Text problems

Classification

I love purple cats. Purple cats are great

(+) Positive sentiment

Text problems

Translation

I love purple cats. Purple cats are great

J'adore les chats violets. Les chats violets sont géniaux

Classification

I love purple cats. Purple cats are great

(+) Positive sentiment

Tagging

I love purple cats. Purple cats are great

[subject] [verb] [adj.] [obj.] [adj.] [sub.] [verb] [adj.]

Generation

I love…

…purple cats. Purple cats are great

Representing Text

I love purple cats. Purple cats are great.

Text encoded as String of characters

- Varible length
- Discrete
- Not ordinal

Is 'a' closer to 'b' Than 'u'?

Ascii or Unicode

'a' = 97
'b' = 98
'c' = 99
⋮
'u' = 117

Representing text

Tokenization

I love purple cats. Purple cats are great.

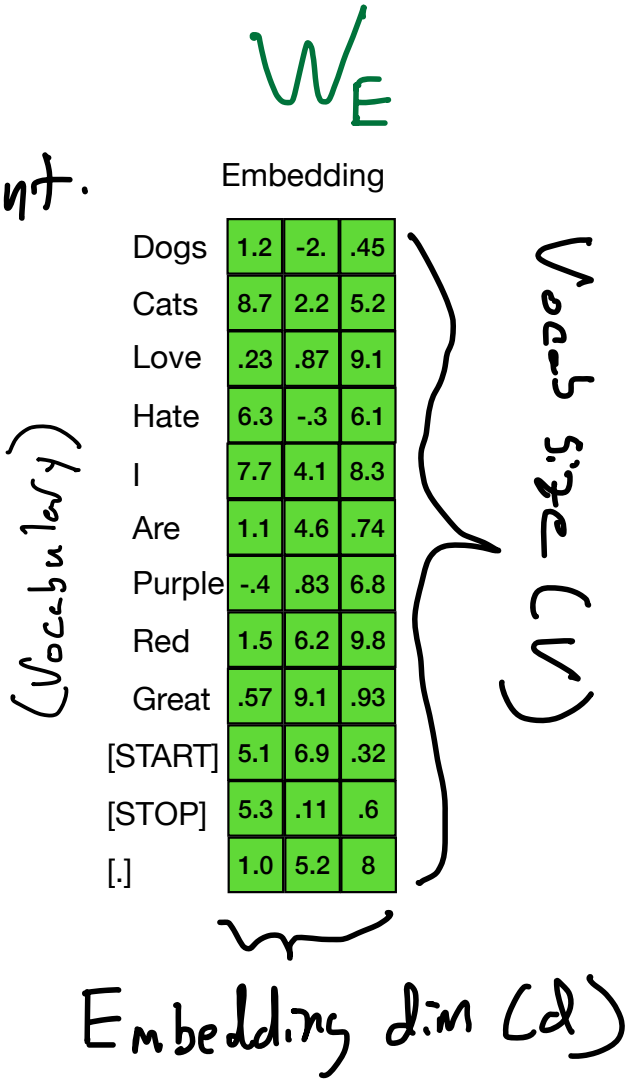
Break text into features

- Character-level: ['I', ' ', 'l', 'o', 'v', 'e'...]
- Word-level: ['I', 'love', 'purple', 'cats', '.'...]

Representing text

Convert tokens to floating-point numbers

- Represent each token with a vector of numbers
- One unique vector for each token in the vocabulary



Representing text

[START] I love purple cats. Purple cats are great. [STOP]

Special tokens for start and stop of text.

(we'll see later why this is useful!)

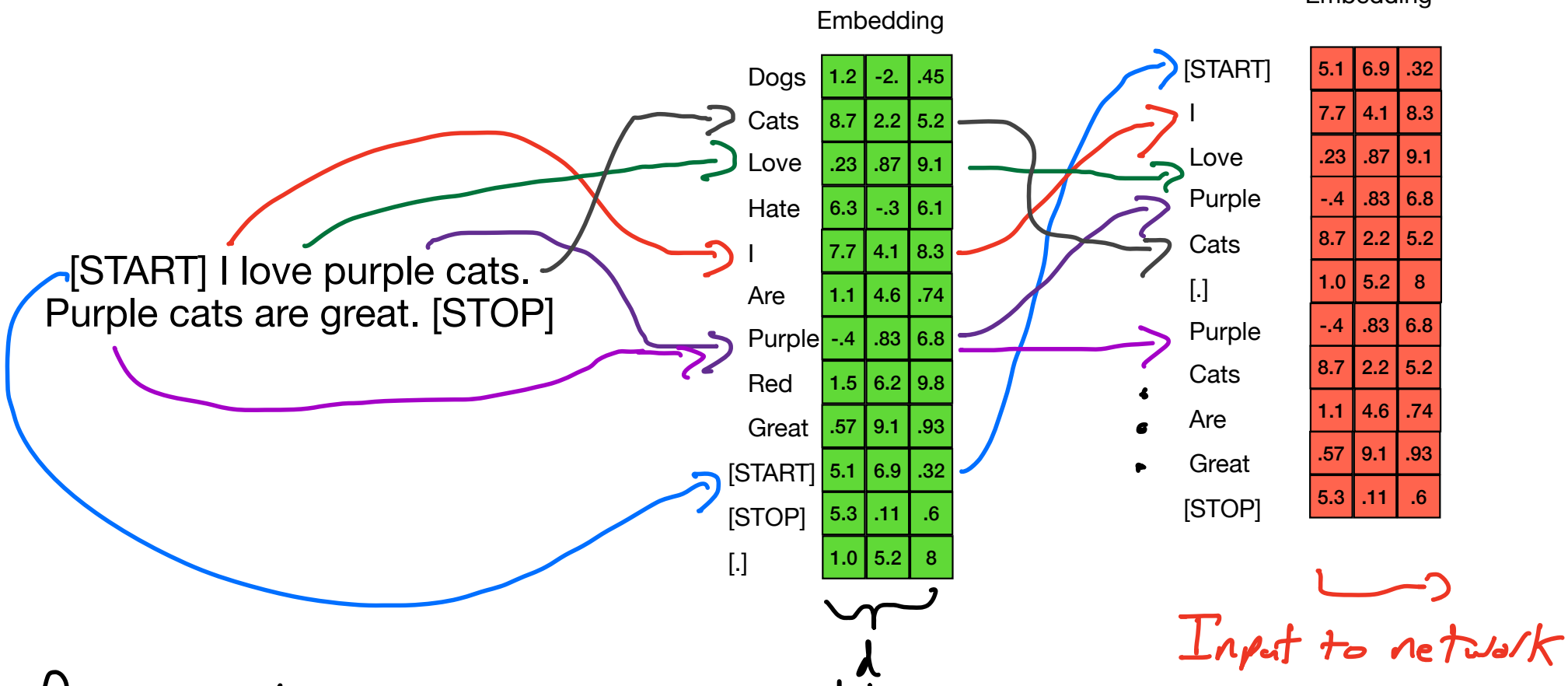
Representing Text

Vocab

$W_E : V \times d$

Embedded text

$X : L \times d$



Replace tokens w/ their embedding vector!

Representing text

Tokenization

I love purple cats. Purple cats are great.

- Character-level: ['I', ' ', 'l', 'o', 'v', 'e'...]
  - Small Vocab (~100)
  - Many features in a single observation (eg sentence)!
- Word-level: ['I', 'love', 'purple', 'cats', '.'...]
  - Large Vocab (~50,000+)
  - Few tokens per obs.
    - What if we encounter a new word, like a name?

Byte-pair embedding

I love purple cats. Purple cats are great.

- Common words should be single tokens
- Limit Vocab size, let rare or unseen words be multiple tokens

e.g.

cats → 'cats'

Purplish → 'Purpl' 'ish'

Gabe → 'G', 'a', 'b', 'e'

Byte-pair embedding

I love purple cats. Purple cats are great.

1) Start w/ all words as character tokens

Cats: 'C' 'a' 't' 's'

great: 'g' 'r' 'e' 'a' 't'

purple: 'p' 'u' 'r' 'p' 'l' 'e'

are: 'a' 'r' 'e'

love: 'l' 'o' 'v' 'e'

2) Combine tokens that appear together most often

'at': 3 'Ca': 2 'pu': 2 'ur': 2 're': 2 'rp': 2 'pl': 2 'le': 2

→ Cats: 'C' 'at' 's'

great: 'g' 'r' 'e' 'at'

3) Repeat until desired Vocab size

Byte - Pair embedding - GPT 3

Tokens	Characters
58	301

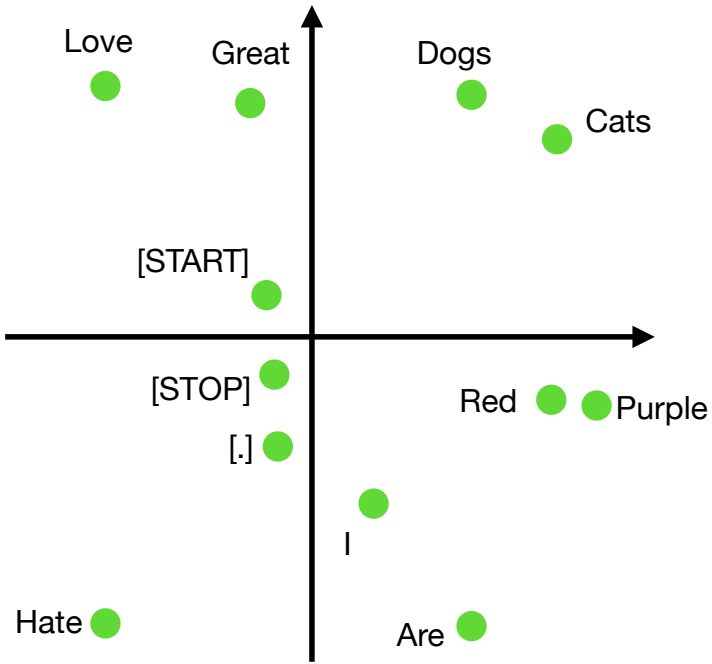
OpenAI's large language models (sometimes referred to as GPT's) process text using tokens, which are common sequences of characters found in a set of text. The models learn to understand the statistical relationships between these tokens, and excel at producing the next token in a sequence of tokens.

```
[5109, 15836, 596, 3544, 4221, 4211, 320, 57753, 14183, 311, 439, 480,
2898, 596, 8, 1920, 1495, 1701, 11460, 11, 902, 527, 4279, 24630, 315,
5885, 1766, 304, 264, 743, 315, 1495, 13, 578, 4211, 4048, 311, 3619,
279, 29564, 12135, 1990, 1521, 11460, 11, 323, 25555, 520, 17843, 279,
1828, 4037, 304, 264, 8668, 315, 11460, 13]
```

Word embeddings

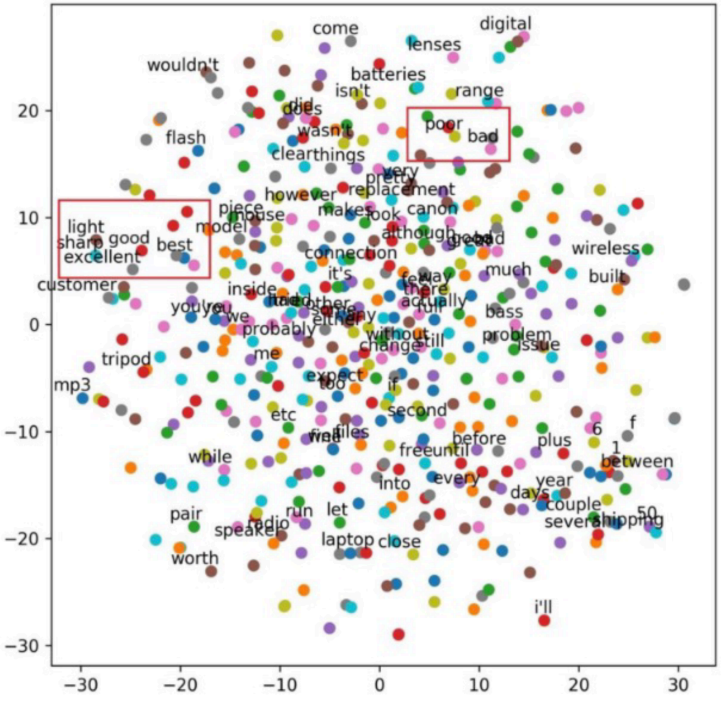
	Embedding		
Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.23	.87	9.1
Hate	6.3	-.3	6.1
I	7.7	4.1	8.3
Are	1.1	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	6.2	9.8
Great	.57	9.1	.93
[START]	5.1	6.9	.32
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

- Embedding map words to points
- often interpretable!



WE ← This is a parameter we can learn with gradient descent!

Word embeddings

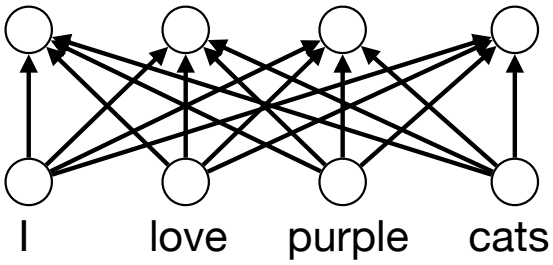


Cross-domain sentiment aware word embeddings for review sentiment analysis

Jun Liu[1] · Shuang Zheng[1] · Guangxia Xu[1,2,3] · Mingwei Lin[4]

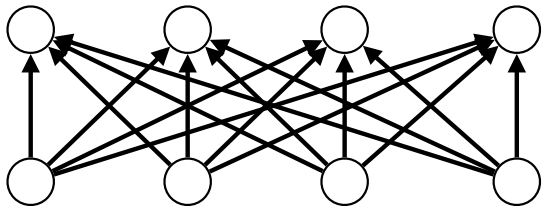
Applying neural networks

Fully-connected

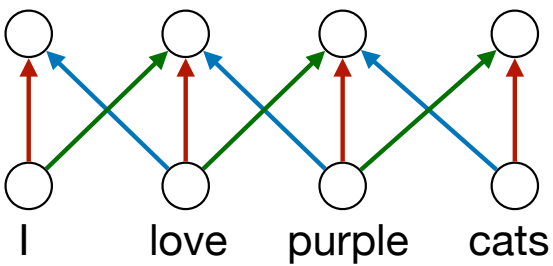


Sentences / text does not have a fixed length!

Fully-connected

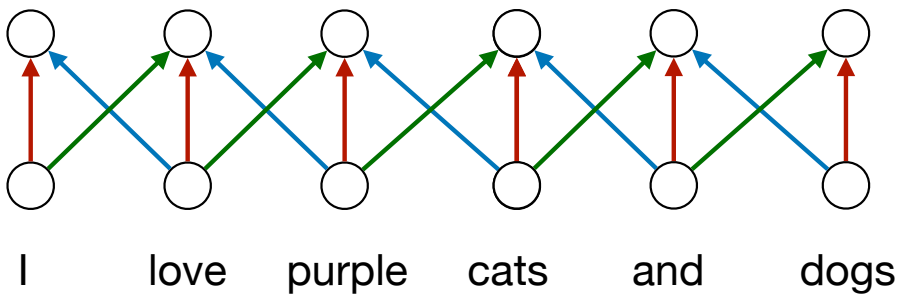


Convolutional



Ok for conv. layers!

Convolutional



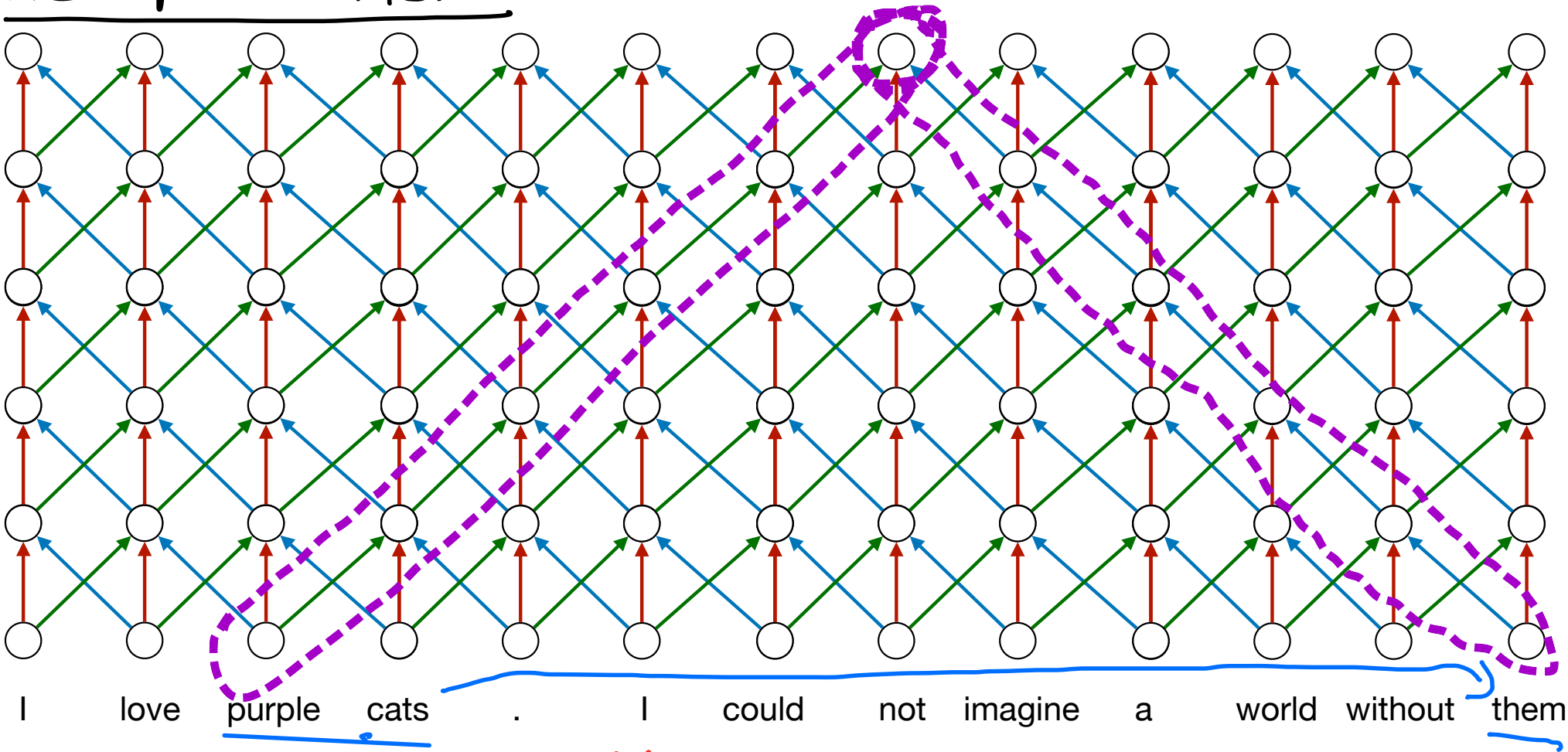
Long range dependencies

Geoffrey Everest Hinton CC FRS FRSC[12] (born 6 December 1947) is a British-Canadian computer scientist and cognitive psychologist, most noted for his work on artificial neural networks. From 2013 to 2023, he divided his time working for Google (Google Brain) and the University of Toronto, before publicly announcing his departure from Google in May 2023, citing concerns about the risks of artificial intelligence (AI) technology.[13] In 2017, he co-founded and became the chief scientific advisor of the Vector Institute in Toronto.[14][15]

- Need to recognize who 'he' refers to

Receptive field

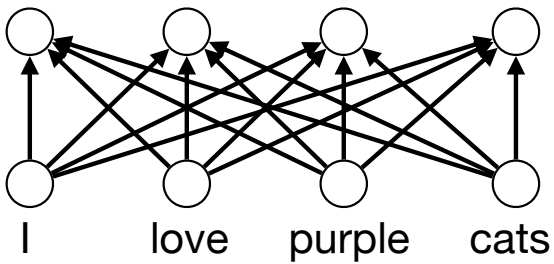
Window of inputs that a neuron depends on



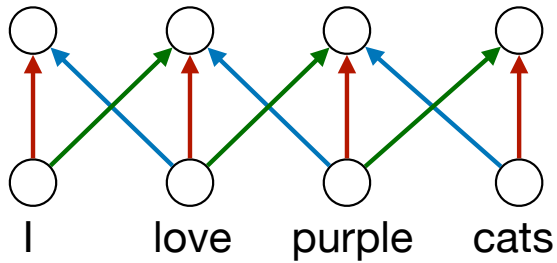
only one neuron could recognize this dependence

Recurrent networks

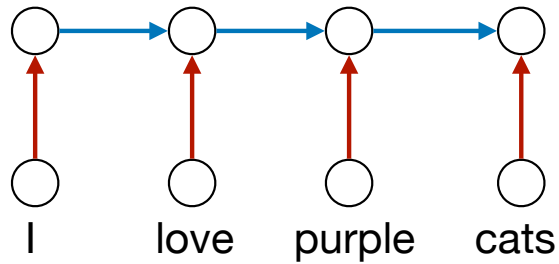
Fully-connected



Convolutional



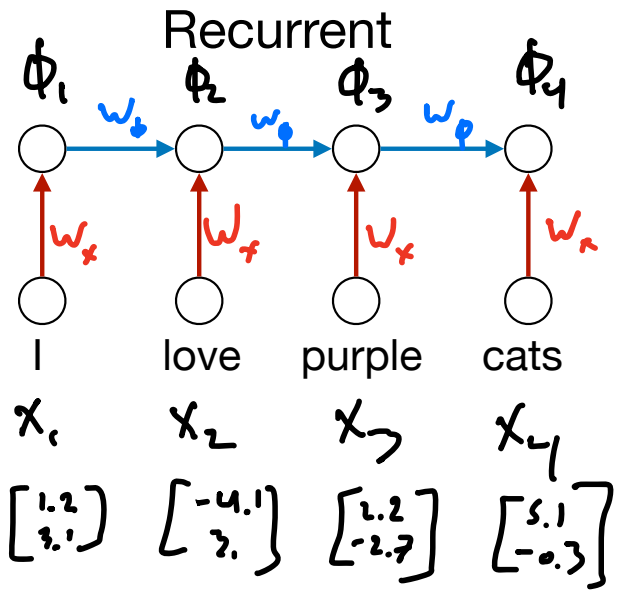
Recurrent



Idea: process features in sequence

↳ share weights like convolutional layers

Recurrent Networks



(+) Arbitrary length input

(+) No fixed receptive field*

(+) Shared weights

(or relu, tanh etc.

