You've learned about <self attention>, you've learned about <multi headed attention>. Let's put it all together to build a transformer network in this video. You see how you can pair the attention mechanisms you saw in the previous videos to build the transformer architecture. Starting again with the sentence Jane Visite the feet on September and its corresponding embedding. Let's walk through how you can translate the sentence from french to English. I've also added the start of sentence and end of sentence tokens here. Up until this point, for the sake of simplicity, I've only been talking about the embeddings for the words in the sentence. But in many sequences sequence translation task, will be useful to also at the start of sentence or the SOS and the end of sentence or the EOS tokens which I have in this example. The first step in the transformer is these embeddings get fed into an encoder block which has a <multi head attention> there. So this is exactly what you saw on the last slide where you feed in the values Q K and V computed from the embeddings and the weight matrices W. This layer then produces a matrix that can be passed into a feed forward neural network. Which helps determine what interesting features there are in the sentence. In the transformer paper, this block, this encoding block is repeated N times and a typical value for N is six.