Decoder Processing Flow

① Context Vector (from Encoder)

Context **c**

Initializes decoder state: s_0 = c

↓

② Initial Input Token

⟨START⟩ Special token to begin generation

↓

③ RNN Processing (Autoregressive)

s_1 t=1	s_2 t=2	s_3 t=3	s_4 t=4

↓

④ Generated Output Tokens

나는 y_1	AI를 y_2	사랑해 y_3	⟨EOS⟩ y_4

Key Components

- **Initial State:** s_0 = context vector c
- **RNN:** LSTM or GRU for generation
- **Softmax:** Predicts next token

Computation

- Each step takes previous output as input
- Updates hidden state and predicts next token

```
s_t = f(y_{t-1}, s_{t-1})
y_t = softmax(W·s_t)
```

Generation Process

- **Autoregressive:** Uses own predictions
- Stops when ⟨EOS⟩ token is generated
- Can use greedy or beam search decoding

⚠ Key Difference

Decoder uses its own previous predictions as input (autoregressive), unlike encoder which

processes fixed input.