Intro to Language Processing

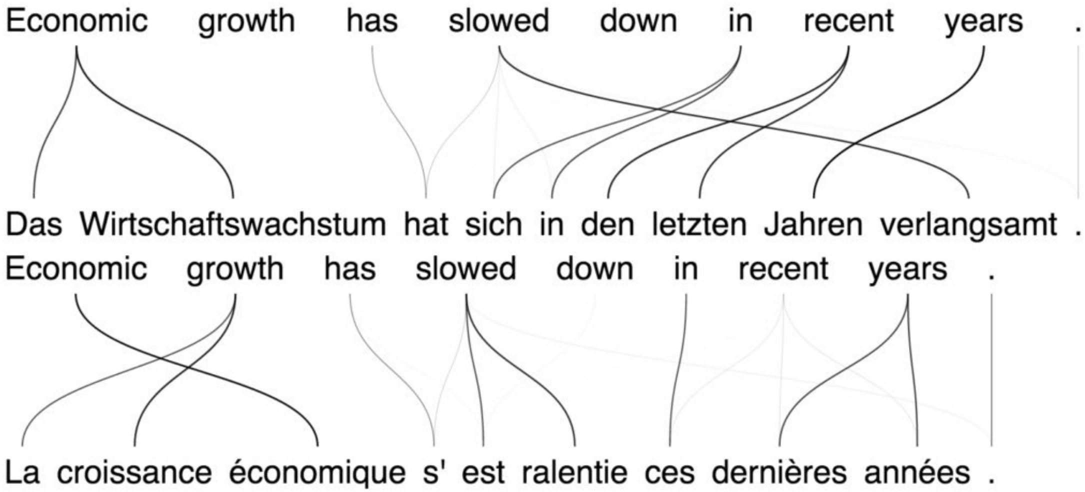
ML@B Workshop
9 October 2018

Clone this:

https://tinyurl.com/ydbk6mc7

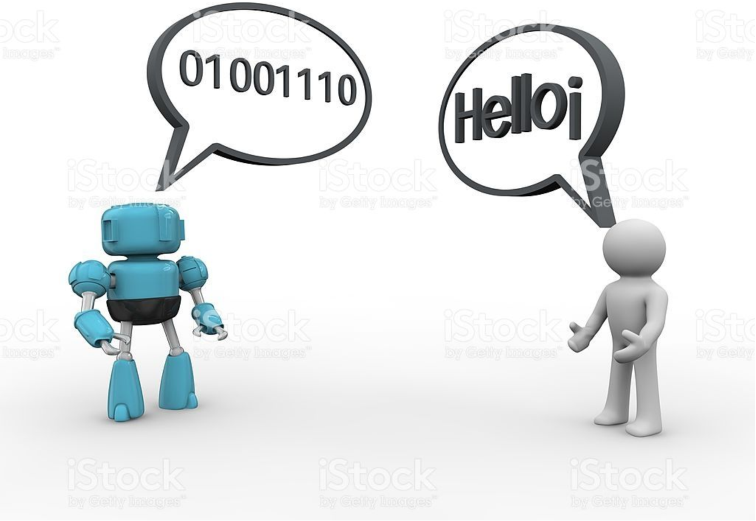
All Around Us

Google Translate



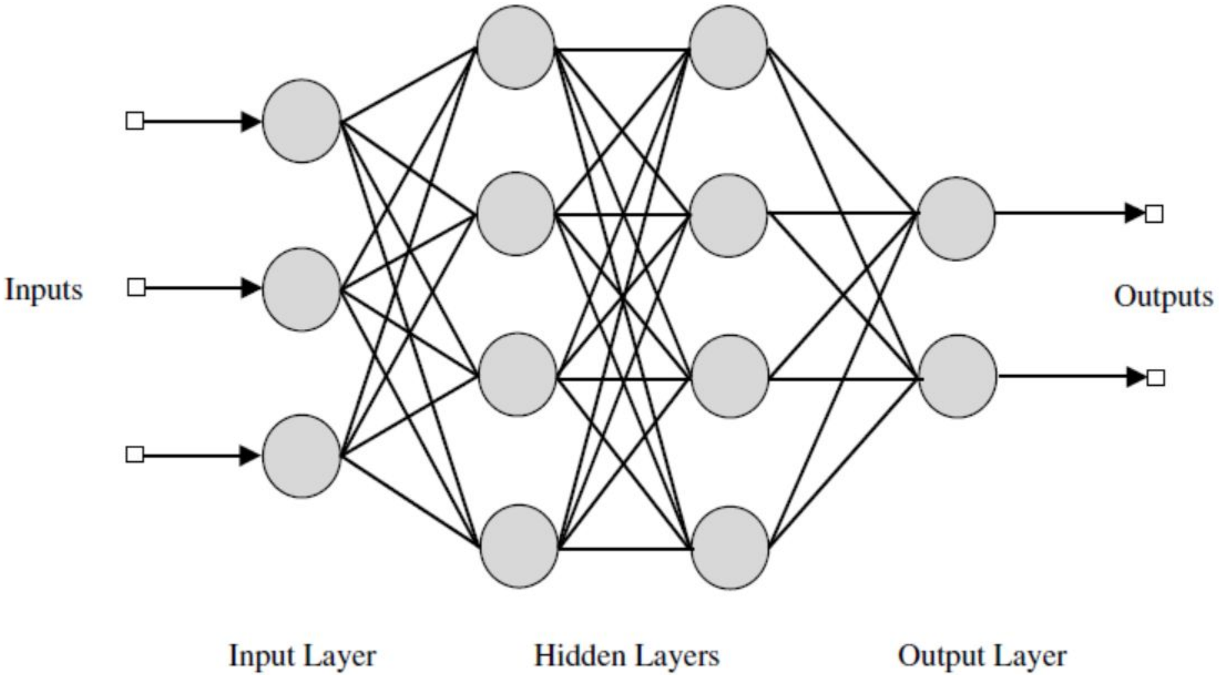
Importance

- Not everyone can code
- Allows broader audience to interact and communicate with machines



