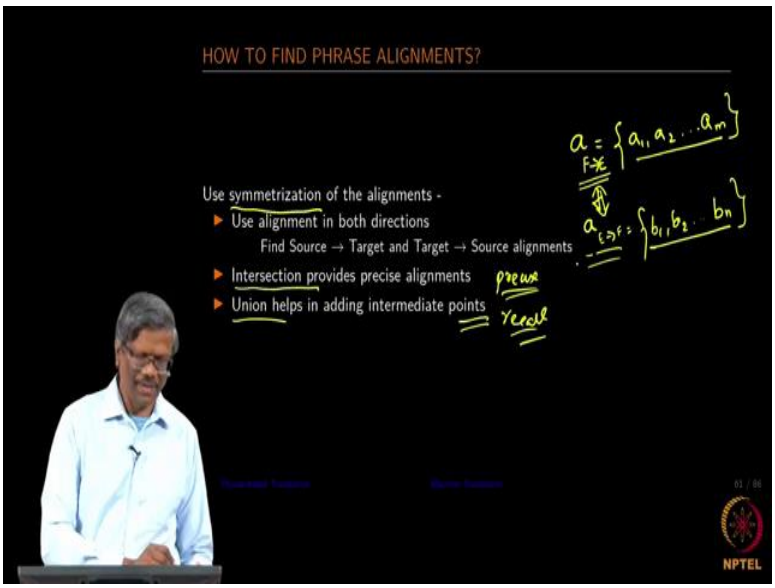
Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute, Madras

Lecture – 70
Symmetrization of alignments

(Refer Slide Time: 00:15)



So, how do we find the phrases in the foreign language as well as in the English language ok? So, there are a few mechanisms available, right. So, one way to do is you just take three grams at a time and then call as a phrase, ok. For example, if I consider this sentence here, I can consider this as a phrase, I can consider this as a phrase or I can consider these four words as a phrase and so on right. Is it the right way to do it?

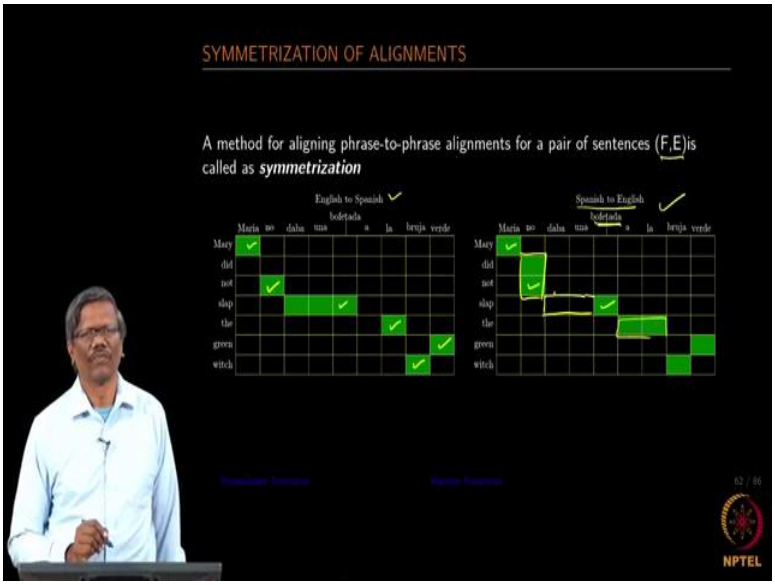
So, as we mentioned right the order of the words in English is not going to be the same in the foreign language as well right. The first word may align to the first word in English, but the fifth word may align to the eighth word in the foreign language or the second word in the foreign language and so on. So, I will show one example when we go into the alignment models, ok. So, what we are going to do is we are going to use a process called symmetrization, this process is very simple.

So, what we do in this process is we take the English sentence and the foreign sentence as a pair and then form the alignment table. Let us say from and then form the alignment table from English to the foreign language in this fashion. And, then take the intersection

of these two. So, when you do the intersection what is going to happen is only those alignment that is common to both this alignment and these alignments are going to remain and then later we take the union of those. So, that will also give you all the alignments possible in both these alignments, correct.

So, when we take the intersection it gives you a precise alignment that means, it gives only those alignment points which are very critical for translation, ok. And, then union gives you a whole lot of alignment that is possible both in this as well as in this case, ok. So, this will be useful for recall and this is for precise. Remember, these terms that used in the information retrieval it is very similar to that ok. So, what we will do is we will take an example and then see how this really works, right.