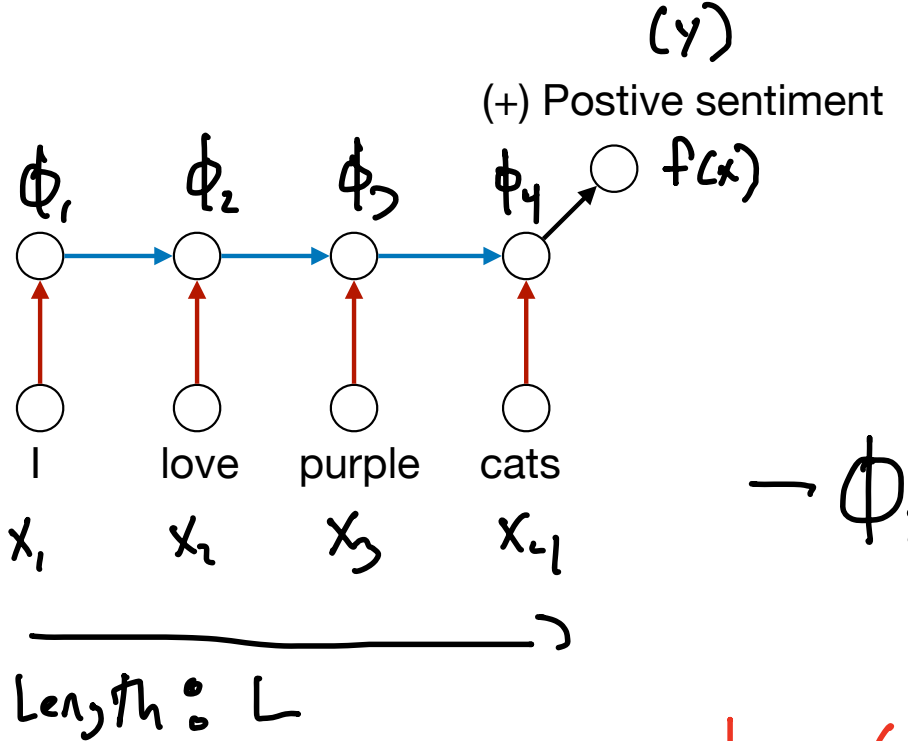
$$\phi_0 = \sigma(x_0^T W_x + b)$$

$$\phi_i = \sigma(x_i^T W_x + \phi_{i-1} W_\phi + b)$$

Each neuron depends on the previous in the same layer!

Downsides?

Classification



- Make prediction using representation at the end.

$$P(y=1|x) = \sigma(\phi_L^T w_o + b_o)$$

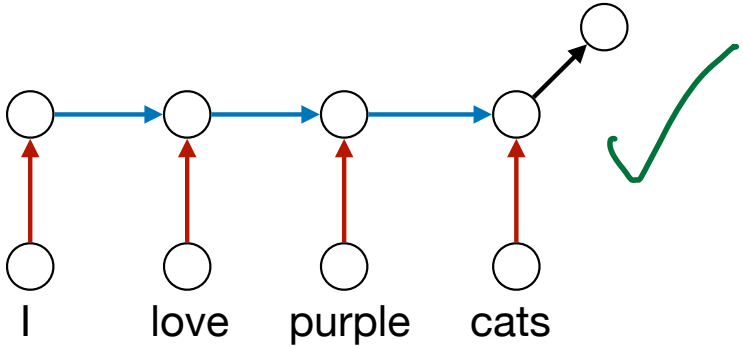
Logistic Regression

- ϕ_L contains info from every feature

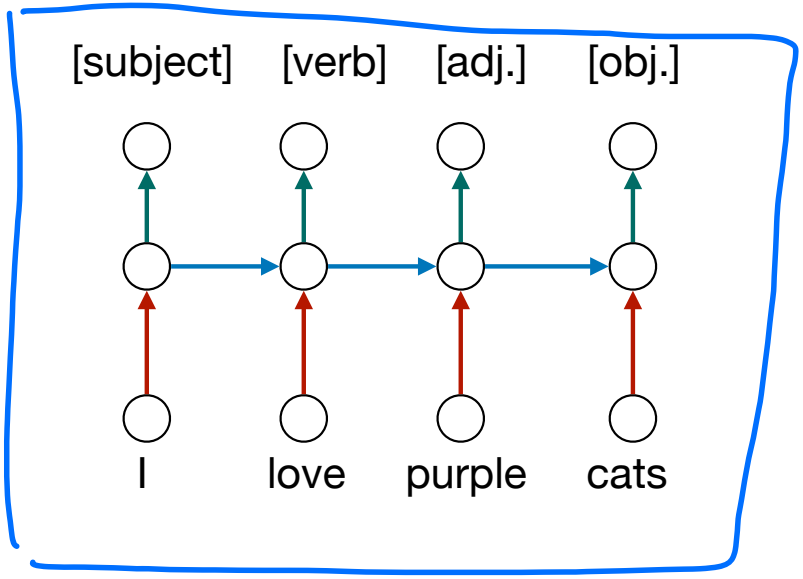
$$Loss(x, y, w) = -\log P(y|x)$$

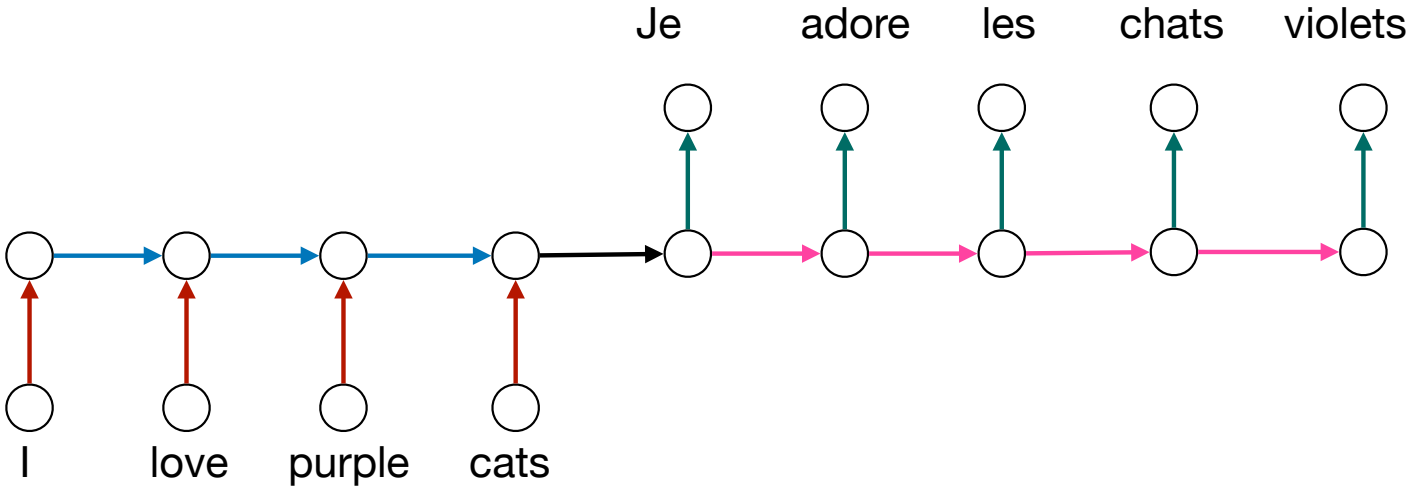
Other problems

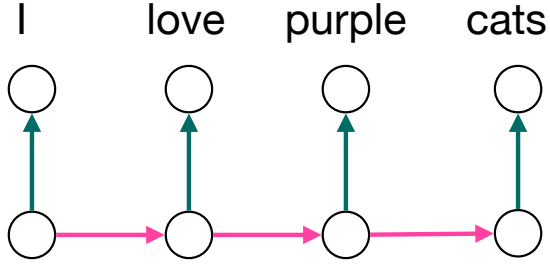
(+) Postive sentiment



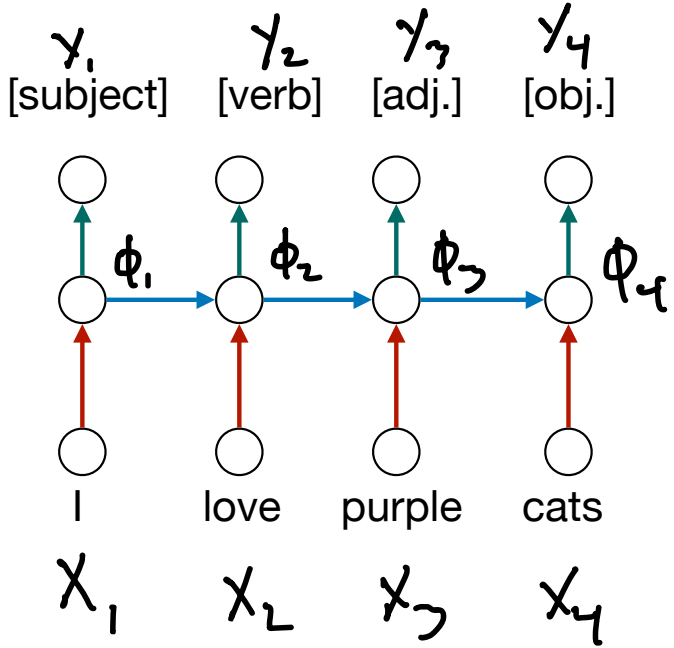
Tagging







Tagging



Make prediction for every word. → Multiple Labels!

E.g. Part-of-Speech tagging

Predict at each location using corresponding representation

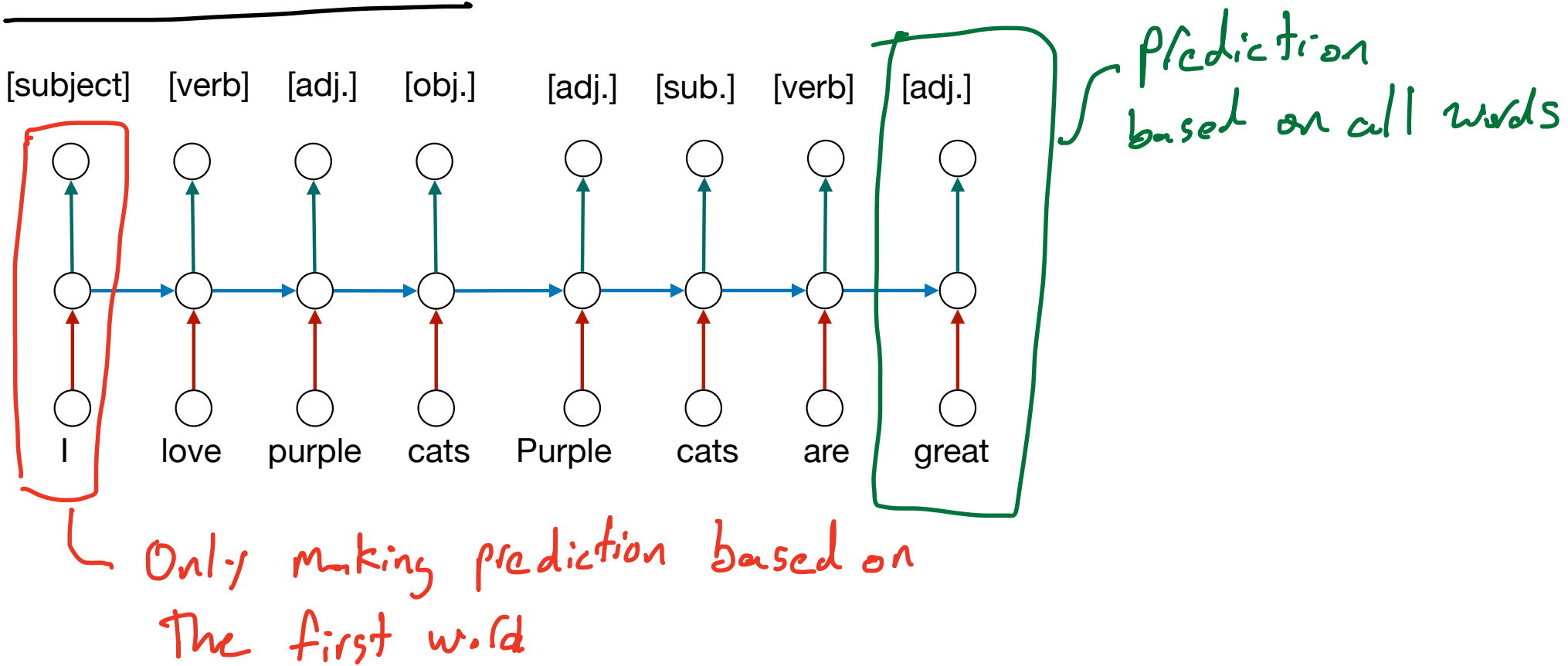
$$P(y_i = c|x) = \text{Softmax}(\phi_i^T W_c + b)_c$$

Multinomial logistic regression

$$\text{Loss}(x, y, w) = -\sum_{i=1}^{L} \log P(y_i | x)$$

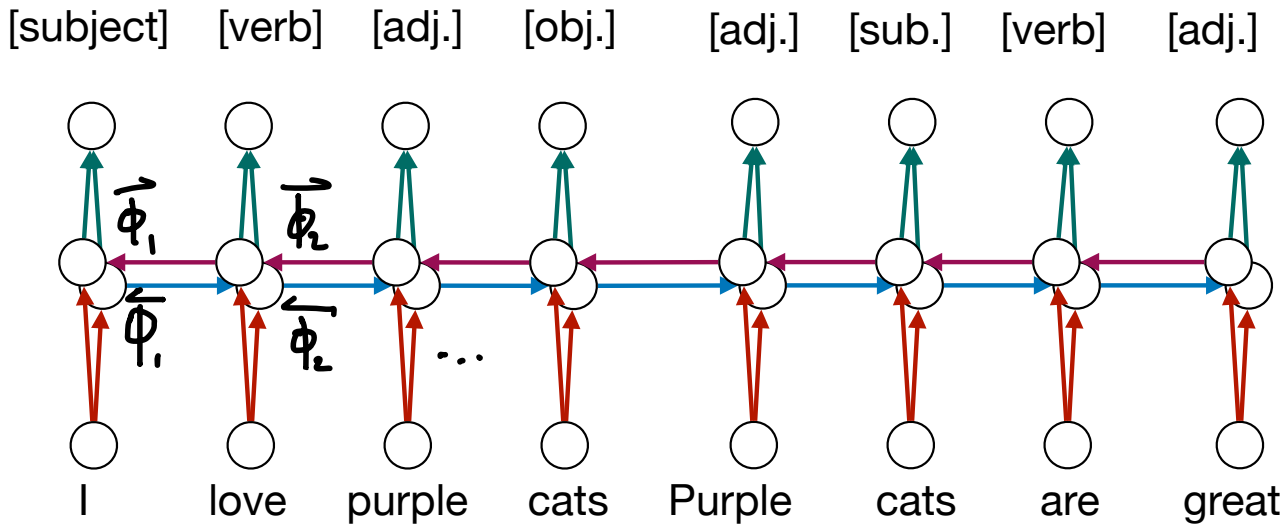
Sum loss over locations!

Backward info



Bi-directional RNNs

Include both forward and backward info



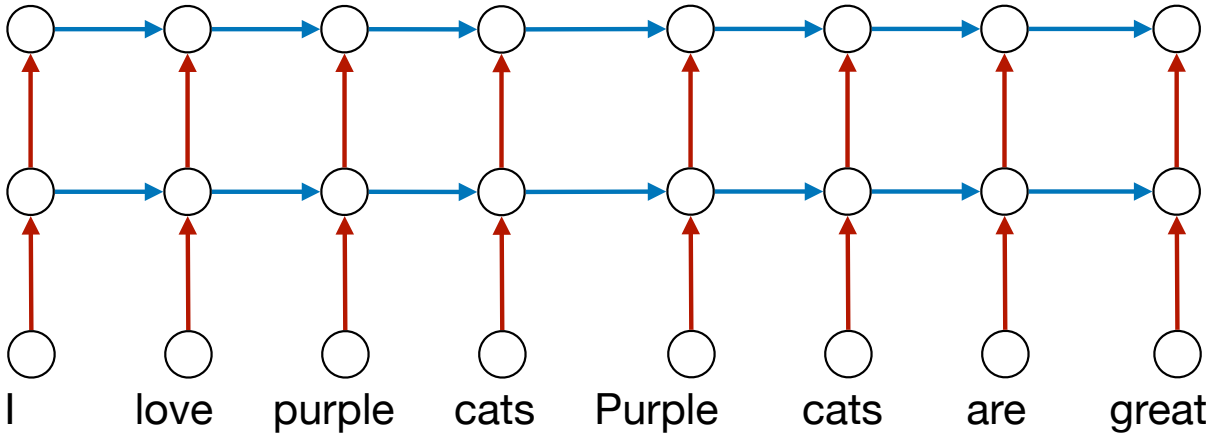
$$\vec{\phi}_i = \sigma\left(x_i^T \vec{w}_x + \vec{\phi}_{i-1} \vec{w}_\phi + \vec{b}\right)$$

$$\overleftarrow{\phi} = \sigma\left(x_i^T \overleftarrow{w}_x + \overleftarrow{\phi}_{i+1} \overleftarrow{w}_\phi + \overleftarrow{b}\right)$$

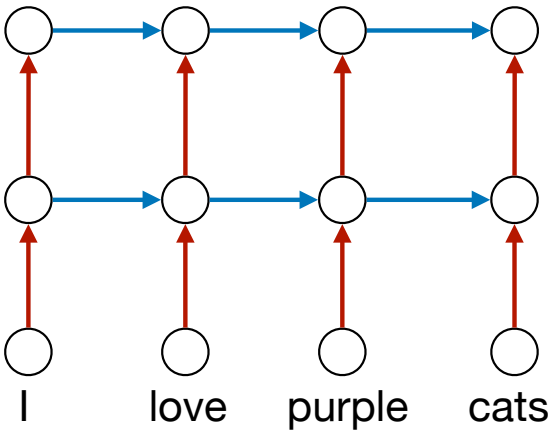
$$P(y_i|x) = f_M\left(\vec{\phi}_i \vec{w}_o + \overleftarrow{\phi}_i \overleftarrow{w}_o + b\right)$$

Multi-Layer RNNs

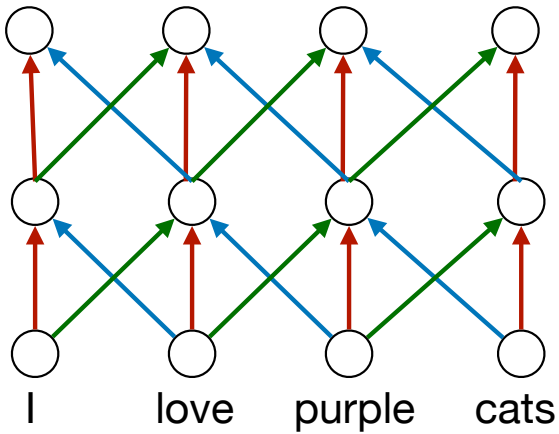
Can stack RNNs as Conv. or f.c. Layers



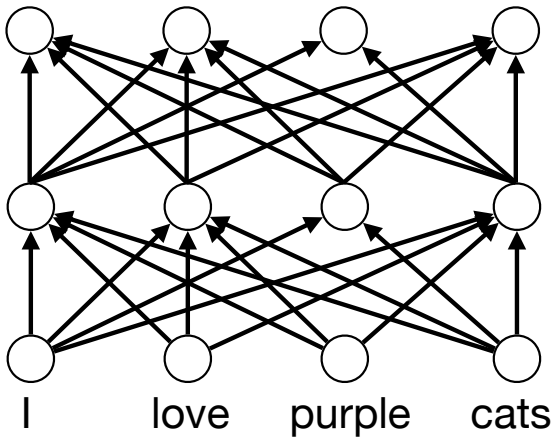
Recurrent



Convolutional

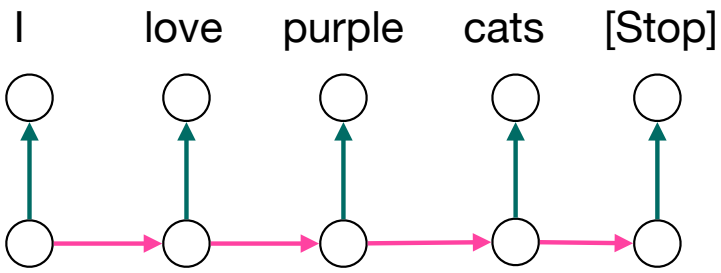


Fully-connected



Language Modelling

How do we learn to generate text?



- Prediction without input?

Language Modelling

(Next word prediction)

Predict next word based on context

{Start} I love purple → Cats

Input

Predict

$f($ [{Start}, I, love, purple] $)$

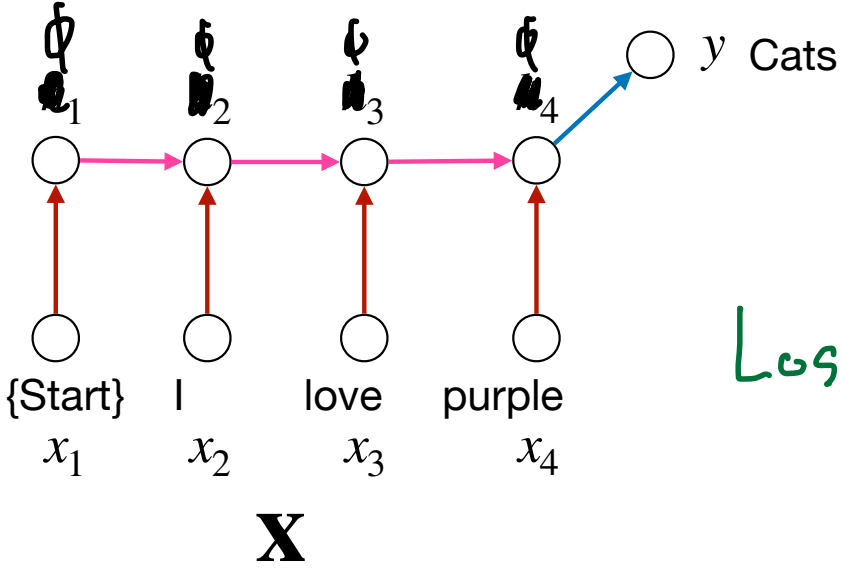
5.1	6.9	.32	5.1
6.9	4.1	8.3	7.7
.32	.87	9.1	.23

→

.02	$p(\text{word} = \text{dog} \mid \text{input})$
.18	$p(\text{word} = \text{colors} \mid \text{input})$
.01	$p(\text{word} = \text{love} \mid \text{input})$
.24	$p(\text{word} = \text{cats} \mid \text{input})$
.51	$p(\text{word} = \text{rain} \mid \text{input})$
.01	$p(\text{word} = \text{I} \mid \text{input})$

⋮

Language Modelling



Prediction

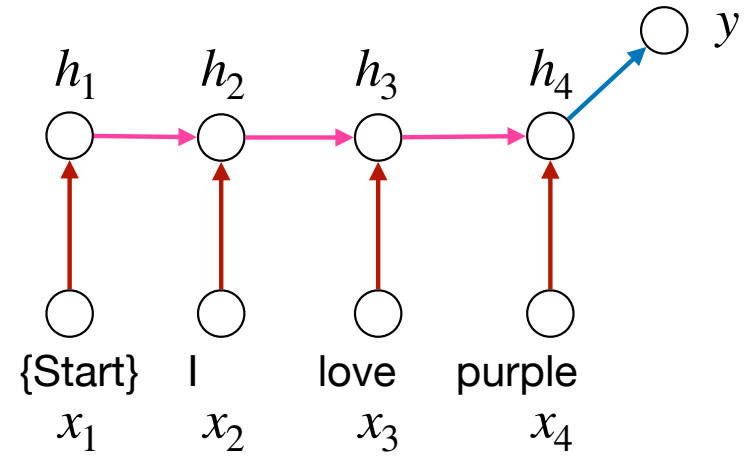
$$P(y=\text{cats} \mid x) = \text{Softmax}(\phi_4^T W + b)_{\text{cats}}$$

W : $d \times V$ → Need 1 output per word in Vocab!