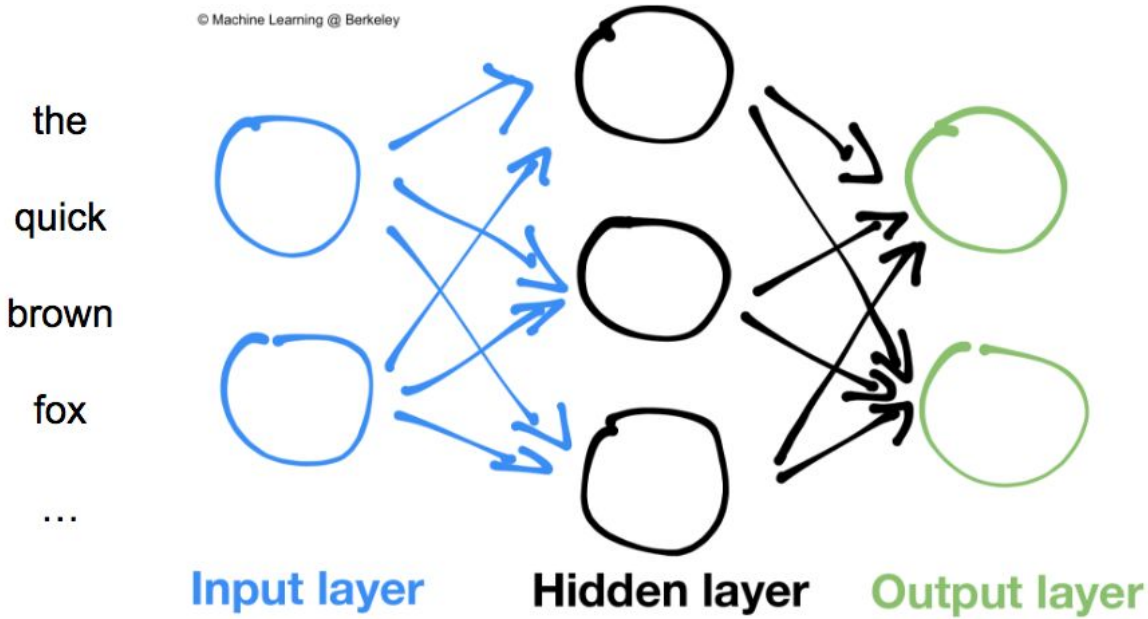
Architectures

Basic Neural Network

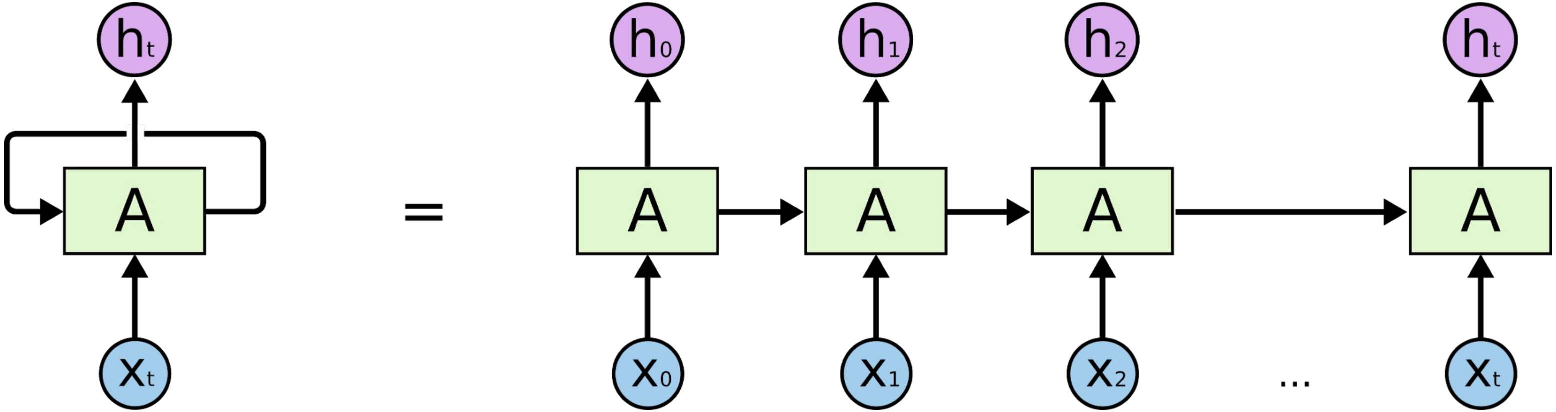


Time

Not able to represent sequential data (No time variable!)



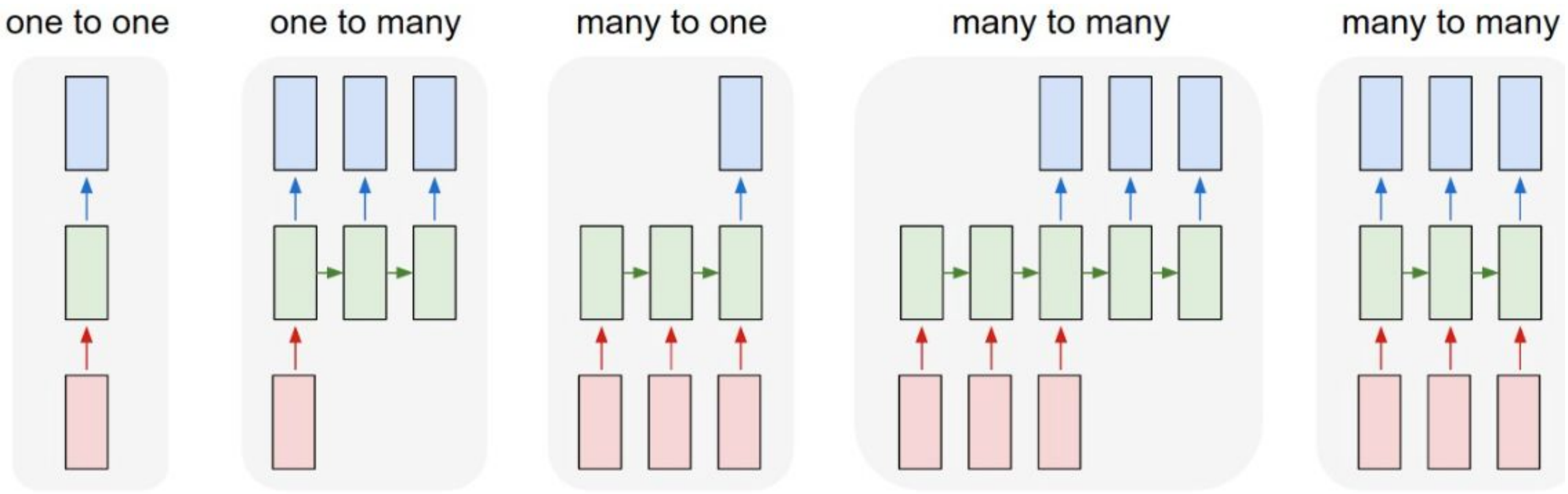
Recurrent Neural Networks (RNNs)



Math Behind the RNN

$$
\begin{aligned}
\boldsymbol{a}^{(t)} &= \boldsymbol{b} + \boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{U}\boldsymbol{x}^{(t)} \\
\boldsymbol{h}^{(t)} &= \tanh(\boldsymbol{a}^{(t)}) \\
\boldsymbol{o}^{(t)} &= \boldsymbol{c} + \boldsymbol{V}\boldsymbol{h}^{(t)} \\
\hat{\boldsymbol{y}}^{(t)} &= \mathrm{softmax}(\boldsymbol{o}^{(t)})
\end{aligned}
$$

