



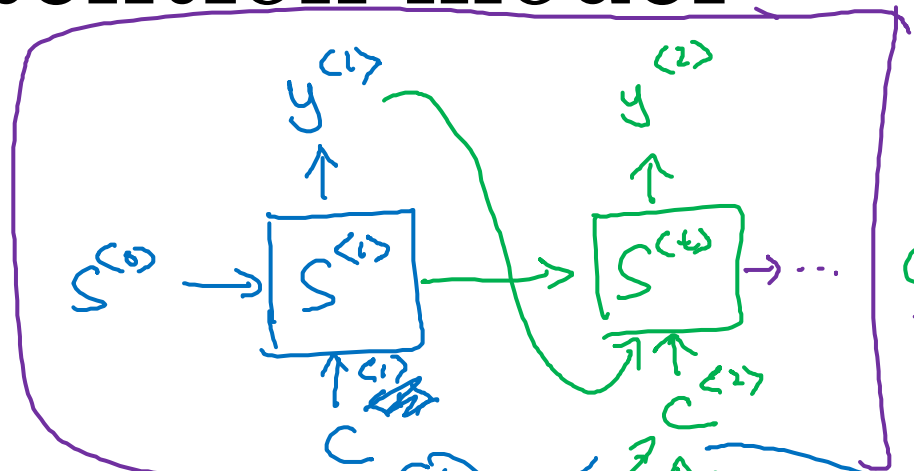
deeplearning.ai

Sequence to
sequence models

Attention model

Attention model

$\alpha^{(t,t')}$ = amount of "attention" $y^{(t)}$ should pay to $a^{(t')}$.

