NLP, Deep Learning & Some Examples

Reyyan Yeniterzi

NLP, Deep Learning & Some E[History]s

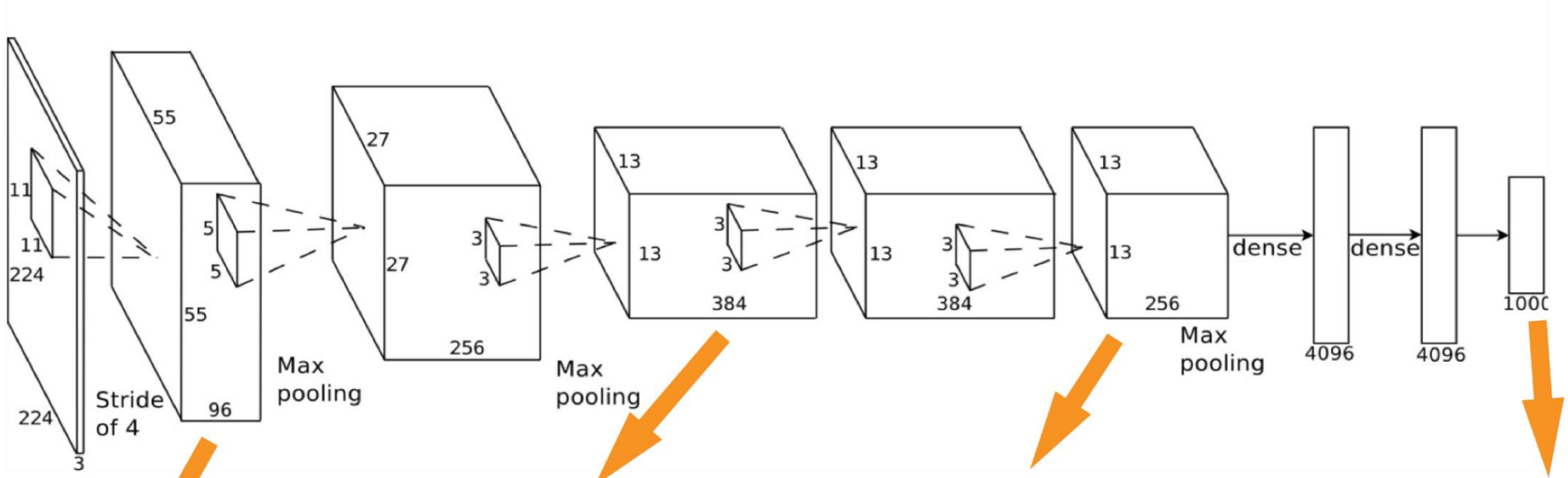
Reyyan Yeniterzi

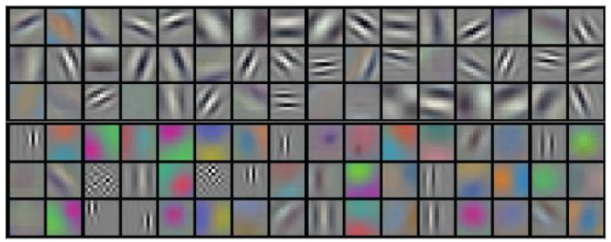
2018
the year of NLP

2018

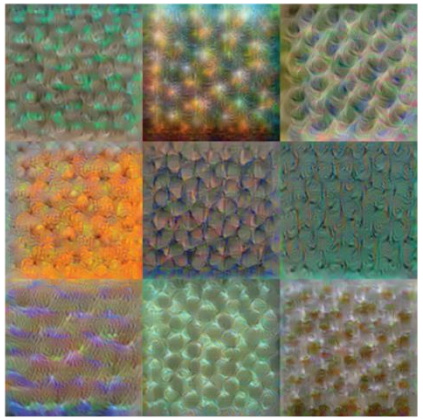
the year of NLP

The ImageNet moment of NLP





Conv 1: Edge+Blob



Conv 3: Texture

Numerical

Data-driven

Conv 5: Object Parts

Fc8: Object Classes

https://www.cc.gatech.edu/~hays/compvision/proj6/


“Transfer learning will be the next driver of ML success.”
Andrew Ng,
NIPS 2016 tutorial

Transfer Learning in NLP?

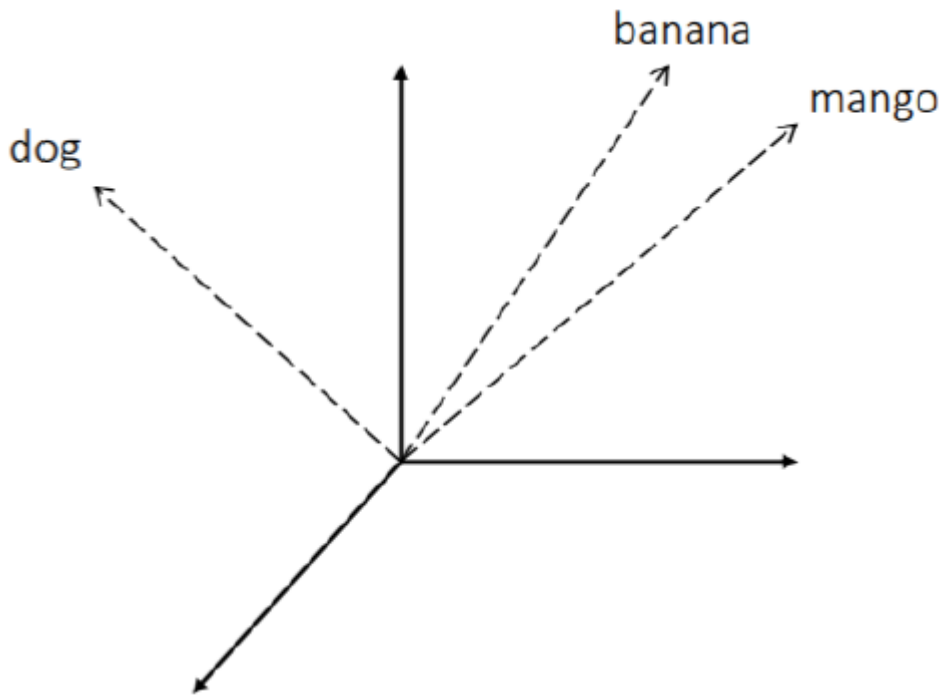
“You shall know a word by the company it keeps”

J. R. Firth, 1957

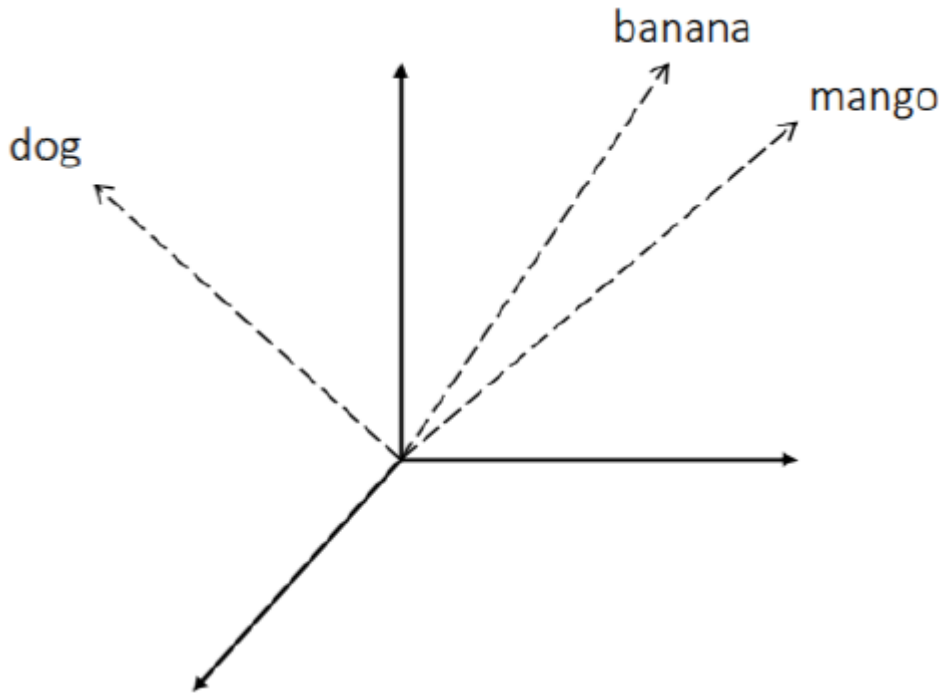
What does *wampimuk* mean?

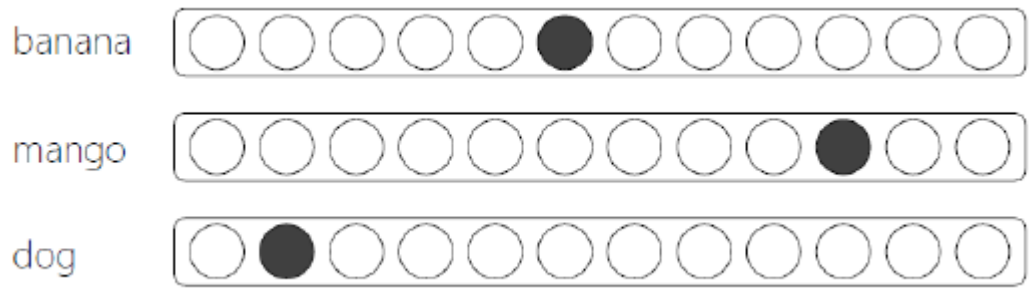
What does *wampimuk* mean?

Marco saw a hairy little wampimuk crouching behind a tree.

