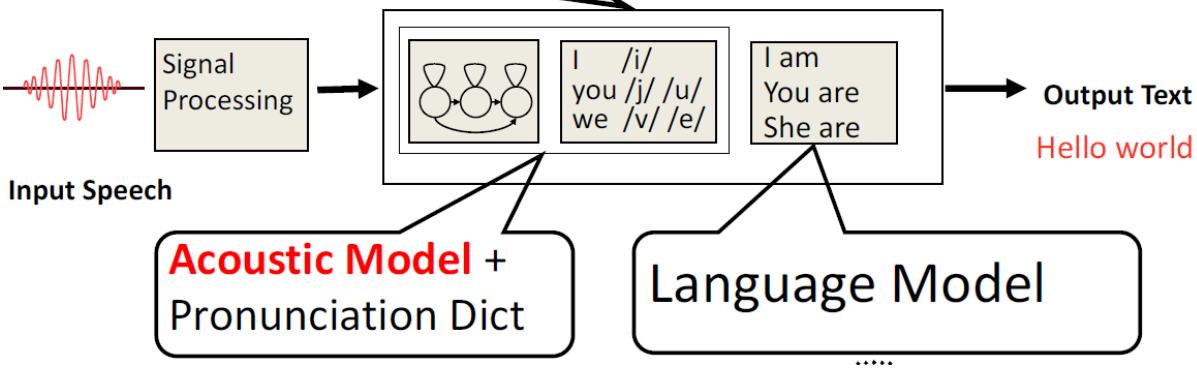
Various Models and Architectures

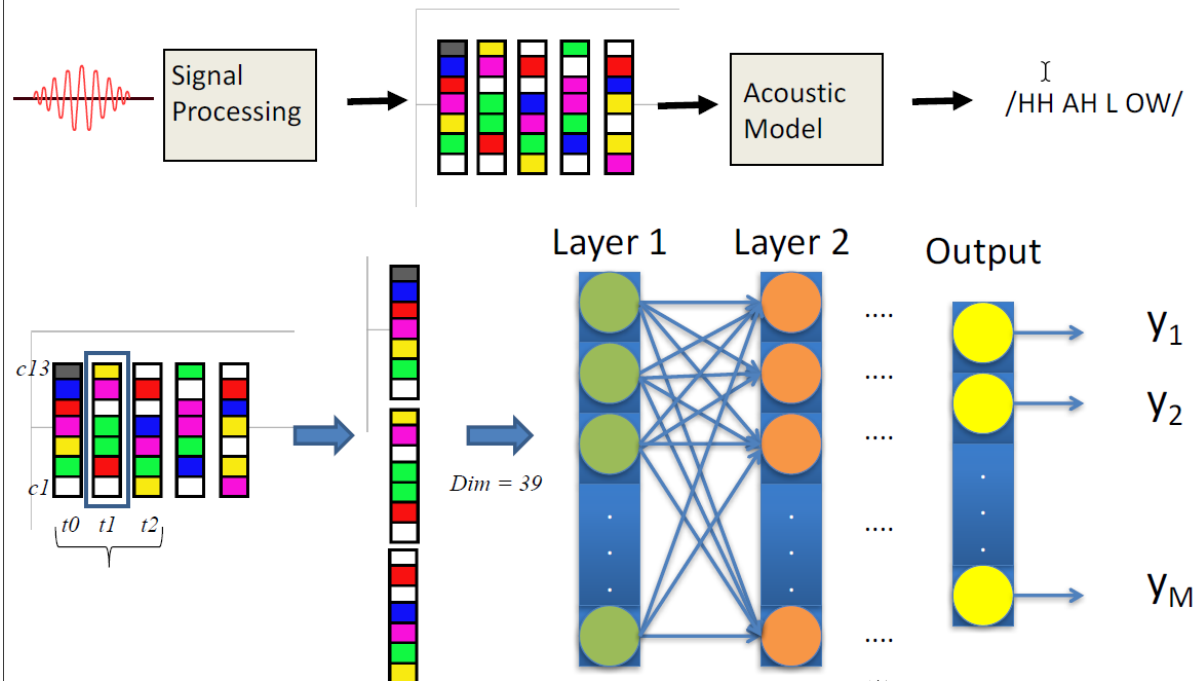
Speech & Language Processing Applications

