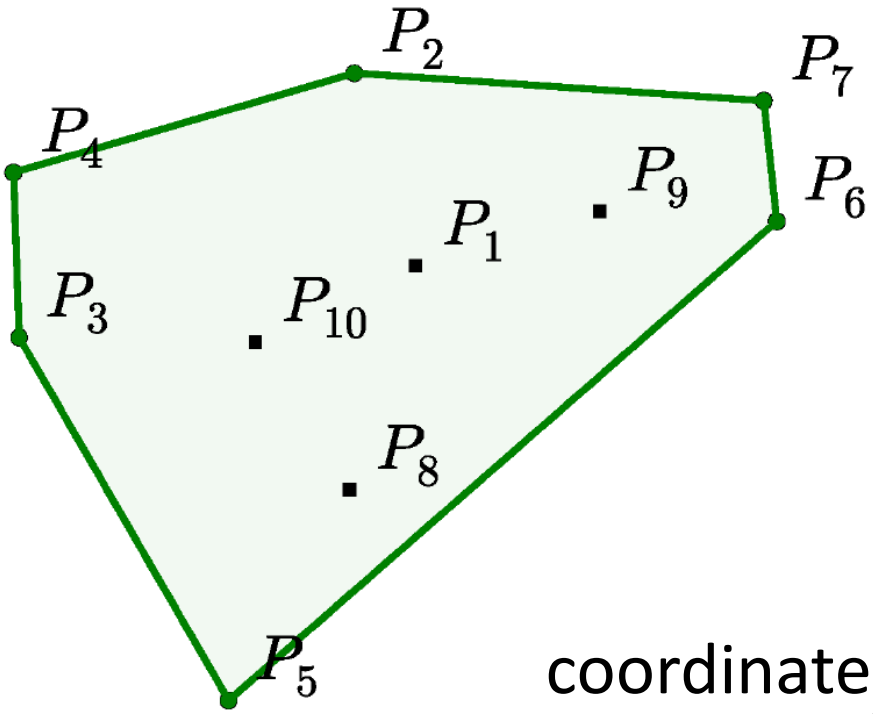
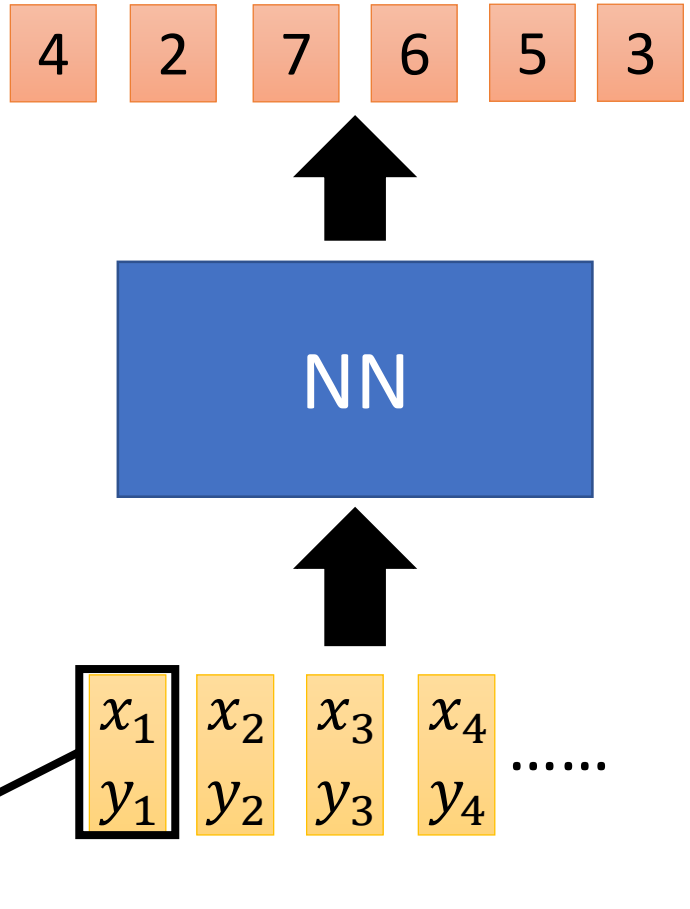
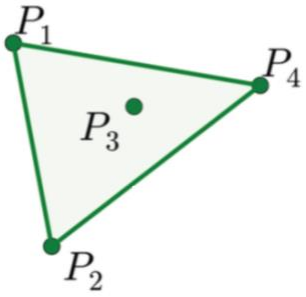
Pointer Network

