# Recitation 6: RNN Basics

Joseph Konan and Kinori Rosnow

# Sequential Data

Sequential Dependence - outcome depends on previous/next steps of sequence

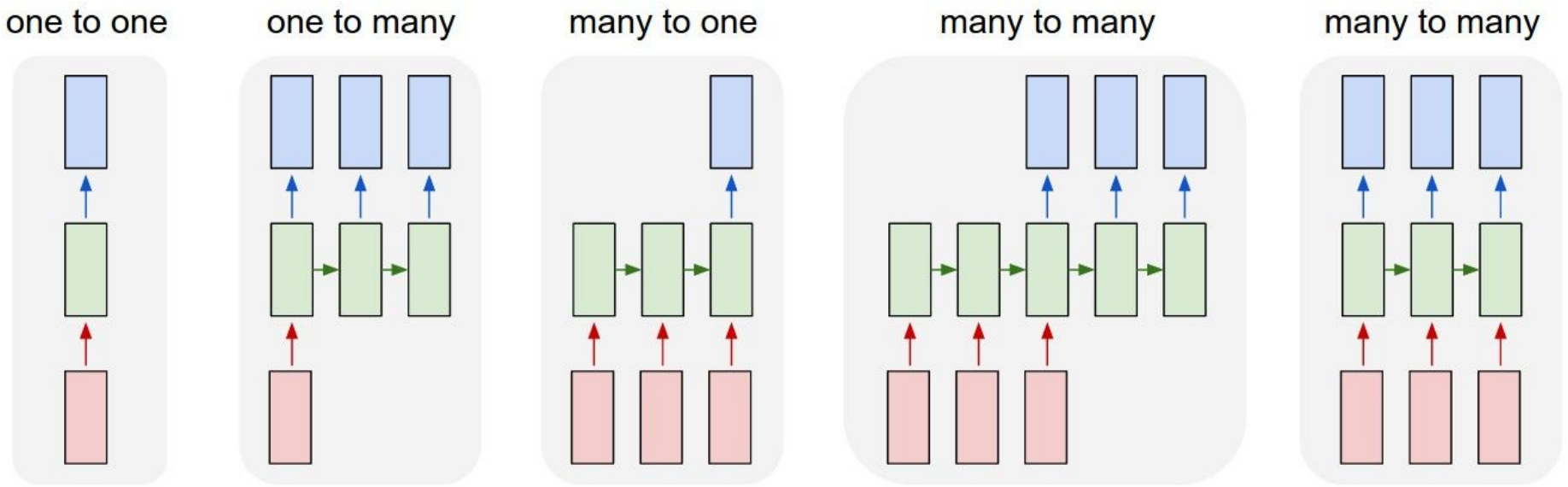
**Finish the lyric:**
“Never gonna to give you up. Never gonna to let you _____.”

