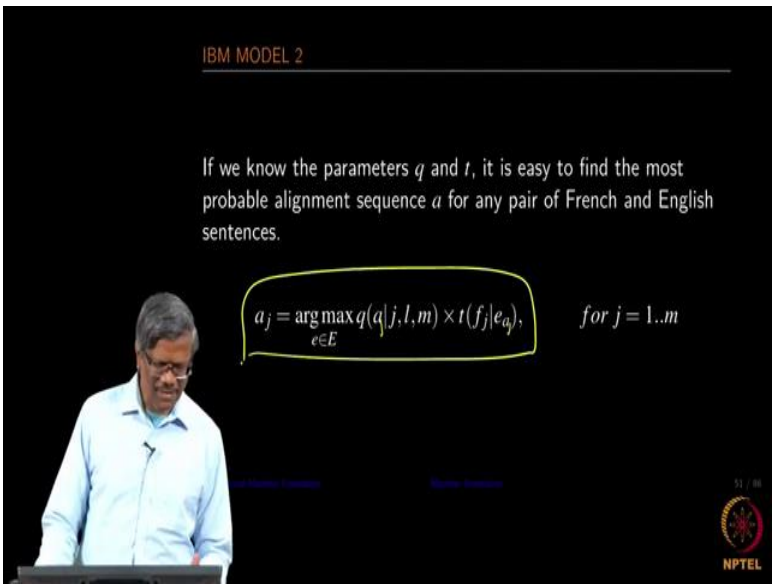
Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute, Madras

Lecture – 69
Introduction to Phrase-based translation

(Refer Slide Time: 00:15)

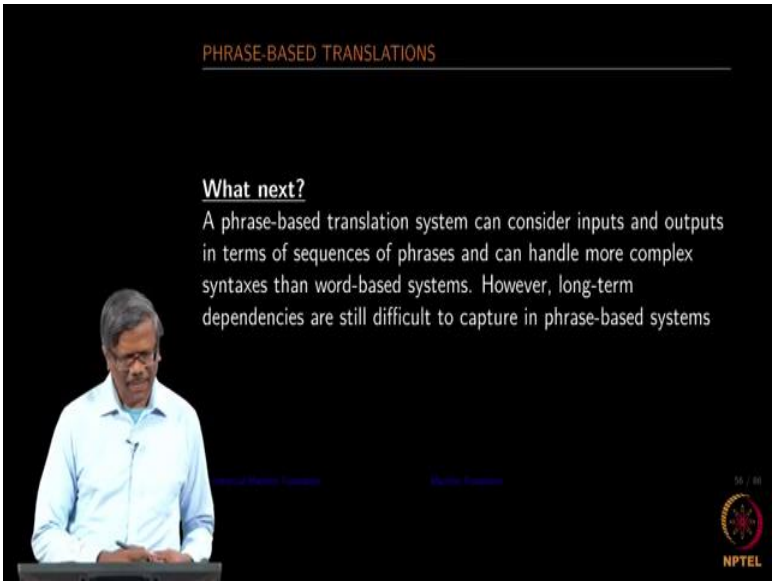


Alright. In continuation of the machine translation using phrases. I just want to refresh you on the IBM model 2; you know in terms of identifying the alignments right.

