MT Summer Term 2021 Ex3: IBM Models 1, 2 and 3; Expectation Maximisation (EM)

1. Why is the following idea of estimating translation probabilities $p(e|f)$, where e is an English sentence and f is a foreign sentence, not a good idea:

$$p(e|f) =_{MLE} \frac{number\ of\ times\ f\ translates\ as\ e}{number\ of\ times\ f\ translates\ into\ anything}$$

2. What strategy can you apply to do better?

3. Given the following two sentences and the alignment vector, what does the alignment vector say?

Mary did not slap the green witch .

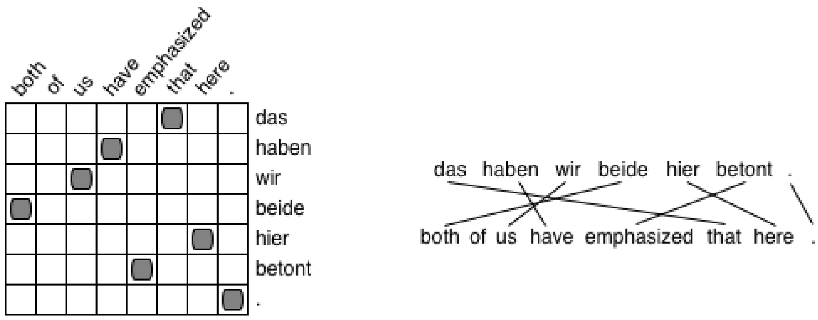
Maria no daba una bofetada a la bruia verde .

$\langle 1,2,4,4,4,0,5,7,6,8 \rangle$

4. Draw the alignment in 3.

5. Given the alignments in 3 and 4, in your own words explain the null element, fertility, reordering and translation parameters.

6. Express the alignment drawn in the pictures below as an alignment vector:



7. Draw the alignment in 3 and 4 above as a two-dimensional grid.

8. Given that your source string has m words and your target string has l words, how many alignments can you have between the source and the target string? (Remember to include the null element …).

9. In what sense are alignments a the hidden structure of translation. In what sense is a a latent variable in IBM models 1 (and the others). Explain in your own words.

10. Given that $\hat{e} = \underset{e}{argmax}\, p(e|f) = \underset{e}{argmax}\, p(f|e)p(e)$, which component part is modeled by IBM model 1?

11. Explain why $p(f,a|e,m) = p(a|e,m) \times p(f|a,e,m)$? What are f, e, a and m?

12. In your own words, explain why: $p(f|e,m) = \sum_{a\in\mathcal{A}} p(f,a|e,m) = \sum_{a\in\mathcal{A}} p(a|e,m) \times p(f|a,e,m)$ What is the technical term designating what is used to get rid of the a's in the right-hand side of this equation?

13. In your own words, explain IBM model 1:

$$p(f,a|e,m) = p(a|e,m) \times p(f|a,e,m) = \frac{1}{(1+l)^m} \prod_{j=1}^{m} \mathbf{t}(f_j|e_{a_j})$$

$$p(f|e,m) = \sum_{a\in\mathcal{A}} p(f,a|e,m) = \sum_{a\in\mathcal{A}} \frac{1}{(1+l)^m} \prod_{j=1}^{m} \mathbf{t}(f_j|e_{a_j})$$

14. IBM Model 2: in your own words, explain the distortion parameter

$$\mathbf{q}(i|j,l,m)$$

15. IBM Model 2: given that

$$p(a|e,m) = \prod_{j=1}^{m} \mathbf{q}(a_j|j,l,m)$$

and the following example

$$\begin{aligned} l &= 6 \\ m &= 7 \\ e &= \text{And the program has been implemented} \\ f &= \text{Le programme a ete mis en application} \\ a &= \{2,3,4,5,6,6,6\} \end{aligned}$$

what is $p(a|e,7)$

16. IBM Model 2: in your own words explain:

$$p(f|a,e,m) = \prod_{j=1}^{m} \mathbf{t}(f_j|e_{a_j})$$

In particular, what are j and a_j above?