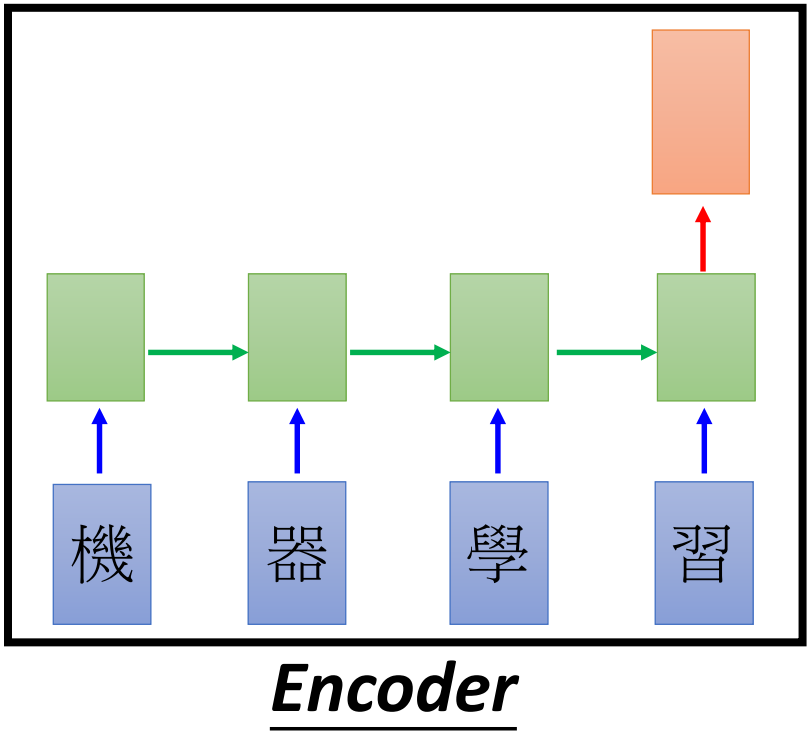
Pointer Network

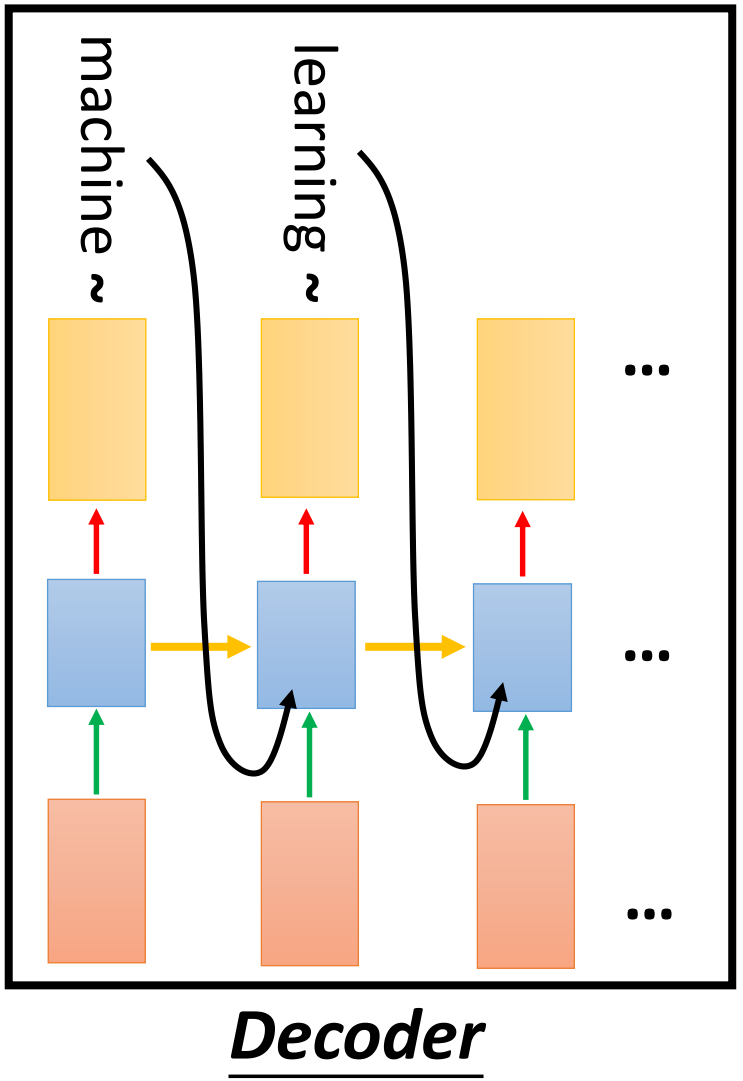




Sequence-to-sequence?

