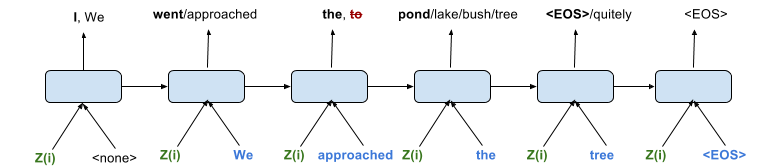
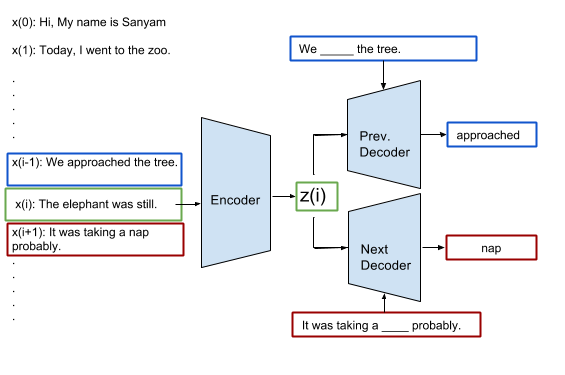


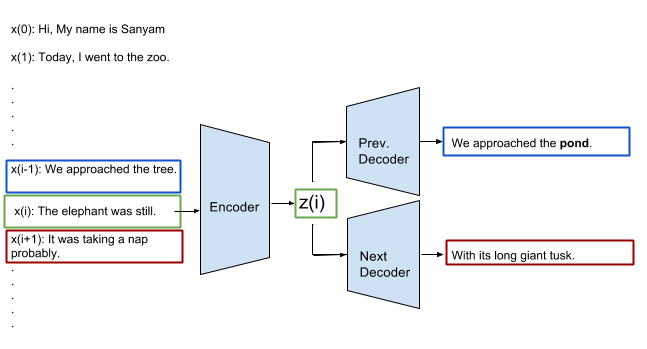
Imputing the missing word: Decoder model



Unrolling the Decoder in time: The input to the Decoder at every time-step is the Z(i) embedding from the Encoder (in green), the actual words in the target sentence (in blue), the previous hidden state. The output is a distribution over the possibilities of words. In the third time-step, it becomes easy for the Decoder to output “the” instead of “to” as it knows for a fact that the actual word preceding it is “approached” and not “went”.



Imputation: Guessing the missing words given all other words and the context of a neighbouring sentence



Skip Thoughts model overview