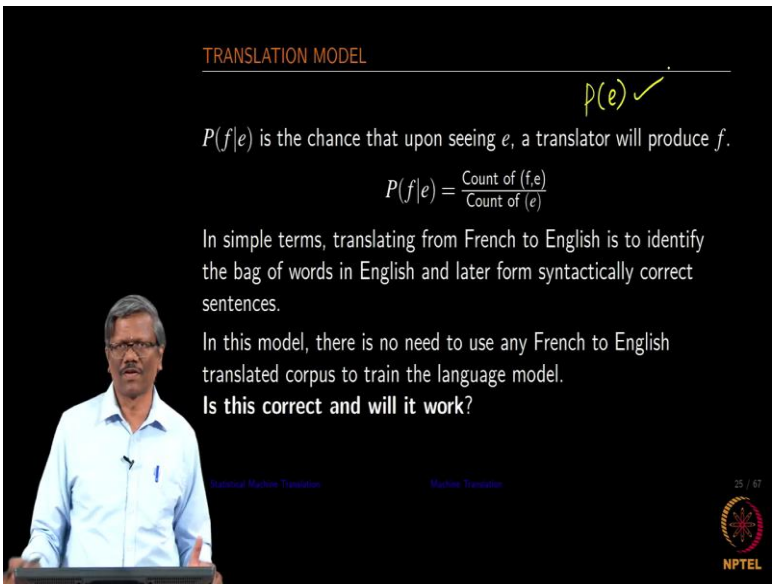
Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute, Madras

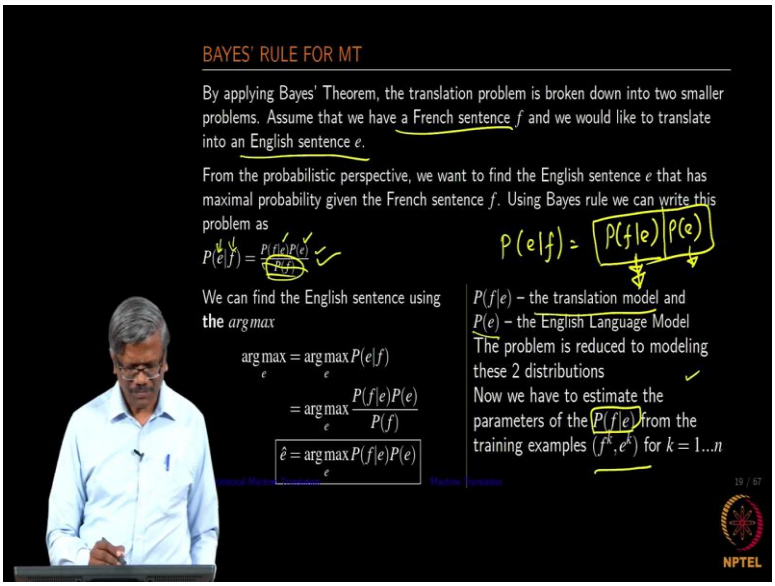
Lecture - 66
Translation Model, Alignment Variables

(Refer Slide Time: 00:15)



The next one is we need to estimate the Translation Model. So, if we, so far we have looked at P_e, right. So, we know how to give a sentence, we know how to really find the probability of a sentence, of a new sentence that is provided, right. So, we have done it earlier. So, we just rush through those slides. Now the next question is how do we really estimate the translation model? So, this is what we are trying to do is not it.

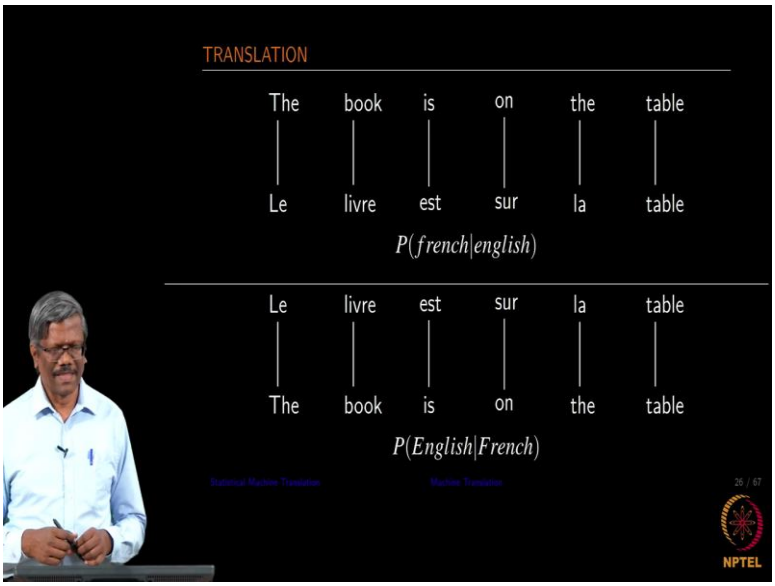
(Refer Slide Time: 00:52)