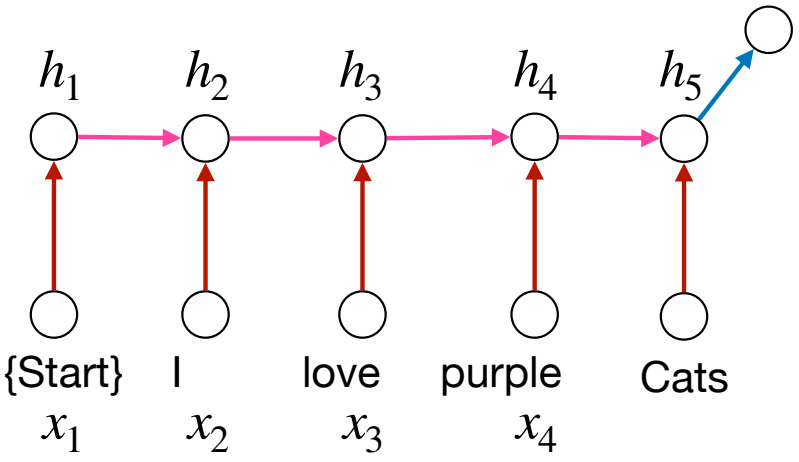
Loss : NLL!

$$\text{Loss} = -\log P(y \mid x)$$

Sampling



X



X

.02	$p(\text{word} = \text{dog} \mid \text{input})$
.18	$p(\text{word} = \text{colors} \mid \text{input})$
.01	$p(\text{word} = \text{love} \mid \text{input})$
.24	$p(\text{word} = \text{cats} \mid \text{input})$
.51	$p(\text{word} = \text{rain} \mid \text{input})$
.01	$p(\text{word} = \text{I} \mid \text{input})$
⋮	

→

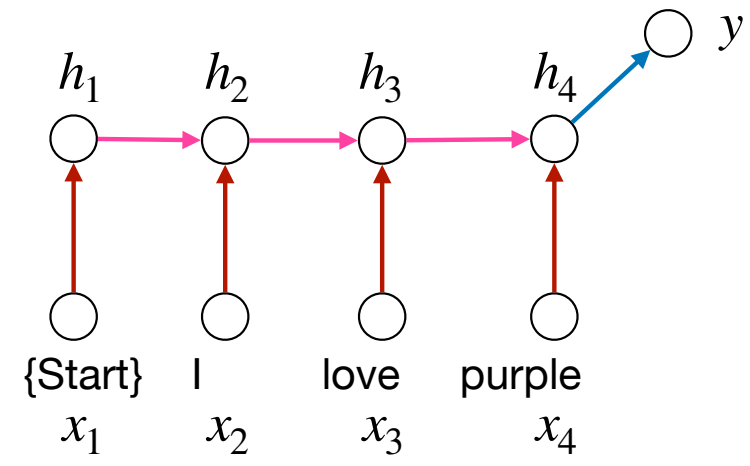
Cats

↳ Pick next word according to pred. probabilities!

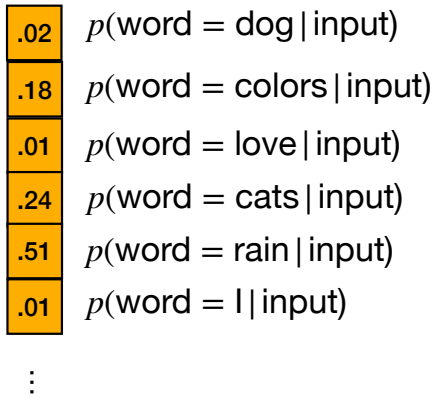
.12	$p(\text{word} = \text{very} \mid \text{input})$
.08	$p(\text{word} = \text{much} \mid \text{input})$
.01	$p(\text{word} = \text{love} \mid \text{input})$
.91	$p(\text{word} = \{\text{Stop}\} \mid \text{input})$
.01	$p(\text{word} = \text{rain} \mid \text{input})$
.01	$p(\text{word} = \text{I} \mid \text{input})$
⋮	

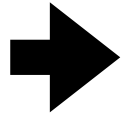
→

Very

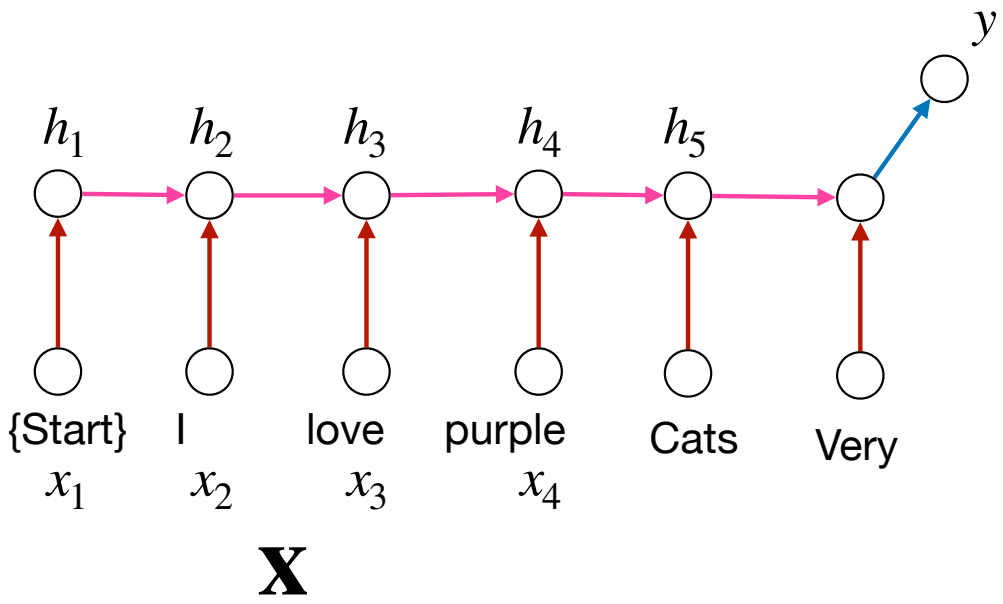


Sampling

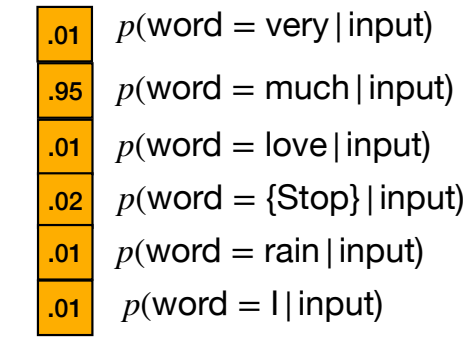


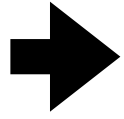


Cats

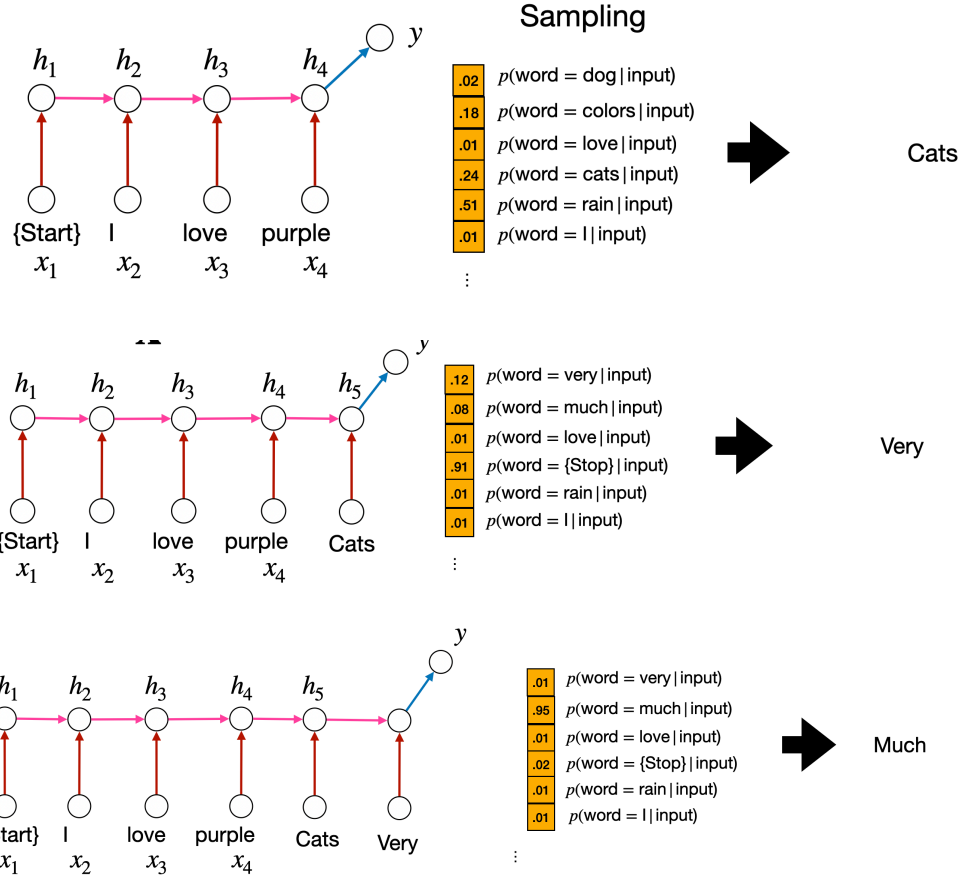


X





Much


Sampling
h_1 h_2 h_3 h_4 y
{Start} I love purple
x_1 x_2 x_3 x_4
.02 p(word = dog | input)
.18 p(word = colors | input)
.01 p(word = love | input)
.24 p(word = cats | input)
.51 p(word = rain | input)
.01 p(word = I | input)
Cats
h_1 h_2 h_3 h_4 h_5 y
{Start} I love purple Cats
x_1 x_2 x_3 x_4
.12 p(word = very | input)
.08 p(word = much | input)
.01 p(word = love | input)
.91 p(word = {Stop} | input)
.01 p(word = rain | input)
.01 p(word = I | input)
Very
h_1 h_2 h_3 h_4 h_5 y
{Start} I love purple Cats Very
x_1 x_2 x_3 x_4
.01 p(word = very | input)
.95 p(word = much | input)
.01 p(word = love | input)
.02 p(word = {Stop} | input)
.01 p(word = rain | input)
.01 p(word = I | input)
Much

Dataset: sentences

{start} I love purple cats {stop} {start} Purple rain is my favorite song {stop} {start} Goldfish crackers are delicious {stop}

{start} Dogs are adorable {stop} {start} Neural networks are cool {stop} {start} I have a pet goldfish {stop}

One sentence: **{start} I love purple cats {stop}**

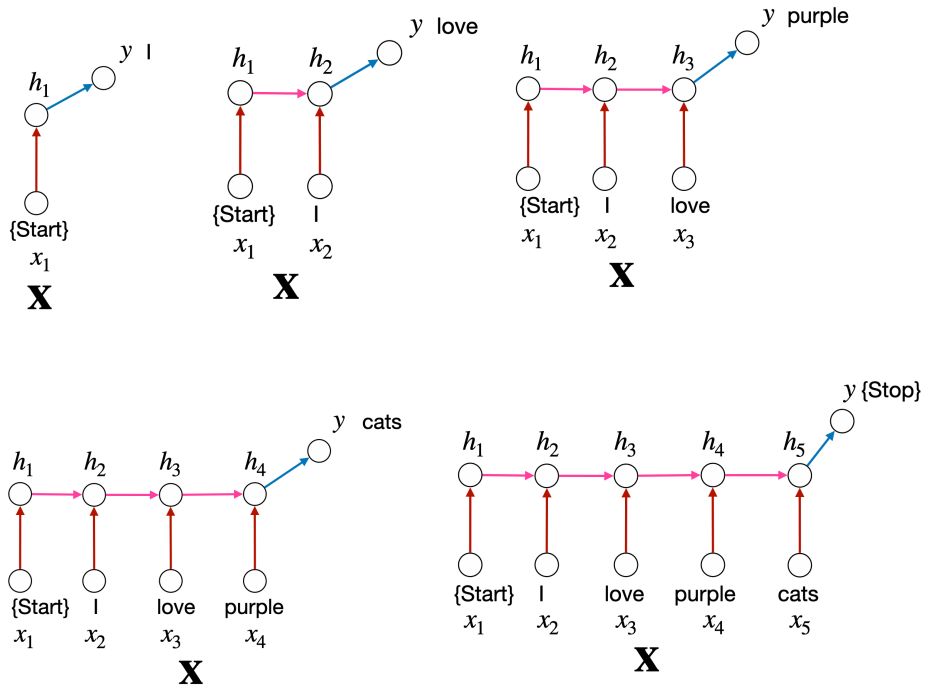
How many observations?

Dataset: sentences

{start} I love purple cats {stop} {start} Purple rain is my favorite song {stop} {start} Goldfish crackers are delicious {stop}

{start} Dogs are adorable {stop} {start} Neural networks are cool {stop} {start} I have a pet goldfish {stop}

One sentence: **{start} I love purple cats {stop}**



Loss:

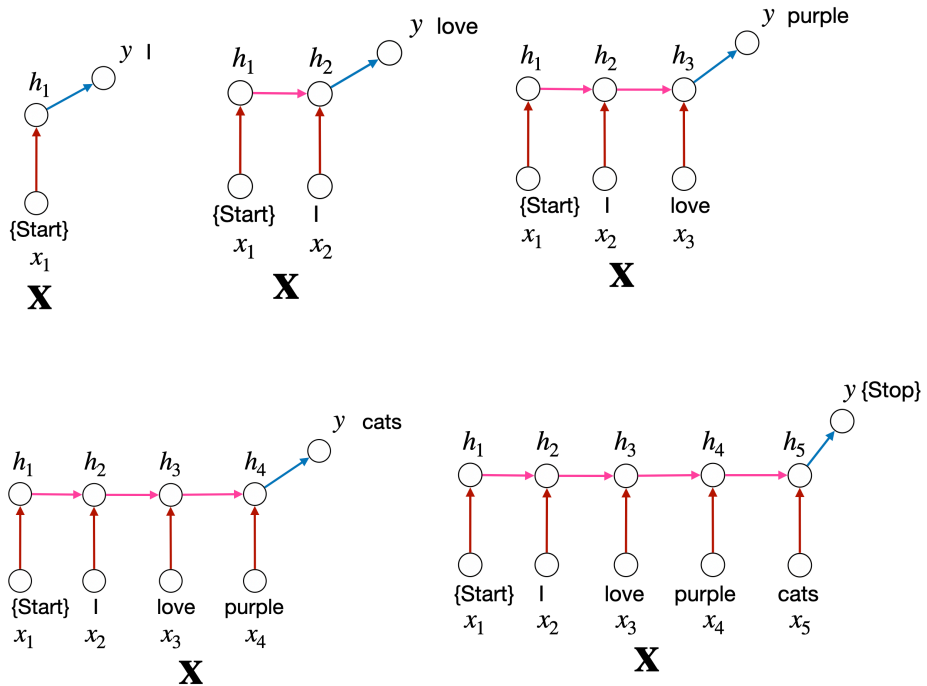
$$-\sum_{i=1}^{L} \log P(x_i \mid x_1 \ldots x_{i-1})$$

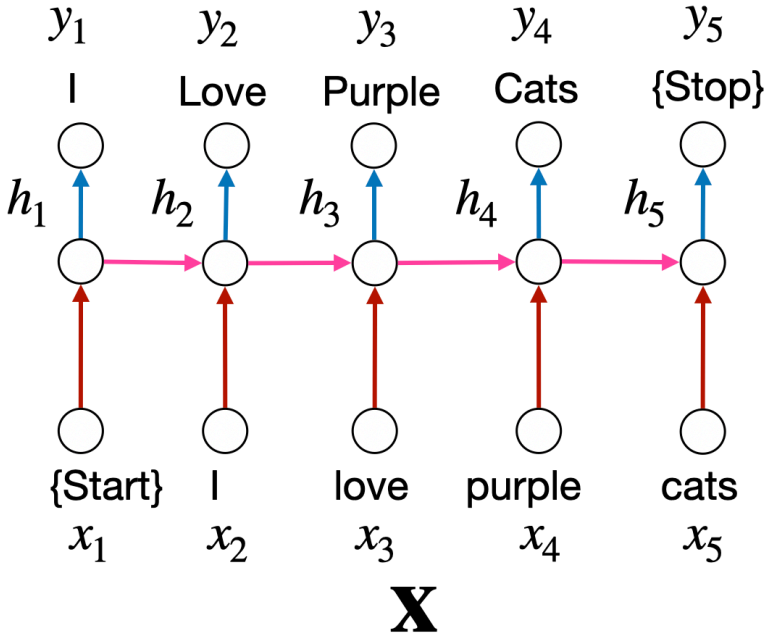
Dataset: sentences

{start} I love purple cats {stop} {start} Purple rain is my favorite song {stop} {start} Goldfish crackers are delicious {stop}

{start} Dogs are adorable {stop} {start} Neural networks are cool {stop} {start} I have a pet goldfish {stop}

One sentence: **{start} I love purple cats {stop}**



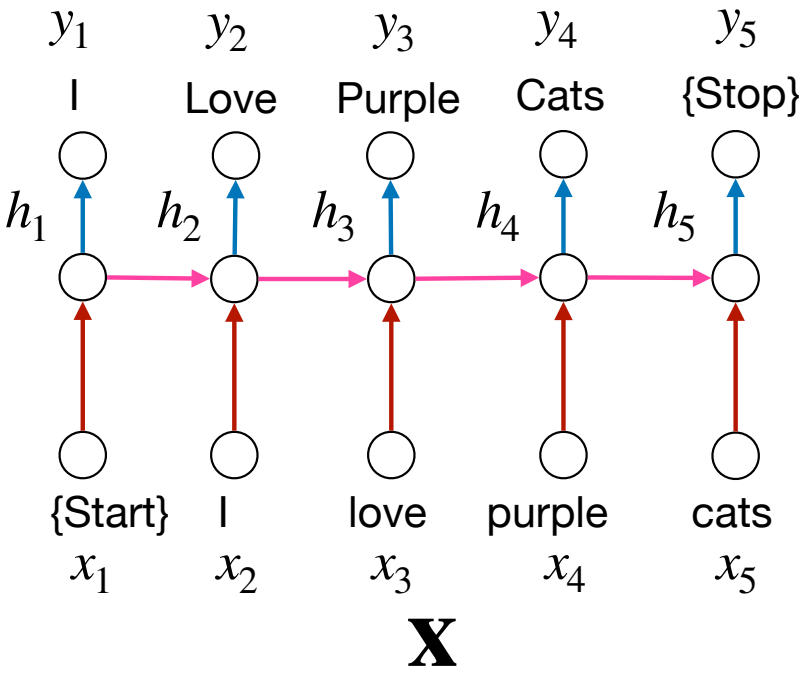


Dataset: sentences

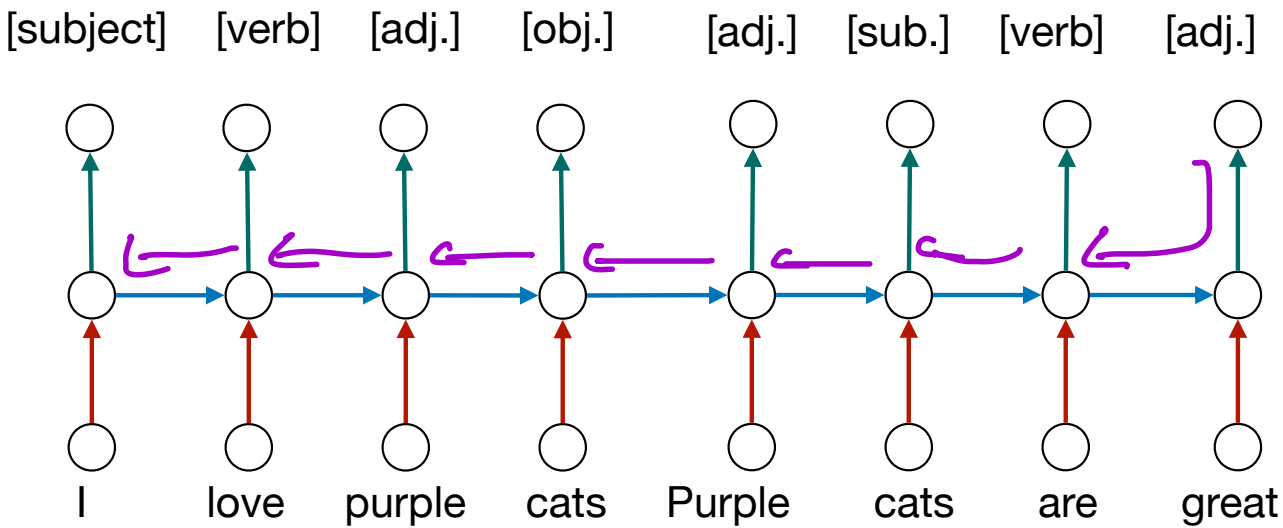
{start} I love purple cats {stop} {start} Purple rain is my favorite song {stop} {start} Goldfish crackers are delicious {stop}

{start} Dogs are adorable {stop} {start} Neural networks are cool {stop} {start} I have a pet goldfish {stop}

One sentence: **{start} I love purple cats {stop}**



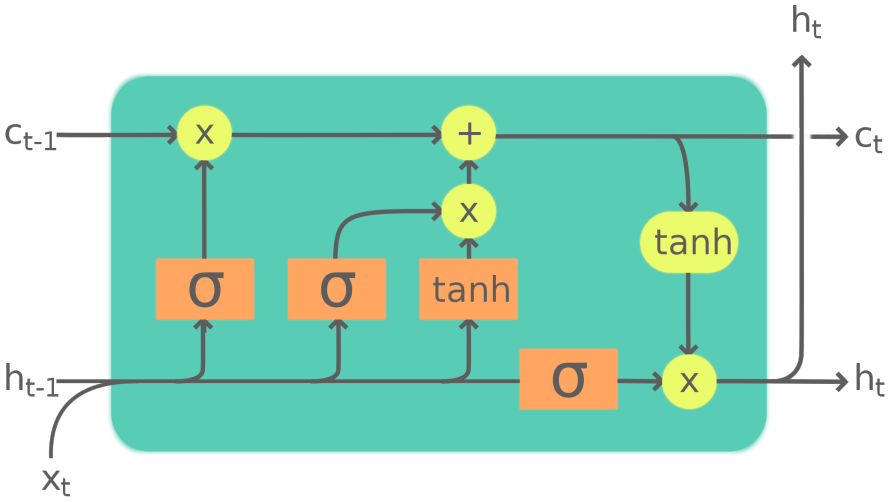
Vanishing / exploding gradients



Equivalent to Many Layers!

