Token Flow Pipeline

Input Text

"The cat sat"

Tokenization

["The", "cat", "sat"]

Token IDs

[546, 1828, 2062]

Embeddings

[[0.2, -0.1, ...],
[0.5, 0.3, ...],
[0.1, 0.7, ...]]

Model Output

Summary tokens

Subword
splitting

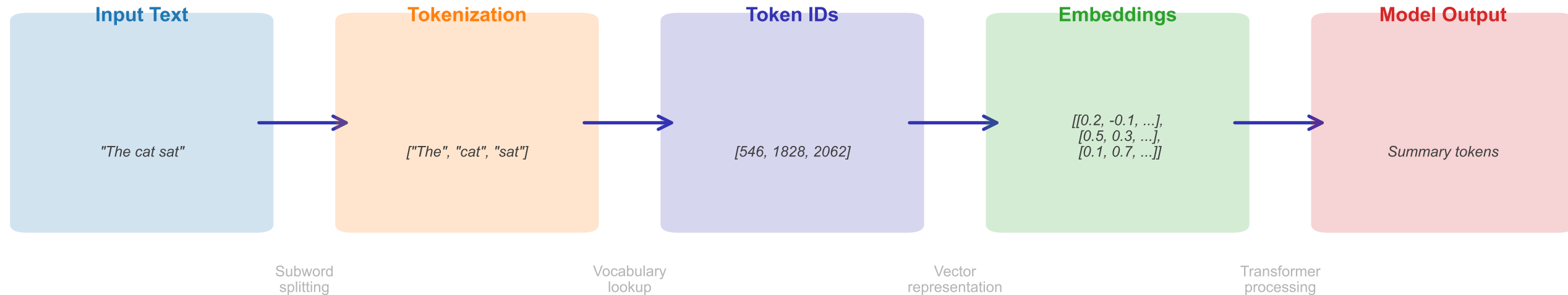Vocabulary
lookup

Vector
representation

Transformer
processing

Each token limited to model vocabulary (typically 30K-50K tokens)