# Ass 1 - Word Alignment

Dorin Keshales

## Part 3

### Introduction:

#### Model 1:

- Direction: French -> English
- I found that after 25 epochs I get the lowest possible AER. Therefore, I set Model 1 to run for 25 epochs.
- I used uniform initialization for the translations table – all t(e|f) were initialized with the value: 1/(the French vocabulary size). Where the French vocabulary contains all the unique French words which were in the corpus.
- The AER on Model 1 is **0.296891**.

#### Model 2:

- Direction: French -> English
- I found that after 8 epochs I get the lowest possible AER. Therefore, I set Model 2 to run for 8 epochs.
- I used uniform initialization for both the translation table and the alignments table. all t(e|f) and q(i| j, l, m) were initialized with the value: 1/(the French vocabulary size). Where the French vocabulary contains all the unique French words which were in the corpus.
- q(i| j, l, m) – where **i** is the word position in the French sentence, **j** is the word position in the corresponding English sentence, l specifies the length of the French sentence(without the NULL symbol) and m specifies the length of the corresponding English sentence.
- The AER on Model 2 is **0.204614**.

## First Experiment

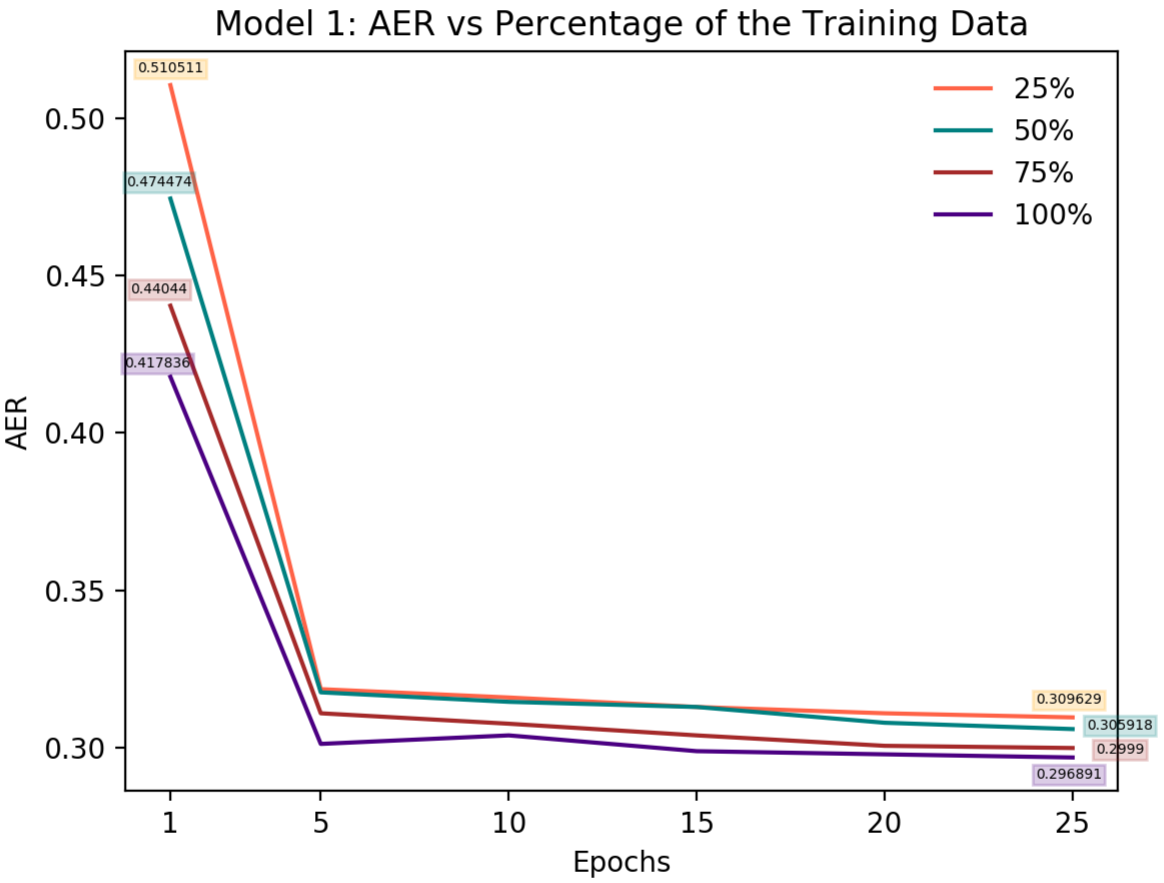
Run your aligner several times, with various percentages of the training data. How does the AER change when you change the amount of training material?