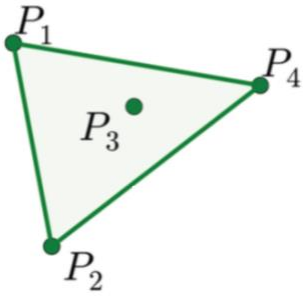




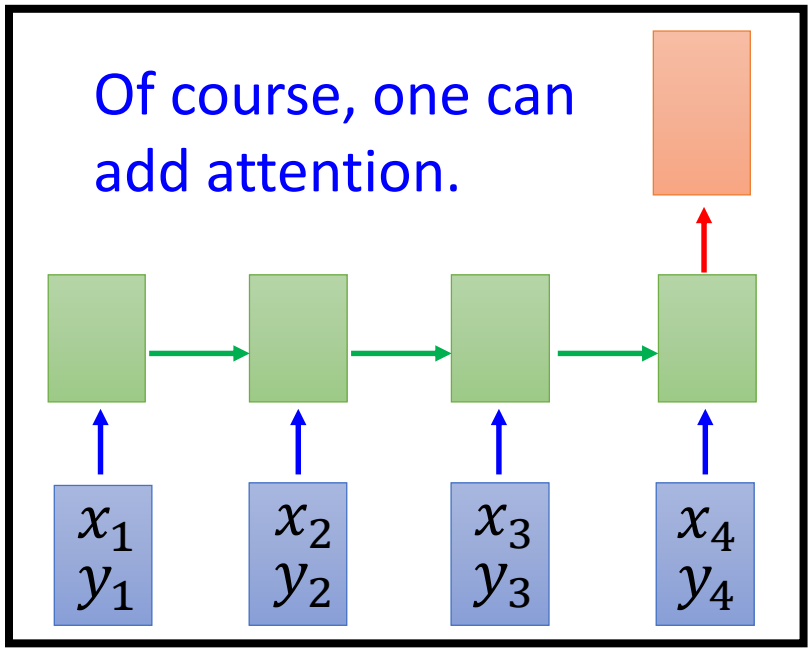


Problem?

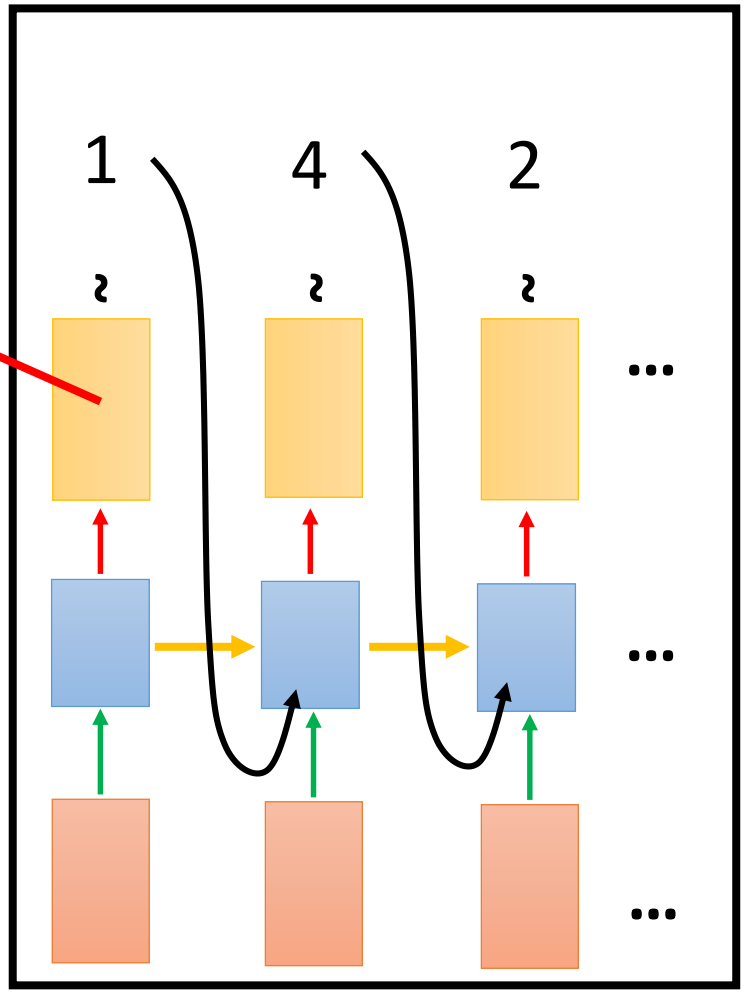
Sequence-to-sequence?



{1, 2, 3, 4, END}



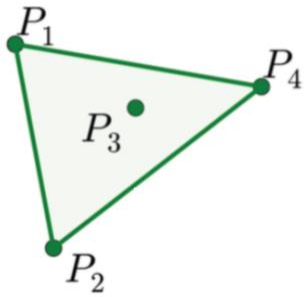
Encoder



Decoder

Pointer Network

x_0
y_0 : END



Attention Weight

key

0.5

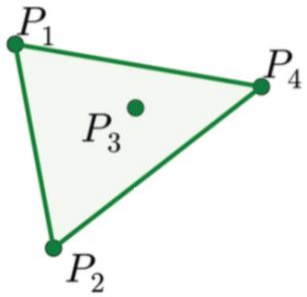
z^0

h^0 h^1 h^2 h^3 h^4

x_0 y_0 | x_1 y_1 | x_2 y_2 | x_3 y_3 | x_4 y_4

Pointer Network

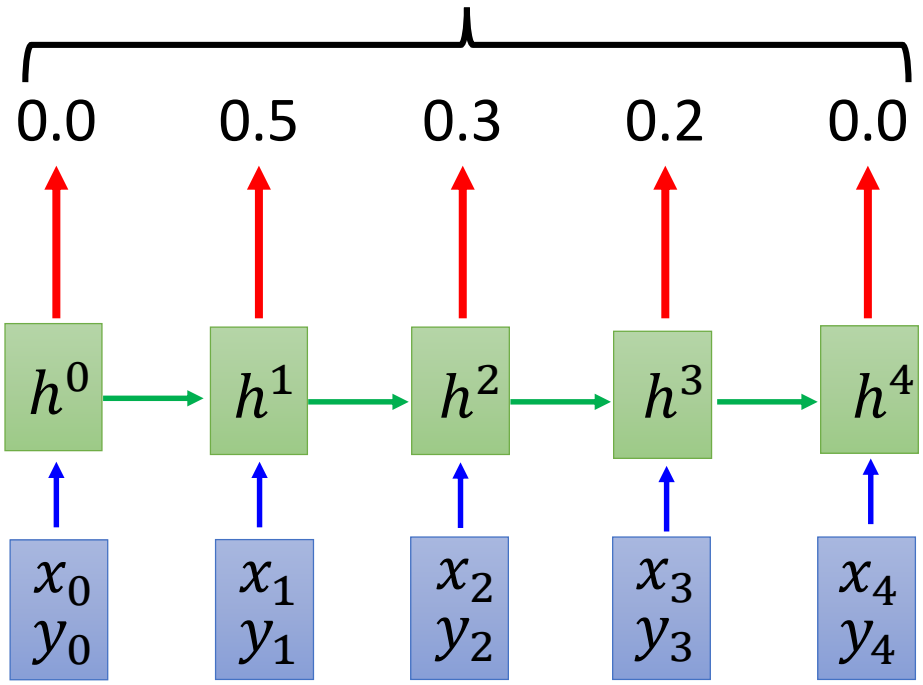
x_0 y_0 : END



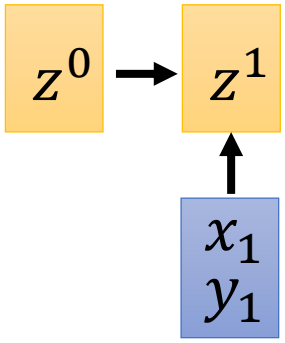
Output: 1

What decoder can output depends on the input.

