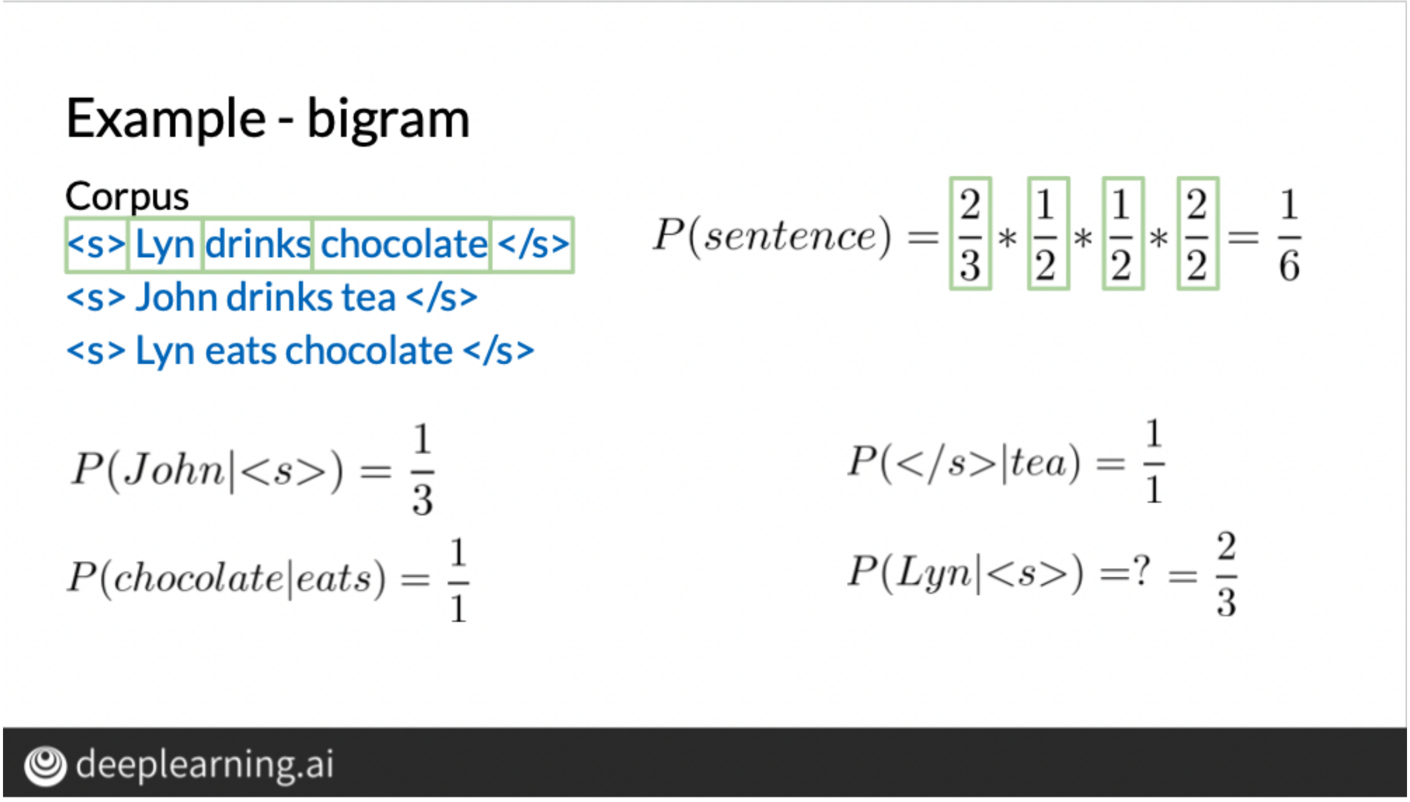
Starting and Ending Sentences

We usually start and end a sentence with the following tokens respectively: <s> </s>.

When computing probabilities using a unigram, you can append an <s> in the beginning of the sentence. To generalize to an N-gram language model, you can add N-1 start tokens <s>.

For the end of sentence token </s>, you only need one even if it is an N-gram. Here is an example:



Example - bigram

Corpus

<s> Lyn drinks chocolate </s>

<s> John drinks tea </s>

<s> Lyn eats chocolate </s>

$$P(sentence) = \boxed{\frac{2}{3}} * \boxed{\frac{1}{2}} * \boxed{\frac{1}{2}} * \boxed{\frac{2}{2}} = \frac{1}{6}$$

$$P(John|<s>) = \frac{1}{3}$$

$$P(</s>|tea) = \frac{1}{1}$$

$$P(chocolate|eats) = \frac{1}{1}$$

$$P(Lyn|<s>) =? = \frac{2}{3}$$

deeplearning.ai