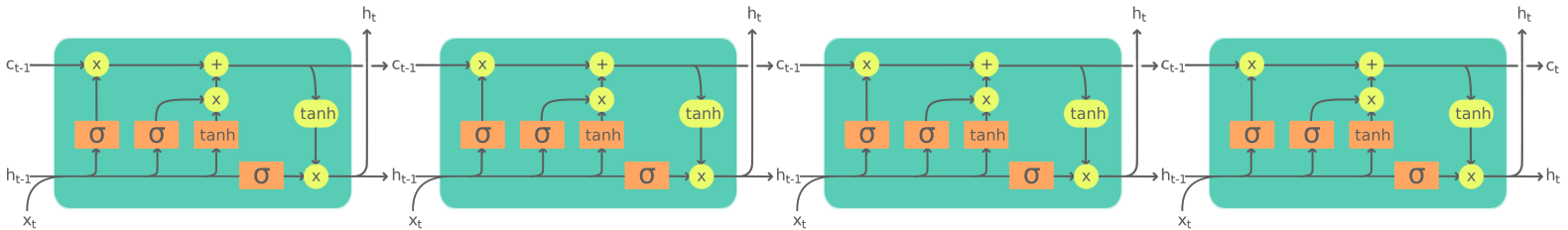
→ Vanishing / exploding gradients

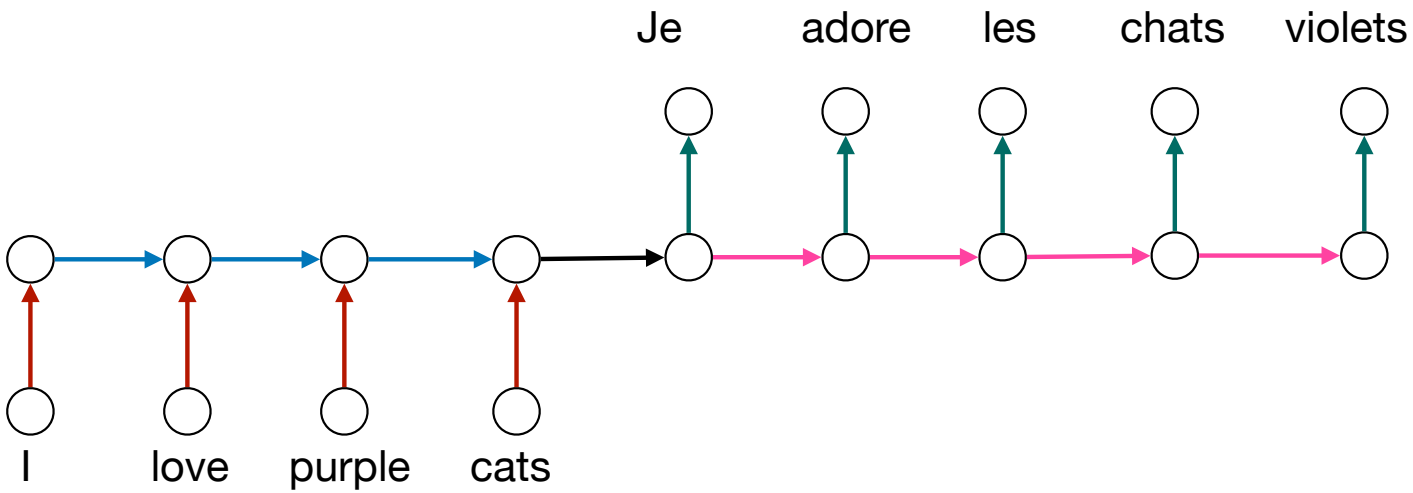
Long-Short term Memory (LSTM)



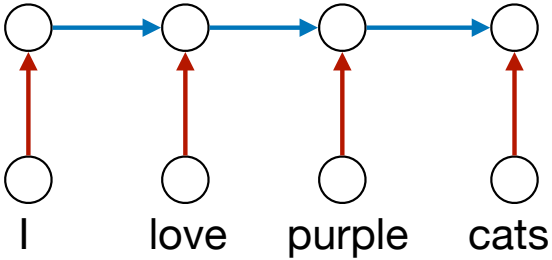
- Similar to residual network

Legend: Layer Componentwise Copy Concatenate

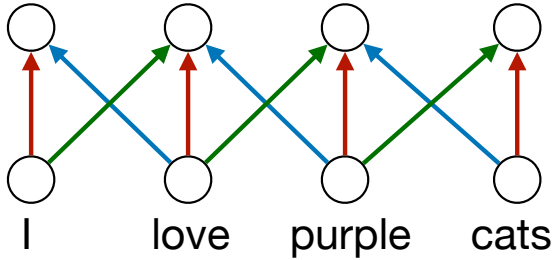



Je
adore
les
chats
violets
I
love
purple
cats

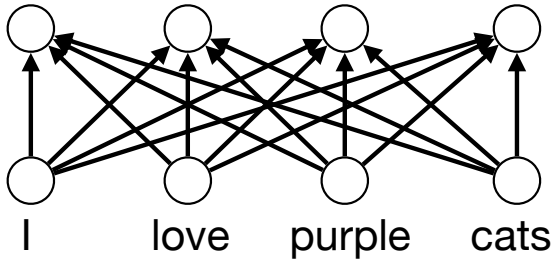
Recurrent



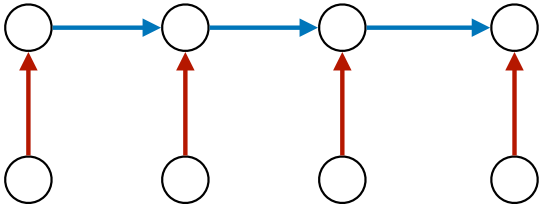
Convolutional



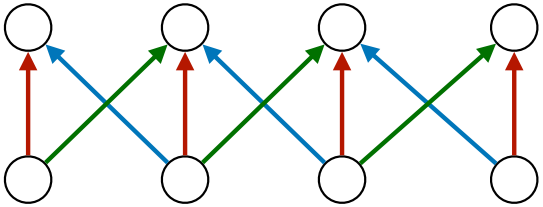
Fully-connected



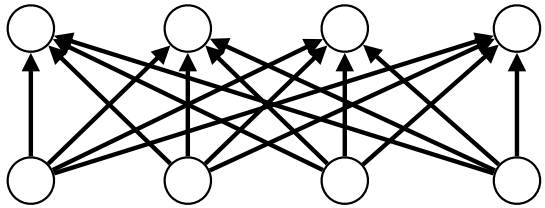
Recurrent



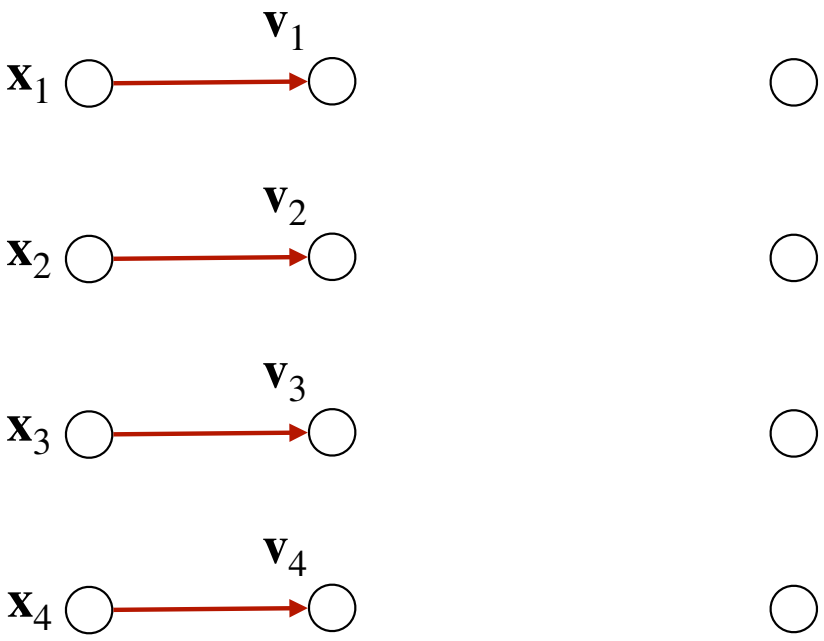
Convolutional



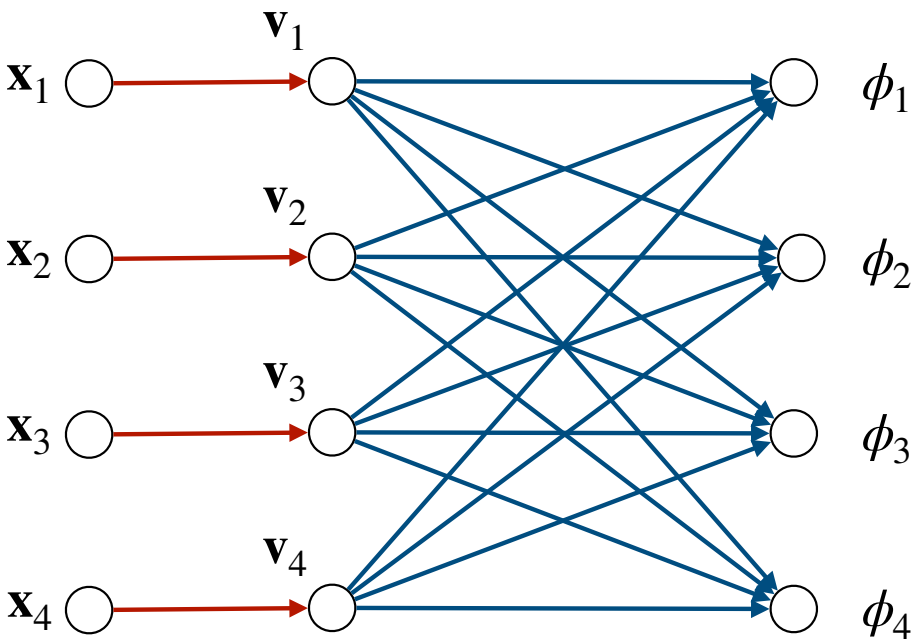
Fully-connected



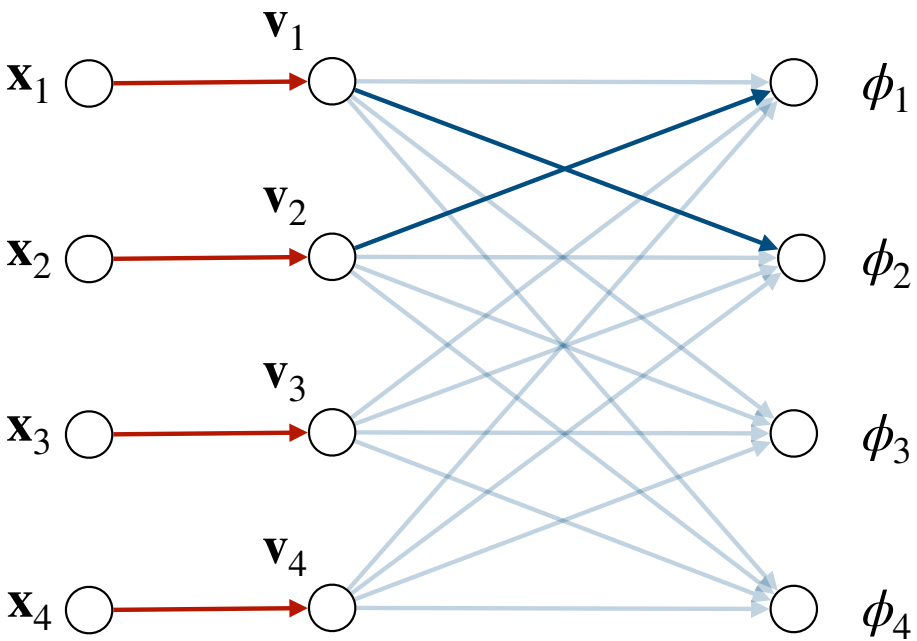
Attention


$\mathbf{v}_1$
$\mathbf{x}_1$
$\mathbf{v}_2$
$\mathbf{x}_2$
$\mathbf{v}_3$
$\mathbf{x}_3$
$\mathbf{v}_4$
$\mathbf{x}_4$

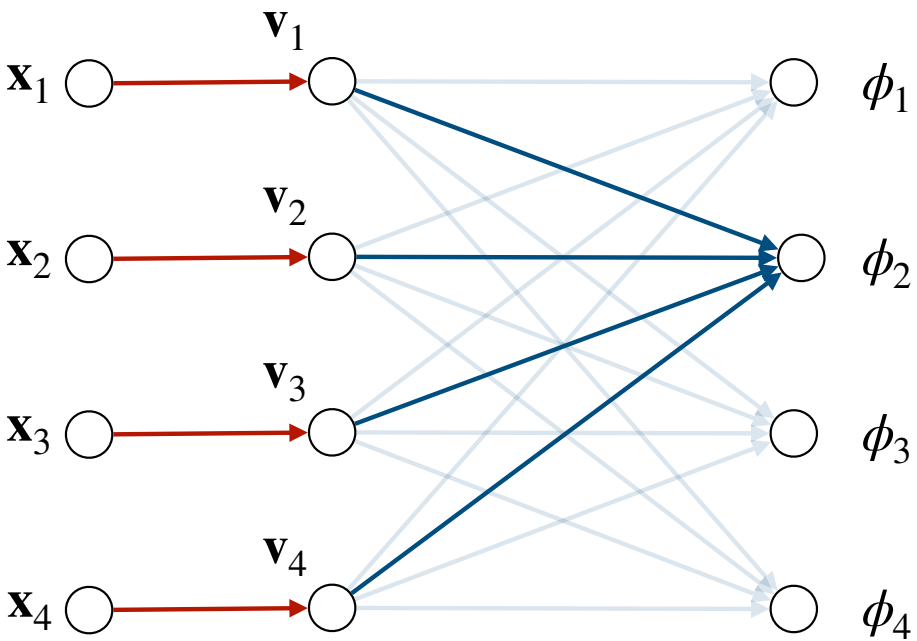
Attention


$\mathbf{v}_1$
$\mathbf{x}_1$
$\mathbf{v}_2$
$\mathbf{x}_2$
$\mathbf{v}_3$
$\mathbf{x}_3$
$\mathbf{v}_4$
$\mathbf{x}_4$
ϕ_1
ϕ_2
ϕ_3
ϕ_4

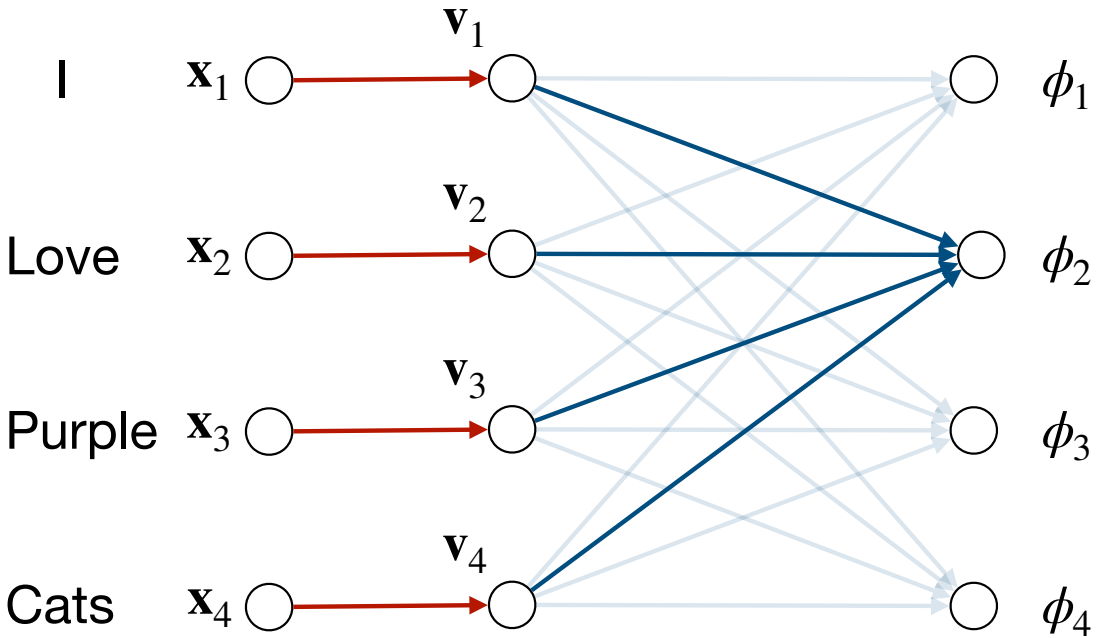
Attention


$\mathbf{v}_1$
$\mathbf{x}_1$
$\mathbf{v}_2$
$\mathbf{x}_2$
$\mathbf{v}_3$
$\mathbf{x}_3$
$\mathbf{v}_4$
$\mathbf{x}_4$
ϕ_1
ϕ_2
ϕ_3
ϕ_4

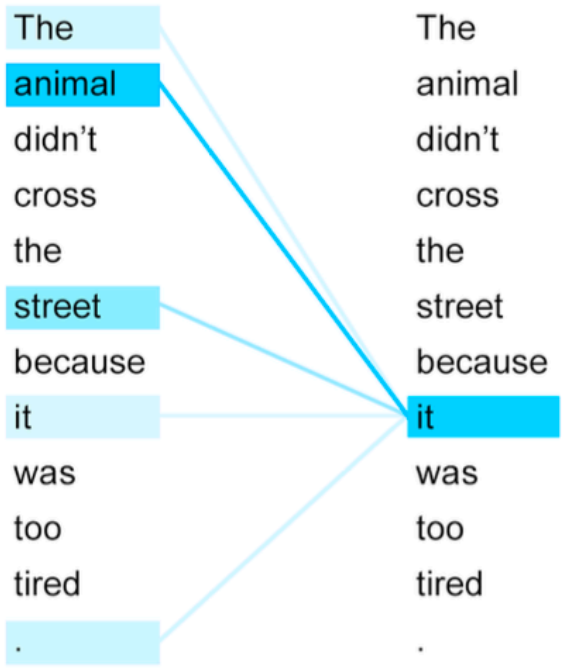
Attention

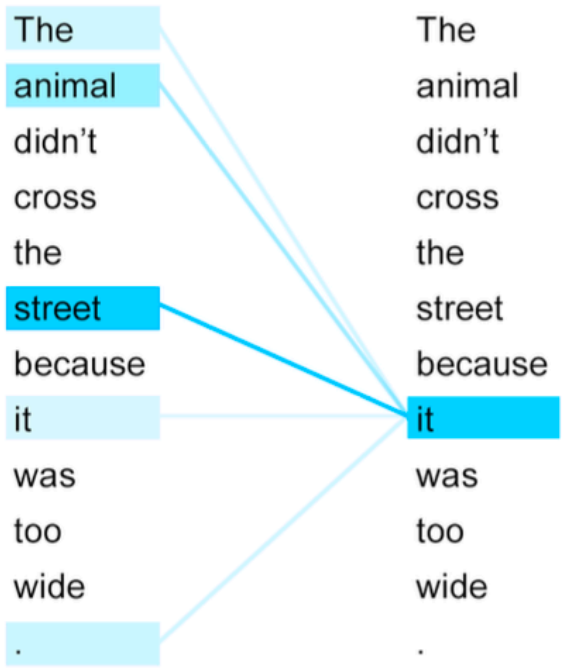

$\mathbf{v}_1$
$\mathbf{x}_1$
ϕ_1
$\mathbf{v}_2$
$\mathbf{x}_2$
ϕ_2
$\mathbf{v}_3$
$\mathbf{x}_3$
ϕ_3
$\mathbf{v}_4$
$\mathbf{x}_4$
ϕ_4

Attention

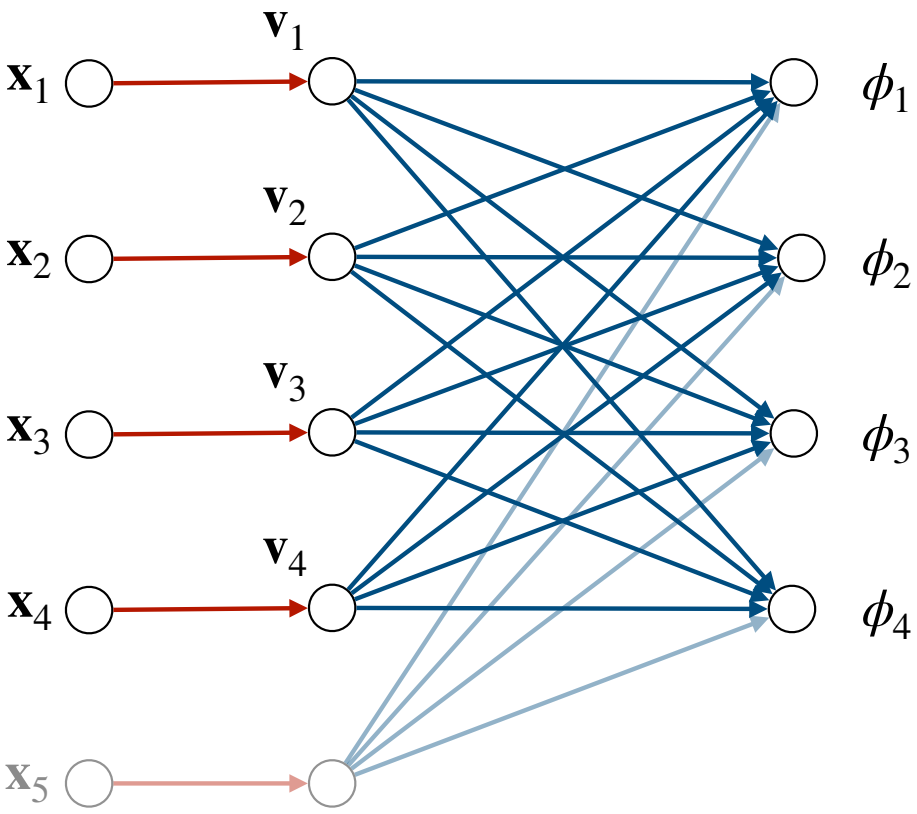

I
Love
Purple
Cats
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{v}_1$
$\mathbf{v}_2$
$\mathbf{v}_3$
$\mathbf{v}_4$
ϕ_1
ϕ_2
ϕ_3
ϕ_4

