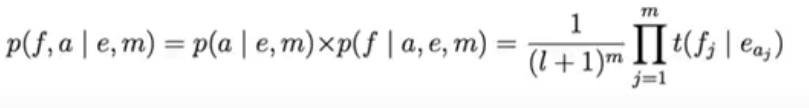
IBM Translation Models

* Research done in late 1980’s ~ early 1990’s.
* Language Model: assigns a probability to each sentence in input language (P(e)).
* Translation Model: assigns a probability to each output sentence in the output language based on the input sentence (P(f|e)).
* Out Model: P(e|f). That is, our model attempts to assign a probability to an English sentence based on a French sentence (we are trying to translate from French to English).
* Out goal: maximize the output probability of the model:

How to model? 🡪 Alignments

* Alignment is the mapping between two equivalent words in the two languages. There could also be a Null word for words that don’t have any alignment.  
  Note that if there are words in the English and word in the French source, then there are pair options to assign (that is for each French word, there are different options to be assigned to). because there is also the Null word.  
  Also, word can be mapped to multiply words in the other language.
* Therefore, we are interested in implementing a model that assigns a probability to each possible assignment.
* That is we want to compute the probability of the pairs and the word in the input language, based on the English sentence and the length of that output sentence. This can be separated into two models: the probability of the pairs based on the English sentence and the length times the probability of the French sentence based on the pairs, the English sentence and the length and then get the probability of the French sentence by summing over all the pair probabilities.
* Alignments are rarely used for translation today.

IBM Model 1:

* , where is an alignment, we assume that each alignment is equally likely.
* , t is the alignment probability.
* Then, the output probability can be calculated:  
  
* Translations steps:
  1. Pick an alignment .
  2. Pick the French words based on the probability computation.

IBM Model 2:

* In this model, IBM only re-defined the alignment algorithm.
* The new alignment method aims to compute the probability that the ’th French word is connected to the ’th English word, given lengths of the French () and English () sentences.
* Then the probability of each word in computed the same way as in the previous model.

The Parameters Estimation Problem:

* We have pairs of translated sentences. We are interested in estimating and to we can compute the probabilities when translating a new sentence.
* The main challenge is that we don’t have the alignment pairs (very hard to annotate).
* Assuming the alignment is known, we can compute the prediction be on maximum likelihood which is a simple count ratio:  
  