# [SEQUENCE-TO-SEQUENCE]

- Input a sequence, output a sequence
- Encoder-decoder are Seq-to-Seq Networks
- We feed the network a pair of sequences

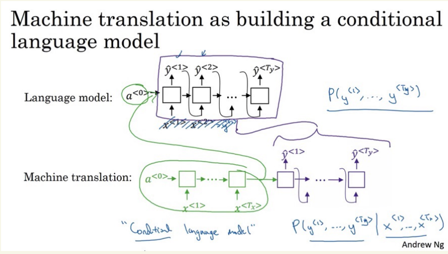
# [ATTENTION MODELS]

- Problem with long sequence: difficult to memorize.
- Solution: attention model. Improve performance on longer sequence

- Language model: allows to estimate the prob. of a sentence

VS

- Machine translation: similar to language model, but instead of starting with vector of zeros, it starts with an encoder network that figures out some repr of the input sentence. Estimates the prob. of a sentence on conditions of an input sequence.



Machine translation as building a conditional language model

-> You still need to pick the most likely output sentence. You shouldn't sample at random. You can pick from algo like:

- Greede search
- Beam search: selects multiple alternatives for an input sequence at each timestep based on conditional probability