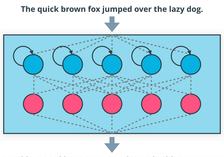
# Machine Translation

Machine Translation can be thought of as a sequence-to-sequence learning problem.




Machine Translation: A Sequence-to-Sequence Learning Problem

You have one sequence going in, i.e. a sentence in the source language, and one sequence coming out, its translation in the target language.

This seems like a very hard problem - and it is! But recent advances in Recurrent Neural Networks have shown a lot of improvement. A typical approach is to use a recurrent layer to encode the meaning of the sentence by processing the words in a sequence, and then either use a dense or fully-connected layer to produce the output, or use another decoding layer.

Experimenting with different network architectures and recurrent layer units (such as LSTMs, GRUs, etc.), you can come up with a fairly simple model that performs decently well on small-to-medium size datasets. Commercial-grade translation systems need to deal with a much larger vocabulary, and hence have to use a much more complex model, along with many optimizations, etc. Training such models requires a lot of data and compute time.

## Neural Net Architecture for Machine Translation

Let's develop a basic neural network architecture for machine translation.

### Input Representation

The key thing to note here is that instead of a single word vector or document vector as input, we need to represent each sentence in the source language as a sequence of word