

Natural Language Processing

GTC 2020

Rachel Hu

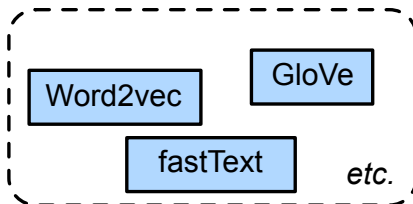
Applied Scientist

Amazon AI

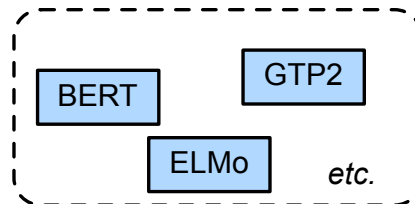


NLP Roadmap

Input
Representations:

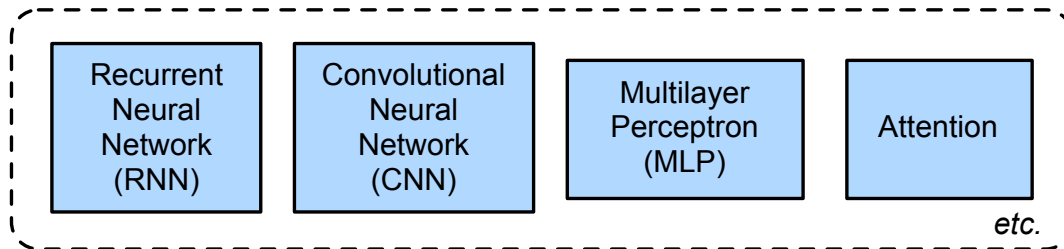


Shallow Representation

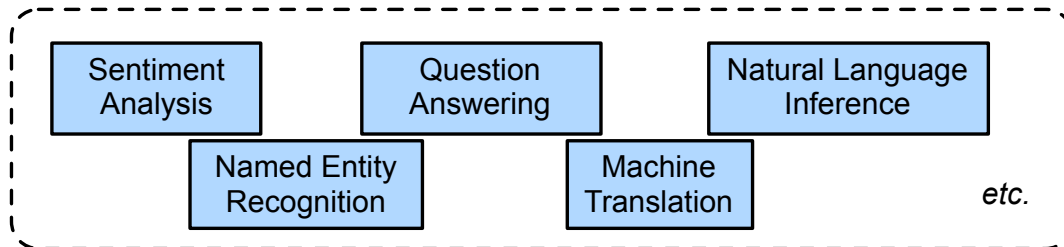


Deep Representation

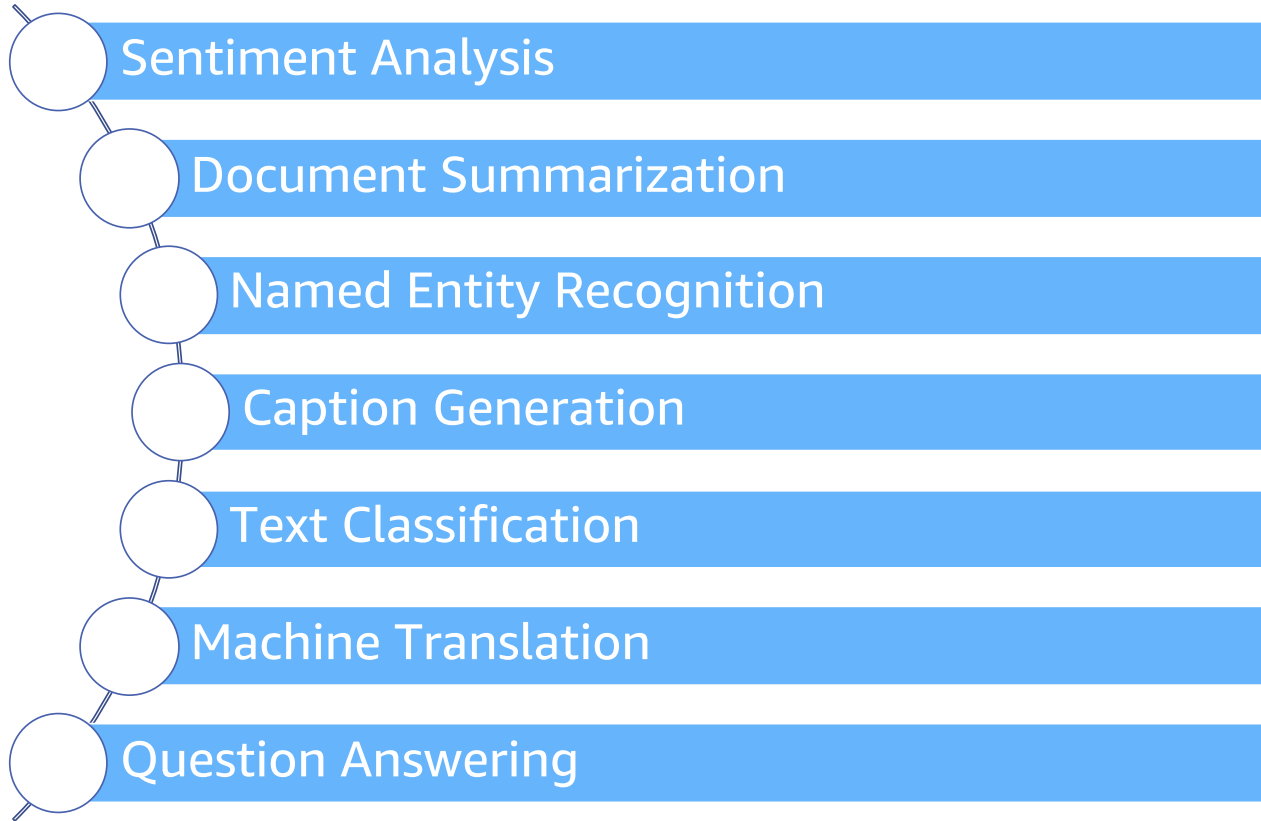
Architectures:



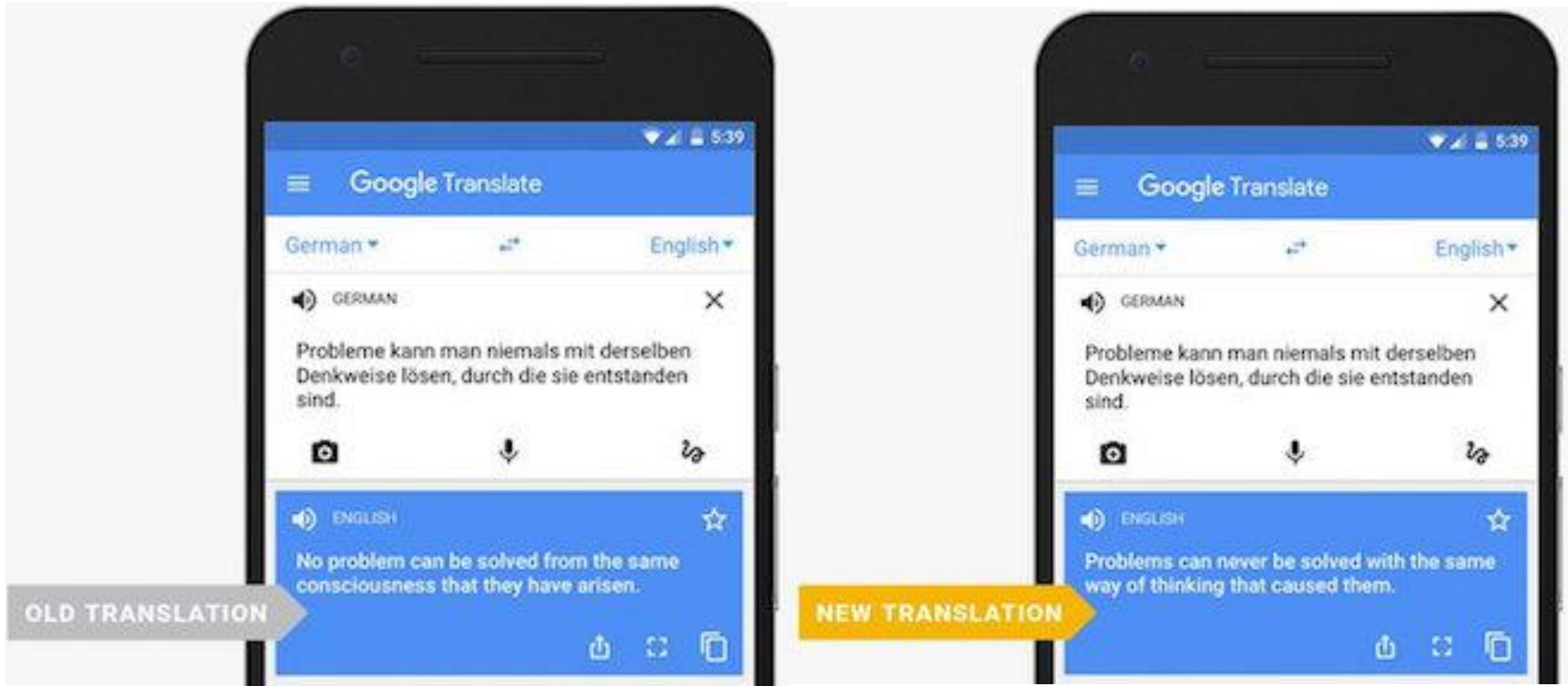
Downstream
Applications:



NLP Applications



Machine Translation



<https://www.pcmag.com/news/349610/google-expands-neural-networks-for-language-translation>

Text Synthesis

Content: Two dogs play by a tree.

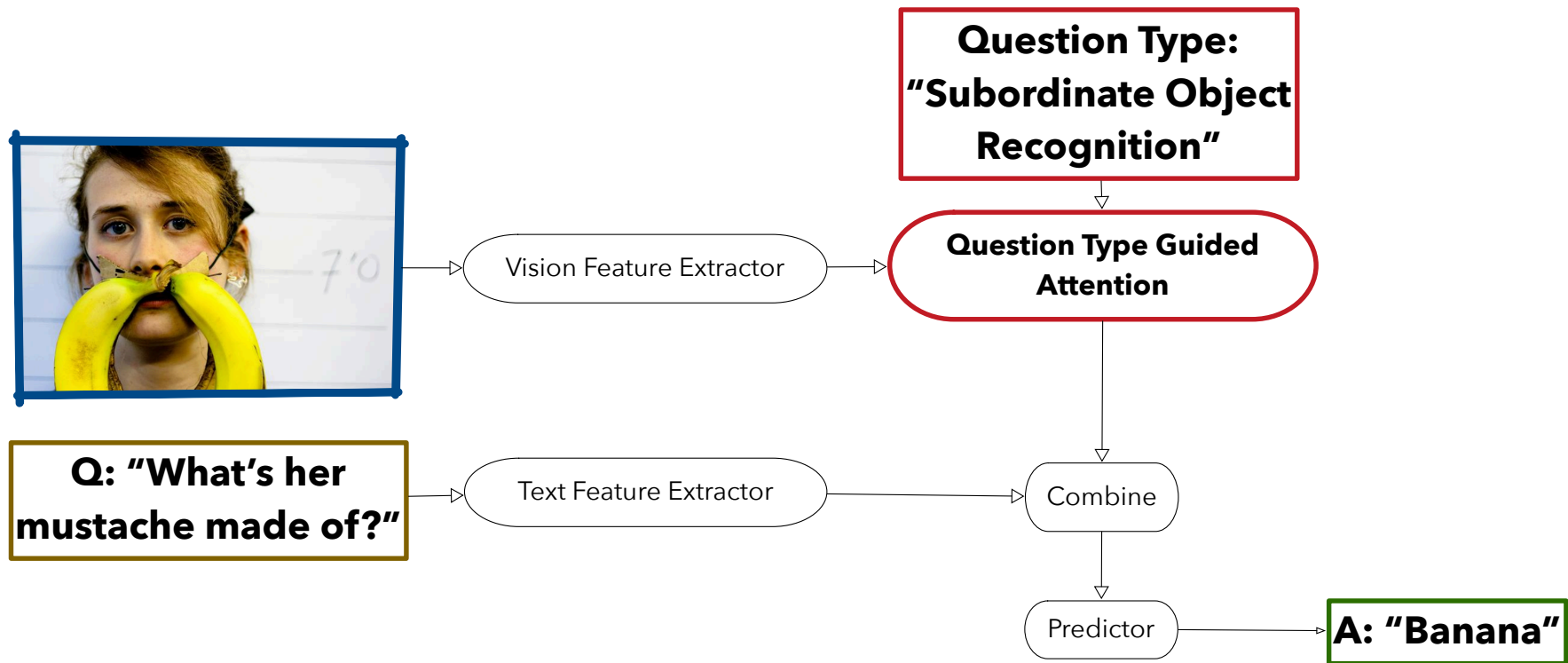
Style: happily, love



RNN

Two dogs in love play happily by a tree.

