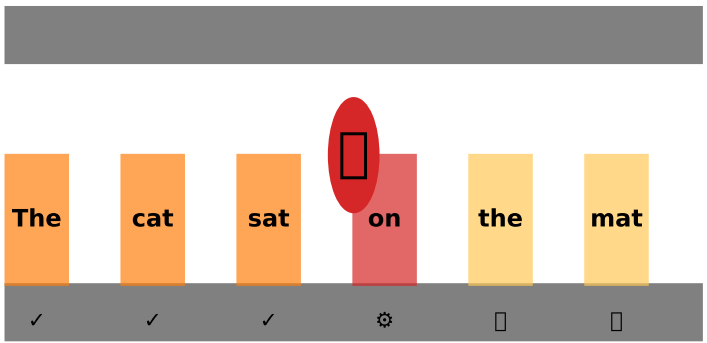
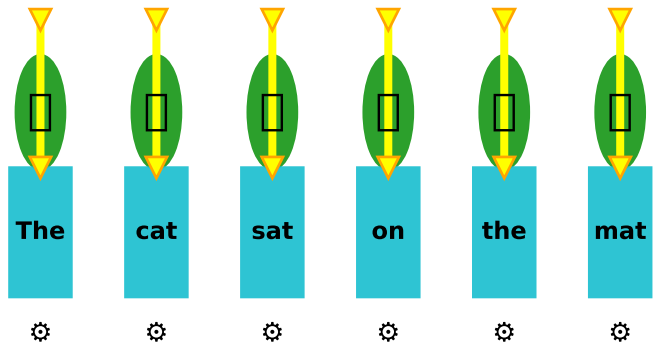
RNN: Single Worker Assembly Line

1 token processed at a time



Transformer: Parallel Processing Factory

All tokens processed simultaneously!