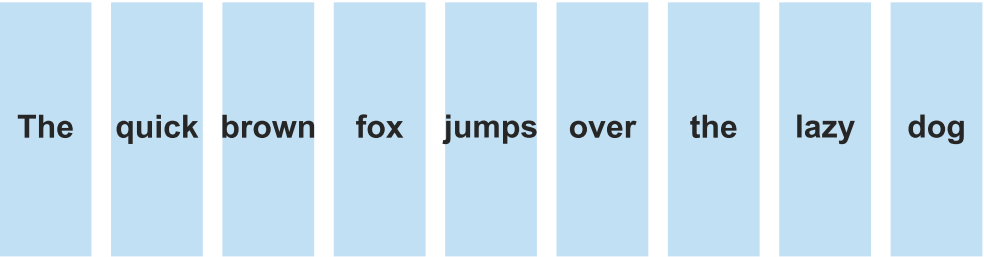


# Masked Language Modeling: How BERT Learns

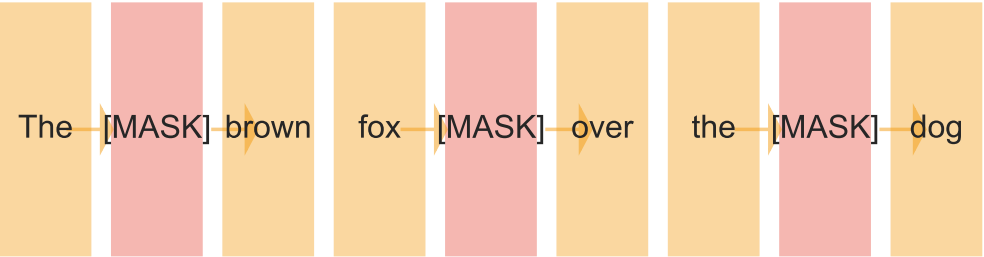
Step 1: Original Text



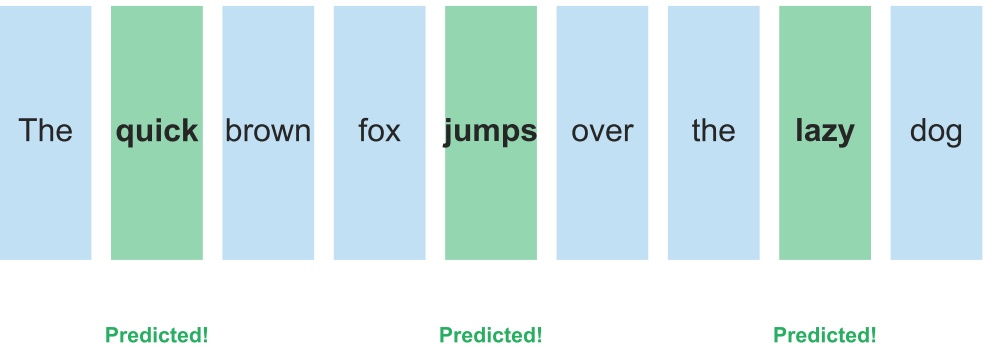
Step 2: Random Masking (15%)



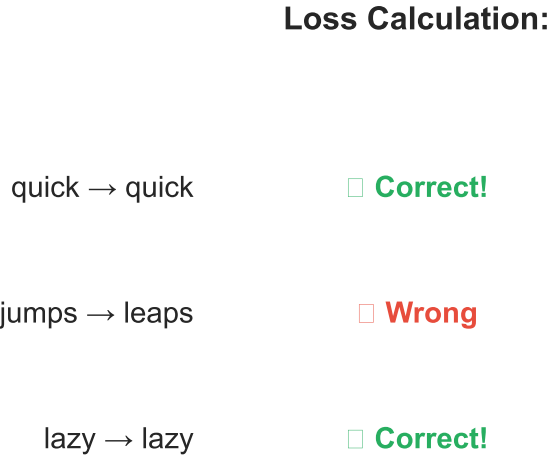
Step 3: Bidirectional Context



Step 4: Prediction








Step 5: Learning from Errors



*Update weights to improve predictions*

What BERT Learns

-  Grammar patterns
-  Word relationships
-  Context understanding
-  Semantic meanings
-  World knowledge