Question answering



Human captions from the training set

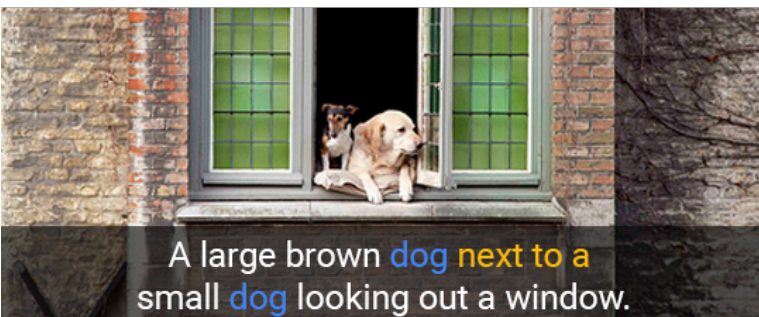
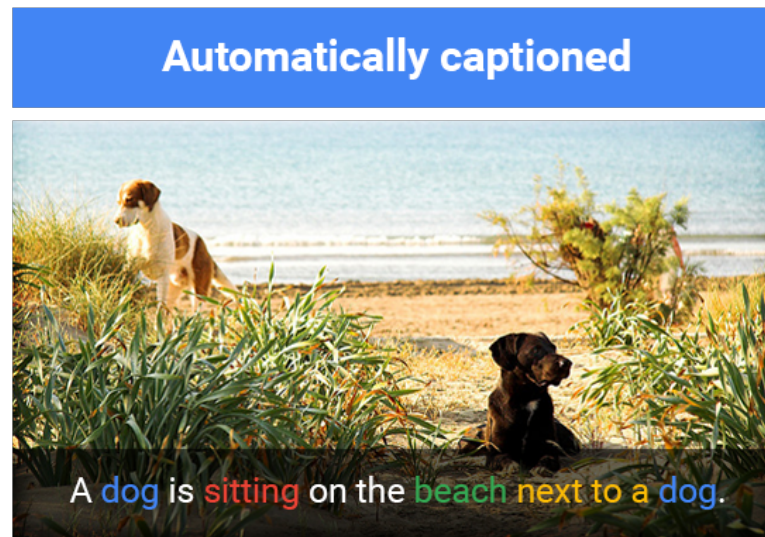


Image captioning

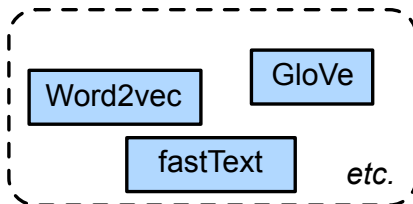


Shallue et al, 2016

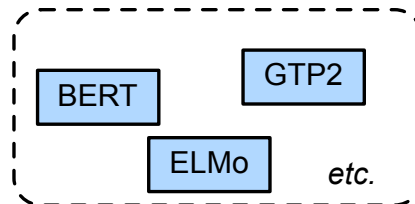
<https://ai.googleblog.com/2016/09/show-and-tell-image-captioning-open.html>

NLP Roadmap

Input
Representations:

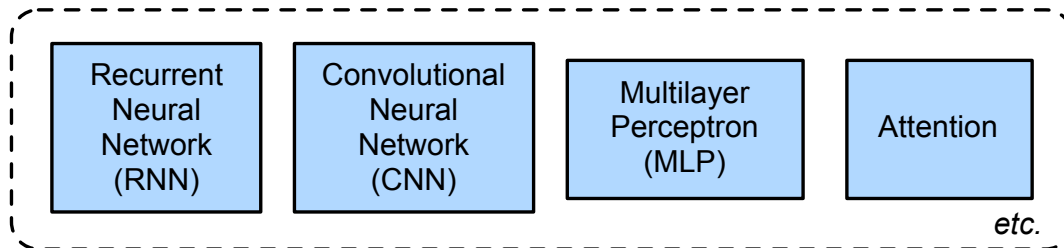


Shallow Representation

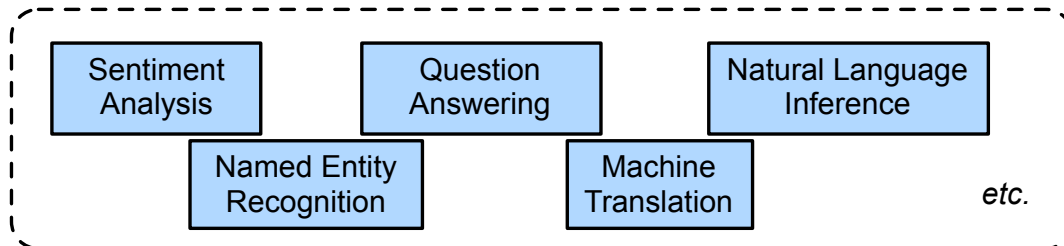


Deep Representation

Architectures:

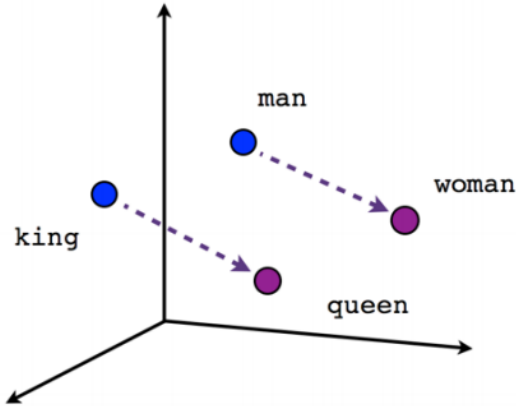


Downstream
Applications:

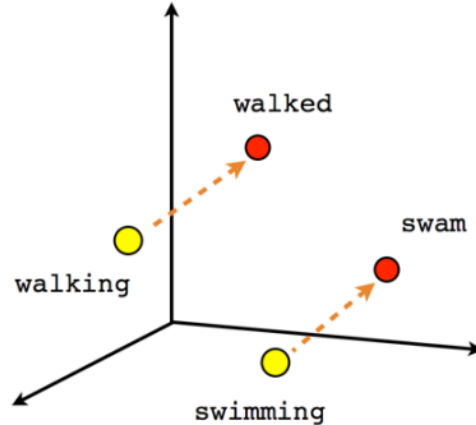


Shallow Word Representation

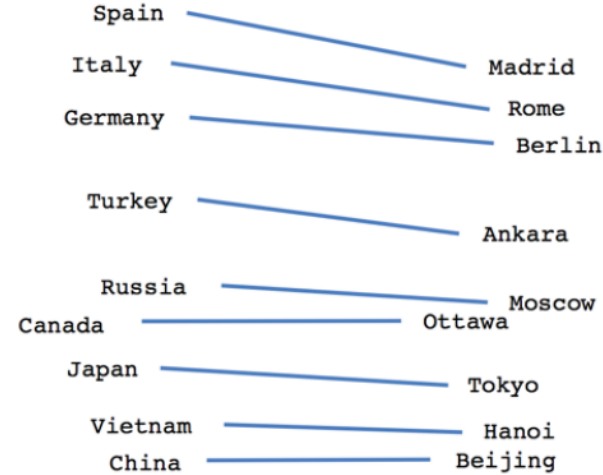
- Represent words by vectors



Male-Female



Verb tense



Country-Capital

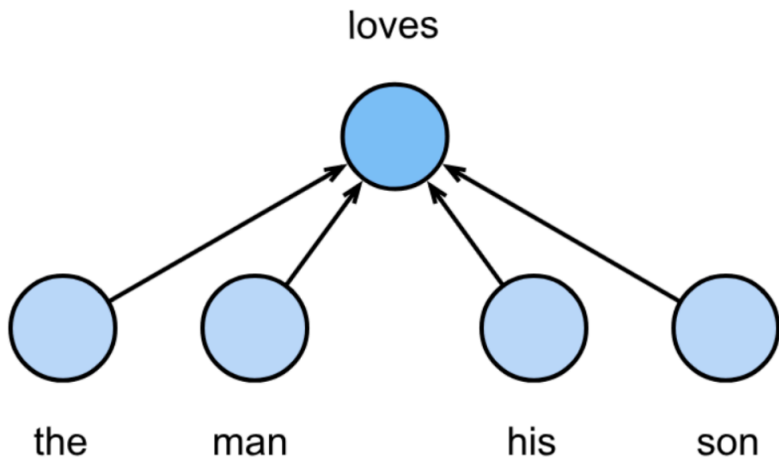
<https://nlp.stanford.edu/projects/glove/>

Continuous Bag Of Words (CBOW)

