingua I can say translate to interlingua to French to German to Spanish to Hindi, Tamil or whatever ok.

So, creating the interlingua is a very critical aspect. So, is it really possible to do that? I think a lot of interesting examples are available on the internet in the scientific research community. You may want to look at that and see how they really create the interlingua its very similar to what we have done with XML or JSON right now you would know.

So, it is earlier now we used to have our own way of passing the packets from one application to the other. So, instead now what we do is we use an XML mechanism or a JSON mechanism to read the XML or decode the write the XML or decode the XML in the same fashion JSON also can be used right. So, in the same fashion we create something here very similar to this model; and then try to use various translators to make use of the interlingua to translate it into their own language. So, this is a very ideal model correct people have not been very successful in doing that so far alright.

(Refer Slide Time: 17:32)



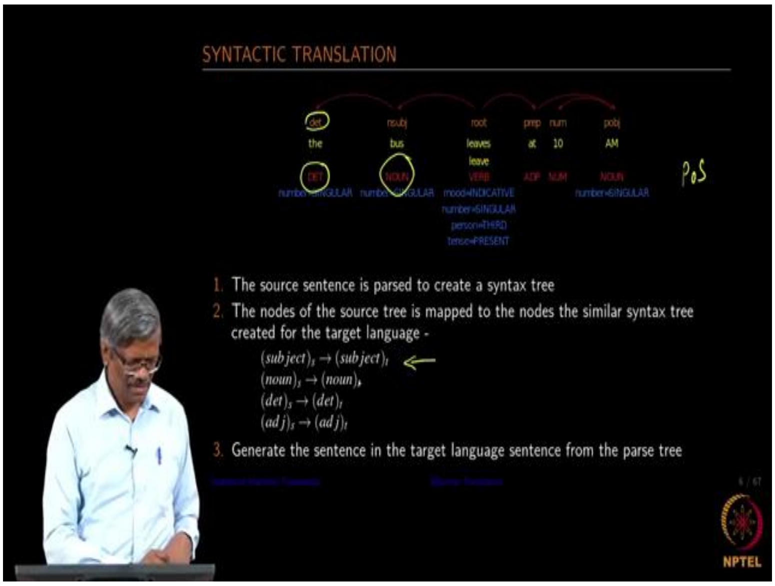
WORD2WORD OR LITERAL TRANSLATION

Every word from the source language is converted into the target language, one word at a time with out considering the whole sentence as context

5 / 67

So, as I mentioned earlier the word to word or literal translation is done, by taking every word from the source or then convert that into the target language. And take one word at a time and try to construct the sentence in the target language without really understanding the context.

(Refer Slide Time: 17:54)



SYNTACTIC TRANSLATION

1. The source sentence is parsed to create a syntax tree
2. The nodes of the source tree is mapped to the nodes the similar syntax tree created for the target language -

   $(subject)_s \rightarrow (subject)_t$ ←
   $(noun)_s \rightarrow (noun)_t$
   $(det)_s \rightarrow (det)_t$
   $(adj)_s \rightarrow (adj)_t$
3. Generate the sentence in the target language sentence from the parse tree

