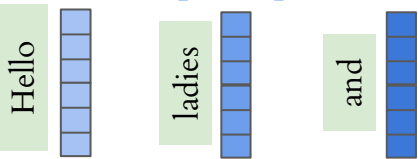


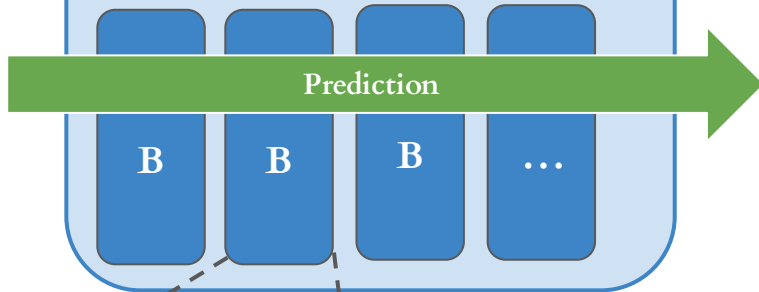
Input

“Hello ladies and ____”

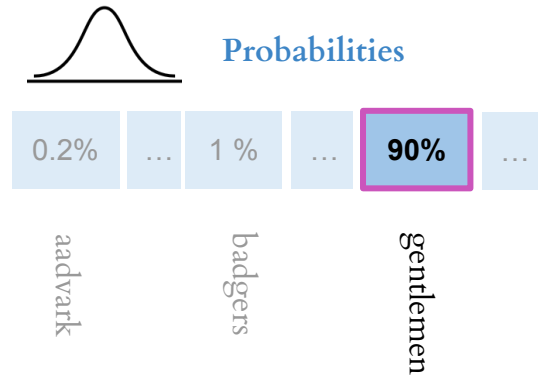
Learned Input Representations



Transformer LM



Probabilities



Corresponding Vocabulary Words

Low-level Information

Transformer Block

Contextual Information