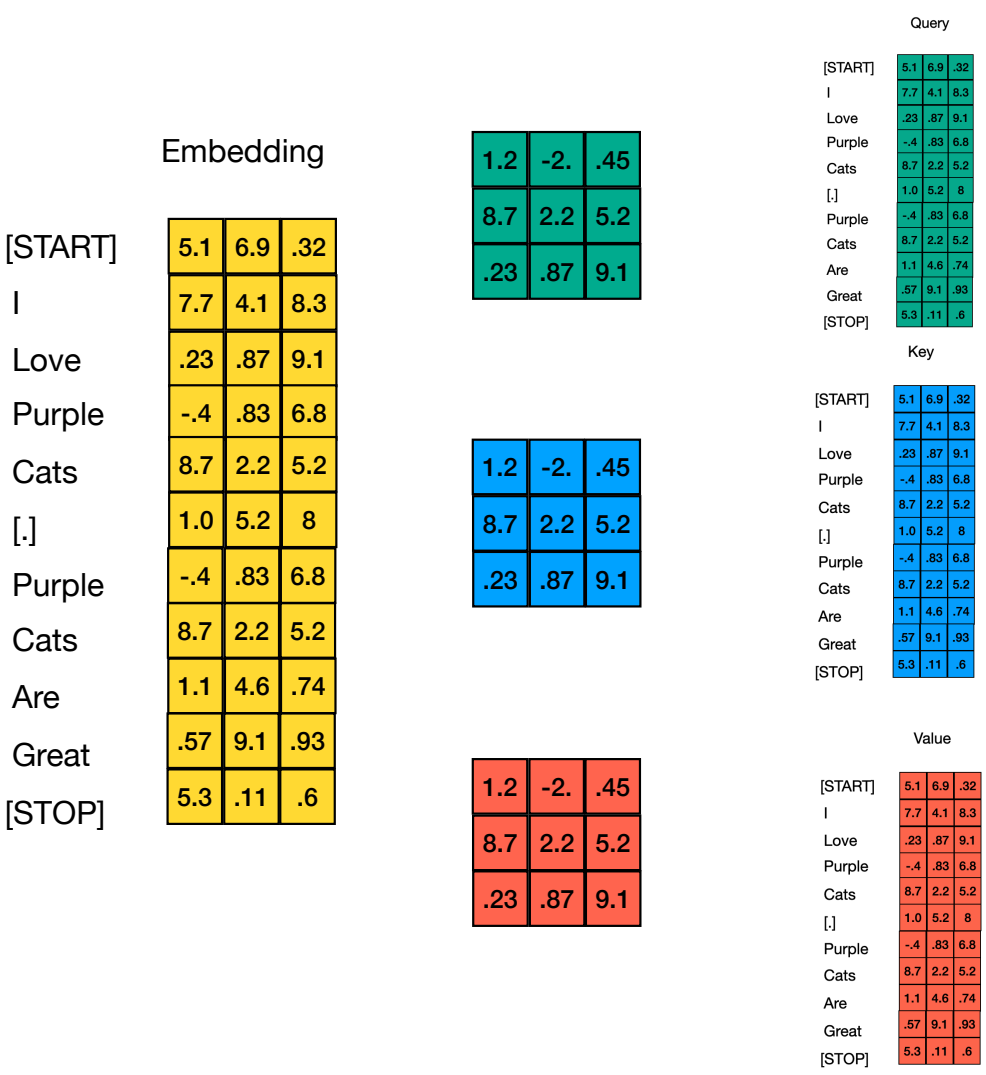

The
Law
will
never
be
perfect
,
but
its
application
should
be
just
-
this
is
what
we
are
missing
,
in
my
opinion
.
<EOS>
<pad>
The
Law
will
never
be
perfect
,
but
its
applicatio
should
be
just
-
this
is
what
we
are
missing
,
in
my
opinion
.
<EOS>
<pad>

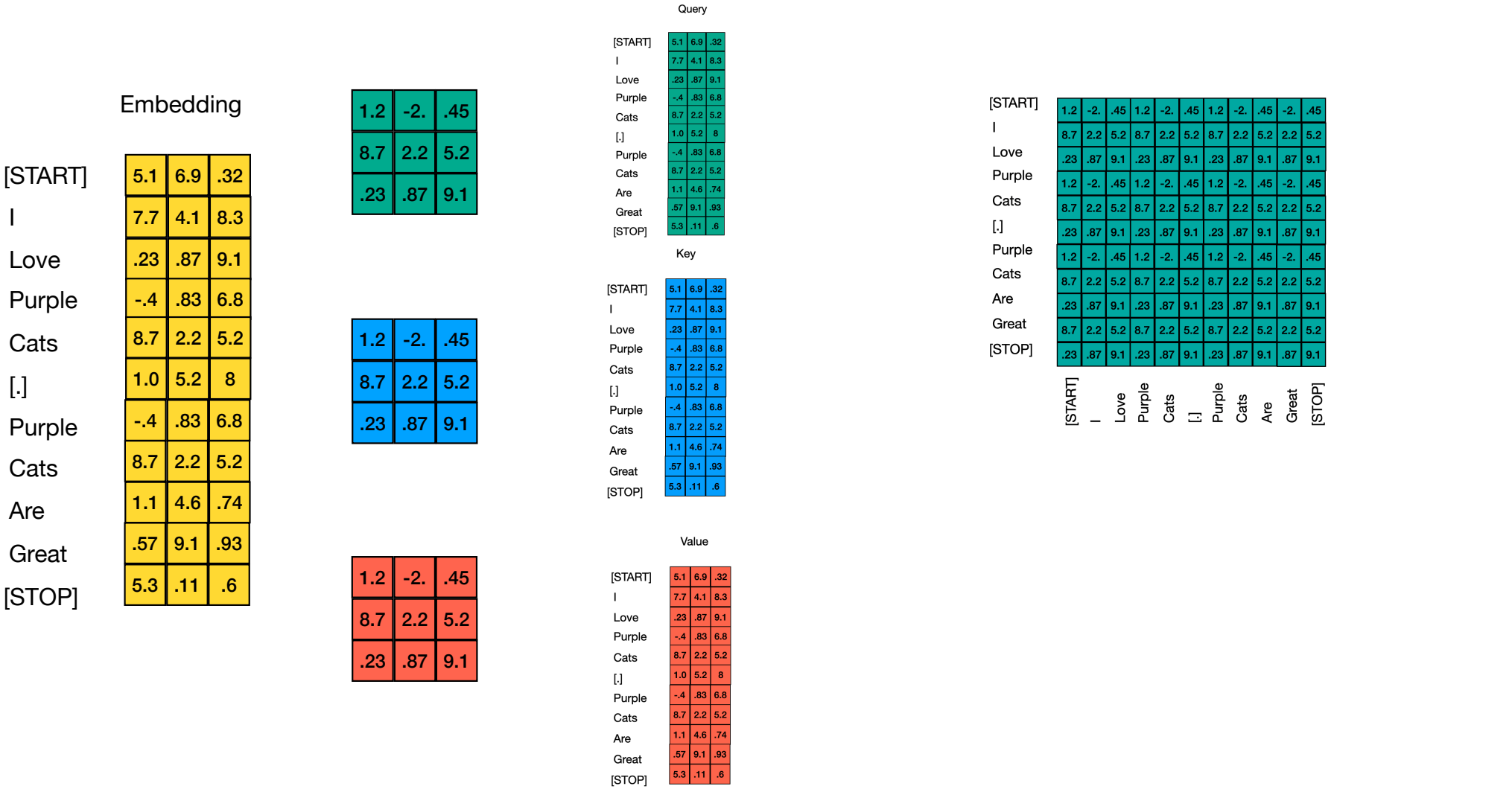

The
animal
didn't
cross
the
street
because
it
was
too
tired
.
The
animal
didn't
cross
the
street
because
it
was
too
tired
.

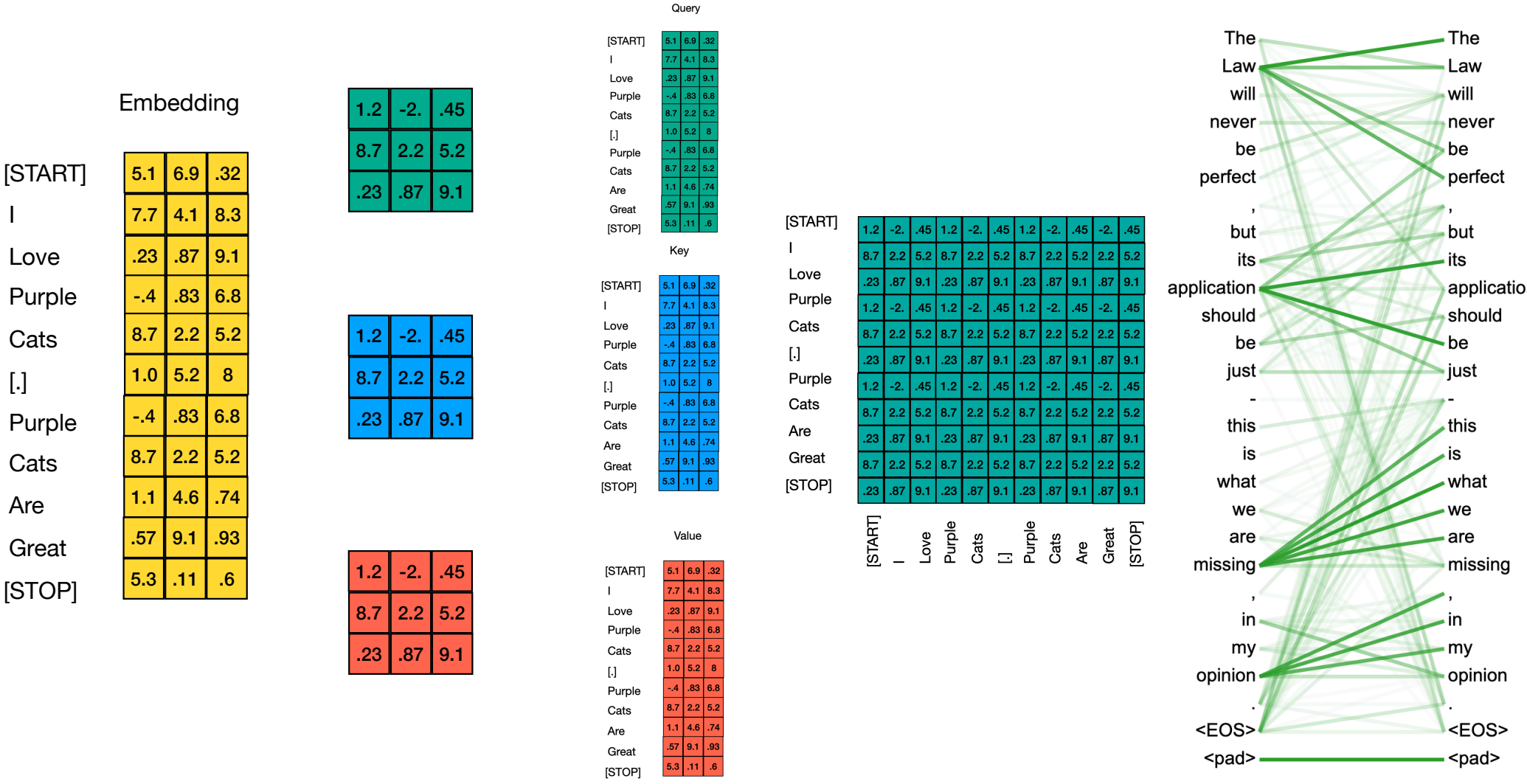

The
animal
didn't
cross
the
street
because
it
was
too
wide
.
The
animal
didn't
cross
the
street
because
it
was
too
wide
.

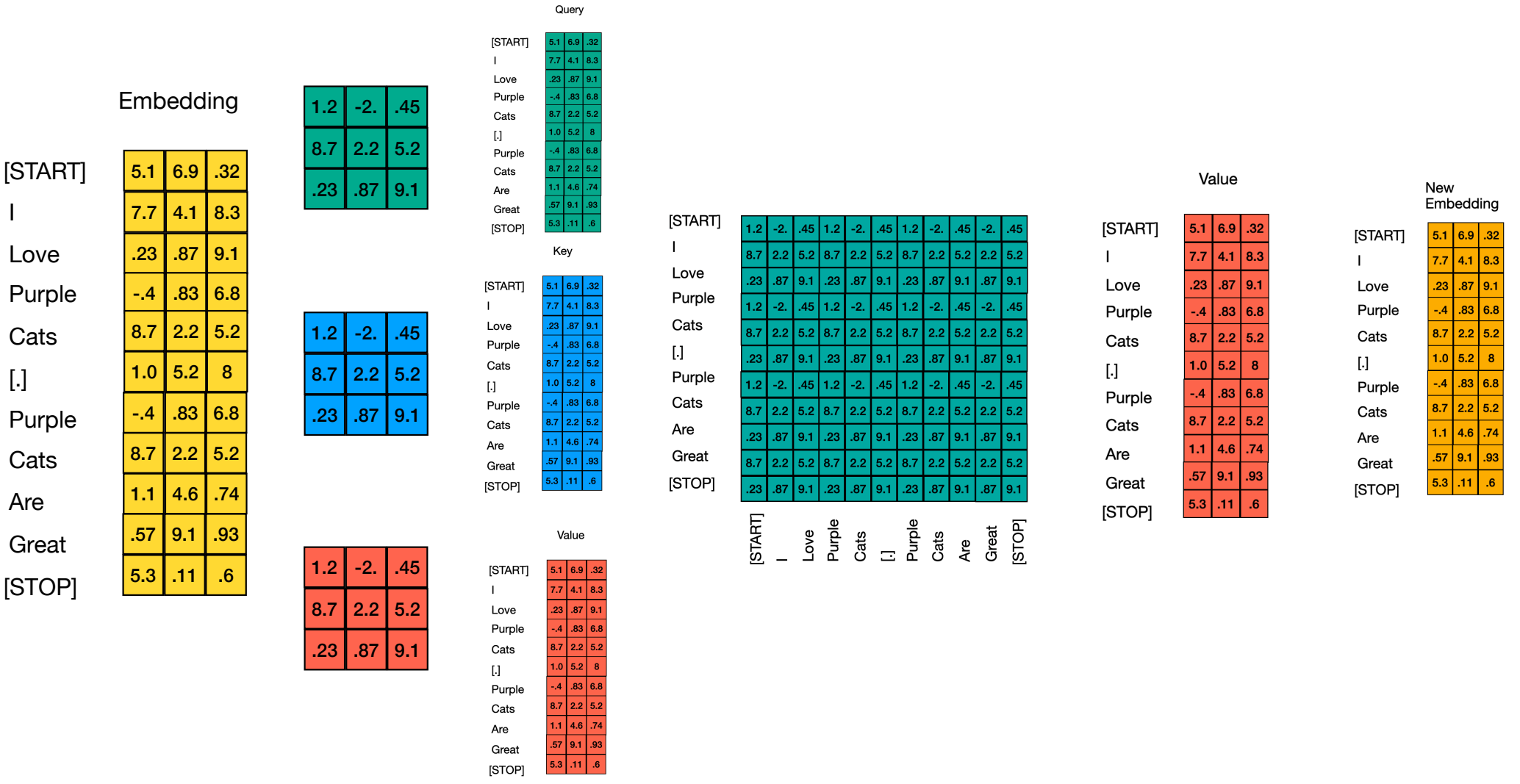
Attention


$\mathbf{v}_1$
$\mathbf{x}_1$
$\mathbf{v}_2$
$\mathbf{x}_2$
$\mathbf{v}_3$
$\mathbf{x}_3$
$\mathbf{v}_4$
$\mathbf{x}_4$
$\mathbf{x}_5$
ϕ_1
ϕ_2
ϕ_3
ϕ_4


Embedding
[START]
I
Love
Purple
Cats
[.]
Purple
Cats
Are
Great
[STOP]
5.1 6.9 .32
7.7 4.1 8.3
.23 .87 9.1
-.4 .83 6.8
8.7 2.2 5.2
1.0 5.2 8
-.4 .83 6.8
8.7 2.2 5.2
1.1 4.6 .74
.57 9.1 .93
5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
Key
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6

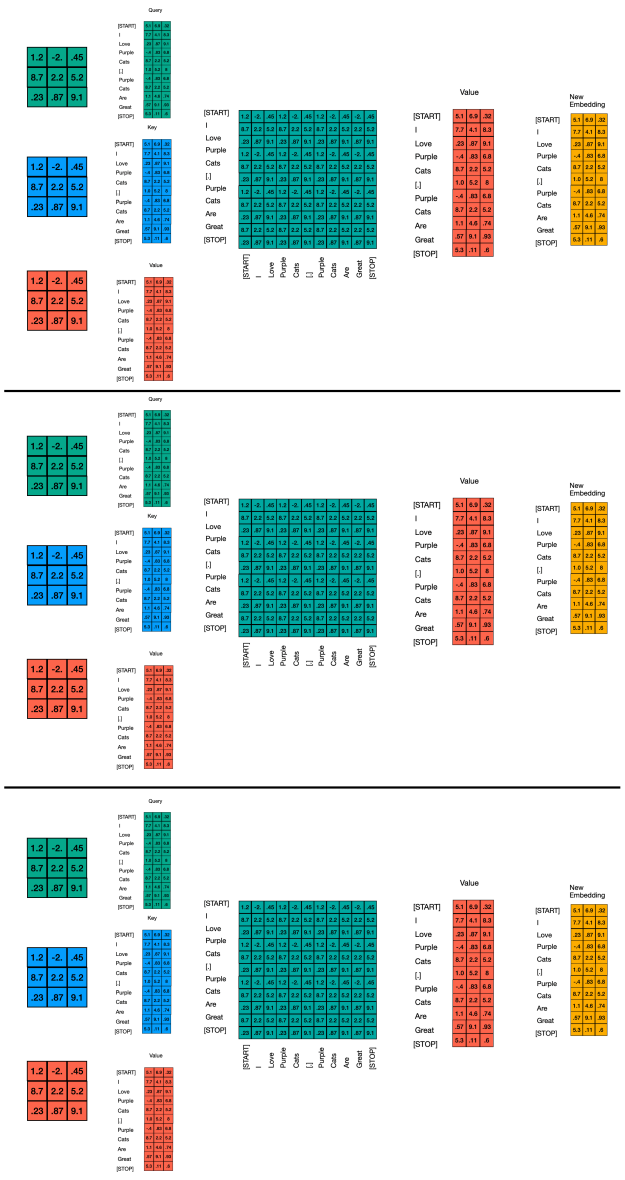

Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
Key
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
[START] 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
I 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Love .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[.] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Are .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Great 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[STOP] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]


Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
Key
Value
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]
The Law will never be perfect , but its application should be just - this is what we are missing , in my opinion . <EOS> <pad>


Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Key
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]
[START] 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
I 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Love .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[.] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Are .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Great 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[STOP] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
New Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6

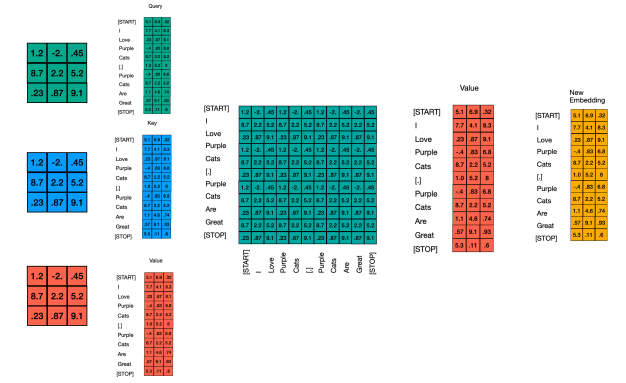
Embedding

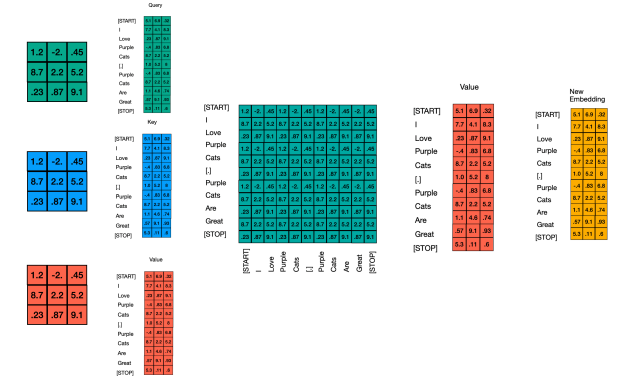
[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[.]	1.0	5.2	8
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
Are	1.1	4.6	.74
Great	.57	9.1	.93
[STOP]	5.3	.11	.6

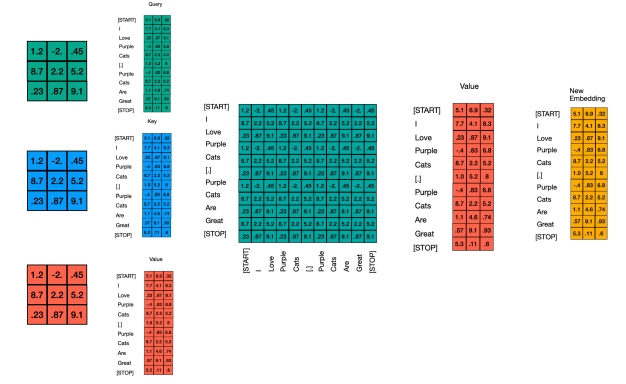


Embedding

[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[.]	1.0	5.2	8
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
Are	1.1	4.6	.74
Great	.57	9.1	.93
[STOP]	5.3	.11	.6