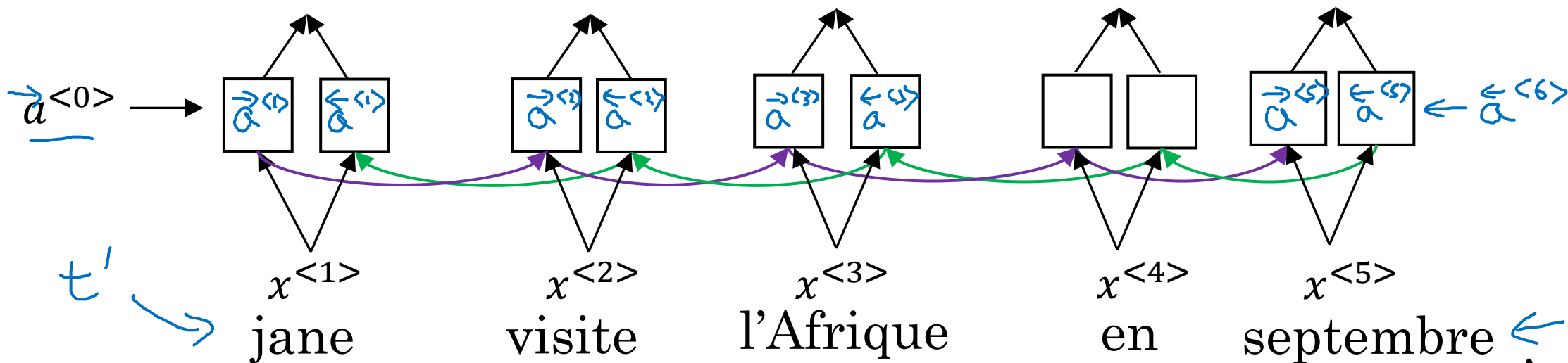


$$C^{(2)} = \sum_{t'} \alpha^{(2,t')} a^{(t')}$$

$$a^{(t')} = (\vec{a}^{(t')}, \overleftarrow{a}^{(t')})$$

$$\sum_{t'} \alpha^{(1,t')} = 1$$

$$C^{(1)} = \sum_{t'} \alpha^{(1,t')} a^{(t')}$$

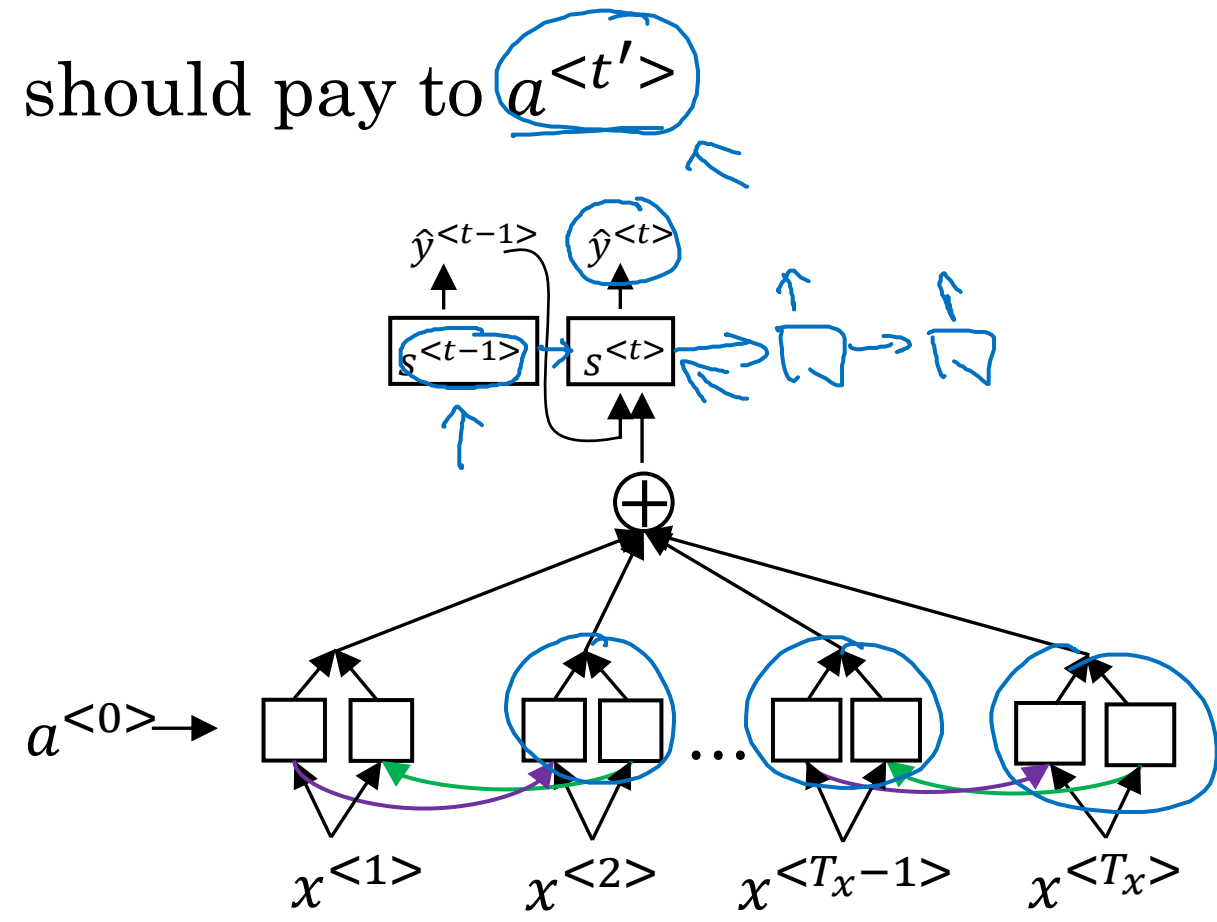
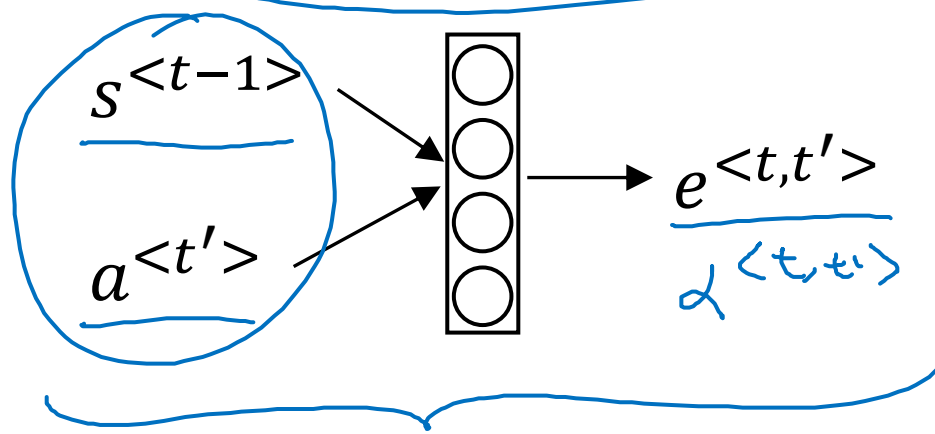


Computing attention $\alpha^{<t,t'>}$

T_x T_y

$\alpha^{<t,t'>} =$ amount of attention $y^{<t>}$ should pay to $a^{<t'>}$

$$\alpha^{<t,t'>} = \frac{\exp(e^{<t,t'>})}{\sum_{t'=1}^{T_x} \exp(e^{<t,t'>})}$$



[Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate]

[Xu et. al., 2015. Show, attend and tell: Neural image caption generation with visual attention]

Attention examples

July 20th 1969 → 1969 – 07 – 20

23 April, 1564 → 1564 – 04 – 23

Visualization of $\alpha^{<t,t'>}$:

