

# Masked Language Modeling: How BERT Learns

Step 1: Original Text

The quick brown fox jumps over the lazy dog

Step 2: Random Masking (15%)

The [MASK] brown fox [MASK] over the [MASK] dog

Step 3: Bidirectional Context

The-[MASK]-brown fox-[MASK]-over the-[MASK]-dog

Step 4: Prediction

The quick brown fox jumps over the lazy dog

Predicted!

Predicted!

Predicted!

Step 5: Learning from Errors

## Loss Calculation:

quick → quick

Correct!

jumps → leaps

Wrong

lazy → lazy

Correct!

What BERT Learns



Grammar patterns



Word relationships



Context understanding



Semantic meanings



World knowledge

Update weights to improve predictions