- **CBOW**

Center word is based on the context

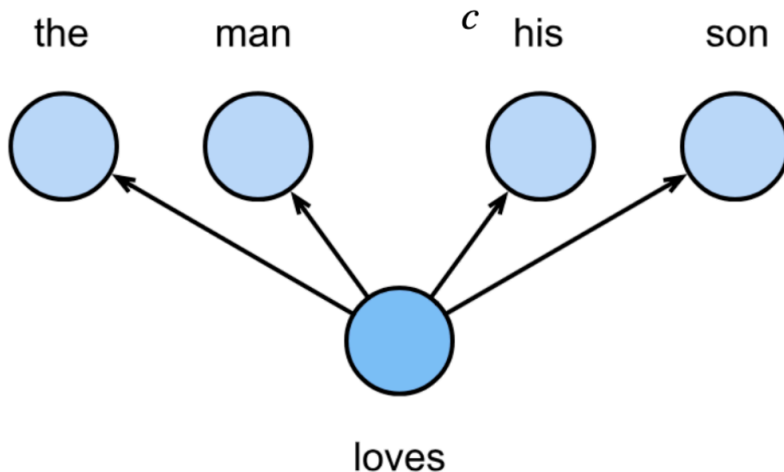
$$p(w | \text{context}) = p(w | \{w_c\})$$



- **Skip-gram**

Context is based on Center word

$$p(\text{context} | w) = \prod p(w_c | w)$$



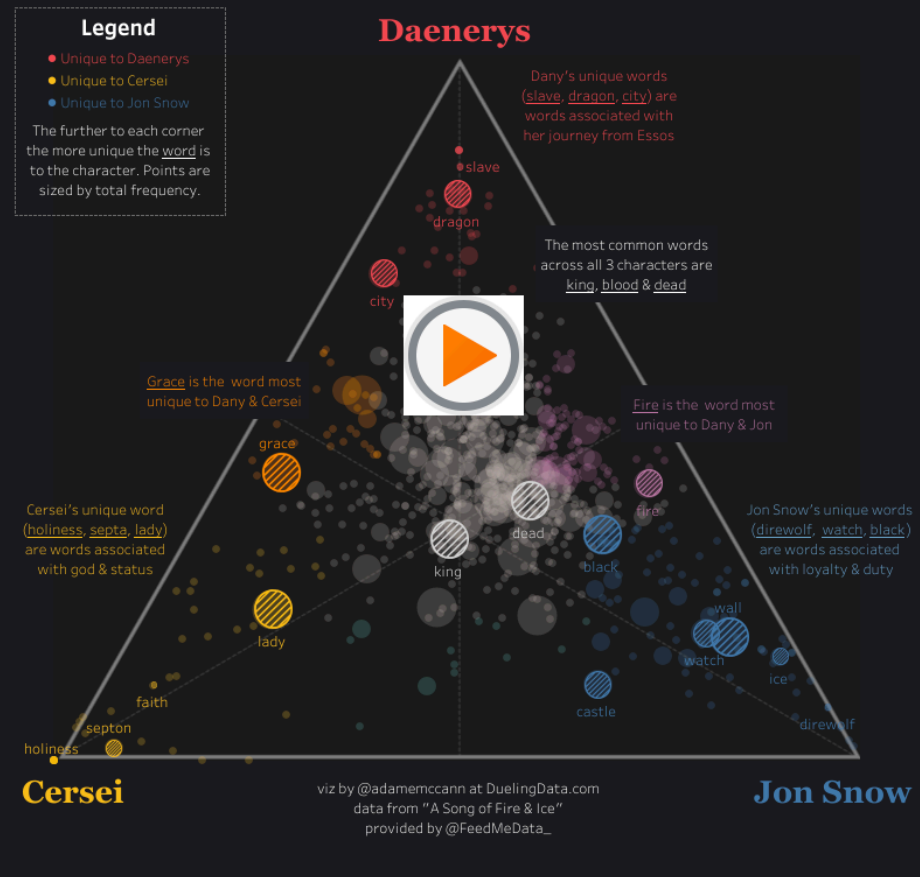
Shallow Word Representation

Eg.

- Word2vec
- GloVE
- fastText
- etc.

GAME OF THRONES™ IN WORDS

This viz shows the most unique words by character for each chapter in the 5 Game of Thrones books



Problem of Shallow Word Representation

M Mongabay.com

Upset about **Amazon fires** last year? Focus on deforestation this year (commentary)

That is, the recent deforestation surge fueled the 2019 Brazilian Amazon fires. The fires were in fact a lagging indicator of recent deforestation. Such information ...