- Automatic Speech Recognition

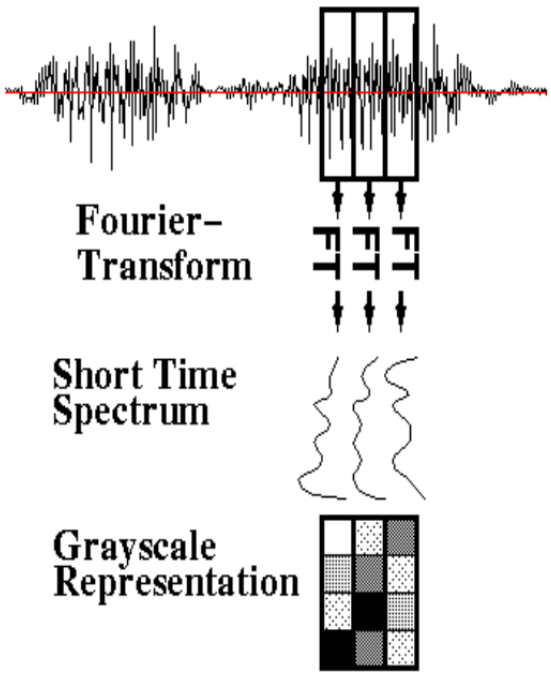
$$\arg\max_W P(W \mid X) = \arg\max_W \frac{P(W)\,p(X \mid W)}{P(X)}$$