**Answer:**

I run Model1 with the following percentages of the training data:
25%, 50%, 75%, 100%.

We can see that the more training data we have, the better our starting point is after just one epoch. In addition, we can notice that all curves have approximately the same trend, but still at the end of the 25th epoch we get that the minimum AER value is different if we train on different percentages of the data. The best AER (=0.296891) was received when I run Model 1 on the entire dataset (100%).

For conclusion, the more training data we have the lower the AER value can get.

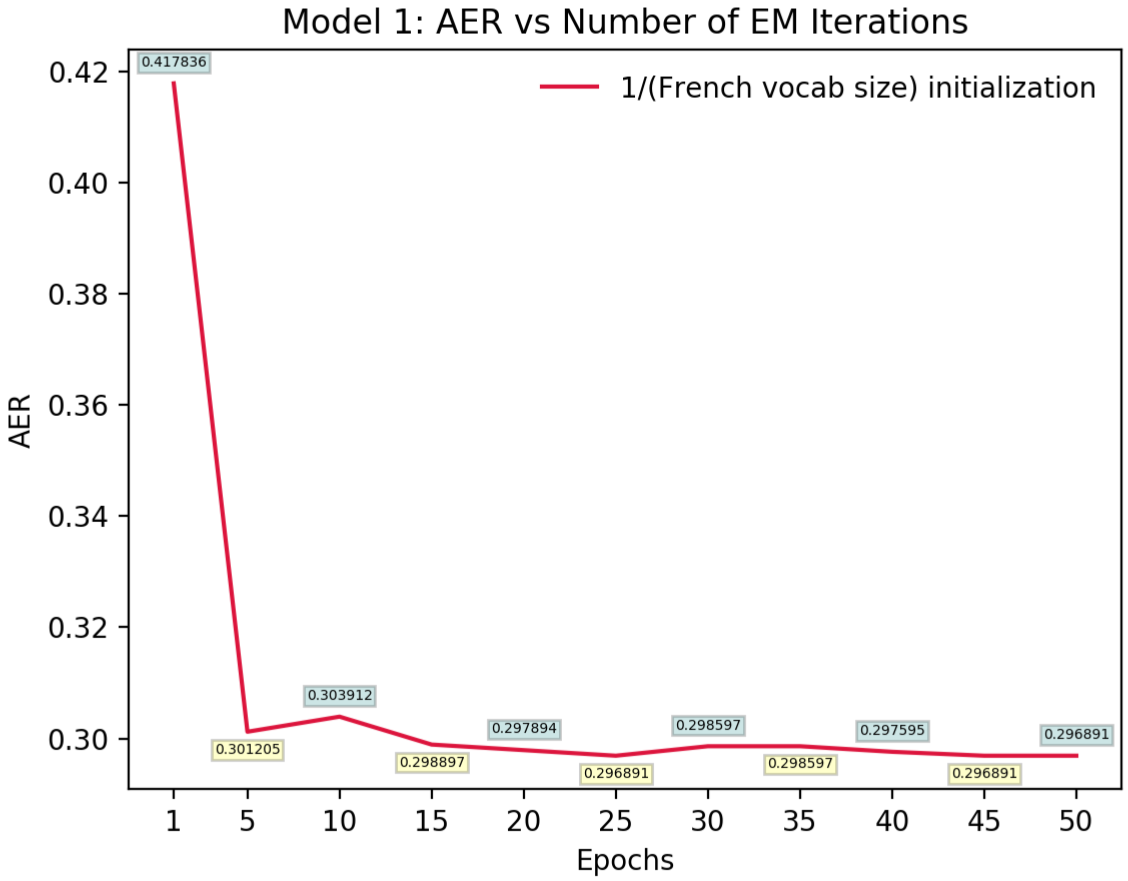


## Second Experiment

How does the AER change when you change the number of EM iterations?

**Answer:**
We can see that 'til epoch number 25,  as long as I increased the number of EM iterations we got lower AER respectively. In epoch no. 25 i got to the lowest AER possible which is 0.296891, on the current data set which contains only 100000 sentences for translation. After the 25th epoch, the AER grew as I added more EM iterations. And in epoch number 30 we can see that the AER starts to decrease again as i added more EM iterations.

In Conclusion, as long as we increase the number of EM iterations and the AER didn’t reached its lowest value (on the dataset) the AER decreases respectively. The moment the AER reached its lowest value, adding more EM iterations will cause the AER to be increased and then maybe after few more epochs, the AER will then decrease again to its lowest value and so on.