(Refer Slide Time: 03:39)



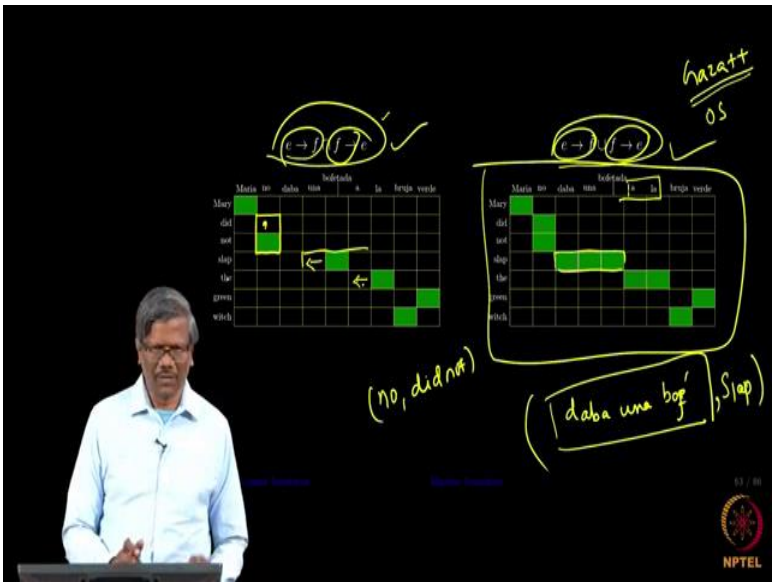
So, as I mentioned earlier method for aligning phrase to phrase alignments for a pair of sentences is called symmetrization, ok. So, we will see how this is done here. So, in this case we have an English to Spanish alignments and the Spanish to English alignment in this case, ok. So, you might also notice that right so, when you do the English to Spanish translation or the alignment you will see all English words are used, right. And, then when you go from the Spanish to English some of the words are not aligned because we just use the direct translation of slap to this word in Spanish, alright.

So, these are the two alignments that are possible. So, how do you take the intersection of this? So, what you do is when you write the alignments you find out what are the

common ones here; so, right. So, these are the words the Maria and Mary and we have no or not and we have this Spanish word for slap, ok. So, you have a phrase for the slap in Spanish ok, and that is what is given here. So, in this case we just take this span intersection. So, this is the value point that is coming in and then we have this point, this point and this point. So, that is going to be your intersection point.

So, when you look at the union of those so, you will have these alright.

(Refer Slide Time: 05:50)



So, what I suggest is if you want to really try this right take any sentence in English, and then maybe you can use Google to translate that into any other language that you do not understand and then try this model, ok. So, maybe you want to take one word at a time in the foreign language and then find out what is the meaning of that and start filling it in the English to Spanish alignment.

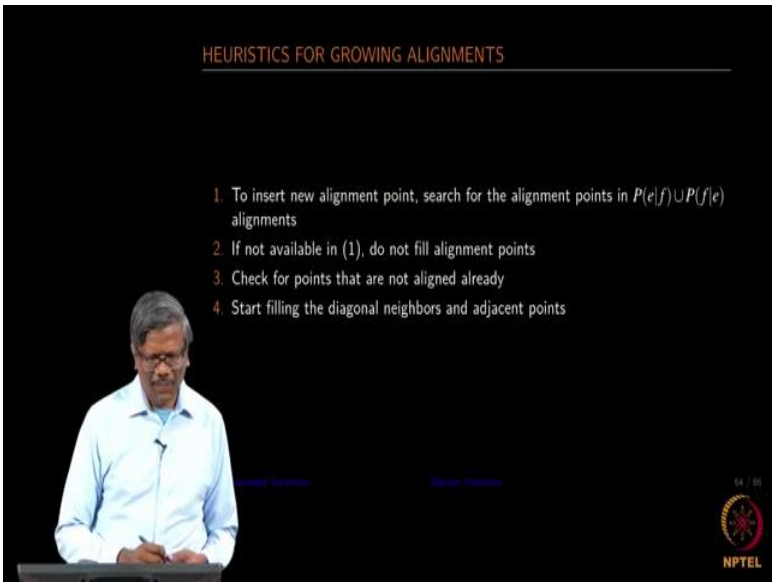
So, that will really help you in terms of understanding this or if you are not really for taking a foreign language that you do not understand take your mother tongue and then take one English sentence or take a sentence in your mother tongue and translate it into English and then try to fill in this table in this form. So, by doing that you will understand the process of symmetrization very well, alright.

So, let us see what is the intersection of the union of this. If you look at this it will be close to this right, but not quite. So, in this case, what happens is there are certain

elements which are aligned properly are coming into play, the element that is not aligned properly would be removed in this process. For example, there is an application called Gaza plus, ok. So, this helps you in terms of aligning two foreign sentences. So, this is open-source and you can download this and then use this application to really do the alignment process.

So, when you align these two words or rather these two sentences using Gaza plus a here is aligned here, ok. So, this is a noisy word or noisy alignment that you have in the alignment process and those noisy alignments would be gone when you do the intersection and union and finally, generate the final alignment table, ok. So, how do we really start creating the alignments? Now, we have in the symmetrization process and intersection of the alignment and the union of the alignment, right.