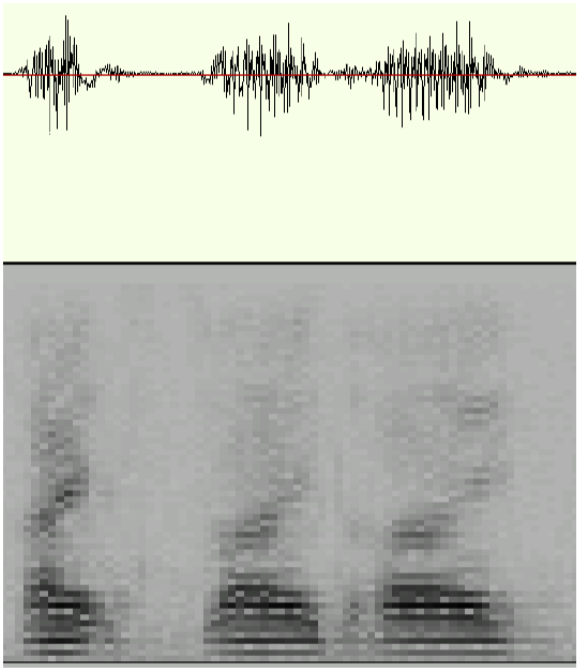


Neural-based Acoustic Models

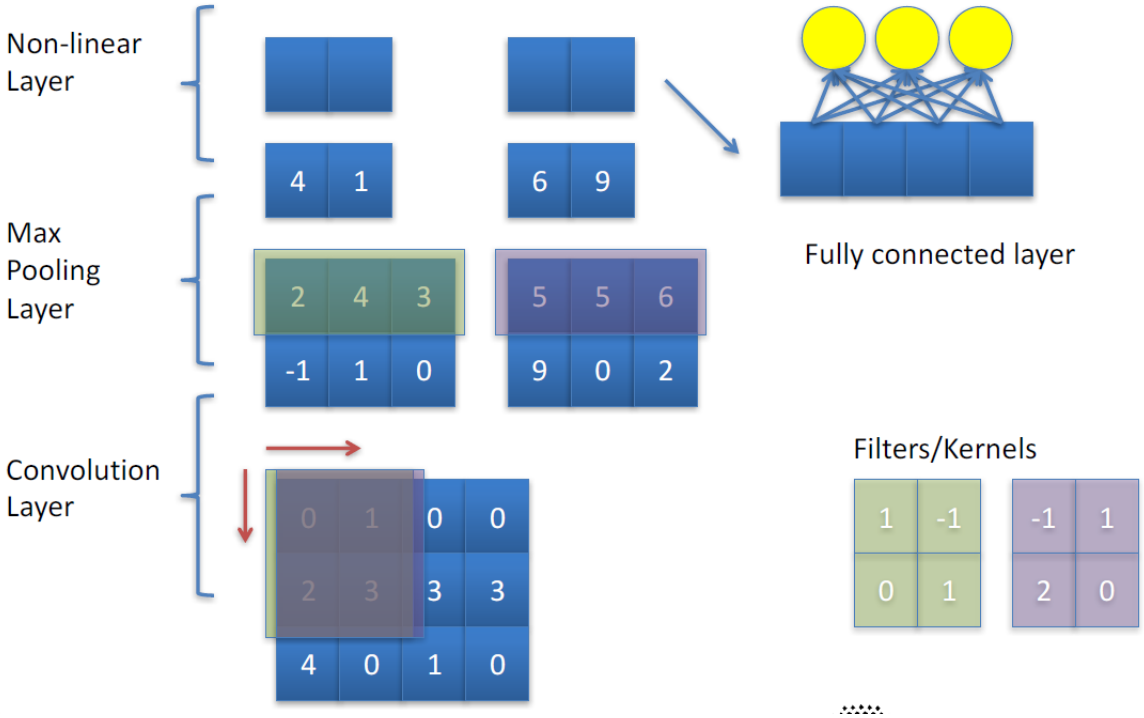


Neural-based Acoustic Models

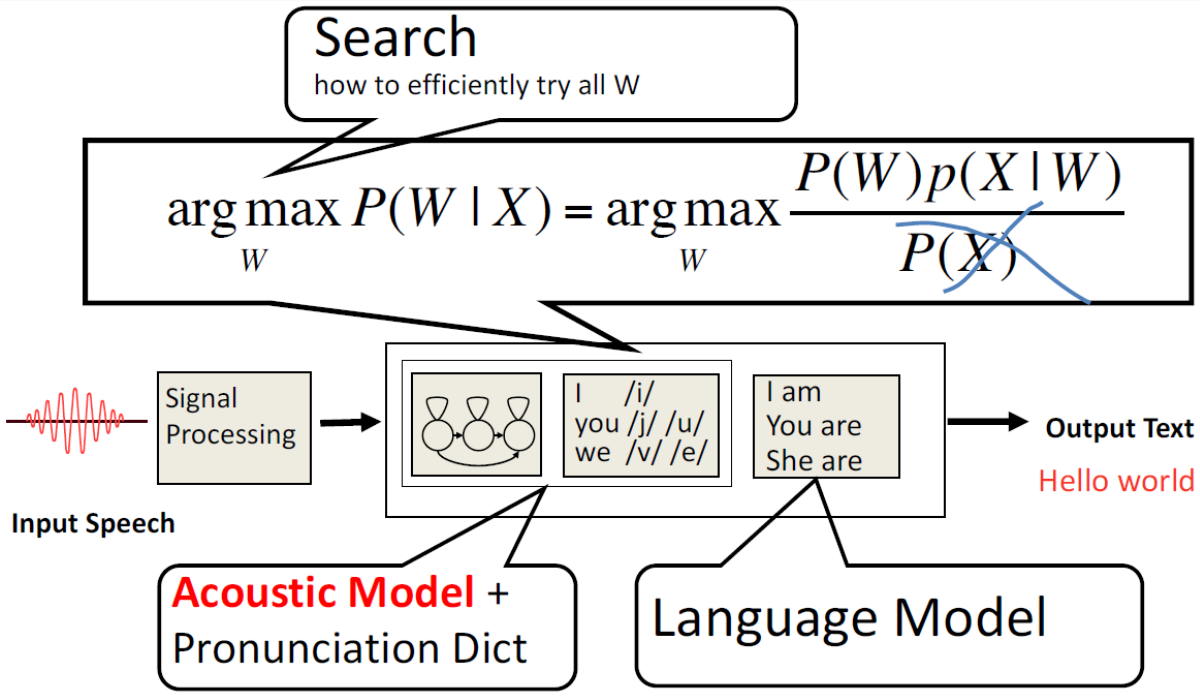




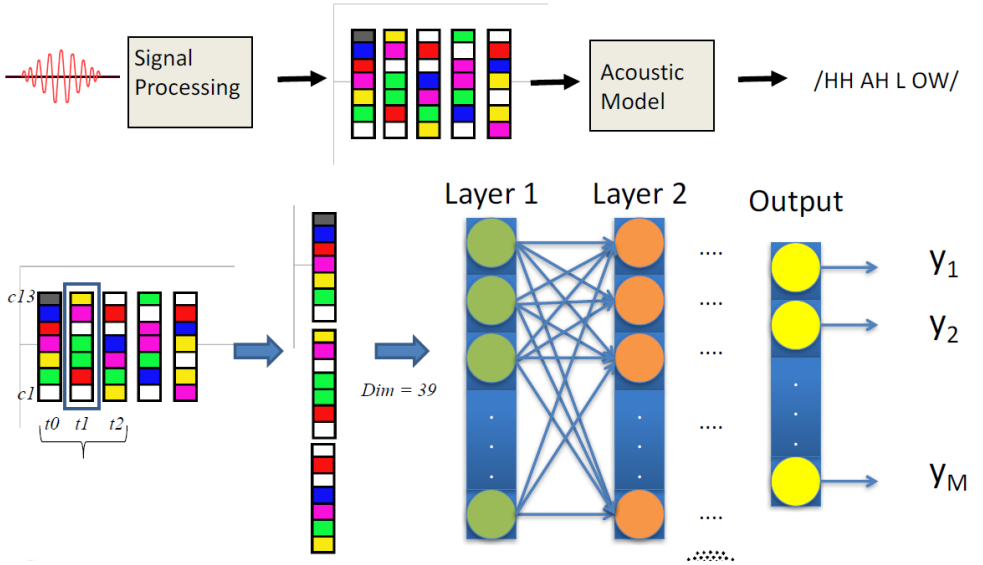
Mel Frequency Cepstral Coefficients (MFCC)
Convolutional Neural Nets