So, here, so this is what we are we have some ideas on the language model. And now we need to really estimate this parameter from the corpus. Let us see how that could be achieved. So, one way is to find the count; the count is, the count of the French and the English sentences and then use the normalizing factor of this and finally, produce a parameter value for this. So, in simple terms translating from French to English is to identify the bag of words in English and later syntactically correct the sentences to produce the right English sentence, ok.

So, in this model there is no need to use any French to English translated corpus to train the link language model. Is this true ok? So, I am just leaving this to you to find out whether this particular model would work. So, is it possible to estimate the probabilities here? So, this is nothing but your joint probability you know that right. So, the question here is, will this model works; is this good enough to really find the translation for us. So, I like you to go through some material that is available in the research community and check whether this particular model will work for us.

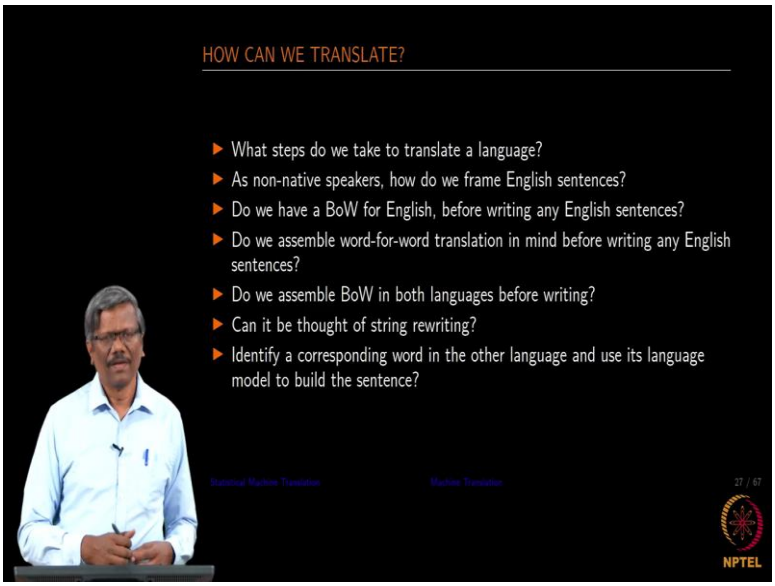
(Refer Slide Time: 02:51)



So, let us now look at the translation that we had done earlier from a different perspective, right. So, earlier we just looked at the word in this Swahili English Model and then try to find out whether it relates to certain words in this Swahili language and so on. So, in this case what we do is, since we do not know those languages that we are translating, we want to find out whether there is one to one correspondence first, right. So, that is how you would have started even in the English to Swahili; here also we have a sentence in English and we have a sentence in French.

So, we try to find the probability of this sentence in French given the English sentence. So, we try to map one on one in this case; and then if we have English to French or the French to English we do this we try to align the words in the French to the English words. We do not know whether that is the right way to do it or not, but let us do that first

(Refer Slide Time: 04:11)



