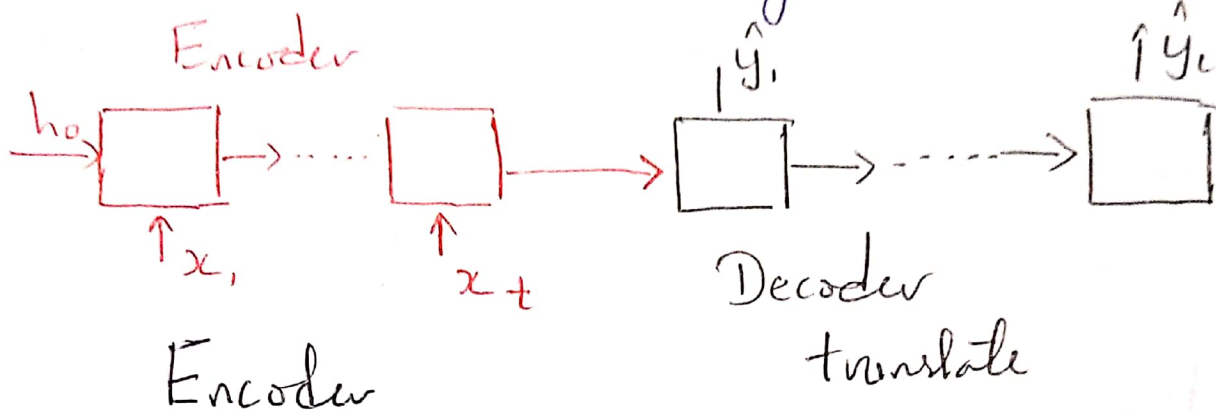


# Attention Model Intuition Page 1

Translate from French to English



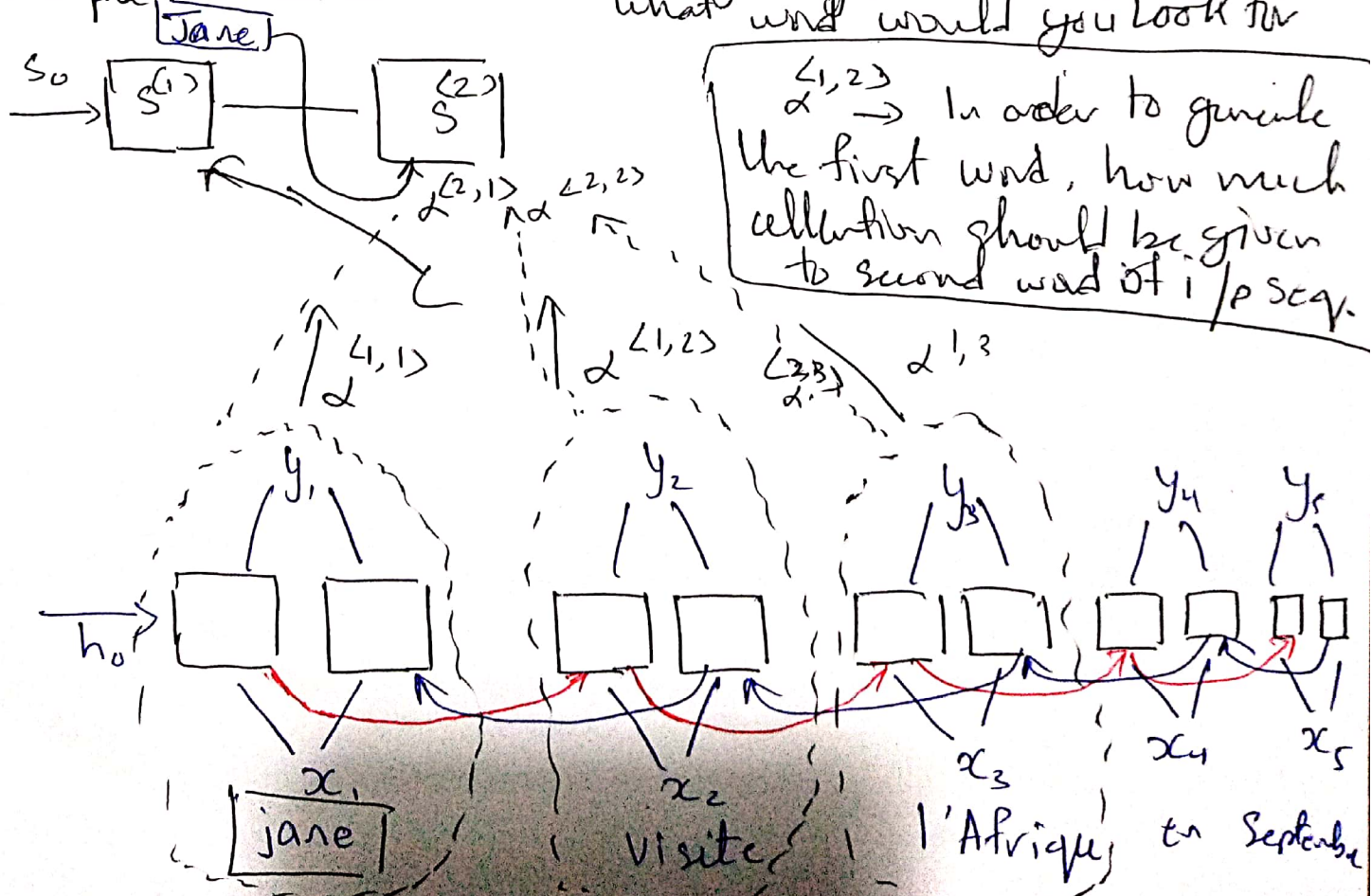
Read  $\rightarrow$  memorize  $\rightarrow$  store

Attention Model:  
gives attention weights

Humans: do not store  
translation is part by part  
b/c difficult to memorize.