**Fill in the blank:**
“Hey, I just met you and this is _____, but here’s my number. So call me, maybe”

**What is this word:**
**“**heh-low-wer-l-d”

**Other Tasks:**
- speech transcription
- text generation
- ...

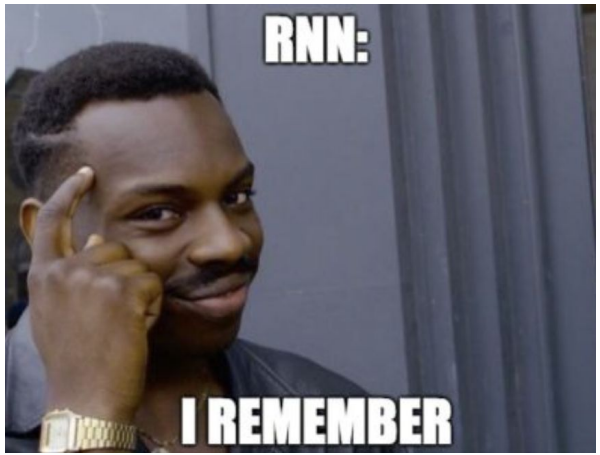
# Data Types

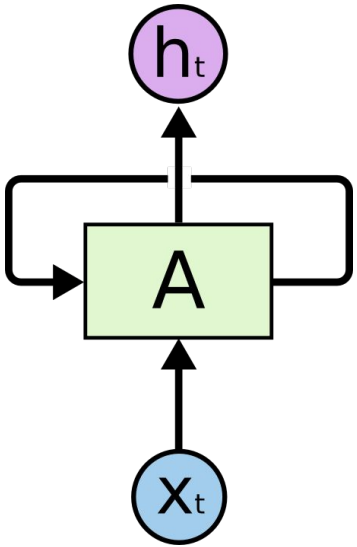



https://i.stack.imgur.com/b4sus.jpg

# Recurrent Neural Networks

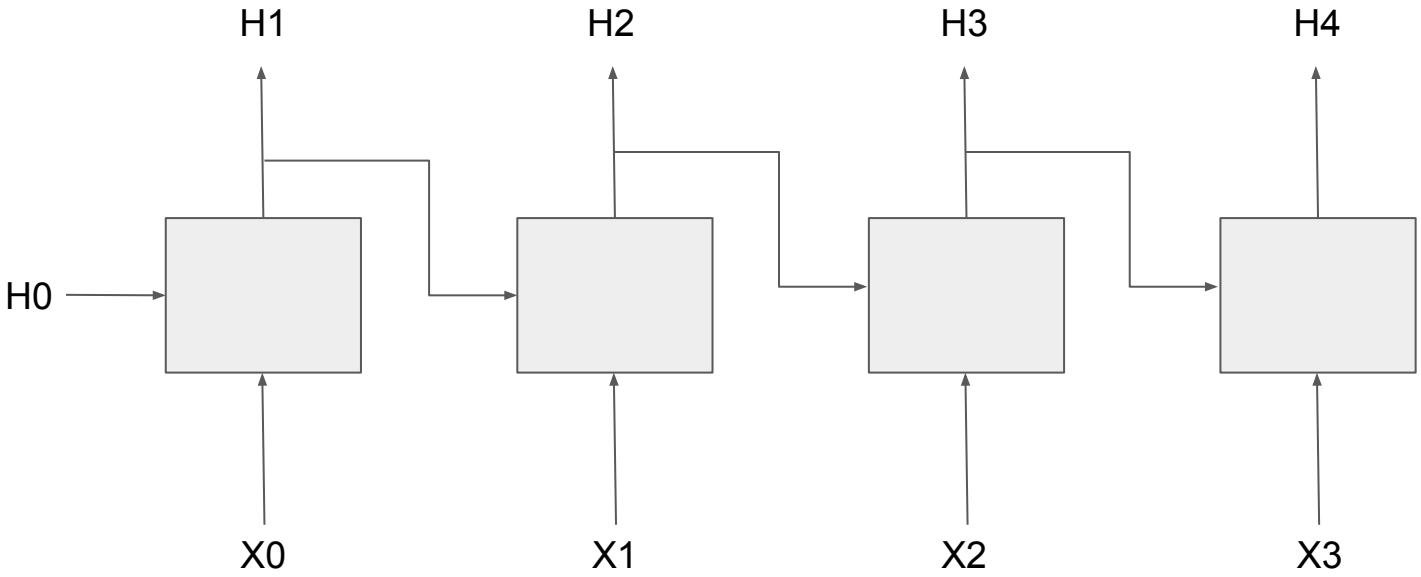
- Cyclic connection
  - Previous step informs next step
- Differs from CNN
  - Considers sequential dependencies
- Architectural variants to address certain issues