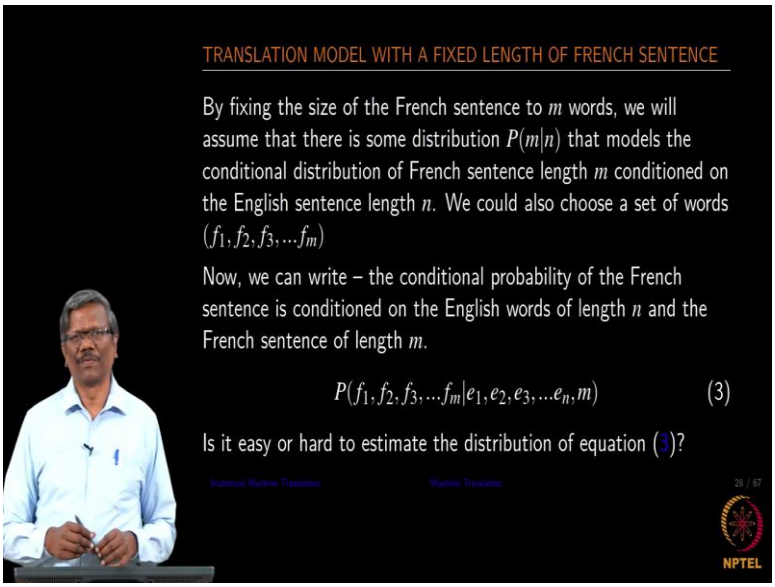
If you do that, is it possible for us to really get the right translated version? So, what steps we normally do, you know in Swahili language somebody is fix in Tamil and you have a friend who does not understand Tamil, you want to give him the English translated version. So, what steps do you take? The first thing is we try to understand the meaning of that sentence since we know that language well and then we also know English.

We probably know all the words from a Tamil to English because you have you are an experienced speaker in English and also in Tamil. Sometimes some words are very difficult to translate. So, at that time you will have some trouble. So, how do you really you know makeup that particular sentence, you know we add additional words in English, so that we convey the meaning. So, we have some ways to do it, but we do not know how we are really doing that, right.

So, it is been somehow inbuilt into us. Do we really form the bag of words when we really want to translate from what he heard to English? I do not think we do any of those, right. Do we assemble word to word? I do not think we do that either when we do the translation. Do we assemble a bag of words in both cases? No, I think we mentioned that earlier. So, how do we do that? So, can we just call it as you know the string of words that are already framed in the mind; and then we are only rewriting the string from one language to the other; can we think of it in that fashion?

So, these are all things you know we even though we have been doing this day in day out, we have lost the ability to really code this insert algorithms to solve the problem, correct.

(Refer Slide Time: 06:35)

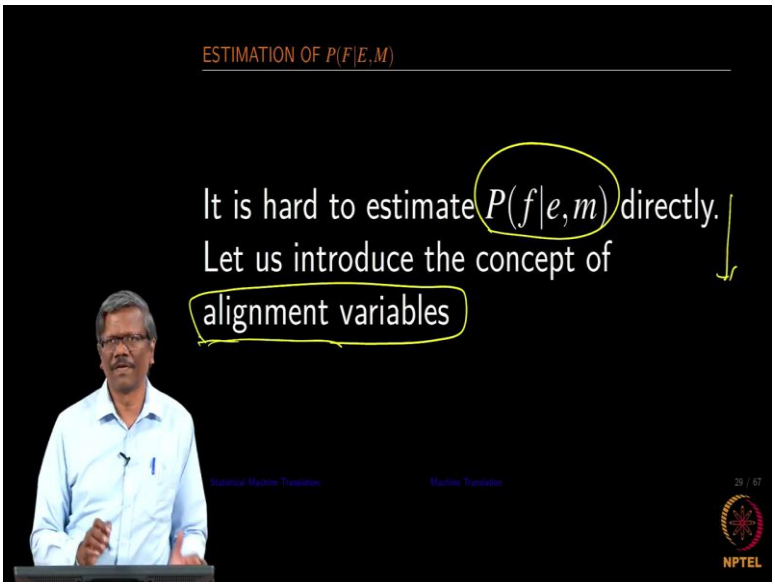


So, that is what we need to really bring in into this mesh. We have to really get a handle on how to translate from one language to the other; applying certain constraints in order for us to translate from French to English. So, in this case what we would do is, we will first restrict the length of the sentence. So now, we need to look at this from this angle of the size.

$$P(f_1, f_2, f_3 \ldots., f_m / e_1, e_2 \ldots . e_n \;\; n, m)$$

So, you look at them, these sentences in English, and then restrict the number of words that you want to form in the translated version for French, and then finally, create a distribution. So, this easy to estimate, I think this is not very easy to estimate, it is a very hard problem to estimate all that.