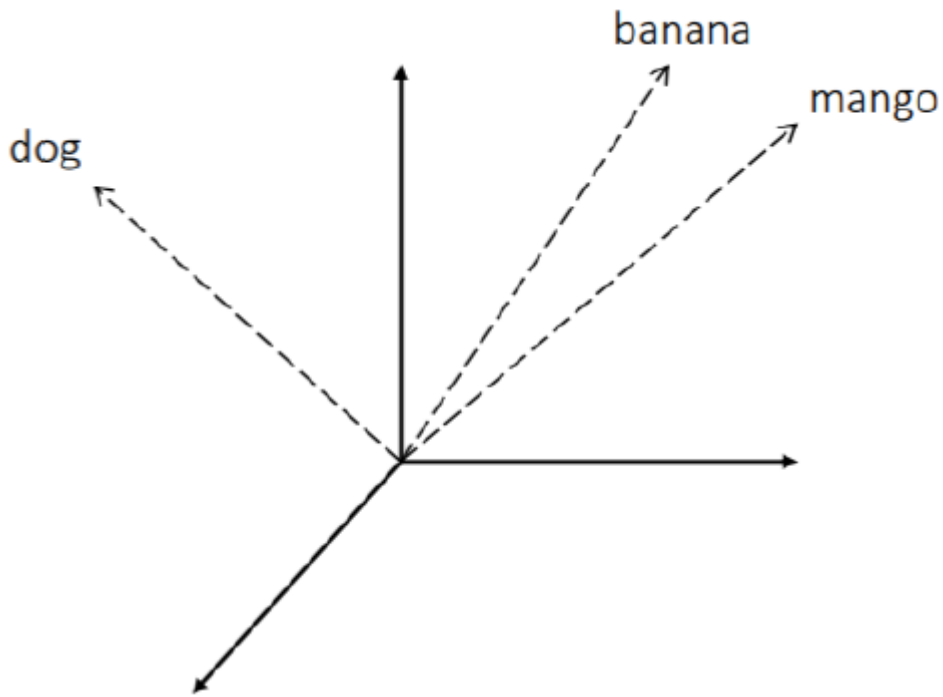


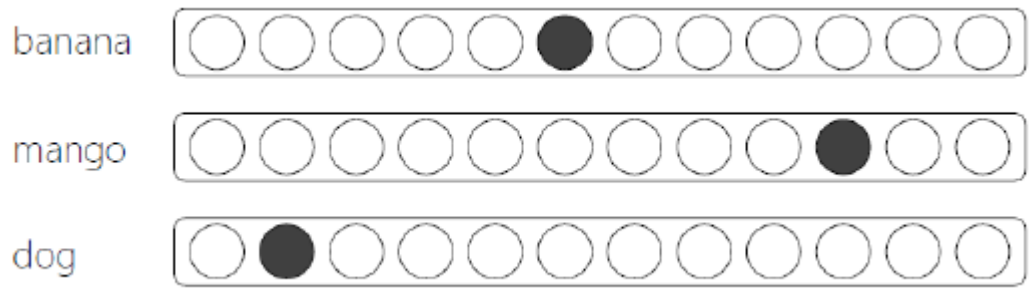
https://www.microsoft.com/en-us/research/uploads/prod/2018/04/NeuralIR-Nov2017.pdf

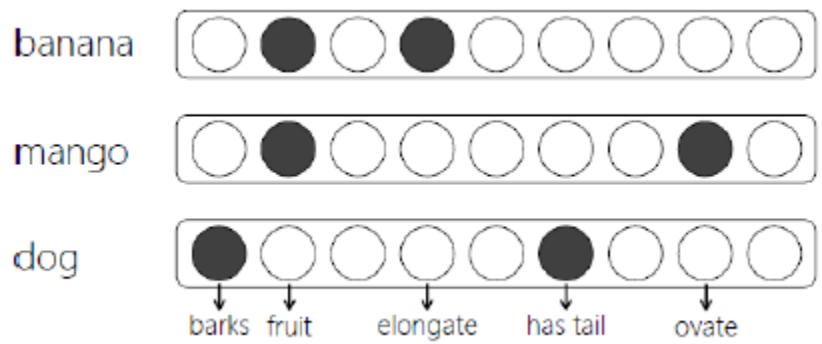




https://www.microsoft.com/en-us/research/uploads/prod/2018/04/NeuralIR-Nov2017.pdf



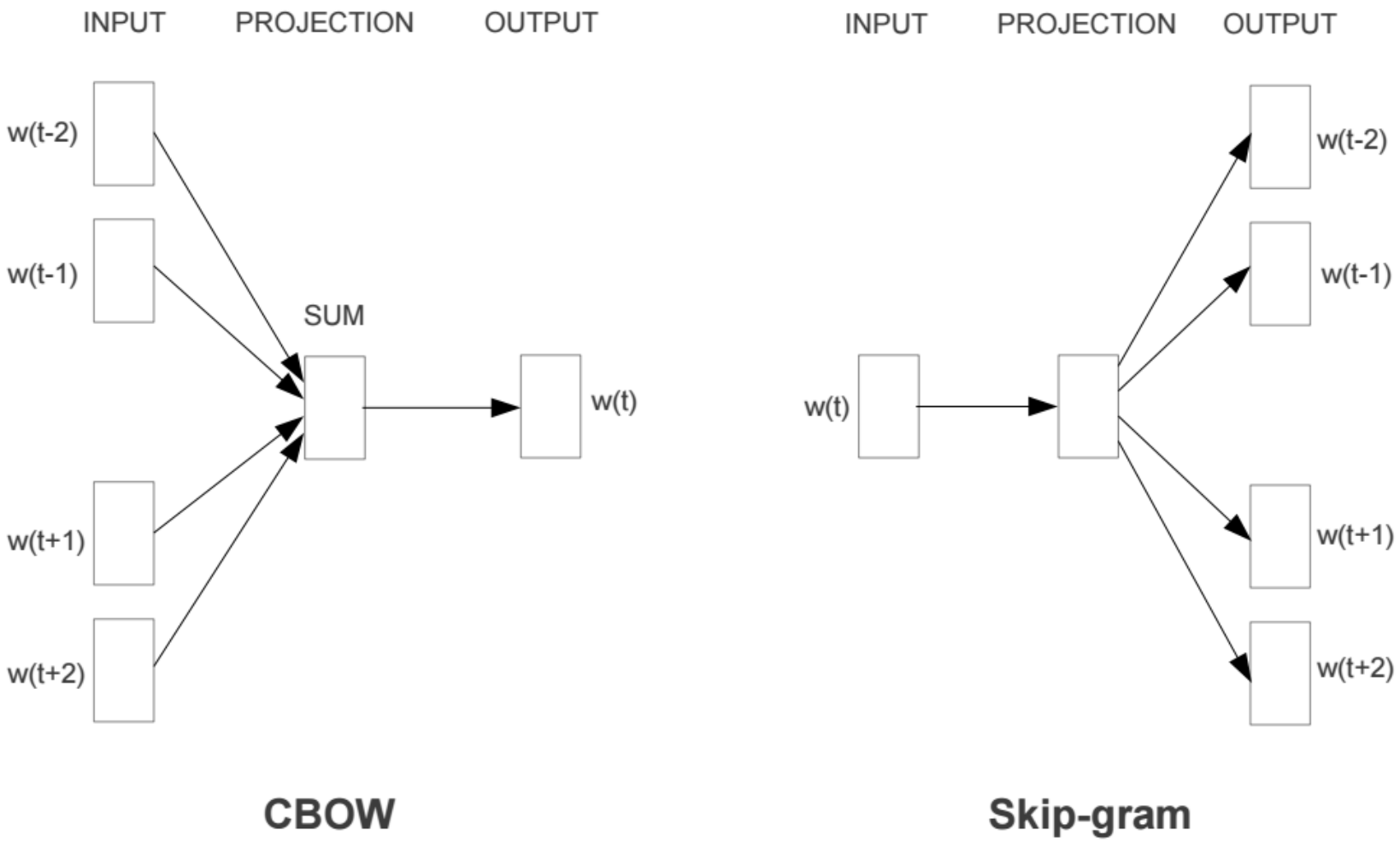




https://www.microsoft.com/en-us/research/uploads/prod/2018/04/NeuralIR-Nov2017.pdf

Word2Vec

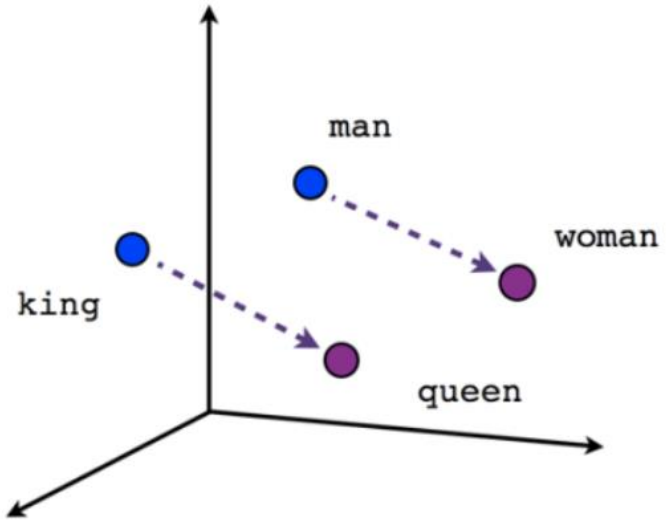
2013



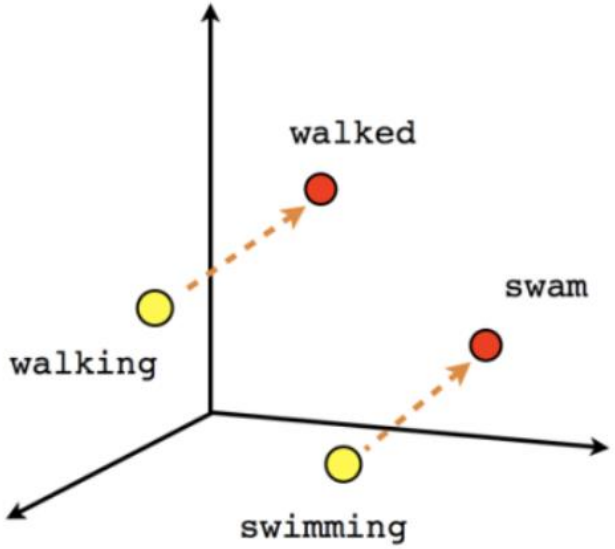
Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space". 2013