(Refer Slide Time: 08:33)



So, now, we need to really form an algorithm really to grow our alignments properly and these are the first steps that you would like to use to really align the foreign sentence and the English sentence, ok. So, to insert a new alignment you always search for the alignment points in the union, and then if it is not available in the union you do not fill that point.

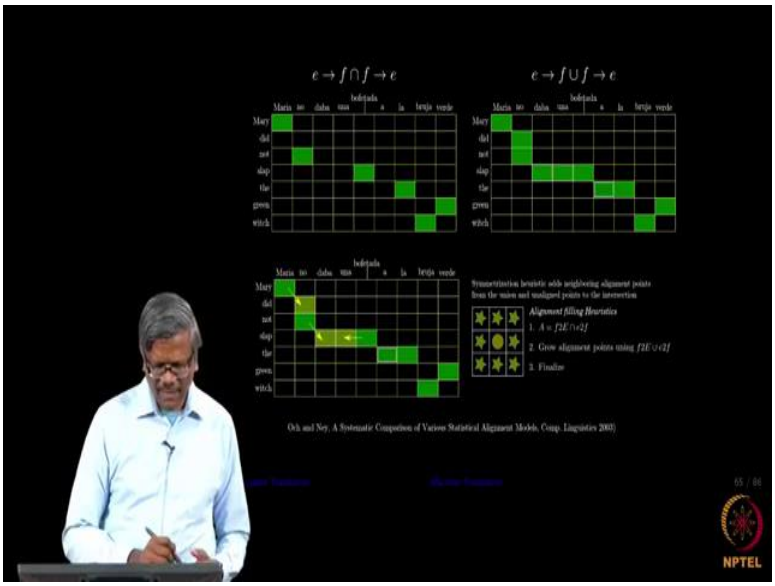
So, you go through the entire table of alignments and then pick each alignment every time and then find out if that alignment point belongs to the union of that and if it's not already filled, ok. If it is not filled, find out if it is available in the union of these two

alignment unions, right. And, then I check the point that is not aligned already alright as I mentioned earlier and then start filling it diagonally and then horizontally in this fashion one point at a time, ok.

So, as I let me go back again. So, here what we want to do is we want to find out phrases right so, that is the idea. So, now, we have this symmetrization process where we have the union or we have the intersection of these two alignments and the union of these two alignments, ok. So, we need to find out the alignments like this, right.

So, for example, no is aligned to did not. So, we need to find out whether it is possible to grow this alignment point, ok. In the same fashion, these three are here, are one phrase, ok. So, this is one phrase that is aligned to slap and then a la is aligned to this. So, is it possible to grow this and then this and this, ok; so, this is the basic idea. So, we want to be able to figure out an algorithm to do this and this is how we do that.

(Refer Slide Time: 11:36)



So, here is an example of how you really fill the alignment. So, we start from the point this is filled and then check whether this point can be brought in as an alignment point since it is not there in the union of that. So, we are not going to use this, right. So, when it comes to this point we check whether this is part of the union? Yes and then see can we grow from here or how does it grow diagonally we start from here, right.

So, this point will get filled because we have a point in the union as well. In the same fashion, we have when you come to this point you will find out whether there is an alignment that is available in the union. So, you grow it and then from here also you fill this point. So, in this fashion you fill these two points, this point and this point using the symmetrization process, ok. This is the alignment that you already have, it finds out these. It is similar to the neighborhood algorithm that you might have seen in the image processing. So, it is the 8th neighborhood.

So, you try to find out whether there is one available here; if this point is available as part of the union fill it up, ok. So, this is how you grow the alignments and when there are no more points to be filled you leave it and then you now have plotted yourself for good alignment table which you may want to use it for the translation purpose. So, when you get this alignment what exactly it means Maria is aligned to Mary; that means, that is the translation of Maria right and then no is aligned to did not and so on so forth.

So, this is the way you can get it, but there are possibilities of more alignments more phrases that you may find. So, this is one phrase that we have identified that did not correspond to no and then this Spanish phrase is aligned to slap or slap has three words in Spanish that are aligned and then that is a la is the and then here there is a single point alignment single point.

So, you can consider these single points as a phrase as well in this case, ok. So, these are all considered as phrases in our case ok. So, there is one phrase 2 3 4 5 and 6. So, is it possible to grow more phrases? So, using this algorithm symmetrization process and the heuristics we would not be able to add any more in this alignment table. Is it possible for us to find more phrases in this, so that is the next step? So, how do you identify phrases in this is the next step, ok.

So, now having obtained an alignment table for the phrases, now we need to find out how many phrases that I can take out of this, so that I will be able to translate multiple phrases into a single word in English or multiple words into multiple words in English, ok.

So, I am so, this is clear now ok. So, I suggest you take another sentence and then try this rather than using it initially you can practice with this sentence that we have here, but you can take a sentence in your mother tongue and then start practicing these all.