(Refer Slide Time: 07:36)

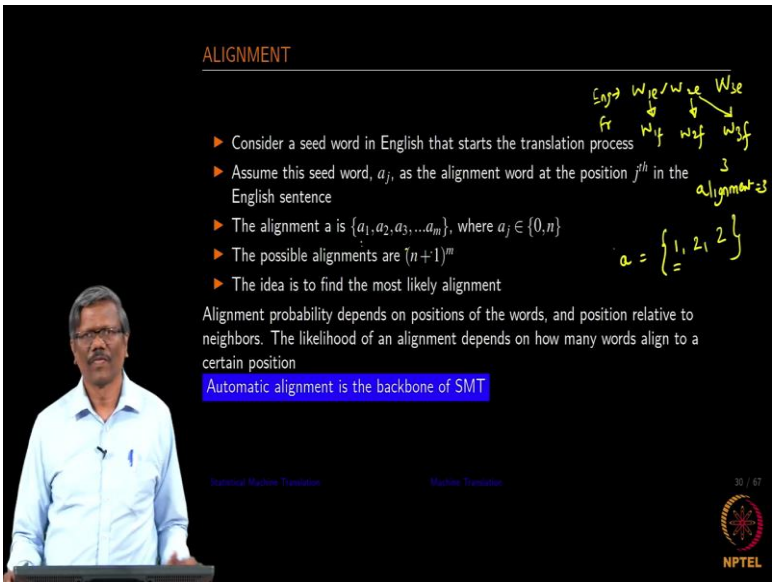


So, it is difficult to estimate directly. So, we introduce the concept of alignment, right. So, there could be so many different possibilities in the earlier case, also because of the restriction that we have imposed. So, we need to split this into various small pieces, so that we can finally, estimate this, ok. So, that is where we bring in what is called an alignment variable.

You remember we have been trying to translate from English to Swahili, we are mentally doing some kind of alignment of word from English to Swahili as we had seen more Swahili sentences, right and this is something we are going to be doing. So, we do this mentally, but we need to really give a formal picture to that, that is where we are going.

So, if we give a formal picture to that and the model to this, then it is possible for you to repeatedly do this; it is not going to just an onetime empirical approach. So, that is where we are bringing in the formal approach, where we try to get the alignment variable.

(Refer Slide Time: 08:58)



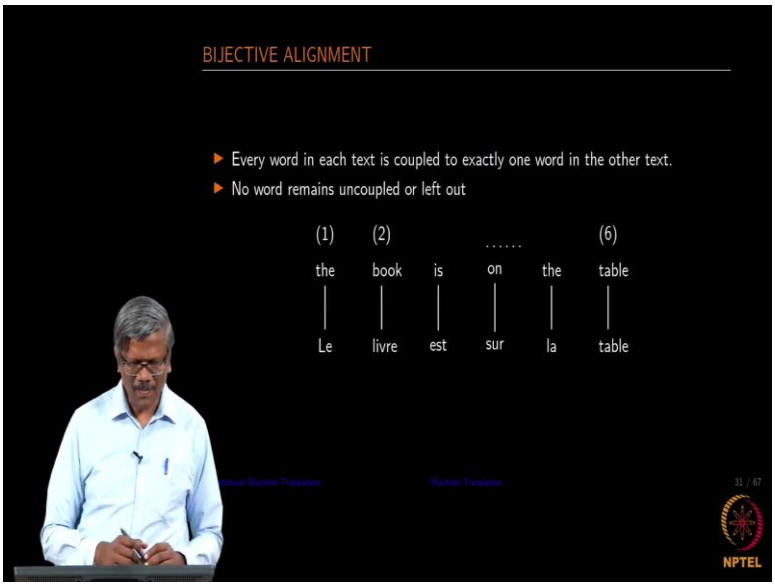
So, let us find out what is the alignment variable? So, let us say that we had a seed word in English that starts the translation process; here we have an English sentence and we want to translate that into French. So, assume that this seed word is zero, let us assume this first and then see the examples in the next slide. When you align the words for example, word 1 2 3 and then this is the English sentence and this is the French sentence we have.

What I mean by the alignment is, trying to map this word of English into the French word. So, we have three alignments here ok. So, in the case where there are m French words, so we will have m alignments ok; in this case this is 3 alignment is 3. So, we have if you align it. So, this is how it is, right. So, the first word this is your alignment for French, that is the first word in English is aligning to the first word in French, so that is given by the index of this array.

So, if you consider this as an array, the index of this array will give you which word we are looking at in French. The first word in English is aligned to the first word in French given by the English word 1 right. So, the second English word is aligned to the second element in this array which is the second word in French. So, in the same fashion this, So, we can also have something like this, alright. So, this is what we are talking about here. So, if you look at the alignment, there are various combinations that you will get into; it is equal to n plus 1 to the power m.