Word2Vec

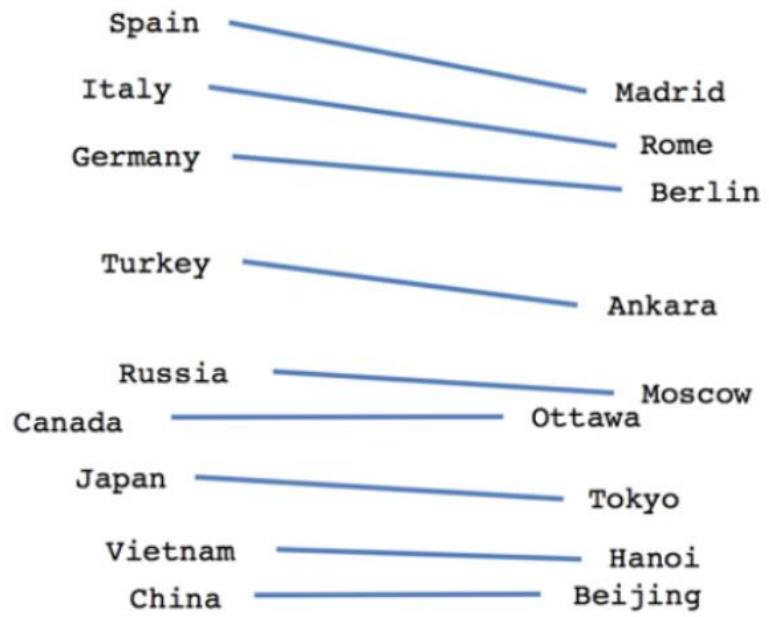
2013



Male-Female



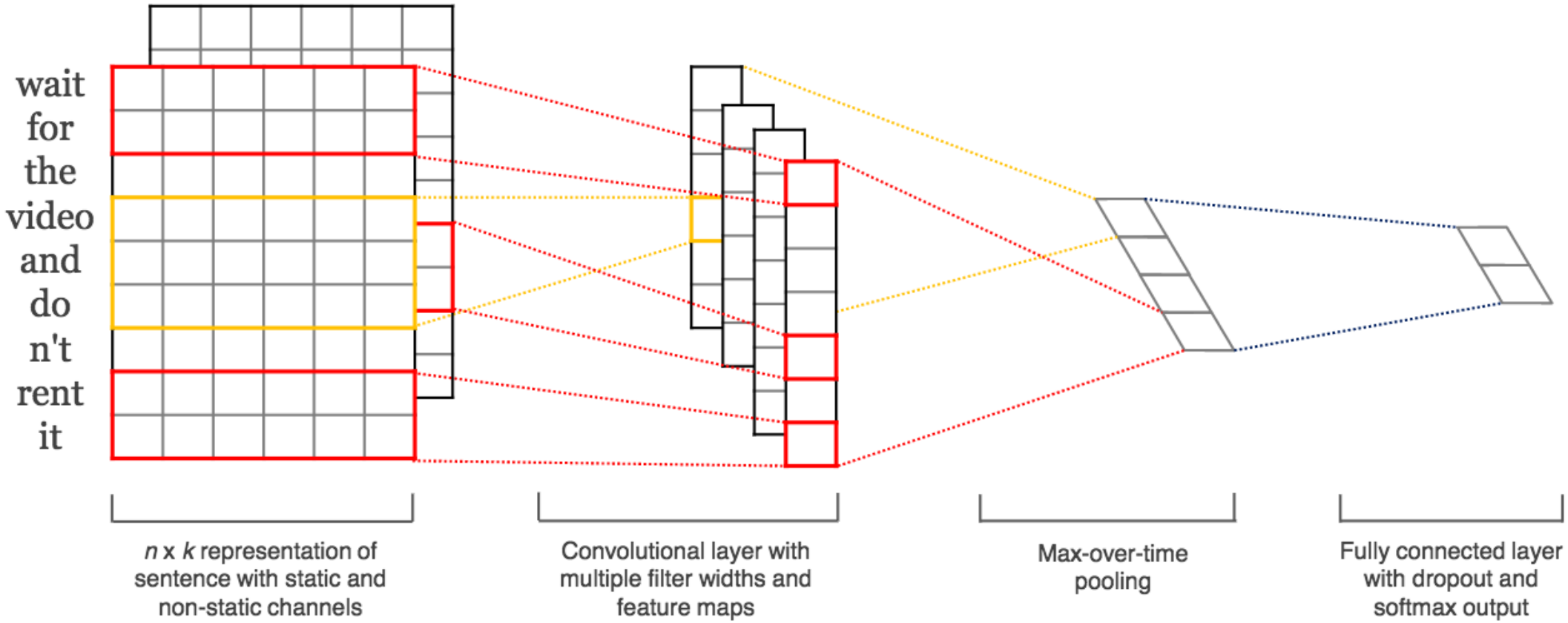
Verb tense



Country-Capital

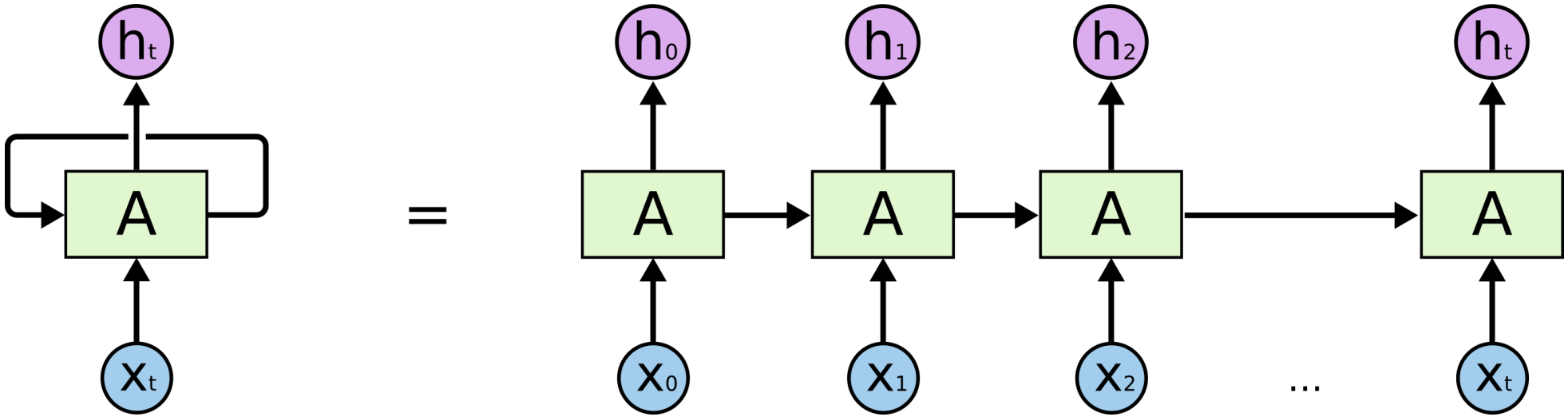
CNN

2014



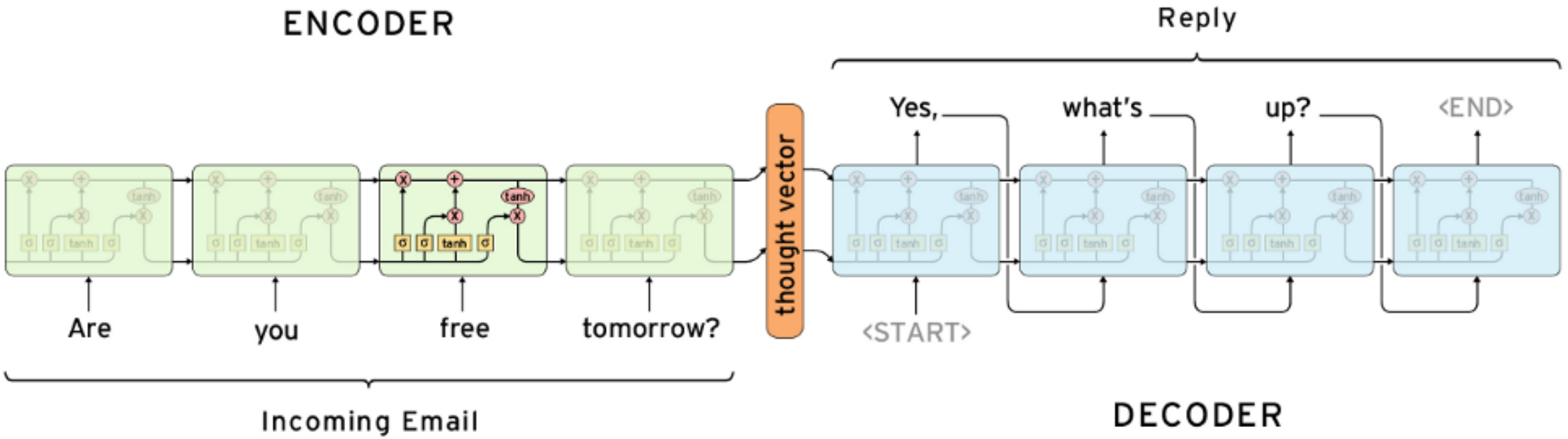
Kim, Yoon. "Convolutional neural networks for sentence classification." 2014.

RNN



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

Encoder&Decoder

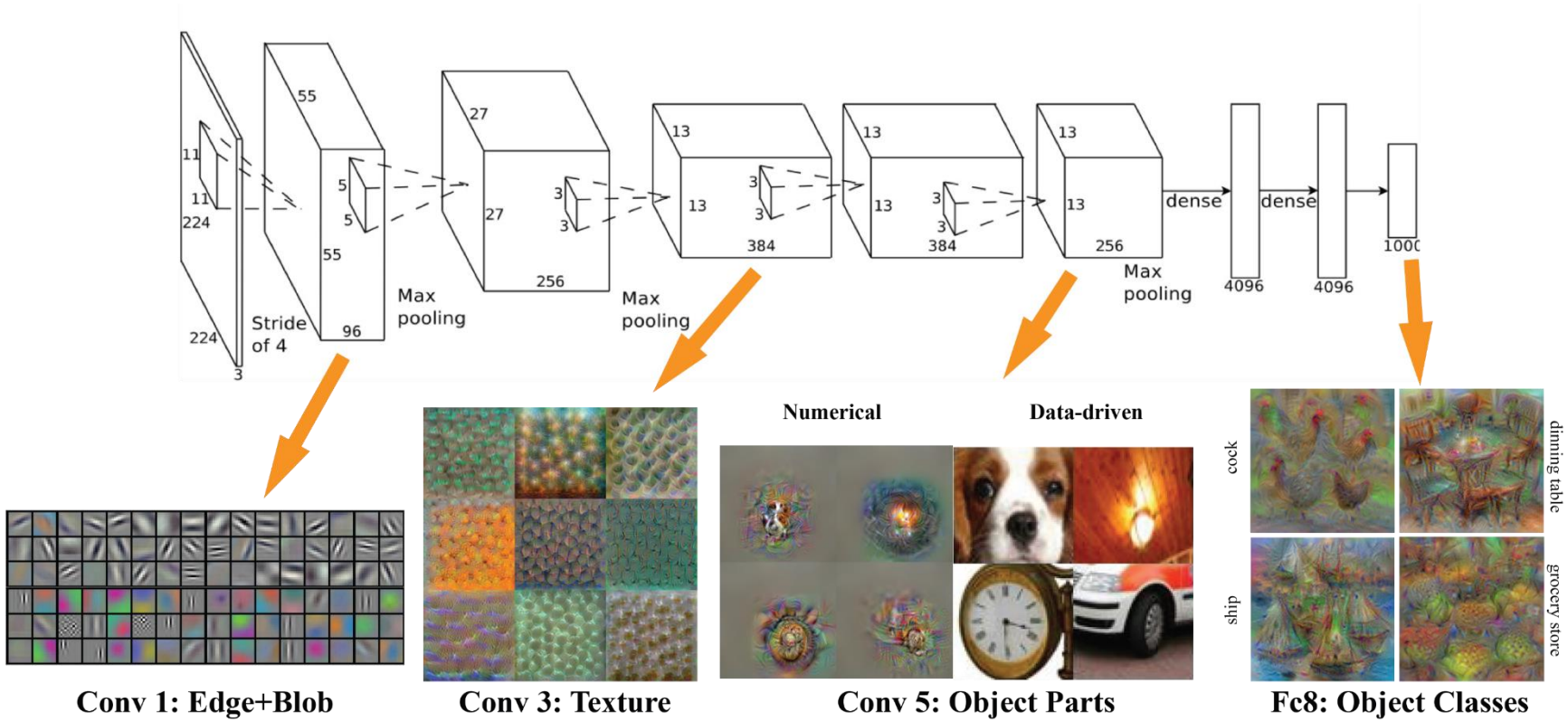


https://mc.ai/multi-modal-methods-image-captioning-from-translation-to-attention/

Transfer learning in NLP through the embedding layer

Shallow represention is not enough,

still require large training sets



Word embeddings are cool,

but context matters!

