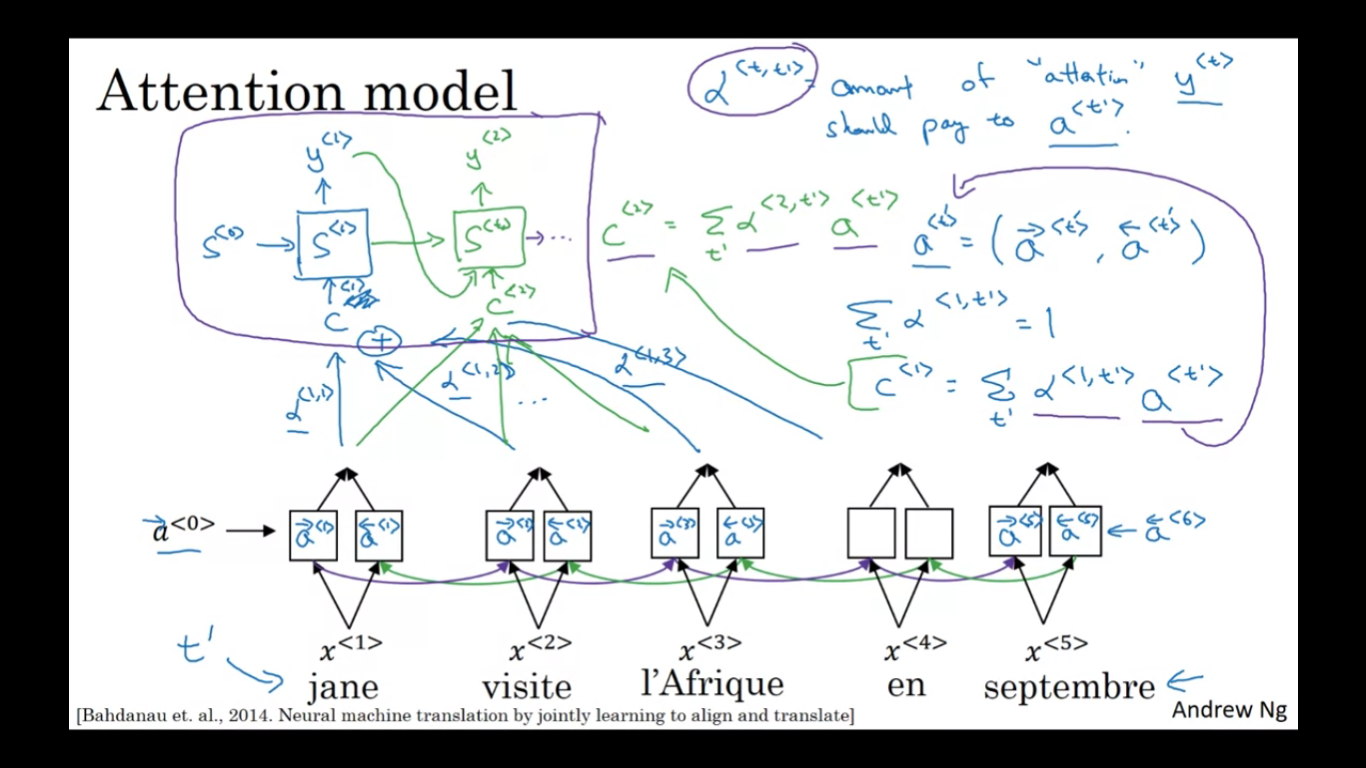
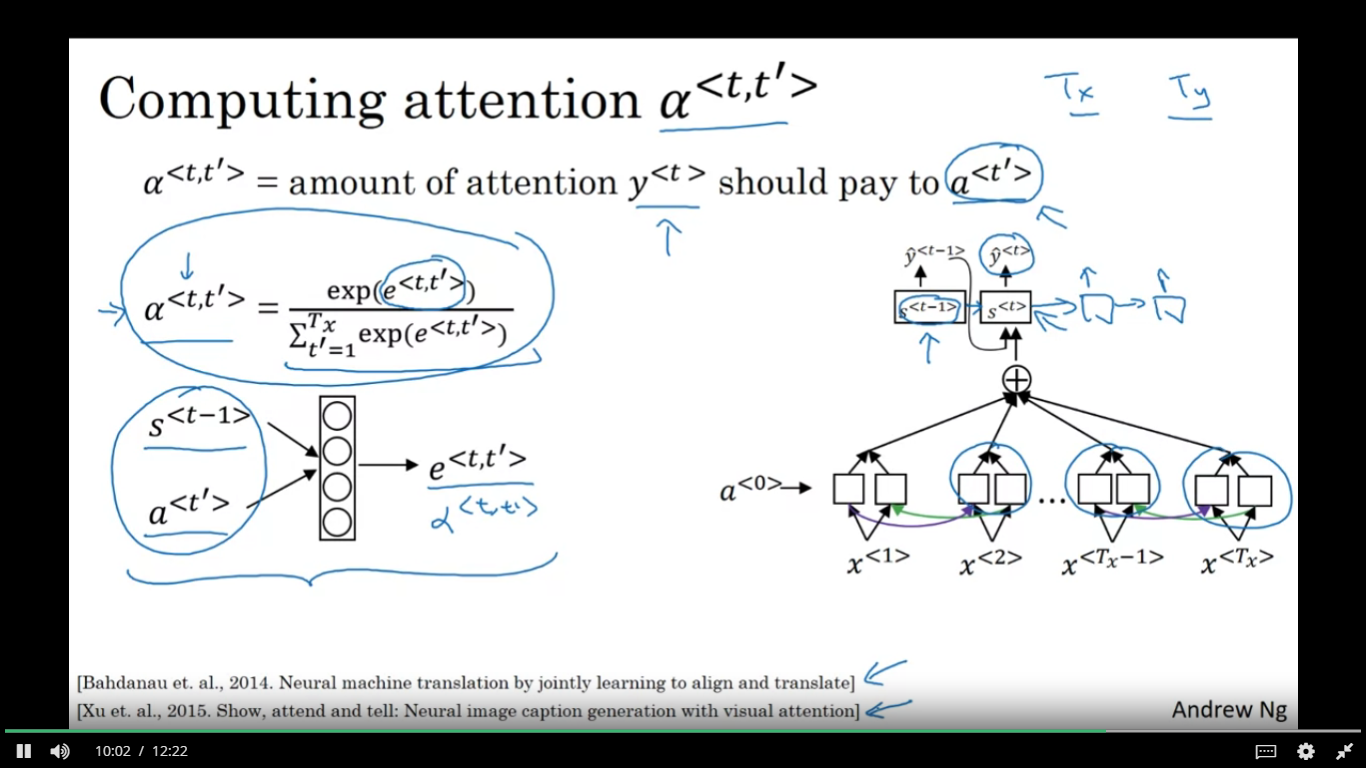
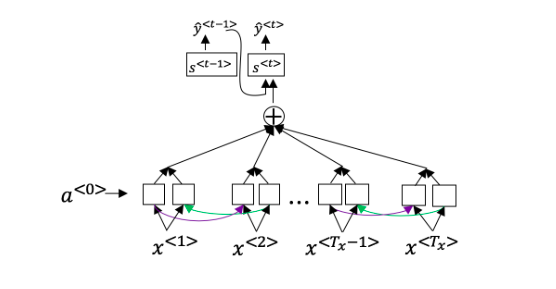
**Attention Model**

