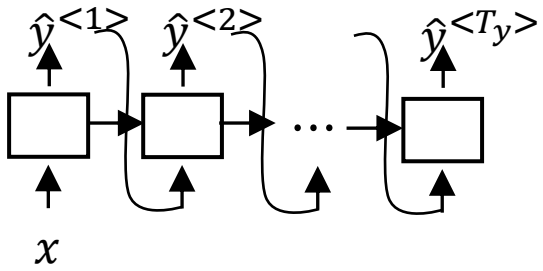
Transformers Intuition

Transformers Motivation

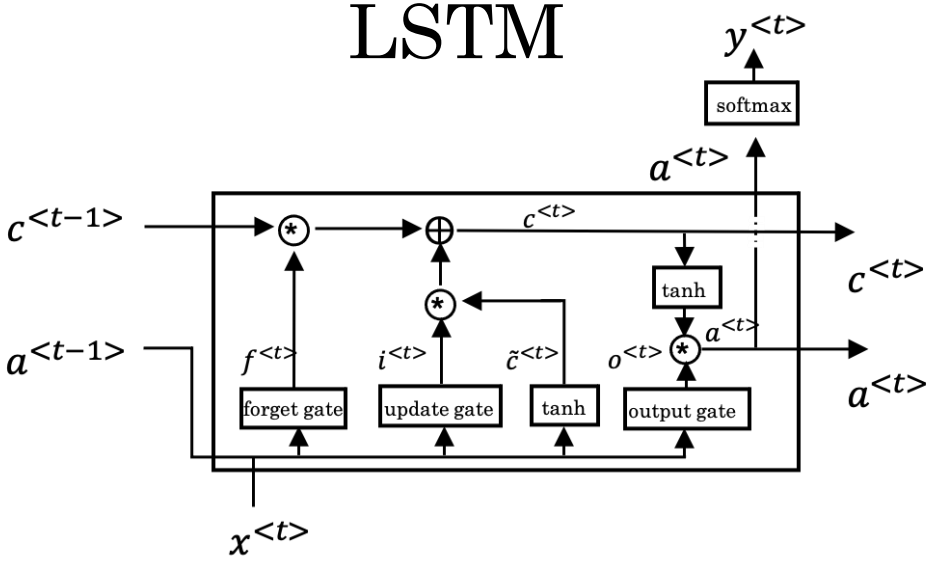
Increased complexity, sequential

RNN



GRU

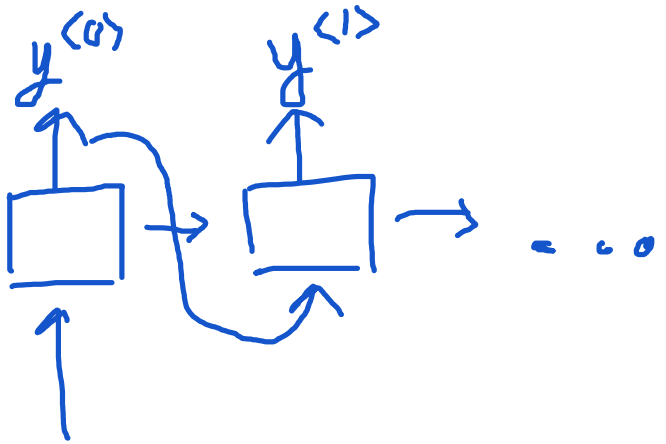
LSTM

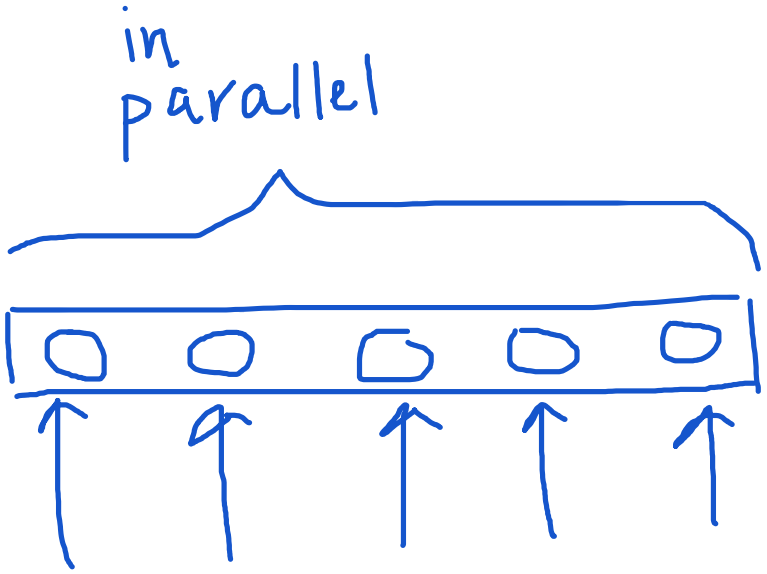


Andrew Ng

Transformers Intuition

- Attention + CNN
  - Self-Attention
  - Multi-Head Attention





[Vaswani et al. 2017, Attention Is All You Need]