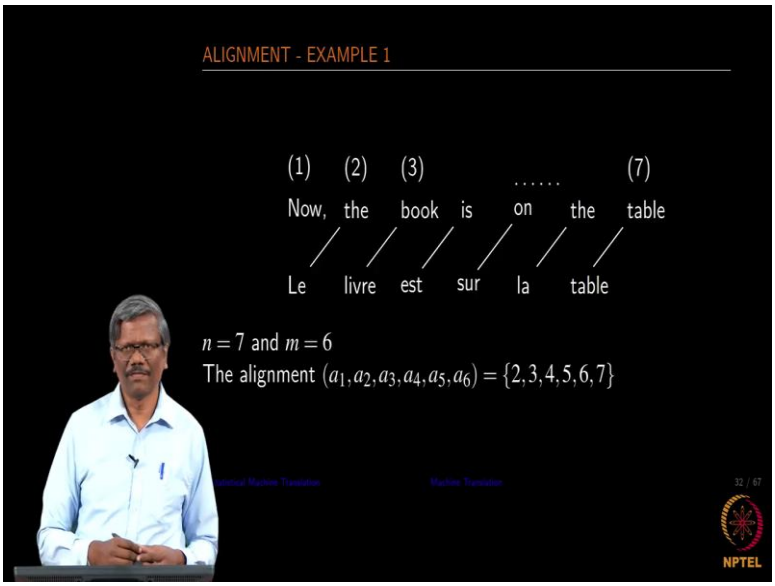
So, that many alignments you will have and then that many parameters that we have to estimate if you have to really translate this. So, the idea is to find what is the most likely alignment, that gives the exact translation for us; out of the so many combinations of alignments, we need to find just one alignment that is most likely. So, here again we are going to be estimating the alignment probability depending on the position of the word. So, this alignment is the backbone for statistical machine translation.

(Refer Slide Time: 12:06)



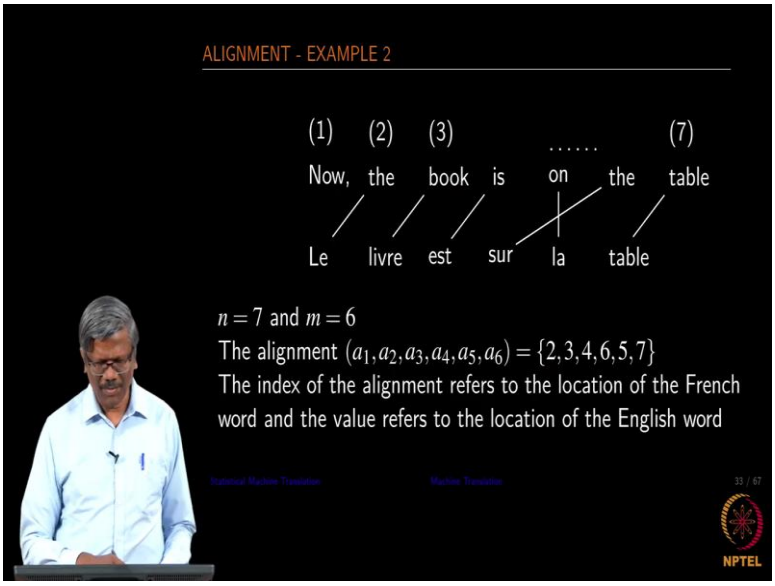
So, we will talk about a few alignments first one is the Bijective alignment, where every word in each text is coupled exactly to one word in the other text. For example, if we have six words here, every word is connected to the other one. It is possible that there could be one English word that could align with two French words as well, alright. So, in this case no word remains uncoupled or left out this is one example. But that is a very ideal situation.

(Refer Slide Time: 12:39)



If you look at the second example here, we have in English sentence: now the book is on the table and then the French equivalent is Le livre est Sur la table. So, when you align it, the word now is not aligned to any of the words, right. The second word is aligned to the first French word and then the third word in English is aligned to the second and that what we are representing here, right; the second word in English, the third word in English in this way.

(Refer Slide Time: 13:24)



So, another example if we look at this, the same English and French sentences I have taken here; and the alignments are little different, the third word in English is aligned to this second word in French, alright. So, there are so many possible alignments that you will get and we need to find the right alignments which really help you in terms of the translation, ok. So, this is another example, the word 3 is aligned to all the words in French.