Automatic Speech Recognition

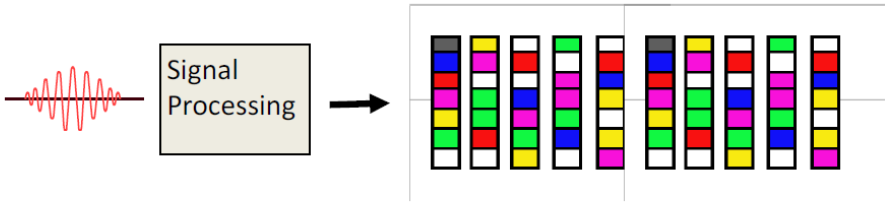


Deep Neural Nets for Acoustic Models

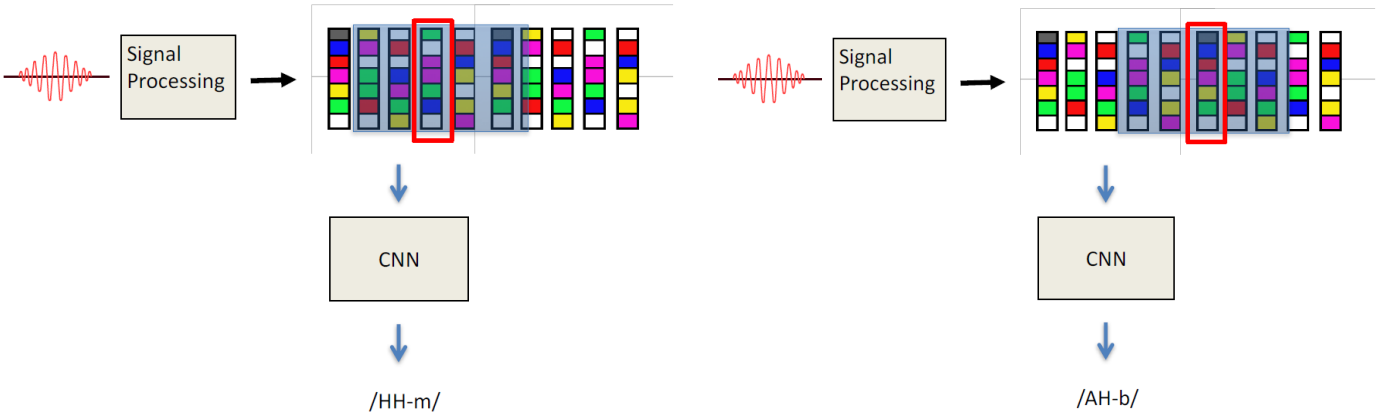


CNN for Acoustic Modelling

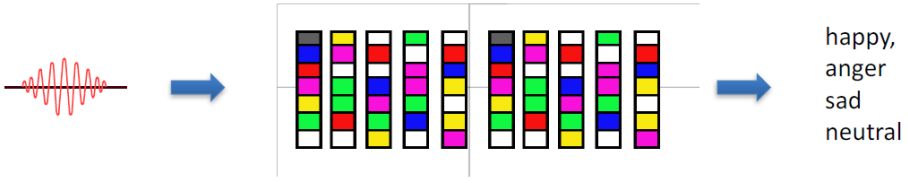
- CNN can be used in speech recognition
- Speech signal ⟶ Sequence of features