Word embeddings are cool,

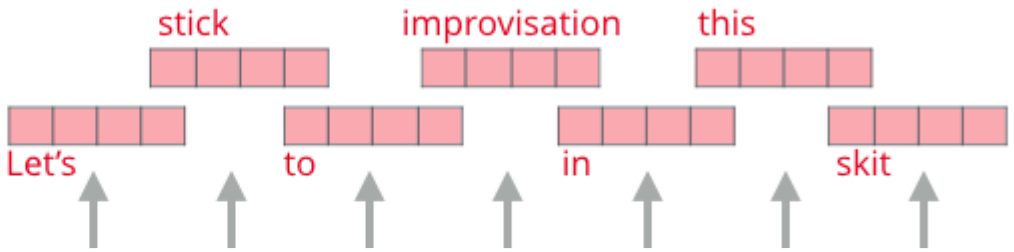
but context matters!

It is time for contextual word embeddings

ELMo

2018

ELMo Embeddings



Embeddings from **L**anguage **Mo**dels

Words to embed

Peters, Matthew E., et al. "Deep contextualized word representations." 2018.

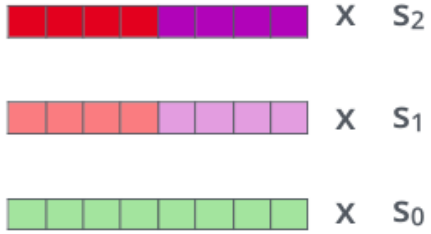
https://jalammar.github.io/illustrated-transformer/

Embedding of “stick” in “Let’s stick to” - Step #2

2018

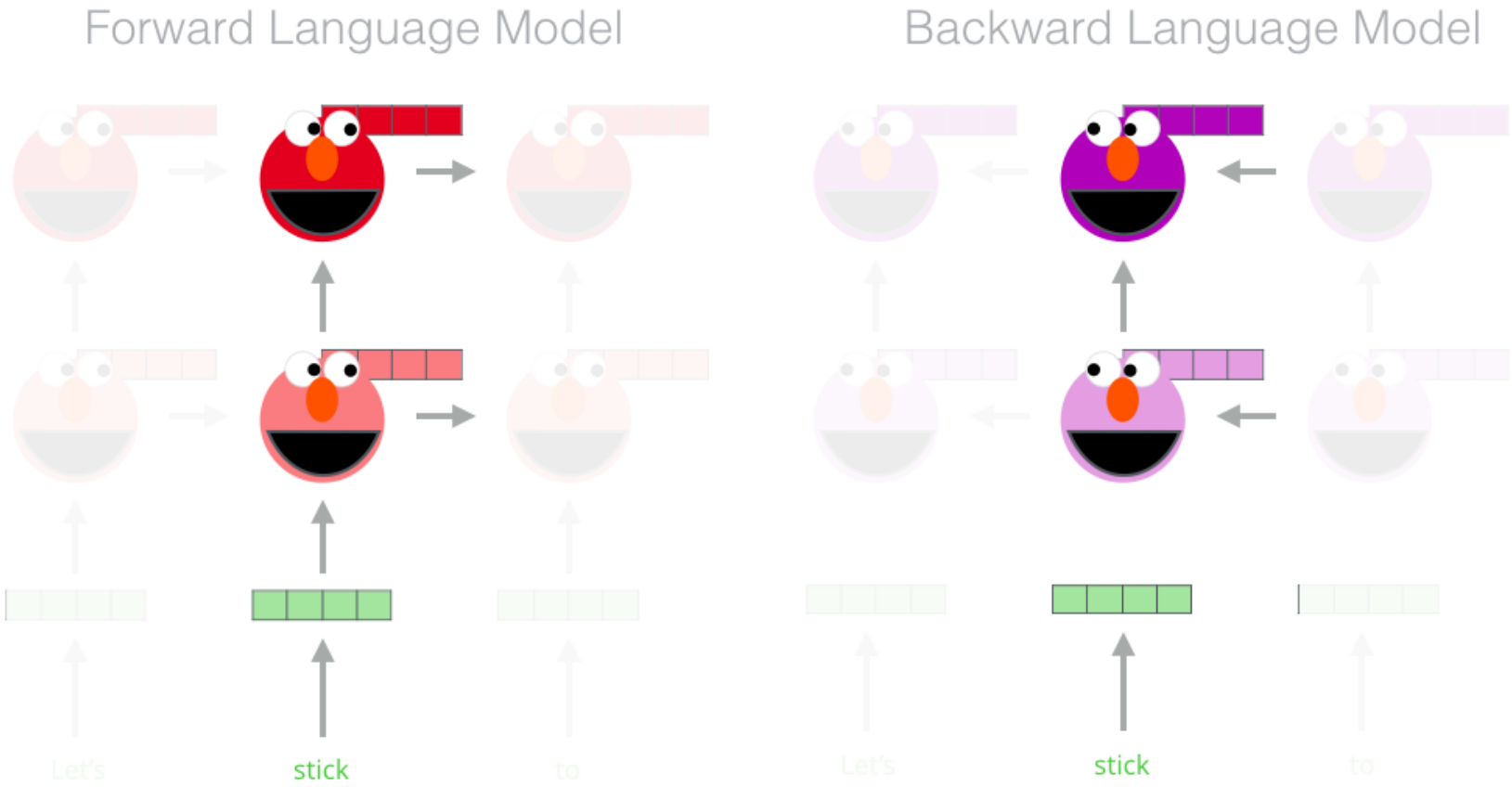
1- Concatenate hidden layers

2- Multiply each vector by a weight based on the task



3- Sum the (now weighted) vectors

ELMo embedding of “stick” for this task in this context



Peters, Matthew E., et al. "Deep contextualized word representations." 2018.