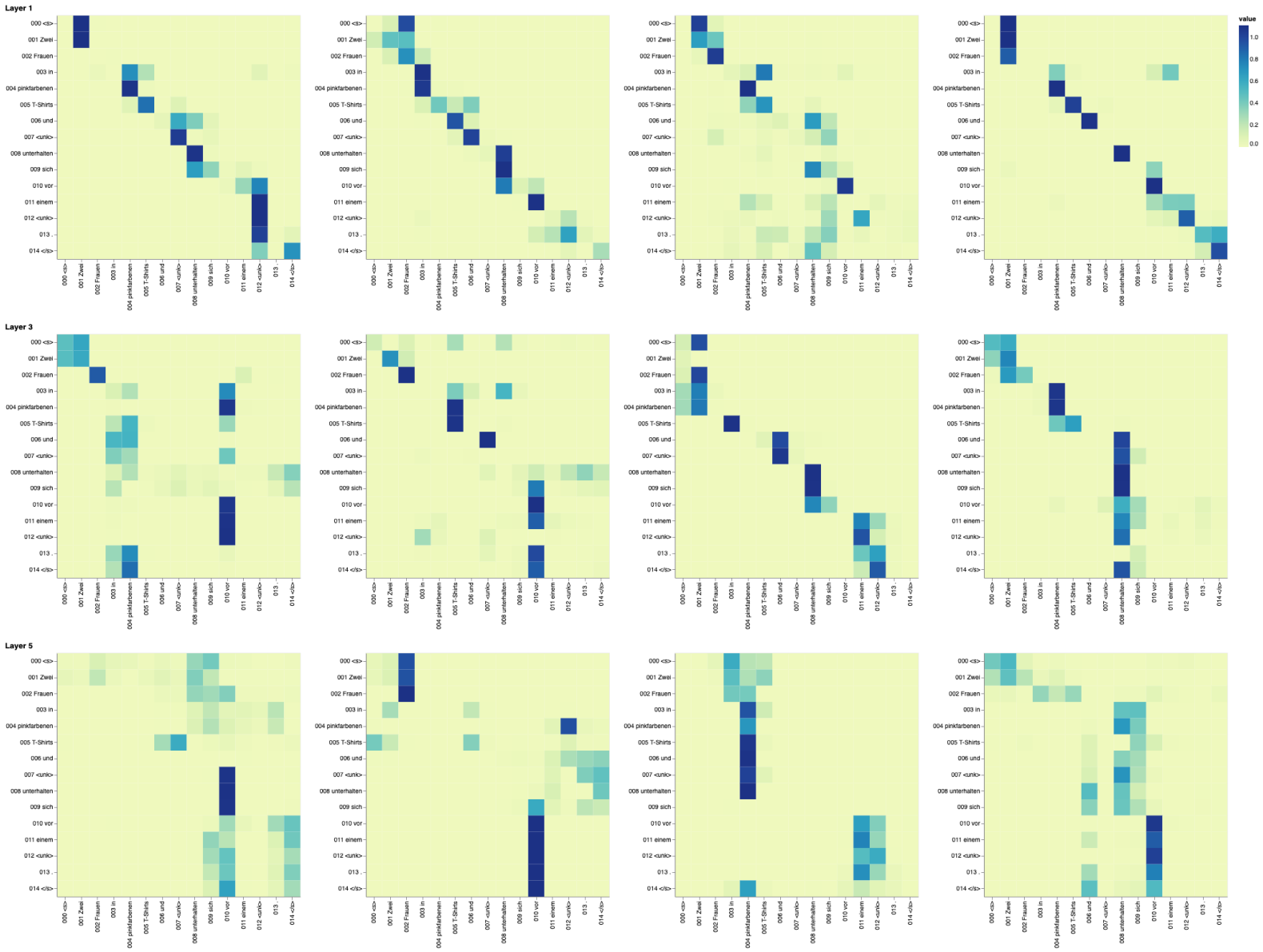


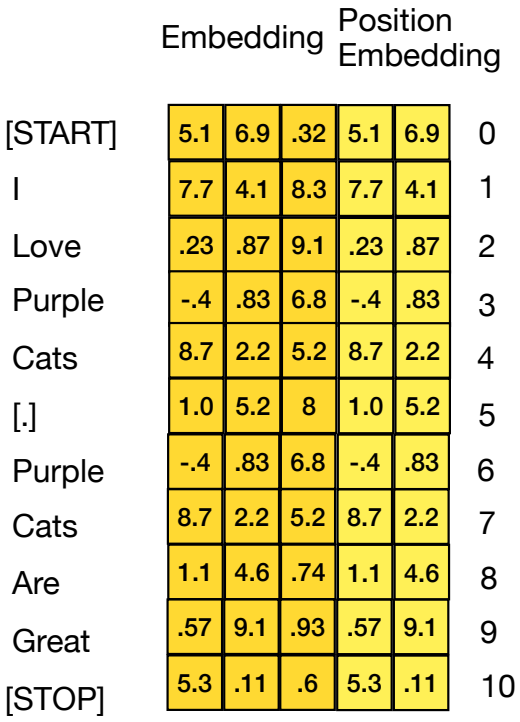


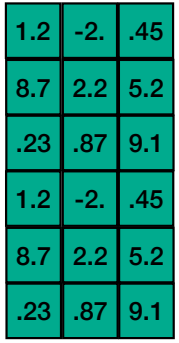


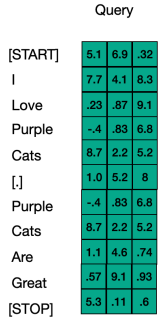
Concatenated Embeddings

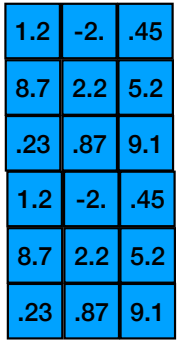
[START]	5.1	6.9	.32	5.1	6.9	.32	5.1	6.9	.32
I	7.7	4.1	8.3	7.7	4.1	8.3	7.7	4.1	8.3
Love	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1
Purple	-.4	.83	6.8	-.4	.83	6.8	-.4	.83	6.8
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2
[.]	1.0	5.2	8	1.0	5.2	8	1.0	5.2	8
Purple	-.4	.83	6.8	-.4	.83	6.8	-.4	.83	6.8
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2
Are	1.1	4.6	.74	1.1	4.6	.74	1.1	4.6	.74
Great	.57	9.1	.93	.57	9.1	.93	.57	9.1	.93
[STOP]	5.3	.11	.6	5.3	.11	.6	5.3	.11	.6

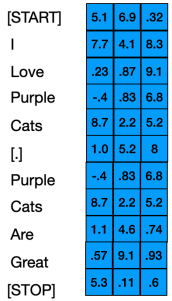

Layer 1
Layer 3
Layer 5
value
1.0
0.8
0.6
0.4
0.2
0.0
000 <s>
001 Zwei
002 Frauen
003 in
004 pinkfarbenen
005 T-Shirts
006 und
007 <unk>
008 unterhalten
009 sich
010 vor
011 einem
012 <unk>
013 .
014 </s>


Embedding
Position Embedding
[START] 5.1 6.9 .32 5.1 6.9 0
I 7.7 4.1 8.3 7.7 4.1 1
Love .23 .87 9.1 .23 .87 2
Purple -.4 .83 6.8 -.4 .83 3
Cats 8.7 2.2 5.2 8.7 2.2 4
[.] 1.0 5.2 8 1.0 5.2 5
Purple -.4 .83 6.8 -.4 .83 6
Cats 8.7 2.2 5.2 8.7 2.2 7
Are 1.1 4.6 .74 1.1 4.6 8
Great .57 9.1 .93 .57 9.1 9
[STOP] 5.3 .11 .6 5.3 .11 10

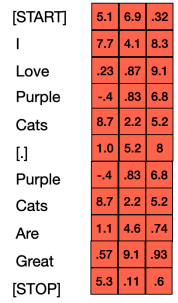

1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1

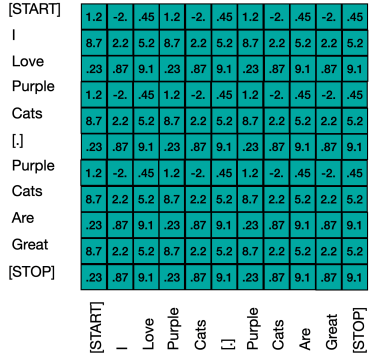

Query
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6

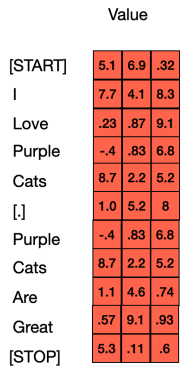

1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1

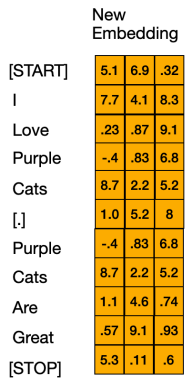

Key
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6


1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1


Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6


[START] 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
I 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Love .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[.] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Are .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Great 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[STOP] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]


Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6


New Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6

Position
Embedding

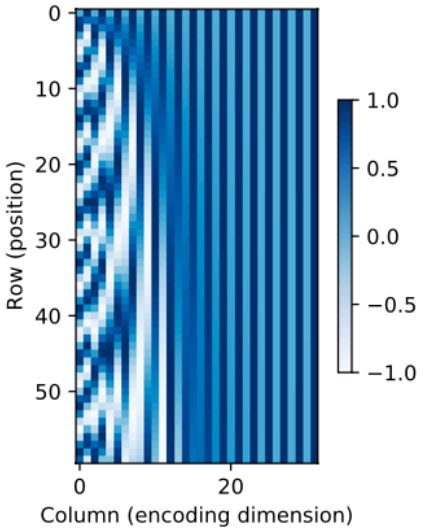
5.1	6.9	5.1	6.9	0
7.7	4.1	7.7	4.1	1
.23	.87	.23	.87	2
-.4	.83	-.4	.83	3
8.7	2.2	8.7	2.2	4
1.0	5.2	1.0	5.2	5
-.4	.83	-.4	.83	6
8.7	2.2	8.7	2.2	7
1.1	4.6	1.1	4.6	8
.57	9.1	.57	9.1	9
5.3	.11	5.3	.11	10
				i

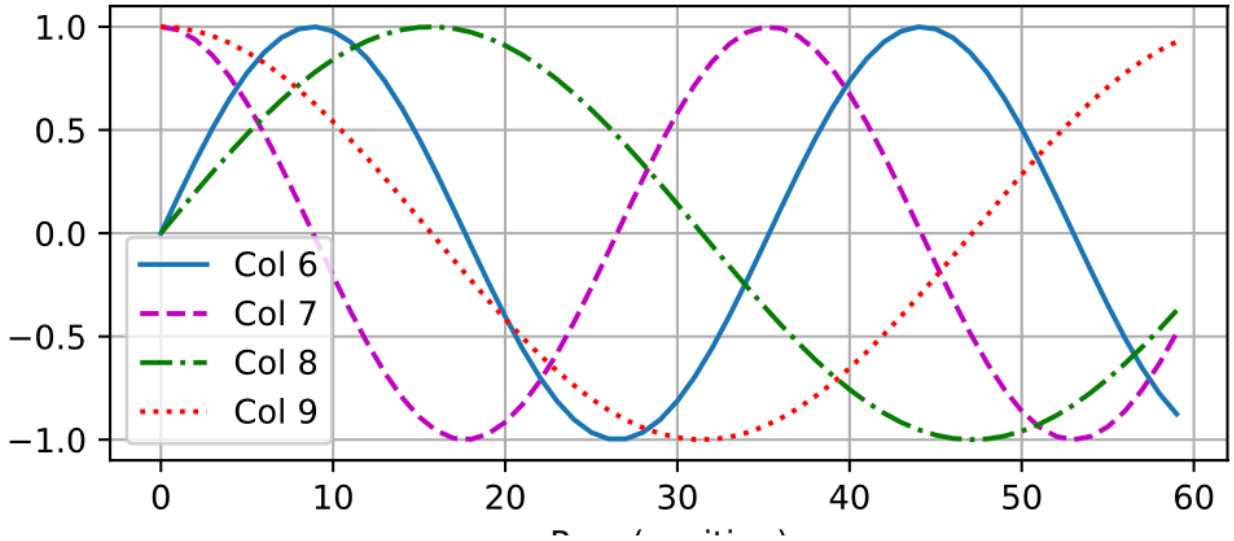
$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

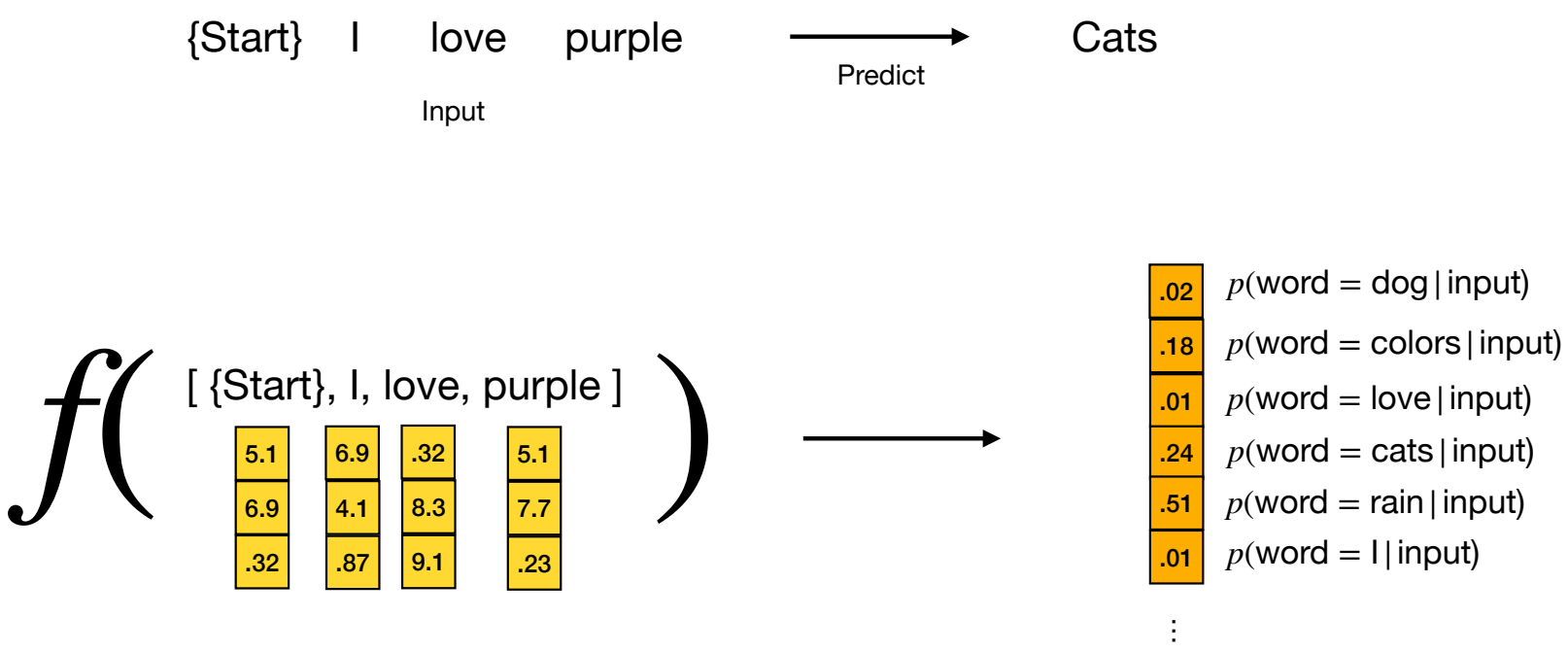
$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

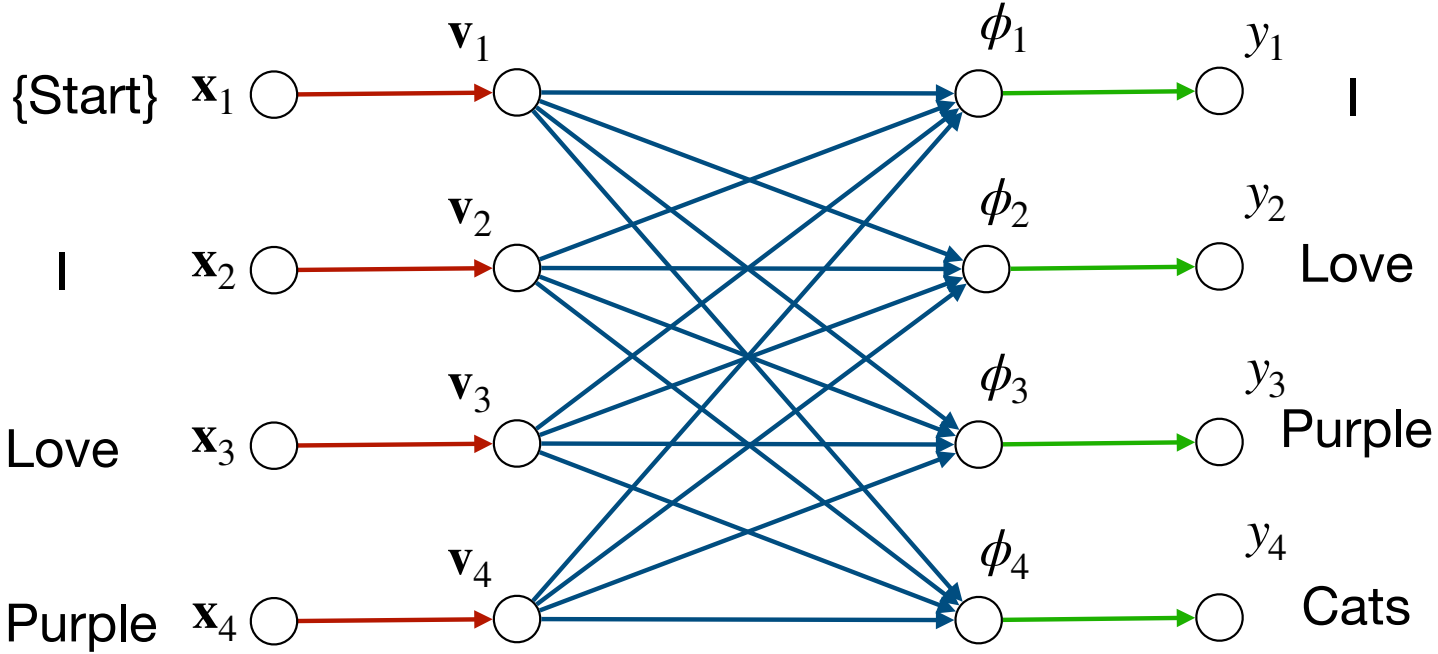


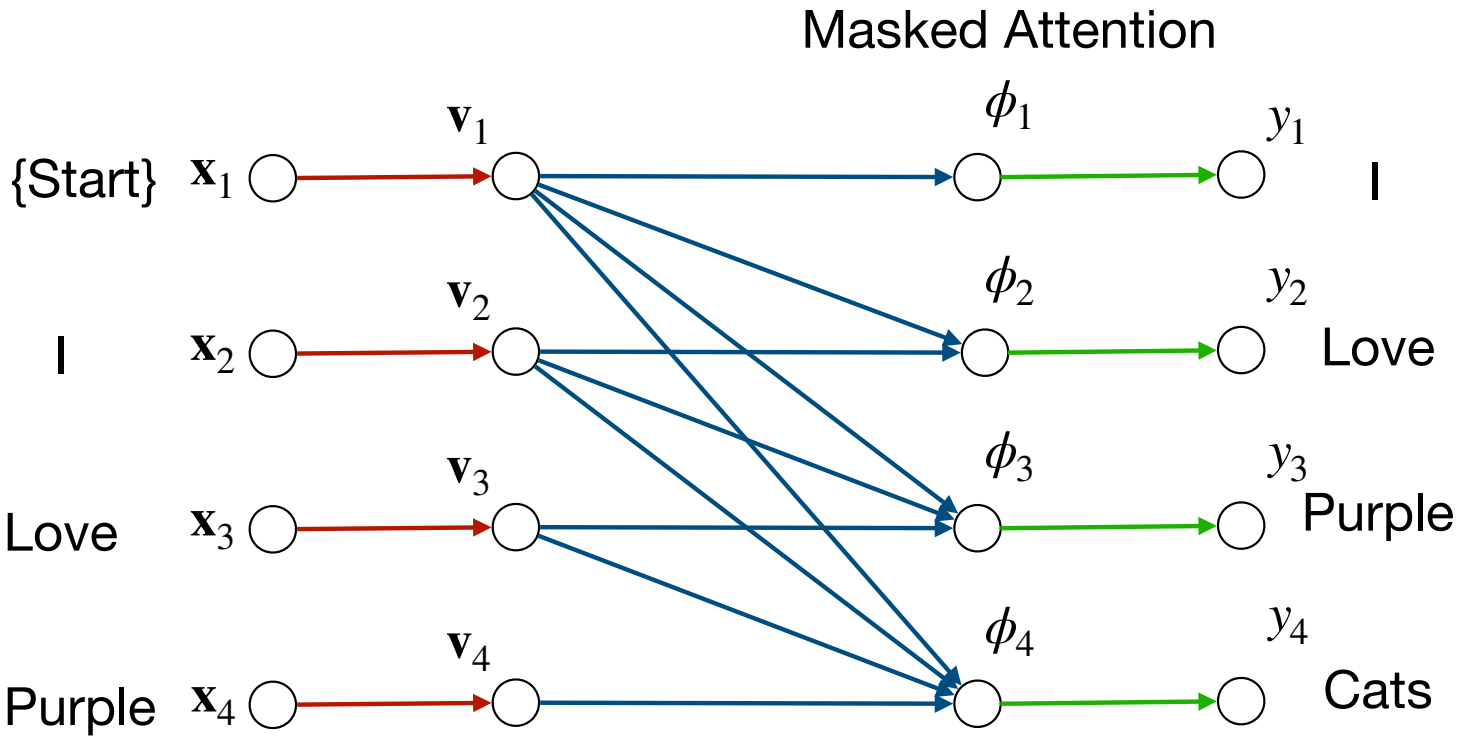


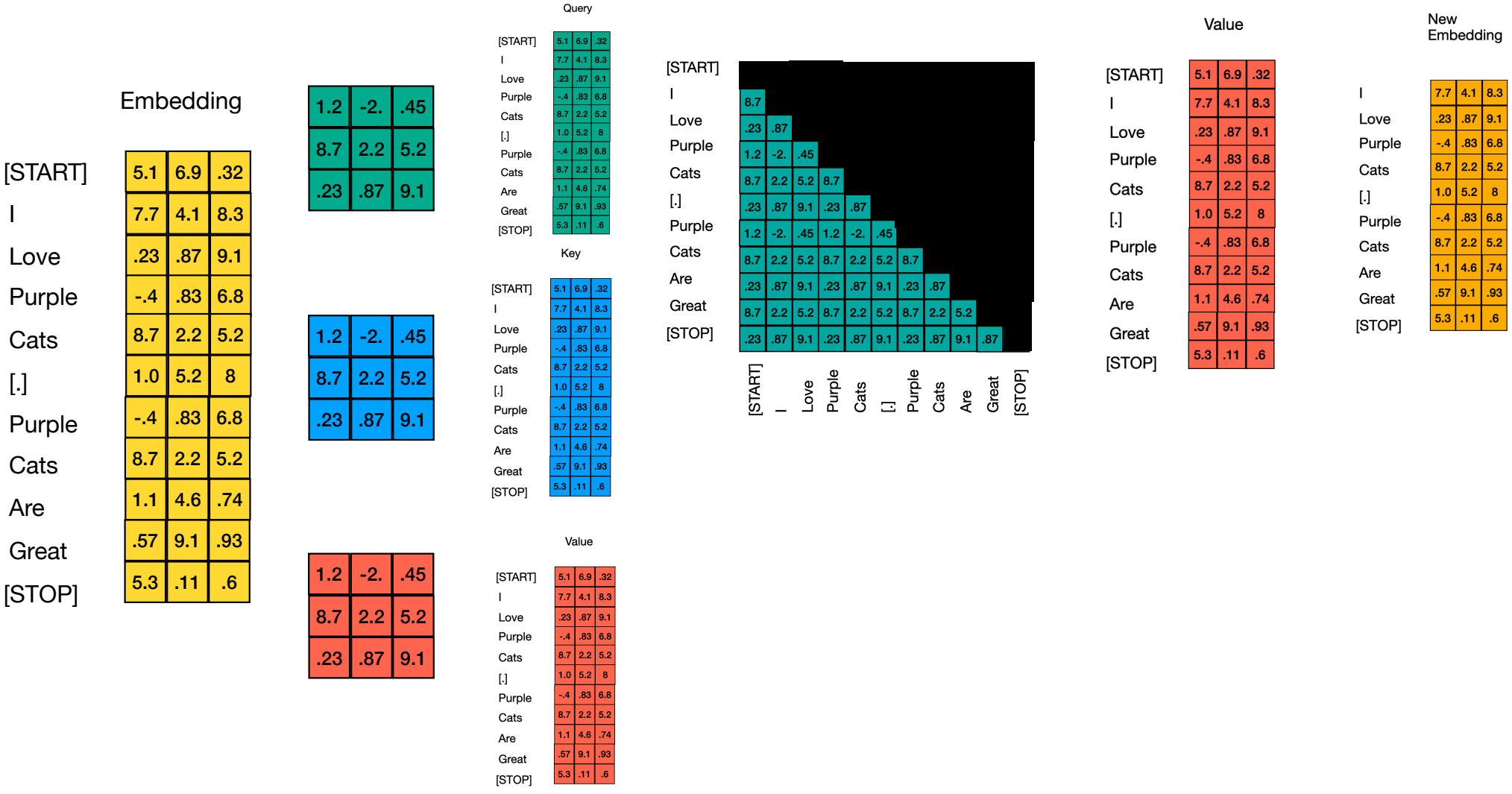
Autoregressive language modeling
(Next word prediction)



Loss:


{Start}
I
Love
Purple
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{v}_1$
$\mathbf{v}_2$
$\mathbf{v}_3$
$\mathbf{v}_4$
ϕ_1
ϕ_2
ϕ_3
ϕ_4
y_1
y_2
y_3
y_4
I
Love
Purple
Cats


Masked Attention
{Start}
I
Love
Purple
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_4$
$\mathbf{v}_1$
$\mathbf{v}_2$
$\mathbf{v}_3$
$\mathbf{v}_4$
ϕ_1
ϕ_2
ϕ_3
ϕ_4
y_1
y_2
y_3
y_4
I
Love
Purple
Cats


Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
Key
Value
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]
8.7
.23 .87
1.2 -2. .45
8.7 2.2 5.2 8.7
.23 .87 9.1 .23 .87
1.2 -2. .45 1.2 -2. .45
8.7 2.2 5.2 8.7 2.2 5.2 8.7
.23 .87 9.1 .23 .87 9.1 .23 .87
8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2
.23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
New Embedding
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6

WaveNet

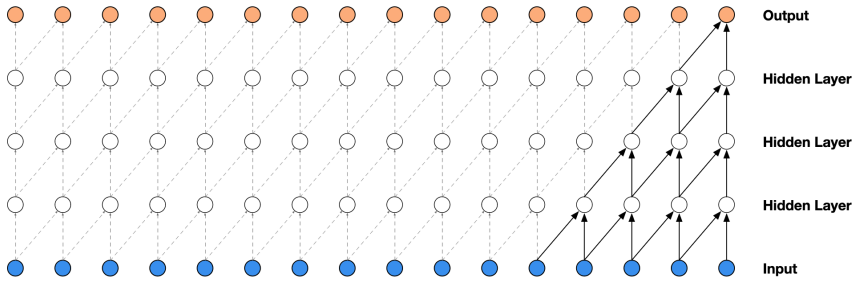
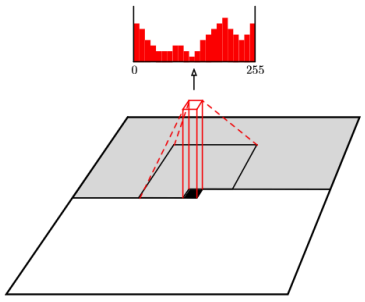
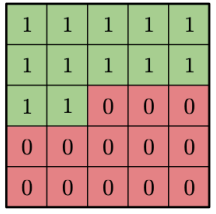
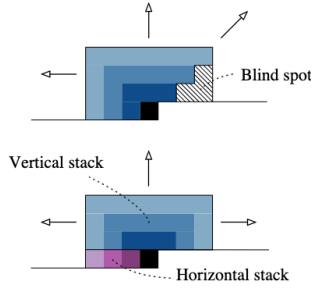


Figure 2: Visualization of a stack of causal convolutional layers.