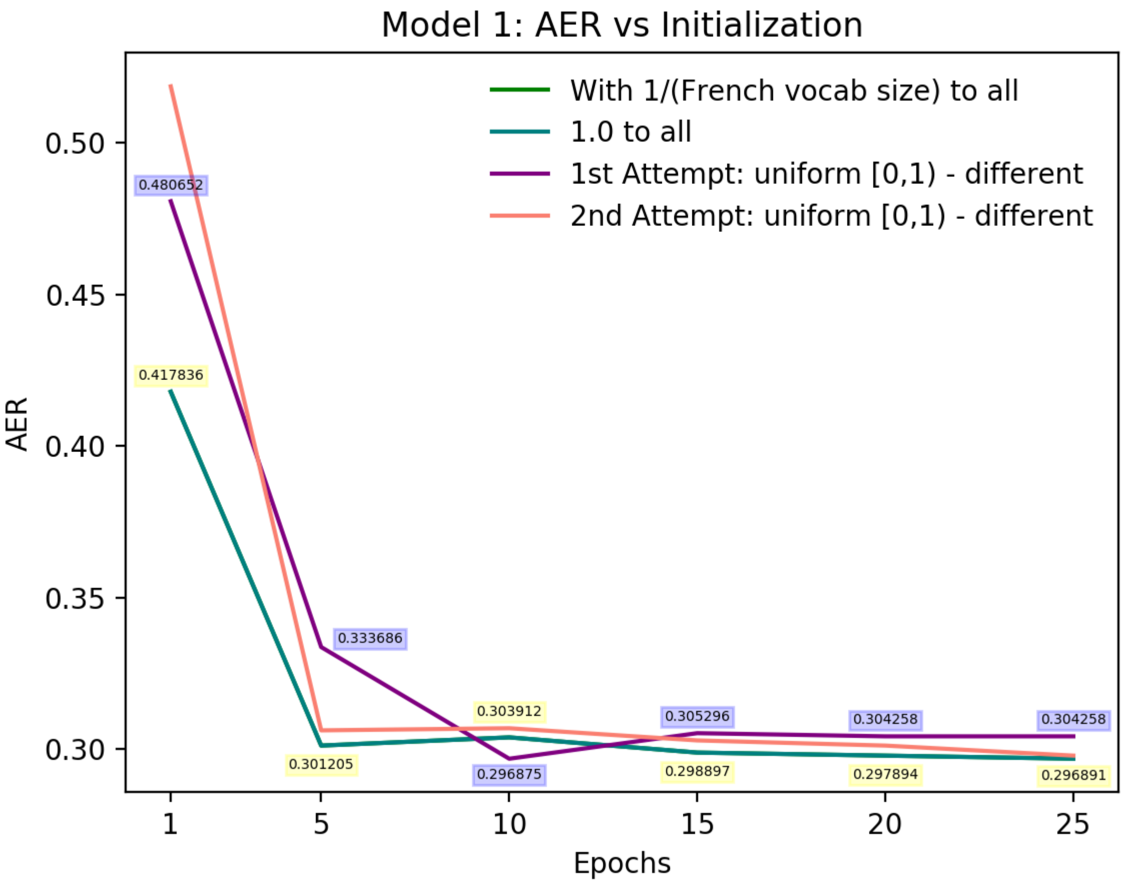
## Third Experiment

Does initialization affect the AER? Run model 1 several times, with different random initializations.

**Answer:**

I tried 4 initializations over all to Model 1. The first 2 initializations were uniform. means that, all t(e|f) were initialized with the same value. In the 1st initialization I chose to initialize with value of 1/(the size of unique French words in the corpus – including the NULL symbol). And in the 2nd initialization I chose to initialize with 1.0 .

We can see that for the uniform initialization method (The first two initializations) we got the same AER in every epoch. So that the green curve and the turquoise curve combined into one visible curve.



The other 2 initializations were practically the same initialization method which is initializing each t(e|f) with random value. I used random.random() so that each t(e|f) was initialized randomly from the range [0.0, 1.0). I run Model 1 twice with this initialization method since different initialization value to each t(e|f) can sometimes bring the model to a very good starting point and at other times brings exactly the opposite effect. And indeed, it can be seen that both curves- the purple and the pink, reflect these two cases.

Over all, after 25 epochs in 3 out of 4 initializations we got to the same AER value. In my opinion the 3rd initialization (the purple curve) would have got too to 0.296891 AER, if I had added more EM iterations.
In conclusion, I can say that the initialization itself doesn’t affect much on the AER and if it does the difference is too small to care about it. Moreover, more EM iterations can fix it.