https://jalammar.github.io/illustrated-transformer/

2018

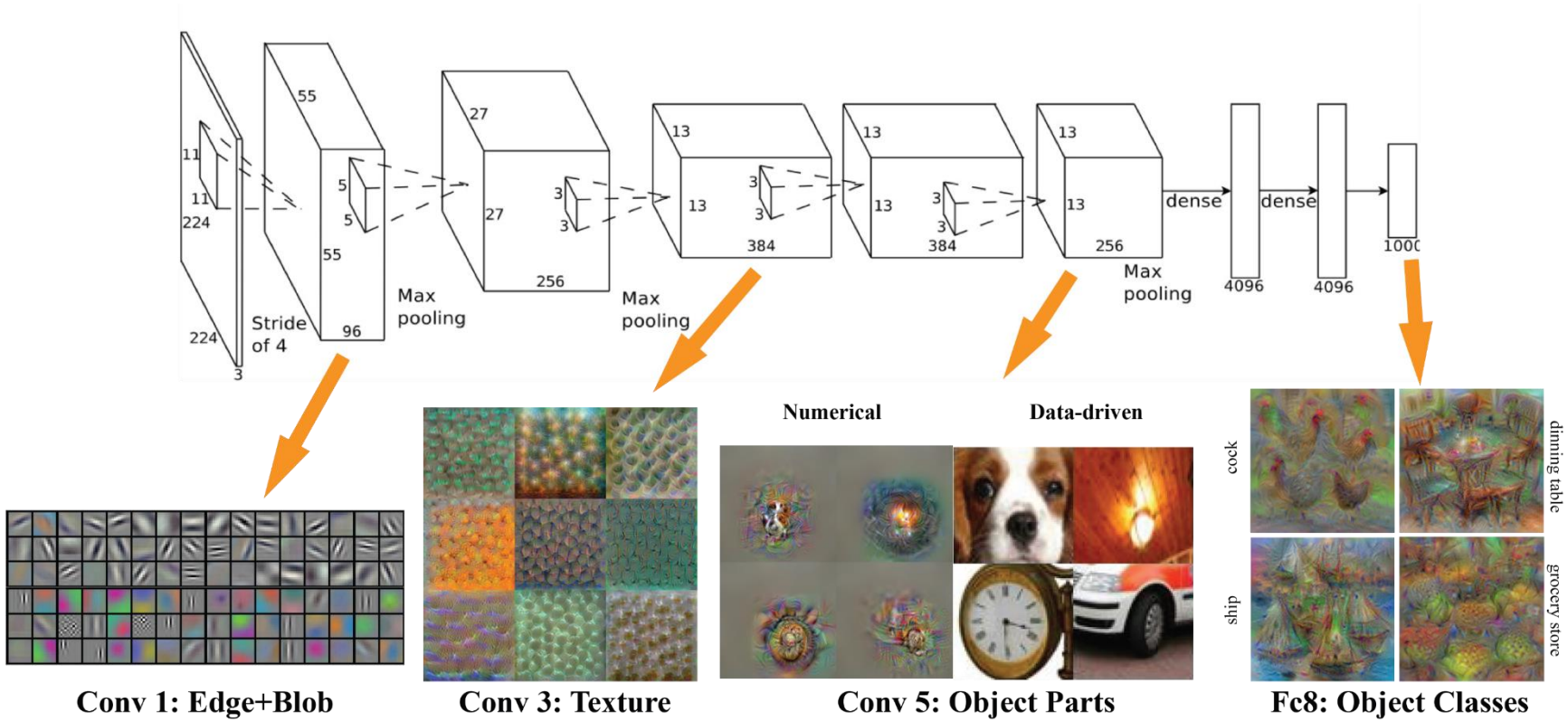
ELMo Contextualized embeddings

2013 Word2Vec Context free embeddings

Transfer learning in NLP through the embedding layer

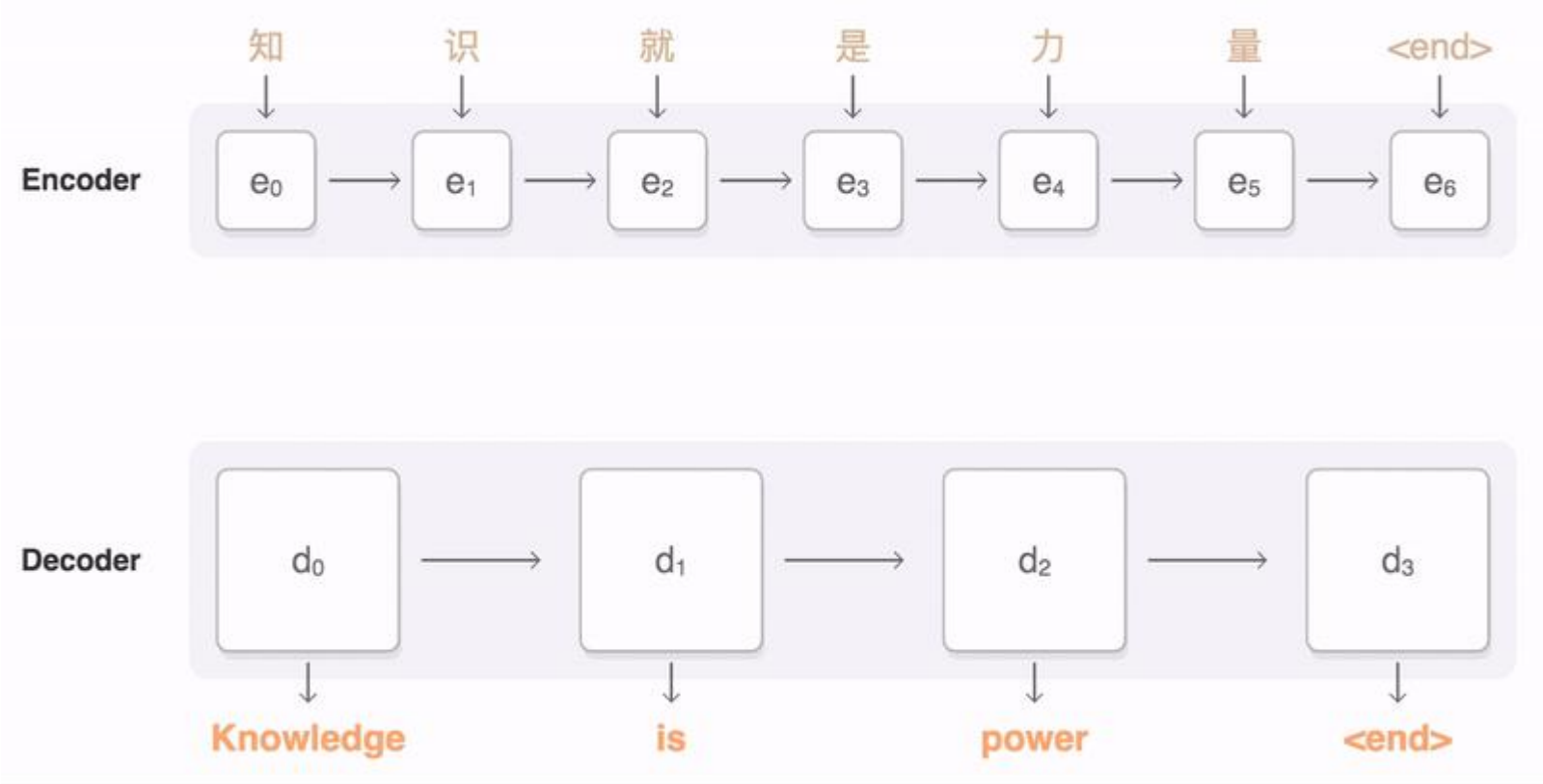
Shallow represention is not enough,

still require large training sets

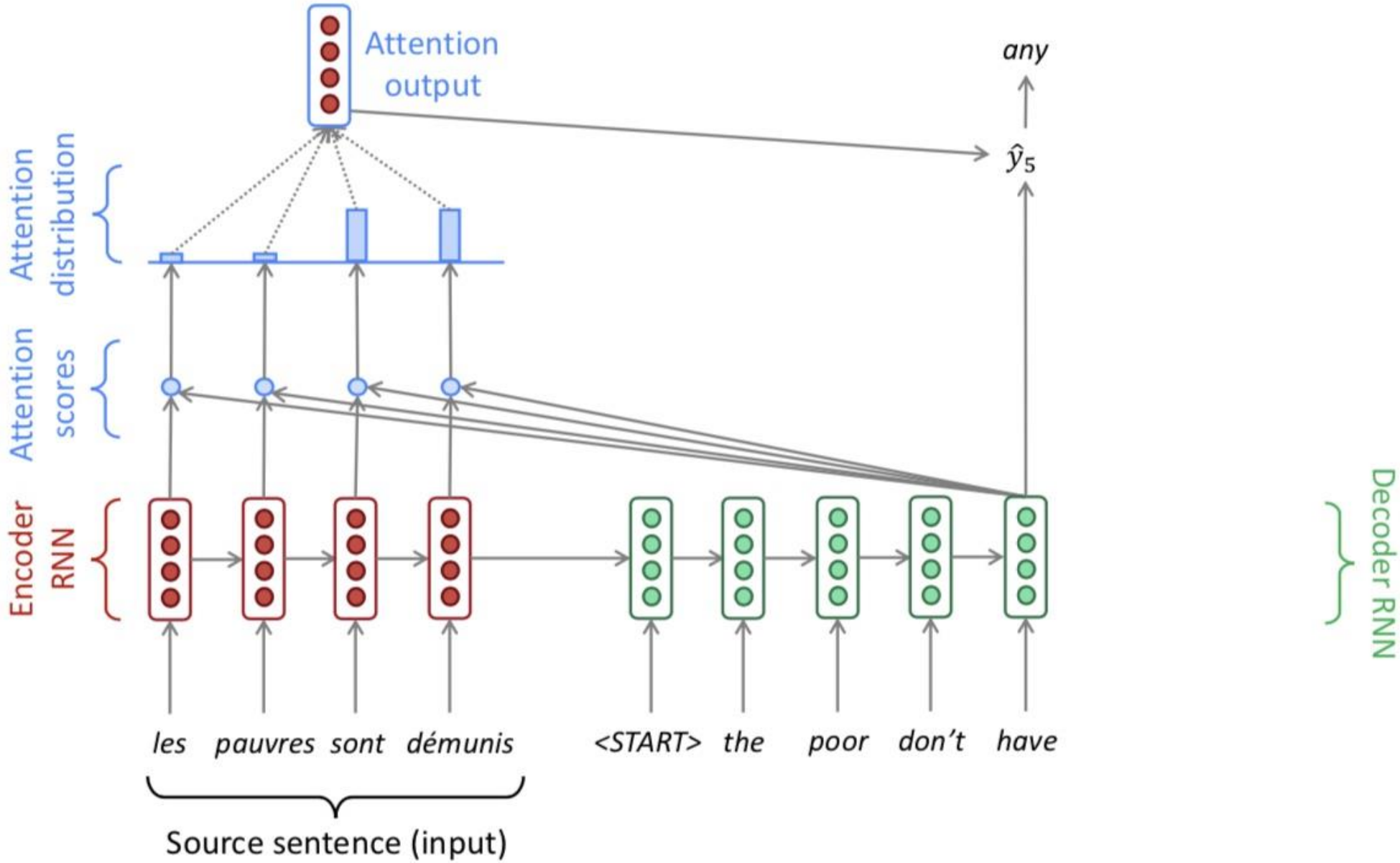


Attention

2014

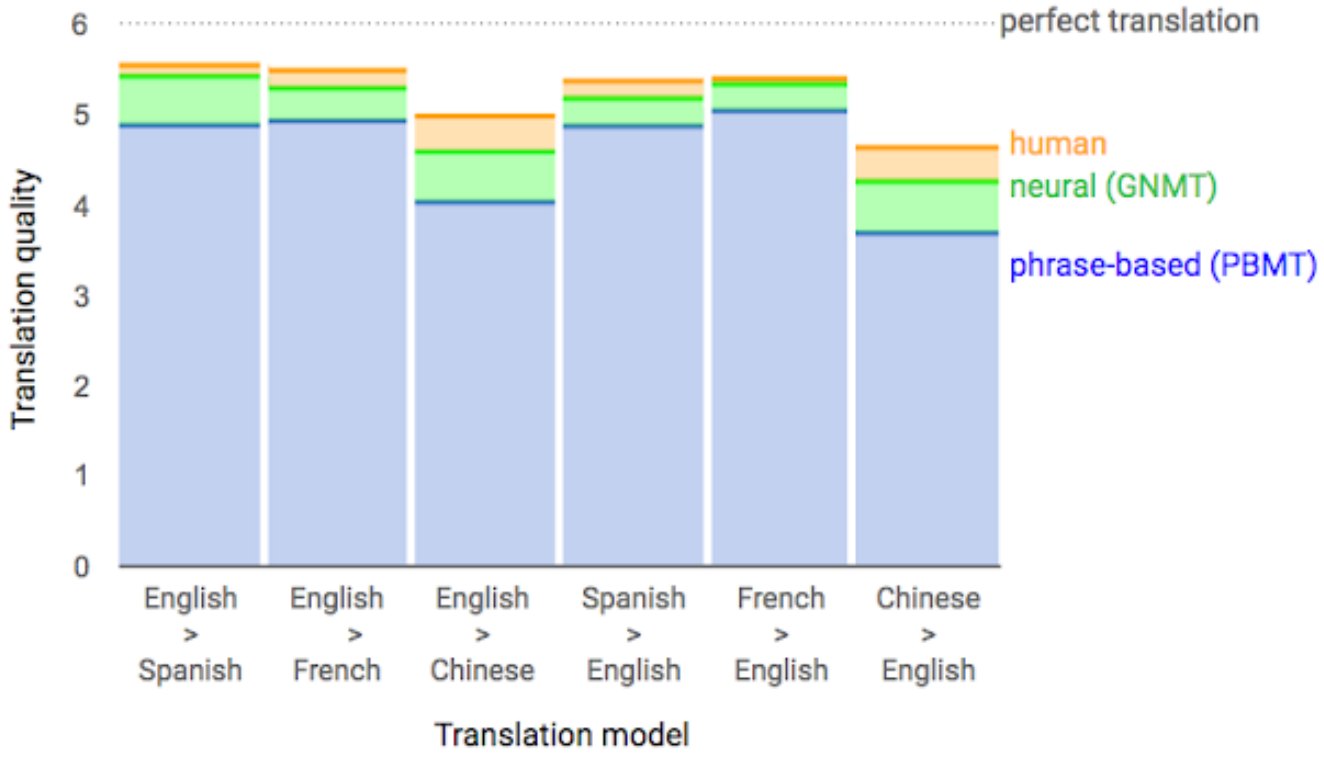


Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." 2014.
https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html