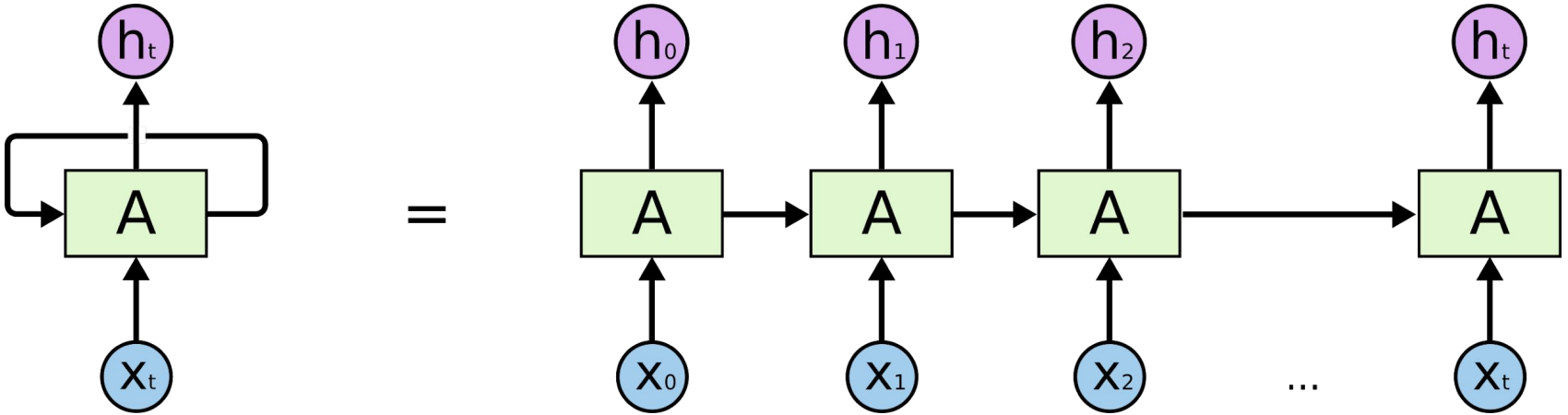




http://colah.github.io/posts/2015-08-Understanding-LSTMs/

# RNN Unrolled - through time/steps



# RNNs Unrolled

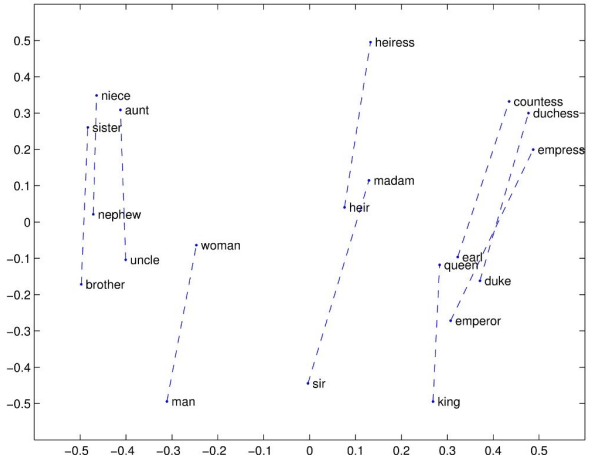


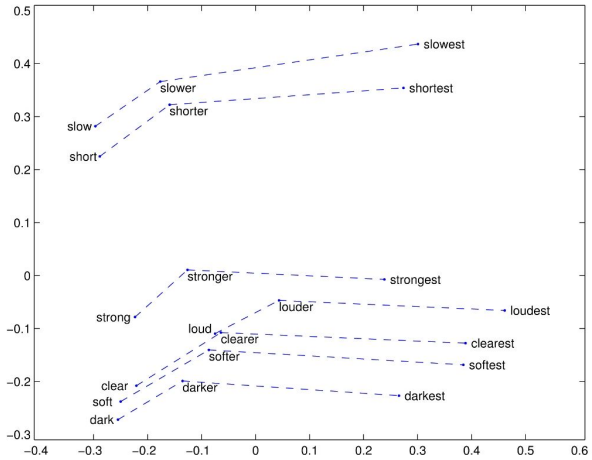

http://colah.github.io/posts/2015-08-Understanding-LSTMs/

# Text vector representation

- Input: One hot encoding
  - “Never going to give you up” {N=6}
    One Hot Encoding: Never = [1, 0, 0, 0, 0, 0]
- Input/Post-processing: Word embedding