Types of RNN Layers

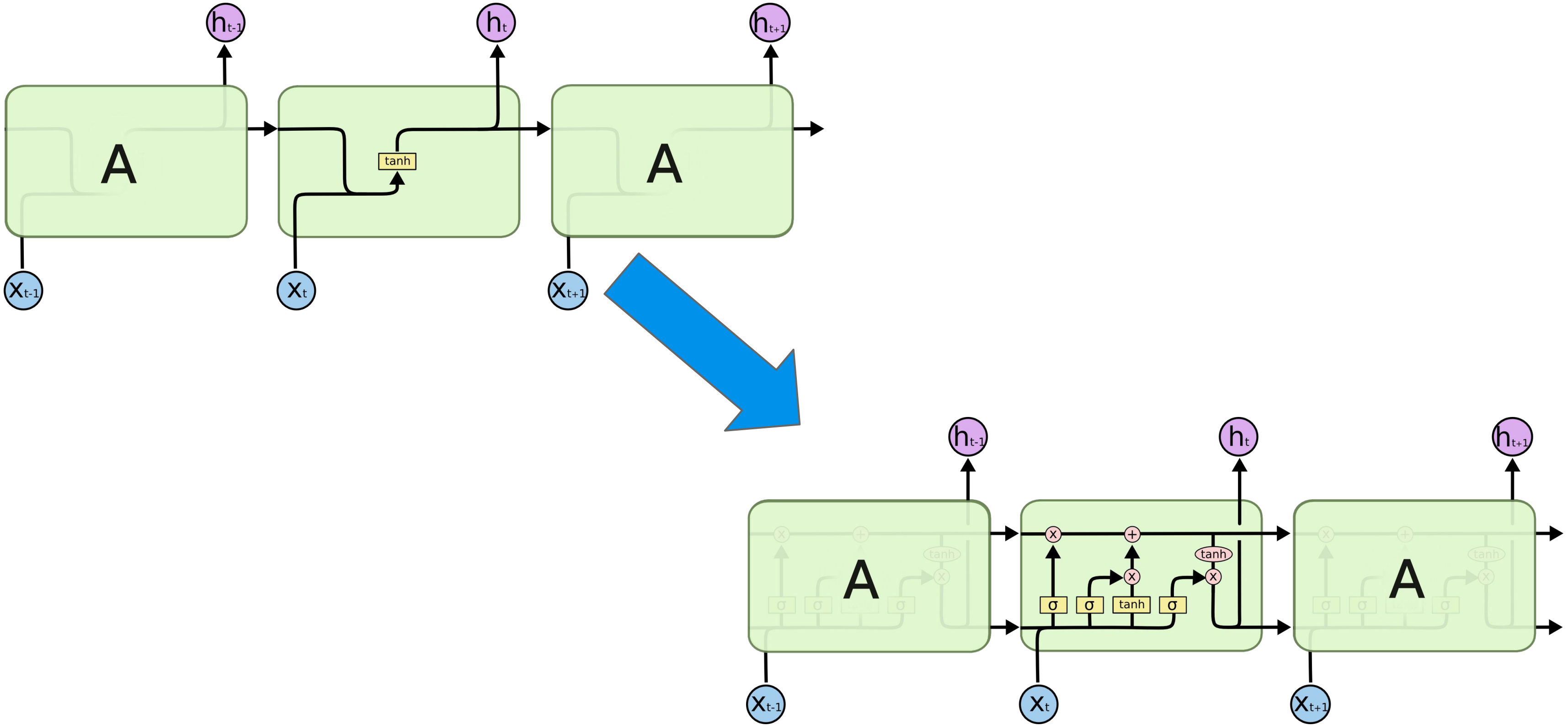


Standard Network | Image Captioning | Sentiment Analysis | Machine Translation | Language Modeling

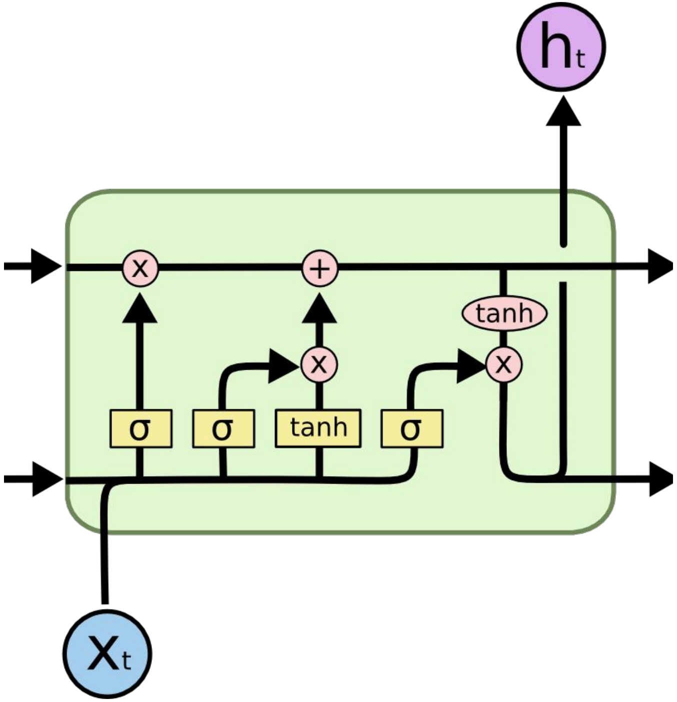
But What’s Even Better?

- Math Reason: RNN’s suffer greatly from vanishing gradient problem
    - Backpropagation doesn’t work very well due to multiplying extremely small numbers by each other over and over again
- Intuitive Reason: RNN’s always pass on the information they have
    - No way to forget old information that is no longer necessary

Long Short-Term Memory (LSTM)



LSTM in Detail



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

$$o_t = \sigma\left(W_o \; [h_{t-1}, x_t] \; + \; b_o\right)$$

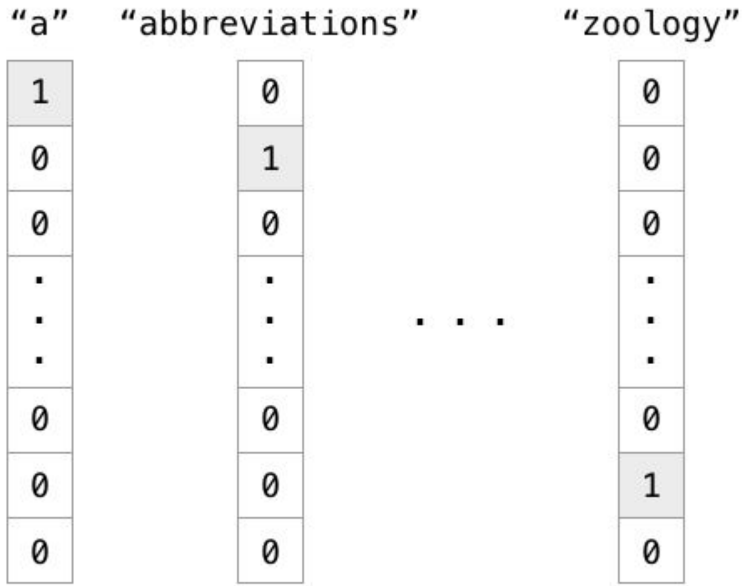
$$h_t = o_t * \tanh\left(C_t\right)$$

Word Embeddings

How could we represent words to computers?

