

## Rodas Jateno, Best Chantanapongvanij

### Lab 2, Language Modeling

pchantal1@swarthmore.edu, rjatenol@swarthmore.edu

March 1, 2022

#### Question 1:

The soup warmed the dog.  
Zhiidh or thir o vozir.  
Alignment:  $\langle 2, 1, 5, 4, 3 \rangle$

#### Question 2:

Let  $n_{eats',mange'}$  represent the number of times eat aligns with mange given some alignment.  
Let  $\tau_{eat',mange'}$  represent the likelihood of us getting this alignment. (Probability of generating mange given eat) It's related to the number of times we saw them.  
We start by assuming the value for  $\tau_{eat',mange'}$  to be 1.

$$n_{eats',mange'} = \frac{\tau_{eats',mange'}}{P_{mange(S1)}} = \frac{1}{?} \quad (1)$$

In order to figure out what to put for  $P_{mange(S1)}$  we have to look at the alignment between 'mange' and everything.

$$\begin{aligned} P_{mange(S1)} &= \tau_{she,mange} + \tau_{eats,mange} + \tau_{bread,mange} \\ &= 1 + 1 + 1 = 3 \end{aligned} \quad (2)$$

$$n_{eats,mange(S2)} = \frac{\tau_{eats',mange'}}{P_{mange(S2)}} = \frac{1}{3} \quad (3)$$

$$n_{eats,mange} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \quad (4)$$

$$n_{eats,0} = \sum_f n_{eats,f} = \frac{8}{3} \quad (5)$$

$$\tau_{eats',mange'} = \frac{n_{eats,mange}}{n_{eats,0}} = \frac{2/3}{8/3} = \frac{1}{4} \quad (6)$$

#### Question 3:

$n_{e,0}$  is the sum of all alignments that the English word maps to the French words. It is not the number of times the word "e" appears in the training corpus.

**Question 4:** This is definitely a valid alignment for a French sentence of length 3. This accounts for when vocabularies don't match, such as when each of the three words in French sentence cannot be translated into English.

**Question 5:** While there are a large number of possible alignments of "the" with "pain", there are also a large number of possible alignments of "the" with every other word. The denominator is big, therefore, the number gets washed out.

**Bonus:**

Let's try to make the  $\tau$  value 3 and see what happens.

Let  $n_{eats',mange'}$  represent the number of times eat aligns with mange given some alignment.

Let  $\tau_{eat',mange'}$  represent the likelihood of us getting this alignment. (Probability of generating mange given eat) It's related to the number of times we saw them.

We start by assuming the value for  $\tau_{eat',mange'}$  to be 1.

$$n_{eats',mange'}(S1) = \frac{\tau_{eats',mange'}}{P_{mange}(S1)} = \frac{3}{?} = \frac{1}{3} \quad (7)$$

In order to figure out what to put for  $P_{mange}(S1)$  we have to look at the alignment between 'mange' and everything.

$$\begin{aligned} P_{mange}(S1) &= \tau_{she,mange} + \tau_{eats,mange} + \tau_{bread,mange} \\ &= 3 + 3 + 3 = 9 \end{aligned} \quad (8)$$

$$n_{eats,mange}(S2) = \frac{\tau_{eats',mange'}}{P_{mange}(S2)} = \frac{3}{9} = \frac{1}{3} \quad (9)$$

$$n_{eats,mange} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3} \quad (10)$$

$$n_{eats,0} = \sum_f n_{eats,f} = \frac{8}{3} \quad (11)$$

$$\tau_{eats',mange'} = \frac{n_{eats,mange}}{n_{eats,0}} = \frac{2/3}{8/3} = \frac{1}{4} \quad (12)$$

In Equation 9 we can see that because of proportionality, whatever we choose for  $\tau$  is going to give us the same  $\frac{1}{3}$ . As a result, our future  $\tau$  value is going to stay the same. However, when we look at the  $P_{mange}$  We can see that it's different now. This value is directly affected by our  $\tau$  value.