PixelCNN







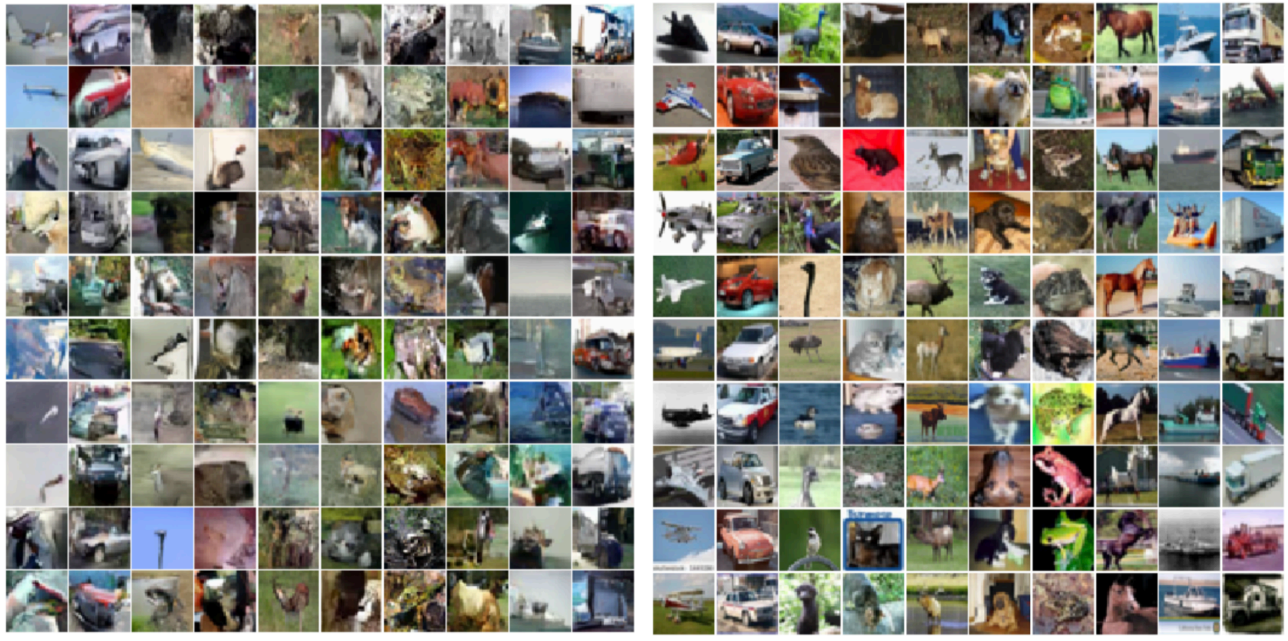
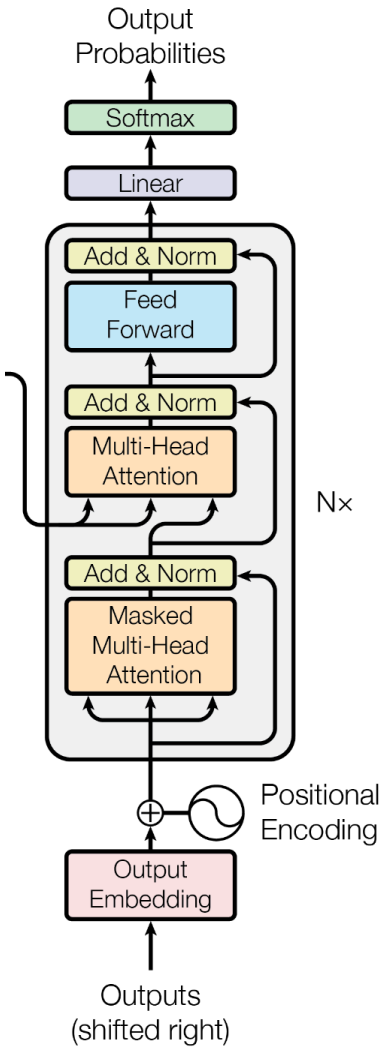
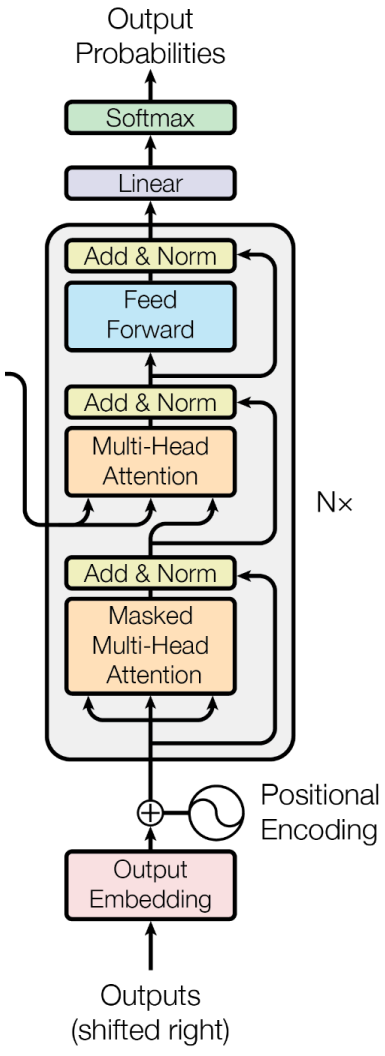


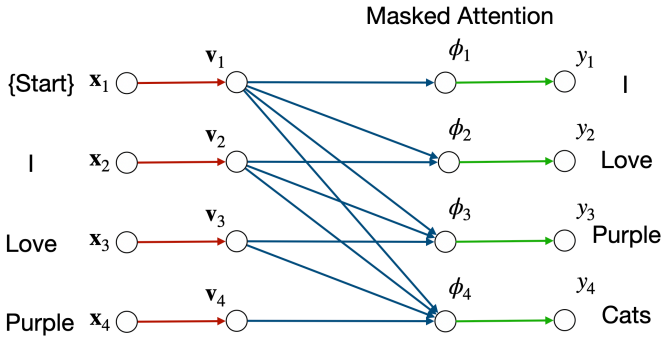
Figure 4: Class-conditional samples from our PixelCNN for CIFAR-10 (left) and real CIFAR-10 images for comparison (right).

I Love Purple Cats {Stop}



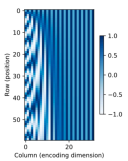
{Start} I love purple cats

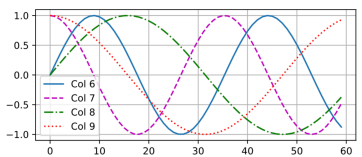


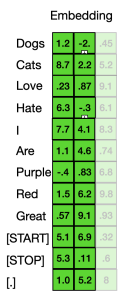


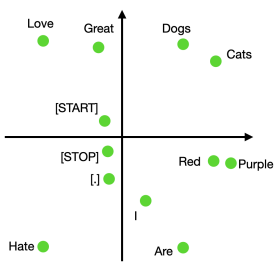
$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

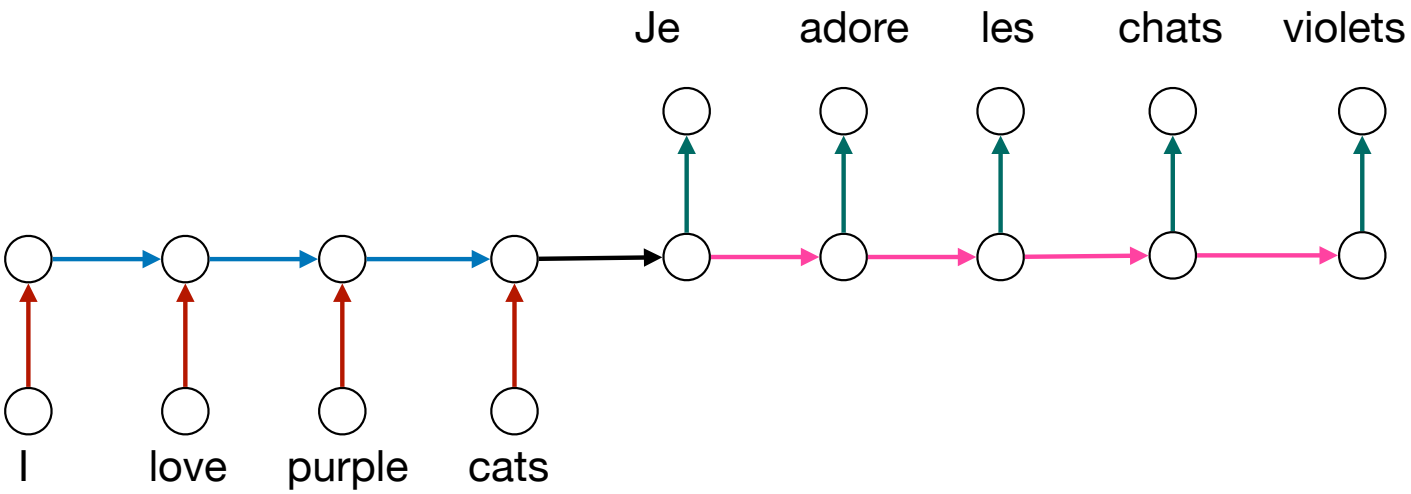
$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

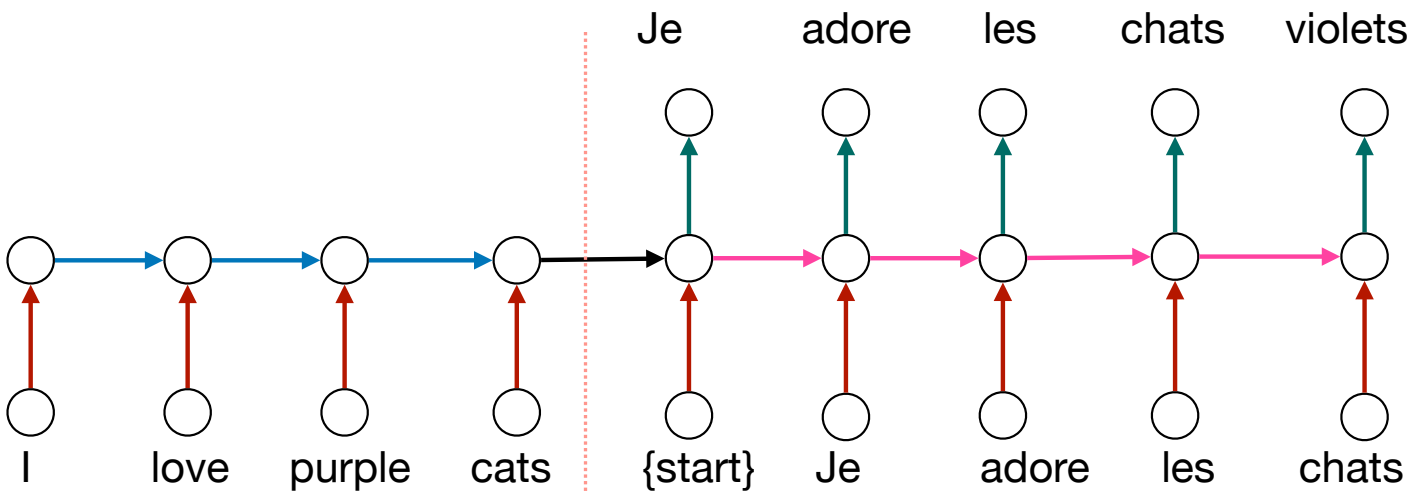




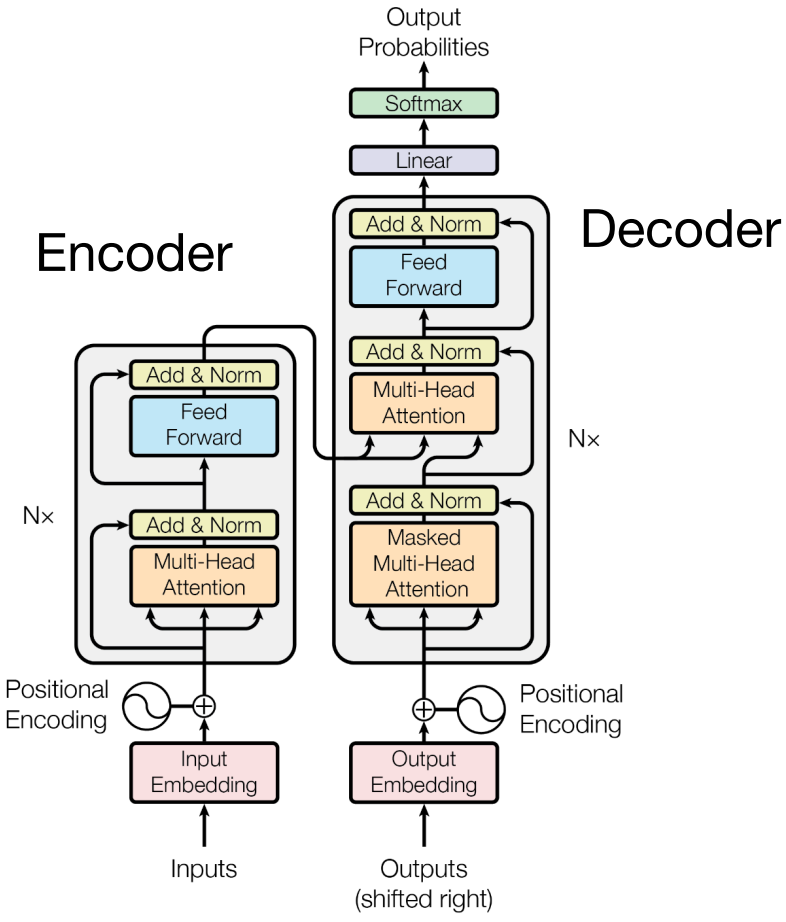





Je
adore
les
chats
violets
I
love
purple
cats


Je
adore
les
chats
violets
I
love
purple
cats
{start}
Je
adore
les
chats

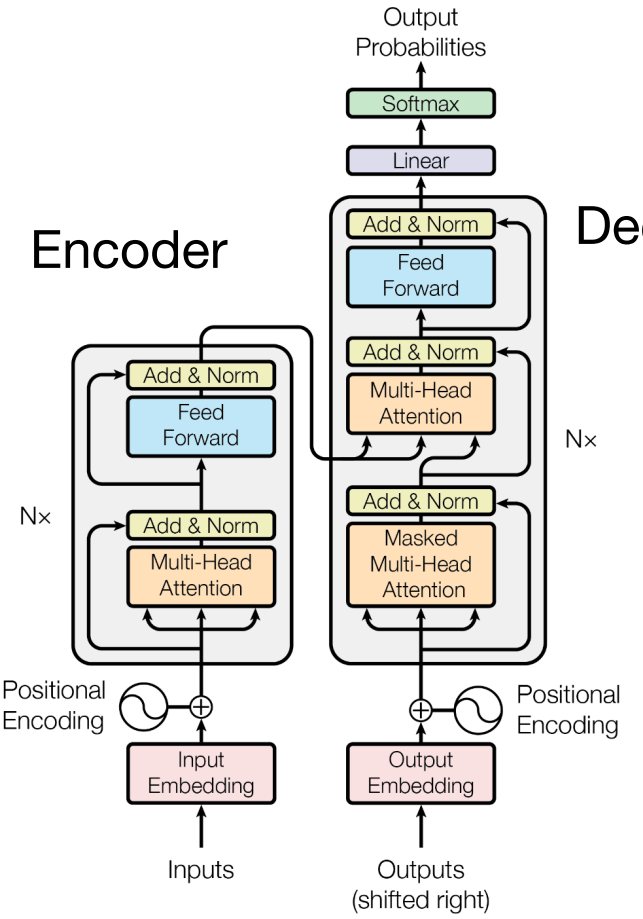
Je adore les chats violets {stop}



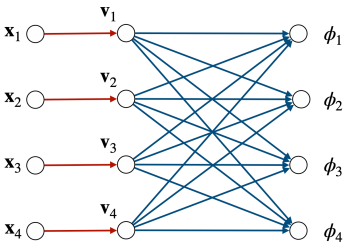
{start} I love purple cats {stop}

{start} Je adore les chats violets

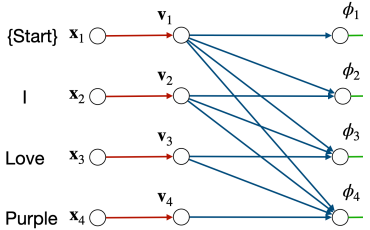
Je adore les chats violets {stop}



Self attention



Masked attention



{start} I love purple cats {stop}

{start} Je adore les chats violets

Decoder Embedding

[START]	5.1	6.9	.32
Je	7.7	4.1	8.3
Adore	.23	.87	9.1
Les	-.4	.83	6.8
Chats	8.7	2.2	5.2
Violet	1.0	5.2	8
[Stop]	-.4	.83	6.8

Encoder Embedding

[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[Stop]	1.0	5.2	8

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

Query

[START]	5.1	6.9	.32
Je	7.7	4.1	8.3
Adore	.23	.87	9.1
Les	-.4	.83	6.8
Chats	8.7	2.2	5.2
Violet	1.0	5.2	8
[Stop]	-.4	.83	6.8

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

Key

[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[Stop]	1.0	5.2	8

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

Value

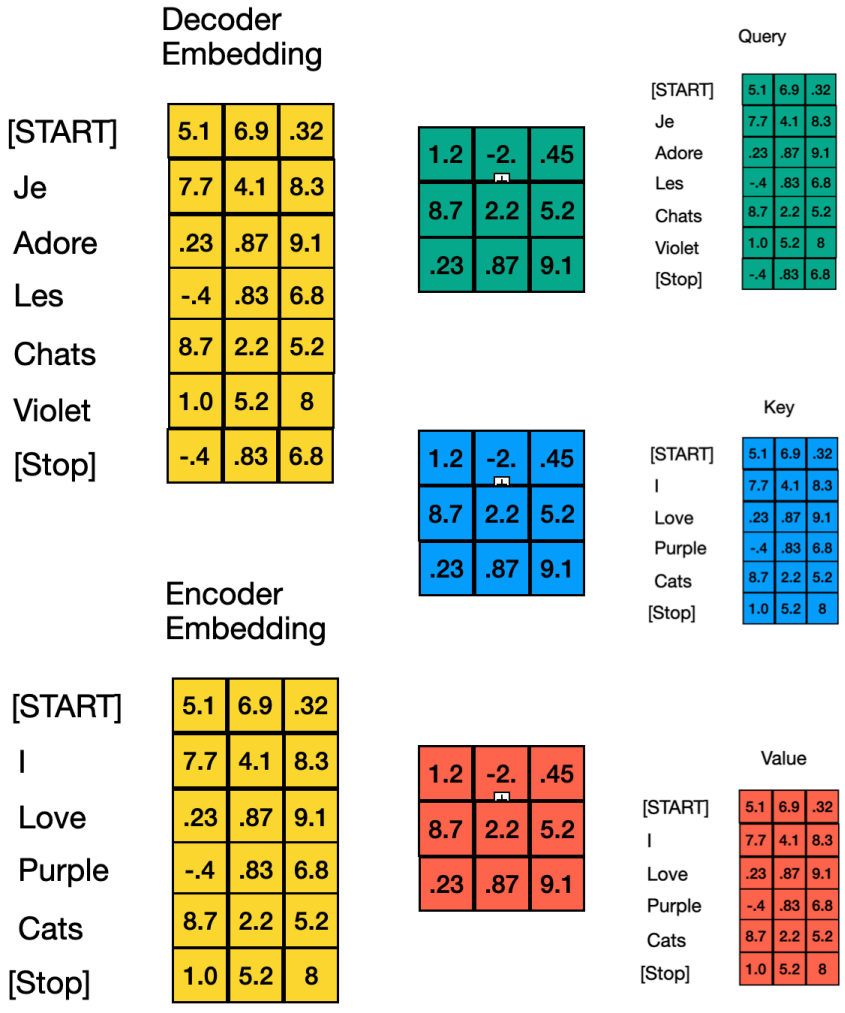
[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[Stop]	1.0	5.2	8

[START]	1.2	-2.	.45	1.2	-2.	.45
Je	8.7	2.2	5.2	8.7	2.2	5.2
Adore	.23	.87	9.1	.23	.87	9.1
Les	1.2	-2.	.45	1.2	-2.	.45
Chats	8.7	2.2	5.2	8.7	2.2	5.2
Violet	.23	.87	9.1	.23	.87	9.1
[STOP]	1.2	-2.	.45	1.2	-2.	.45

