Problem 1

Consider the following rather contrived dataset:
<s> Kim I am </s>
<s> am I Kim </s>
<s> Kim I am </s>
<s> I </s>

The vocabulary size |V|=3. Consider the second sentence "<s> am I Kim </s>". Under model U, the probability

$$p_u = \text{p(am|<s>)p(I|am)p(Kim|I)p(</s>|Kim)}$$

$$\text{p(am|<s>)} = \frac{1}{4}$$

$$\text{p(I|am)} = \frac{1}{3}$$

$$\text{p(Kim|I)} = \frac{1}{4}$$

$$\text{p(</s>|Kim)} = \frac{1}{3}$$

However, under the model S, we have

$$\text{p(am|<s>)} = \frac{2}{7} > \frac{1}{4}$$

$$\text{p(Kim|I)} = \frac{2}{7} > \frac{1}{4}$$

So in this case, we have $p_s > p_u$.