How could we represent words to computers?

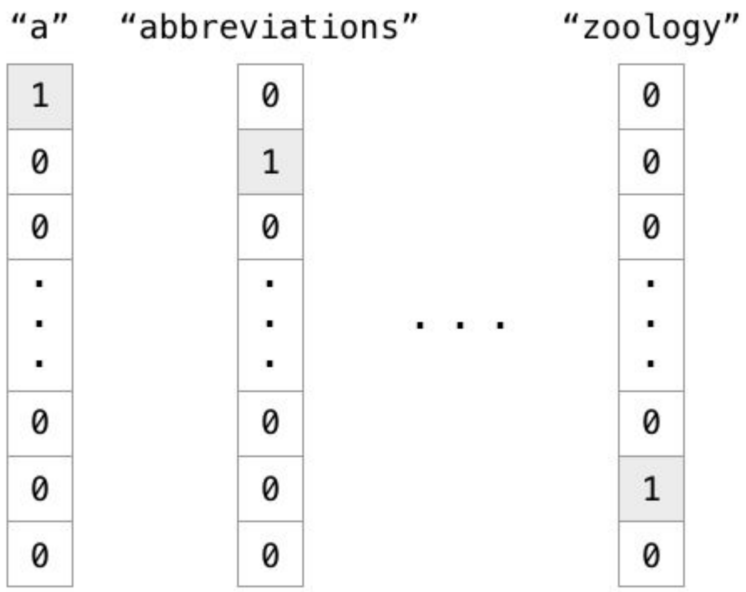
One-Hot Encoding?



How could we represent words to computers?

One-Hot Encoding?

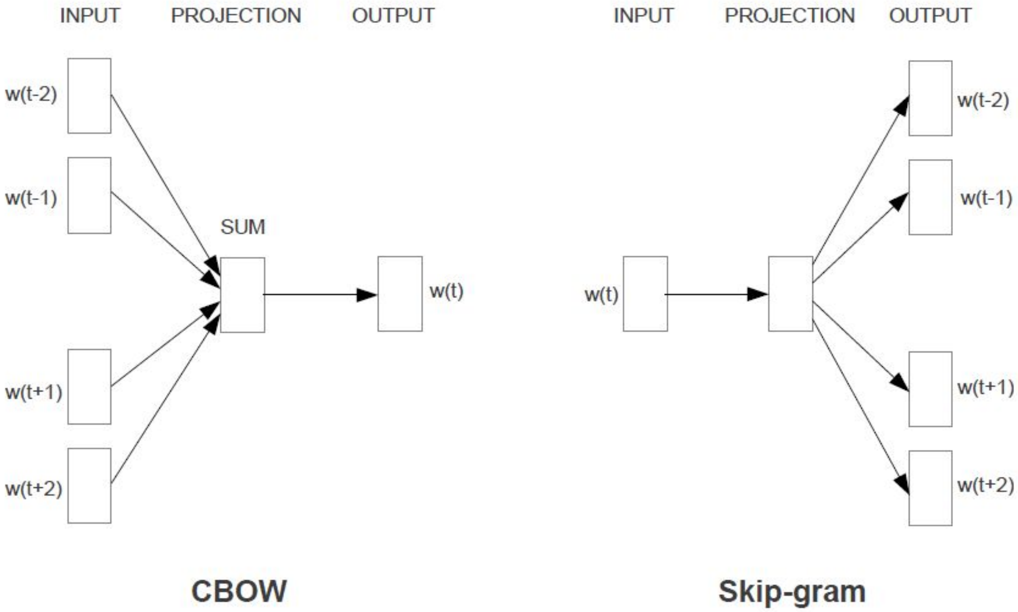
Every vector would need to have the length of our entire vocabulary size. Horribly inefficient



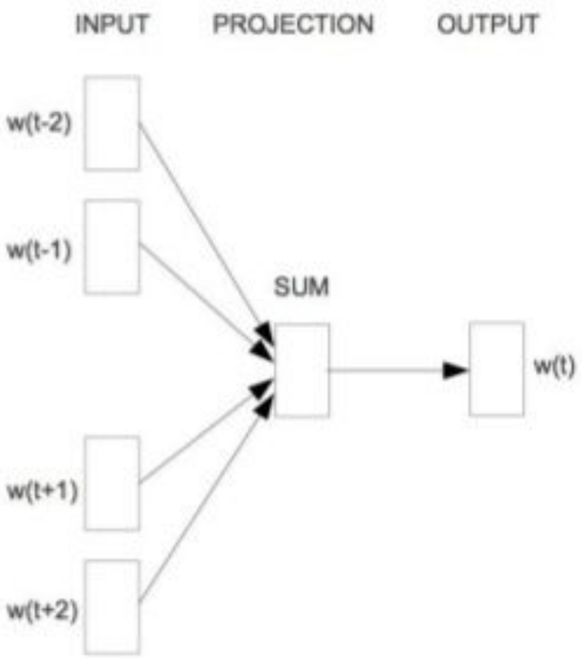
Word Embeddings!

- Every word is a vector, can be around size 200-300 rather than entire vocabulary
- Where do the vectors come from?

