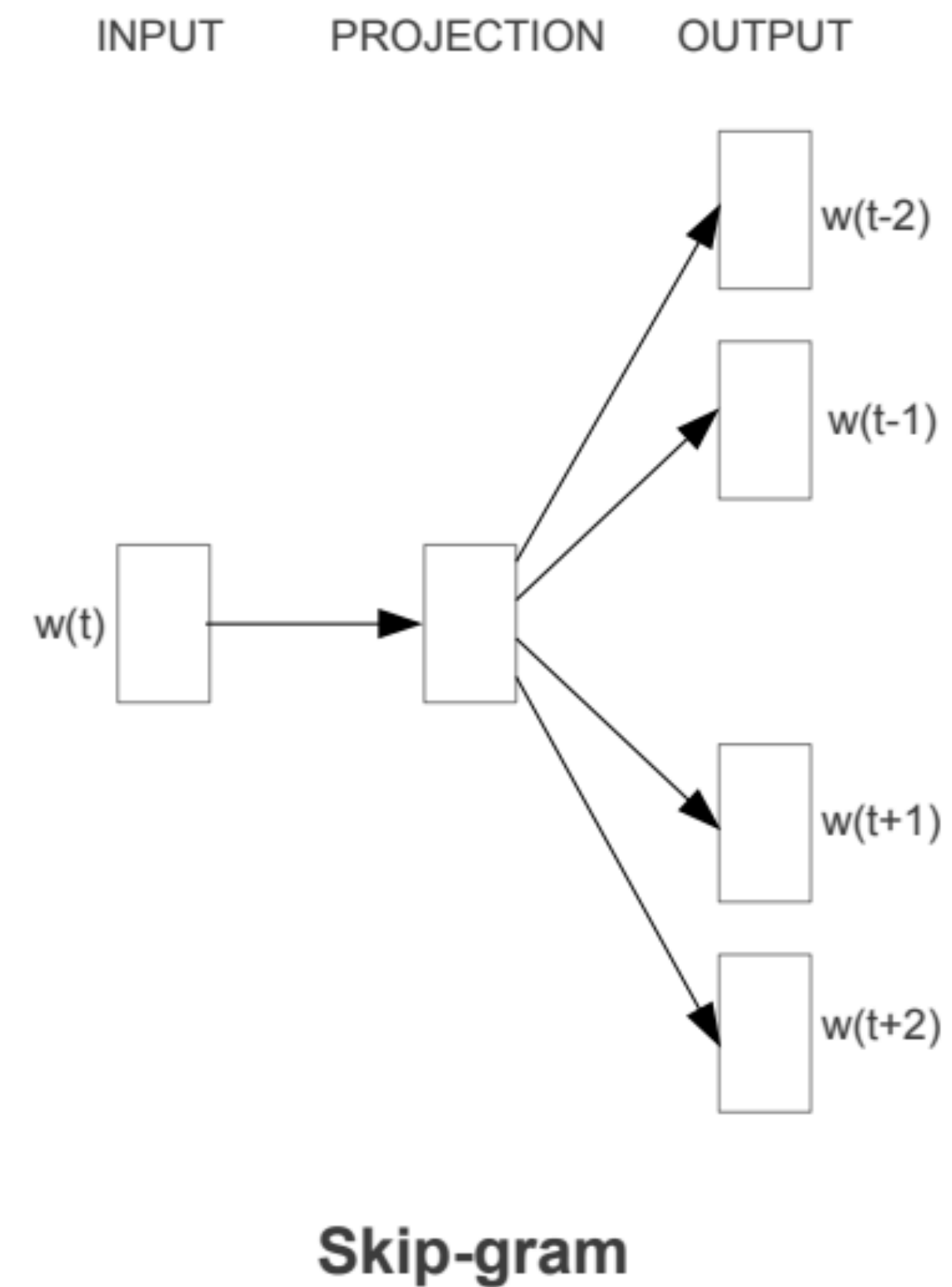
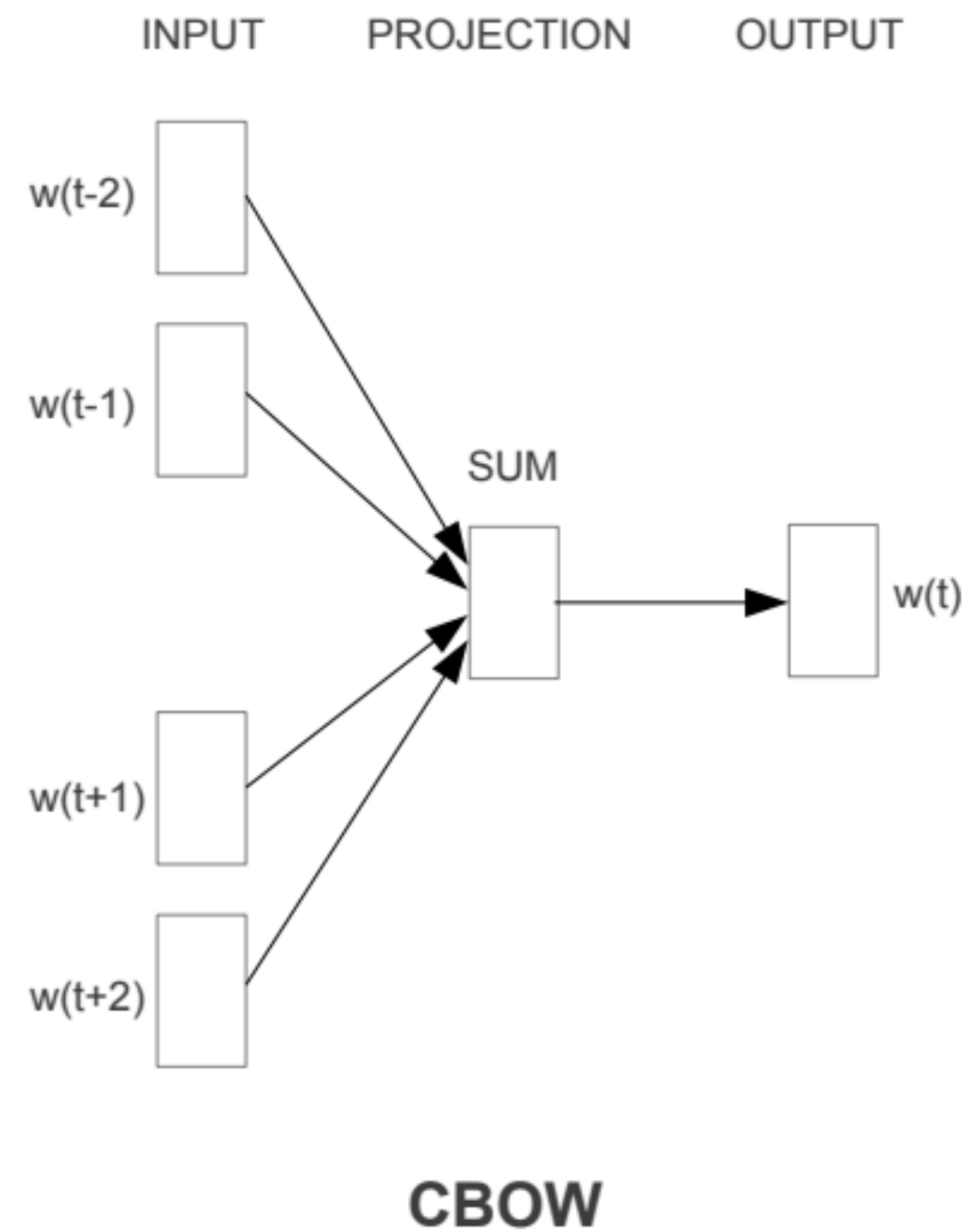
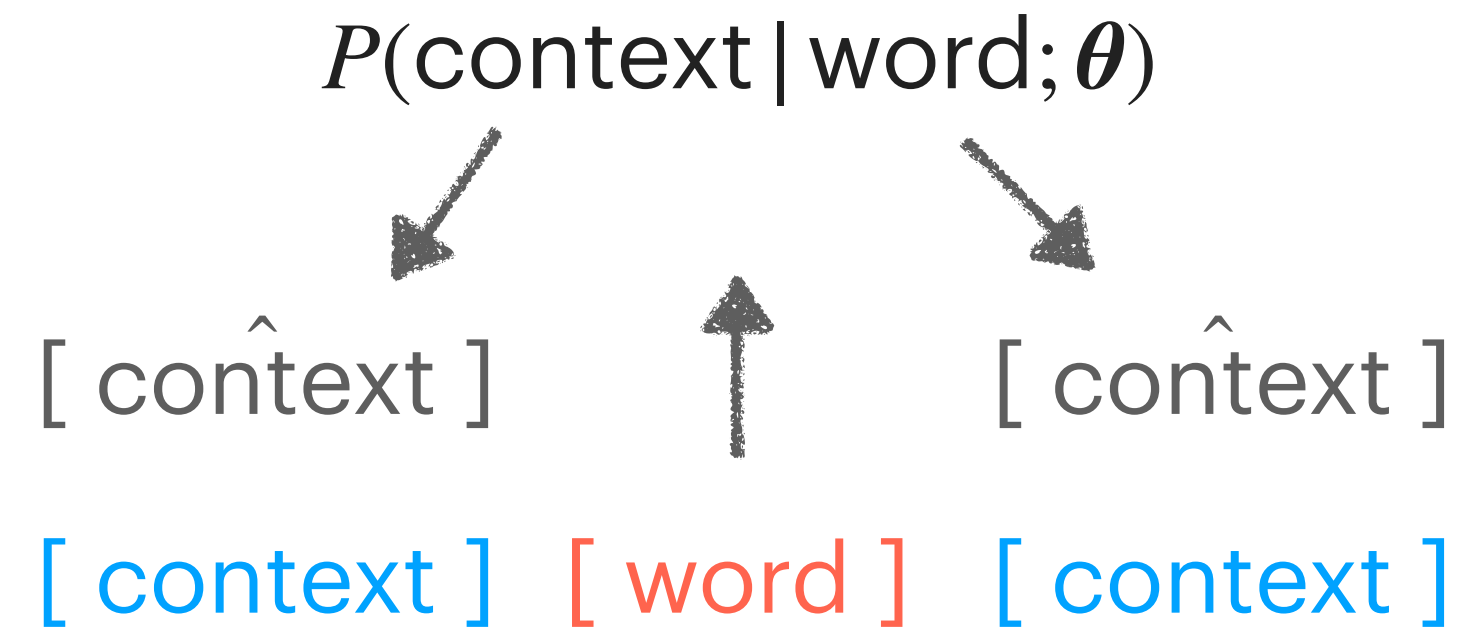


# Continuous Bag of Words (CBOW) & Skip Gram modeling



- Taken from Mikolov et al., *Efficient Estimation of Word Representations in Vector Space* (2013)

# The skip gram model



Example: the quick brown fox jumped over the lazy dog

$\mathbf{x}_{t-c} \quad \dots \quad \mathbf{x}_t \quad \dots \quad \mathbf{x}_{t+c}$

Learning problem: 
$$\hat{\theta} = \arg \max_{\theta} \prod_{t=1}^T \prod_{-c \leq j \leq c, j \neq 0} \mathbf{x}_{t+j} \cdot p(\mathbf{x}_{t+j} | \mathbf{x}_t; \theta)$$
$$= \arg \min_{\theta} - \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \mathbf{x}_{t+j} \cdot \log p(\mathbf{x}_{t+j} | \mathbf{x}_t; \theta)$$

where

- $T$  = corpus length
- $c$  = context window
- $N$  = vocabulary size
- $\mathbf{x} \in \{0,1\}^N$
- $\theta$  = parameterization