1 week ago



BI Business Insider

How to restart your **Amazon Fire Stick** in 3 different ways - Business Insider

The Amazon Fire Stick is one of the most popular media streaming devices available right now. Not only can you shop on Amazon using a Fire Stick, but you can ...

10 hours ago



Problem of Shallow Word Representation

M Mongabay.com

Upset about Amazon fires last year? Focus on deforestation this year (commentary)

That is, the recent deforestation surge fueled the 2019 Brazilian Amazon fires. The fires were in fact a lagging indicator of recent deforestation. Such information ...

1 week ago



BI Business Insider

How to restart your Amazon Fire Stick in 3 different ways - Business Insider

The Amazon Fire Stick is one of the most popular media streaming devices available right now. Not only can you shop on Amazon using a Fire Stick, but you can ...

10 hours ago



- **Cons** for shallow word representation:
 - Word ambiguity: Polysemy
 - Ignores sequential information(Word2vec)
 - Language model only looks in a single direction
 - Often don't update the pre-trained models

Deep Word Representation

BERT

Bidirectional Encoding
Representation
from Transformer



BERT Architecture

BERT Base

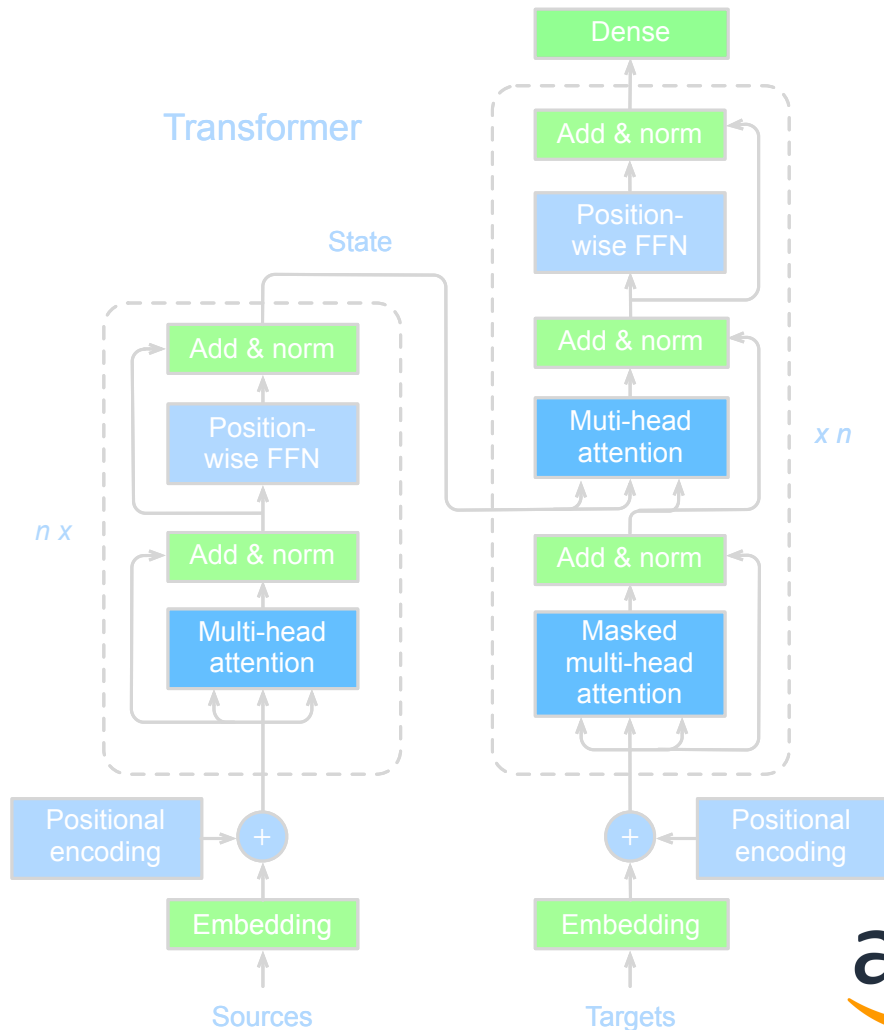
blocks = 12

parameters = 110M

BERT Large

blocks = 24

parameter = 340M



BERT Family