argmax from this distribution

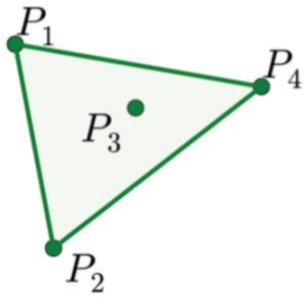


key



Pointer Network

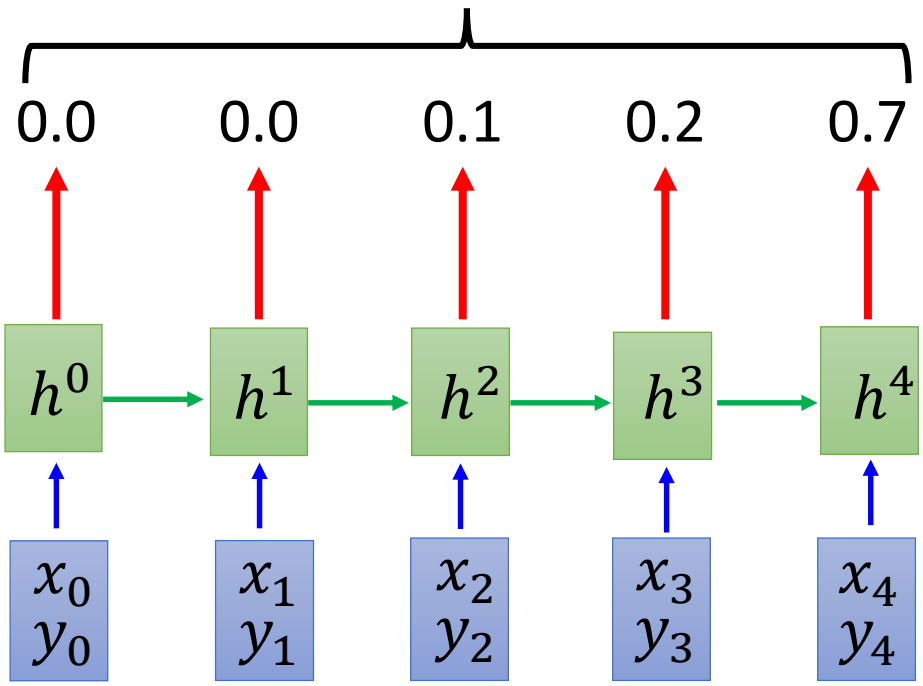
x_0 y_0 : END



Output: 4

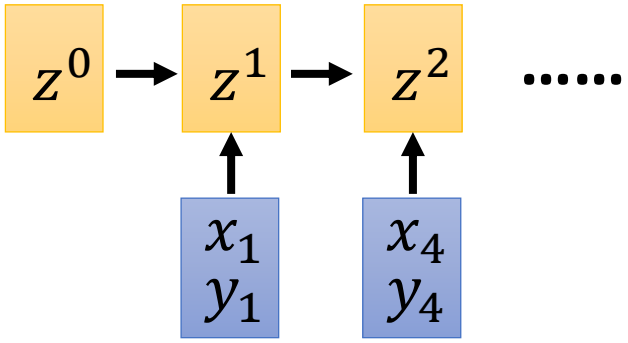
≈

argmax from this distribution



What decoder can output depends on the input.

