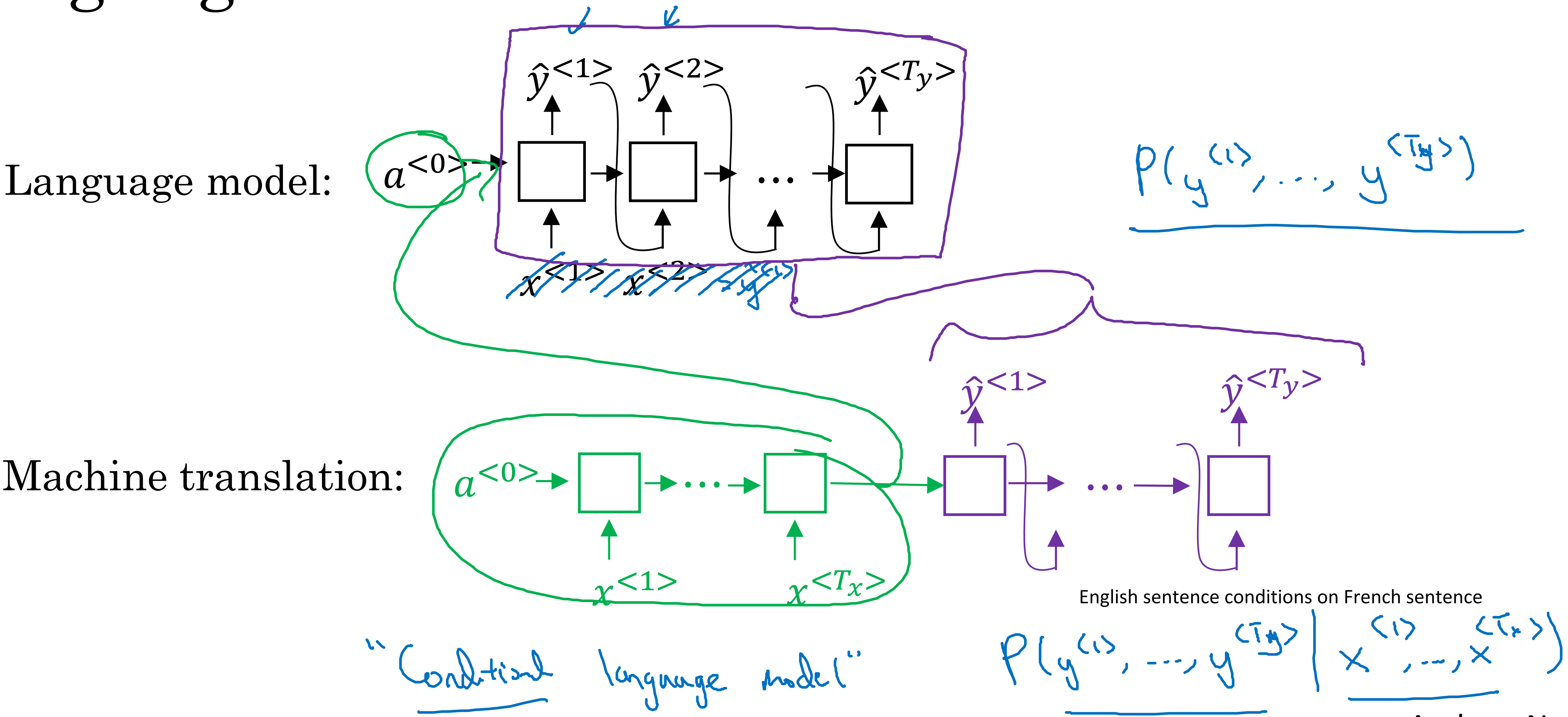
deeplearning.ai

Sequence to sequence models

Picking the most likely sentence

Machine translation as building a conditional language model



English sentence conditions on French sentence

Andrew Ng

Finding the most likely translation

Jane visite l'Afrique en septembre.

$$\underline{P(\overbrace{y^{<1>},\ldots,y^{<T_y>}}^{\text{English}} \mid \overset{\overset{\text{French}}{\downarrow}}{x})}$$

→ Jane is visiting Africa in September.

→ Jane is going to be visiting Africa in September.

→ In September, Jane will visit Africa.

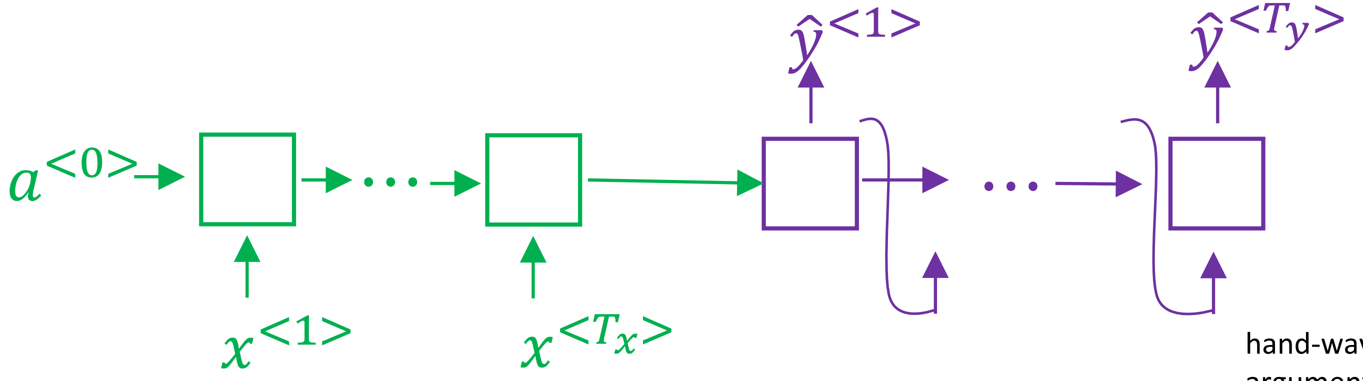
→ Her African friend welcomed Jane in September.

$$\underset{y^{<1>},\ldots,y^{<T_y>}}{\arg\max}\; P(\underline{y^{<1>},\ldots,y^{<T_y>}} \mid x)$$

maximum this term
Bean search

Why not a greedy search?

$P(\hat{y}^{<1>} | x)$



And what an approximate search algorithm does is it will try, it won't always succeed, but it will try to pick the sentence y that maximizes that conditional probability. And, even though it's not guaranteed to find the value of y that maximizes this, it usually does a good enough job.

$$\arg\max_{y} P(\underbrace{\hat{y}^{<1>}, \hat{y}^{<2>},}\ \ldots, \hat{y}^{<T_y>} | x)$$

hand-wavey argument, but this is an example of a broader phenomenon

$10{,}000$

10

$10{,}000^{10}$

$P(y|x)$

So, this is just a huge space of possible sentence and it's impossible to rate them all, which is why the most common thing to do is use an approximate search algorithm

if you want to find the sequence of words, y1, y2 all the way up to the final word that together maximize the probability, it's not always optimal to just pick one word at a time.

→ Jane is visiting Africa in September.

→ Jane is going to be visiting Africa in September.

$P(\text{Jane is going} | x) > P(\text{Jane is visiting} | x)$

Andrew Ng