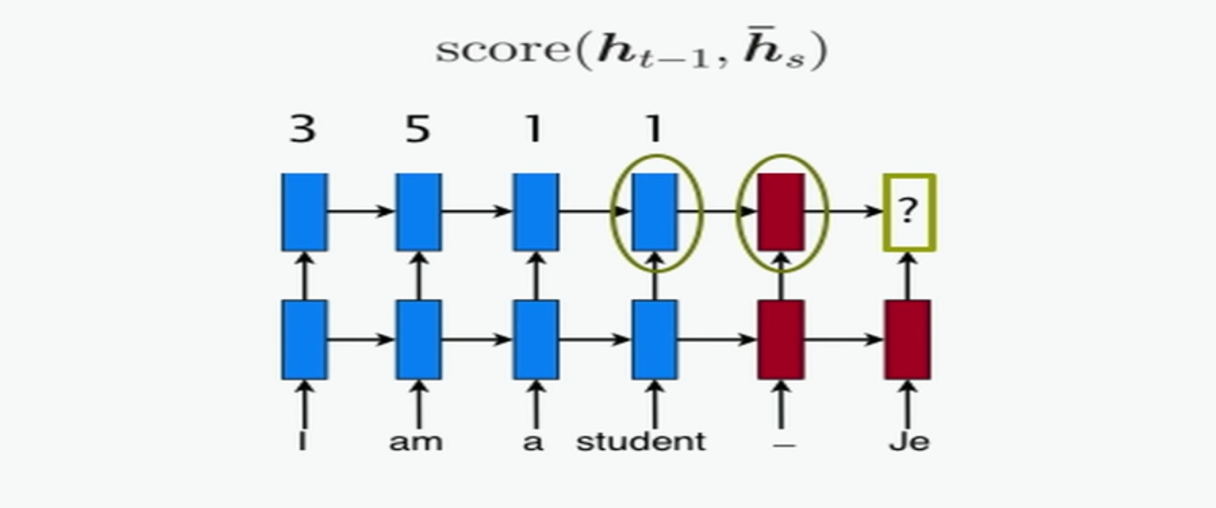
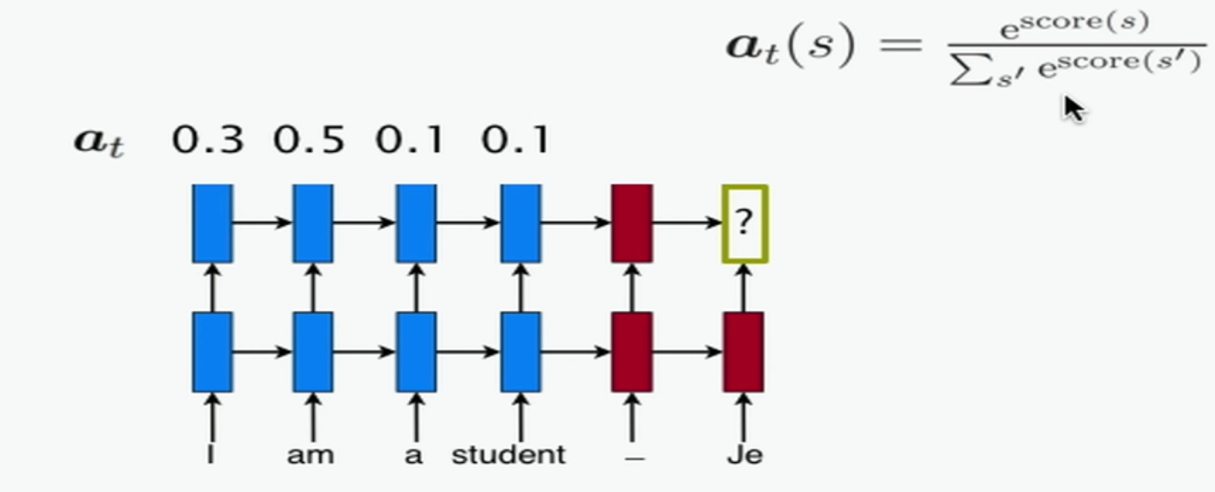
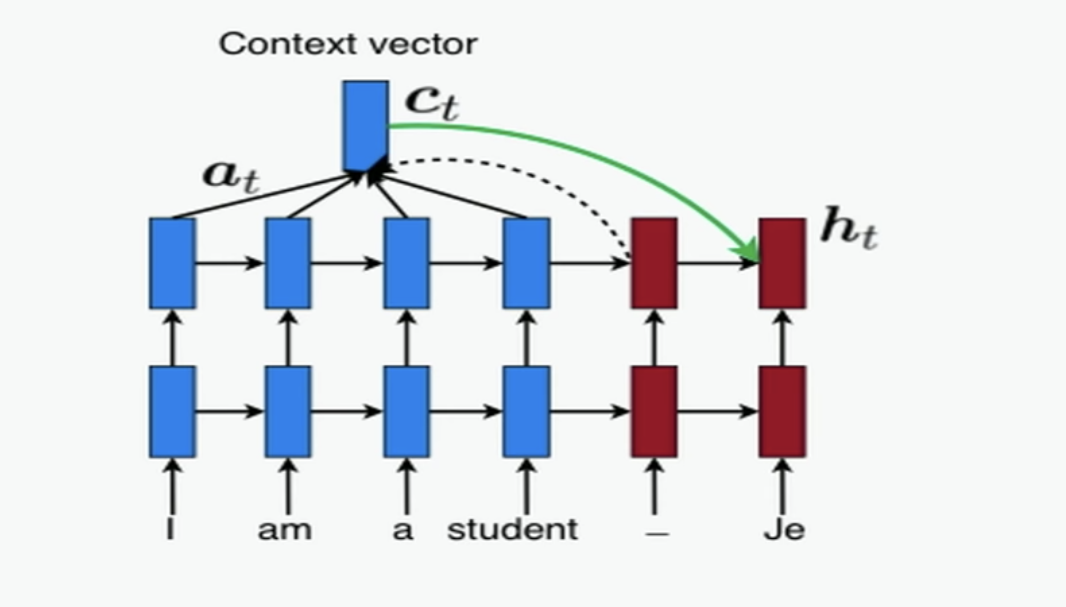
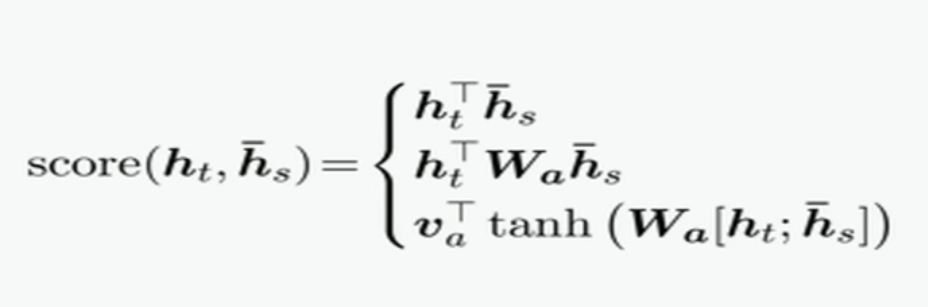
Lecture 10 | Neural Machine Translation and Models with Attention