Self-Attention

Self-Attention Intuition

$A(q, K, V)$= attention-based vector representation of a word

→ calculate for each word

Transformers Attention

$$A(q, K, V) = \sum_i \frac{\exp(e^{<q \cdot k^{<i>}>})}{\sum_j \exp(e^{<q \cdot k^{<j>}>})} v^{<i>}$$

$x^{<1>}$	$x^{<2>}$	$x^{<3>}$	$x^{<4>}$	$x^{<5>}$
Jane	visite	l'Afrique	en	septembre

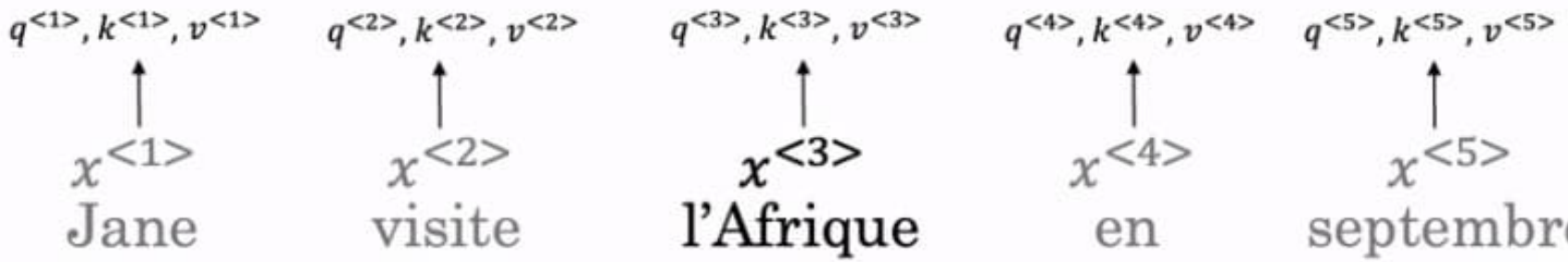
[Vaswani et al. 2017, Attention Is All You Need]

Self-Attention

$$A(q,K,V) = \sum_i \frac{\exp(q \cdot k^{<i>})}{\sum_j \exp(q \cdot k^{<j>})} v^{<i>}$$

$A^{<3>}$

Query (Q)	Key (K)	Value (V)
$q^{<1>}$	$k^{<1>}$	$v^{<1>}$
$q^{<2>}$	$k^{<2>}$	$v^{<2>}$
$q^{<3>}$	$k^{<3>}$	$v^{<3>}$
$q^{<4>}$	$k^{<4>}$	$v^{<4>}$
$q^{<5>}$	$k^{<5>}$	$v^{<5>}$



$q^{<3>} = W^Q \cdot x^{<3>}$
$k^{<3>} = W^K \cdot x^{<3>}$
$v^{<3>} = W^V \cdot x^{<3>}$