http://wangyp.tech/2019/01/05/Transformer/

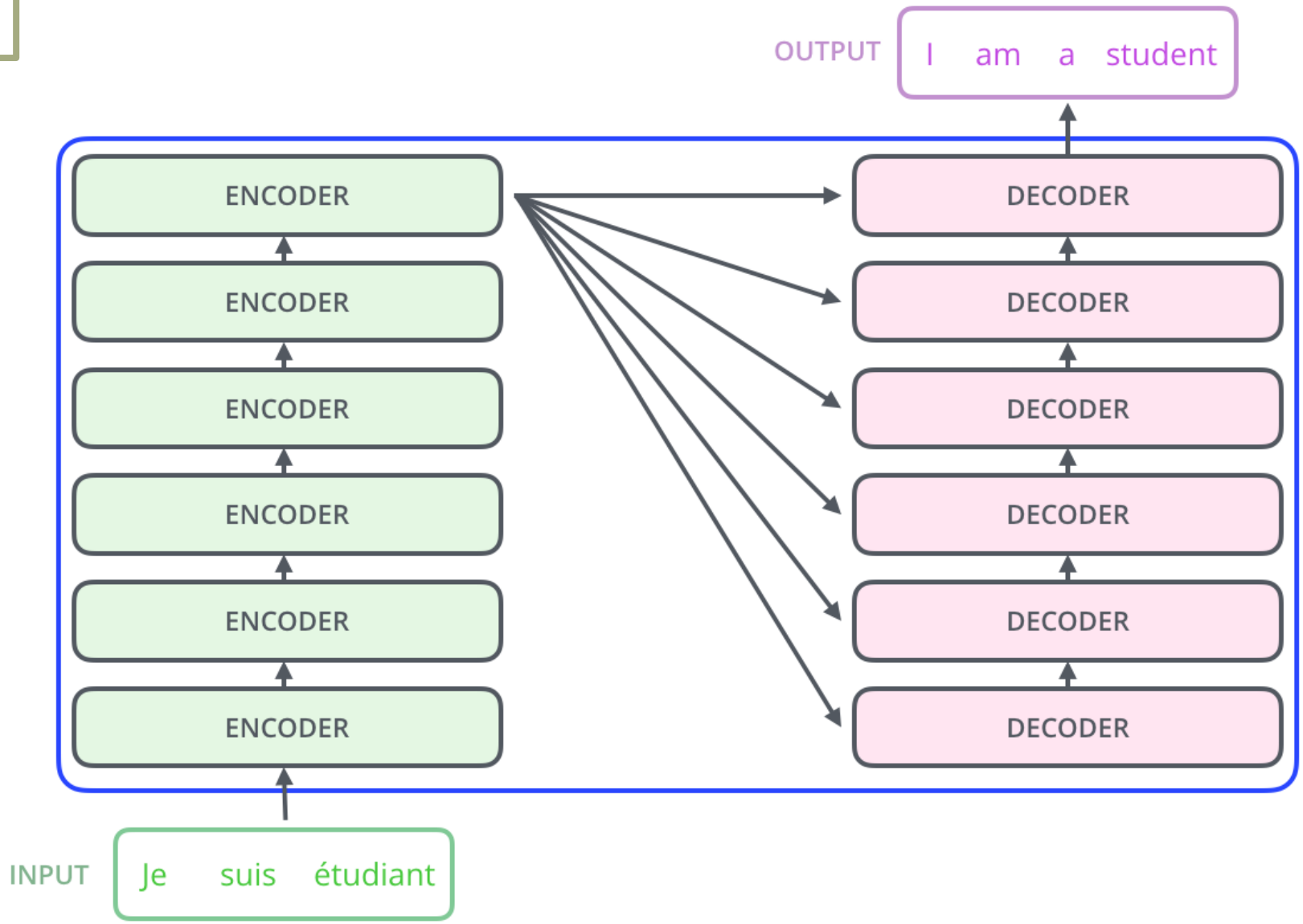
2016



https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Transformer

2017



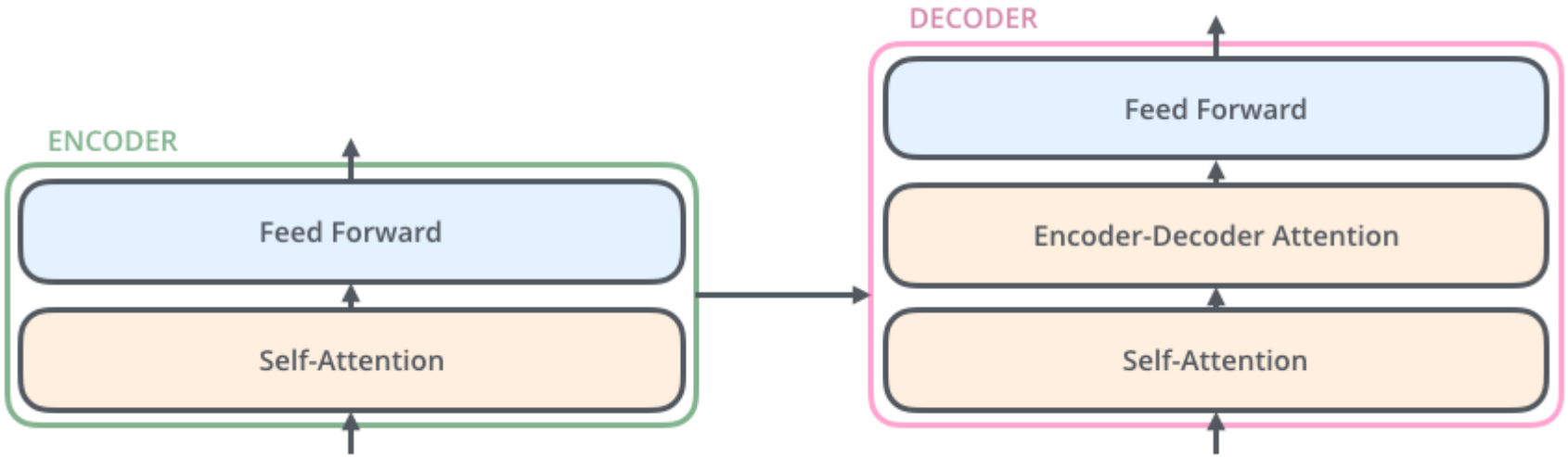
Vaswani, Ashish, et al. "Attention is all you need." *NIPS*. 2017.

https://jalammar.github.io/illustrated-transformer/

Transformer

Self-Attention

2017

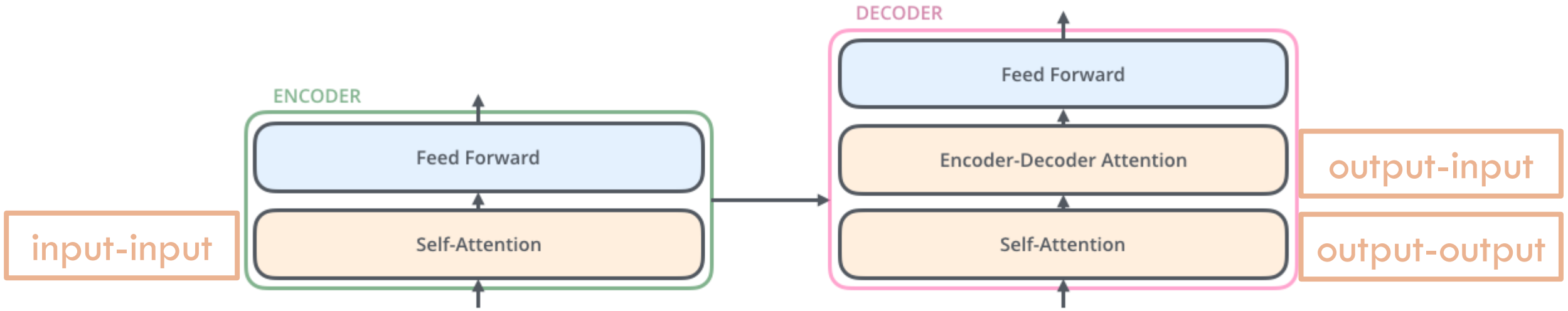


Vaswani, Ashish, et al. "Attention is all you need." *NIPS*. 2017.
https://jalammar.github.io/illustrated-transformer/

Transformer

Self-Attention

2017



Vaswani, Ashish, et al. "Attention is all you need." *NIPS*. 2017.

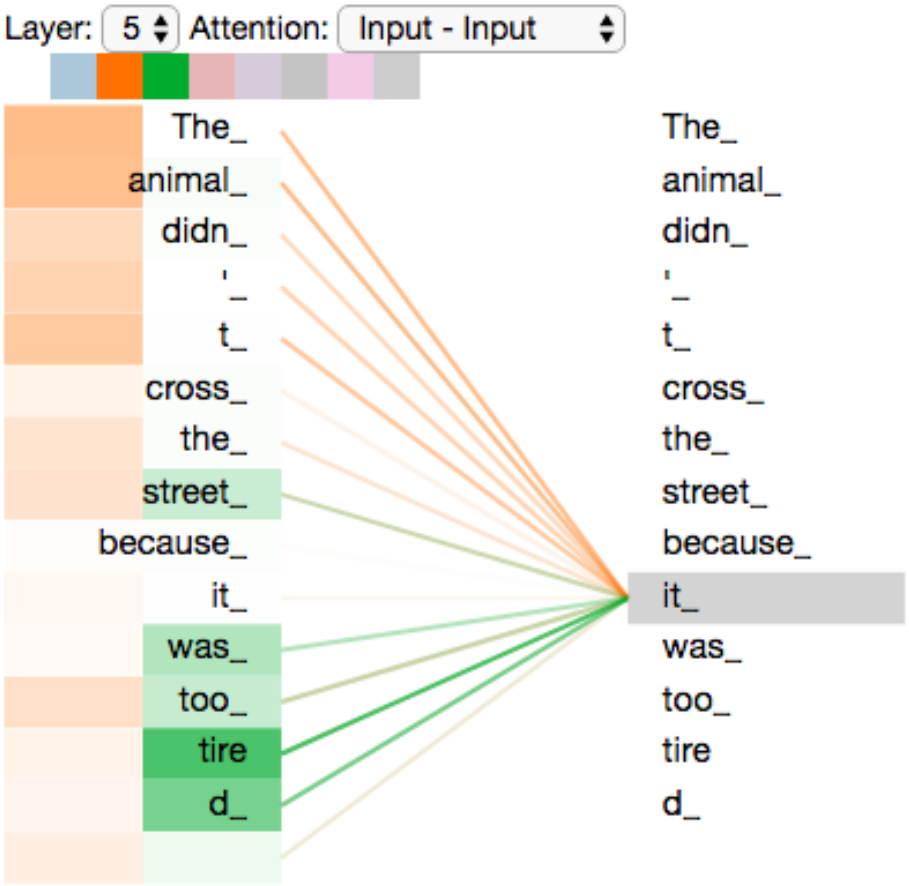
https://jalammar.github.io/illustrated-transformer/

Transformer

Self-Attention

2017

input-input



https://jalammar.github.io/illustrated-transformer/

Transformers > RNNs (LSTMs)

Attention is all you need!