* NMT is the approach of modelling the entire MT process via one big ANN
* Decoder can be seen as a conditional recurrent language model
* 4 big advantages of Neural MT
  + End-to-end training
    - All parameters are simultaneously optimised to minimise a loss function on the network’s output
  + Distributed representations share strength
    - Better exploitation of word and phrase similarities
  + Better exploitation of context
    - NMT can use a much bigger context, both source and partial target text, to translate more accurately
  + More fluent text generation
    - Text generation is much higher quality
* Models with Attention : Vanilla seq2seq and long sentences
  + Instead of the final unit of encoding layer being the only input state into the decoding layer, we now have a pool of source states where the decoding units have access to all the units within the encoding layer and learn to use the relevant encoding unit to translate
    - Random access memory where you only retrieve the relevant unit as needed
  + Attention model tend to replicate the alignment model, translate as you go rather than read whole sentences then translate

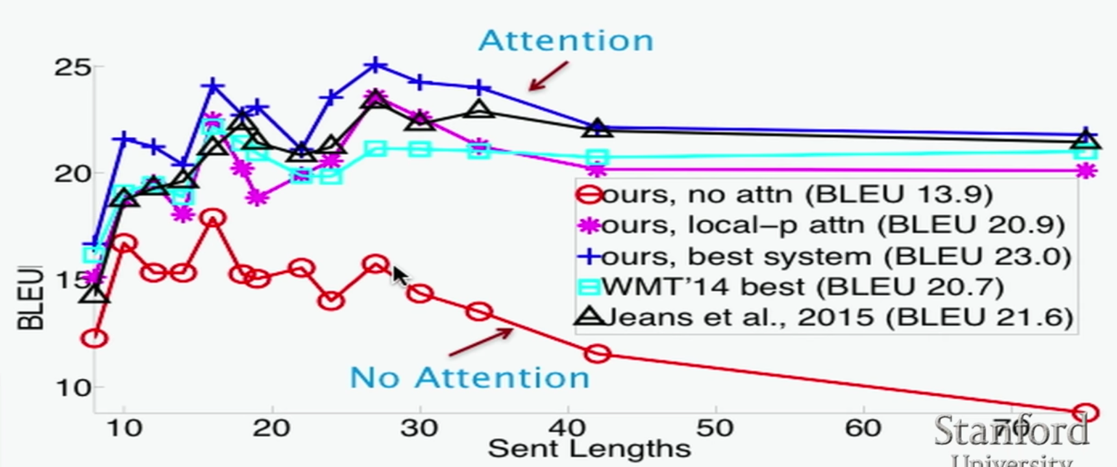




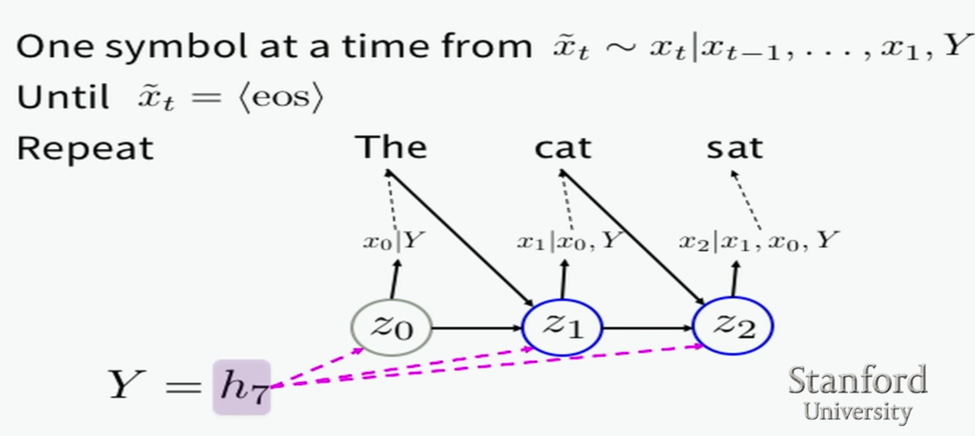




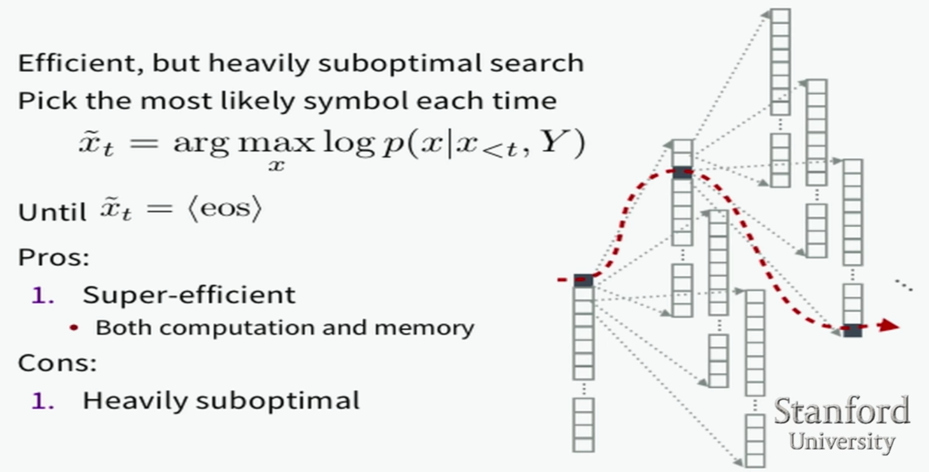
* + You can avoid focusing on everything in the encoding later at each time. You can have local focus where you have subset of source states and this is good for long sequences



* Decoding
  + Ancestral Sampling



* + Greedy search (common technique)



* + Beam search