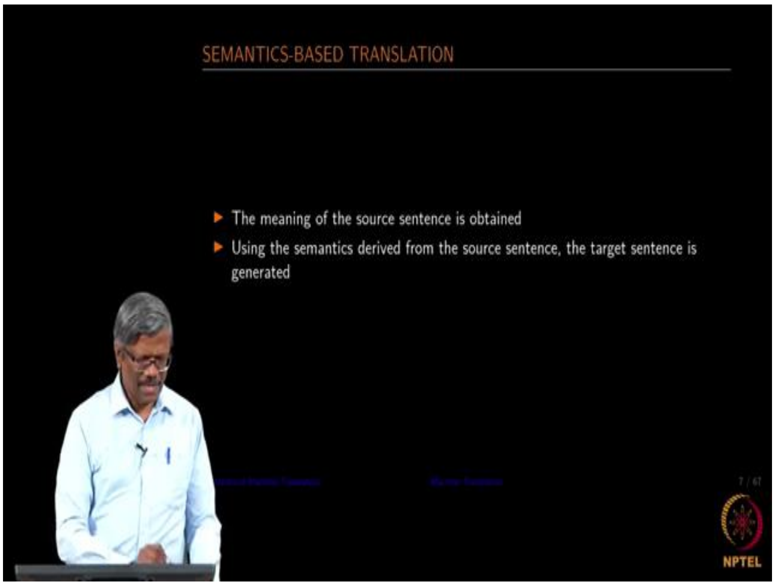
6 / 67

This is more of a dependency graph rather than a syntax tree it is a little more advanced where you can also find out how these parts of speech are connected ok. So, for example, the determiner here I have the bus leaves at 10 AM. So, I have actually used Google API to do this, I gave the sentence to Google API and then it gave me this graph ok.

So, the sentence I used was the bus leaves are 10 AM. So, you can see that its ability to really create the parts of speech right. So, in the syntactic translation what you do is you have another dependency graph created for the other language in a similar fashion. And then move the words into the respective slots ok. So, as I have shown here you move the source of the subject of this subject to the target subject of the target.

And then noun from the source to the node where the noun is represented in the other graph and so on. So, we can generate a sentence without really understanding what that sentence is talking about again the context is not considered. But this is somewhat better than the first model where we did the literal translations using words.

(Refer Slide Time: 19:31)



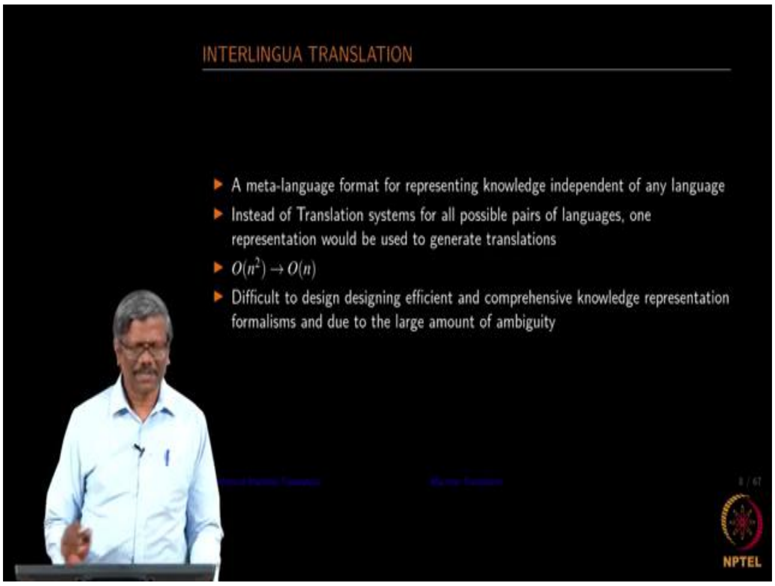
SEMANTICS-BASED TRANSLATION

- The meaning of the source sentence is obtained
- Using the semantics derived from the source sentence, the target sentence is generated

7 / 67

The third one as I mentioned it's about understanding the meaning of the sentence and then using the semantics, you construct the target language.

(Refer Slide Time: 19:43)



INTERLINGUA TRANSLATION

- A meta-language format for representing knowledge independent of any language
- Instead of Translation systems for all possible pairs of languages, one representation would be used to generate translations
- $O(n^2) \rightarrow O(n)$
- Difficult to design designing efficient and comprehensive knowledge representation formalisms and due to the large amount of ambiguity

8 / 67

The next one is I say again we discussed that, considering an interlingua as an intermediary which we can use it to translate. And then from here, we can translate it to multiple languages, the idea of this is to have multiple systems using the same knowledge base right. So, if you are able to construct the interlingua from the source language.

Now, it is possible for us to relieve generate the target sentence in any language. So, instead of having several translators we just have one model where we create the intermediary knowledge from the source language and then using that knowledge construct the sentence for the other languages. So, this is another mechanism in by doing that we are also reducing the complexity in this fashion right. So, as I mentioned its very difficult for design. So, this is what has been attempted by people earlier.