Sample Sentence:

When you generate translation  
what word would you look for



$\alpha^{2,1} \rightarrow$  In order to generate second word  
how much attention should be given  
to the second word of input

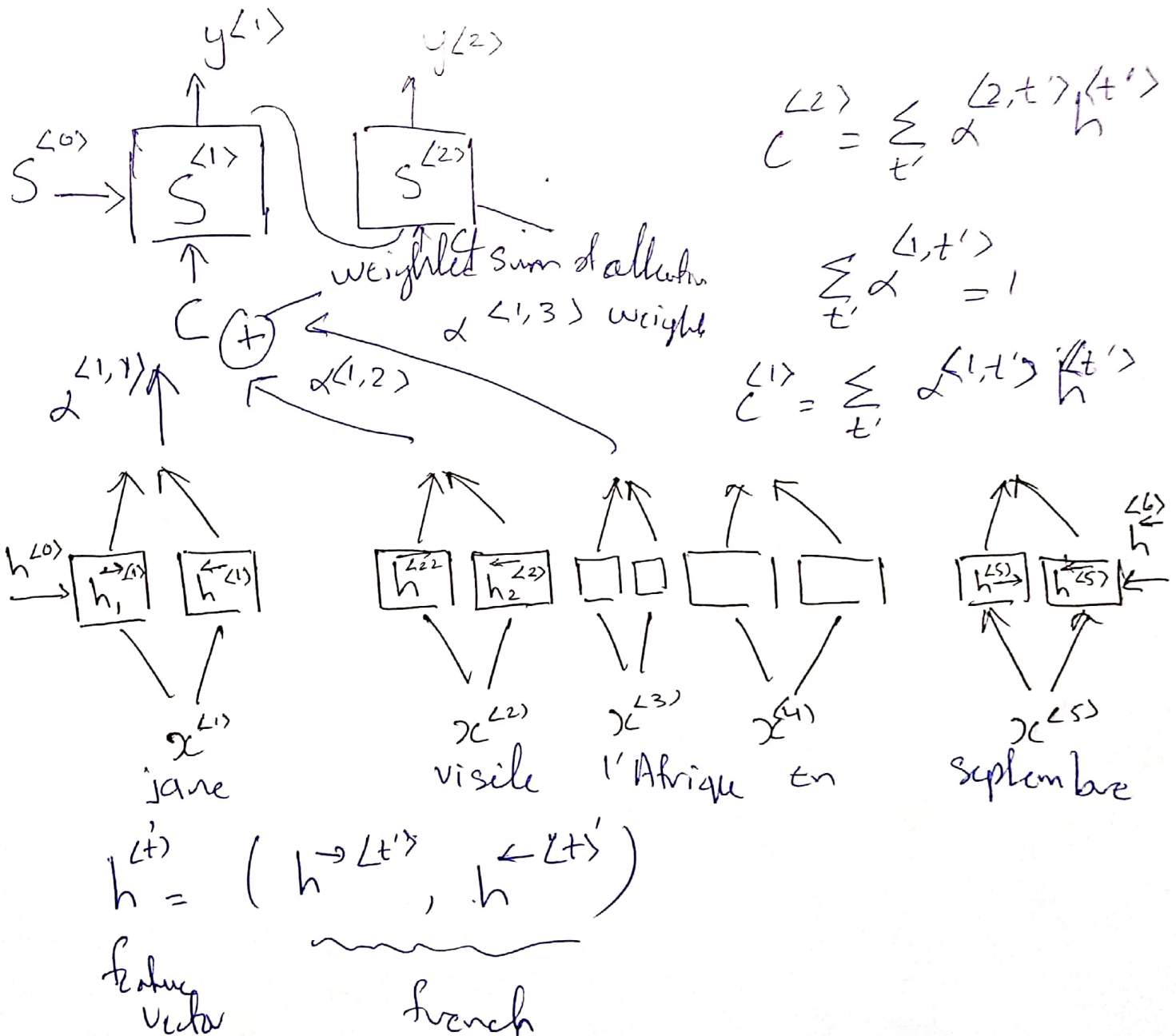
How to compute weights  $\alpha^{(t,t')}$   
Context depends upon  $\alpha^{(t,t')}$

In general,

$$\alpha^{(3,t)} = \left\{ \begin{array}{c} s^{(2)} \\ \left[ \begin{array}{cc} \rightarrow & \leftarrow \\ h_t & , h_t \end{array} \right] \end{array} \right\}$$

Given Input Sequence

Use bi-directional RNN/GRU/LSTMs to compute features



$\alpha^{L,t,t'}$  = amount of attention  $y^{L,t}$  should pay to  $h^{L,t'}$

Context vector  $c, c, \dots$

How to compute  $\alpha^{L,t,t'}$ , amount of attention  $y^{L,t}$  should pay to  $h^{L,t'}$ .

$$\alpha^{L,t,t'} = \frac{\exp(c^{L,t,t'})}{\sum_{t'=1}^{T_h} \exp^{L,t,t'}}$$

Compute a NN

|