$$a_j = arg \max_{e \in E} q(a|i,j,m) \times t(f_j|e,m) for\ j = 1 \ldots m$$

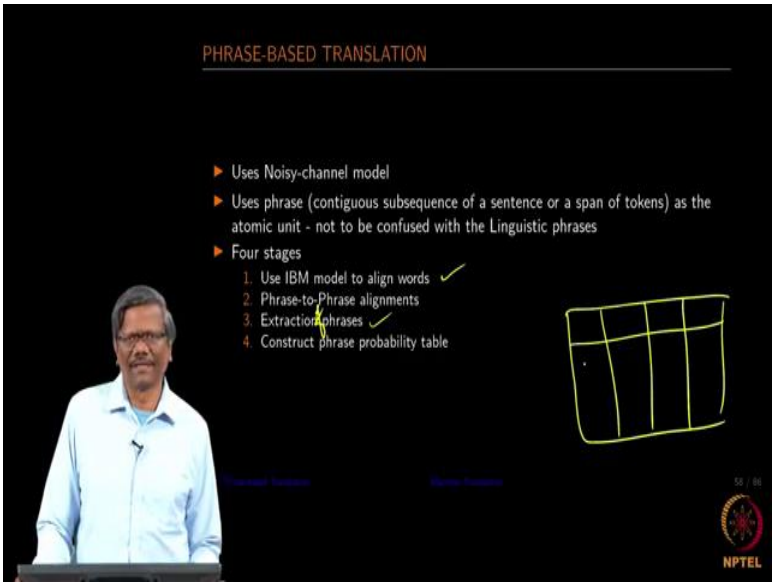
So, we saw that the alignments can be obtained using this equation right. So, we have a set of alignment that you get for every foreign word to the English word right, and this gives you the probability and this is a big table of alignment that you will have. So, this is something that we are going to be using in the phrase-based model.

(Refer Slide Time: 00:58)



And this section of the video we will be talking about the phrase-based model where instead of considering just one word we going to be considering more than one word at a time in order to do the translation ok. This is something that Google and Bing had been using for a long time successfully ok.

(Refer Slide Time: 01:22)



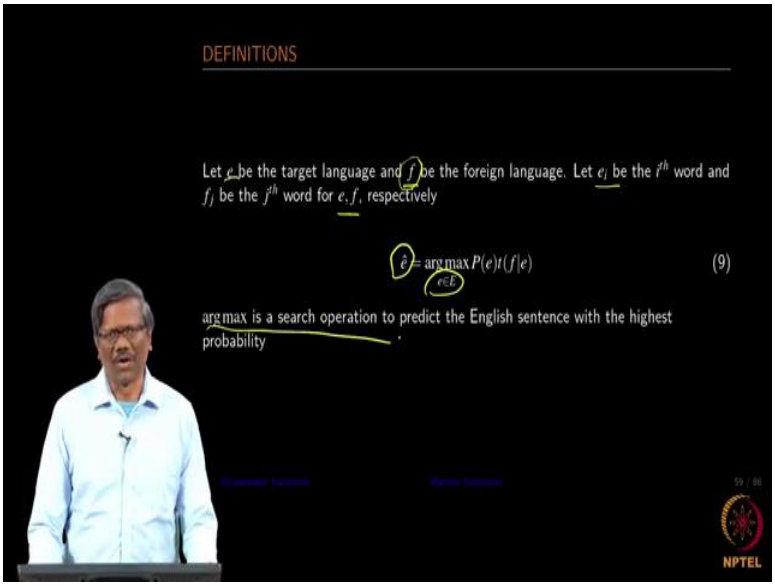
So, we would like to get into this little bit more and see how this functions ok. So, this is not going to be a very elaborate exercise with respect to phrase-based translation. I will be talking about how the phrases could be constructed and then how can we learn the phrase-based learning ok during the training process and then the decoding process is very similar to what we had discussed earlier for IBM models alright.

So, in this case, again we are going to be using the Noisy-channel model. We going to use phrase contiguous of a sequence of a sentence or a span of tokens as in the atomic unit ok. This is not to be confused with the linguistic phrases. The phrases that we are going to be generated using this we would also contain the linguistic phrases.

This is going to have four stages: one, it's going to use as I mentioned earlier IBM model to align words ok. So, that is where we have to start first. So, and then the second one is a phrase to phrase alignment. So, from the word that we have obtained. how do we really construct phrases? And then later how do we align them? And then once we have aligned those phrases: some phrases are useful some phrases are not useful or some phrases would not be consistent with what we are looking at.

So, we want to look at the extraction of phrases in a way that its very consistent with our model and then later construct the phrase probability table the same thing that we have done earlier to write with the word. So, we going to be constructing the phrase probability table that will contain the alignment of various phrases and it will be a huge table of this type like what we are seen earlier for every combination of the English and foreign words there will be a score obtained and that score would be later used for decoding the news foreign sentence that comes into the decoder alright ok.