2018

ELMo

2017

Transformer

Attention is all you need
No need for RNNs

2014

Attention

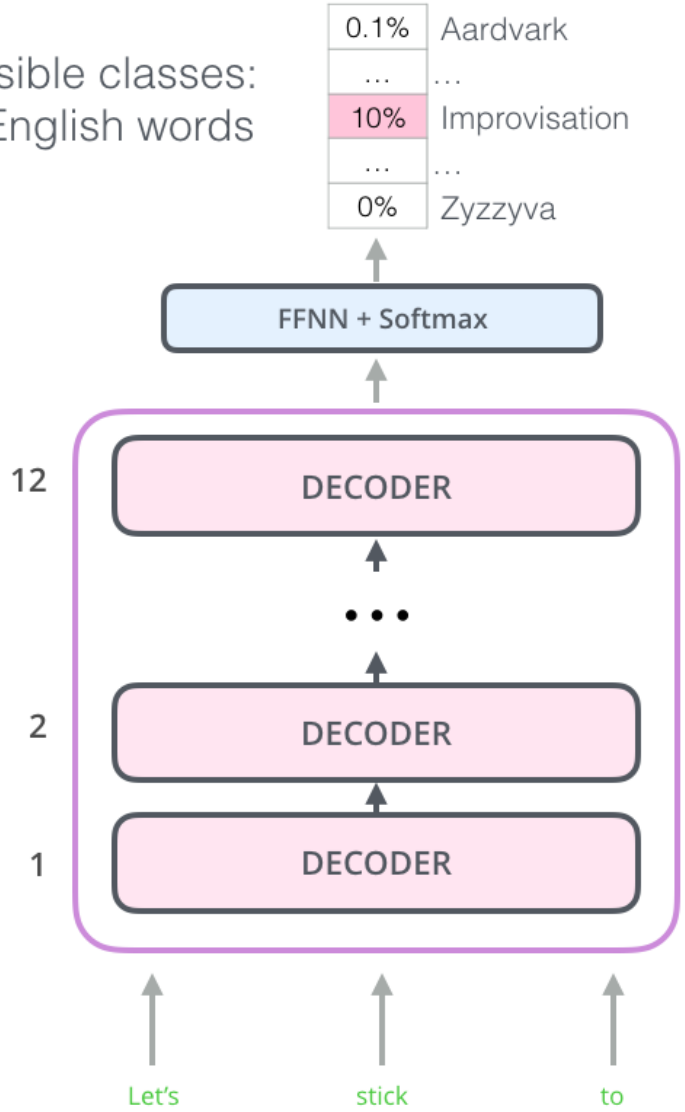
2013

Word2Vec

OpenAI Transformer

2018

Unsupervised
LM Training



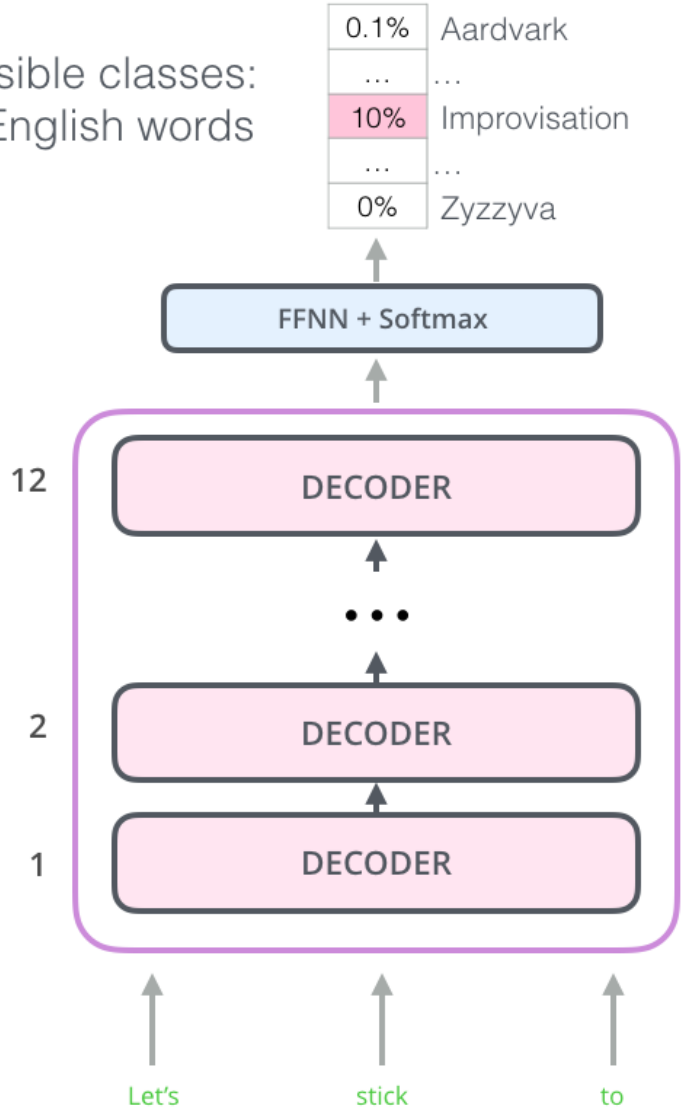
Radford, Alec, et al. "Improving language understanding by generative pre-training." 2018.

http://jalammar.github.io/illustrated-bert/

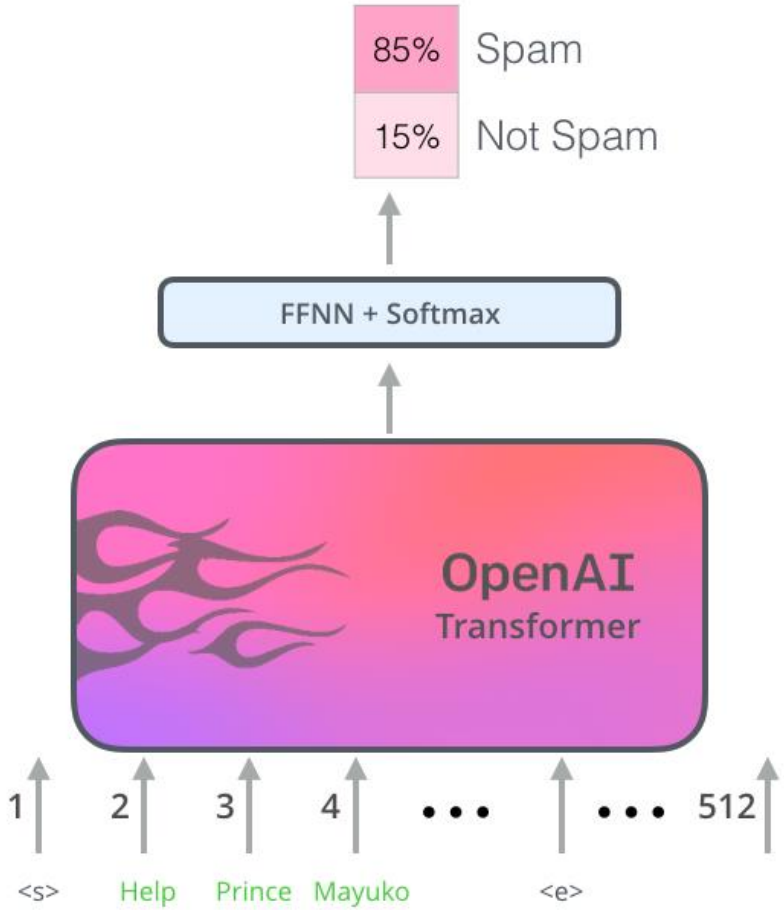
OpenAI Transformer

2018

Unsupervised
LM Training

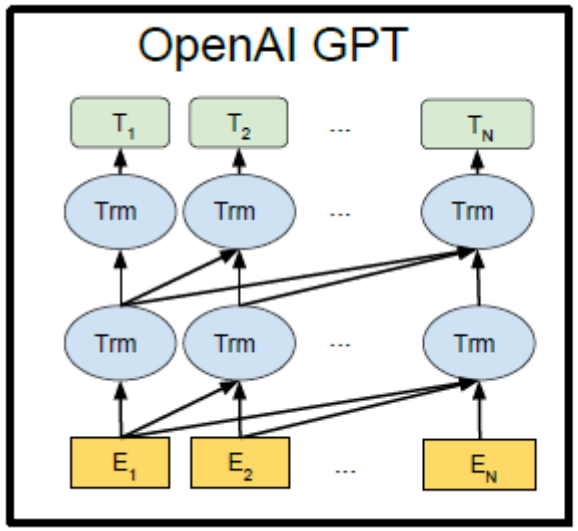


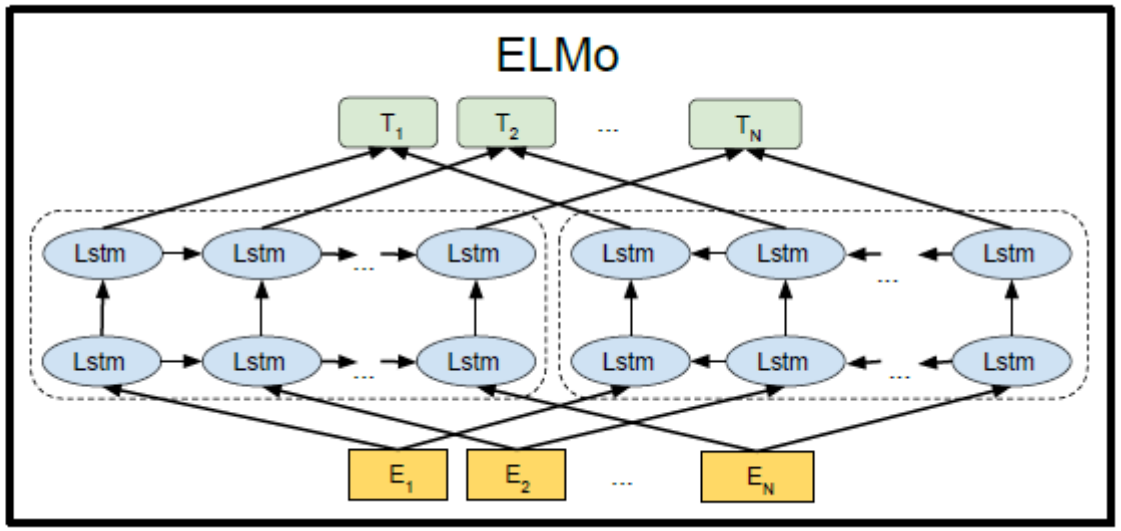
Supervised
Fine-Tuning



Radford, Alec, et al. "Improving language understanding by generative pre-training." 2018.