Learning Word Embeddings: CBOW and Skip-Gram



CBOW (Continuous Bag-Of-Words) Example



`the cat climbed a tree`

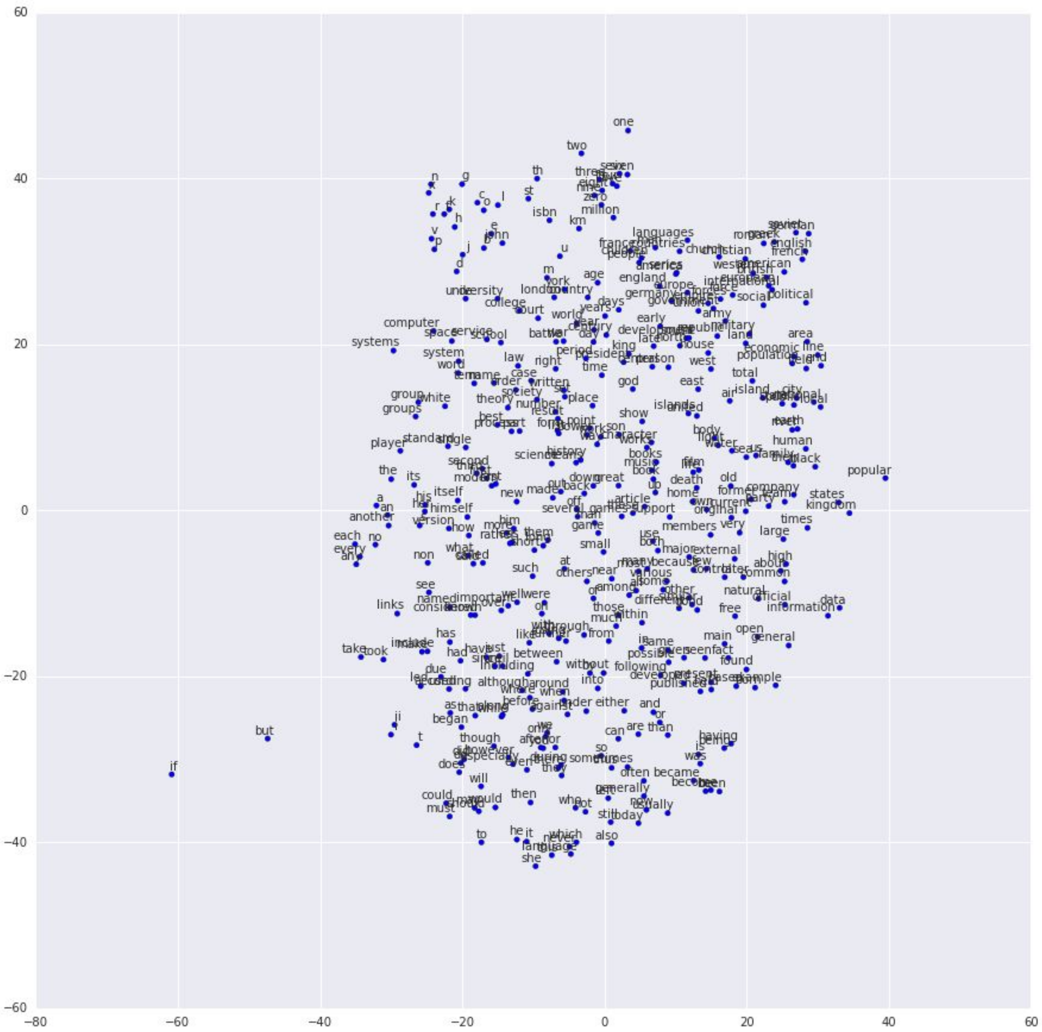
Given context:

`a, cat, the, tree`

Estimate prob. of

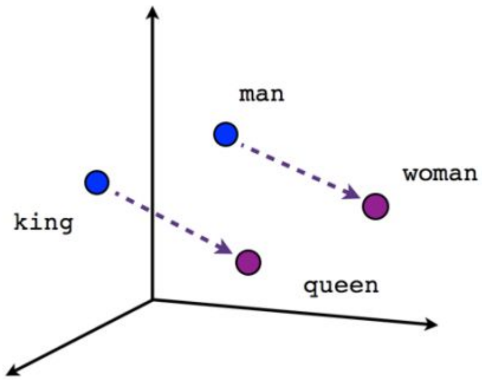
`climbed`

Word Embedding Demo

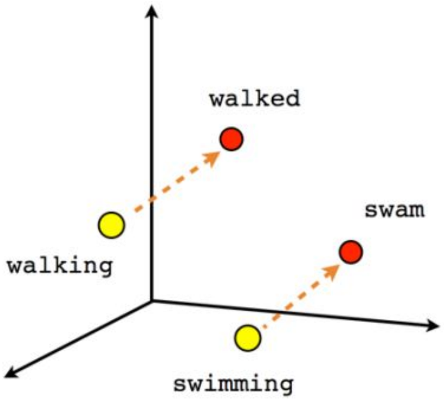


60
40
20
0
−20
−40
−60
−80
−60
−40
−20
0
20
40
60
one
two
th
st
isbn
km
million
zero
languages
countries
university
college
london
england
germany
europe
american
english
french
political
social
computer
systems
system
word
space
service
school
world
years
days
army
military
area
total
economic
population
law
right
time
case
president
king
early
house
west
east
god
theory
number
place
society
group
groups
white
islands
show
body
water
human
popular
player
standard
single
history
science
books
music
book
film
death
old
company
states
kingdom
the
its
a
an
another
each
every
any
no
his
himself
itself
version
new
made
back
great
game
small
members
very
large
times
high
about
common
data
information
natural
official
free
general
open
main
fact
found
example
non
see
links
such
others
near
among
those
within
from
between
having
being
take
took
has
had
began
ii
t
but
if
though
however
especially
does
will
could
would
must
should
then
who
not
now
usually
today
still
he
it
which
never
also
to
she
and
or
than
are
can
is
was
became
generally
often
became
been
before
against
either
into
following
possible
published