- CNN has shown to improve the ASR performance
  - IBM Paper, 2013
  - Now, it belongs to one of the state-of-the-art techniques

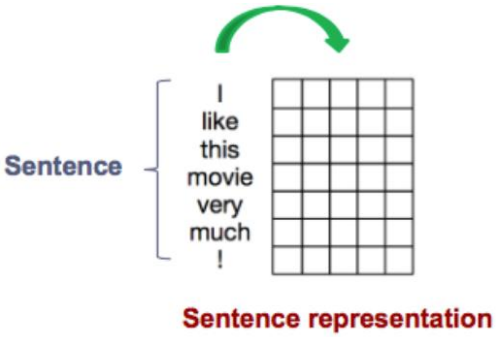


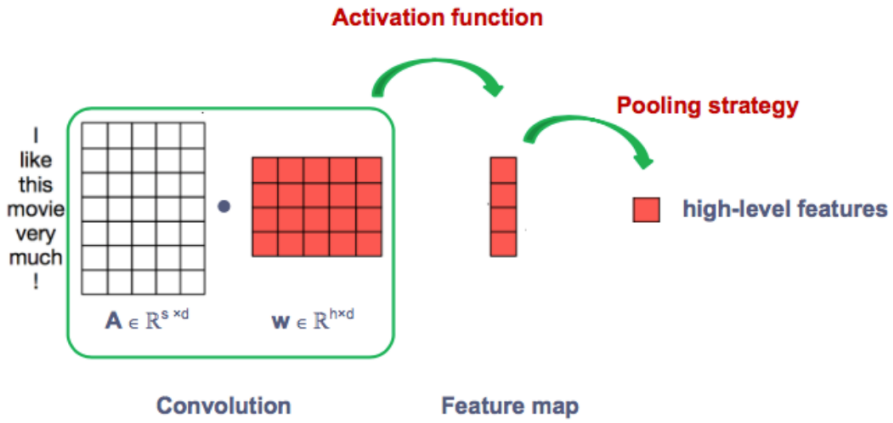
- Extract emotion from speech



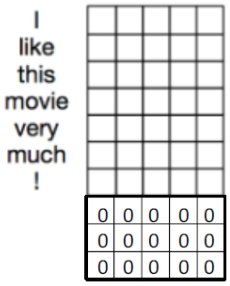
CNN for Natural Language Processing

- If the input is a sentence, where can we get the image for CNN?
  - Word representations (word vectors / word embeddings): a word can be represented as a vector



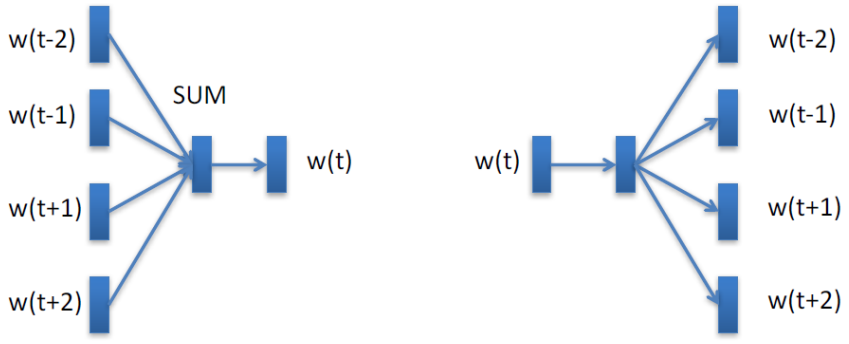


- What if the sentences have different length?
  - Define a max sentence length
  - Pad the sentence with zero vectos or cut the sentence to the max sentence length
  - E.g. with a max sentence length of 10
    - ‚I like this movie very much!‘ can be represented as follows



Word Embeddings

- Another way was proposed by Mikolov et al, (2013)
- https://code.google.com/p/word2vec/