(Refer Slide Time: 03:40)

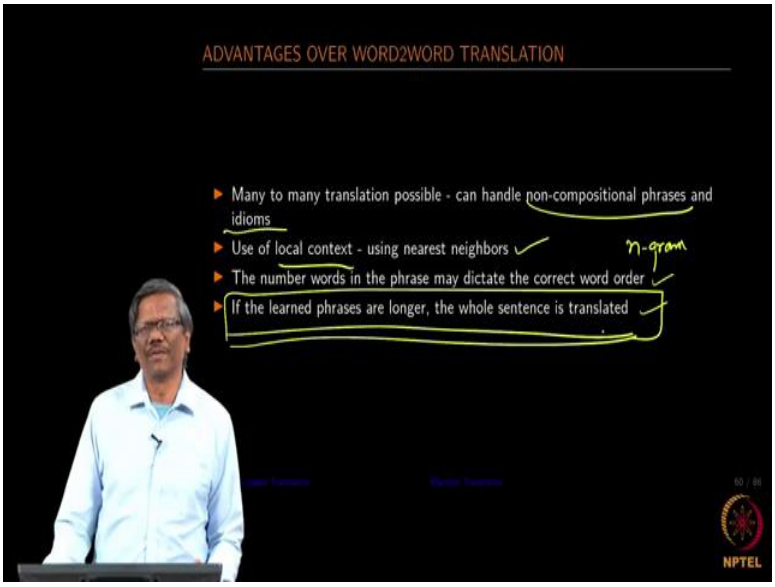


So, the same definition that we had used earlier, I am going to be again using them for the sake of formality. We going to have the language e as the target language and then f

as our foreign language and e I will be the I the word and f I will be the jth word for e and f respectively ok. At the end of the day what we need to find out is; probable English sentence from the given foreign sentence that we have chosen right and this is going to be a list of translated statements, and we need to find the best of what is available in that list ok.

So, as we mentioned earlier this is going to be a big search operation we will talk about the search operation when we really go to the neural translation models alright ok.

(Refer Slide Time: 04:46)



So, what do we do with this right? So, why should we really go for the phrase-based translation model? Is there any advantage that we going to have the phrase-based models? There are a lot of advantages with respect to the phrase-based translation models, one is we will be having a lot of possible translation from foreign to the English language and we can handle a lot of non-compositional phrases as well as idioms we often do.

So, earlier in the case of IBM models which was only about word to word right. So, I would be not really be handling the phrases very well. So, in this case, since we are going to be constructing words that are going to be more than two or three, we will have better translation possible ok.

So, why is it possible to have a good translation because it going to be using the neighborhood words right. So, remember in the language model as you increase the number of context words or the n-gram right. We were able to really get a good handle on the generation of the symptoms.

So, that is why even in this case since, we are going to be using more than one word which can call this as a local context we going to be using the neighborhood words. So, we might be able to bring in some kind of semantic information as part of this right. So, we expect that the phrase-based model would do really better; when we incorporate the local context into the translation model ok. The number of words in the phrase may dictate the correct word order ok.

Sometimes, you know when you translate from foreign language to the English language is not necessary that the phrases are going be contagious in the foreign language as well as in the English language right. So, the phrases that you will find in the foreign language may be split into various words in English ok. So, we will look at that little later ok how that word order is coming into play, and then if you have the learn phrases which are longer, the whole sentence could be translated.

So, this is proven to be right in some cases when the sentence is very short, but when the sentences are longer and when the phrases are almost equal to the size of this sentence the translation is not really improving better than let say a three-word phrase or four-word phrase ok. In some research paper they have found out that beyond three you do not find a lot of I mean when I say beyond three words when a phrase contains more than three words the translation is not really improving much better.

So, much phrase-based translation you will see about three-word phrases or max four-word phrases coming into play ok. So, this is only theoretical knowledge, but as we get deeper into the translation process we understand that it's not necessarily true.