  - Efficient use of space (denser)
  - Can represent relationships (information rich)
- Output: Probability Distribution
  - “Never going to give you up” {N=6}
    [Never, going, to, give, you, up]
    P(w) = [0.01, 0.01, 0.03, 0.04, 0.05, 0.86]

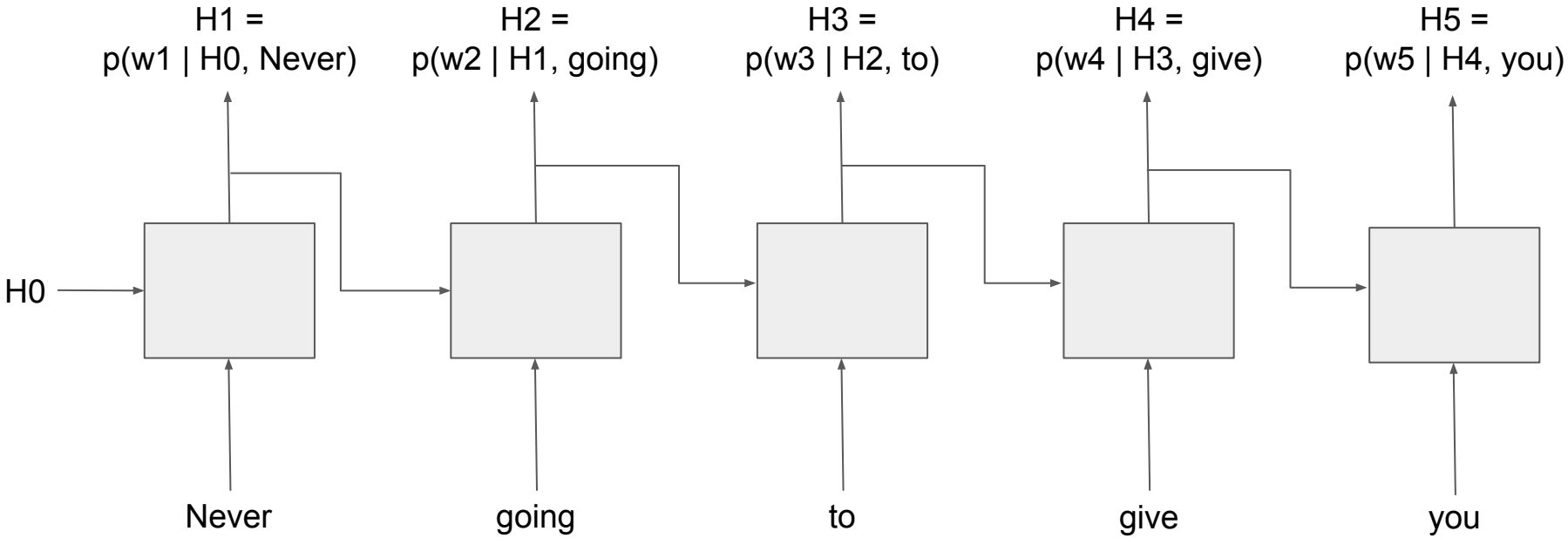
    “Never going to give you up. Never going to let you down.” {N=8}
    [Never, going, to, give, you, up, let, down]
    P(w) = [0.01, 0.01, 0.02, 0.03, 0.03, 0.43, 0.03, 0.44]