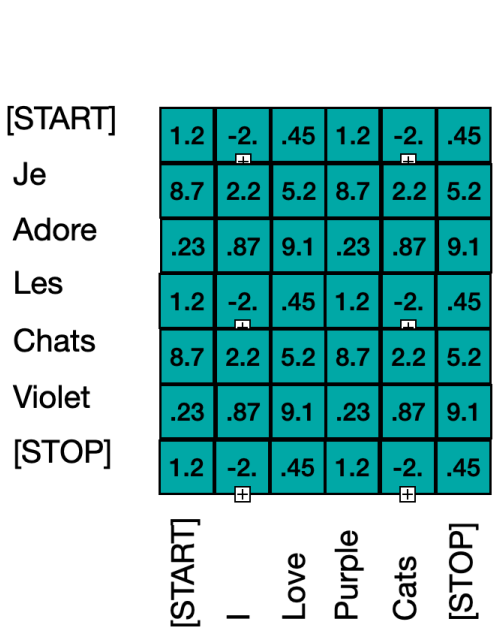
[START] I Love Purple Cats [STOP]

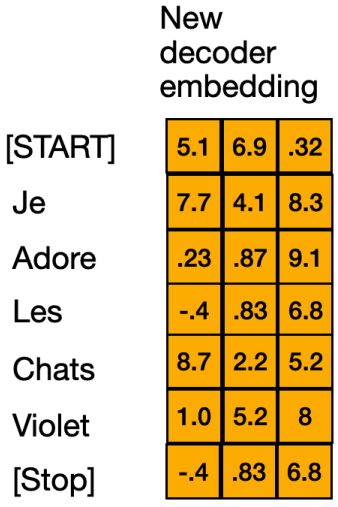
New decoder embedding

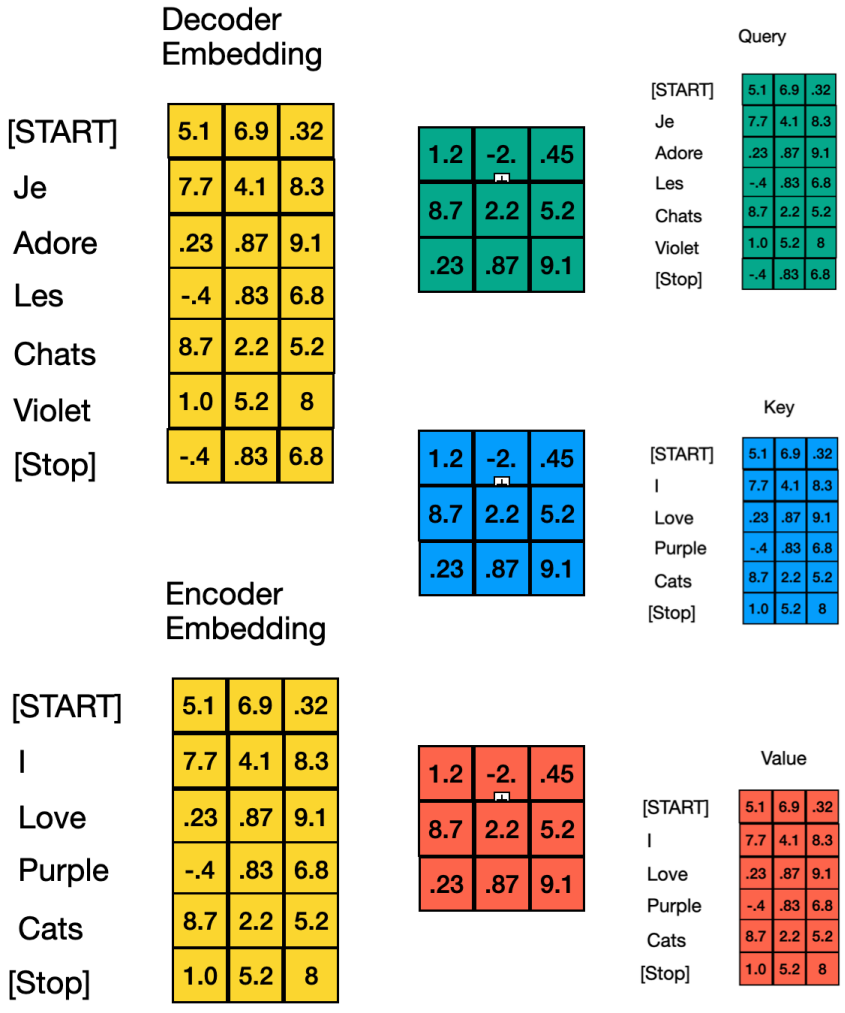
[START]	5.1	6.9	.32
Je	7.7	4.1	8.3
Adore	.23	.87	9.1
Les	-.4	.83	6.8
Chats	8.7	2.2	5.2
Violet	1.0	5.2	8
[Stop]	-.4	.83	6.8



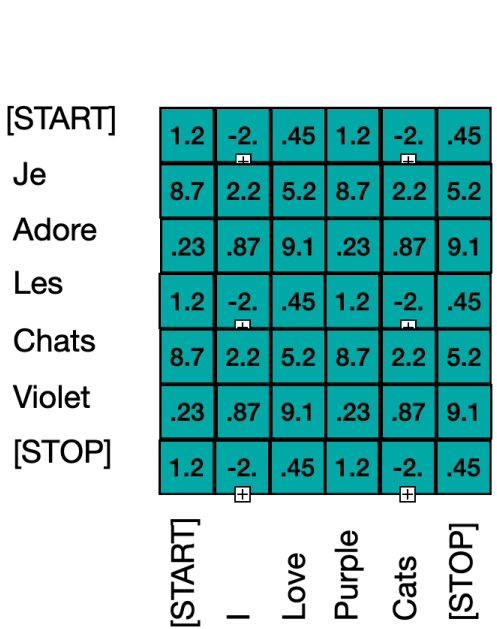

```
def attention(query, key, value, mask=None, dropout=None):
    "Compute 'Scaled Dot Product Attention'"
    d_k = query.size(-1)
    scores = torch.matmul(query, key.transpose(-2, -1)) / math.sqrt(d_k)
    if mask is not None:
        scores = scores.masked_fill(mask == 0, -1e9)
    p_attn = scores.softmax(dim=-1)
    if dropout is not None:
        p_attn = dropout(p_attn)
    return torch.matmul(p_attn, value), p_attn
```

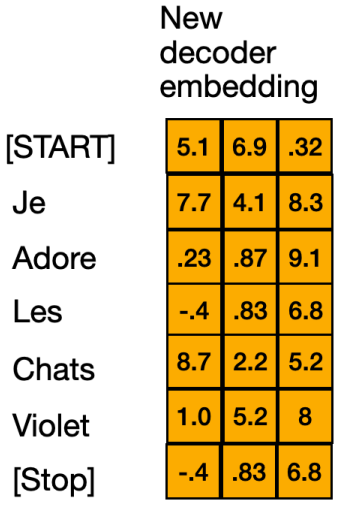


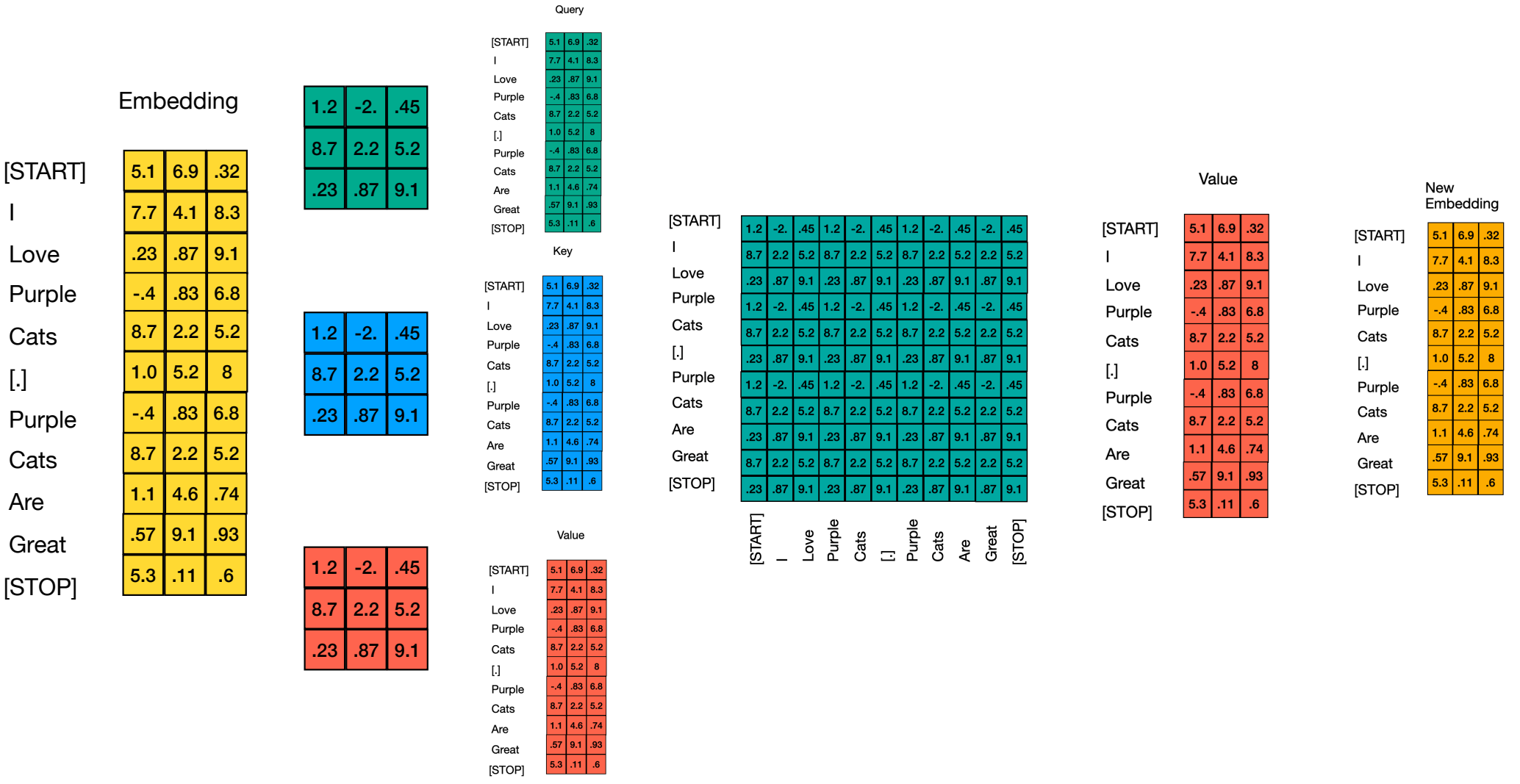


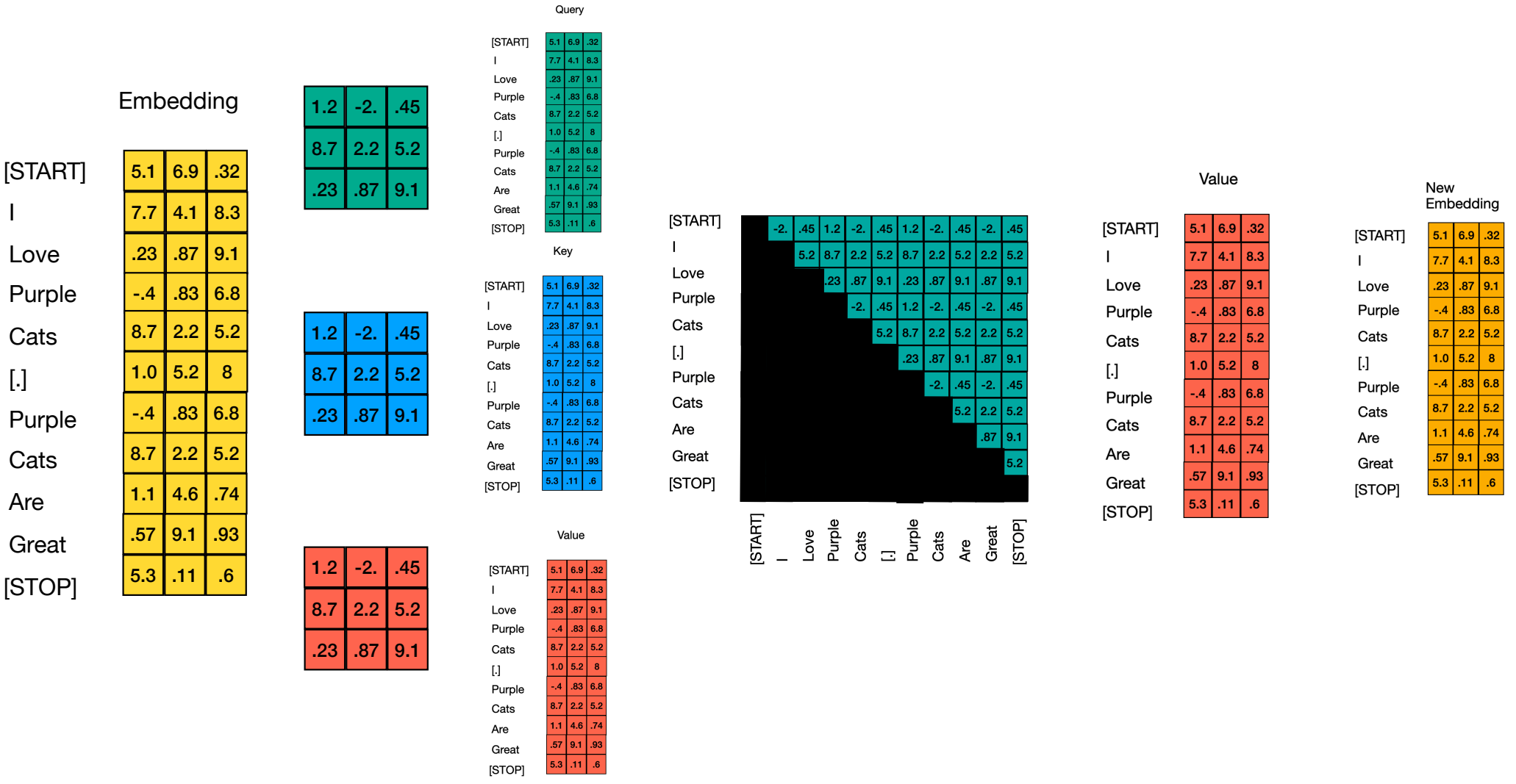



```
def attention(query, key, value, mask=None, dropout=None):
    "Compute 'Scaled Dot Product Attention'"
    d_k = query.size(-1)
    scores = torch.matmul(query, key.transpose(-2, -1)) / math.sqrt(d_k)
    if mask is not None:
        scores = scores.masked_fill(mask == 0, -1e9)
    p_attn = scores.softmax(dim=-1)
    if dropout is not None:
        p_attn = dropout(p_attn)
    return torch.matmul(p_attn, value), p_attn
```






Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Key
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]
[START] 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
I 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Love .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[.] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Are .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Great 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[STOP] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Value
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
New Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6


Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1
Query
Key
Value
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]
-2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
.23 .87 9.1 .23 .87 9.1 .87 9.1
-2. .45 1.2 -2. .45 -2. .45
5.2 8.7 2.2 5.2 2.2 5.2
.23 .87 9.1 .87 9.1
-2. .45 -2. .45
5.2 2.2 5.2
.87 9.1
5.2
Value
New Embedding

Embedding

[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[.]	1.0	5.2	8
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
Are	1.1	4.6	.74
Great	.57	9.1	.93
[STOP]	5.3	.11	.6

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

Query

Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

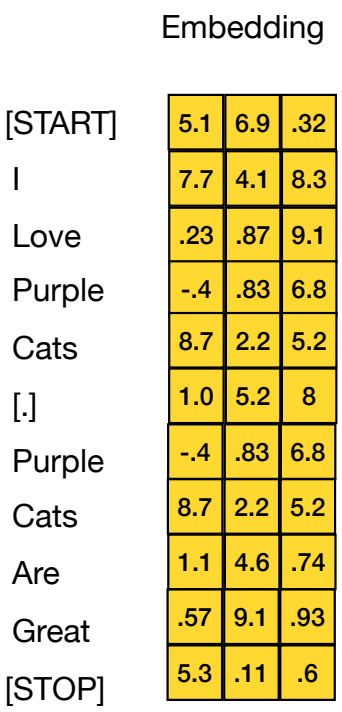
Key

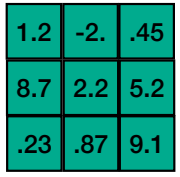
Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

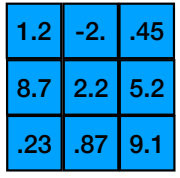
Value

Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

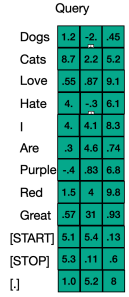
	[START]	I	Love	Purple	Cats	[.]	Purple	Cats	Are	Great	[STOP]
[START]	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
I	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
Love	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Purple	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
[.]	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Purple	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
Are	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Great	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
[STOP]	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1

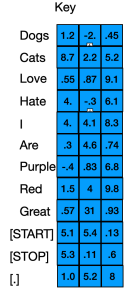

Embedding
[START] 5.1 6.9 .32
I 7.7 4.1 8.3
Love .23 .87 9.1
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
[.] 1.0 5.2 8
Purple -.4 .83 6.8
Cats 8.7 2.2 5.2
Are 1.1 4.6 .74
Great .57 9.1 .93
[STOP] 5.3 .11 .6


1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1

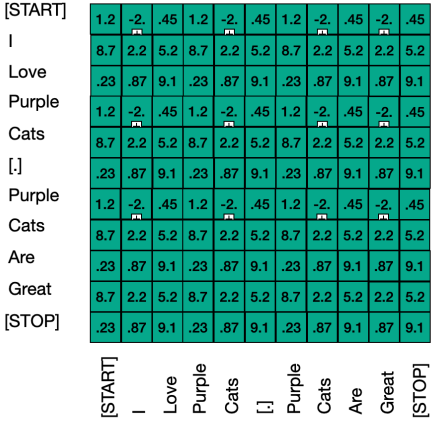

1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1


1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1


Query
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


Key
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


Value
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


[START] 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
I 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Love .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[.] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Purple 1.2 -2. .45 1.2 -2. .45 1.2 -2. .45 -2. .45
Cats 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
Are .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
Great 8.7 2.2 5.2 8.7 2.2 5.2 8.7 2.2 5.2 2.2 5.2
[STOP] .23 .87 9.1 .23 .87 9.1 .23 .87 9.1 .87 9.1
[START] I Love Purple Cats [.] Purple Cats Are Great [STOP]


Value
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8

Embedding

[START]	5.1	6.9	.32
I	7.7	4.1	8.3
Love	.23	.87	9.1
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
[.]	1.0	5.2	8
Purple	-.4	.83	6.8
Cats	8.7	2.2	5.2
Are	1.1	4.6	.74
Great	.57	9.1	.93
[STOP]	5.3	.11	.6

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

1.2	-2.	.45
8.7	2.2	5.2
.23	.87	9.1

Query

Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

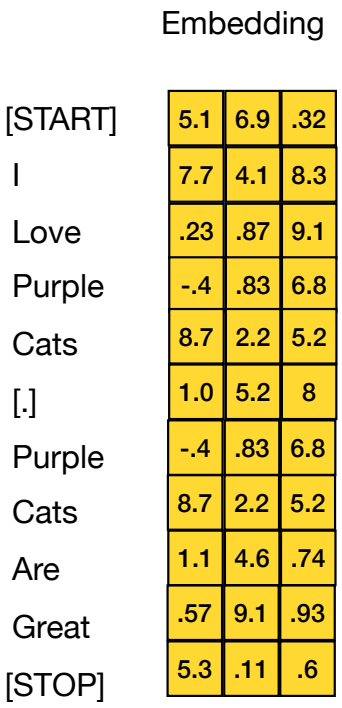
Key

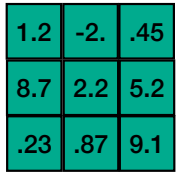
Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

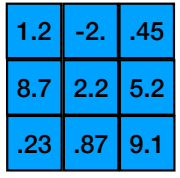
Value

Dogs	1.2	-2.	.45
Cats	8.7	2.2	5.2
Love	.55	.87	9.1
Hate	4.	-.3	6.1
I	4.	4.1	8.3
Are	.3	4.6	.74
Purple	-.4	.83	6.8
Red	1.5	4	9.8
Great	.57	31	.93
[START]	5.1	5.4	.13
[STOP]	5.3	.11	.6
[.]	1.0	5.2	8

	[START]	I	Love	Purple	Cats	[.]	Purple	Cats	Are	Great	[STOP]
[START]	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
I	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
Love	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Purple	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
[.]	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Purple	1.2	-2.	.45	1.2	-2.	.45	1.2	-2.	.45	-2.	.45
Cats	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
Are	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1
Great	8.7	2.2	5.2	8.7	2.2	5.2	8.7	2.2	5.2	2.2	5.2
[STOP]	.23	.87	9.1	.23	.87	9.1	.23	.87	9.1	.87	9.1

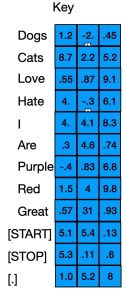

Embedding
[START]
I
Love
Purple
Cats
[.]
Purple
Cats
Are
Great
[STOP]
5.1 6.9 .32
7.7 4.1 8.3
.23 .87 9.1
-.4 .83 6.8
8.7 2.2 5.2
1.0 5.2 8
-.4 .83 6.8
8.7 2.2 5.2
1.1 4.6 .74
.57 9.1 .93
5.3 .11 .6


1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1

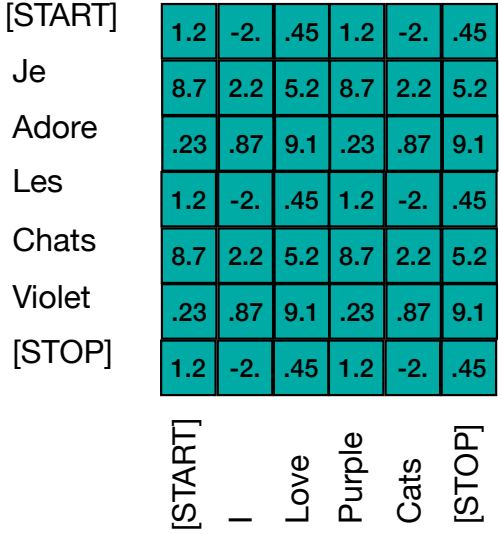

1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1


1.2 -2. .45
8.7 2.2 5.2
.23 .87 9.1


Query
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


Key
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


Value
Dogs 1.2 -2. .45
Cats 8.7 2.2 5.2
Love .55 .87 9.1
Hate 4. -.3 6.1
I 4. 4.1 8.3
Are .3 4.6 .74
Purple -.4 .83 6.8
Red 1.5 4 9.8
Great .57 31 .93
[START] 5.1 5.4 .13
[STOP] 5.3 .11 .6
[.] 1.0 5.2 8


[START]
Je
Adore
Les
Chats
Violet
[STOP]
1.2 -2. .45 1.2 -2. .45
8.7 2.2 5.2 8.7 2.2 5.2
.23 .87 9.1 .23 .87 9.1
1.2 -2. .45 1.2 -2. .45
8.7 2.2 5.2 8.7 2.2 5.2
.23 .87 9.1 .23 .87 9.1
1.2 -2. .45 1.2 -2. .45
[START]
I
Love
Purple
Cats
[STOP]