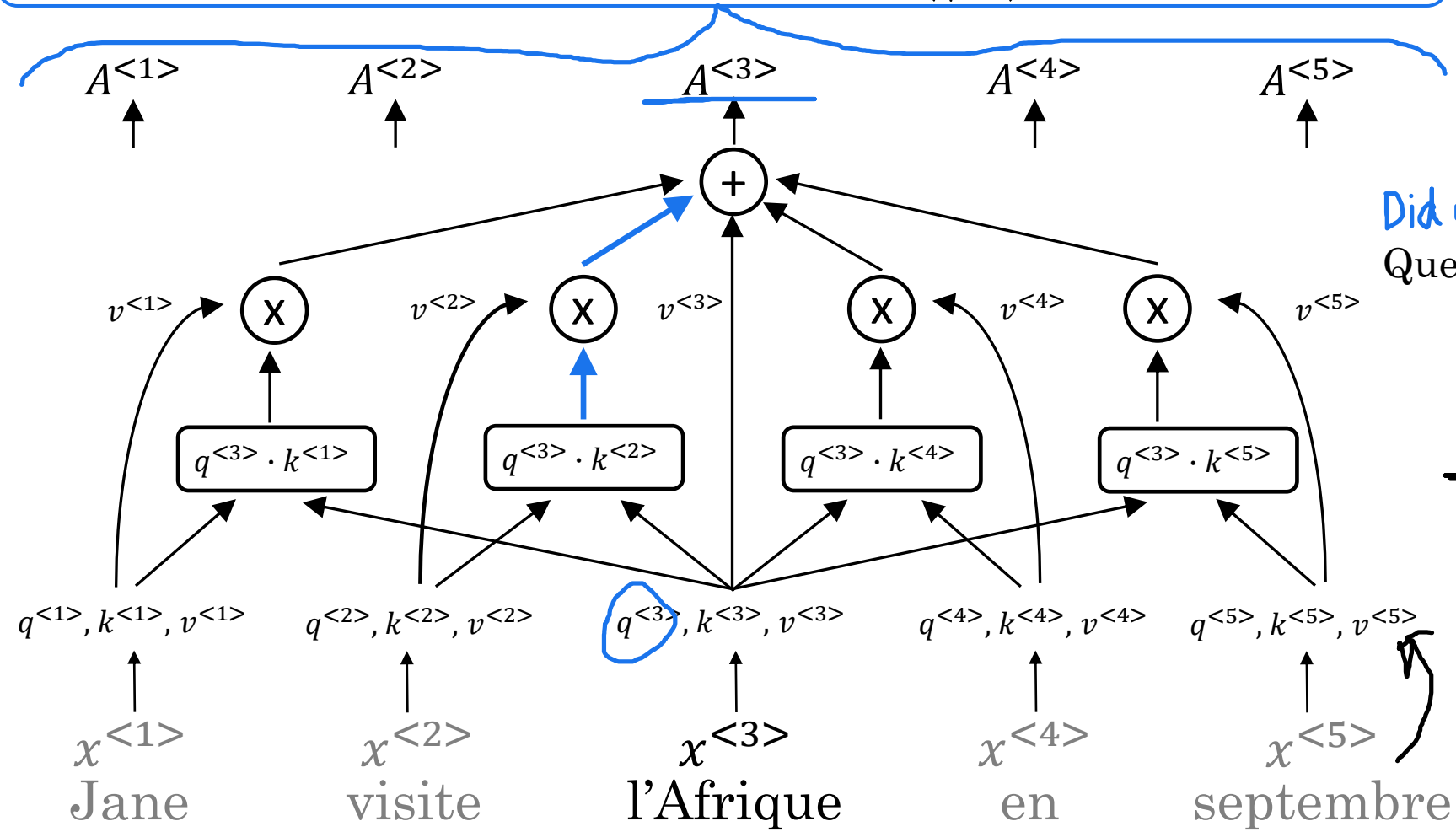
[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng

Self-Attention

$$A(q, K, V) = \sum_i \frac{\exp(e^{<q \cdot k^{<i>}>})}{\sum_j \exp(e^{<q \cdot k^{<j>}>})} v^{<i>}$$

$$\boldsymbol{Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V}$$



Did what?

Query (Q)	Key (K)	Value (V)
$q^{<1>}$	$k^{<1>}$	$v^{<1>}$
$q^{<2>}$	$k^{<2>}$	$v^{<2>}$
$q^{<3>}$	$k^{<3>}$	$v^{<3>}$
$q^{<4>}$	$k^{<4>}$	$v^{<4>}$
$q^{<5>}$	$k^{<5>}$	$v^{<5>}$

What's happening there?