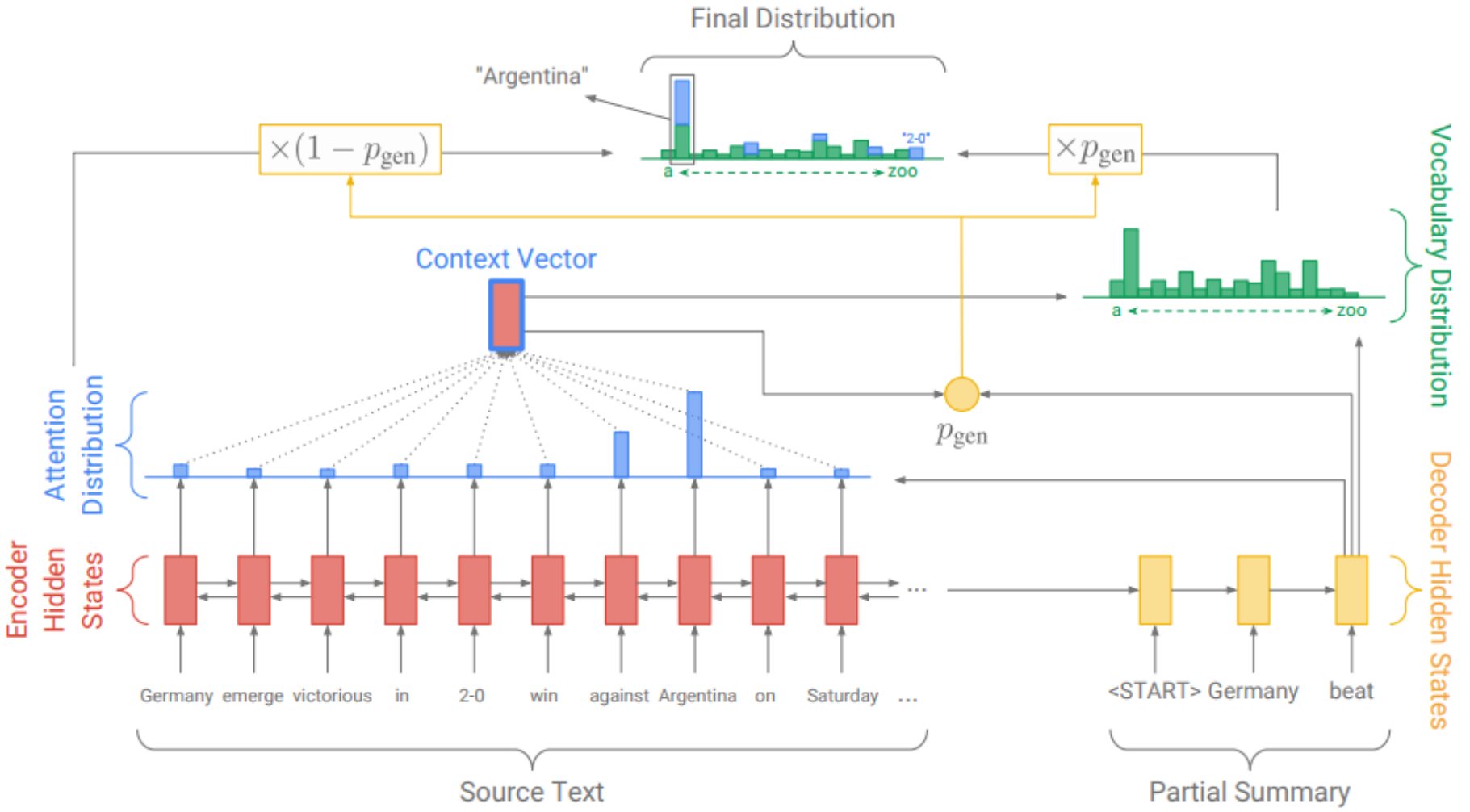
key



The process stops when "END" has the largest attention weights.

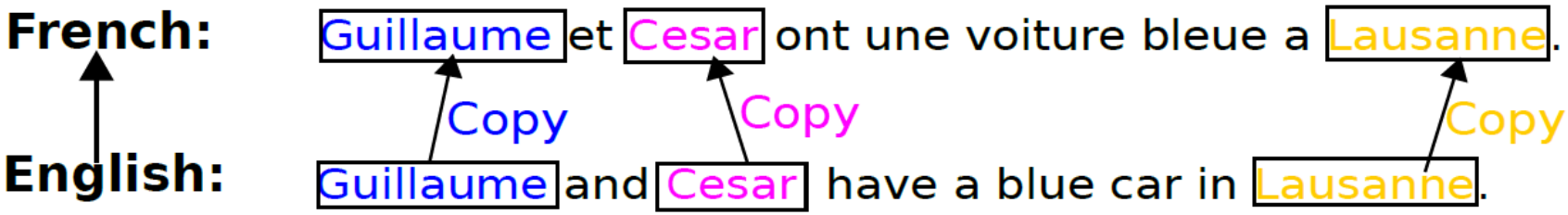
https://arxiv.org/abs/1704.04368

Applications - Summarization



More Applications

Machine Translation



Chat-bot

User: X寶你好，我是庫洛洛

Machine:庫洛洛你好，很高興認識你