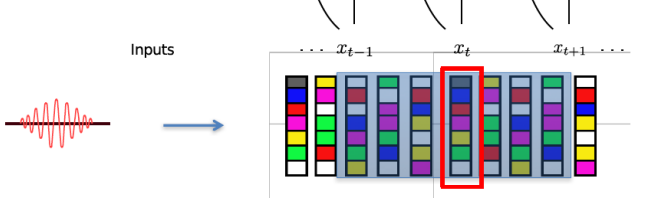


CBOW

Skip-gram

RNN for Acoustic Modelling

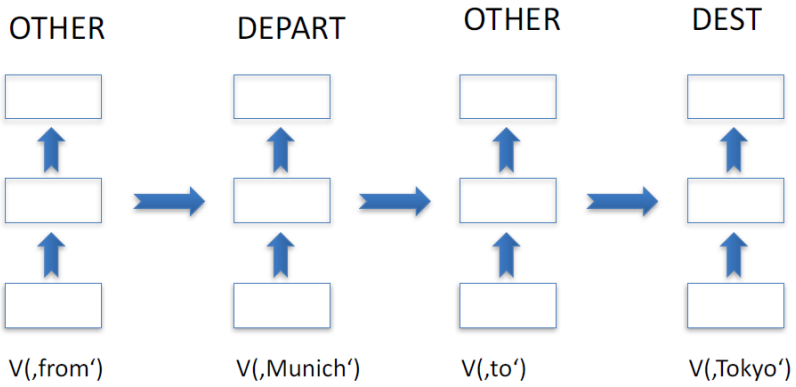
- RNN can be used in speech recognition
- Speech signal ⟶ Sequence of features



- RNN has shown to improve the ASR performance
  - Graves et al, 2013 – 2966 citations
  - Now, it belongs to one of the state-of-the-art techniques

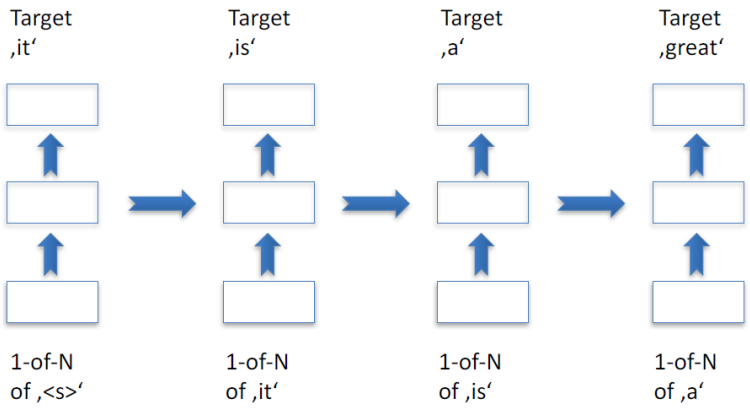
RNN for Slot filling

- e.g. A ticket from Munich to Tokyo please



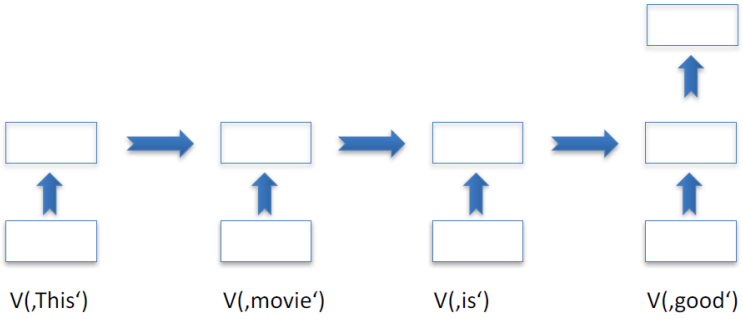
RNN for Language Modelling

- Training data: ‚It is a great day‘



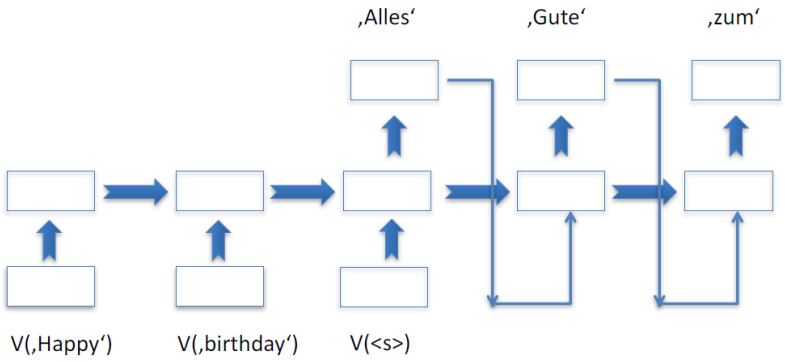
RNN for Sentiment Analysis

- Input is a sentence
- Output is positive, negative or neutral



RNN for Machine Translation

- Input is a sentence in the source language
- Output is also a sentence but in the target language



RNN for Chit-chat Dialog Modelling

- Input is fed to the source side
- Output of the system in the target side