person action

Jane visit

W^Q, W^K, W^V

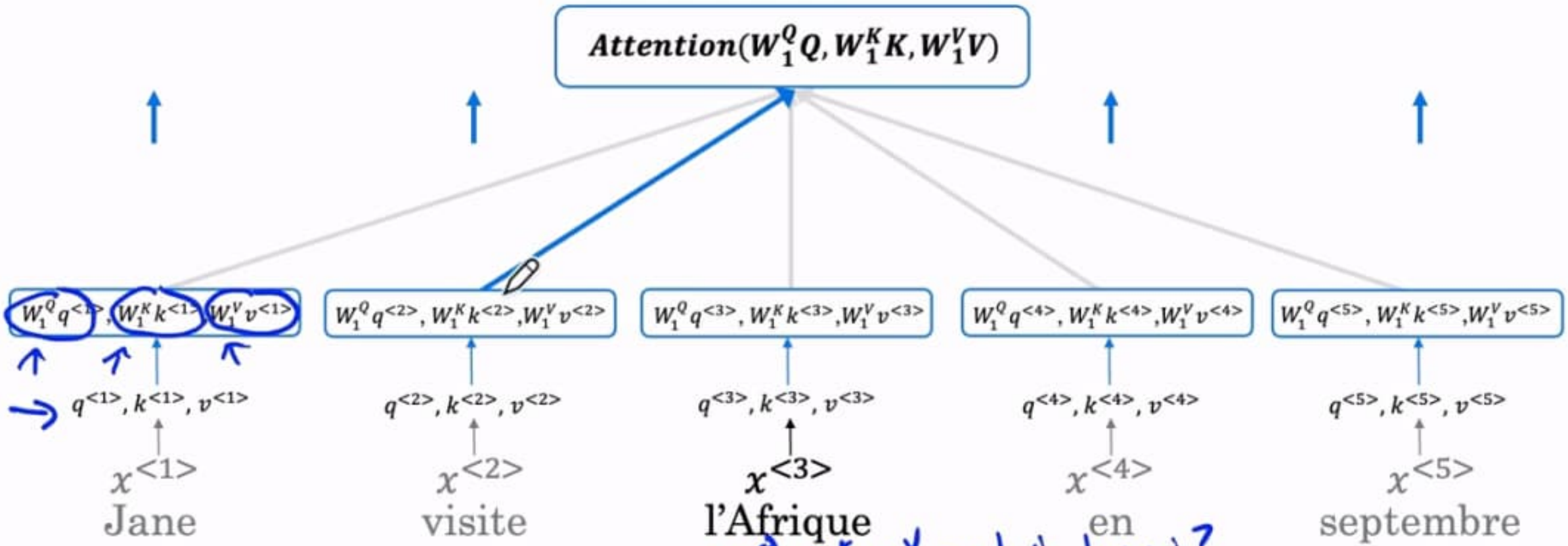
[Vaswani et al. 2017, Attention Is All You Need]

Multi-Head Attention

Multi-Head Attention

"head"

$MultiHead(Q, K, V)$



[Vaswani et al. 2017, Attention Is All You Need]

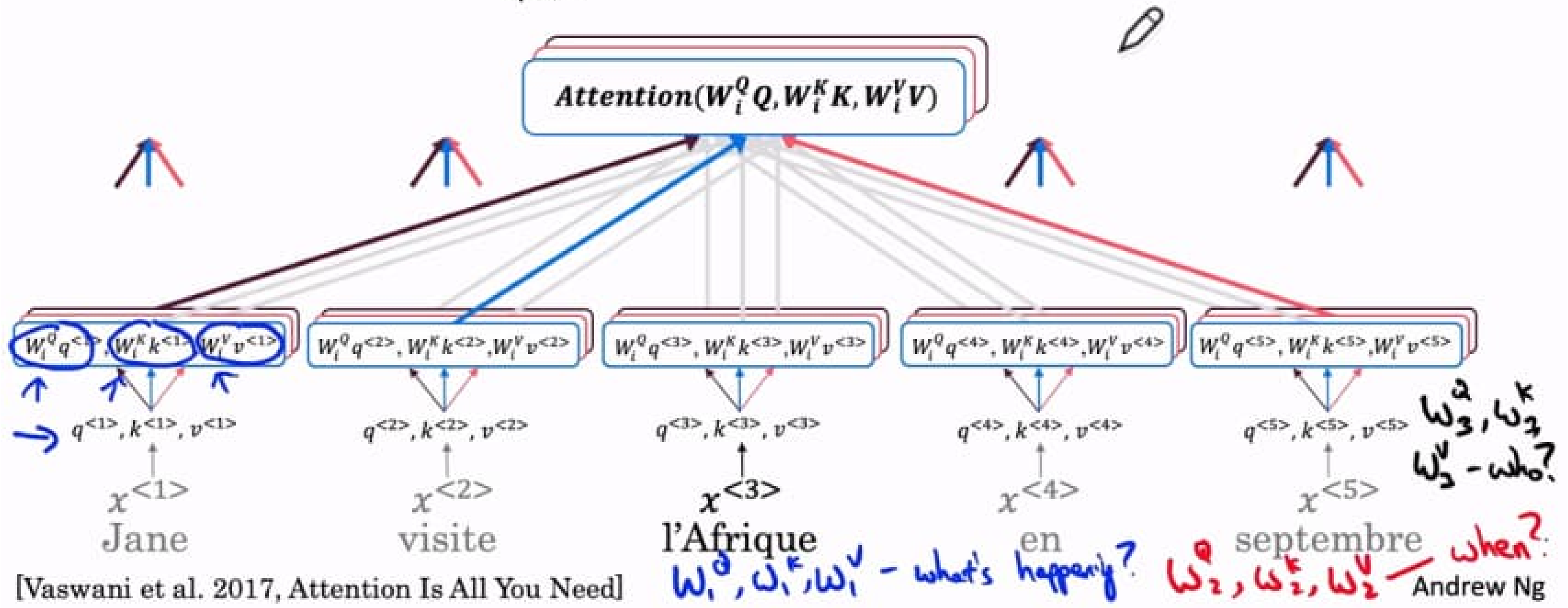
W_1^Q, W_1^K, W_1^V - what's happening?