https://nlp.stanford.edu/projects/glove/

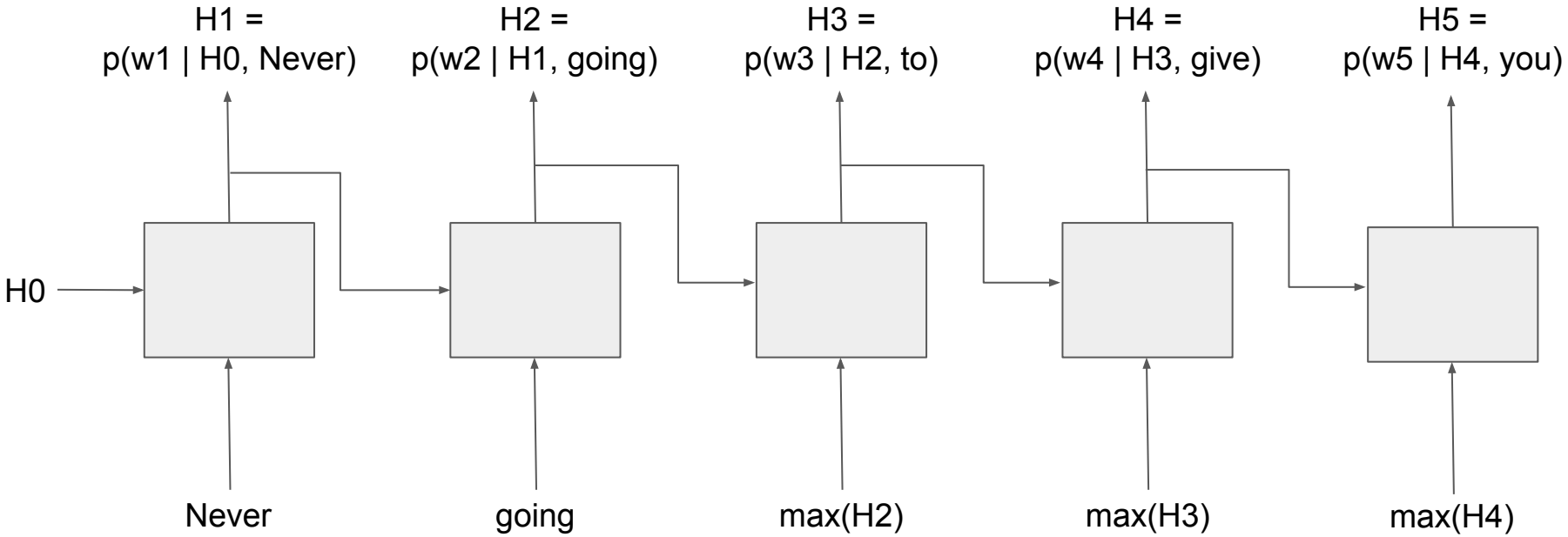
# Prediction Example




Never going to give you ___

Never going to give you up

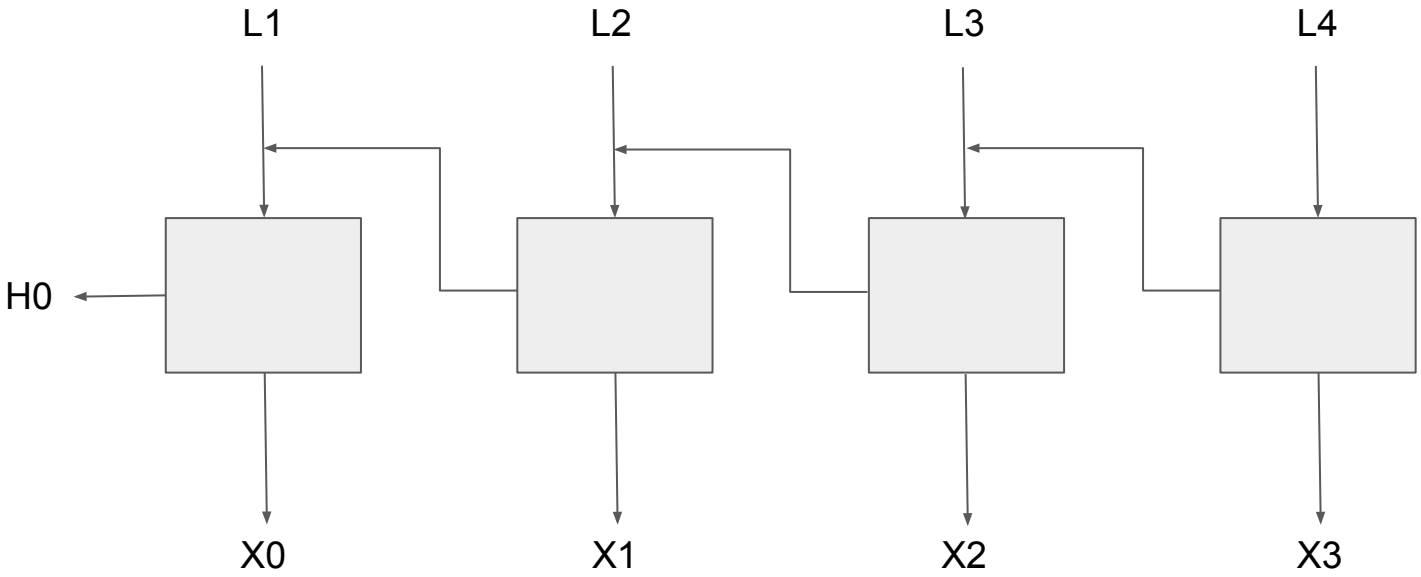
# Generation Example




Never going __ ____ __ __

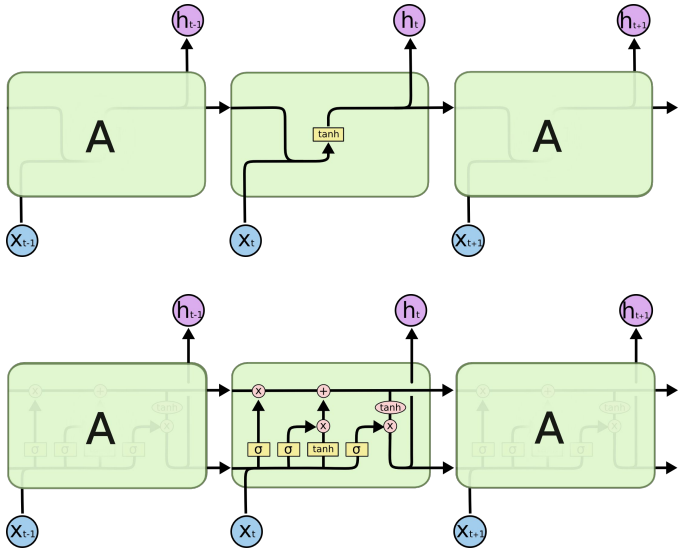
Never going to give you up

# Backpropagation



# Problems Training

- After many iterations
  - Short term memory
  - Vanishing Gradients
  - LSTMs and GRUs combat these issues
- Early training for tasks like generation
  - Cold start - can use teacher forcing
  - Lack of exploration - can use noise
- Long term/non-local dependencies may be reduced or lost
  - Attention




http://colah.github.io/posts/2015-08-Understanding-LSTMs/

# Resources


1. http://colah.github.io/posts/2015-08-Understanding-LSTMs/
2. Vanishing/Exploding Gradients (C2W1L10) - YouTube - https://www.youtube.com/watch?v=qhXZsFVxGKo
3. GloVe: Global Vectors for Word Representation (stanford.edu) - https://nlp.stanford.edu/projects/glove/