http://jalammar.github.io/illustrated-bert/





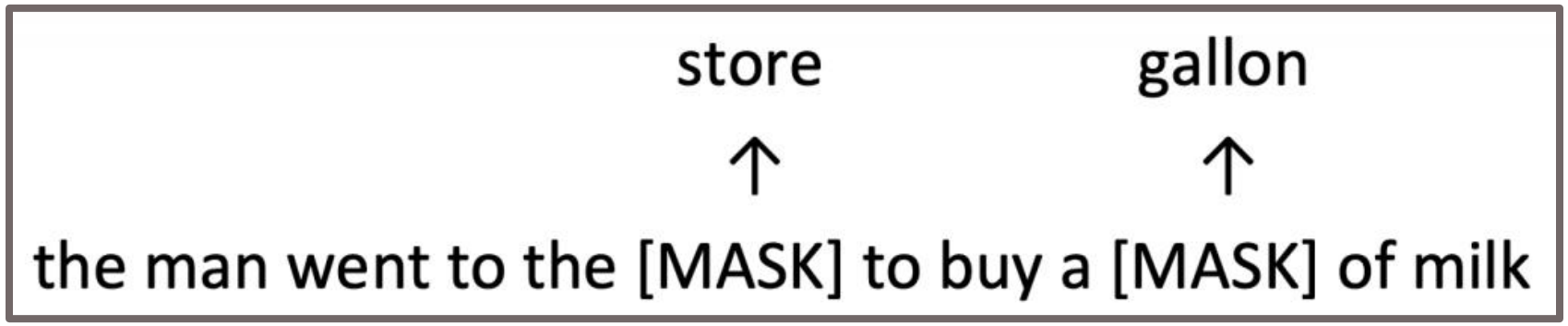
Unidirectional Transformer (Transformer Decoder)

Shallow BiDirectional LSTM

Training a LM using both left and right context jointly

Training a LM using both left and right context jointly

Masked Language Model + Transformer Encoders



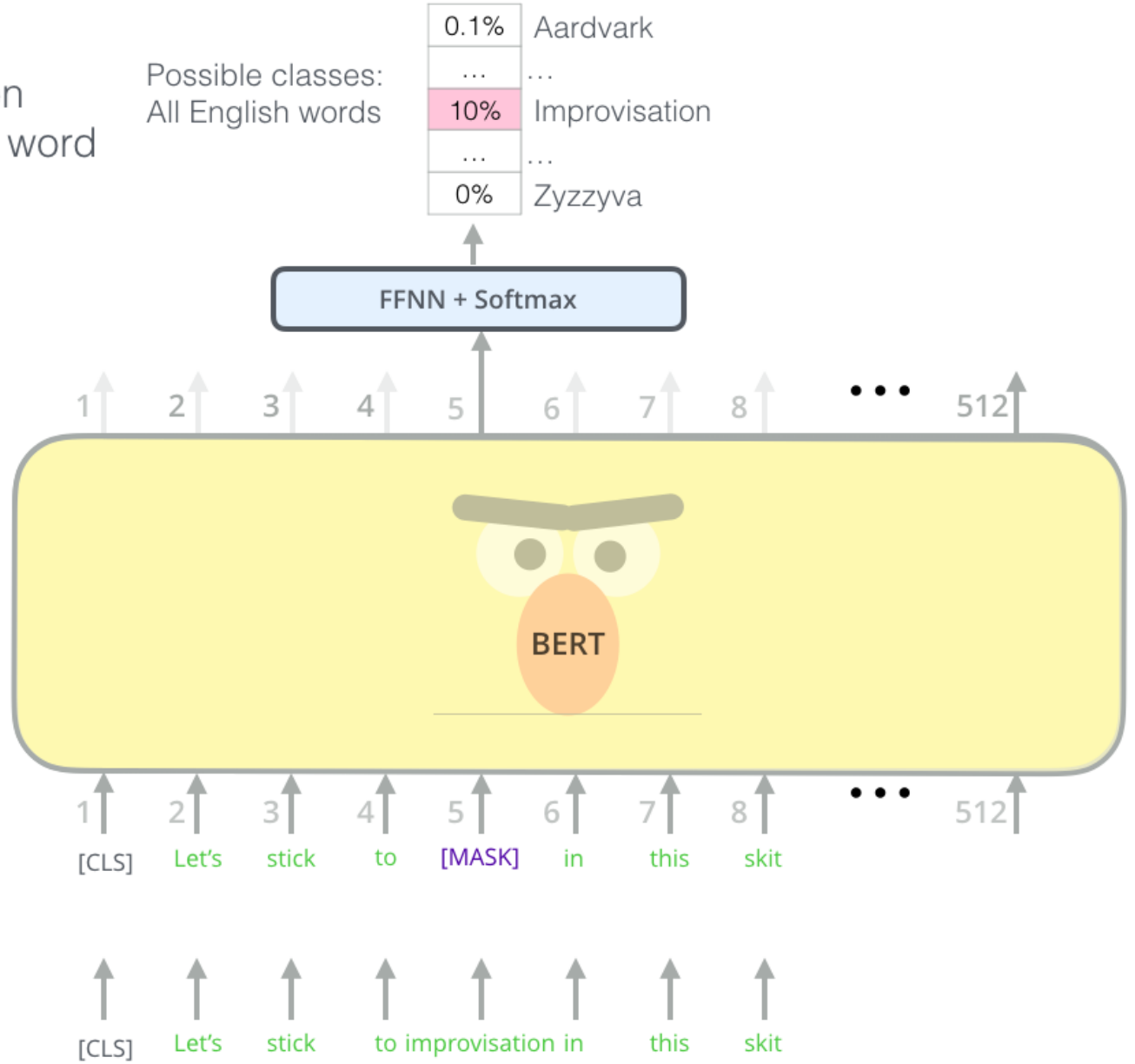
https://blog.feedly.com/nlp-breakfast-2-the-rise-of-language-models/

BERT

2018

Use the output of the masked word's position to predict the masked word

Possible classes: All English words



Bidirectional
Encoder
Representations from
Transformers

Randomly mask 15% of tokens

Input

http://jalammar.github.io/illustrated-bert/

BERT

2018

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:

Dataset:

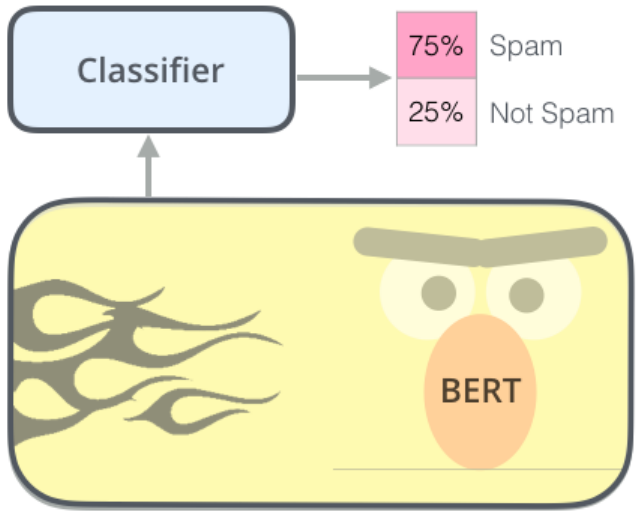
Objective:

Predict the masked word (langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step

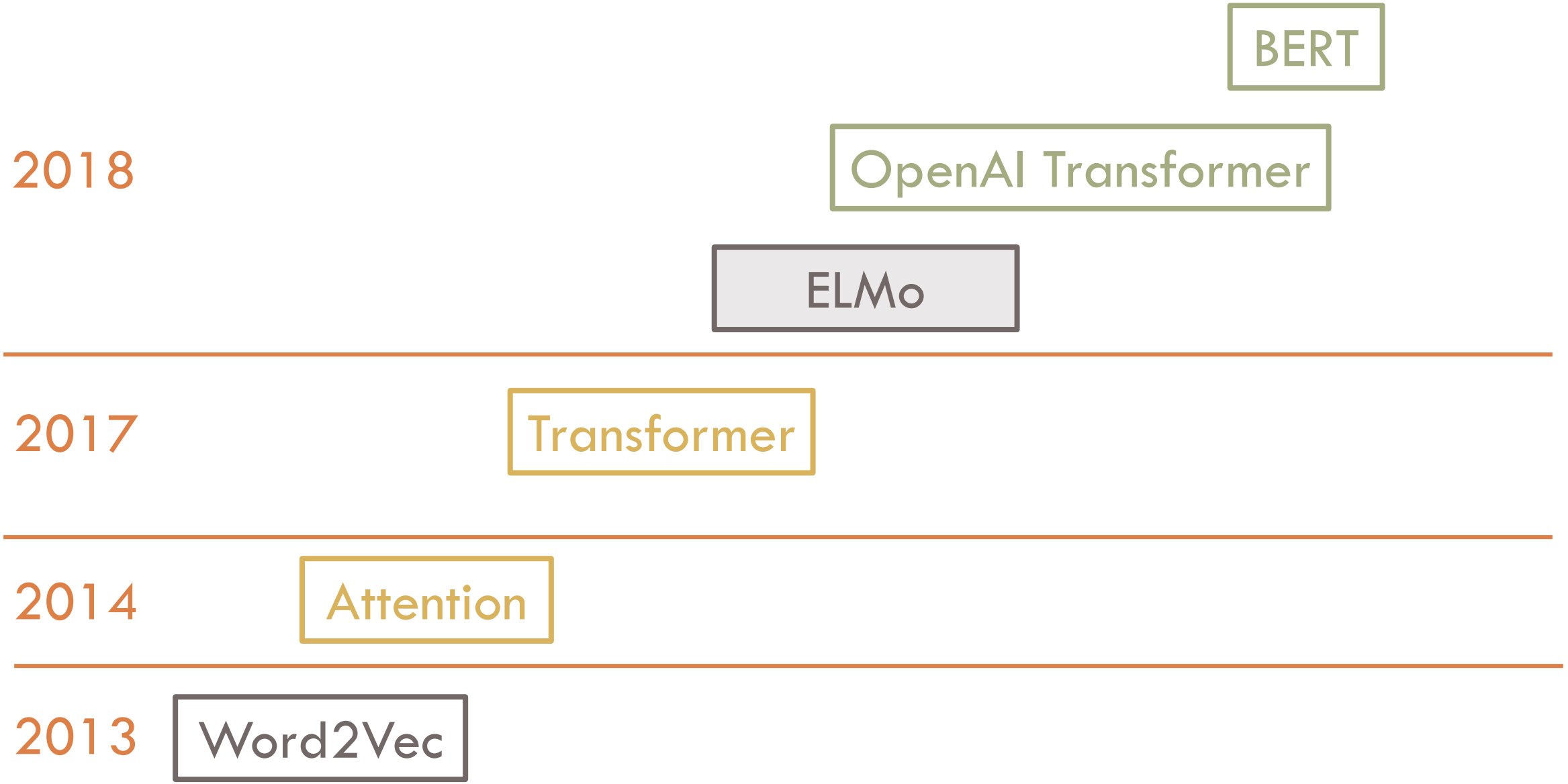
Model: (pre-trained in step #1)



Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

http://jalammar.github.io/illustrated-bert/


BERT
2018
OpenAI Transformer
ELMo
2017
Transformer
2014
Attention
2013
Word2Vec


123
SESAME STREET

Any Questions?