17. IBM Model 2: in your own words, explain

$$p(f,a|e,m) = p(a|e,m) \times p(f|a,e,m) = \prod_{j=1}^{m} \mathbf{q}(a_j|j,l,m)\, \mathbf{t}(f_j|e_{a_j})$$

18. Express $p(f|e,m)$ in terms of the formula in 17 above, by marginalizing over the alignments.

19. In your own words, please explain IBM Model 3:

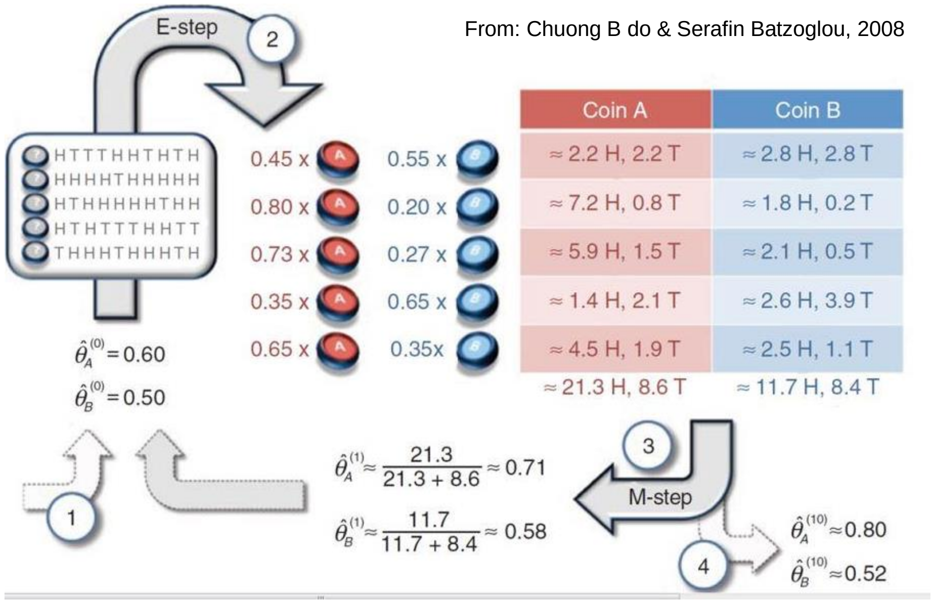
$$P(a,f|e) = \binom{m-\varphi_0}{\varphi_0} \times p_0^{(m-2\varphi_0)} \times p_1^{\varphi_0}$$

$$\times \prod_{i=1}^{l} n(\varphi_i|e_i) \times \prod_{j=1}^{m} t\left(f_j\middle|e_{a_j}\right)$$

$$\times \prod_{j:a_j\neq 0}^{m} d(j|a_j, l, m) \times \prod_{i=0}^{l} \varphi_i! \times \frac{1}{\varphi_0!}$$

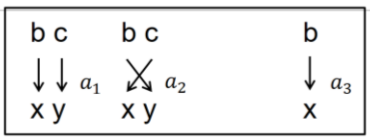
Recall that: $P(f|e) = \sum_a P(a, f|e)$ and $P(a|e, f) = \frac{P(a,f|e)}{\sum_a P(a,f|e)}$

20. In your own words, explain the main idea about Expectation Maximisation (EM):



21. Please do Expectation Maximisation (EM) to estimate $\widehat{\theta_A}$ and $\widehat{\theta_B}$ (the probability of A producing a head, and the probability of B producing a head) under the initial random assignments $\widehat{\theta_A^{(0)}} = 0.6$ and $\widehat{\theta_B^{(0)}} = 0.4$:

22. Please estimate translation parameters t using EM given the following data



and uniform intial translation parameters:

$$t(x|b) = t(y|b) = t(x|c) = t(y|c) = \frac{1}{2}$$