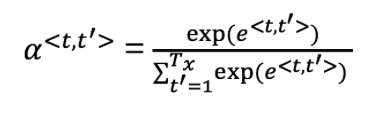




**6.Consider the attention model for machine translation.**



**Further, here is the formula for****α****<****t****,****t****′****>.**



**Which of the following statements about****α****<****t****,****t****′****> are true? Check all that apply.**

a)We expect α<t,t′> to be generally larger for values of a<t′> that are highly relevant to the value the network should output for y<t>. (Note the indices in the superscripts.)

b)We expect α<t,t′> to be generally larger for values of a<t> that are highly relevant to the value the network should output for y<t′>. (Note the indices in the superscripts.)

c)∑tα<t,t′>=1 (Note the summation is over t.)

d)∑t′α<t,t′>=1 (Note the summation is over t'.)

Answer: a, d

**7.The network learns where to “pay attention” by learning the values e<t,t′>, which are computed using a small neural network:**

**We can’t replace****s****<****t****−****1****> with****s****<****t****> as an input to this neural network. This is because s^{<t>} depends on****α****<****t****,****t****′****> which in turn depends on****e****<****t****,****t****′****>; so at the time we need to evaluate this network, we haven’t computed****s****<****t****> yet.**

a)True

b)False

Answer: a