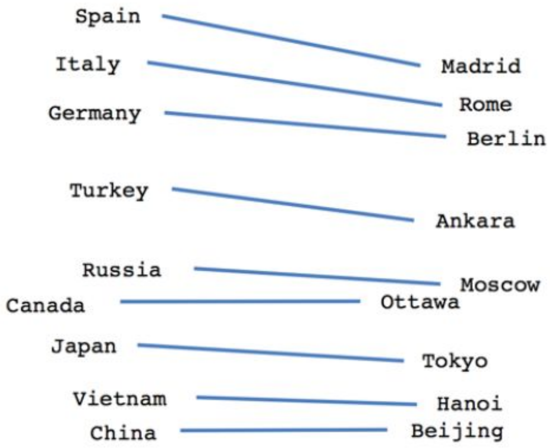
Cool Properties



Male-Female



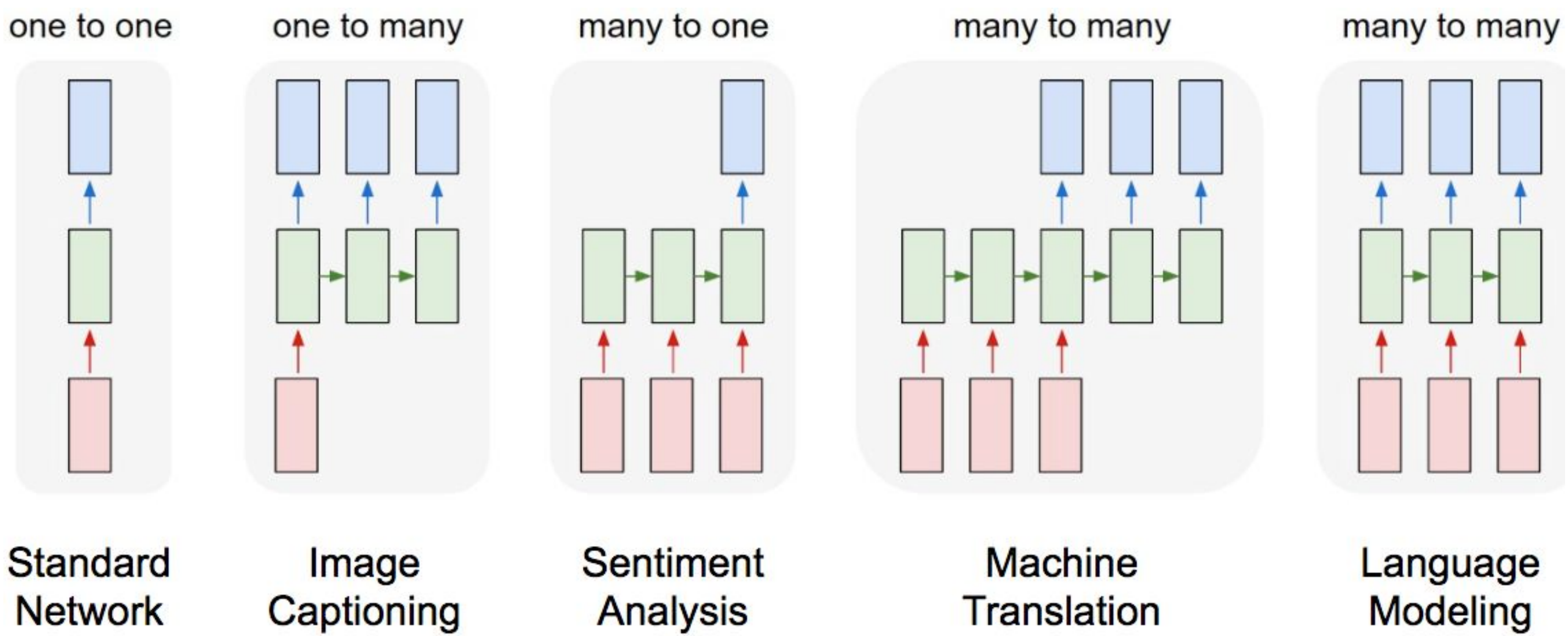
Verb tense



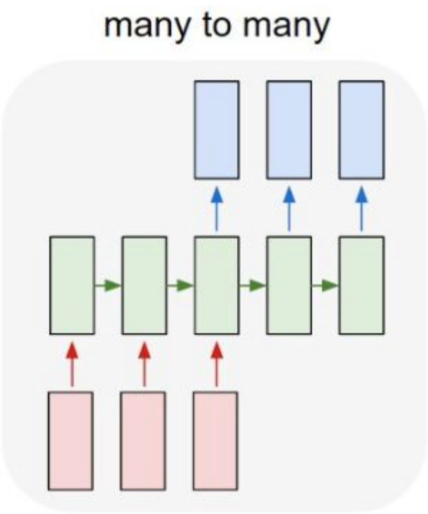
Country-Capital

Machine Translation

Translation is a sequence-to-sequence problem.



Translation is a sequence-to-sequence problem.



Machine
Translation

Variants of “Languages”

Languages can actually be quite general!

Variants of “Languages”: Summarization

“Argentina coach Alejandro Sabella believes Lionel Messi’s habit of throwing up during games is because of nerves. The Barcelona star has vomited on the pitch during several games over the last few seasons and appeared to once again during Argentina’s last warm-up match against Slovenia on Saturday.”

“Argentina coach Sabella believes Messi’s habit of being sick during games is down to nerves.”

CNN / Daily Mail Dataset.

Variants of “Languages”: Chatbots

“Hey Alexa, what is the weather like tomorrow?”

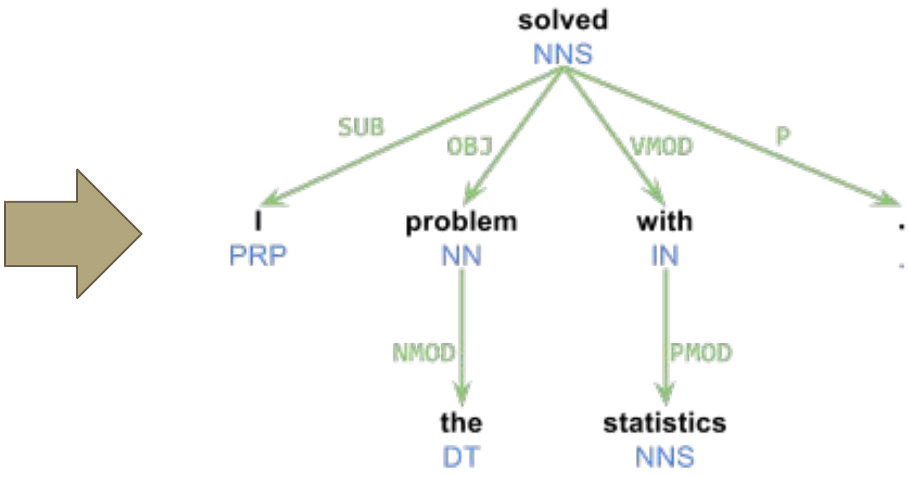
“It’s seventy-five degrees and cloudy tomorrow.”

Variants of “Languages”: Part-of-Speech Tagging

“Machine learning is cool!” → “ADJ NN VBZ JJ”

Variants of "Languages": Dependency Parsing

"I solved the problem with statistics."



Variants of "Languages": Dependency Parsing

[ROOT solved NNS
    [SUB I PRP]
    [OBJ problem NN
        [NMOD the DT]
    ]
    [VMOD with IN
        [PMOD statistics NNS]
    ]
    [P . .]
]

Variants of “Languages”: Dependency Parsing

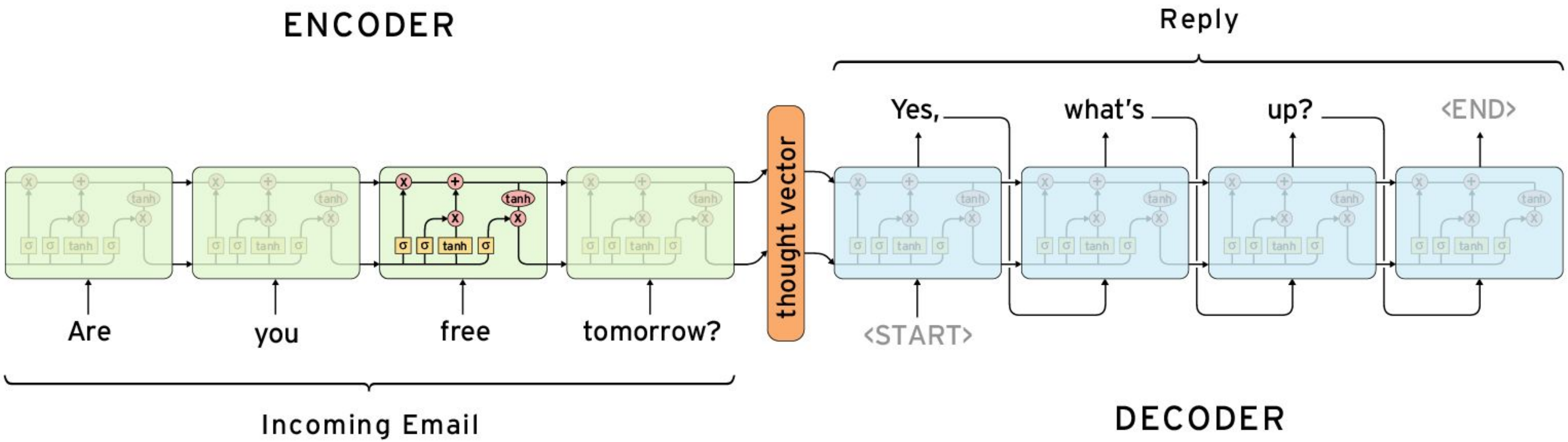
[ROOT solved NNS [SUB I PRP] [OBJ problem NN [NMOD the DT]] [VMOD with IN [PMOD statistics NNS]] [P . .]]

Variants of “Languages”: Dependency Parsing

“I solved the problem with statistics.”

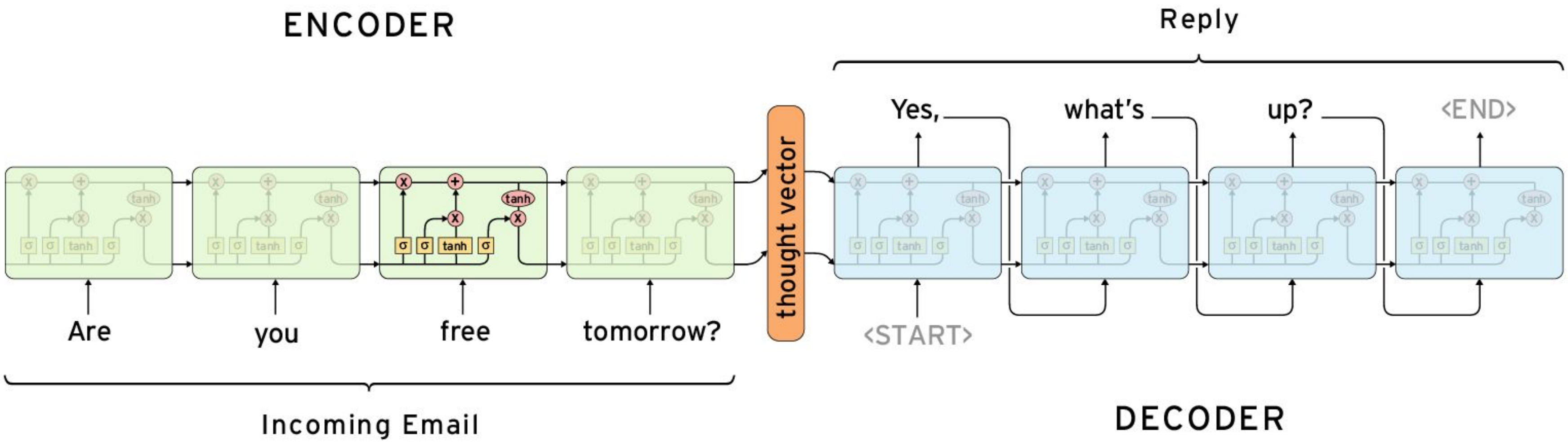
[ROOT solved NNS [SUB I PRP] [OBJ problem NN [NMOD the DT]] [VMOD with IN [PMOD statistics NNS]] [P . .]]

Encoder-Decoder Model

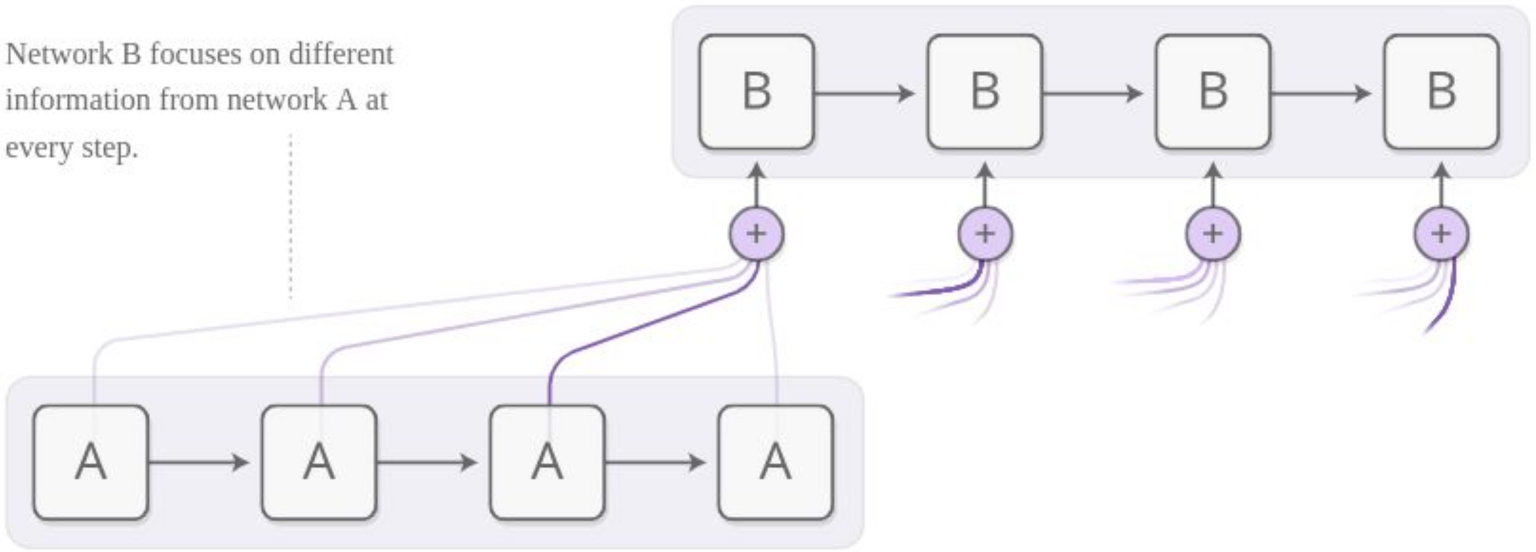


Attention

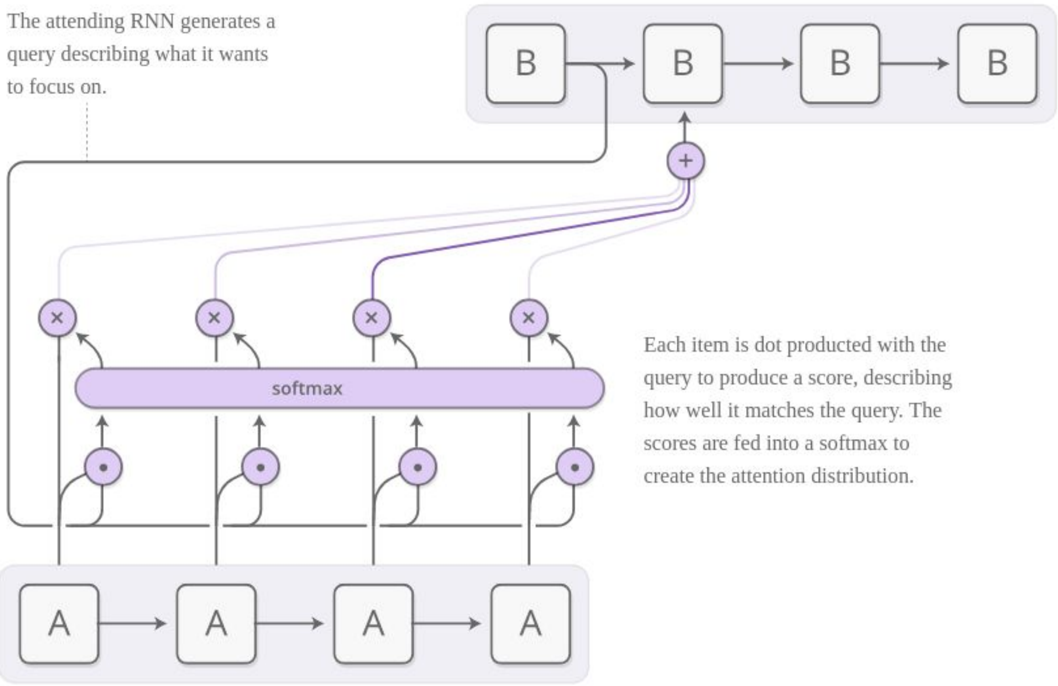
Attention: Motivation



Attention: Focus on the entire input.

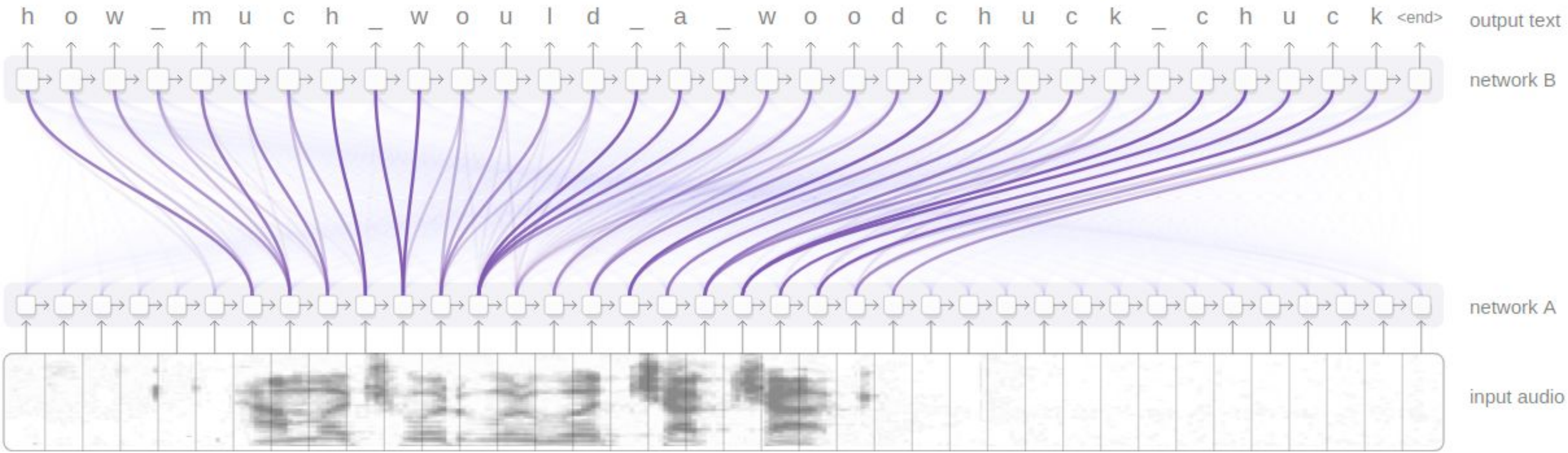


Attention: Architecture



Translation with Attention Demo

Attention: Also used for Audio



Attention: Also used for Image Captioning

Further Reading

Transformers

Attention Is All You Need

Ashish Vaswani[*]
Google Brain
avaswani@google.com

Noam Shazeer[*]
Google Brain
noam@google.com

Niki Parmar[*]
Google Research
nikip@google.com

Jakob Uszkoreit[*]
Google Research
usz@google.com

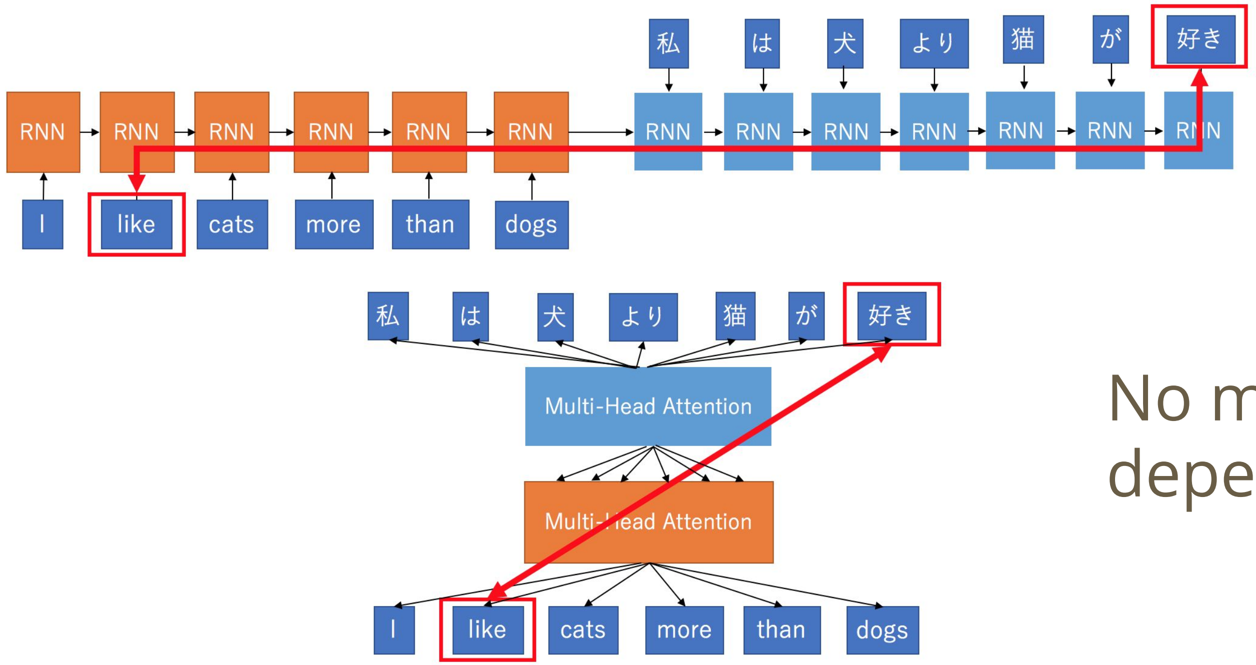
Llion Jones[*]
Google Research
llion@google.com

Aidan N. Gomez[*, †]
University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser[*]
Google Brain
lukaszkaiser@google.com

Illia Polosukhin[*, ‡]
illia.polosukhin@gmail.com

Transformers



No more long-term dependencies!