Multi-Head Attention

"head"

$$MultiHead(Q, K, V)$$

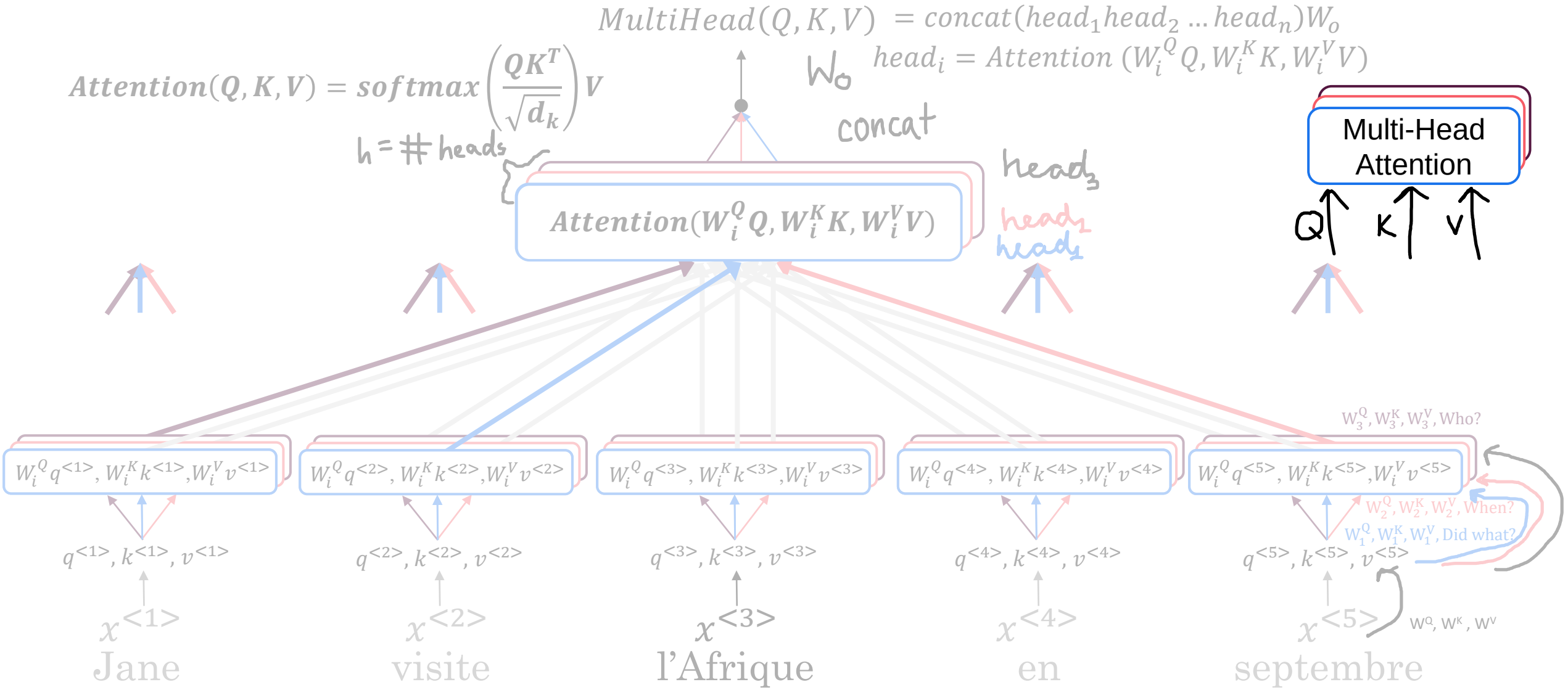
$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng

Multi-Head Attention

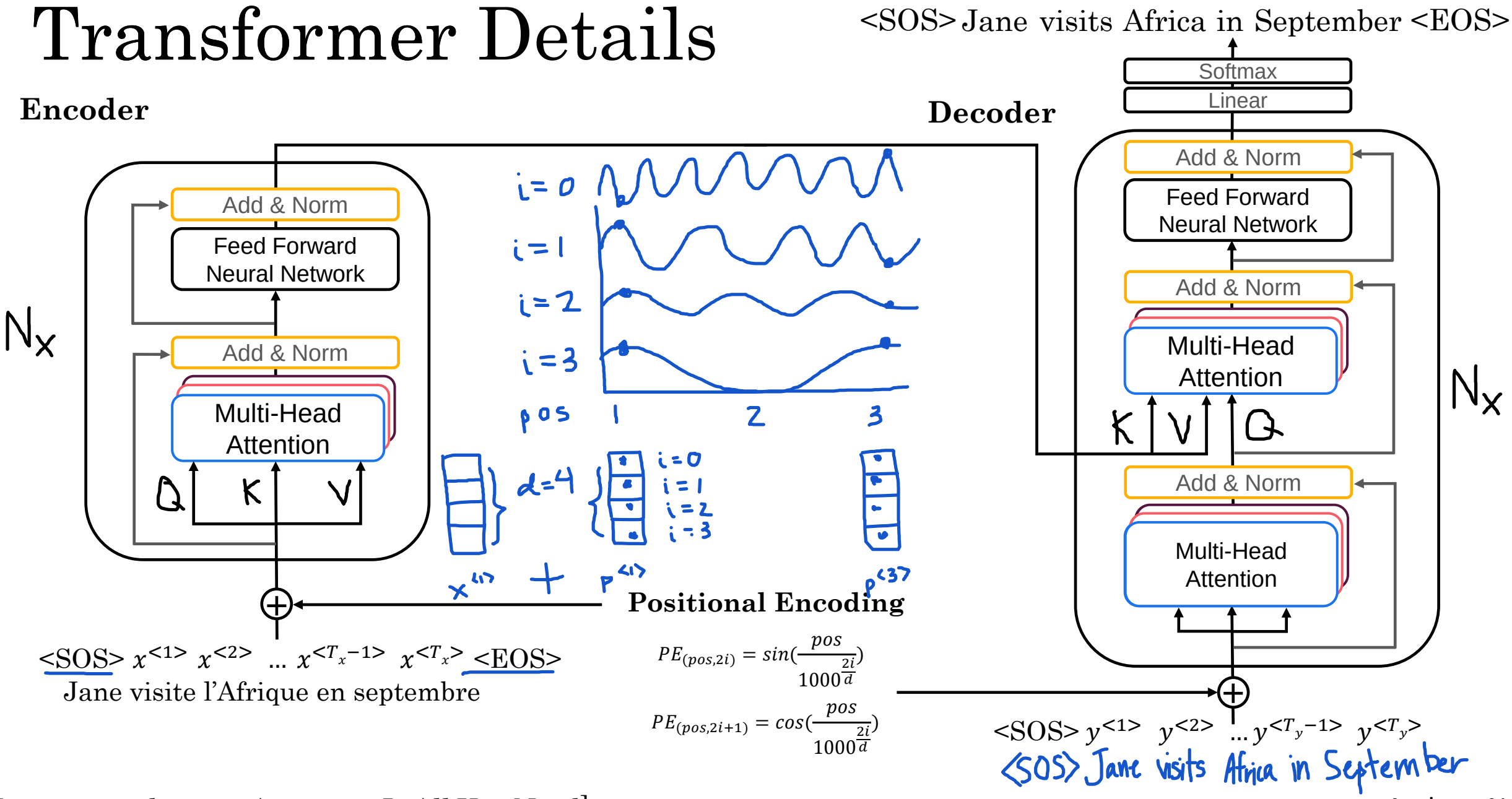


[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng

Transformers

Transformer Details



[Vaswani et al. 2017, Attention Is All You Need]

Andrew Ng